WILEY

ESD in Silicon Integrated Circuits

Second Edition

AJITH AMERASEKERA | CHARVAKA DUVVURY

ESD in Silicon Integrated Circuits

Second Edition

ESD in Silicon Integrated Circuits

Second Edition

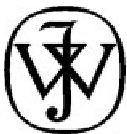
Ajith Amerasekera
Charvaka Duvvury
Texas Instruments, Inc., USA

With

Warren Anderson
Compaq Computer Corporation, USA

Horst Gieser
Fraunhofer Institute for Reliability and Microintegration IZM ATIS, Germany

Sridhar Ramaswamy
Texas Instruments, Inc., USA



JOHN WILEY & SONS, LTD

Copyright © 2002 by John Wiley & Sons, Ltd.,
Baffins Lane, Chichester,
West Sussex PO 19 1UD, England

Phone (+44) 1243 779777

e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk

Visit our Home Page on http://www.wileyeurope.com or http://www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a license issued by the Copyright Licensing Agency Ltd., 90 Tottenham Court Road, London W1P 0LP, UK without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons, Ltd., Baffins Lane, Chichester, West Sussex PO 19 1UD, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770571.

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Jossey-Bass, 989 Market Street,
San Francisco, CA 94103-1741, USA

WILEY-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

John Wiley & Sons Australia, Ltd., 33 Park Road,
Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd., 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada, Ltd., 22 Worcester Road,
Etobicoke, Ontario, Canada M9W 1L1

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0 470 49871 8

Typeset in 10/12pt Times by Laserwords Private Limited, Chennai, India.
Printed and bound in Great Britain by Antony Rowe Ltd., Chippenham, Wiltshire.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

Contents

Preface ix

1 Introduction 1

1.1 Background 1
1.2 The ESD Problem 3
1.3 Protecting Against ESD 4
1.4 Outline of the Book 4
Bibliography 7

2 ESD Phenomenon 8

2.1 Introduction 8
2.2 Electrostatic Voltage 9
2.3 Discharge 11
2.4 ESD Stress Models 12
Bibliography 15

3 Test Methods 17

3.1 Introduction 17
3.2 Human Body Model (HBM) 18
3.3 Machine Model (MM) 27
3.4 Charged Device Model (CDM) 28
3.5 Socket Device Model (SDM) 40
3.6 Metrology, Calibration, Verification 42
3.7 Transmission Line Pulsing (TLP) 47
3.8 Failure Criteria 58
3.9 Summary 60
Bibliography 61

4 Physics and Operation of ESD Protection Circuit Elements 68

4.1 Introduction 68
4.2 Resistors 68
4.3 Diodes 70

4.4 Transistor Operation 77
4.5 Transistor Operation under ESD Conditions 85
4.6 Electrothermal Effects 95
4.7 SCR Operation 98
4.8 Conclusion 101
Bibliography 102

5 ESD Protection Circuit Design Concepts and Strategy 105
5.1 The Qualities of Good ESD Protection 106
5.2 ESD Protection Design Methods 109
5.3 Selecting an ESD Strategy 123
5.4 Summary 124
Bibliography 124

6 Design and Layout Requirements 126
6.1 Introduction 126
6.2 Thick Field Device 127
6.3 nMOS Transistors (FPDs) 132
6.4 Gate-Coupled nMOS (GCNMOS) 138
6.5 Gate Driven nMOS (GDNMOS) 149
6.6 SCR Protection Device 150
6.7 ESD Protection Design Synthesis 155
6.8 Total Input Protection 164
6.9 ESD Protection Using Diode-Based Devices 172
6.10 Power Supply Clamps 176
6.11 Bipolar and BiCMOS Protection Circuits 179
6.12 Summary 183
Bibliography 184

7 Advanced Protection Design 188
7.1 Introduction 188
7.2 PNP-Driven nMOS (PDNMOS) 188
7.3 Substrate Triggered nMOS (STNMOS) 189
7.4 nMOS Triggered nMOS (NTNMOS) 192
7.5 ESD for Mixed-Voltage I/O 200
7.6 CDM Protection 214
7.7 SOI Technology 215
7.8 High-Voltage Transistors 216
7.9 BiCMOS Protection 218
7.10 RF Designs 219
7.11 General I/O Protection Schemes 220
7.12 Design/Layout Errors 221
7.13 Summary 223
Bibliography 224

8 Failure Modes, Reliability Issues, and Case Studies 228
8.1 Introduction 228
8.2 Failure Mode Analysis 229
8.3 Reliability and Performance Considerations 238
8.4 Advanced CMOS Input Protection 239
8.5 Optimizing the Input Protection Scheme 242
8.6 Designs for Special Applications 249
8.7 Process Effects on Input Protection Design 253
8.8 Total IC Chip Protection 255
8.9 Power Bus Protection 256
8.10 Internal Chip ESD Damage 258
8.11 Stress Dependent ESD Behavior 263
8.12 Failure Mode Case Studies 267
8.13 Summary 271
Bibliography 272

9 Influence of Processing on ESD 282
9.1 Introduction 282
9.2 High Current Behavior 284
9.3 Cross Section of a MOS Transistor 287
9.4 Drain-Source Implant Effects 288
9.5 *p*-Well Effects 293
9.6 *n*-Well Effects 294
9.7 Epitaxial Layers and Substrates 295
9.8 Gate Oxides 298
9.9 Silicides 300
9.10 Contacts 304
9.11 Interconnect and Metalization 305
9.12 Gate Length Dependencies 306
9.13 Silicon-on-Insulator (SOI) 310
9.14 Bipolar Transistors 312
9.15 Diodes 314
9.16 Resistors 315
9.17 Reliability Trade-Offs 316
9.18 Summary 317
Bibliography 320

10 Device Modeling of High Current Effects 326
10.1 Introduction 326
10.2 The Physics of ESD Damage 327
10.3 Thermal (“Second”) Breakdown 330
10.4 Analytical Models Using the Heat Equation 335
10.5 Electrothermal Device Simulations 339

10.6 Conclusion 344
Bibliography 345

11 Circuit Simulation Basics, Approaches, and Applications 350

11.1 Introduction 350
11.2 Modeling the MOSFET 351
11.3 Modeling Bipolar Junction Transistors 367
11.4 Modeling Diffusion Resistors 371
11.5 Modeling Protection Diodes 375
11.6 Simulation of Protection Circuits 376
11.7 Electrothermal Circuit Simulations 382
11.8 Conclusion 385
Bibliography 389

12 Conclusion 394

12.1 Long-Term Relevance of ESD In ICs 394
12.2 State-of-the-Art for ESD Protection 395
12.3 Current Limitations 396
12.4 Future Issues 398
Bibliography 399

Index 401

Preface

In the seven years since the first edition of this book was completed, Electrostatic Discharge (ESD) phenomena in integrated circuits (IC) continues to be important as technologies shrink and the speed and size of the chips increases. The phenomena related to ESD events in semiconductor devices take place outside the realm of normal device operation. Hence, the physics governing this behavior are not typically found in general textbooks on semiconductors. Similarly the circuit design issues involve nonstandard approaches that are not covered in general books on electronic design. There has been a large amount of work done in the areas of ESD circuit design and the physics involved, most of which has been published in a number of papers and conference proceedings. This book covers the state-of-the-art in circuit design for ESD prevention as well as the device physics, test methods, and characterization. We also include case studies showing examples of approaches to solving ESD design problems.

For the second edition, we have completely revised a number of chapters and brought other chapters up to date with the latest learning. The last seven years have seen many developments in the understanding of ESD phenomenon and the issues related to circuit and transistor design, as well as to modeling and simulation.

The book is intended for those working in the field of IC circuit design and transistor device design. In addition, the basics presented in this book should also appeal to graduate students in the field of semiconductor reliability and device/circuit modeling. As the problems associated with ESD become significant in the IC industry the demand for graduates with a basic knowledge of ESD phenomena also increases. We hope that this book will help students meet the demands of the IC industry in terms of understanding and approaching ESD problems in semiconductor devices.

There are many companies and research institutes that have made it possible to understand and solve the majority of ESD problems in ICs. Some of the companies that have been particularly active in recent years are Texas Instruments, Philips Semiconductors, Lucent, Rockwell, IBM, Motorola, DEC/Compaq, David Sarnoff Labs, and Intel. Research Institutes that have made significant contributions in recent years are Sandia National Labs, Clemson University, Stanford University, the University of California in Berkeley, the University of Western Ontario in Canada, the University of Illinois at Urbana-Champain, Twente University in The Netherlands, the Technical University of Munich and the Fraunhofer Institute both in Germany, and IMEC in Belgium.

We have many people to thank for their contributions to our personal knowledge and understanding in this area. We would particularly like to thank Robert Rountree,

Thomas Polgreen, and Amitava Chatterjee for their contributions both at the circuit design and at the device level. Ping Yang and William Hunter have provided excellent technical guidance during the evolution of the work on ESD, and without their management support this work would not have been undertaken in the first place. Many of our colleagues here at Texas Instruments have done the groundwork, which has helped us expand our understanding in this area. We are especially grateful for the contributions of Kuen-Long Chen, David Scott, Vikas Gupta, Mike Chaine, Karthik Vasanth, Vijay Reddy, Tom Diep, Steve Marum, and Julian Chen, in this respect. In the area of device physics and modeling, the contributions of Mi-Chang Chang, Kartikeya Mayaram, Jue-Hsien Chern and Jerold Seitchik have been invaluable. We have had the pleasure of working closely with many academic institutions, and we thank Professors Henry Domingos at Clarkson University, Ken Goodson, Robert Dutton, Kaustav Banerjee at Stanford University, Chenming Hu at UC Berkeley, Elyse Rosenbaum and Steve Kang at University of Illinois at Urbana-Champain, and Jan Verweij, and Fred Kuper at the University of Twente, for their collaboration over the years. We greatly appreciate the significant contributions that Carlos Diaz and Sridhar Ramaswamy (University of Illinois at Urbana-Champain), Kaustav Banerjee (UC Berkeley), Xin Yi Zhang (Stanford), Sungtaek Ju (Stanford), and Gianluca Boselli (University of Twente), during their PhD studentships, have made to our understanding of the many issues related to ESD in silicon integrated circuits.

Ajith Amerasekera
Charvaka Duvvury
Dallas, November 2001.

1 Introduction

1.1 BACKGROUND

The phenomenon of electrostatic discharge (ESD) gives rise to images of lightning strikes or the sparks that leap from one's fingertips when touching a doorknob in dry winter. The sparks are the result of the ionization of the air gap between the charged human body and the zero-potential surface of the doorknob. Clearly a high voltage discharge takes place under these circumstances with highly visible (and sometimes tangible) effects. In the semiconductor industry, the potentially destructive nature of ESD in integrated circuits (IC) became more apparent as semiconductor devices became smaller and more complex. The high voltages result in large electric fields and high current densities in the small devices, which can lead to breakdown of insulators and thermal damage in the IC. The losses in the IC industry caused by ESD can be substantial if no efforts are made to understand and solve the problem [Wagner93]. Figure 1.1 shows that the distribution of failure modes observed in silicon ICs and ESD is observed to account for close to 10% of all failures [Green88]. The largest category is that of electrical overstress (EOS), of which ESD is a subset. In many cases, failures classified as EOS could actually be due to ESD, which would make this percentage even higher [Merrill93].

The significance of ESD as an IC failure mode has led to concerted efforts by IC manufacturers and university research workers in the US, Europe, and Japan to study the phenomena. Progress has been made in understanding the different types of ESD events affecting ICs, which has enabled test methods to be developed to characterize their ESD [Bhar83][Greason87]. ESD prevention programs have been put in place during IC manufacturing, testing, and handling, which have reduced the buildup of static and the exposure of ICs to ESD. Studies have been made of the nature of destruction in IC chips and, based on this work, techniques for designing protection circuits have been implemented, which has made it possible for the present generation of complex ICs to be robust for ESD.

The introduction of each new generation of silicon technology results in new challenges in terms of ESD capability and protection circuit design. Figure 1.2 shows how ESD performance for specific protection circuits has changed over time. Initially the ESD performance improves as the circuit designs mature and problems are solved or debugged. After a certain time the technology changes

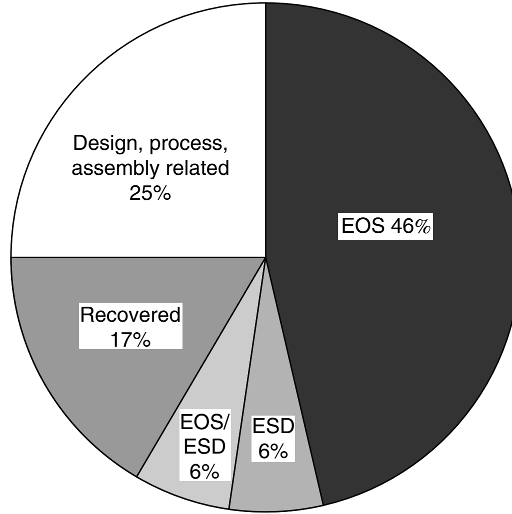


Figure 1.1 Distribution of failure models in silicon ICs. ESD accounts for approximately 10% with EOS responsible for close to 50% of the failures (After [Green88])

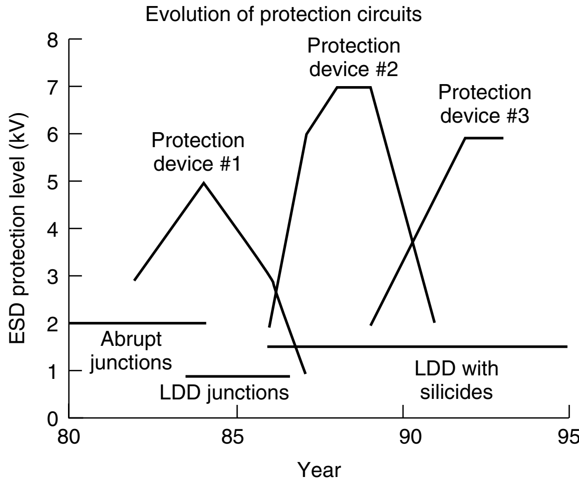


Figure 1.2 ESD protection levels as a function of time. As technologies change, new protection devices are needed to reach the same ESD levels as before

(i.e., LDD, silicides) cause the circuit to no longer function to its original capability, and the introduction of new protection techniques are needed to restore good ESD performance. CMOS ICs in automotive environments require very high ESD protection levels, which places an even higher demand on the design of protection circuits. The speed with which new technologies are introduced have reduced the

available time for protection circuit development. In fact it is becoming more and more important to design circuits that can be transferred into the newer technologies with minimum changes. Hence, it is necessary to understand the main issues involved in ESD protection circuit design and the physical mechanisms taking place in order to ensure that the design can be scaled or transferred with minimum impact to the ESD performance. The purpose of this book is to provide an introduction to the basic mechanisms involved in ESD events, the physical processes taking place in the semiconductor, and the design and layout approaches to obtain good ESD performance.

The importance of building-in reliability demands design approaches that include ESD robustness as part of the technology roadmap.

The design and optimization of circuits with ultrasmall transistors (sub-0.25 μm) use a large number of simulation tools prior to committing the circuits to silicon. Thus, modeling and simulation of ESD effects in the protection circuit is important; we discuss the main approaches here. The book is aimed at providing an overall picture of the issues involved in ESD protection circuit design and analysis. It is intended to provide a basis in this field for circuit design and reliability engineers as well as process and device design engineers who have to deal with ESD in integrated circuits.

1.2 THE ESD PROBLEM

ESD is the transient discharge of static charge, which can arise from human handling or contact with machines. The mathematics of the generation of static electricity has been presented in some detail in previous works [Bhar83][Greason87]. In a typical work environment, a charge of about 0.6 μC can be induced on a body capacitance of 150 pF, leading to electrostatic potentials of 4000 V or greater. Any contact by the charged human body with a grounded object such as an IC pin can result in a discharge for about 100 ns, with peak currents in the ampere range. The energy associated with this discharge could mean failure to electronic devices and components. Typically, the damage is thermally initiated in the form of device or interconnect burnout. The high currents could also lead to on-chip voltages that are high enough to cause oxide breakdown in thin gate MOS processes. The latter form of damage requires a large amount of energy. Many semiconductor devices can be damaged even at a few hundred volts, but the damage is too weak to be detected easily, resulting in what is known as *walking wounded* or *latency effects* [McAteer82]. A device can be exposed to undetected ESD events, starting in the fabrication area during process [Hill85] and extending through the various manufacturing stages up to the system level. Thus, precautions to suppress ESD become important through all phases of an IC's life.

As mentioned earlier, ESD is a subset of the broad spectrum of EOS, where the EOS family includes lightning and electromagnetic pulses (EMP). EOS, in general, commonly refers to events other than ESD that encompass time scales in the

microsecond and millisecond ranges compared to the 100 ns range associated with ESD. EOS events can occur due to electrical transients at the board level or the system level. They can also occur during device product engineering characterization or during the burn-in test. Although much of the reliability focus has been on ESD, EOS is being increasingly considered to be a major issue demanding more attention as it becomes a significant failure mode in the IC industry. Much of the device physics and analytical modeling discussed here will be equally applicable to EOS stress conditions.

1.3 PROTECTING AGAINST ESD

The main ESD problem in a wafer fabrication area is static charge generation, which needs to be suppressed. Prevention methods include the use of antistatic coatings to the materials or the use of air ionizers to neutralize charges. Damage caused by human handling can be reduced by proper use of wrist straps for grounding the accumulated charges and shielded bags for carrying the individual wafers. Static control and awareness are two important programs to combat ESD in the semiconductor-manufacturing environment [McAteer79][Dangelmayer85].

As a second step to reduce ESD effects, protection circuits are implemented within the IC chip [Lindholm85]. With effective protection circuits in place, the packaged device can be handled safely from device characterization to device application. However, the packaging procedure itself can cause serious damage; antistatic precautions are also needed during the wire bonding and assembly phases. Even with good protection circuits, devices are not necessarily immune to ESD once they are on the circuit boards. Other forms of ESD from the charged boards are possible. Thus ESD precautions are important during system assembly as well. Finally, the implementation of effective on-chip protection is a continuous learning experience. Even if not very effective, a relatively weak protection circuit is better than none. A good protection design would be capable of surviving the ESD event and protect the internal transistors connected to the IC pin.

It is a challenging task to design effective protection circuits, and several design iterations can be required to optimize them.

1.4 OUTLINE OF THE BOOK

During an ESD event, the on-chip components operate outside their usual range. The behavior of semiconductor circuit elements is not covered by standard texts on semiconductor device physics. A general understanding of this behavior can be obtained from publications on the high current behavior of bipolar devices [Ghandhi77]. Similarly, circuit design and layout for ESD robustness require particular guidelines that have evolved through years of experimental work in this field. The same is true for test methods and characterization. In this book we have

attempted to present coverage of all these aspects, which would enable the reader to gain a broad understanding of ESD in ICs and the main issues involved in improving ESD performance. The book draws from a large publication base in this area, the bulk of which is available through the *Proceedings of the EOS/ESD Symposium*, which is held annually and deals with all areas of ESD. Much of the detailed understanding of ESD in ICs has been presented at this symposium. Brief outlines have been presented in review papers which have presented the state-of-the-art regarding ESD at the time of publication [Amerasekera92][Duvvury93].

The book consists of 12 chapters and an outline of the contents of each chapter is given later. Chapter 2 first presents the details of the ESD phenomena introducing the 'charge' and 'discharge' effects. With this background, Chapter 3 discusses the various appropriate test models and the test methods. These phenomenas are in terms of the voltages, currents, and pulse durations, whereas the test methods are described in terms of the simulations of the events arising from the different stress models. The test methods shown to approximate the phenomena consist of the Human Body Model to represent the human handling, the Machine Model to emulate machine contact, and the Charged Device Model to determine the effects of field-induced charging of the packaged IC. The issues dealing with the accuracy of these models and the commercial testers available to simulate them are also discussed.

To understand the mechanisms of device failures and operation of the semiconductor protection devices under the high current short duration ESD pulses, the device physics behind these will be considered in Chapter 4. The protection device design requires an understanding of the physics involved in resistors, reverse-biased PN diodes, the parasitic *npn* operation of an nMOS transistor, or the latchup operation of a PNP device.

As a new addition to the book, Chapter 5 describes the ESD protection design concepts outlining the general principles used to construct ESD protection circuits and the necessary strategy. This basic background is deemed to be necessary before delving into the protection circuit designs themselves. In Chapter 6, the design requirements for effective protection circuits that can perform without degrading the IC chip functions are discussed. For example, a protection at the input should not affect the gate-oxide reliability, an output protection should have no impact on the output buffer performance, and neither should result in an increase in the leakage current in the chip. The approach taken here will be to demonstrate a synthesis of the protection circuit design needs while considering the optimum design compatible with complex internal IC chip current paths during ESD, or the function of the chip, that is, whether it is floating substrate DRAM or a grounded substrate logic chip. Each individual protection element will first be discussed separately before combining them to form effective protection schemes. Just as important as the protection design is its implementation. The layout of a protection device plays a crucial role in its effectiveness. Both the design and layout techniques are discussed in Chapter 6. As will be demonstrated, effective protection circuit schemes can perform far below the expected

level mainly because of poor implementation. The chapter will focus on the design practices and guidelines for the protection design layouts. Even after an effective design and layout, the full ESD robustness of the IC cannot always be guaranteed.

With the recent advances in technologies, the protection circuit design has become even more challenging. Many of these latest developments are dealt with in Chapter 7. For example, completely new protection concepts had to be introduced to be compatible with high-performance transistors for the deep submicron technologies, or novel concepts had to be developed to accommodate the new IC circuit designs such as mixed-voltage applications and high-voltage applications. The concepts described in this chapter represent the very latest and form the basis for protection strategy across many companies.

To illustrate the transistor phenomena and the design techniques discussed in the earlier chapters, the main failure modes observed in advanced silicon ICs will be discussed in Chapter 8, together with case studies related to the effects of design and layout on ESD performance. This analysis involves a thorough stress methodology for characterization and a full study of the failure modes. Several actual case studies will be presented, which indicate the common, and some times more bizarre, ESD problems. A brief summary of the failure analysis techniques useful for ESD as well as the poststress failure criterion will be reviewed.

The development of newer protection techniques are needed because of the degradation of the existing protection devices with advances in process technologies as shown in Figure 1.2. In many cases, process dependence of ESD performance can frustrate any attempts to achieve the specified ESD levels for the product. Chapter 9 discusses the principal aspects related to process effects, such as the impact of LDD junctions or silicided diffusions on ESD performance. The specifics of the process effects and methods to monitor these process effects will be reviewed.

In Chapter 10, a review of the device modeling techniques based on the high current behavior of the protection circuits is given. These look at the approaches used in analytical and numerical modeling of the ESD phenomena in semiconductor devices. This continues to be an evolving field and a lot of work is currently being done to uncover the underlying mechanisms involved and identify the main predictive indicators to be used. The eventual goal is the capability to develop and evaluate high-performance ESD protection circuits in new processes using simulation techniques.

To enable more specifically the design of ESD protection circuits, there have been some recent advances in simulation methods. Chapter 11 is a new addition which gives details of circuit simulations that can be used for protection circuit development as well as in the analysis of the protection circuit behavior under ESD conditions. The latter capacity is increasingly becoming important as demand grows for efficient protection designs with minimum iterations in silicon. The eventual goal is to achieve first pass success which is within reach with the methods described in Chapters 10 and 11.

Finally, in Chapter 12 a summary of the main issues is given, together with an evaluation of the state-of-the-art with regard to protection techniques and future ESD requirements and directions for further work in this area.

BIBLIOGRAPHY

[Amerasekera92] A. Amerasekera and J. Verwey, "ESD in integrated circuits", in *Qual. Rel. Eng. Int.*, **8**, 259–272, 1992.

[Bhar83] T.N. Bhar and E.J. McMahon, in *Electrostatic Discharge Control*, Hayden, New Jersey, 1983.

[Dangelmayer85] G.T. Dangelmayer and E.S. Jesby, "Employee training For successful ESD control", in *Proceedings of the 7th EOS/ESD Symposium*, pp. 20–23, 1985.

[Duvvury93] C. Duvvury and A. Amerasekera, "ESD: a pervasive reliability concern for IC technologies", in *Proc. IEEE*, **81**, 690–702, 1993.

[Greason87] W.D. Greason, in *Electrostatic Damage in Electronics: Devices and Systems*, Research Studies Press, London, 1987.

[Green88] T. Green, "A review of EOS/ESD field failures in military equipment", in *Proceedings of the 10th EOS/ESD Symposium*, pp. 7–14, 1988.

[Hill85] H. Hill and D.P. Renaud, "ESD in semiconductor wafer processing", in *Proceedings of the 7th EOS/ESD Symposium*, pp. 6–9, 1985.

[Ghandhi77] S.K. Ghandhi, in *Semiconductor Power Devices*, John Wiley & Sons, New York, 1977.

[Lindholm85] A.W. Lindholm, "A case history of an ESD problem", in *Proceedings of the 7th EOS/ESD Symposium*, pp. 10–14, 1985.

[McAteer79] O.J. McAteer, "An effective ESD awareness training program", in *Proceedings of the 1st EOS/ESD Symposium*, pp. 1–3, 1979.

[McAteer82] 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.

[Merrill93] R. Merrill and E. Issaq, "ESD design methodology", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 233–237, 1993.

[Wagner93] R.G. Wagner, J. Soden and C.F. Hawkins, "Extent and cost of EOS/ESD damage in an IC manufacturing process", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 49–55, 1993.

2 ESD Phenomenon

Horst Gieser

2.1 INTRODUCTION

The phenomenon of electrostatic discharge (ESD) occurs when an electrostatic voltage slowly develops between an object and its surrounding environment, commonly referred to as earth or ground, then spontaneously discharges as an electrical current impulse. Therefore, the laws of electrostatics and electrodynamics apply. They are well described in various textbooks [Jonassen97] and are not the focus of this chapter. Instead, this chapter focuses on the specific aspects of ESD in an IC environment. O. McAteer and T. Dangelmayer provide additional information on the nature and control of such electrostatic discharges [McAteer90][Dangelmayer99].

Although many processes in our human body are controlled by means of electricity, we do not have any specific sense for it. Therefore, our experience of strong electrostatic fields is limited to raised hair. Even the discharge of our body capacitance at several kV via a low resistive path is recognized just as some discomfort. In particular, during dry winter, this experience is very common while walking across a synthetic carpet and touching a doorknob or when leaving a car. One measure to control the level of charge and electrostatic voltage as well as the magnitude of the current transient is to provide a safe path for the flow and recombination of the charge by means of a moderate surface resistance in the range of some kΩ/sq ... MΩ/sq. Sufficient air humidity also helps to generate this property. Another method, applicable to insulators, is the use of well-adjusted air ionizers.

If any of these ESD controls fail, the electrostatic voltages can increase, causing spontaneous high-current impulses with a duration in the range of 1 ns to 100 ns, which can either charge the sensitive IC or discharge through it. While there is no indication at all that ESD-protected ICs fail due to the pure presence of an electrostatic field, there are many cases where the discharge current impulse through the IC results in both a voltage drop and power dissipation, causing devices within the IC to fail. Depending on the pre-charge voltage capacitance, resistance, and inductance of the discharge, this current impulse may easily reach approximately 10 A. Therefore, the voltage drop across a 2 Ω power bus of a 0.13 μm CMOS device, designed for an operation at 1.2 V, may well exceed 20 V, putting the

ultra-thin gate oxides at severe risk. A person might not even recognize a discharge from an equivalent pre-charge voltage level.

In an IC environment there are a multitude of processes that may generate charge onto or from persons or objects, such as parts of machines, ICs, modules, packages, and CRT screens. The following chapter discusses the mechanisms for the generation of the electrostatic voltage and for the fast current impulses in an environment in which ICs are manufactured and handled by persons or machines. Knowledge of the charging mechanisms increases the reader's awareness for avoiding ESD control problems and, in cases where they do occur, to trace them back to standard ESD stress models. This chapter also provides the background behind the ESD test methods used for product qualification and behind the pulsed characterization techniques (see Chapter 3).

2.2 ELECTROSTATIC VOLTAGE

The electrostatic voltage resulting from the separation of charge is the driving force for the discharge current. The voltage on a charged object relative to earth ground can be easily measured by means of an electrostatic voltmeter. If a discharge takes place between two objects, the voltage difference and the capacitance between these objects must be considered at the actual instant of discharge. Decreasing the distance between the objects or adding a third object at a lower potential increases the capacitance and thus reduces the voltage. This process is called *capacitive voltage suppression*. In an IC-handling environment, the four basic mechanisms generating electrostatic voltages are triboelectric charging, ionic charging, direct charging, and field-induced charging. The first two mechanisms are slow processes. The current impulses of the latter two depend on the impedance of the charge path and may stress the IC.

Triboelectric charging results from the mechanical contact and separation of two surfaces with different electron affinity. The object with the higher affinity acquires the electron. After separation it will remain negatively charged with respect to the object that had spent the electron. If the charges cannot immediately recombine, additional instances of contact and separation increase the amount of charge, which builds up a higher voltage. Actually, no friction or rubbing is necessary to generate and separate charge. The more rapid the separation of the objects carrying the charges occurs, the less the chance to recombine. Any contamination of the surface, humidity, temperature, and the roughness and pressure of the surface contact have a significant influence. Humidity, in particular, increases the surface conductivity, raising the rate of recombination. In general, highly insulating hydrophobic materials such as PTFE, better known as Teflon™ or silicone rubber, are the most susceptible to charging and can carry the charge for nearly an infinite time.

The material that gives electrons or acquires electrons when rubbed against another material has resulted in the frequently cited *triboelectric series*, which has been generated for various materials on the basis of many laboratory experiments

[Moss82]. The underlying theory of contact and frictional electrification is well explained by Harper [Harper98]. The technical advisory ADV11.2-1995: 'Triboelectric Charge Accumulation Testing,' published by the ESD Association, provides good insight in the complex phenomenon of triboelectric charging. It reviews procedures and problems associated with various test methods that are often used to evaluate triboelectrification. The test methods reviewed indicate gross levels of charge and polarity, but are unrepeatable if used to give more exact values in *real-world* situations [ESDA-Tribo95].

In *real-world* situations, tribocharging occurs, for example, when walking or rising from a chair. It can charge moving parts of machines as well as IC packages. It also results from spray cleaning, for example, with pressurized high-resistive deionized water or from blasting with carbon dioxide pellets. Without proper grounding, for example, with wrist straps or controlled conductivity, tribocharging can generate electrostatic voltages of up to some 10 kV on persons. Voltages on parts of machines, ICs, and modules are usually lower. Corona discharges or residual conductivity may also limit the voltage.

Ionic charging, the second process, is associated with the use of air ionizers that are inevitable for the neutralization of immobile charge on insulating surfaces. It occurs only if the flow of ionized gas molecules is not properly balanced or adjusted to the charging properties of the individual manufacturing process step. The resulting voltage can easily exceed some 100 V, but should be controlled to below this level.

Direct charging, the third process, occurs if mobile charge is directly transferred from a charged object, such as a cable into an IC. The amplitude and duration of the current pulse depend on the voltage difference, the capacitance of the IC or voltage source with respect to the environment, and the impedance of the charge path. It may be associated with the insertion of a device into a test socket.

Field-induced charging, the fourth process, is closely related to direct charging. In this case, a neutral IC is brought slowly into an external electrostatic field, or the electrostatic field increases [Speakman74][Bossard80]. This external field causes the separation of mobile charge on the conductive parts of the IC, in particular, on its lead frame and the semiconductor chip itself. As soon as this still neutral IC makes contact to another conductive object at a different voltage, a very narrow, very high current pulse charges the IC.

The consecutive discharge of the charged device to another object stresses it again, but with the opposite polarity of the current flow [Lafferty84]. Most often, tribocharging is the root cause for the generation of the external electrostatic field. Immobile charge may be sitting either on the package of the IC, on a part of a machine or any other item in the close vicinity of the IC. Obviously, in an environment in which sensitive microelectronic components are handled, the generation of electrostatic voltages must be strictly controlled and ideally minimized in all instances of the process.

Whether constant or slowly changing electrical fields without discharge events are dangerous for integrated circuits and/or discrete devices will be discussed.

Only devices with dielectric layers that are not shunted by any pn-junction or other conductive path could be affected in principle as eventual voltage differences of the external leads acting as antennas could become present across the internal dielectrics. Modern examples for such devices are discrete RF-MOS-transistors. Sometimes this pure electrostatic instance without rapid charge or discharge processes taking place is called *field-induced model, FIM* [Unger81]. If there are any shunt paths such as diodes between the leads of the device, there is good evidence that voltage differences can almost completely equalize in the time constants associated with mechanical motion. The resulting low displacement currents are not likely to damage any pn-junctions or dielectrics of ICs.

2.3 DISCHARGE

The discussion of the discharge mechanism applies to any spontaneous transfer of charge between two conductive objects at a different electrostatic voltage in order to establish the same electrostatic voltage on both objects. For simplification, we assume that the discharge takes place between one charged conductive object and a grounded conductor that it approaches. The isolator between the two electrodes may either be air or a controlled gas atmosphere in a relay switch of an ESD test system. The transition from isolation to conduction requires the breakdown of the isolator or, if the breakdown condition is not fulfilled during the approach of the electrodes, the final direct contact of the electrodes. While the approach takes some milliseconds, the breakdown and the consecutive discharge occurs in the ns-domain. For high voltages, an air gap breaks down starting with a single electron that generates an avalanche. Finally, a plasma channel of ionized gas develops to a low resistance. The formation of the resistive phase is accompanied by a visible and audible spark and may take at least some 100 ps for a high-voltage ESD in an IC environment [Renninger91][Hyatt93][Lin92]. For low voltages or in a relay, either evacuated or filled with a highly pressurized gas, the avalanche cannot develop easily. Therefore, the gap closes until field emission and direct contact establish the conduction. This mechanism results in transitions of few 10 ps [Pommerenke93][Boenisch01], which is one order less than the system rise time of the current oscilloscopes for single events. Therefore, the measurement accuracy for such fast impulses is very limited. In particular for voltages in the kV regime, a complex set of parameters have a more or less significant influence on the trigger phase and the resistive sustaining phase of the discharge: Parameters are the electrostatic voltage, the polarity, the distance between the electrodes, their shape, material, and surface layers, their speed of approach, their illumination, the composition of the gas, and its pressure. After a conductive stage has been reached employing the locally available mobile charge, the amplitude and the waveform of the discharge current is strongly influenced by the time- and current-dependent resistance of the plasma channel, the external resistance, the capacitance, and the inductance of the discharge circuit. The *lumped element* approach for the discharge

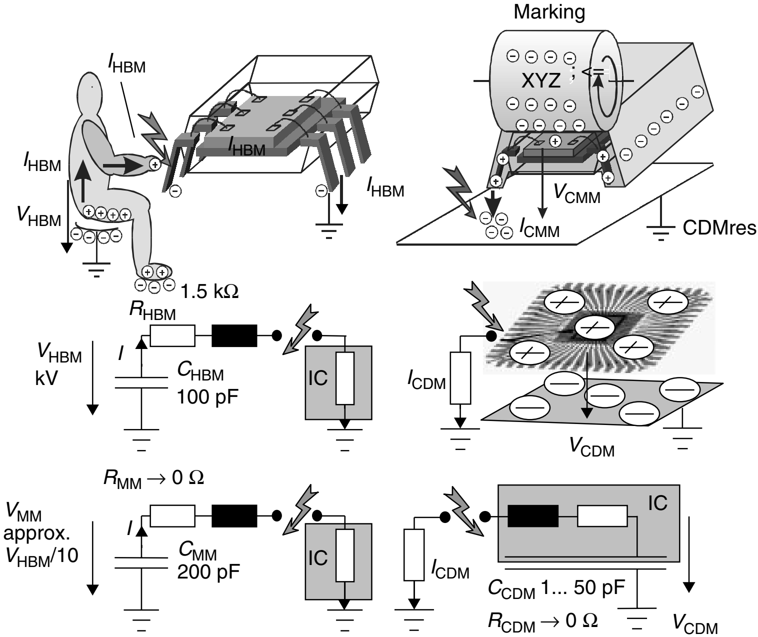


Figure 2.1 ESD stress models HBM, MM, and (F)CDM with typical parameters. (After [Gieser99], reproduced by permission of Shaker Verlag)

circuit is only valid if the geometric size is smaller than 20% of the shortest wavelength in the spectrum of the discharge impulse. The interdependencies of these parameters are very complex and yet not fully understood for the ESD voltage domain. The major influence on the reproducibility of a discharge across a closing air gap results from the *statistical time lag* between the time the strength of the electrical field fulfills the requirement for breakdown and the time the *lucky* electron actually starting the avalanche becomes available. During this time lag, the closing of the gap continues and the electrical field strength increases further. The avalanche multiplication factor and thus the discharge current depends exponentially on the field strength [Renninger91]. The reproducibility is a major issue of all ESD test methods.

After the introduction into the principle process of electrostatic charging and discharging, the next section introduces the ESD stress models that can be seen as representative cases of *real-world* ESD events.

2.4 ESD STRESS MODELS

Mainly two types of ESD-phenomena occurring in an IC environment are distinguished and simplified into stress models. The first assumes a charged person

approaching a grounded IC. When the air breaks down between the finger and one pin of the IC, the protection structures in the IC turn on and the capacitance of the person is discharged via the IC and the grounded pin into ground. Without the protection structure, voltage can build up across the gate oxides of a MOS-structure until it breaks down and closes the discharge path. This model is called *Human Body Model HBM*. Originally, the *Machine Model MM* has been introduced in Japan as a more severe HBM with intentionally $0\,\Omega$-resistance and a larger capacitance. It may simulate stress from sources with low impedance that are not necessarily ESD. It is not representative for ESD in an automated handling environment. Both models, HBM and MM, involve at least two pins and typically generate power-related failures in pn-junctions.

ESD-mechanisms in manufacturing and automatic handling are typically associated with a charged piece of equipment, a charged IC-package or the charged conductive parts of the IC itself. Depending on the charge-up method, this model is referred to as *charged device model CDM* [Speakman74][Bossard80] or *Field-Induced Charged Device Model FCDM* [Renninger89][Lafferty84]. In the (F)CDM, the capacitance of the lead frame and chip is charged or discharged, with respect to the environment, as soon as the electrical breakdown of the residual gap is initiated while it approaches a conductive object at a different electrostatic voltage. In the worst case, the discharge is determined by the capacitance of the device, the inductance of the pin, and the resistance of the ionized channel, resulting in an extremely narrow pulse with a high peak current even for voltages around 1 kV. In most cases, the voltage drop across protection elements and power bus resistance raises across internal gate oxides and causes damage.

Enoch and Lin have investigated the effects of the field-induced charging and discharging through devices soldered on printed circuit boards as *Field-Induced Model FIM* [Enoch86] and *Charged Board Model CBM* [Lin94]. It should be noted that FIM has been introduced by Unger for the damage of unprotected MOS-transistors caused by the presence of an electrostatic field without any discharge event [Unger81]. In comparison with CDM of single ICs, the capacitance of the charged board as well as the inductances of the metal traces are significantly larger and the discharge circuit is even more complex. One important finding of Lin was that even short circuits temporarily attached to the edge connector of the board could not fully protect sensitive devices against the very fast, high-current discharges.

The situation that a charged person is putting an IC on a low resistive tabletop is actually a combination of a fast high-current CDM-impulse followed by a HBM-discharge. The large capacitance of the tabletop to ground makes the existence of a direct connection to ground optional.

For the ideal stress models HBM, MM, and CDM with an assumed RLC circuit, the discharge current can be easily calculated from the solution of the second order differential equation.

$$\frac{d^2(i)}{dt^2} + \frac{R}{L} * \frac{di}{dt} + \frac{1}{LC} * i = 0 \tag{2.1}$$

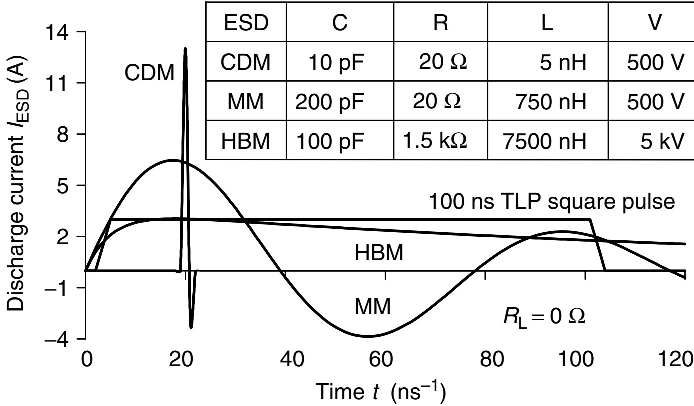


Figure 2.2 RLC-Discharge current waveforms of the three basic ESD stress models HBM, MM, and CDM for typical circuits in comparison with a TLP square pulse. (After [Gieser99], reproduced by permission of Shaker Verlag)

If the oscillation frequency $\omega = \frac{1}{\sqrt{(L*C)}}$ exceeds the damping coefficient $\alpha = \frac{R}{2*L}$, including the load resistance, the discharge is an oscillation as observed for the MM with low resistive loads. Otherwise it is aperiodically damped, like the HBM. The CDM discharge varies.

These three stress models describe the hazardous discharge through or into an IC that is not powered. Other ESD stress cases consider ICs in a module or system that may even be connected to the power supply. They describe a charged person discharging its capacitance via a metallic tool into a grounded object or an object with a large capacitance to ground. This discharge results in an initial narrow peak rising in less than 1 ns that discharges the local capacitance of the tool followed by a longer period in which the person is discharged. This more severe two-terminal stress model, typically used with a main capacitor of 150 pF and a resistor of 330 Ω, is called *System Level HBM*. If ESD is applied to a CMOS or BiCMOS-IC connected to the power supply, there is a good chance that the discharge may trigger a latch-up in the parasitic npnp-structures of the device. This mechanism is called *Transient Latch-Up (TLU)*. The ESD Association is working on the definition of a standard test practice [ESDA-TLU].

The reader can find a concise terminology for the different aspects of ESD and the protection measures in [ESDA-Glossary94], edited and published by the ESD Association.

After this brief introduction into the charging and discharging mechanisms leading to ESD stress models for integrated circuits, the following Chapter 3 Test Methods will detail how the ESD stress models have been implemented in testers for qualification, and discuss techniques for the precise measurement of impulses

and for the pulse characterization of integrated circuits and structures, whether for qualification or development of the ESD protection.

BIBLIOGRAPHY

[McAteer90] O. McAteer, *Electrostatic Discharge Control*, McGraw-Hill, New York, 1990.

[Bossard80] P. Bossard, R. Chemelli, and B. Unger, "ESD damage from triboelectrically charged pins", in *Proc. 2nd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 17–22, 1980.

[Boenisch01] S. Boenisch, W. Kalkner, and D. Pommerenke, "Broadband measurement of ESD risetimes to distinguish between different discharge mechanisms", in *Proc. 23th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 373–384, 2001.

[Dangelmayer99] T. Dangelmayer, *ESD Program Management: A Realistic Approach to Continuous, Measurable Improvement in Static Control*, 2nd ed., Kluwer Academic Publishers, Dordrecht, NL, 1999.

[Enoch86] R.D. Enoch and R.N. Shaw, "An Experimental Validation of the Field Induced ESD Model", in *Proc. 8th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 224–231, 1986.

[ESDA-Glossary94] ESD Association, *ESD ADV1.0-1994: Glossary of Terms*, 1994.

[ESDA-TLU] ESD-Association WG5.4, *ESD SP5.4: Standard Practice for Latch-Up Stressing CMOS/BiCMOS ICs Using Transient Stimulation,* Work in Progress 2001.

[ESDA-Tribo95] ESD-Association, *ESD ADV11.2-1995: Triboelectric Charge Accumulation Testing*, 1995.

[Gieser99] H. Gieser, "Verfahren zur Charakterisierung von integrierten Schaltungen mit sehr schnellen Hochstromimpulsen (Methods for the characterisation of integrated circuits employing very fast high current impulses)" in *Dissertation Technische Universitaet Muenchen TUM*, Shaker-Verlag, Aachen, Germany, 1999.

[Harper98] W.R. Harper, *Contact and frictional electrification,* Laplacian Press, 1998.

[Hyatt93] H.M. Hyatt, "The resistive phase of an air discharge and the formation of fast rise time ESD pulses", in *J. Electrostatics*, **31**, 339–356, 1993.

[Jonassen97] N. Jonassen, *Electrostatics*, Kluwer Academic Publishers, Dordrecht, NL, 1997.

[Lafferty84] D. Lafferty, "Secondary discharge: a new jeopardy and a new tool", in *Proc. 6th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 131–135, 1984.

[Lin92] D.L. Lin and T.L. Welsher, "From Lightning to Charged-Device Model Electrostatic Discharges", in *Proc. 14th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 68–75, 1992.

[Lin94] D.L. Lin and M.-C. Jon, "Off-Chip Protection: Shunting of ESD Current by Metal Fingers on Integrated Circuits and Printed Circuit Boards", in *Proc. 16th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 279–285, 1994.

[Moss82] R.Y. Moss, "Caution — Electrostatic Discharge at Work", in *IEEE Trans. Comp. Hybr. and Man.*, **CHMT-5**, 512–515, 1982.

[Pommerenke93] D. Pommerenke, “On the influence of the speed of approach, humidity and arc length on ESD breakdown”, in *Tagungsband 3rd ESD-Forum*, pp. 103–111, 1995.

[Reninger89] R.G. Reninger, M.-C. Jon, D.L. Lin, T. Diep and T.L. Welsher, “A field-induced charged device model simulator”, in *Proc. 11th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 59–71, 1989.

[Renninger91] R. Renninger, “Mechanisms of charged device model electrostatic discharges”, in *Proc. 13th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 127–143, 1991.

[Speakman74] T.S. Speakman, “A model for the failure of bipolar silicon integrated circuits subjected to electrostatic discharge”, in *Proc. 12th Annual Symposium on Reliability Physics*, IEEE, pp. 60–69, 1974.

[Unger81] B. Unger, “Electrostatic Discharge Failures of Semiconductor Devices”, in *Proc. 19th IRPS*, pp. 193–199, 1981.

3 Test Methods

Horst Gieser

3.1 INTRODUCTION

Manufacturers and users of ICs have derived the ESD test methods from the basic ESD stress models: the Human Body Model (HBM) and the Charged Device Model (CDM). The ESD test should reproduce the different failure signatures and quantify the sensitivities attributable to the various types of ESD in an IC environment. ESD test standards specify how an IC has to be stressed in an ESD test system. They specify the discharge current waveforms for a given precharge voltage, acknowledging that the discharge current through the IC, generating voltage differences and heating up structures, is responsible for the majority of ESD failure mechanisms. In principle, the ESD sensitivity levels should allow a comparison with the levels of electrostatic voltage measured in a fabrication process.

A key issue of the ESD test and its standardization is the reproducibility of test results on the same test system and the correlation between different test systems. This requires repeatable stress conditions at least on the same system. While the basic ESD stress models are simple lumped RLC circuits with ideal switches, their implementation in real ESD test systems is associated with additional distributed parasitic elements connected to the stressed ICs. In particular, the ultrashort, varying air discharges of CDM challenge the available metrology. High-pin count devices call for any possible reduction in test time. Failure criteria strongly influence the failure thresholds, too. With respect to these issues, this chapter discusses the test methods used for the ESD qualification of products, which are the traditional Human Body Model (HBM), the Machine Model (MM), the Charged Device Model (CDM), and its derivative — the Socket Discharge Model (SDM). Although the underlying physics is the same, different standardization groups continuously review and re-edit the standards. Their goal is to specify globally applied, cost-effective test methods with a high level of reproducibility and correlation across IC device types and ESD testers.

Tightened specifications contribute to improved test systems but can make them more difficult, if not impossible, to design and expensive to build, characterize, and verify during daily operation. This chapter provides the basic understanding of the test methods and their correlation issues. The process of continuous improvement makes it necessary to look into the most recent release of each standard document. Most of them can be downloaded for free from the issuing organizations.

In addition to the RLC-type stress methods, square pulse methods have been used in order to characterize and optimize ESD protection. These methods are extremely valuable for the analysis of poorly performing ESD protection in products. Only square pulses can provide detailed insight into the transient and quasistatic current versus voltage characteristics of an ESD protection element and of the elements to be protected in the ESD-relevant time and current domain. Solid-state pulse generators are used for longer pulse durations at lower amplitudes. Currently, the ESD Association is working on a standard for the square pulse characterization method [ESDA-TLP].

3.2 HUMAN BODY MODEL (HBM)

The HBM is the traditional ESD testing standard, originally defined in the MIL-STD-883x method 3015.7. This standard defines the current waveform for the discharge of a 100 pF capacitor through a 1.5 kΩ resistor and a 0 Ω load for different discharge voltages. Figures 2.1 and 2.2 compare it to other stress models. C_{HBM} stores the charge. R_{HBM} limits the current. L_{HBM} is the effective inductance of the discharge path in a real tester, which, together with R_{HBM}, determines the rise time, specified as between 2 ns and 10 ns from 10% to 90% of the amplitude I_{max}. Within the next 150 ns it has to decay to $1/e * I_{\mathrm{max}}$. Transforming this ideal impulse into the frequency domain shows that an oscilloscope with a minimum bandwidth for single shot events of 350 MHz is sufficient for the almost ideal pulse.

Stressing integrated circuits according to this standard has unveiled serious correlation issues, as shown in Figure 3.1.

They were attributed to additional parasitic elements in real testers and to their effect on the discharge waveform [Chemelli85][Lin87][Strauss87][Roozendaal90][Verhaege93]. They were attributed to the inadequate description of a real HBM-test system by the lumped element model shown in Figure 2.1. Measurement of the discharge current waveforms in the real test systems with a 500 Ω-resistor helped identify additional parasitic elements drawn in Figure 3.2. C_s is the parasitic stray capacitance of R_{HBM} and the interconnect. C_t is the parasitic capacitance of the test board and R_L is the resistance of the load or device under test (DUT). The RLC circuit can be modeled numerically or even analytically up to the fourth order to obtain waveforms for different values of the elements [Roozendaal90][Verhaege93].

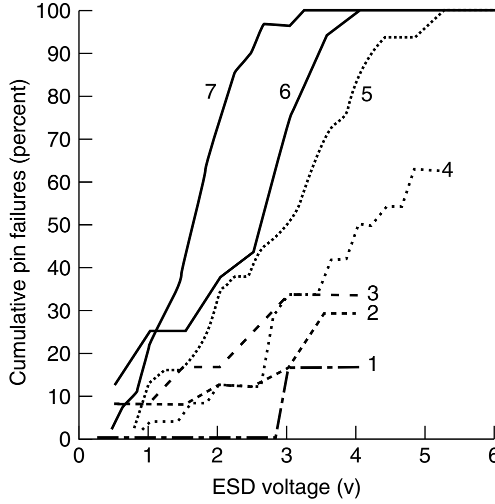


Figure 3.1 Cumulative pin failures versus ESD voltage for the same product tested with different HBM testers 1 to 7

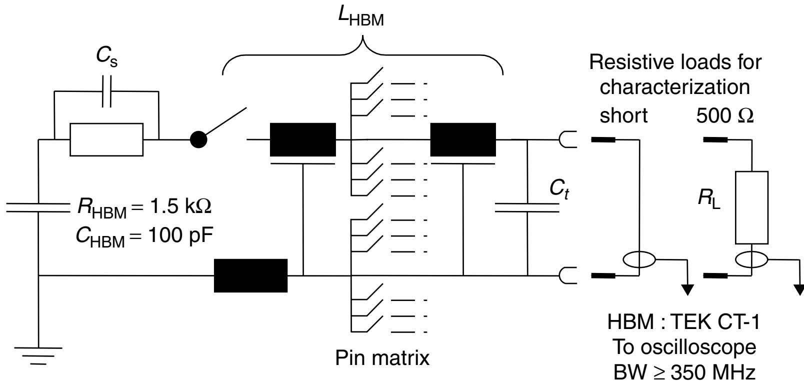


Figure 3.2 Schematics of an HBM-tester with two resistive loads for the characterization of the waveform

A simple description of the HBM current waveform can be obtained from a simplified solution for the differential equation of the RLC circuit.

$$I_{\mathrm{HBM}}(t) = V_{\mathrm{HBM}} C_{\mathrm{HBM}} \frac{\omega_0^2}{\sqrt{a^2 - \omega_0^2}} \mathrm{e}^{-\frac{R_{\mathrm{HBM}}}{2L_{\mathrm{HBM}}} t} \sinh\left(\sqrt{a^2 - \omega_0^2}\right) t \qquad (3.1)$$

with $a = \frac{R_{\mathrm{HBM}}}{2L_{\mathrm{HBM}}}$ and $\omega_0 = \frac{1}{\sqrt{L_{\mathrm{HBM}} C_s}}$, and $a > \omega_0$.

From this equation, an estimate of the rise time is given by

$$t_{\text{rise}} = \frac{2L_{\text{HBM}}}{R_{\text{HBM}}} \tag{3.2}$$

For $t_{\text{rise}} = 10\,\text{ns}$, L_{HBM} is required to be about $7.5\,\mu\text{H}$.

Even if the $1.5\,\Omega$ resistor implies a current source characteristic for the low impedances of a fully conducting protection element, a complex interaction between tester and IC may take place during the turn-on phase of the protection element. As an example, Figure 3.3 shows simulated waveforms for voltage, current, and power at a snap-back protection element which is used in the schematic of Figure 3.4 to protect an IC-input. In this case, a circuit model for the tester was combined with a circuit model for an active protection transistor. The physics and function of this gg-nMOS-protection transistor are explained

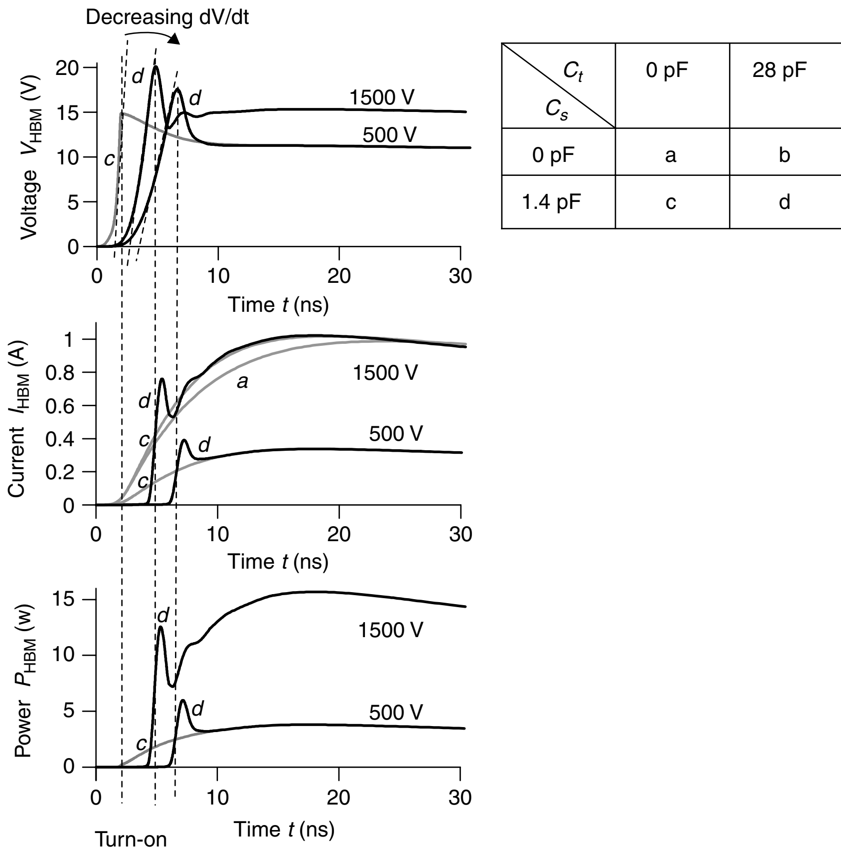


Figure 3.3 Influence of voltage V_{HBM}, stray capacitance C_s and test board capacitance C_t on the stress parameters voltage, current and power at a snapback protection transistor

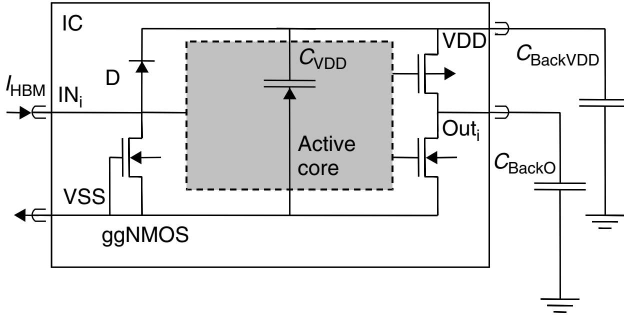


Figure 3.4 Schematics of an IC under HBM-stress in an HBM-tester

in detail in Chapters 4 and 11. The complexity of the interaction motivates the preference of simulation over analytical solutions. Most findings are valid also for the MM and TLP described in Sections 3.3 and 3.7 and the correlation of failure thresholds between the different test methods. Therefore, for a detailed analysis of the ESD behavior of a device it must always be looked at in combination with the tester and the resulting failure signature. Circuit simulation with valid models and backed by precise metrology can provide the necessary insight.

With respect to the reference waveform Figure 3.3 (a), the stray capacitance C_s of the resistor R_{HBM} reduces the rise time and causes an overshoot in current (c) with additional stress on any type of protection device in the early phase of the discharge. The test board capacitance C_t increases the rise time and reduces the peak current I_{max}. It has no effect on a discharge into a short $R_L = 0\,\Omega$ used for calibration in the MIL-standard. Figure 3.3 shows that the HBM-discharge raises the voltage across the parasitic test board capacitance C_t until the trigger condition of the protection device, in this case a snapback transistor, is reached. The turn-on of this transistor is followed by a discharge of C_t with a strong current peak not controlled by R_{HBM} to the clamping voltage at this current level as shown by the waveforms (d). This current peak adds to the regular HBM-discharge (c) and may exceed the regular peak current for low stress voltages V_{HBM} in particular. The associated peak current or peak power applied within a short period may at least initiate a localized heating and destructive breakdown mechanism depending on the internal structure of the device. The construction of the device determines the balance of the influence of the displacement current called *dV/dt-triggering [Kropf93]* with respect to the base transit time and therefore the resulting difference between the turn-on voltage to which C_t has been charged and the holding voltage to which C_t is discharged [Wolf99]. The transit time reduces with gate length and thus the dV/dt influence increases for modern technologies. The bigger the difference, the more power the protection device must handle. Snap back devices with soft leakage effects [Ohtani90], discussed in Chapter 8, and inhomogeneous triggering are more sensitive to C_t than diodes. Whether the few lumped RLC-elements of the

model are sufficient to describe the fast transitions associated with the switching of active devices in a test system of a distributed nature requires careful consideration [Russ94].

The test standards from ESD Association, JEDEC, and the AEC for HBM [ESDA-HBM01][JEDEC-HBM00][AEC-HBM01] request to measure the discharge current with a $0\,\Omega$ load and, additionally, with a low-inductive $500\,\Omega$-resistor in order to identify the effect of C_t. For the development of these standards, Verhaege *et al.* [Verhaege93][Verhaege96] have made a detailed mathematical analysis of the influences of the effective lumped elements on the HBM-waveform employing the analytical solution of the differential equation for the fourth-order lumped element HBM after Roozendaal *et al.* [Roozendaal90]. Waveforms are simulated with random parameter sets and compared to the specifications in the standard. The result is a probability curve for the selectivity of a specification in the standard (e.g., $t_{r500\Omega}$, $I_{pshort}/I_{p500\Omega}$) with respect to a certain parameter such as the test board capacitance C_t. A very steep selectivity curve for C_t demonstrates that the ESD Association standard [ESDA-HBM01] is more selective in comparison with the JEDEC-specification. The development of standards continues to employ this method and should ideally end in a single document. Large effective C_t in the order of 40 pF have been found by means of this extraction method in 512 pin testers (year 1996) and are a serious concern in particular for high pin count test systems. Whether correlation between these machines and manual testers should be expected and whether it can be enforced by a single test standard is under discussion. Nevertheless, these standards implemented in current HBM-testers have already significantly improved the correlation.

3.2.1 HBM Correlation Issues

Several detailed studies of the interaction between the tester and the semiconductor DUT have identified further correlation issues and potential test artifacts. The major influences on correlation are listed below.

- The rate dV/dt [Kropf93][Russ94][Musshoff96][Barth01] at which the voltage at the DUT initially rises until the dV/dt-dependent trigger condition is reached (Figure 3.3). While there is little or no effect on diodes until an instable runaway condition is reached, the turn-on characteristics of each individual finger of a modern protection transistor may depend on this rate. For sensitive devices only sufficiently fast pulses may trigger all fingers simultaneously. As this initial rise is not necessarily related to the 10%/90% rise time of the current pulse into the $0\,\Omega$-load or the $500\,\Omega$-resistor, an additional specification may be introduced into future HBM-standards. One rule delaying technical advance is that a new standard may not rule existing testers out. One other problem is that the subtle detail of the very initial rise may be difficult to measure in the noise floor and vary from pin to pin. It would be better to develop protection elements and transistors that are as little as possible sensitive to this effect.

- Different types of protection elements have a different sensitivity to the influence of tester parasitics. In general, diodes are less sensitive with respect to tester parasitics compared to snapback devices.
- The method to determine the peak current of the HBM in the presence of overshoot or clipping. Although the easier procedure would be to use the measured peak current, the terms of failure physics related to the energy of the pulse that is more meaningful to extrapolate the exponential decay to the time at which the maximum of the waveform is displayed and to use this value as maximum. If this value has been determined for qualification, the displayed maximum might be used for daily verification.
- The increased ESD robustness of the input IN_i of a protection scheme similar to Figure 3.4 due to the large capacitance of up to a few nF of the supply net C_{VDD} charged via the forward-biased diode D. It is sometimes overlooked that this effect may be increased for small networks by the total of the background capacitance C_{Back} of the tester that is charged via the IC [Russ93].
- The choice of the step width of the stress pulses is critical for ESD tests in general. The actual failure threshold may be very close to one or to the next step.
- The effect called *stress hardening* [Hull88]. Devices may fail if they are stressed once at the high level, but pass if this level is reached by step-stressing starting at low levels.
- The *window effect* of devices such as SCR-structures. They may trigger at low and at high stress levels satisfyingly, but fail between them. dV/dt effects and the supply of hold currents are a likely root cause [Duvvury88].
- That the grounding switch drawn in the standard has not been implemented in all testers. Without grounding or leakage measurement between pulses, a series of pulses may charge up the test board to voltages critical for the device, for example, the gate of PowerMOSFETs [Brodbeck00]. Despite the additional test time, leakage measurement between pulses is recommended for sensitive devices.
- That pins may not be properly electrically connected during the discharge event. Pins named NC, implying to be not connected to the chip, can still be bonded for undocumented test purposes and thus be damaged by ESD. If they are not connected but tested, a very severe type of test board discharge may occur when the dielectric between adjacent pins finally breaks down and the current flows into the adjacent pin. This discharge may damage the adjacent pin below its individual HBM-failure level. A similar discharge occurs if, for example, a bad contact in the test socket disrupts the interconnect between the HBM-source and the DUT and finally breaks down at a certain level [Reiner01]. Therefore, continuity tests driving current through a forward-biased junction of each pin before the stress cycle starts should guarantee this integrity.
- The failure criterion and its limits having a very significant influence on the failure threshold are further discussed in Section 3.8.
- The maximum applicable stress level may differ between different systems resulting in a misleading comparison of pass levels with failure thresholds.

3.2.2 Test Procedure

Most semiconductor and system manufacturers have in-house test procedures that are adapted to meet their own requirements. For the most part, they follow the general test procedure defined by the ESD Association, JEDEC, and the AEC [AEC-HBM01][ESDA-HBM01][JEDEC-HBM00], which have their common origin in the MIL-STD 883 C method 3015. The procedure specifies the calibration of the tester, the number of samples to be tested, and the pin combinations to be tested. The standardized failure criterion is the data sheet specification. The HBM test procedure applies in principle to all stress test methods that stress one pin with respect to a single pin, other pins, or a group of pins.

In principle, the real HBM discharge may occur between any combination of 2 pins. Therefore, the ideal coverage of sensitive combinations would be achieved stressing each pair individually. For devices with more than 64 pins this would be endless. The standard test methods had to select meaningful combinations that reflect the design of the protection scheme. Each pin is stressed with respect to one of the supply pins, V_{psi} grounded connecting either the substrate or the circuit. Similar named supply pins V_{ddi}, V_{ssi}, V_{cci}, ... may be grouped together, if the resistance between them is less than an arbitrary value of 2 Ω, causing a voltage drop during stress. Although in principle there are little differences between the documents, the JEDEC standard does not have the 2 Ω restriction [JEDEC-HBM00]. All other pins are left floating. In a next step the standards require to stress the pin with all the other nonsupply pins grounded. The effect is the same as a pin-to-pin test, but it provides a number of return paths for the stress current, which reduces the severity of the test compared to a direct pin-to-pin test. Pins such as offset adjust, compensation, clocks, control, address, data, V_{ref}, no connects (NC) are considered to be nonpower supply pins. For example, a programming power pin, usually called V_{pp}, shall be considered to be a nonsupply pin because it does not supply current to or interface with any other pins, and is not a diode drop away from any nonpower pins [ESDA-HBM01].

For evaluation purposes of small circuits one should consider to stress all the pins against every other pin in turn. For higher pin counts a reasonable judgment must be made of which pins will provide the worst-case conditions under pin-to-pin testing in order to reduce testing time. These conditions will vary according to the type of silicon (*n*-substrate or *p*-substrate), the type of IC (e.g., nMOS, CMOS, bipolar), and the circuit design itself. As a rule of thumb, the pins furthest apart from each other (i.e., diagonally opposite) and the pins adjacent to the stressed pin, would provide an indication of the worst-case performance. However, it may be necessary to check the layout of the IC and the bonding of the package to ensure that these are indeed the pins with the largest and smallest resistances to the stressed pin.

In addition to the improvement of correlation, the standardization effort is driven by the need to save test time as devices with more than 1000 pins and with a large number of power supply pins need to be qualified. These devices can require test

times of up to a few weeks. After intensive investigations, the reduction of the number of pulses per stress condition from three positive and three negative pulses to one positive and one negative pulse, and of the interval between two pulses from 1 s to 300 ms have been introduced in the most recent revision of the HBM-standard [Verhaege96][ESDA-HBM01]. Originally, three pulses were motivated by the poor repeatability of the relay and by the expectation that any possible cumulative degradation as a result of repeated pulsing will manifest itself. Electrothermal simulation shows that the stressed device should return into thermal equilibrium in far less than 1 ms [Verhaege96]. However, it cannot be excluded — even for a delay of the traditional 1 s and the leakage measurement between pulses — that residual charge remains on internal nodes of the device [Gieser95]. For further time saving for high pin count devices, the selection of representative groups of pins is discussed. With respect to the existing database, still not all companies have adopted the new standard into their own procedures.

For qualification purposes, three new components may be used at each stress voltage level or pin combination in order to avoid stress hardening or cumulative effects. Step stressing starting at any stress voltage is also permissible in order to reduce the number of devices; however, it may produce a different result as discussed in the previous Subsection *Correlation Issues*.

A minimum of three devices must be tested for qualification purposes in order to identify variations in the withstand capability. Looking at the distribution of the failure thresholds for HBM, or even better, TLP described in Section 3.7 for many more devices across the wafer gives much more detailed insight in process related ESD issues. It is also recommended to carry out persistence tests for protection elements intended for use in supply clamps with some hundreds or even thousands of pulses at 90% of the failure threshold level.

The standards, for example, from ESD Association [ESDA-HBM01] define a geometrical series of ESD stress levels between 250 V and 8 kV. They do not request that a robust IC must pass a specific level. The user of the standard and his customer must make this decision. Some advice for typical levels follows.

Company internal testing specifications usually require circuits to have a minimum pass threshold voltage of ±2 kV HBM stress on all pins, as they can be handled in an ESD-protected environment with no significant loss due to HBM-type discharges. ESD pass voltages of ±1 kV HBM are still acceptable for more complex ICs. Even lower voltages may be acceptable for few RF-pins of RF-components handled in well-protected areas. On the other side, manufacturers have developed ±4 kV ESD robust ICs, even for advanced circuits in state-of-the-art technologies, creating a target level that other manufacturers are urged to match. Originally, the ±4 kV target has been motivated by a US military specification that requests more sensitive devices to be labeled as ESD-sensitive and handled in more expensive ESD-protected areas. Many commercial specifications have similar standards.

Increasing the ESD threshold beyond 4 kV does not significantly increase the yield in order to justify the additional cost required. However, specific applications such as interface circuits and the hostile automotive environment [Sullivan85] might

even ask for HBM-withstand voltages above 10 kV [Lee88] or for tests with a system level HBM.

3.2.3 System Level HBM

After the discussion of device testing standards, the reader should be aware that the EN 61000-4-2 standard [EN-HBM96] also specifies an HBM-discharge waveform with a very fast-rising first peak followed by a longer discharge of a nominal 150 pF-capacitance of the human body through a 330 Ω-resistor. Although this test is intended for system level ESD, it is sometimes required for the test of the interface ICs for a harsh environment such as the automobile or the office. In this case, the discharge current may be directly injected into one pin of an interface that is directly connected to an IC and return via the system. Off-chip ESD protection by means of varistors, inductors, and series resistance may have been added to reach an increased immunity. Cost issues shift the demand to the protection of the IC.

3.2.4 Package Issues for HBM

In principle, HBM-thresholds are not dependent on the package by itself. However, packages are related to their specific test sockets and test boards. Although, they are specified to deliver an HBM-impulse within the specification, minor differences of their parasitics may interact with the DUT, resulting in a different failure threshold. This is in particular probable if actual failure threshold of the device is close to the selected stress level or the failure criterion is ambiguous.

3.2.5 High Pin Count Devices

Most commercial HBM-testers have 256 or 512 test channels for devices up to 256 or 512 pins sufficient to stress the majority of ICs. If the pin count of ICs exceeds even the 1400 pins available in most advanced HBM-testers, four test methods have been suggested and used by the industry and are discussed by ESD Association, JEDEC, and Sematech for future standardization.

- Three different test fixture boards could be used. The first connect all IO-pins to the tester, the second all supply pins V_{xxi} and a 1st group of IO-pins, and the third all supply pins V_{xxi} and the rest of IO-pins. While all groups can be tested almost in conformance with the standard, the overhead necessary for testing 5% to 30% more pins than tester channels are available with a product specific set of test fixture boards is significant and reduces the applicability in industry.
- The *ganging* method connects all common ground and all supply pins without the use of the relay matrix of the tester on the test fixture board together. While all groups can be tested almost in conformance with the standard, the overhead

for testing 5% to 30% additional pins with a product specific set of test fixture boards is also significant.

- The *rotation* test method is applicable to square-shaped DUT packages only and if in each quadrant is at least one pin of each group of supply pins V_{xxi}. Only one quadrant is wired and connected to the tester. After stress of this quadrant, the other three quadrants are stressed sequentially with an individual pin setup. The number of pins can be increased by an n factor of up to 4 on a nonproduct-specific test fixture board. The method cannot be used for all ICs and requires significant programming.
- The *split-IO* test method requires at least two test-fixture boards with all V_{xxi} or at least with one pin of each V_{xxi}-group and a part of the IO-pins. With simple HBM-programming, this method increases the available pin count significantly and may be combined with ganging. Equivalent to the rotation method, the IO-pin test cannot be carried out arbitrarily between different groups on different boards.

Brodbeck concludes that the split-IO method possibly combined with ganged supplies is most applicable, conceding that there is always a risk that the weak path may be in a combination of IO pins not on the same test fixture board [Brodbeck01]. In this combination, the stress may result in internal current paths that could be missed.

Other methods under discussion are to select worst-case combinations covering all types of IO-cells and protection circuits, and supply networks as well as stress conditions from the *a priori* knowledge of the design. Although feasible in the results, this approach can only be followed by the manufacturer of the IC.

Despite the reduction in test time in the current releases of the standards, as explained in Subsection *Test Procedure*, it can be an effort of weeks to qualify a single type of high pin count ICs. Together with the value of the DUTs, this should further increase the sensitivity for test time and correlation.

3.3 MACHINE MODEL (MM)

As explained in the previous chapter, the machine model, MM, intended by the Japanese IC manufacturers to be a severe HBM, is a low-impedance, high-current discharge that oscillates for a low impedance of the load. The discharge circuit is similar to the HBM-circuit of Figure 3.2. The capacitance C_{MM} is defined as 200 pF, while R_{MM} is nominally 0 Ω. In a real ESD tester, R_{MM} will be greater than 0 Ω, and during a discharge, the dynamic impedance can be much higher depending on the actual design of the tester and the type of relay. In comparison to the HBM, it generates a very similar type of power-related failure in the pn-junctions at lower precharge voltages. Minor differences may be attributed to the turn-on characteristics of alternative current paths within the protection element that may vary for different MM-testers. Depending on the device and the tester,

the failure threshold V_{HBM} is approximately $10 \ldots 20 * V_{MM}$. Gross differences in the failure signature between HBM and MM were, as an example for HBM, attributable to the breakdown of the field oxide beneath a poly silicon resistor at the input. During MM, the protection element failed.

ESD Association [ESDA-MM98], JEDEC [JEDEC-MM97], and AEC [AEC-MM01] have improved the correlation of this test defining a discharge current waveform that implies an effective inductance of 0.75 µH and an effective resistance of 10 Ω in the discharge path. The effective inductance in earlier, and now outdated, specifications was 0.5 µH and 2.5 µH. Other testers may still exist. The correlation between different 0.75 µH-testers is still worse than for the HBM test method with the 1.5 kΩ-resistor. The parasitic elements of the MM-tester have a much stronger influence on the discharge current waveform in particular through an IC with its dynamically changing impedance. The pin-to-pin variation of the impedance may increase with the pin count requiring high density interconnect load boards. Only few incidences of potential MM cases in the real world were discussed. Gate oxide failures, typical for automatic handling, could not be reproduced by MM, but by CDM [Gieser94B][Gieser94A][Chaine94]. Some semiconductor manufacturers, particularly in Japan, use the MM test mainly for internal purposes. Some manufacturers of automotive electronics and of RF devices prefer the MM intending to simulate high current events associated with a low source impedance. The high costs for the qualification of an IC with respect to a specific model are a strong motivation to minimize the number of such tests. Moreover, the failure mechanisms with the MM are often very similar or identical with those observed with HBM. For all of these reasons, it is expected that the use of the MM and the effort for developing advanced testers and standards would further decrease in the years to come.

The test procedures, pin combinations, and correlation issues are equivalent or worse in comparison with HBM. The reduction of test time is discussed in [ESDA-MM-TR00]. Typically required withstand voltage levels V_{MM} are ±200 V for regular and ±400 V for increased demands. The virtual package influence due to differences in the test boards may even be more significant as the influence of parasitics on the waveform is more significant in comparison with HBM.

3.4 CHARGED DEVICE MODEL (CDM)

The CDM either with direct or with field-induced charging is the most frequent discharge mechanism in an automated handling environment. For CDM with current pulses of few ns duration and with amplitudes of up to some 10 A generated in most cases, gate oxide failures that may not be reproduced by HBM or MM are often seen. They are commonly found in the periphery or at interfaces between power networks [Gieser94B][Gieser94A][Chaine94][Olney96][Brodbeck96]. Transistors with minor gate oxide leakage in the order of 1 µA and less caused by CDM have been shown to fail reproducibly in consecutive life tests or from 50 V

HBM raising serious concerns of latency [Gieser94B][Colvin93][Reiner95]. Recognizing the absolute necessity of a CDM-test complimenting the HBM-test, the standardization of these extremely short, difficult to measure ESD pulses has allocated significant resources around the world. It is the common understanding that well-known capacitances needs to be charged to a certain voltage level and discharged measuring the resulting current transient in order to obtain correlation. Unlike for the 1.5 kΩ current source characteristic of HBM with a nearly ideal or at least appropriate capacitor, resistor, and relay of the CDM-specific problem is that the discharge circuit depends strongly on the device and on several environmental conditions. The existing documents from JEDEC [JEDEC-FCDM00], ESD Association [ESDA-CDM99], and AEC [AEC-CDM01] are still debated between the physical background of the event, the ease and cost of use, and the freedom of designing and building a test system that the manufacturers ask for. Therefore, after a short introduction, this section discusses the major issues of CDM-testing and the correlation between different testers. It should enable the reader to understand the background, get the best from the current testers and standards, and follow or even contribute to further international standardization activities.

3.4.1 CDM Testers and Methods for Charging and Discharging

The schematic of the first CDM-test setup developed by Bossard *et al*. is shown in Figure 3.5. The device was lying in "Dead Bug" position on a ground plane to achieve a well-defined, large capacitance and contacted with a charge relay and a discharge relay [Bossard80]. Because of the relays, this mode is referred to as *contact mode*. The capacitance between the lead frame with the chip and the ground plane was charged via one pin and discharged via another pin. Typically, the pin

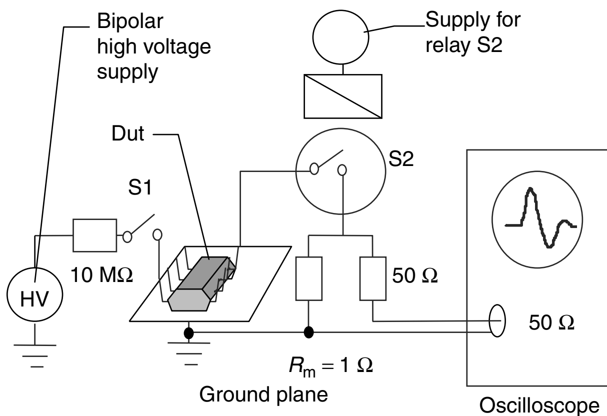


Figure 3.5 First CDM-test system. (After [Bossard80], reproduced by permission of ESD Association)

with the best contact to the substrate was chosen as the charge pin. For a *p*-type substrate this is the V_{SS}-pin. The capacitance of the device to ground depends strongly on the package and on any air gap or other dielectric between the package and the ground plane. The actual voltage is determined by the resistance of the voltage source and the isolation resistance between the chip and the ground plane as well as by capacitive voltage suppression if the device is disconnected previous to discharge. The recovery of the potential due to charge retention of insulators may also have an influence [Renninger89]. The inductance of the discharge path depends on the area that is surrounded by the discharge loop consisting of the bond wire, the pin, the relay, and of any additional wiring to the ground plane. A similar, but automated, contact mode tester was built by Shaw *et al.* [Shaw86]. Avery has minimized the parasitic inductances of the tester putting the device on a spring-driven horizontal ground plane slider and employing a discharge through an air gap during the approach of the discharge tip [Avery87].

High-pin count packages with array contacts such as PGA or BGA or bare dies require an automatic CDM-tester with an *xy*-positioning and a vertically approaching discharge tip preferably with an integrated resistive current sensor. The discharge pin is frequently referred to as *pogo pin* as it should be spring loaded to avoid mechanical damage after contact. The standardization focuses on these testers. Figure 3.6 shows a schematic of the first tester of this type employing a disk resistor with extremely low series inductance that has been introduced by Renniger *et al.* [Renninger89]. The name *Field-induced CDM (FCDM)* reflects the method of field-induced charging and discharging of the device (IC). This machine together with the test background of the former AT&T became the basis for the JEDEC standard [JEDEC-FCDM00]. Correlation was mainly an issue of copying the original machine and setup precisely. For running the test, the IC is fixed by vacuum in "Dead Bug" position on the charge plate. The charge plate can be alternatively switched to high voltage or to ground via a high ohmic resistor.

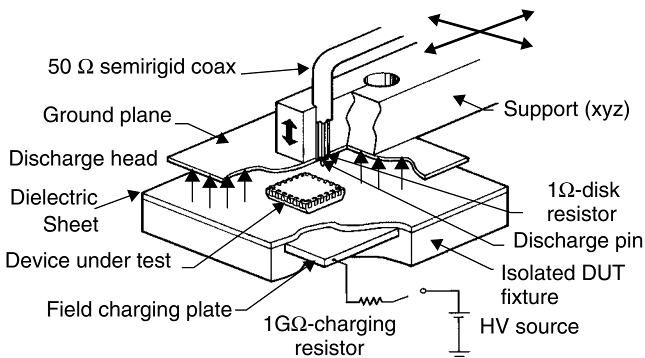


Figure 3.6 FCDM test system with disk resistor. (After [Renninger91], reproduced by permission of ESD Association)

Four procedures of (F)CDM-charging and stress can be found implemented in the different testers — some of them selectable by the user.

- The JEDEC-procedure [JEDEC-FCDM00] stresses the device with two pulses of similar amplitude but of opposite polarity. Lifting the discharge pin before the field-charging plate is grounded leaves the device in a charged state at a high potential to ground. Now, the discharge pin approaches the device again in order to discharge it. This method is very time effective, but limited to qualification test (Figure 3.7, top).
- In the ESD Association FCDM case [ESDA-CDM99], the field-charging plate is contacted to the high voltage while the device is not grounded via the discharge pin. Then the discharge pin approaches the device and a stress impulse charges the device. A positive potential of the charge plate results in a stress current impulse with positive polarity. While the device remains grounded, the charge plate is slowly discharged to ground via the 1 GΩ-resistor.

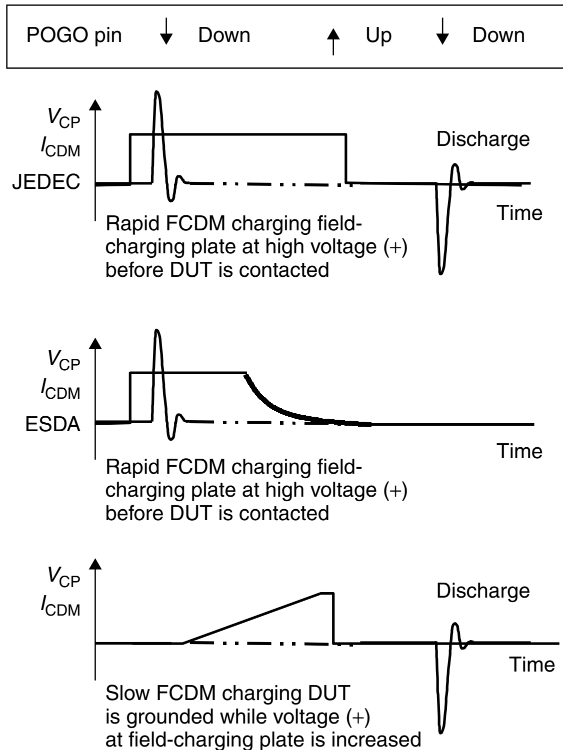


Figure 3.7 Comparison of three FCDM charging sequences resulting in a double stress of both polarities and in a single-stress impulse during either the charge or the discharge. Depicted are the voltage of the field-charge plate and the CDM current for the different contact states of the discharge pin

- In an alternative *slow charge FCDM* mode, the device is contacted first, then it is charged by connecting the field-charging plate to the high-voltage HV via the 1 GΩ-resistor or raising the voltage. Then the discharge pin is raised and the charge plate is grounded. In this case, a positive voltage at the field-charge plate causes a negative discharge current (Figure 3.7, bottom).
- In the *direct charging CDM* specified by the ESD Association, the field-charging electrode remains connected to ground. Therefore, some CDM publications refer to the field-charge plate as ground plane. Independent from the discharge pin, a dedicated charge pin is located at one edge of the ground plane and connected to the HV-supply via a high-ohmic resistor. Moving the discharge head, this HV-pin is brought into contact with the IC for charging. A positive charge polarity results in a positive stress current.

The field-induced charging process that has been implemented in the tester can be identified by changing the trigger polarity of the oscilloscope and by watching the motion. It should be noted that the failure thresholds depend strongly on the polarity of the stress current. For analysis purposes it is mandatory to know the thresholds for both polarities and the associated failure sites.

For the very fast transients, the ground plate above the IC is mainly coupled by its capacitances to the field-charge plate below the IC and to the IC itself. The DC connection may be established by the HV power supply and the outer conductor of the coaxial cable connecting the disk resistor to the oscilloscope. This questions the applicability of the simple RLC-circuit for the CDM-tester.

In order to obtain reproducible and comparable results, each (F)CDM tester needs a periodic qualification and verification. The standards specify the type and capacitance of a set of reference capacitors referred to as *CDM verification standard test modules* or *CDM test modules*. The module with the large capacitance shall indicate the performance of the coupling capacitance between the charge plate and the ground plane. After verification of the capacitance at 1 MHz these modules are charged to specified voltage levels. The discharge current pulse is measured employing a calibrated metrology chain. Parameters that are expected to have an influence on the failure threshold of ICs are extracted from the waveforms and compared with the specification. As such, peak current, rise time, the full width at half maximum (FWHM), and ratios of the second and third peak to the maximum are considered. JEDEC specifies for the FCDM two gold-plated brass disks that must have 6.8 pF ±5% and 55 pF ±5% measured at 1 MHz relative to the field-charge plate covered by a dielectric during qualification and device test [JEDEC-FCDM00]. ESD Association specifies two gold-plated or nickel-plated disks etched in the center of a single-sided clad square of FR-4 printed circuit board [ESDA-CDM99]. They have a capacitance of 4.0 pF ±5% and 30 pF ±5%. Note, if any dielectric material is used to isolate the IC from the field-charge plate, it must be added under the modules for characterization, too.

The actual resistance R_m of the nominal 1 Ω disk resistor used in most testers for measuring the discharge current may change due to electrical and mechanical

stress. It must be measured employing a 4-terminal method before and after testing and used for calculating the discharge current $I_{\mathrm{CDM}}(t) = a * V_{\mathrm{m}}(t)/R_{\mathrm{m}}$. $V_{\mathrm{m}}(t)$ is the measured voltage and a is the attenuation factor.

For characterization, the oscilloscope must have a bandwidth of 3.5 GHz, while for daily verification a less expensive 1-GHz oscilloscope is allowed. As an example, the peak current for a 500 V discharge of the small module is specified to 5.75 A ± 15% (1 GHz) by JEDEC and 7.5 A ± 20% (3.5 GHz) or 4.5 A ± 20% (1 GHz) by the ESD Association. The variation between pulses is accounted for by the tolerance band, although ±20% results in an overlap for the high voltages above 1 kV and the distribution of the peak current is not symmetrical to the nominal value. From pulse term definitions, needle-shaped pulses should preferably be specified in full width at half maximum t_{d} as there is no flat top line present that would be necessary for a valid rise time specification. ESD Association specifies this pulse width to be less than 400 ps (3.5 GHz) or 600 ps (1 GHz). JEDEC specifies it to (1 ± 0.5) ns.

Figure 3.8 shows the variations in two series of 50 consecutive CDM discharge impulses each at $V_{\mathrm{CDM}} = -500$ V and $V_{\mathrm{CDM}} = -2000$ V captured with a high-quality metrology chain (3.5 GHz) for a 4 pF-module discharged on a homemade tester with very small inductance [Gieser94A] as recommended by the ESD Association. Obviously, the amplitudes generated with this tester and a calibrated metrology chain exceeds the specification. Pulses that are far below the average pulse amplitude are referred to as *runts*. They have been triggered at a distance without excess field or the ionization in the arc channel has been disturbed. Because of the strong variations of the waveforms, no wide band sampling techniques can be used.

In Japan, other models have been developed and used in order to address CDM-like failures: the charged package model (CPM) [Fukuda86], the small capacitance model (SCM) [Wada95], and a contact mode CDM employing a relay [Tanaka97]. In the SCM, a 10 pF capacitor is discharged into 1 pin of an IC in a test socket.

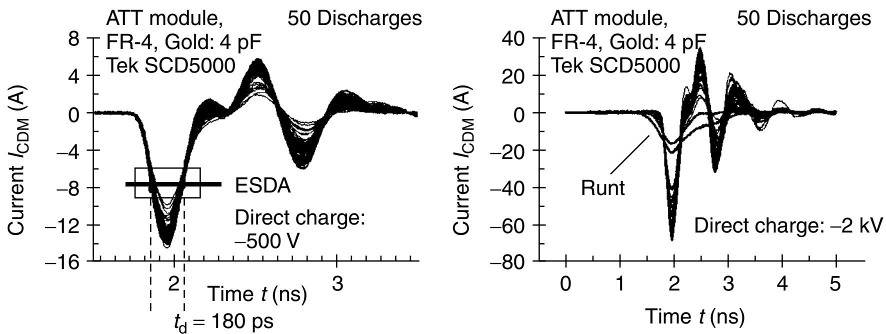


Figure 3.8 Two series of CDM-discharge waveforms at $V_{\mathrm{CDM}} = -500$ V and $V_{\mathrm{CDM}} = -2000$ V. (After [Gieser99], reproduced by permission of Shaker Verlag)

Japanese ESD standards are published by the Electronic Industries Association of Japan EIAJ.

3.4.2 Correlation Issues

This subsection discusses correlation issues that need attention or have still to be resolved in standardization. Most of them have been continuously addressed in many publications and discussions during the development of the standards.

- *Metrology:* A well-characterized *metrology* chain, which will be discussed in more detail in Section 3.6, is a general requirement for any meaningful comparison [Gieser94B][Gieser94A][Gieser99][Henry96]. The performance of state-of-the-art metrology for single pulses is not sufficient to capture the *real*, ultrafast CDM-pulses equalizing small voltage differences of small objects. The situation becomes even worse, if the metrology chain in total is not characterized by independent measures before any discharge is evaluated. Otherwise, a low reading amplitude due to a wrong attenuation factor a or resistance value R_m may actually overstress an IC with respect to a different tester that pretends to generate a high stress amplitude. Adjustments in the size of the ground plane or the length of the discharge pin or the precharge voltage V_{CDM} must not compensate errors of the metrology chain. The influence of adjustments of the discharge head above the DUT are shown in [Kagerer98]. ESD Association pays most attention to this issue requiring at least the use of calibrated high-quality components, such as precision attenuators specified to ± 0.1 dB at 3.5 GHz, $\pm 5\%$ for DC and an impedance of $50 \pm 3\,\Omega$ [ESDA-CDM99]. Still specified tolerances may sum up to a significant uncertainty. Better, the whole metrology chain and at least the permanently installed set of cable, attenuator and probe must be calibrated. Ideally, sampling techniques and deconvolution techniques should be used for characterization and qualification. Then the use of correction factors would be sufficient for daily verification.
- *Influences on CDM capacitance:* Uneven device surfaces, loose fixture during contact, and vacuum cavities beneath the device reduce the capacitance. Therefore, they increase the failure threshold voltage V_{CDM}. Vacuum cavities are effective for single side-clad modules too, and are in particular critical for the meaningful test of small devices. If vacuum fixation is inappropriate, it may help for direct charging to glue the devices temporarily with wax or Crystal Bond on a temporarily heated sheet of metal of the same shape as the grounded field-charge plate and to put this populated carrier plate into the CDM-tester. The definition of capacitances C_{CDM} of devices with heat sink fins on the backside or with other irregular shape is very questionable in the CDM test in general and asks for socketed solutions with a defined background capacitance as discussed in Subsection *SDM*.
- *Inductive Verification Modules (CIM) for relaxed metrology requirements:* Approaches [Henry99] to slow the discharge of the modules by means of

additional inductance have been made in order to relax the requirements for the metrology. They do not reflect the trend to impedance-controlled and chip-size packages or even bare dies with no additional pin inductance or transmission line such as the traces of the lead frame. However, slowing the discharge by increasing the inductance of the discharge path, for example, by means of a relay switch would lead away from the intention to simulate worst-case conditions and closer back to the original *contact mode* CDM-test setup of Bossard *et al.* [Bossard80].

- *Dielectrics covering the field-charge plate or isolating the verification modules:* The dielectric losses are frequency-dependent and may have a significant influence on the discharge. FR-4 printed circuit boards as used for the verification modules are not intended for RF-application. Better results have been shown for Alumina Al_2O_3 [Henry99].
  After the strong electrical field has been turned-off, depolarization of the dielectric causes the capacitance to decrease and therefore the static voltage to rise again due to charge recovery effects [Renninger89]. The consequence would be that the discharge current depends on the duration the precharge voltage V_{CDM} has been applied. How to deal with this issue for verification modules and ICs requires further research.
- *Charging method:* For correlation between CDM and FCDM, attention should be paid to the possibly changed polarity of the FCDM stress current with respect to CDM and to the doubling of stress events depending on the charge procedure although one polarity should be the more critical one.
- *Charge leakage:* Charge may leak from the field-charge plate to the DUT or from the DUT to the grounded field-charge plate. This leakage may depend on humidity. In worst case, a resistive voltage divider reduces the nominal precharge voltage. A clean dielectric sheet on the field-charge plate and/or cleaning the DUTs in the test procedure can avoid this error. At very high voltages of some thousand volts, depending on the actual pin geometry, corona effects may limit the voltage.
- *Misalignment:* The development of the resistance of discharge plasma and therefore the discharge current is very sensitive to the strength of the electric field. This in turn depends on the geometry of the two electrodes the IC-pin and the discharge pin approaching each other. The consequence is that the ICs must be perfectly aligned with the coordinate system of the tester, and the most position-insensitive, flat area of the IC-pin should be selected.
- *Influence of the size of the DUT:* The size of the DUT that may be a small transistor or a large multichip module relative to the field-charge plate under and the ground plane above it together with the dielectrics and their thickness (distance) determine its capacitance and to some extend inductance [Carey98]. Therefore, ESD Association requires the field-charging plate to be at least seven times larger in area than the component to be tested, in order to avoid fringe effects from the field-charge plate edges [ESDA-CDM99]. However, the ground plane in a distance of the length of the discharge pin above the device may still not fully cover devices larger than 30 mm. Fringe effects make the

appropriate DUT-size for a more homogenous E-field above the DUT even smaller.

- *Influence of the height of the DUT:* It is well known that the geometry of the ground plane and the length of the discharge pin have a strong influence on the waveform and in particular on the pulse amplitude [Carey98][Henry96]. They are used to tune the waveform into specification. One issue may arise if the height of the verification modules differs from the height of the IC pins above the field-charge plate. This issue asks for a solution similar to the 500 Ω characterization for HBM — an additional verification module with a different height or a different approach in general [Gieser99].
- *Variation of the discharge current:* In order to reduce the variation of the discharge current, the statistical time lag must be controlled either by suppressing early discharges by means of dry nitrogen or supporting early discharges by means of UV illumination or increased humidity without generating a leakage path. The alignment between discharge pin and all device pins to be stressed must be such that the discharge pin should hit the IC pins at a flat portion of the surface and not at a sharp wedge or even point.

3.4.3 Test Procedure

A minimum of three devices must be electrically characterized with respect to their static and dynamic parameters specified in the data sheet. At least, during development, data logging of these characterization tests is recommended for a direct comparison after stress. Before starting the CDM test, the operation of the tester needs to be verified. Any conductive contamination needs to be removed from the tester and the devices and any direct skin contact must be avoided. The devices should be cleaned in an ultrasonic bath filled with isopropanol. All pins of the device must be precisely aligned with the xy-axis of the CDM-tester. Any misalignment has a significant influence on the discharge current. Further, the height z must be adjusted such that in its lowered position the discharge pin at least touches the device pin or solder ball. For CDM, one pin after the other is stressed charging and discharging the device three times for each polarity. Then the devices are electrically characterized again. If all of them pass, the procedure is repeated with the next higher stress level.

After qualification and verification, the test setup including the metrology chain (except oscilloscope for verification) must not be changed, in particular, no dielectric layers beneath the DUT and the charge plate must be added. The dielectric would reduce the capacitance and thus increase the failure threshold.

It is recommended to measure and study the discharge current at least for analysis purposes at some pins of an IC as it helps to detect improper stress as a result of charge loss or misplacement, it provides good insight into the behavior of ICs under CDM-stress, and if systematically carried out helps identify under which circumstances ICs are CDM-sensitive to which discharge parameter peak current, full

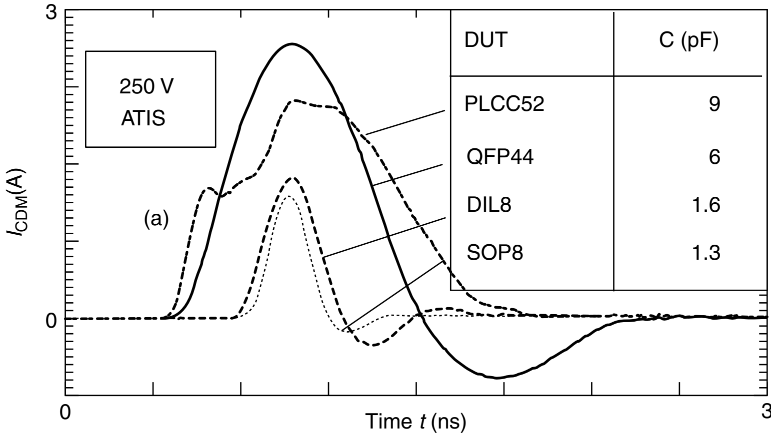


Figure 3.9 Influence of the package on CDM discharge current waveforms. (After [Gieser95], reproduced by permission of ESD Association)

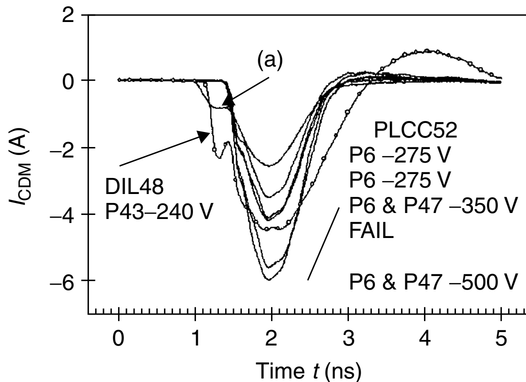


Figure 3.10 Comparison of CDM-discharge current waveforms and failure thresholds for equivalent IC-inputs in PLCC52-packages and in DIL48. (After [Gieser99], reproduced by permission of Shaker Verlag)

width at half maximum FWHM, oscillation and/or voltage as shown in Figures 3.9 to 3.11 [Gieser99][Kagerer98].

For devices with dielectrically isolated circuitry such as fully isolated SOI, in principle, all pins need to be charged simultaneously. Current commercial test equipment does not fulfill this requirement.

By no means, the critical discussion of the issues that have been slowing the standardization until now should discourage anybody to use the current (F)CDM method and standards in order to identify weak designs before they become a statistically significant number of field failures.

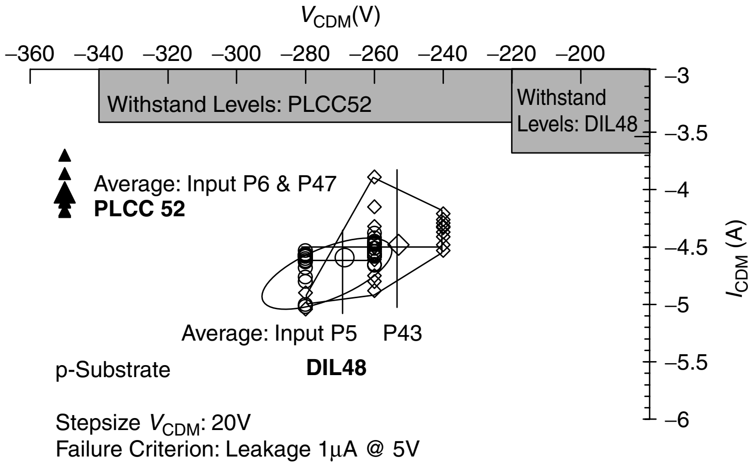


Figure 3.11 CDM-IV-Failure plot as a result of analytical CDM-testing in small step sizes and capturing the discharge current. (After [Gieser99], reproduced by permission of Shaker Verlag)

Experience over years with various (F)CDM-testers, that were or were not compliant to a specific standard, has shown that devices passing ±1 kV or even better ±1.5 kV did not fail in a protected manufacturing and handling environment, while critical devices have merely passed a ±500 V (F)CDM-test on any system [ESDA-CDM]. Ongoing standardization activities will resolve further issues [ESDA-CDM].

3.4.4 Package Issues

Different from HBM in CDM the package determines the capacitance and contributes to the impedance of the discharge path. Therefore, the failure thresholds depend directly on the package and any change of the package or its material requires a new CDM qualification with respect to the current qualification practise. The variety of devices to be tested reaches from bare dies and chip size packages over minimum size packages with discrete transistors to large packages with one or even several chips inside. A second trend, driven by mobile applications, is the reduction of the thickness of the package.

In the simple RLC model, the increase of size and decrease of thickness increases the capacitance and therefore should increase the discharge current. The increase of the size may increase the length of the leads, which could be seen as an increase of the inductance slowing the discharge and reducing the peak current. However, this view is oversimplifying as the leads may form a transmission line–like geometry, in particular with the field-charge plate beneath them and are mutually, capacitively, and inductively coupled with their neighboring leads. A preferred expression would be *trace*. The concept of transmission lines is even more valid, looking at packages

for high-speed digital applications and RF that cannot tolerate excess inductance or capacitance for signal integrity and noise reasons.

The following discusses the different phases of the (F)CDM-discharge, focusing on the interaction between package, tester, and the circuit. After the contact has been established between the charged IC-pin and the grounded discharge pin by means of a plasma channel or directly, the local capacitance of the IC-pin discharges. As the voltage falls, additional charge flows through the lead. The current is limited by the time-invariant impedance of the trace, the current-dependent resistance of the plasma channel and the time-invariant impedance of the test head consisting of the discharge pin, the current monitor, and the ground plane. Travelling along the lead, the more or less deteriorated voltage pulse front reaches the input gate of chip and turns on the protection element within finite time. During this turn-on of the protection, the voltage is already present, for example, across the gate oxide of the input inverter. This is due to the fact that the capacitance of the IC and the substrate is at high potential while the gate remains grounded via the lead. After turn-on of the protection element, the path between the IC capacitance and the ground plane remains established with a current-dependent impedance until the DUT capacitance has been discharged or one of the nonlinear elements in the path turns off again. Other protection elements and the power bus resistance are involved in the discharge too, explaining why the pMOS gate of an input is likely to fail under positive stress polarity and the probability of a failing nMOS increases for the negative polarity [Gieser94A][Gieser99].

Figure 3.9 gives an example for representative nondestructive discharge current waveforms of devices in four different packages. The same type of IC has been assembled in the packages DIL8 and SOP8. The initial rise of the PLCC52-package demonstrates clearly the discharge of the lead followed by a turn-on phase of the protection and the discharge of the Dead Bug IC capacitance C_{CDM}. Obviously, the peak current of the SOP8 and DIL8 ICs with the small capacitances did not reach the amplitude of the 9 pF PLCC52 package. The capacitances C_{CDM} were measured at 1 MHz between the backside of the die attachment and a ground plane in intimate contact with the topside of the IC employing an HP4280 CV-plotter and a homebuild fixture. For this purpose, a small hole was milled into the package [Gieser94A]. The length of the lead and the metrology chain determine which details of the discharge can be resolved. For regular ICs, there is no direct indication of the gate-oxide breakdown visible in the discharge as the broken gate oxide is in parallel to the main discharge path through the protection element. What amount of voltage stress has been applied to the gate oxide for a duration can only be deducted from circuit simulation employing accurate models, at least for the package and the protection elements. A relative simulation technique comparing precharge voltages and resulting failure current waveforms for similar inputs of the same IC helps to increase the significance of the simulation [Gieser94A] in the CDM-domain. In addition, the turn-on of the protection transistors and the breakdown of the gate oxides needs to be characterized independently from CDM-tests employing very fast rising square pulses (VF-TLP) [Gieser96]. The waveforms may depend on the

precharge voltage, and protection elements of modern submicron technologies turn on significantly faster in comparison with the shown example.

Figures 3.10 and 3.11 illustrate the analytical insights of stepping the voltage up in small increments of, for example, 25 V and measuring the discharge current waveform. They show the package influence on failure thresholds of equivalent inputs of ICs packaged in plastic DIL48 ($C_{\mathrm{CDM@1MHz}} = 17.4\,\mathrm{pF}$) and in PLCC52 ($C_{\mathrm{CDM@1MHz}} = 8.9\,\mathrm{pF}$). Two effects can be seen. The turn-on characteristic of the protection element seems to improve with increased voltage as the initial peak visible in the DIL48-curve, and in curve (a) of the PLCC52 merges with the main discharge for higher voltages. Second, the increased capacitance results in a higher peak current at a lower voltage. In comparison with PLCC52, the longer leads of the stressed DIL48-inputs result in a more oscillating discharge. If the individual failure thresholds of several input pins P6 and P47 for the PLCC52 and P5 and P43 in terms of precharge voltage and resulting peak discharge current are plotted into the CDM-IV-diagram of Figure 3.11, the sensitivity of this type of input to the peak current becomes obvious. The input fails if a peak current of approximately −3.5 A has been exceeded. The PLCC52-package had to be precharged to at least −350 V, while the DIL52-package just needed a precharge voltage of −240 V in order to generate this current [Gieser99]. This current sensitivity of the failure is also supported by another study varying the capacitance without changing the inductance by means of thinning the package or adding dielectric sheets. It suggests at least for the same CDM-test system that knowing the critical discharge current for one package, the extrapolation of the failure threshold to other packages should be possible in most cases [Kagerer98].

However applying the same analytical technique in a different case study to different inputs of the same ICs, such CDM-IV-failure plots have also shown the influence of the design on the failure threshold. They have clearly identified both a current threshold sensitivity for two similar of three inputs and a mixed current and voltage sensitivity for one of them [Gieser99].

Further package-related issues that should be considered are:

- For high discharge voltages and very fine pitch packages, neighboring pins may be stressed unintended due to the physics of the air discharge.
- A weak pin next to a strong pin may fail due to mutual coupling. However, during a qualification the weak pin would define the failure threshold [Verhaege94].

3.5 SOCKET DEVICE MODEL (SDM)

The socket device model, SDM [ESDA-SDM-TR00] is an approach to simulate the extremely narrow, fast-rising high-current pulses of the Charged Device Model in an automated ESD test system with a relay matrix and test sockets for the device. It has been implemented and used mainly in one commercial

system although others exist. There is no commonly agreed test standard that defines the model, the test method and the calibration of the tester by means of a waveform. Test experience has shown that in most cases, the method is able to produce CDM-like failure signatures and to identify CDM-sensitive devices [Olney96][Brodbeck96][Kagerer98][Chaine94][Verhaege94]. The socketed technique is in particular useful for devices with heat sink fins and other uneven nonflat backsides. However, few detailed investigations have also shown a more pronounced failure signature and even a failure that could not be reproduced by SDM-tests [Egger93][Olney96][Gossner97][Verhaege94]. Failure threshold voltages did not correlate with package capacitance in the CDM-tester. The majority of devices has failed at a lower failure threshold voltage for SDM in comparison with CDM. This suggests a method employing SDM for searching sensitive components and CDM for reproducing field failure signatures and analyzing sensitive pins [Olney96][Kagerer98]. However, some devices have been shown to fail at higher SDM-voltage thresholds [Gieser94A][Chaine94][Olney96]. A better correlation can be expected if the peak currents of SDM and CDM are compared [Kagerer98]. The duration of the stress pulse is also expected to influence the threshold.

For a better explanation of the principle, Figure 3.12 shows an extremely simplified schematic of an SDM-tester with only two channels — one trace for the discharge and only one for initially storing the charge in the background of the DUT. In a real SDM-tester, for precharge, all traces are connected to VSDM — either directly or via the device in the socket [Gieser94A][Gieser95][Chaine98]. The term “trace” has been introduced for the interconnects to differentiate them from transmission lines with a controlled impedance. The discharge is initiated in a relay at the end of one of the traces and a relatively narrow pulse with a short rise time travels to one pin of the DUT kept at elevated potential by the other charged traces. Arriving at the DUT the pulse causes a strong discharge current to flow through the device. Reflections between the discharge trace terminated with R_T and the other

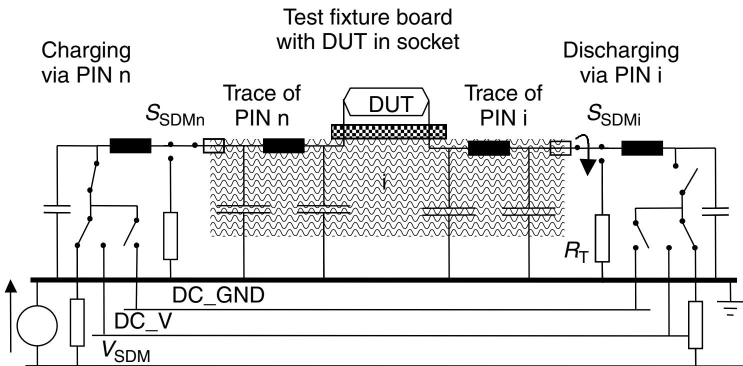


Figure 3.12 Schematic of a SDM tester drawn for only two of all traces connected to the pins of a DUT. (After [Gieser94A], reproduced by permission of ESD Association)

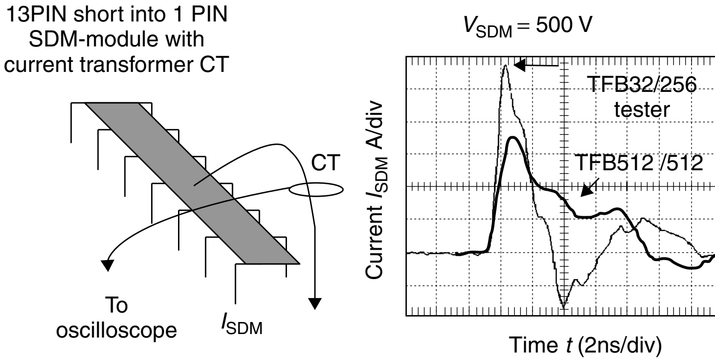


Figure 3.13 Comparison of the current waveforms discharging 13 shorted background pins into 1 pin in a 32-pin test fixture board on a 256-pin and a 512-pin test fixture board on a 512-pin SDM-tester. (After [Chaine98], reproduced by permission of ESD Association)

open ended traces seem to determine the pulse width. The peak current is limited by the impedance of the discharge trace, its resistive termination and the device. Therefore, the current first increases with the number of pins in the background until it saturates. The specific design of the test equipment in particular of the discharge paths including the traces in the background of the device strongly affects critical discharge waveform parameters like rise time, peak currents and pulse duration. Unlike CDM, the waveform of the pulse depends more on the layout of the test board very specific to a certain package than on the device itself. There is no defined capacitance of the device to the tester environment. The waveform shown as an example in Figure 3.13 is very difficult to be captured and verified consistently for different combinations of one discharge pin and several background pins. Understanding that the charged capacitance in the background and the access path to this charge plays an important role in the SDM-test, the SDM-evaluation of single protection elements is of limited value. The ESD Association has summarized the current experience on SDM in a technical report [ESDA-SDM-TR00].

3.6 METROLOGY, CALIBRATION, VERIFICATION

Errors can easily occur in the measurement of the extremely fast-changing electrical quantities associated with ESD. This section briefly discusses the principles and terminology of ESD metrology, together with some major pitfalls.

3.6.1 Fundamentals and Terms

The laws of nature should yield the same results for the same experiments wherever and whenever they are carried out. Metrology can be seen as the art of measuring

a physical quantity reproducibly with a well-known *uncertainty*. Uncertainty is the possible difference between the indicated and the true value. It is the result of systematic and statistical errors. It should be carefully analyzed to exclude the uncertainty that dominates the results and the approach may becomes worthless. A standard is a well-defined reference for a method and, if necessary, an object.

Standardization authorities such as the National Institute of Standards and Technology (NIST) in the United States have developed and maintained extremely precise *primary standards* for fundamental physical quantities such as length, temperature, time, and current, as well as combinations of them. Precision equipment used by other calibration laboratories is called a *secondary standard* or *transfer standard* and is referenced to these primary standards in order to trace measured quantities back to the primary standards. *Calibration* is the quantitative comparison of a piece of metrology or test equipment under certain environmental conditions relative to standards. For example, the equipment manufacturer calibrates a voltage meter at a certain temperature using a secondary standard voltage source. For a specified period after calibration, usually a year, and the specified environmental conditions after warm up, this entitles the user to quantify the uncertainty of his measurement, for example, 5.00 (+3%, −2%) V, as long as the equipment has not been damaged, namely, by electrical or mechanical overload or wear. In the *specification* of the equipment, the manufacturer also sets maximal limits to the uncertainty. Calibration data quantify the uncertainty traceably, but if the limits are exceeded, the equipment needs adjustment or even repair and a new qualification by the manufacturer covering all ranges and specified parameters. In order to gain evidence for a proper operation, the user must verify a subset of functions and electrical parameters employing a reference source or carrying out a relative comparison on a frequent, regular basis and document the results. *Verification* should yield whether the equipment may already be out of specification or will soon be. Therefore, proper documentation of the results must show the trends. For self-test purposes or temperature compensation rather precise sources may be built into the equipment. The objective of frequent verification together with the calibration is to minimize the resources lost for carrying out all tests again that have been done since the last qualification.

3.6.2 ESD Time Domain Metrology

Time domain metrology is necessary in order to measure the waveform of a voltage or current impulse and to characterize the components of a metrology chain. In the case of ESD, this consists of a current probe or a voltage probe, coaxial cables, attenuators, and the oscilloscope itself. For calibration and verification, a pulse generator producing well-known repetitive pulses may be necessary as a transfer standard.

For these repetitive pulses sampling techniques can be used with a bandwidth of up to 50 GHz or a rise time of 7 ps. In this case, the sampling of 512 points of a waveform requires 512 at least nearly identical pulses with very low jitter

to the trigger signal. It should be kept in mind, that any measured waveform is the convolution of the pulse with the whole metrology chain. In the case of a very fast rising step pulse, the oscilloscope has to respond to the transition and to settle with or without ringing to the final plateau. The closer the rise time of the pulse comes to the rise time of the metrology chain, which may be limited by a single element, the more information of the pulse is hidden in the response function of the chain. For linear, bandwidth-limited systems, some information can be restored by means of numerical postprocessing, called deconvolution, employing the step-response function of the system. Excellent publications on ultrafast domain metrology and traceability were written by W.L. Gans, N.S. Nahman, and J.R. Andrew [Gans90][Nahman83][Andrew94].

General sources for unintentionally attenuated or distorted waveforms with respect to the true pulse are

- step response of the metrology chain.
- interference between probe and signal, for example, by adding probe resistance into the current path or probe capacitance to the voltage node.
- single reflection at a mismatched interface.
- multiple, overlapping reflections at different mismatched interfaces.
- dispersion in cables as a result of the frequency-dependent propagation velocity.
- frequency-dependent losses of cables.
- dependency of device properties on voltage, current, power, and energy of the pulse resulting in self heating and breakdown.
- electromagnetic interference of the signal with the components of the metrology chain requiring additional shielding.

This shows why, in particular for CDM high quality, high bandwidth components must be used for the whole metrology chain and why reference experiments must be made to identify erroneous signals. In the wide-band, time domain, the accounting and compensation of frequency dependencies is far more difficult than in the bandwidth-limited frequency domain. Therefore, the calibration of the metrology chain should be done with relevant signal sources in time domain.

The following gives some advice for the components of the metrology chain and their proper application.

Electrostatic discharges having rise times in the range from some 10 ps of the low-voltage CDM up to 10 ns of the HBM must be evaluated. This requires a *single shot* time-domain metrology chain, that should be calibrated itself by means of using fast reference pulse generators and sampling techniques. In history of ESD work, the amount of insight was directly related to the available bandwidth of the single-shot oscilloscopes. As an example, the -3dB-bandwidth of current high-end single-shot oscilloscopes such as the Tektronix TDS7404 typically exceeds 4 GHz above 10 mV/div. A 500 mV/23 ps step pulse generated by means of the TDR-source of an Tektronix TDS8000 sampling oscilloscope has been measured by one individual TDS7404 to rise within 90 ps, with an overshoot of about 10% settling

within the next 200 ps to the nominal value. In the example, the settling error of this oscilloscope has been specified to reach less than 2% within less than 20 ns. The 1-GHz bandwidth TDS7104 has a calculated rise time of 400 ps. It should be noted that the accuracy of a voltage measured with an oscilloscope is typically specified for DC. The uncertainty depends on the measurement range as well as the amplitude and nature of the signal measured. Therefore, it is always good practice to characterize the oscilloscope with well-known reference impulses similar to the application as shown in Figure 3.14 for the example of CDM. Ideally, two-thirds of the vertical range should be used. For reference, the impulse was characterized with the Tektronix TDS8000 20-GHz sampling oscilloscope. The other curves show the reference impulse captured with the TDS7404 (4 GHz) and TDS7104 (1 GHz) in *equivalent time sampling (ET)* mode and with significantly less time resolution in the *real-time sampling (RT)* mode for single events. The waveform of the outdated Tektronix SCD5000 with an SHF-delay line is resolved in steps of 5 ps by means of the analog to digital scan conversion technique. The indicated waveform was almost identical for these individual TDS7404 and SCD5000 setups. The position of the sample relative to the maximum depends on the trigger level. The display of the sampled data without an interpolating line increases the awareness of the potentially under-sampled waveforms.

The specified bandwidth of attenuators, cables, and connectors should be at least 12 GHz. An uncertainty of 1 dB in the logarithmic scale translates into 12% in the linear scale. The attenuation factor of resistive wide band attenuators should be determined by means of DC measurements. Other sources of attenuation that add to uncertainty and can only be determined by system characterization are the mismatch of the impedance and the losses of cables. Therefore, the general rule applies to keep cables as short as possible, to minimize the number of connectors, including

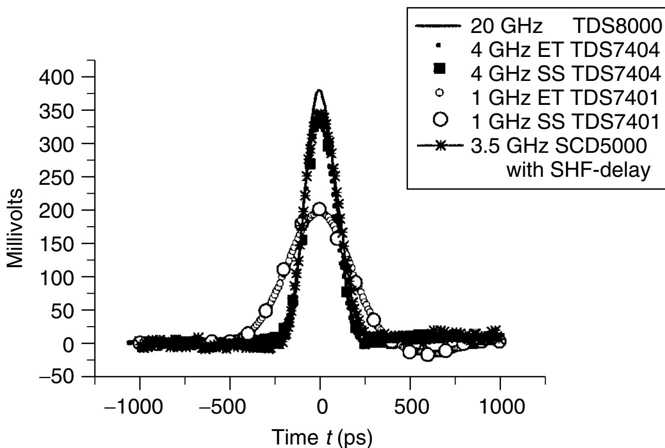


Figure 3.14 Measuring a repeatable, 195-ps-wide Gaussian impulse with four different oscilloscopes

adapters, and to maximize the diameter. In the time domain the excitation of higher modes limiting the diameter is much less of an issue than for the standing waves in frequency domain. For CDM-work SMA or precision N-connectors should be used with PTFE-foam isolated cables.

For HBM-like discharges, the discharge current is measured by means of a current transformer such as the Tektronix CT-1 probe with a bandwidth of 1 GHz. Side effects of this probe are the dynamic insertion impedance of 1 Ω, the voltage sensitivity that can be reduced by shielding and by insertion at the point closest to ground, and the saturation for very strong, long current pulses. Therefore, the CT-2 probe is used for MM-discharges. CDM-like discharge currents are commonly captured by means of a 1 Ω disk resistor [Renninger88] with far less serial inductance than a chip resistor. This resistive probe can be characterized in a face-to-face arrangement. As an example for the characterization with sampling techniques Figure 3.15 shows the transmitted response of two different CDM-test heads to a 30 ps system rise time pulse of a TDR-sampling oscilloscope HP54120T and to a Gaussian impulse formed from a step pulse of the PSPL4050 generator. Employing the rise time formula $t_r^2 = t_{measured}^2 - t_{rsystem}^2$, the calculated rise times of the CDM probes are 30 ps and 13 ps for the probe showing an overshoot of up to 10%. Using a standard voltage probe for measuring the voltage across a DUT during ESD stress requires careful consideration of the ground path of the probe. In particular, electromagnetic fields can couple into the loop of pigtail ground straps. Therefore, the ground strap must be minimized [Anderson98]. If the probe ground is connected to the ground pin of a DUT-test fixture it represents an alternative current path to ground. Differential voltage measurement techniques should preferably be used in order to minimize this influence [Smith93][Maloney01]. In addition, depending on the source impedance, the capacitance of the voltage probe increases the rise time and the resistance reduces the amplitude of the signal.

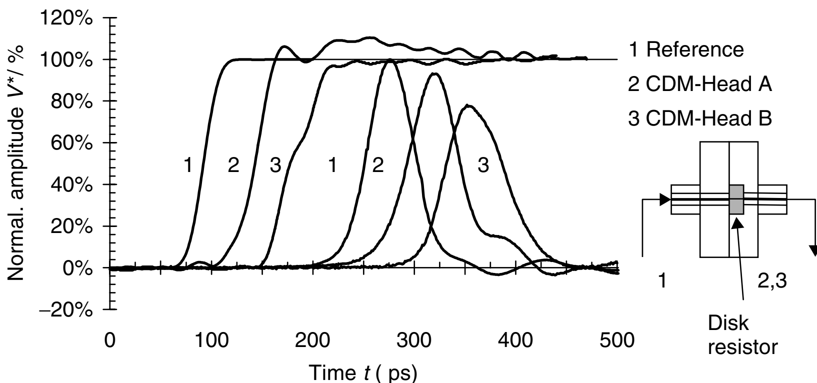


Figure 3.15 Response of different CDM-current probes to a very fast rising step and a Gaussian impulse. (After [Gieser99], reproduced by permission of Shaker Verlag)

Errors from electromagnetic coupling of noise can be identified removing the probe from the discharge circuit to install it in close vicinity of the discharge while strong discharges are triggered. For discharges with high energies and very short rise times it may be necessary to increase the distance using longer cables, delaying the signal to be measured or even to shield the oscilloscope in a Faraday cage [Smith93].

3.7 TRANSMISSION LINE PULSING (TLP)

ESD qualification tests yield only failure thresholds. More detailed information on the ESD-behavior of protection elements and schemes are required for their optimization. The safety margin of an ESD protection design is defined by the sensitivity of the elements, which must be protected with respect to the performance of the protection elements. In view of shrinking safety margins for most advanced technologies and applications, measuring the exact parameters in the ESD-relevant regime is of utmost importance. As the DC-characterization causes strong self-heating for higher currents and does not address the transient behavior, pulsed characterization techniques are necessary in addition to the DC-characterization. These techniques measure the dynamic and the quasistatic device behavior of protection elements and sensitive elements for ESD-relevant times and currents directly. The results are used for modeling and numerical simulation. This explains, why since Wunsch, Brown, and Maloney *et al.* the *Transmission Line Pulsing TLP* technique became an indispensable tool for the development of ESD protection in ICs [Wunsch68][Brown72][Maloney85][Amerasekera91]. Mainly applied to single structures, the TLP method has also successfully been used for the study of I/O cells and full circuits. In addition to the stress function, most TLP systems measure failure criteria, in particular the leakage current, before and after each pulse.

Because of the well-defined physics of generation and conduction of such square pulses, TLP has even a high potential to replace traditional methods for the qualification of products. It can be well repeated on the same system and reproduced on different systems and it delivers more complete and accurate results. In particular, high-pin count devices might drive this evolution. Until the final step of replacement is reached, the question of correlation in terms of failure mechanisms and thresholds between TLP and the traditional qualification stress tests will continue.

3.7.1 Pulsed Characterization

Before the fundamentals and different setups for the generation and measurement of the square pulses are compared, the principle and terminology of the quasistatic pulsed device characterization is explained in Figure 3.16 for the example of an nMOS transistor as a snapback protection element. Other examples for the transient pulsed characterization are the turn on of an nMOS protection transistor and the breakdown of a gate oxide shown in Figures 3.17 and 3.18.

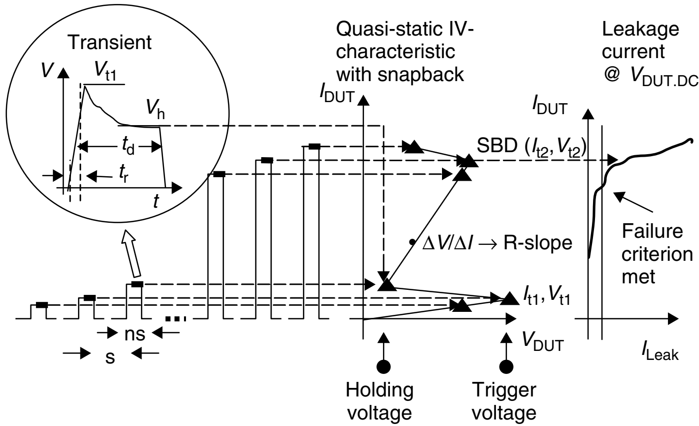


Figure 3.16 Principle of the pulsed characterization with a series of square pulses

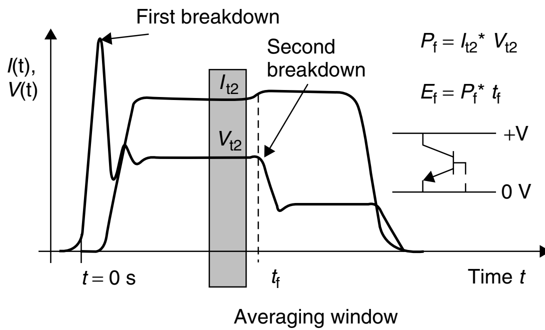


Figure 3.17 TLP-waveforms of current and voltage leading to a second breakdown failure of a bipolar snapback device

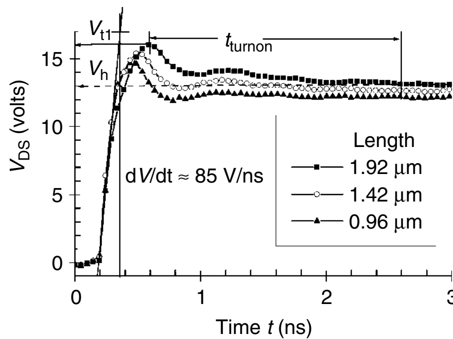


Figure 3.18 Turn-on of protection transistors. (After [Gieser99], reproduced by permission of Shaker Verlag)

As will be introduced in Chapter 4, the high current behavior of the nMOS transistor is very critical to understand its ESD behavior and for designing effective protection circuits. The important parameters are V_{t1} the trigger voltage, V_h the holding voltage, V_{t2} the failure voltage, I_{t1} the trigger current, and I_{t2} the failure current. The reader should specifically refer to the device physics section in Chapter 4 that describes these in detail. The important point to note here is that with the TLP method, the information of the nMOS ESD behavior can be obtained by measuring these values. An illustration of this powerful method is briefly described hereunder.

Traditionally, the TLP method is known for the plot of the quasistatic IV-characteristic in combination with the evolution of the leakage current for the amplitude of the square pulses increasing in steps. The quasistatic values for the voltage between the terminals of the DUT for the current through the DUT should be obtained by means of averaging a certain region in the second half of the transient waveforms. There, most oscillations as a result of the parasitic capacitance and inductance of the TLP setup should have settled. The transistor turns on when V_{t1} is reached and snaps back to the quasistatic holding voltage V_h. The turn-on time of the protection should be defined to start when V_{t1} is reached and to end when the voltage reaches V_h. Other definitions exist. For currents above I_{t1}, the device is in the resistive phase until the condition for the second breakdown failure is reached at the threshold (I_{t2}, V_{t2}). After a certain time at this power level the second breakdown is associated with a second snapback of the voltage and a significant increase of the leakage current depicted in Figure 3.17. These parameters are called time-to-failure, power-to-failure, and energy-to-failure [Wunsch68][Diaz92]. Depending on the definition of the failure criterion discussed in Section 3.8 one may refer (I_{t2}, V_{t2}) to the first onset of leakage or a certain level of leakage. This definition should always be reported together with (I_{t2}, V_{t2}).

How well the actual (I_{t1}, V_{t1}), V_h, and (I_{t2}, V_{t2}) parameters of the device are represented by the quasistatic IV-curve, depends on the chosen step size and resolution and accuracy of the system. It may be necessary to measure V_{t1} directly in the transient voltage waveform. For structures with a dV/dt-sensitive trigger condition the parameters also depend on the rise time of the voltage pulse. V_{t1} decreases for faster rising pulses and the power to failure may increase due to a more homogeneous triggering of the transistor. The obtainable resolution for low currents is a function of the source impedance of TLP-tester. The higher the source impedance with respect to the dynamic impedance of the device, the better the stress current is forced. Some devices may show several snapbacks between different resistive phases. Possible reason for this behavior may be the sequential turn-on of additional fingers, inhomogeneous conduction, or additionally triggered conduction paths in the investigated structure. Examples are devices with a resistive ballast of the transistor fingers or combinations of protection elements with output drivers, a combination of a lateral and a vertical transistor, or the parasitic transistor to a nearby guard

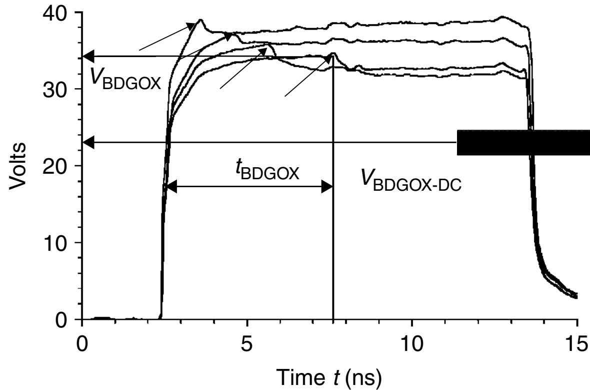


Figure 3.19 Voltage/time dependence for the breakdowns of the 20 nm thick gate oxides of small 20/1 μm-nMOS-transistors within less than 5 ns. (After [Gieser96], reproduced by permission of ESD Association)

ring. The reader is referred to the Subsection *Correlation Issues* for additional information.

The evaluation of the transient behavior requires a TLP setup with negligible parasitic capacitances and inductances. This is of utmost importance if the CDM domain is addressed. If the impedance is exactly controlled to the interface with the DUT, very fast rising narrow pulses of less then 5 ns can be used in order to obtain information from gate oxides and protection elements [Gieser96][Wu00]. Figure 3.18 shows the transient response of a snapback protection element to a square pulse with an amplitude of trigger voltage (V_{t1}). In finite time, the transistors turn on and the holding voltage V_h-level is reached. Figure 3.19 shows the voltage/time dependence for the breakdowns of the 20 nm thick gate oxides of small 20/1 μm-nMOS-transistors within less than 5 ns.

3.7.2 Fundamentals and Implementations

In order to generate a square pulse of selectable duration and amplitude, the distributed capacitance of a transmission line (TL) is charged and discharged. High voltages are necessary in order to generate current pulses of several Amperes. A transmission line is a waveguide with a characteristic impedance Z_0 that only depends on the material and the geometry of the conductors and the dielectric isolating them. The distributed capacitance and inductance are expressed in F/m and H/m. For negligible losses the electric and the magnetic field are transversal to the direction of propagation. If a pulse travels along a transmission line, any discontinuity of the impedance $Z(x) \neq Z_0$ causes a partial reflection of the energy of the incident pulse. The reflection is of the same polarity as the incident pulse if the impedance difference $Z(x) - Z(x-1)$ is positive and of opposite polarity else.

This effect is used for the characterization of unknown devices in the time domain reflectometer but must be minimized in the rest of the system. For TLP-testing of integrated structures shown in Figures 3.20 through 3.23 a high-voltage source is used to charge the distributed capacitance of the transmission line TL1 via a high-ohmic resistor while the coaxial switch S1 is open. After the switch closes, the discharge of such a transmission line (TL1) into a resistive load or into TL2 produces a square pulse. The duration of the square pulse is equal to the length of the charged line divided by the velocity the signal *"switch closed"* propagates from the switch to the high-ohmic end of this line and back to the switch. 10 m of the typical RG58 transmission line with a propagation velocity of 20 cm/ns generate a 100-ns wide pulse. The amplitude of the voltage pulse V is determined

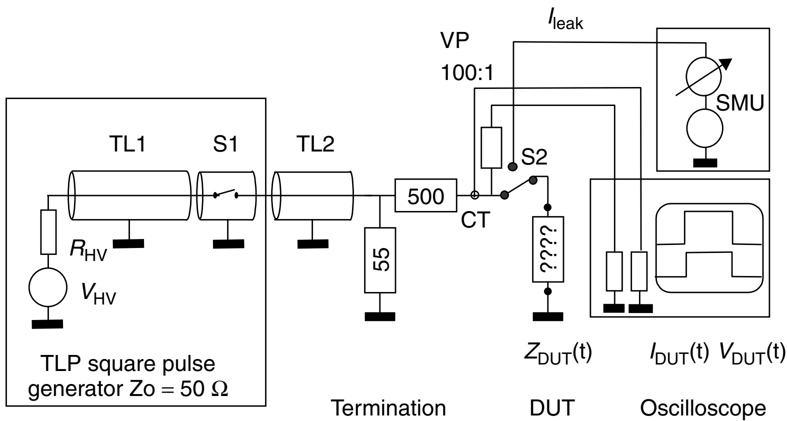


Figure 3.20 Current Source TLP $Z_S = 500\,\Omega$. (After [Maloney85], reproduced by permission of ESD Association)

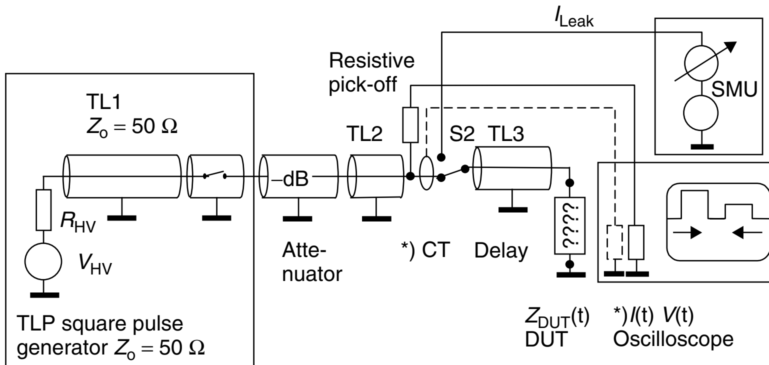


Figure 3.21 Time Domain Reflectometer TLP. (After [Bridgwood85], reproduced by permission of ESD Association)

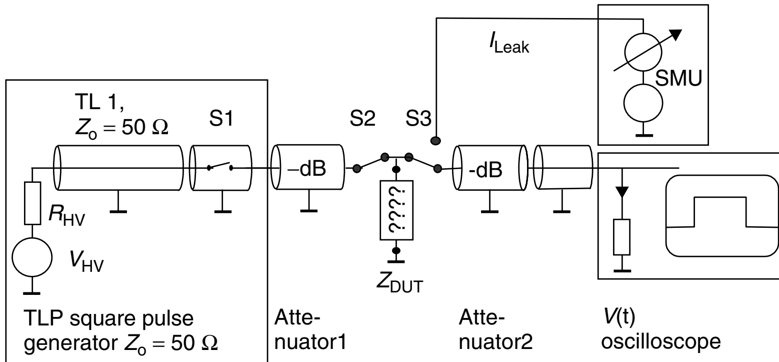


Figure 3.22 Time domain transmission TLP (TDT TLP). (After [Hyatt00], reproduced by permission of Shaker Verlag)

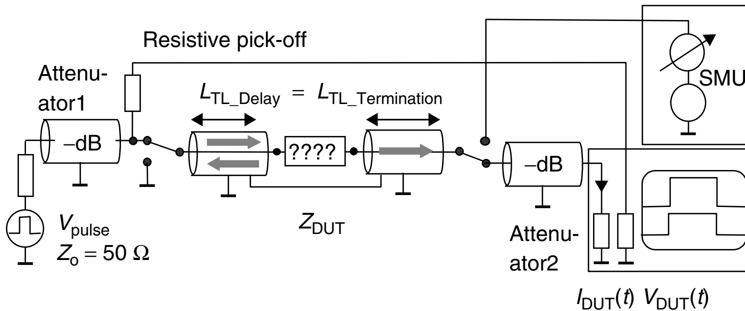


Figure 3.23 Time domain transmission reflectometer TLP or *Embedded DUT*™ TLP (TDRT TLP)

by the precharge voltage V_0 and the impedances of the source Z_S and the load Z_L. $V = V_0 \cdot Z_L/(Z_L + Z_S)$ For a matched impedance in the switch and the load, the amplitude of the voltage pulse would be half of the precharge voltage. Deviations from the ideal square shape of the pulse result from resistive and dielectric losses that are frequency-dependent as well as from variations in the impedance along the line through the whole system. Therefore, it is mandatory to employ cables and components that are well matched, as short as necessary, and with low losses throughout the whole system. However, a dedicated long transmission line may be employed to tailor the rise time in order to comply with HBM or MM. The homogeneous turn-on of protection devices and consecutively their clamping and failure thresholds may depend on the rising edge. The most critical element of the TLP-system is the relay switch. Implementations with a Reed relay surrounded by a metal sleeve or cylinder are commonly in use, although with standard dimensions of commercially available Reed relays the impedance

is in the order of 60 Ω. At least, for a significant pulse duration, for example, 100 ns, the resulting degradation of the leading corner of the pulse top and after the falling edge is acceptable. Another issue is the repeatability of the switching over the full voltage range. A low series resistance by means of a metallic contact should be established instantaneously and be maintained for the full duration of the pulse and possible reflections. Otherwise, a lower amplitude with a trailing pulse of the same polarity or even steps on the top of the pulse can be found. Mechanical contact bouncing is not an issue as it takes place in ms. The shorter the pulses become, the more emphasis needs to be put on the impedance of the signal path. Only few commercial coaxial components are explicitly rated for high voltages. Thus, it is a matter of experience and trial to find the most suitable relays, attenuators, and other components. Local safety regulations for the design and operation of high voltage equipment must be followed. In principle, the different TLP systems can be seen as a modular combination of impedance controlled RF components. Software controls the equipment and extracts the actual current through a device $I_{\mathrm{DUT}}(t)$ and the voltage across the device $V_{\mathrm{DUT}}(t)$ and derives the various data from these measured data. Necessary equipment are the controller, the high voltage source, the oscilloscope and a source measuring unit for leakage measurement and for optional additional voltage bias of the DUT. In combination with a wafer, prober devices can be characterized at wafer level, reducing the cost and providing additional insight in the distribution of the parameters across the wafer. Four implementations of a TLP-system will be discussed in the following paragraphs.

Figure 3.20 shows the *Current Source TLP*, sometimes referred to as TLP500, and mainly used to simulate HBM-stress by means of a 100 ns wide pulse. It is characterized by the termination resistor and the source resistor that forces the current trough a low ohmic DUT. Voltage and current are measured independently as close as possible at the DUT, what makes the method rather tolerant to pulse variations. Additional switches disconnect the DUT from the stress circuit and connect it to the DC-parameter analyzer. High parasitic elements at the DUT slow the achievable dV/dt and increase the probability for ringing. The matched termination eliminates multiple reflections between the open end of TL1 and the DUT. It generates 1 A/kV precharge voltage. The oscilloscope and the probes determine the accuracy nearly independent from the load impedance Z_{DUT}. For low DUT voltages associated with low ohmic protection elements, the voltage probe should be placed between the current probe and the DUT. Otherwise all measured and applied DUT-voltages require correction for the dynamic impedance of 1 ohm in the ground path. Resistors with low inductance can be used to verify the system.

Figure 3.21 shows two types of a Time Domain Reflectometer TDR-TLP-systems with and without a current transformer CT in the signal path. In order to measure the voltage and current at the DUT, they employ the principle that if an incident square pulse reaches the DUT at the end of a transmission line it is reflected depending on the impedance $Z_{\mathrm{DUT}}(t)$ of the DUT relative to the

impedance Z_0 of the transmission line according Equation 3.3.

$$V_{\text{reflected}}(t) = \frac{Z_{\text{DUT}}(t - t_{\text{delay}}) - Z_0}{Z_{\text{DUT}}(t - t_{\text{delay}}) + Z_0} * V_{\text{incident}}(t - t_{\text{delay}}) \tag{3.3}$$

This setup can maintain the 50 Ω-impedance from the generator to the device with minimum parasitic elements and pulse distortion. The system with the current transformer uses an oscilloscope with two channels and measures voltage and current reflected from the DUT independently with the accuracy provided by the probes and the oscilloscope after a calibration of the attenuation factors of the system. Alternatively, for a known impedance Z_0, the current $I(t)$ can be calculated from the relation $i(t) = V_{\text{incident}}(t)/Z_0$ for the incident pulse and $I(t) = -V_{\text{reflected}}(t)/Z_0$. Transmission Line TL3 between the resistive pick-off and the DUT delays the reflected pulse with respect to the incident pulse. Using equation 3.4 and 3.5, voltage $V_{\text{DUT}}(t)$ and current $V_{\text{DUT}}(t)$ at the DUT are calculated from the measured incident and reflected voltage pulse after a shift of the reflected pulse to the left by twice the one-way delay time t_{delay} and a correction for the attenuation of the resistive voltage pick-off and transmission lines [Gieser96][Gieser99].

$$V_{\text{DUT}}(t) = V_{\text{incident}}(t) + V_{\text{reflected}}(t - 2 * t_{\text{delay}}) \tag{3.4}$$

$$I_{\text{DUT}}(t) = \frac{V_{\text{incident}}(t) - V_{\text{reflected}}(t - 2 * t_{\text{delay}})}{Z_0} \tag{3.5}$$

Some uncertainty results from numerical effects for DUT impedances close to open and short and from distortion on the delay line TL3 in the phase of transition. Calibration to 0 Ω and a resistor together with correction improves the accuracy. If the reflected pulse is not completely separated from the incident pulse the incident pulse must be very repeatable and flat. An attenuator between the pulse generator and the DUT is recommended in order to reduce multiple stress caused by multiple reflections that depend on the DUT impedance. Other concepts employ a diode in series with a termination resistor in order to reduce reflections from low-impedant DUTs ($Z_{\text{DUT}} \leq Z_0$) at the open end of the transmission line [Maloney85]. This termination, that may even be switched in order to generate bipolar pulses, depends on polarity and should not be used for the characterization of oxides. A coaxial relay in the delay line allows to connect a DC-parameter analyzer to the DUT.

The *Time Domain Transmission method* [Hyatt00], shown in Figure 3.22, avoids the uncertainties associated with the dispersion of the reflected pulse signal; however, it requires that first a reference voltage pulse $V_{\text{chg}}(t)$ is captured for every voltage step without a DUT in place. The voltage $V_{\text{chg}}(t)$ equals the measured voltage $V(t)$ times the attenuation factor a. Then the device is inserted in the fixture and the resulting pulses $V_{\text{DUT}}(t)$ are captured for the same precharge voltage levels. This method can also be implemented with minimum parasitic elements and in particular distortion. It has to be proofed in advance that the impedance of the relay in the conductive stage is constant and repeatable, generating repeatable

pulses in the full voltage range. An attenuator between the pulse source and the DUT reduces multiple reflections and stabilizes the source impedance Z_0. For the first pulse of a series of decaying reflections the following Equations 3.6 and 3.7 are used to calculate the current through the DUT for the directly measured voltage $V_{\mathrm{DUT}}(t)$ attenuated to an amplitude safe for the oscilloscope. Two relays are necessary in order to isolate the DUT for leakage measurements between stress pulses.

$$Z_{\mathrm{DUT}}(t) = \frac{V_{\mathrm{DUT}}(t)}{V_{\mathrm{chg}}(t) - V_{\mathrm{DUT}}(t)} * Z_0 \tag{3.6}$$

$$I_{\mathrm{DUT}}(t) = \frac{V_{\mathrm{DUT}}(t)}{Z_{\mathrm{DUT}}(t)} \tag{3.7}$$

Figure 3.23 shows the *time domain transmission reflectometer, TDTR or "Embedded DUT*™*"* (*Trademark of Oryx Instruments Inc.* developed by Larry Edelson) embeds the DUT between the center conductors of two transmission lines of equal length and requires a two-channel oscilloscope. The grounded outer conductors of the two lines are connected to each other. The equal length of the lines is necessary in order to align the transmitted and the reflected signal on the screen of the oscilloscope. The length of the delay and termination line should exceed the length of the pulse generating TL1 in order to separate the reflected from the incident pulse at the pick-off. Employing Equation 3.8, the voltage $V_{\mathrm{DUT}}(t)$ is calculated from the reflected pulse considering attenuation factor of the pick-off and the transmission lines. With Equation 3.9 the current $I_{\mathrm{DUT}}(t)$ is calculated from the transmitted pulse voltage at the 50 Ω input resistor of the oscilloscope. Both signals may need some additional attenuation. This system has a source impedance of 100 Ω, as the DUT is in series with the 50 Ω-line connected to the oscilloscope. Although even a short does not generate reflections of opposite polarity, an attenuator between the pulse generator and the pick-off is recommended. In order to test the leakage, additional coaxial relays may be necessary for DC insulation. The calibration of the system is done at least with a short. With the symmetrical outline of the controlled impedance paths, this TLP method has a high potential for an implementation in an automated multipin TLP-test system.

$$V_{\mathrm{DUT}}(t) = a^* 2^* V_{refl}(t) \tag{3.8}$$

$$I_{\mathrm{DUT}}(t) = \frac{V(t)}{Z_0} \tag{3.9}$$

3.7.3 Correlation Issues

In the process of optimization, repeatable stress pulses in combination with the measurement of current and voltage would in principle be sufficient for the relative comparison of two protection elements. Looking at the HBM qualification test of

the product, correlation between TLP and the standard test methods employed for qualification became a concern for the worth of the TLP method. With the extended application of the TLP method also the results obtained with different TLP testers should be comparable. The TLP standard, which is currently developed by the ESD Association, will significantly contribute to this comparability as this standard is expected to be released before a large number of commercial test systems is employed. In view of equivalent turn-on characteristics and energy-related failure mechanisms, TLP-pulses with a rise time in the order of 5 ns and a pulse duration of 100 ns are typically used. They translate 0.66 A into almost 1 kV HBM. Depending on the device and its electrical and physical failure signature, a more detailed analysis may be necessary. The correlation issues already discussed in the Section 3.2 and Section 3.8 are also relevant for the TLP test.

A higher source impedance of the TLP system increases the resolution particularly close to V_{t1} and I_{t1} in principle. Additional parasitics associated with the high impedance setup affect the initial dV/dt at low currents. Oscillations result when the device IV and the load line multiply intersect [Maloney85].

The interconnect between the TLP-generator and both probe needles, which has no controlled impedance, must be kept to a minimum of few centimeters and also be in place for calibration. Even for longer pulse duration, it may affect the waveform and therefore the turn-on of the protection or the indicated voltage and current. Whether the pulse is injected via two symmetrical terminals employing a balun or asymmetrical with one coaxial line may have an influence on the device behavior. This effect should be negligible, if coaxial RF probes are used to contact the DUT and the substrate contact is always connected to the grounded shield of the probe.

In most TLP testers, the pin count and the parasitic elements in the discharge path and in the background of the IC are lower in comparison with HBM testers. Commercial TLP testers for higher pin counts are under development. Rise-time filters may be inserted into the transmission line behind the TLP pulse source in order to study rise-time effects. Ideally, the rising edge has a Gaussian shape. As a low-cost alternative, long transmission lines may be sufficient. Depending on the setup and the device to be stressed, electromagnetic interference between the stress terminals and control terminals, as an example the gate terminal, of the DUT can become an issue.

For each TLP type, the appropriate calibration technique must be applied in order to gain the correct attenuation factors of the system. Testing on the wafer, the contact resistance of the probe needles to the pads must be included in the calibration. For ultrafast pulses the techniques discussed in Section 3.6 should be employed.

Another correlation issue is the elimination or attenuation of multiple reflections, which may cause additional stress to the DUT. How severe the influence is depends on the tested structure.

References about the correlation of HBM and TLP can be found in [Maloney85] [Pierce88][Abderhalden91][Musshoff96][Barth00][Barth01][Keppens01]. Maloney

also discusses in detail critical issues with poly silicon resistors and aluminum lines [Maloney85].

The correlation between very fast rising, very short TLP pulses (VF-TLP) and CDM was investigated for a product like IC by Gieser *et al.* [Gieser96]. The conclusion was that the pulses of 3.3 ns duration generated the CDM-like failure signature of input leakage due to a broken gate oxide. No damage was found in the protection structure. Looking at the failure thresholds of the three inputs no correlation was found between CDM and VF-TLP. A detailed analysis of the circuit topography came to the conclusion that no correlation may be expected between the 1-pin CDM and the VF-TLP which involves two pins. Similar to HBM, the current flows between the stress pin and the reference pin, for example, V_{SS}. Very fast rising pulses may be injected into a single stress pin in order to charge the IC rapidly and with controlled impedance. This capacitively coupled TLP method (CC-TLP Pat. Pend.) will be further studied.

3.7.4 Test Procedure

The key application of the TLP is still the characterization of single protection elements, which may have additional control terminals. Many techniques for the characterization of protection elements and elements to be protected can be found in the literature. Some of them are in close relation with the extraction of parameters for the numerical simulation and require a specific test structure design. For example, applications that require gate or substrate bias and that measure the substrate current in addition to the stress current can be found in Russ and Wolf [Russ99][Wolf98]. Wolf demonstrates the extraction of the base transit time employing the very fast rising pulses of a VF-TLP [Wolf99]. Applications oriented to the analysis of a weak protection scheme can as an example be found in [Smedes01][Ting01].

The control program of a TLP-tester should provide excellent handling and analysis of significantly more data in comparison with HBM. It should be highly automated for routine characterization or process monitoring and must be very flexible for the expert user.

After calibration or verification of the setup, the DUTs are inserted into the test fixture or contacted on the wafer. The leakage current must be measured before and after each pulse. The reverse bias voltage must be carefully selected in a range that makes changes well visible and avoids additional stress during the leakage measurement. Typically, 110% V_{DD} are chosen. One pulse per step should be sufficient. The step size may start small in order to catch V_{t1}, I_{t1} and may be increased until V_{t2}, I_{t2} are reached.

It is recommended to test at least three structures, preferably from different positions on the wafer as subtle technological variations may strongly effect the ESD performance. Depending on the goal, the whole wafer may be mapped.

Typically, after checking the width scaling, the current I_{t2} is reported normalized to the width $A/\mu m$ or area of the protection element $A/\mu m^2$.

The rise time and the duration of the pulses can be varied either to study rise time or self-heating effects until failure occurs.

Overall, the TLP or equivalent techniques should be used to characterize also the structures to be protected, for example, the gate oxides of small transistors and transistors of outputs or clock drivers.

As TLP has a good chance to become accepted for the product qualification, test procedures equivalent to the HBM qualification are expected to emerge in the future.

3.8 FAILURE CRITERIA

The nature of ESD damage in ICs is such that the selection of the failure criteria and of their limits can have a significant influence on the ESD failure threshold and to the identification of an ESD problem in a design or technology. Issues to be resolved are whether the test method and the applied criterion are able to indicate the failure consistently at a certain stress level and whether a subtle degradation might disappear or lead to an early functional failure. This section should be read in view of the ESD-related failure mechanisms in Chapter 8.

Traditionally, attention has been paid to an increase of leakage current, for example, 1 μA@3.5 V for a specified reverse bias voltage at the pin under test caused by ESD. Others are the increase of the power supply current caused by an internal leakage path beyond the periphery or the shift of, for example, reference voltages. Leakage can change the performance of the integrated circuit such as access times in SRAMS, offsets in differential amplifiers and power consumption. Reduced metallization cross-sections in new technologies increase the risk of an open or near-open due to fused metal. At least for the ESD qualification of a product [ESDA-HBM01], the stressed circuit must be subjected to a full-functional production test including all AC- and DC parameters to ensure it still meets all the product specifications as they are the common reference for manufacturer and user. Despite the big effort, even this test cannot always discover a walking wound condition that could lead to an early failure. In CMOS-circuits I_{ddq} tests have been very helpful to identify internal leakage paths beyond the periphery of the circuit. For the latest technologies, the applicability of I_{ddq} tests is limited, as leakage through ultrathin gate oxides and extremely short channels may hide subtle ESD-related leakage. It may not always be possible to do functional tests after each ESD stress in the development phase.

It is recommended that during ESD protection circuit evaluation and design development, all changes in leakage current are carefully monitored. It is still debated whether any significant increase of leakage out of the noise floor of the tester should be considered as a fail or a low fixed threshold should be set with respect to the reaquirements of the application and possible reliability issues. It has been shown that the distribution of the leakage current at a given stress level is linked to the distribution of the ESD failure threshold

[Amerasekera90][Keppens01][Notermans98][Stadler97][Russ99], and a monitor of the leakage current can be used to characterize the ESD behavior of an IC.

In most of the automated ESD test systems for HBM, MM, SDM, the test of the leakage criterion at the stressed pin and at other pins is implemented. Because of the above mentioned, it should just be used as an indication for failure of an IC. In particular, the leakage testing of output pins requires to switch them into an off or Hi-Z state to become able to measure the leakage in a more sensitive voltage range, well above the forward bias voltage 0.5 V. The leakage criterion is standard for the characterization of test structures by means of TLP, HBM, or MM. Other criteria and methods might be necessary for the development of RF-structures.

Looking briefly into the root cause for leakage and its effects. In most cases, leakage current has been attributed to more or less subtle damage in pn-junctions [Amerasekera90][Ohtani90][Amerasekera92] or in dielectric layers. As an example for the leakage in pn-junctions Figure 3.24 shows a series of typical $I - V$ curves measured after ESD stress. Curve 1 shows a typical $I - V$ curve for an undamaged device. Curves 2, 3, and 4 show varying degrees of damage after different ESD stress. Typical monitoring voltages for the leakage current are either 0.4 V or 5.5 V (or 3.6 V for 3.3 V parts), while the choice of leakage current can range from 100 nA to 10 µA. It is also possible to select a percentage change in leakage current (10% to 50%). From Figure 3.24, we see that for Curve 3, the choice of a 1 µA leakage current will signal failure at 5.5 V, but 100 nA at 0.4 V will not show a failure. Similarly, Curve 2 will pass most failure criteria, but in reality the device has been damaged when compared to Curve 1. There is no certainty that such a pin will show latent effects and be a reliability hazard in device operation, but it has definitely been damaged. Hot carriers injected from the pn-junction into the gate

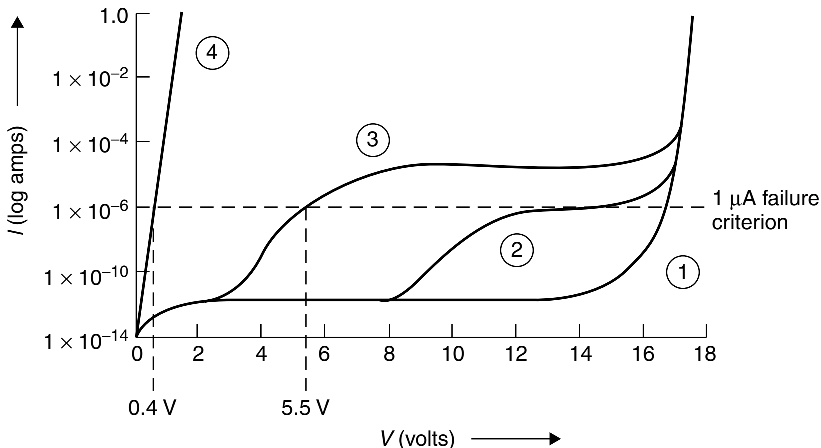


Figure 3.24 Leakage curves measure after ESD stress across an nMOS transistor. Curve 1 to indicate an increased level of damage severity

oxide during an ESD stress may not only shift the threshold voltage [Aur88] but also cause early failures of the gate oxide during life test [Reiner00].

In the case of gate oxide leakage, even very low leakage currents of some 100 pA after CDM-stress, well below data sheet specification, have significantly increased failure rates in life tests [Gieser94B][Colvin93]. Such low currents are difficult and time consuming if not impossible to detect during production test of a digital circuit. When metal gates were used in old CMOS-processes it was found that one pulse has shorted the gate oxide while the next fused this short and the leakage current dropped again.

Subtle damaged pn-junctions as well as gate oxides show a decrease of leakage current over time during storage at room and even more significant at high temperatures. Prestress levels are not reached; however, devices may recover into specification. This process is called *cold healing* and motivates a time limit between ESD test and product test for qualification. In this view, the product test should always start at the lowest temperature, if different temperatures have to be applied during the production test.

In summary, one must be careful when comparing ESD failure thresholds obtained with different failure criteria, and with the choice of the failure criteria itself. Failure analysis efforts benefit from a reproducible failure criterion. In some cases it was useful to stress devices with significantly higher stress currents in order to obtain a more significant physical failure signature for localization. This failure site can be studied in more detail for devices with subtle leakage.

3.9 SUMMARY

Test methods and their implementation in ESD testers have a major impact as well on the results of the product qualification as on the results and the efficiency of the ESD protection development. In view of correlation and standardization, the metrology of very fast rising pulses has been discussed. The awareness has been raised that failure criteria beyond the data sheet specification would be necessary in order to protect against walking wounded devices, although no general and practical solution to this problem could be offered. For pin-to-pin HBM-testing, high-pin count devices require special attention and solutions. The parasitics in the discharge path increase with pin count and the matching between the different testers becomes an even bigger challenge. Despite improvement due to better ESD test standards the parasitics in the discharge path and their interaction with the DUT are still the major issue for correlation problems. As MM does not reproduce most (F)CDM-type field failures, it may be used only for specific purposes in the future. The necessity of performing an ESD qualification, at least in the two distinct domains of the pin-to-pin HBM and the ultrafast single-pin (F)CDM, was pointed out in order to cover the variety of field failures. Some background was provided identifying a general weakness of current (F)CDM standards in the lack of a traceable metrology for ps-discharges as well as in the physics of the

air discharge. Ongoing work of the standardization bodies and individuals has and will identify and solve other issues. One possible solution for a better correlation and repeatability of CDM could be to control the impedance of the discharge path, including the switch, and possibly to slow the waveform to a degree that it can be measured traceably by existing equipment. This could be in the order of 200 ps, or faster, if repetitive techniques can be applied. Higher but more reproducibly obtainable pass levels for ICs with respect to CDM may be expected as the peak current for a given voltage decreases. The different influence of packages for HBM and (F)CDM was discussed. It has been clearly demonstrated for HBM that circuit simulation of the combination of the tester and the DUT can provide much deeper insight into the interaction actually taking place during the stress. A prerequisite are circuit models verified by means of a reliable time domain measurement and reference experiments with well-defined source and load conditions. An inevitable tool for the high-current characterization of the integrated structures in ESD relevant time domains of the HBM and (F)CDM is the transmission line pulse generator. An integral part of the transmission line pulsing procedure is the measurement of the leakage current development for increasing stress. Beyond the selection of the best suitable protection element for a specific application, the parameters for the modeling can be obtained. Four modular TLP setups were discussed and compared in principle. And although square pulses are not expected to occur frequently in reality, TLP with well-defined transmission lines has some long-term potential to become a qualification method and to replace the traditional stress methods.

BIBLIOGRAPHY

[Abderhalden91] J. Abderhalden, “Untersuchungen zur Optimierung von Schutzstrukturen gegen elektrostatische Entladungen in integrierten CMOS-Schaltungen”, Dissertation ETH No. 9251, Eidgenoessische Technische Hochschule, Zuerich, Schweiz, 1991.

[AEC-CDM01] Automotive Electronics Council, Charged Device Model Electrostatic Discharge Test; AEC-Q100-011 Rev A, 2001.

[AEC-HBM01] Automotive Electronics Council, Human Body Model Electrostatic Discharge Test; AEC-Q100-002 - REV-C, 2001.

[AEC-MM01] Automotive Electronics Council, Machine Model Electrostatic Discharge Test; AEC-Q100-03 Rev E, 2001.

[Amerasekera90] E.A. Amerasekera, L.J. van Roozendaal, J. Abderhalden, J.J.P. Bruines and L. Sevat, “An analysis of low voltage ESD damage in advanced CMOS processes”, in *Proc. 12th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 143–150, 1990.

[Amerasekera91] A. Amerasekera, L. van Roozendaal, J. Bruines and F. Kuper, “Characterization and modeling of second breakdown in NMOST’s for the extraction of ESD-related process and design parameters”, in *IEEE T. Electron Dev.*, **ED-38**, 2161–2168, 1991.

[Amerasekera92] E.A. Amerasekera, W. van den Abeelen, L.J. van Roozendaal, M. Hannemann and P.J. Schofield, "ESD failure modes: Characteristics, mechanisms and process influences", in *IEEE T. Electron Dev.*, **ED-39**(2), 1992.

[Anderson98] W.R. Anderson, J.J. Montanaro and N.J. Howorth, "Cross-Referenced ESD Protection for Power Supplies", in *Proc. 20th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 86–95, 1998.

[Andrew94] J.R. Andrew, "PSPL Calibration and NIST traceability", Picosecond Pulse Lab PSPL-Application Note AN-6, April 1994.

[Aur88] S. Aur, A. Chatterjee and T. Polgreen, "Hot electron reliability and ESD latent damage", in *IEEE T. Electron Dev.*, **35**(12), 2189–2193, 1988.

[Avery87] L. Avery, "Charged device model testing: Trying to duplicate reality", in *Proc. 9th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 88–92, 1987.

[Barth00] J. Barth, J. Richner, K. Verhaege and L.G. Henry, "TLP-Calibration, correlation, standards, and new techniques", in *Proc. 22nd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 85–96, 2000.

[Barth01] J. Barth and J. Richner, "Correlation considerations: Real HBM to TLP and HBM testers", in *Proc. 23rd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 453–460, 2001.

[Bossard80] P.R. Bossard, R.G. Chemelli and B.A. Unger, "ESD Damage from Triboelectrically Charged IC Pins", in *Proc. 3rd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 17–22, 1980.

[Bridgwood85] M. Bridgwood and R. Kelley, "Modeling the Effects of Narrow Impulsive Overstress on Capacitive Test Structures", in *Proc. 7th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 84–91, 1985.

[Brodbeck96] T. Brodbeck, H. Bauch, X. Guggenmos and R. Wagner, "Reproducibility of Field Failures by ESD Models — Comparison of HBM, Socketed CDM and Non-Socketed CDM", in *Microelectronics Reliability*, Vol. 36, No.11/12, pp. 1719–1722, 1996.

[Brodbeck00] T. Brodbeck, "Influence of the charging effect on HBM ESD Device testing", in *Proc. 22nd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 66–71, 2000.

[Brodbeck01] T. Brodbeck, "HBM ESD Test von elektronischen Bauelementen mit hoher Anschlusszahl", in *Proc. 7th ESD-Forum*, Interessengemeinschaft ESDe.V., Nördlingen, pp. 77–85, 2001.

[Brown72] W.D. Brown, "Semiconductor device degradation by high amplitude current pulses", in *IEEE Trans. Nucl. Sci.*, **NS-19**, 68–75, 1972.

[Carey98] R. Carey and L. DeChiaro, "An experimental and theoretical consideration of physical design parameters in field-induced charged device model simulators and their impact upon measured withstand voltages", in *Proc. 20th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 40–53, 1998.

[Chaine94] M. Chaine, C.T. Liong and H.F. San, "A Correlation Study Between Different Types of CDM Testers And Real Manufacturing In-Line Leakage Failures", in *Proc. 16th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 63–68, 1994.

[Chaine98] M. Chaine, K. Verhaege, L. Avery, M. Kelly, H. Gieser, K. Bock, L.G. Henry, T. Meuse, T. Brodbeck and J. Barth, "Investigation into

Socketed CDM (SDM) Tester Parasitics", in *Proc. 10th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 301–310, 1998.

[Chemelli85] R. Chemelli and L. De Chiaro, "The Characterization and Control of Leading Edge Transients from Human Body Model ESD Simulators", in *Proc. 7th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 155–162, 1985.

[Colvin93] J. Colvin, "The identification and analysis of latent gate oxide leakage on CMOS-input gates", in *Proc. 15th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 109–116, 1993.

[Diaz92] C. Diaz, S. Kang, C. Duvvury and L. Wagner, "Electrical Overstress (EOS) Power Profiles: A Guideline to Qualify EOS Hardness of Semiconductor Devices", in *Proc. 14th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 88–94, 1992.

[Duvvury88] C. Duvvury, R. Rountree and O. Adams, "Internal chip ESD phenomena beyond the protection circuit", in *Proc. 26th IRPS*, pp. 19–25, 1988.

[Egger93] P. Egger and H. Gieser, "Mehrdimensionale Studie zum Charged Device Model Test", in *Proc. 3rd ESD-Forum*, pp. 194–202, 1993.

[EN-HBM96] European Standard, "EN 61000-4-2: Part 4, Test and Measurement Techniques — Section 2: Electrostatic Discharge Immunity Test" based on IEC 1000-4-2 (1992), 1996.

[ESDA-CDM99] ESD Association, "ESD STM5.3.1-1999: Standard Test Method for Electrostatic Discharge Sensitivity Testing-Charged Device Model (CDM) Component Level", 1999.

[ESDA-HBM01] ESD Association WG 5.1, "ESD STM5.1-2001: Standard Test Method for Electrostatic Discharge Sensitivity Testing- Human Body Model (HBM) Component Level", 2001.

[ESDA-MM98] ESD Association WG 5.2, "ESD STM5.2-1998: Standard Test Method for Electrostatic Discharge Sensitivity Testing- Machine Model (MM) Component Level", 1998.

[ESDA-MM-TR00] ESD Association WG 5.2, "ESD TR 10-00: Machine Model (MM) Electrostatic Discharge (ESD) Investigation — Reduction in pulse number and delay time", 2000.

[ESDA-SDM-TR00] ESD Association WG 5.3.2, "ESD TR 08-00: Socket Device Model (SDM) Tester, an applications and technical report", Revision 1.1, 2000.

[ESDA-TLP] ESD Association WG5.5, "ESD DS5.5-200x: Standard Practice for Electrostatic Discharge Sensitivity Testing-Transmission Line Pulse (TLP) Component Level" Work in Progress, 200x.

[Fukuda86] Y. Fukuda, S. Ishiguro and M. Takahara, "ESD Protection Network Evaluation by HBM and CPM (Charged Package Method)", in *Proc. 8th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 193–199, 1986.

[Gans90] W.L. Gans, "Dynamic calibration of waveform recorders and oscilloscopes using pulse standards", in *IEEE Trans. on Instrumentation and Measurement*, **39**(6), 952–958, 1990.

[Gieser94A] H. Gieser and P. Egger, "Influence of Tester Parasitics On Charged Device Model Failure Thresholds", in *Proc. 16th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 69–84, 1994.

[Gieser94B] H. Gieser, P. Egger, M. Herrmann, J. Reiner and A. Birolini, "CDM-only field degradation and its reliability aspect", pp. 371–378, 1993 and *Quality and Reliability Engineering International*, **10**(4), pp. 341–350, July, August, 1994.

[Gieser95] H. Gieser, "Tutorial Device Level Testing", in *Tutorial of 7th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, 1995.

[Gieser96] H. Gieser and M. Haunschild, "Very-fast transmission line pulsing of integrated structures and the charged device model", in *Proc. 18th EOS/ESD Symposium*, pp. 85–94, 1996 and in *Proc. 7th Reliability Conference of Japan*, pp. 17–26, 1997 and in *IEEE Trans. on Components, Packaging, and Manufacturing Technology*, Part C, Vol. 21, No. 4, pp. 278–285, 1998.

[Gieser99] H. Gieser, "Verfahren zur Charakterisierung von integrierten Schaltungen mit sehr schnellen Hochstromimpulsen (Methods for the characterisation of integrated circuits employing very fast high current impulses", in *Dissertation Technische Universitaet Muenchen TUM*, Shaker-Verlag, Aachen, Germany, 1999.

[Gossner97] H. Gossner and T. Brodbeck, "Dangerous parasitics of Socketed CDM ESD testers", in *Microelectronics Reliability*, Vol. 37, No. 10/11, pp. 1465–1468, 1997.

[Henry96] L.G. Henry, H. Hyatt, J. Barth, M. Stevens and T. Diep, "Charged Device Model (CDM) metrology: limitations and problems", in *Proc. 18th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 167–179, 1996.

[Henry99] L.G. Henry, M. Kelly, T. Diep and J. Barth, "Issues Concerning CDM ESD Verification Modules- The Need to Move to Alumina", in *Proc. 21st EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 203–211, 1999.

[Henry98] L.G. Henry, M. Kelly, T. Diep and J. Barth, "The importance of standardizing CDM ESD test head parameters to obtain data correlation", in *Proc. 20th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 72–84, 1998.

[Hull88] R. Hull, "Stress-Hardening of Damage Thresholds due to Cumulative ESD Pulse Testing", in *Proc. 10th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 39–46, 1988.

[Hyatt00] H. Hyatt, J. Harris, A. Alanzo and P. Bellew, "TLP-Measurements for verification of ESD protection device response", in *Proc. 22nd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 111–120, 2000.

[JEDEC-FCDM00] JEDEC Solid State Technology Association, "JESD22-C101-A: Field-Induced Charged Device Model Test Method for Electrostatic Discharge-Withstand Thresholds of Microelectronic Components", 2000.

[JEDEC-HBM00] JEDEC Solid State Technology Association, "JESD22-A114-B: Electrostatic Discharge (ESD) Sensitivity Testing Human Body Model (HBM)", 2000.

[JEDEC-MM97] JEDEC Solid State Technology Association, "JESD22-A115-A: Electrostatic Discharge (ESD) Sensitivity Testing Machine Model (MM)", 1997.

[Kagerer98] A. Kagerer and T. Brodbeck, "Influence of the device package on the results of CDM tests — consequences for tester characterization and test procedures", in *Proc. 20th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 320–327, 1998.

[Keppens01] B. Keppens, V. de Heyn, M. Natarajan Iyer and G. Groeseneken, "Contribution to Standardization of Transmission Line Pulse Methodology", in *Proc. 23rd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 461–467, 2001.

[Kropf93] R. Kropf, C. Russ, R. Kolbinger, H. Gieser and S. Irl, "Zeitaufgeloeste Untersuchung des Snapback-Verhaltens eines ESD-Schutztransistors", in *Proc. 3rd ESD-Forum*, VP-Verlag, Herrenberg, pp. 19–26, 1993.

[Lee88] T. Lee, G. Thome and T. Guthrie, "Discrete Semiconductor Testing to 40 kV", in *Proc. 10th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 119–123, 1988.

[Lin87] D.L. Lin, M. Strauss and T. Welsher, "On the Validity of ESD Threshold Data Obtained Using Commercial Human Body Model Simulators", in *Proc. 25th IRPS*, pp. 77–87, 1987.

[Maloney85] T. Maloney and N. Khurana, "Transmission Line Pulsing Techniques for Circuit Modeling of ESD Phenomena", in *Proc. 8th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 49–54, 1985.

[Maloney01] T.J. Maloney, D.H. Cho, S.S. Poon and B. Lisenker, "Improving the balanced coaxial differential probe for high voltage pulse measurements", in *Proc. 23rd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 398–407, 2001.

[Musshoff96] C. Musshoff, H. Wolf, H. Gieser, P. Egger and X. Guggenmos, "Risetime effects of HBM and square pulses on the failure thresholds of GGNMOS transistors", in *Proc. 9th ESREF Symposium*, pp. 1743–1746, 1996.

[Nahman83] N.S. Nahman, "Picosecond-domain waveform measurements: Status and future directions", *IEEE Trans. on Instrumentation and Measurement*, **IM-32**(1), 117–123, 1983.

[Notermans98] G. Notermans, P. de Jong and F. Kuper, "Pitfalls when correlating TLP, HBM and MM testing", in *Proc. 20th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 170–176, 1998.

[Ohtani90] S. Ohtani and M. Yoshida, "Model of Leakage Current in LDD Output MOSFET Due to Low-Level ESD Stress", in *Proc. 12th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 177–181, 1990.

[Olney96] A. Olney, "A Combined Socketed and Non-socketed CDM Test Approach for Eliminating Real-World CDM Failures, EOS/ESD Symposium Proceedings", in *Proc. 18th EOS/ESD-Symposium*, ESD Association, Rome, NY, USA, pp. 62–75, 1996.

[Pierce88] D.G. Pierce, W. Shiley, B.D. Mulcahy, K.E. Wagner and M. Wunder, "Electrical overstress testing of a 256 K UVEPROM to rectangular and double exponential pulses", in *Proc. 10th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 137–146, 1988.

[Reiner95] J.C. Reiner, "Latent gate oxide damage induced by ultra-fast electrostatic discharge", in *Dissertation Eidgenoessisch Technische Hochschule (ETH) Zuerich*, Hartung Gorre Verlag, Konstanz, Germany, 1995.

[Reiner00] J.C. Reiner, T. Keller, H. Jaeggi and S. Mira, "Microelectronics and Reliability", Elsevier Science Ltd., Vol. 40, No. 8–10, Aug.–Oct. 2000, pp. 1619–1630.

[Reiner01] J. Reiner and T. Keller, "Relevance of contact reliability in HBM-ESD test equipment", in *Microelectronics and Reliability*, Elsevier, Vol. 41, No. 9–10, Sept.–Oct. 2001, pp. 1397–1401.

[Renninger88] R.G. Renninger, M.-C. Jon, D.L. Lin, T. Diep and T.L. Welsher, "A Microwave-Bandwidth Waveform Monitor for Charged Device

Model Simulators", in *Proc. 10th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 162–171, 1988.

[Renninger89] R.G. Renninger, M.-C. Jon, D.L. Lin, T. Diep and T.L. Welsher, "A field-induced charged device model simulator", in *Proc. 11th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 59–71, 1989.

[Renninger91] R.G. Renninger, "Mechanisms of Charged Device Model Electrostatic Discharges", in *Proc. 13th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 127–143, 1991.

[Roozendaal90] L.J. van Roozendaal, E.A. Amerasekera, P. Bos, W. Baelde, F. Bontekoe, P. Kersten, E. Korma, P. Rommers, P. Krys, U. Weber and P. Ashby, "Standard ESD testing", in *Proc. 12th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 119–130, 1990.

[Russ93] C. Russ, H. Gieser, P. Egger and R. Kropf, "Compact electrothermal simulation of ESD protection elements", in *Proc. 6th ESREF-Symposium*, pp. 395–400, 1993.

[Russ94] C. Russ, H. Gieser and K. Verhaege, "ESD protection elements during HBM-ESD-stress tests -further numerical and experimental results", in *Proc. 16th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 96–105, 1994.

[Russ99] C. Russ, "ESD Protection Devices for CMOS Technologies:Processing Impact, Modeling, and Testing Issues", in *Dissertation Technische Universitaet Muenchen TUM*, Shaker-Verlag, Aachen, Germany, 1999.

[Shaw86] R.N. Shaw, "A Programmable Equipment for ESD Testing to the Charged Device Model", in *Proc. 8th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 232–237, 1986.

[Smedes01] T. Smedes, R.S. Ruth, A.J. Huitsing and R.M.D.A. Velghe, "The application of transmission line pulse testing for the ESD analysis of integrated circuits", in *Proc. 23rd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 426–434, 2001.

[Smith93] D.C. Smith, *High frequency measurements and noise in electronic circuits: a practical guide of successful techniques for designing, debugging, and reducing noise*, Kluwer Academic Publishers, Dordrecht, NL, 1993.

[Stadler97] W. Stadler, X. Guggenmos, P. Egger, H. Gieser and C. Musshoff, "Dopes the ESD failure current obtained by Transmission Line Pulsing always correlate to Human Body Model tests?", in *Proc. 19th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 366–372, 1997.

[Strauss87] M. Strauss, D. Lin and T. Welsher, "Variations in Failure Modes and Cumulative Effects Produced by Commercial Human Body Model ESD Simulators", in *Proc. 9th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 59–63, 1987.

[Sullivan85] S. Sullivan and D. Underwood, "The Automobile Environment: Its Effects on Human Body ESD Model", in *Proc. 7th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 103–106, 1985.

[Tanaka97] M. Tanaka, K. Okada and M. Sakimoto, "Study of correlations between real ESD phenomena and test methods for semiconductor devices", in *Proc. 7th Reliability Conference of Japan*, pp. 163–170, 1997.

[Ting01] L.-M. Ting, C. Duvvury, O. Trevino, J. Schichl and T. Diep, "Integration of TLP analysis for ESD troubleshooting", in *Proc. 23rd EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 445–452, 2001.

[Verhaege93] K. Verhaege, P. Roussel, G. Groesenecken, H. Maes, H. Gieser, C. Russ, P. Egger, X. Guggenmos and F. Kuper, "Analysis of HBM ESD Testers and Specifications Using 4th Order Lumped Element Model", in *Proc. 15th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 129–138, 1993.

[Verhaege94] K. Verhaege, G. Groeseneken, H. Maes, P. Egger and H. Gieser, "Influence of Tester, Test Method and Device Type on CDM ESD Testing", in *Proc. 16th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 49–62, 1994.

[Verhaege96] K. Verhaege, C. Russ, G. Groeseneken, D. Robinsohn-Hahn, D. Lin, M. Farris, J. Scanlon and J. Veltri, "Recommendations to further improvements of HBM ESD Component Level Test Specification", in *Proc. 18th EOS/ESD-Symposium*, ESD Association, Rome, NY, USA, pp. 40–53, 1996.

[Wada95] T. Wada, "Study of ESD Evaluation Methods for Charged Device Model", in *Proc. 17th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 186–193, 1995.

[Wolf98] H. Wolf, H. Gieser and W. Stadler, "Bipolar Model Extension for MOS Transistors Considering Gate Coupling Effects in the HBM ESD Domain", in *Proc. 20th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 271–280, 1998.

[Wolf99] H. Wolf, H. Gieser and W. Wilkening, "Analyzing the switching behavior of ESD-protection transistors by very fast transmission line pulsing", in *Proc. 21st EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 28–37, 1999.

[Wu00] J. Wu, P. Juliano and E. Rosenbaum, "Breakdown and Latent Damage of Ultra-Thin Gate Oxides under ESD", in *Proc.22th EOS/ESD Symposium*, ESD Association, Rome, NY, USA, pp. 287–295, 2000.

[Wunsch68] D.C. Wunsch and R.R. Bell, "Determination of threshold failure levels of semiconductor diodes and transistors due to pulse voltages", in *IEEE Trans. Nucl. Sci.*, NS-15, p. 244, 1968.

4 Physics and Operation of ESD Protection Circuit Elements

Ajith Amerasekera

4.1 INTRODUCTION

Devices subjected to high current and high-voltage stress operate well outside their normal operating ranges and their modes of operation are very different under these conditions compared to normal operating conditions. An understanding of the high current device behavior is essential in analyzing the phenomena taking place in the IC during an ESD or high current stress event. In this chapter we will look at the behavior of the main circuit elements used in ESD protection circuits. We will discuss the physics of these devices under high current conditions and the main parameters governing their performance. ESD is a very high current event that occurs for very short durations, and in this chapter we will direct the analysis to that region of operation. A basic understanding of semiconductor device physics is assumed throughout this chapter [Ghandhi77][Sze81][Muller86].

The principles of high current operation under ESD conditions are similar to power semiconductor devices [Ghandhi77]. The important devices to be considered are PN diodes under both forward- and reverse-biased conditions, the nMOS and pMOS transistor, the bipolar transistor *npn* and *pnp*, and semiconductor resistor elements. In addition, silicon-controlled rectifiers (SCR) are used as ESD protection circuits, and are also present as parasitic elements that trigger under ESD conditions.

4.2 RESISTORS

We begin the analysis with semiconductor resistors as they provide an opportunity to look at the basic device physics that also apply to diodes and transistors.

In general, N-type resistors are preferred for use with ESD circuits in processes that use P-substrates and P-type resistors with N-substrates. The conductivity of a semiconductor is given by

$$\sigma = nq\mu_n + pq\mu_p \tag{4.1}$$

where μ_n and μ_p are the electron and hole mobilities and n and p are the electron and hole concentrations, respectively. $q = 1.602\times 10^{-19}$, C is the electronic charge, n and p are given by the equations

$$n = n_i \exp\left[\frac{q(\psi - \phi)}{kT}\right] \tag{4.2}$$

$$p = n_i \exp\left[\frac{q(\phi - \psi)}{kT}\right] \tag{4.3}$$

ϕ and ψ are the potentials corresponding to the Fermi energy, E_F, and the Fermi energy for the intrinsic semiconductor, E_i, respectively. Hence, $\phi \equiv -E_\mathrm{F}/q$ and $\psi \equiv -E_i/qn_i$ is the intrinsic carrier concentration given by

$$n_i = \sqrt{N_\mathrm{C}N_\mathrm{V}} \exp\left(\frac{-E_\mathrm{g}}{2kT}\right) \tag{4.4}$$

where N_C and N_V are the effective density of states in the conduction and valence bands, respectively. For silicon, we get

$$n_i = 1.69 \times 10^{19} \left(\frac{T}{300}\right)^{3/2} \exp\left(\frac{-E_\mathrm{g}}{2kT}\right) \tag{4.5}$$

In N-type resistors at low injection current levels, the current density is given by

$$J = J_n = N_\mathrm{B} q \mu_n E \tag{4.6}$$

$$= N_\mathrm{B} q v_\mathrm{d} \tag{4.7}$$

as the hole current is negligible. N_B is the background doping concentration and E is the electric field. $v_\mathrm{d} = \mu E$ is the drift velocity of carriers. As the voltage is increased E increases and so does J and the resistor is *ohmic*, that is, it shows a linear dependence between current and voltage. As the voltage is raised further the field increases and so does J in accordance with Equation 4.6 until at $E = 10^4\,\mathrm{V\,cm^{-1}}$ the electron drift velocity saturates at $v_\mathrm{s} \approx 10^7\,\mathrm{cm\,s^{-1}}$. Further increasing the voltage serves only to increase the electric field with no increase in J and

$$J = J_\mathrm{sat} = N_\mathrm{B} q v_\mathrm{s} \tag{4.8}$$

J_sat is a function of the doping concentration and will not be observed for either low $N_\mathrm{B}(< 10^{14}/\mathrm{cm}^3)$ or very high doping levels when $N_\mathrm{B} \approx 10^{20}/\mathrm{cm}^3$ [Hower70][Caruso74][Amerasekera93]. N-well resistors with $N_\mathrm{B} \approx 10^{17}/\mathrm{cm}^3$ will have $J_\mathrm{sat} \approx 10^5\,\mathrm{A\,cm^{-2}}$, which translates to a current of about 10 mA in a 20-µm

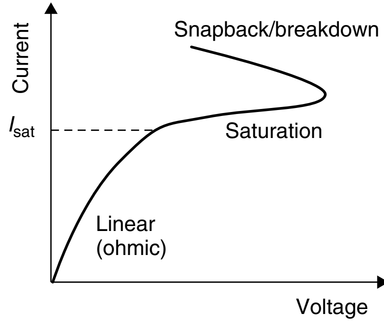


Figure 4.1 The high current $I-V$ curve for an n-type diffused resistor showing the ohmic and saturated regions

wide resistor. For very $n+$ diffusions with $N_B = 10^{20}/\text{cm}^3$, $J_{sat} \approx 10^8\,\text{A cm}^{-2}$, which gives I_{sat} for a 20 μm wide resistor of $\approx$ 10 A, the saturation effect will not be seen. The high current $I-V$ curve for a 20-μm wide n-well resistor is shown in Figure 4.1. As the voltage is increased even further in the saturation region, the E-field eventually reaches the impact ionization threshold and holes are generated. When the generated hole current becomes large enough to contribute to the total current the voltage decreases and a negative resistance or *snapback* characteristic is observed as shown in Figure 4.1. A rough calculation of the maximum voltage across the resistor may be done by assuming that impact ionization begins at $\approx$ 150 kV cm^{-1}. Then the maximum voltage for a length of 2 μm is 30 V. The amount of impact ionization required to cause snapback depends on N_B, which would influence the maximum voltage across the resistor before snapback occurs.

Snapback in the resistor could also occur due to heating in the saturation region [Amerasekera93]. For more highly doped resistors, self-heating is more likely to result in snapback than avalanche breakdown. Once snapback occurs the current in the resistor is carried by both holes and electrons. Further heating will eventually result in the melt temperature of silicon being reached and damage occurring. Analytical studies and simulations have also shown that current constrictions in the resistor may occur in the negative resistance region [Khurana66][Hower70][Yang93]. Current constrictions would result in filaments forming and silicon melting taking place at lower injection current levels.

4.3 DIODES

The simplest voltage clamping device is the diffusion diode. Bias conditions and the associated $I-V$ curve for a forward-biased diode are shown in Figure 4.2 and for a reverse-biased diode in Figure 4.3. In the forward direction the diode begins appreciable current conduction at $\approx$ 0.5 V, and has an on-resistance R_{on} per unit width of between 20 Ω μm^{-1} and 100 Ω μm^{-1}. Under reverse-biased conditions,

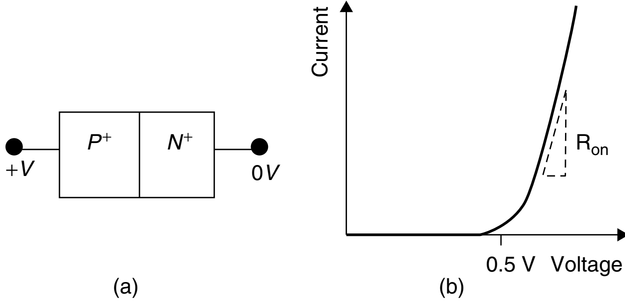


Figure 4.2 (a) A simple PN diode structure in the forward-biased condition. (b) Forward-biased $I-V$ curve for a PN diode. Appreciable current conduction begins at ~ 0.5 V and the dynamic on-resistance is given by $R_{on} = \Delta V/\Delta I$

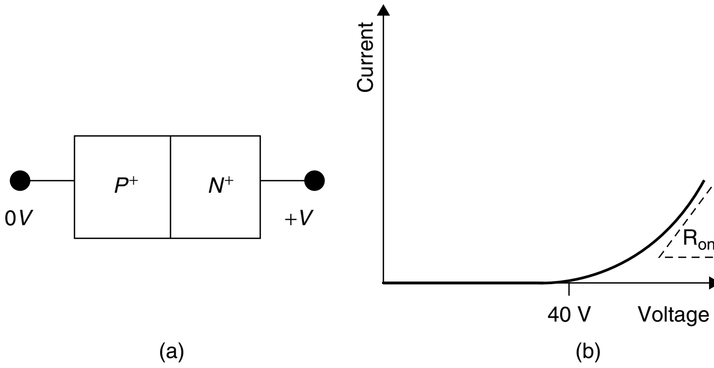


Figure 4.3 (a) Reverse-bias condition for a simple PN diode. (b) Reverse-biased $I-V$ curve for a PN diode. Avalanche breakdown begins at about 40 V for this diode, with a dynamic on-resistance of R_{on}

the current conduction begins when the junction goes into avalanche breakdown. The avalanche breakdown voltage, BV_{av}, is a function of the n and p doping concentrations and in a submicron process BV_{av} is around 10 V.

4.3.1 Forward Bias

The ideal diode equation gives the current flowing through the diode as a function of applied bias

$$I = I_0(T) \cdot \left[\exp\left(\frac{qV}{kT}\right) - 1\right] \tag{4.9}$$

$k = 1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$ is Boltzmann's constant, V is the voltage across the junction and T is the temperature. I_0 is the saturation current given by

$$I_0(T) = \frac{qAD(T)n_i^2(T)}{N_B L_d} \tag{4.10}$$

and $D(T)$ and N_B are the minority carrier diffusion coefficient and the doping concentration, respectively. L_d is the diffusion length. This equation is applicable to diodes operating in the regime of low level injection. At room temperature the current increases by an order of magnitude for each 60 mV increase in forward bias. At high current levels, the forward diode current equation is modified to become

$$I_{high} = I_0' \exp\left(\frac{qV}{2kT}\right) \tag{4.11}$$

where,

$$\begin{aligned} I_0' &= \left(\frac{2N_B}{n_i}\right) I_0 \\ &= \left(\frac{2qDn_i}{L_d}\right) I_0 \end{aligned} \tag{4.12}$$

which is independent of the background doping concentration N_B. This is an indication that the region has become conductivity modulated. In a conductivity modulated region, the n and p concentrations are both higher than the background doping concentration. Hence,

$$p = n = n_i \exp\left(\frac{qV}{2kT}\right) \tag{4.13}$$

and the conductivity is a function of the number of carriers crossing the junction rather than the doping level of the well region. The transition between low- and high-level injection occurs at

$$V_0 = \frac{2kT}{q} \ln\left(\frac{2N_B}{n_i}\right) \tag{4.14}$$

For the case of $N_B = 10^{17}/\mathrm{cm}^3$ at room temperature, $V_0 \sim 0.8\,\mathrm{V}$. This high-level injection region is rarely observed because the resistance of the heavily doped regions becomes important at about the same bias.

In lightly doped substrates and diodes built in a well, the current flow is almost entirely lateral between the adjacent highly doped regions as indicated in Figure 4.4(a). The diode area is defined by the junction sidewall area; that is, proportional to the junction depth and the peripheral length. Diodes built in thin epitaxial substrates will have current flow paths that are vertical as shown in Figure 4.4(b). In this case the bottom-wall area of the junction defines the diode area; that is, the diode area is proportional to the area of the diffusion.

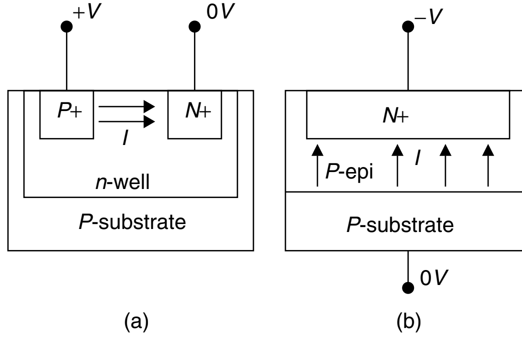


Figure 4.4 (a) Forward-biased current in an *n*-well diode. The current flow is almost entirely lateral. (b) Forward-biased current flow in a substrate diode. The current flow is vertical

Diodes used in ESD circuits will carry very high current densities [Boselli01]. The structures shown in Figure 4.4 have heavily doped anode and cathode regions and lower doped regions that can be *P* or *N* type depending on the diode. The on-resistance in low current injection will be dominated by the doping concentration of this region, however, for high current densities the conductivity modulation condition results in a lowering of the on-resistance to about $50\,\Omega\mu\text{m}^{-1}$. Conductivity modulation under transient conditions occurs after a finite time approximately equivalent to the carrier transit time in the lower doped region. Hence the initial resistance will be large and will reduce with time (τ) within ≈ 1 ns. τ is a function of the width of the low-doped region. This time-dependent clamping voltage can be significant in protection circuits that will be affected by high voltages at the pad even for very short durations, or for very short duration ESD events, such as with the *Charged Device Model* (CDM) discharge.

4.3.2 Reverse Bias

When a diode junction is reversed-biased more than a few tenths of a volt, the reverse current is defined by

$$I_R = \frac{qADN_C N_V}{L_d N_B}\exp\left(\frac{-E_g}{kT}\right) + \frac{qW}{\tau_e}\sqrt{N_C N_V}\exp\left(\frac{-E_g}{2kT}\right) \quad (4.15)$$

where N_C and N_V are the density of states in the conduction band and the valence band, respectively. W is the width of the depletion region and τ_e is the effective carrier lifetime. The assumption in the above equation is that the field across the junction depletion region at the applied voltage is below the critical field necessary for avalanche breakdown ($\sim 10^5\,\text{V cm}^{-1}$). The reverse current is thus solely due to the sum of thermally generated carriers in the depletion region given by the second term on the right-hand side of Equation 4.15 and the diffusion component

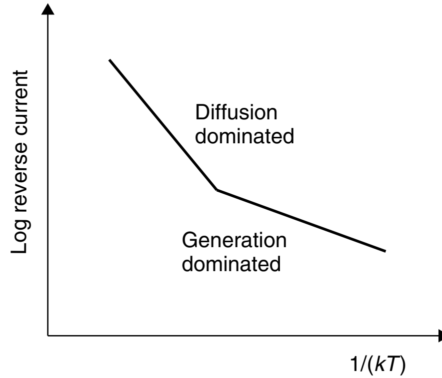


Figure 4.5 Current in a reverse-biased junction, I_R, as a function of temperature. The diffusion dominated region where $\ln(I_R)$ is proportional to $1/kT$ is observed at higher temperatures, and the generation-dominated region where $\ln(I_R)$ is proportional to $1/2kT$ is observed at lower temperatures

of the carriers in the neutral region given by the first term on the right-hand side of Equation 4.15. Figure 4.5 shows a plot of $\ln(I_R)$ as a function of $1/kT$. At low temperatures I_R will be dominated by the thermal generation, and shows a slope of 1/2. At higher temperatures, the diffusion current dominates and I_R has a slope of 1.

As the reverse voltage is increased and the electric field across the junction approaches $10^5\,\mathrm{V\,cm^{-1}}$, the carriers in the depletion region can impart enough energy in a collision with the lattice to generate electron–hole pairs, which become free carriers. These new carriers in turn are accelerated, collide with the lattice and create more carriers. The process is known as *avalanche multiplication*. The hole and electron currents, I_{p0} and I_{n0}, flowing into a high field region are multiplied so that the currents exiting the region are:

$$I_p = I_{p0} + \alpha_p I_{p0} + \alpha_n I_{n0} \tag{4.16}$$

$$I_n = I_{n0} + \alpha_p I_{p0} + \alpha_n I_{n0} \tag{4.17}$$

$\alpha_{n,p}$ are the electron and hole impact ionization coefficients and are temperature dependent. Okuto and Crowell [Okuto75] give an empirically determined form for the impact ionization coefficients for electrons and holes, $\alpha_{n,p}$/cm, as a function of temperature;

$$\alpha_{n,p} = A_{n,p}\cdot\{1+C_{n,p}\cdot 10^{-4}(T-300)\}\cdot E\cdot\exp\left(\frac{-B_{n,p}^2\cdot[1+D_{n,p}\cdot(T-300)]^2}{E^2}\right) \tag{4.18}$$

$A_n = 0.426/\mathrm{V}$, $A_p = 0.243/\mathrm{V}$, $B_n = 4.81\times 10^5\,\mathrm{V\,cm^{-1}}$, $B_p = 6.53\times 10^5\,\mathrm{V\,cm^{-1}}$, $C_n = 3.05\times 10^{-4}$, $C_p = 5.35\times 10^{-4}$, $D_n = 6.86\times 10^{-4}$, and $D_p = 5.87\times 10^{-5}$ are the coefficients for electrons and holes; E is in $\mathrm{V\,cm^{-1}}$. The above equations

can be simplified to obtain $\alpha_{n,p}$ by reducing to

$$\alpha_{n,p} = A_i \cdot \exp\left(\frac{-B_i}{E}\right) \tag{4.19}$$

where $B_i = E_g/q\lambda$ and λ is the mean free path of the carrier [Crowell66]

$$\lambda = \lambda_0 \tanh\left(\frac{E_{r0}}{2kT}\right) \tag{4.20}$$

$\lambda_0 \approx 50$ Å and $E_{r0} = 50\,\text{meV}$ are λ and the optical phonon energy, E_r, at 0 K. A reasonable fit to experimental results has been obtained by using [Grant73]

$$\alpha_{n,p} = A_i \cdot \exp\left(\frac{-B_i(T)}{E}\right) \tag{4.21}$$

where the coefficient A_i remains constant as a function of temperature and the major variation with temperature is assumed to occur in the exponent $B_i(T)$.

The multiplication factor for holes and electrons, $M_{n,p}$, is an important parameter in applying impact ionization models to device behavior.

$$M_{n,p} = \frac{I_{n,p}(\text{out})}{I_{n,p}(\text{in})} \tag{4.22}$$

$I_{n,p}(\text{out})$ and $I_{n,p}(\text{in})$ define the currents at the edges of the depletion region.

For $\alpha_n \approx \alpha_p \approx \alpha$, $M_{n,p}$ can be written as

$$M_{n,p} = \frac{1}{1 - \int_0^w \alpha dx} \tag{4.23}$$

Avalanche breakdown occurs at the voltage when $M_{n,p}$ approaches infinity, that is,

$$\int_0^w \alpha = 1 \tag{4.24}$$

Empirically, the relationship between M and the voltage across the junction V_j has been described in the form [Miller57]

$$M = \frac{1}{1 - (V_j/V_{av})^n} \tag{4.25}$$

where V_{av} is the avalanche breakdown voltage and n is a fitting parameter ranging from 2 to 6 depending on the type of junction being considered. The increase in M with applied voltage is very sharp as V_j approaches V_{av}.

Another approach is to substitute for α from Equation 4.21 in Equation 4.23, to give

$$M = \frac{1}{1 - A_i \exp(-B_i/V_j)}. \tag{4.26}$$

The fitting parameters $A_i \approx A \cdot x_\mathrm{d}$ and $B_i \approx B \cdot x_\mathrm{d}$ provide better empirical matching over a range of V_j compared to Equation 4.25, especially when comparing graded and abrupt junctions. As V_j approaches V_av, M approaches ∞, and Equation 4.26 can be rewritten as

$$V_\mathrm{av} = \frac{B_i}{\ln(A_i)} \tag{4.27}$$

This equation can be used to relate A_i and B_i if V_av is already known.

The only limiting factor is the resistance of the neutral regions outside the depletion region. The temperature dependence of α is such that as T increases, the impact ionization decreases and M goes down, which means that the avalanche breakdown voltage increases with temperature.

4.3.3 *p-i-n* Diode

In state-of-the-art silicon IC technologies, diodes have highly doped p and n diffusions separated by a low-doped region of either p type in n/p-substrate diodes or n type in the case of n-well diodes as shown in Figure 4.6. At low current levels the low-doped or intrinsic (i) region acts as a resistor in series with either the anode or cathode of the diode. However, at high current levels, the i-region becomes strongly conductivity modulated, that is, the n and p concentrations are significantly greater than their equilibrium values.

At low injection levels, the current density through the diode is dominated by the drift current in the lightly doped region. At moderate and high injection levels, J is dominated by diffusion currents in the vicinity of the junctions. It can be shown [Boselli01] that under these conditions the resistance changes with current and varies as $R \propto 1/\sqrt{I}$. Therefore, the modulation phenomenon results in a current-dependent resistor in this region, with R decreasing as I increases. Under high injection conditions the device behaves as a series combination of an ideal diode and a variable resistor whose resistance is dependent on the direction of current flow. When the diode is forward-biased, the resistor is modeled as $R = K/\sqrt{I}$, where K is proportional to the separation of the contact from the junction. The on-resistance of the *p-i-n* diode at high current levels is much lower than at low current levels, which is beneficial for ESD protection circuits.

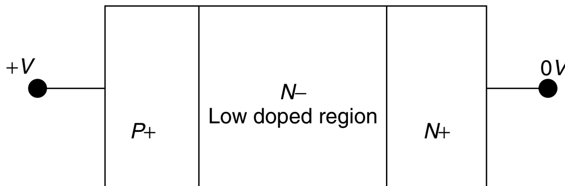


Figure 4.6 A simple P+/N−/N+ diode typical of that in an n-well, in forward-bias condition

As the injection levels are increased further, the E-field in the n region builds up until the carrier velocities reach v_S and the current saturates. At the same time the field now allows holes to cross the n region and reach the n/n^+ junction reducing the space charge at this junction and allowing more electrons to be injected into the n region. The voltage required to maintain the same I can begin to decrease and a negative resistance region is observed as the voltage reduces to a lower sustaining level [Ghandhi77][Chatterjee88].

4.4 TRANSISTOR OPERATION

4.4.1 Bipolar Transistors in Typical Operating Conditions

A bipolar transistor in normal operation has a forward-biased junction that enables minority carriers to be injected into the vicinity of a reverse-biased junction. As currents in reverse-biased junctions are carried by minority carriers, the increase in minority carrier concentration at the junction edge will result in an increase in the current across the junction. The current flow across the reverse-biased junction can be modulated by controlling the injection of minority carriers from the forward-biased junction, and this forms the basis of the bipolar transistor.

A simple *npn* bipolar structure is shown in Figure 4.7. This is a three terminal device consisting of two *pn* junctions. Electron injection takes place from the *emitter n* region when a positive voltage V_{be} is applied across the base-emitter junction. The carriers are injected into the *base p* region. The electron current reaching the reverse-biased junction will be enhanced if the width of the base is small enough that there is no significant loss of electrons in the base due to recombination with holes. Electrons cross the reverse-biased junction into the *collector n* region. The number of carriers reaching the reverse-biased collector junction are determined

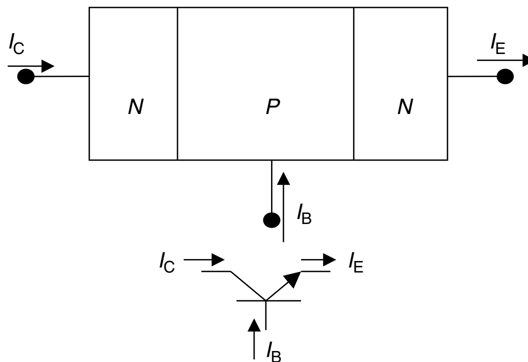


Figure 4.7 A simple *npn* transistor showing the collector current I_C, emitter current I_E, and base current I_B

by the width of the base region, W, and the recombination rate in the base, which is dependent on the hole concentration in the base. The hole current density in the *npn* transistor, J_p, is negligible and the electron current density, J_n, is given by the equation,

$$J_n = J_s \left[\exp\left(\frac{qV_{bc}}{kT}\right) - \exp\left(\frac{qV_{be}}{kT}\right)\right] \tag{4.28}$$

V_{bc} is the base-collector voltage and J_s is the saturation current density given by

$$J_s = \frac{qn_i^2 \tilde{D}}{N_B} \tag{4.29}$$

n_i is the intrinsic carrier concentration and N_B is the total number of impurities per unit area of the base. $\tilde{D}$ is the effective minority carrier diffusion coefficient, which assumes that the diffusion coefficient D is not strongly dependent on position.

To turn on an *npn* bipolar transistor, V_{be} needs to be positive and greater than kT/q. J_n then becomes a function of the most positive voltage. Under active bias the term $\exp(qV_{bc}/kT)$ is negligible and the collector current in an *npn* bipolar transistor is given by

$$I_c = I_s \exp\left(\frac{qV_{be}}{kT}\right) \tag{4.30}$$

where $I_s = J_s \times A$ and A is the base-emitter junction area.

Figure 4.8 shows how I_c varies with V_{ce} for an *npn* transistor as the base current I_b is increased. Initially, with $I_b = 0$, the only current flow is that across the reverse-biased collector-base junction until breakdown of the junction takes place at BV_{cbo}. As I_b is increased, that is, V_{be} is made more positive, the transistor turns on. Higher I_b results in higher I_c, both in the saturation and linear regions.

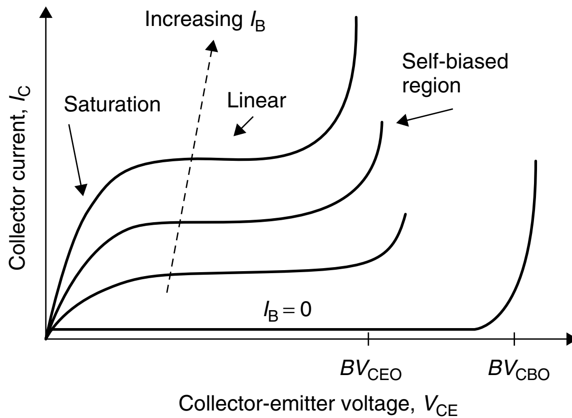


Figure 4.8 I_c as a function of the collector-emitter voltage V_c for an *npn* transistor with increasing I_b. BV_{cbo} indicates the collector-base breakdown voltage with the emitter open circuit and BV_{ceo} is the collector-emitter voltage when the base is open circuit

The *current gain* of the transistor is a measure of its effectiveness and is determined by the ratio of the output current to the input current for a specific bias condition. As I_c is an exponential function of V_{be}, the smaller the current between the base and emitter for a given V_{be} the more effective the transistor. The base-emitter current, I_b, consists of: (1) the current due to recombination of injected electrons with holes in the base; (2) the current due to recombination in the space charge region; and (3) the current due to hole injection from the base into the emitter. The sum of these currents defines I_b. The gain is improved if recombination is reduced (i.e., lifetime increased) in the base for electrons and in the emitter for holes.

The two important current gain definitions for circuits used under ESD type conditions are the common-base current gain, α, and the common-emitter current gain, β. α and β are given by

$$\alpha = \frac{\partial I_c}{\partial I_e} \sim \frac{I_c}{I_e} \tag{4.31}$$

$$\beta = \frac{\partial I_c}{\partial I_b} \sim \frac{I_c}{I_b} \tag{4.32}$$

α and β are related by

$$\beta = \frac{\alpha}{1-\alpha} \tag{4.33}$$

Typically, $\alpha \approx 1$ and $\beta \gg 1$.

Two other parameters of importance are the base transport factor α_T and the emitter efficiency γ, which is the effectiveness of the emitter junction in injecting electrons into the base. α_T is a measure of the loss of carriers due to recombination in the base region

$$\alpha_T = \frac{I_n(W)}{I_n(0)} \tag{4.34}$$

$$\approx 1 - \frac{W^2}{2L_B^2} \tag{4.35}$$

L_B is the minority carrier diffusion length in the base.

γ is given by the ratio of the incremental electron current from the emitter to the incremental total electron current,

$$\gamma = \frac{I_{ne}}{I_{ne} + I_{pe}} \tag{4.36}$$

I_{ne} is the electron current across the emitter-base junction and I_{pe} is the hole current across the emitter-base junction.

The recombination current only flows between the emitter and the base. I_c consists almost entirely of the collected electrons, which were injected at the base-emitter junction. As V_{be} is decreased, I_c follows Equation 4.30 until injection is so low that generation in the space charge region begins to dominate. At low currents,

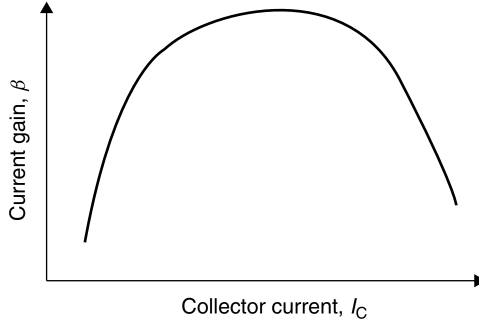


Figure 4.9 The current gain β as a function of I_c for a bipolar transistor showing the low β regions at low I_c and high I_c

therefore, I_c is a smaller fraction of I_e and β is low. The variation of β with collector current is shown in Figure 4.9. In the ideal region, β is more or less constant, and then as I_c increases further, β begins to decrease rapidly. β degradation at high I_c occurs when the injected minority carrier density into the base approaches the majority carrier density thereby increasing the majority carrier charge in the base. Hence, the injected carriers effectively increase the base doping, which reduces the emitter efficiency and β. A second effect at high injection levels is current crowding at the emitter, which is essentially because of two-dimensional effects. The flow of majority carriers in the base region leads to a potential drop across the width of the emitter and causes variation in the forward-biased voltage across different regions of the emitter. This results in a variation in the injected current density across the emitter junction. For example, a lateral ohmic drop of $\approx 26\,\text{mV}$ will result in a reduction in the emitter current density of $1/e$. At high currents, therefore, the effective emitter area is reduced and contribute to the reduction in β.

An important parameter for time-dependent operation of bipolar transistors is the *base transit time*, τ_B [Muller86][Krieger89]. τ_B determines how long it takes for electrons injected at the emitter to reach the collector junction and turn on the transistor. If the injected minority carriers into the base have a charge Q_{nB}, then

$$Q_{nB} = \int_0^W qAn'(x)dx \tag{4.37}$$

where A is the area of the emitter and $n'(x)$ is the concentration of the excess electrons in the base at a point x. For a collector current, I_c, τ_B is given by

$$\tau_B = \frac{Q_{nB}}{I_c} \tag{4.38}$$

If $n'(x)$ is assumed to be linear across W and substituting for I_c then

$$\tau_B = \frac{W^2}{2\tilde{D}_n} \tag{4.39}$$

At high current levels, a field is set up between the emitter and collector, which aids the flow of minority carriers towards the collector and [Muller86]

$$\tau_B = \frac{W^2}{4\tilde{D}_n} \tag{4.40}$$

Typical values of τ_B for a 1-μm wide base are about 250 ps, which are well within the rise-times associated with ESD stress currents.

4.4.2 MOS Transistors in Typical Operating Conditions

A schematic cross section of an nMOS transistor is shown in Figure 4.10 and consists of two *n* diffusion regions in a *p* substrate. A pMOS transistor is the complement of the nMOS and the description presented here is applicable to the pMOS with changes in the polarity of the majority and minority carriers. The relationship between the drain voltage V_D and the drain current I_D for different gate voltages V_G is shown in Figure 4.11. During normal operation, the transistor operates in either the linear region *Region 1* or the saturation region *Region 2*. Upon application of a positive voltage, V_D, at the drain (D), with the gate (G), source (S) and substrate (B), connected to zero volts, no current will flow until the reverse-biased drain-substrate junction goes into avalanche breakdown at V_{av} as depicted by the curve $V_G = 0$ V in Figure 4.11. When a positive voltage, V_G, is applied at the gate,

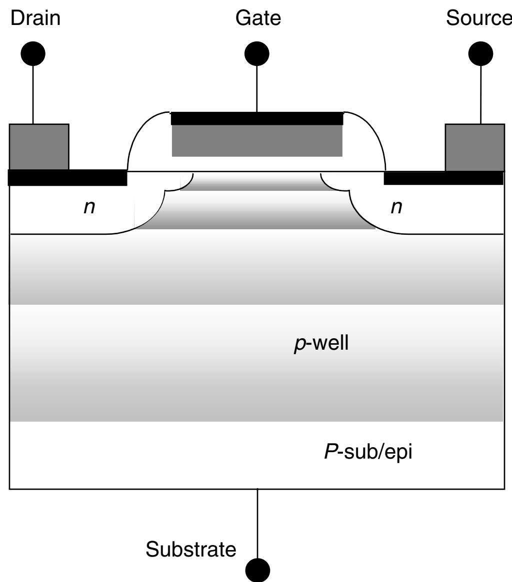


Figure 4.10 Schematic cross section of an nMOS transistor

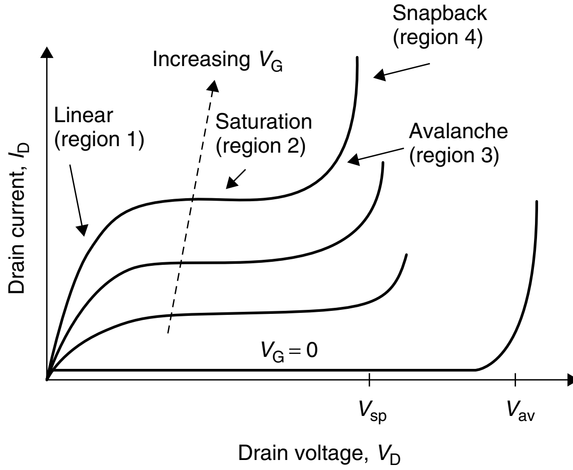


Figure 4.11 The drain current I_D of an nMOS transistor as a function of the drain-source voltage V_{DS} for varying V_G. V_{av} is the drain-substrate breakdown voltage with $V_G = 0\,V$

the p region between the drain and source first becomes depleted and then, as V_G is increased further, the p-region becomes inverted and an n channel is formed. The transistor is now in the on-state. The gate voltage at which the transistor turns on is called the threshold voltage V_T. In the subthreshold (linear) region, the drain current is given by

$$I_D = \frac{aW}{L} \cdot (V_G - V_T) \cdot V_D - b \cdot V_D^2 \tag{4.41}$$

a is a constant dependent on the electron mobility in the channel and the capacitance between the gate and the semiconductor. W is the width of the transistor, L is the channel length and b is a constant dependent on the gate oxide capacitance and channel doping concentration.

In the saturation region, the current is given by

$$I_{Dsat} \approx \frac{b'W}{L}(V_G - V_T)^2 \tag{4.42}$$

b' is similar to b but includes a doping dependent parameter [Sze81]. From these two equations we see that the subthreshold current is a linear function of the gate voltage, while the saturation current is a quadratic function of V_G.

For a constant V_D, I_D varies as a function of V_G as shown in Figure 4.12. It is also interesting to observe the current at the fourth (substrate) terminal as V_G is varied, also shown in Figure 4.12. The substrate current I_{sub} is generated by impact ionization of the channel carriers at the drain-substrate junction and is a function of the magnitude of the channel current I_D as well as the drain-substrate junction voltage V_j. I_{sub} is observed to increase to a maximum and then decrease

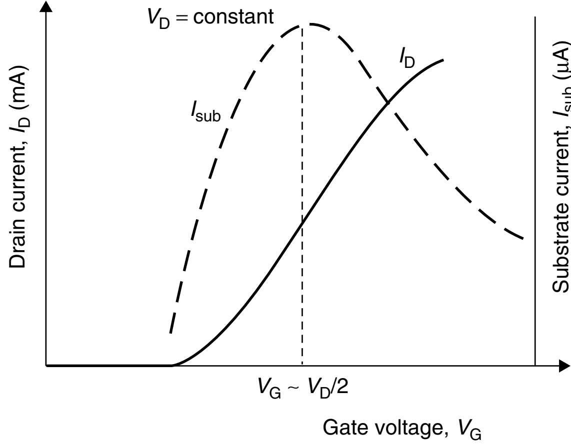


Figure 4.12 Drain current I_D vs. gate voltage V_G for a constant drain-source voltage V_{DS}. Also shown is the substrate current I_{sub} as a function of V_G

as V_G is increased. The increase in I_{sub} is attributed to the increase in I_D with V_G. However, as I_D is increased further, the voltage in the channel V_{DCH} gets large and causes the junction voltage V_j to decrease for the same V_D, because $V_j = (V_D - V_{DCH})$. The behavior of I_{sub} under different gate, drain, and substrate bias conditions is very important to the high current performance of the MOS transistors and we will return to the topic in more detail later (Section 4.5.2) [Ramaswamy97].

The linear current and saturation current regions are the normal operating regions for the MOS transistor, where we have control over the drain current through V_G. At ESD current levels, the MOS transistor is forced to carry current that could be much higher than can be supported in these regions. In such cases, we end up in the avalanche region (Region 3) and into the *snapback* region (Region 4) as shown in Figure 4.11.

4.4.3 Avalanche Conditions

Both bipolar and MOS transistors can be biased into avalanche during very high current injection. In this section we outline the main features in a transistor operating with one avalanching junction [Dutton75][Reisch92]. For simplicity consider the one-dimensional bipolar transistor shown in Figure 4.13. α is the common-base gain and M is the avalanche multiplication factor in the collector-base depletion region defined by ratio of the output current I_{out} to the input current I_{in} at the avalanching junction, such that

$$M = \frac{I_{out}}{I_{in}} \tag{4.43}$$

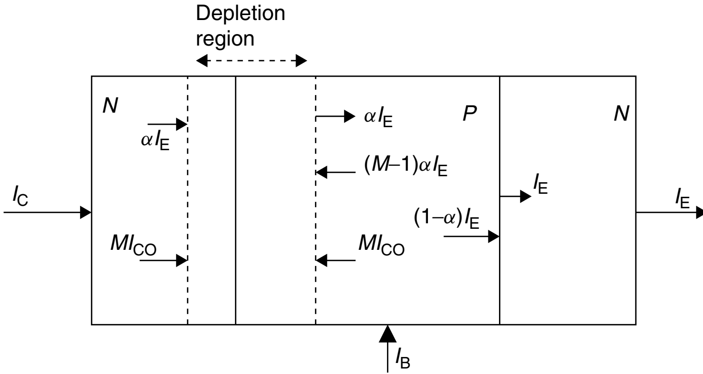


Figure 4.13 Simple one-dimensional model of a bipolar transistor with avalanching in the reverse-biased collector-base depletion region. M is the avalanche multiplication factor in the depletion region and α is the common-base current gain

The collector current, I_c, is given by

$$\begin{aligned} I_c &= \alpha I_e + (M-1)\alpha I_e + M I_{CO} \\ &= \alpha M I_e + M I_{CO} \end{aligned} \tag{4.44}$$

where I_e is the emitter current and I_{CO} is the thermal generation current across the depletion region as described for reverse-biased junctions in Equation 4.15. As the electrons recombining in the base must equal the holes recombining in the base

$$(1-\alpha) = M I_{CO} + (M-1)\alpha I_e + I_b \tag{4.45}$$

and I_b is the base current being injected into the transistor. Now from Equation 4.33

$$(1-\alpha) = \frac{\alpha}{\beta} \tag{4.46}$$

and

$$\alpha I_e = \frac{I_c}{M} - I_{CO} \tag{4.47}$$

we get

$$I_c = \frac{M\beta I_b + M I_{CO}(1+\beta)}{1-\beta(M-1)} \tag{4.48}$$

As the avalanche generation at the collector-base junction increases the hole current entering the base terminal I_b goes to zero and eventually becomes negative. The condition for $I_b \leq 0$ and $I_c > 0$ is given by

$$\beta(M-1) \geq 1 \tag{4.49}$$

$$\beta M \geq (\beta + 1) \tag{4.50}$$

In the open-base condition $I_b = 0$ and Equation 4.45 reduces to [Sze81]

$$I_c = I_e = I = \frac{MI_{CO}}{1 - \alpha M} \tag{4.51}$$

Equation 4.48 describes the current in a floating base transistor and combining with Equation 4.25 can be used to determine the collector-emitter breakdown voltage, BV_{ceo}, in terms of the collector-base breakdown voltage, BV_{cbo},

$$BV_{ceo} = BV_{cbo}(1 - \alpha)^{1/n} \tag{4.52}$$

4.5 TRANSISTOR OPERATION UNDER ESD CONDITIONS

4.5.1 Bipolar Transistor under ESD Conditions

The bipolar transistor can be designed to carry high currents under normal conditions, but such devices require a high collector-emitter voltage V_{ce} and dissipate a lot of power. ESD protection circuits with such devices are used in high-voltage and high-power applications [Corsi93]. However, transistors designed for low voltage (≤ 5 V) applications cannot carry large currents and the behavior deviates from the standard operation.

Under ESD conditions, the base terminal is connected to the emitter terminal either directly or through an external resistor R_{ext} as shown in Figure 4.14. Two turn-on modes are possible; the first trigger condition is shown in Figure 4.14 and the second trigger condition is shown in Figure 4.15. The first type operates

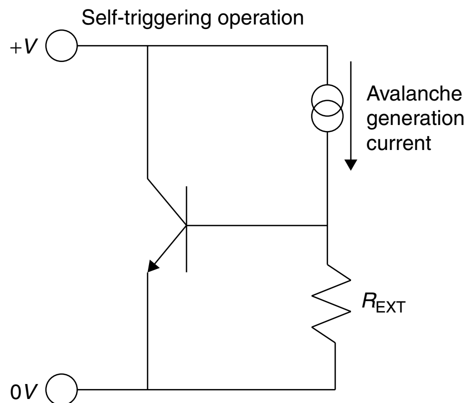


Figure 4.14 Schematic of a bipolar *npn* transistor operating in the self-triggering mode. R_{ext} is an external resistor connected between the base and the emitter to ensure *npn* turn-on when avalanche generation occurs at the collector-base junction

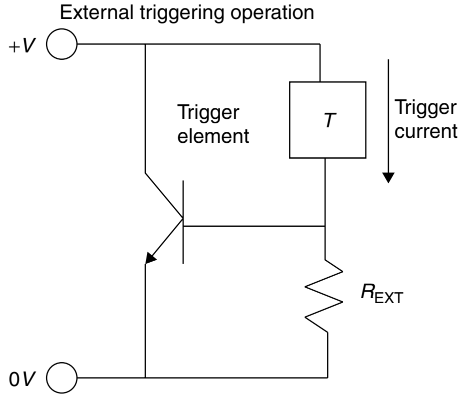


Figure 4.15 Schematic of a bipolar *npn* connected in the external triggering mode. R_{ext} is an external resistor and the bipolar is triggered when the current through the trigger element T is sufficient to forward-bias the emitter-base junction of the *npn*

through self-biasing using the internal avalanche generation of carriers to turn on the bipolar transistor. As will be discussed below, the bipolar is triggered at the collector-base breakdown voltage, which is $\approx BV_{cbo}$. In the second type the transistor is provided some forward biasing by means of an external current source, T, in Figure 4.15, and the voltage drop across the resistor R_{ext}. The advantage of this approach over the first is that it reduces the voltage at which the bipolar turns on in the self-triggering mode [Chatterjee91A][Amerasekera92]. We will discuss the mechanisms of the bipolar operation in more detail in this section.

In the first type, a positive current is forced into the collector and Figure 4.16 shows the variation of the forced I_c with the collector voltage, V_c. Initially the reverse-biased collector-base junction is in high impedance and the reverse current is given by Equation 4.15. The collector voltage rises until it reaches the collector-base junction breakdown voltage, BV_{cbo}. Impact ionization then takes place in the junction and electron-hole pairs are generated. The electric field is such that the electrons enter the collector and increase I_c, and the holes drift to the base terminal generating a negative base current. The hole current into the base contact will result in a potential drop in the base, and when the emitter-base voltage V_{be} reaches $\sim 0.5\,\text{V}$ the *npn* begins to turn on. *Note*: The value of V_{be} at which turn on takes place is not a precise value. Under ESD conditions, turn-on will occur when V_{be} is sufficiently forward-biased to support the ESD stress current injected at the collector. We will use $V_{be} \approx 0.5\,\text{V}$ in order to simplify the analysis.

The generated hole current, I_B, required to forward-bias the emitter-base junction will depend on the internal base resistance, R_B, as well as R_{ext}. I_B required to turn on the *npn* can be calculated from,

$$V_{be} = I_B \cdot (R_B + R_{ext}) \tag{4.53}$$

$$V_{be} > 0.5\,\text{V} \tag{4.54}$$

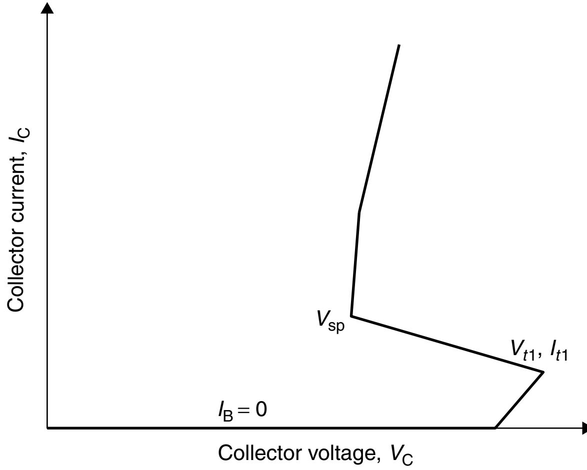


Figure 4.16 High current $I-V$ curve for an *npn* transistor showing the trigger voltage and holding condition for self-biased operation under high current conditions

The function of the generated I_B under ESD conditions is to increase V_{be} and forward bias the emitter-base junction thus providing the electrons necessary for the *npn* action to take place. Once the *npn* turns on, the electron current reaching the reverse-biased collector-base junction increases the number of generated electron-hole pairs since [Dutton75][Reisch92],

$$I_B \propto (M - 1) \cdot I_e \tag{4.55}$$

M is a function of the voltage across the junction, and as I_e increases M can decrease and V_{cb} can also decrease. The voltage at which V_c begins to decrease is the snapback trigger voltage denoted by V_{t1} in Figure 4.16. The snapback trigger current is given by $I_c = I_{t1}$. The multiplication factor, M, can be estimated from

$$\frac{I_B}{I_{t1}} = 1 - \frac{1}{M} \tag{4.56}$$

V_c now decreases until eventually a stable condition is reached whereby the generated I_B satisfies Equation 4.50. The condition for snapback, which requires that I_B flows out of the base terminal and I_C is positive has been previously defined in Equation 4.49 as

$$\beta(M - 1) \geq 1 \tag{4.57}$$

The V_{be} necessary to support the injected, I_c, is obtained from Equation 4.30. Further increase in I_c results in an increase in V_c as conductivity modulation of the internal base resistance necessitates an increase in I_B to sustain the transistor in the on-condition. The snapback holding voltage is given by V_{sp}, and is a function

of the base-width, W, the multiplication factor, M, for the collector-base junction, R_B and R_{ext}.

In the second operating mode (Figure 4.15), T is a current source formed either by a transistor or a diode operating in reverse-bias with a breakdown voltage much less than BV_{cbo}. The current from T goes through the resistor R_{ext} and increases V_{be}. As V_{be} becomes more positive, the electrons entering the base from the emitter can contribute to the avalanche generation at the collector junction. Thus the voltage (and M) required for a given I_B is lower and, since V_{be} is already positive, the I_B required to fully turn on the bipolar is reduced. V_{t1} is, therefore, lower than for the first bias condition without external triggering. The lower V_{t1} is extremely desirable in ESD protection circuits because it ensures that the protection device will trigger before the device being protected.

4.5.2 MOS Transistors under ESD Conditions

Under ESD stress conditions the MOS transistor in the ESD path is required to carry amperes of current. It should also have a low clamping voltage because of the very thin gate oxides (a few nanometers thick) in advanced CMOS technologies. This low impedance clamp can be obtained by triggering the inherent lateral bipolar transistor present in both nMOS and pMOS transistors. Figure 4.17 shows a cross section of an nMOS transistor and the associated lateral *npn* transistor. The collector is formed by the drain of the nMOS, the emitter is formed by the source, and the base is the substrate.

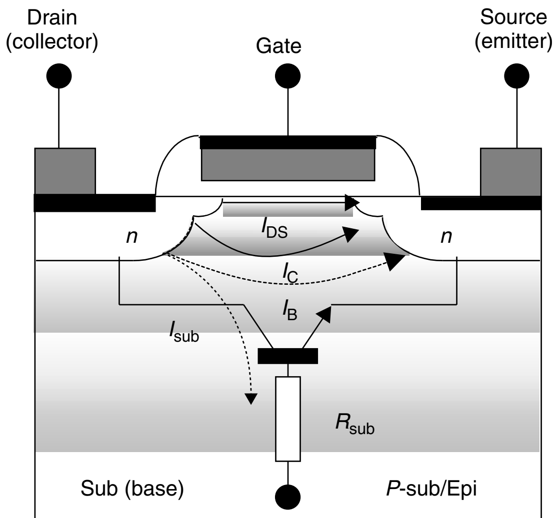


Figure 4.17 Cross section of an nMOS transistor showing the parasitic lateral *npn* bipolar transistor and associated currents

Under high current conditions, the nMOS transistor operates in Region 3 and Region 4 of the $I-V$ curve in Figure 4.11. The mechanisms of operation in these regions involve both avalanche breakdown and turn-on of the parasitic lateral bipolar transistor (LNPN). For better comprehension of the high current operation, we consider the transistor with gate, source, and substrate at 0 V, which results in the high current $I-V$ curve shown in Figure 4.18. As the drain current is increased, the reverse-biased drain-substrate junction is initially in high impedance. The only current is the reverse current given by Equation 4.15. Eventually the drain-substrate junction begins to avalanche due to the high voltage across it, and electron-hole pairs are generated. The electrons are swept across the drain junction towards the drain contact, adding to the drain current, while the holes drift towards the substrate contact giving rise to a substrate current, I_{sub}, similar to the case of the base current for the bipolar transistor described in the previous section. The effective substrate resistance is denoted by R_{sub} in Figure 4.17, which also shows schematically the currents in the nMOS. As I_{sub} increases, the potential at the source-substrate junction increases and forward biases this junction causing electrons to be emitted into the substrate. The electron current density from the source begins to contribute to the drain current and the parasitic bipolar transistor can be considered to be turned on. This is effectively *self-biased* bipolar operation, since the bias current is generated by the intrinsic avalanching at the drain-substrate junction. In Figure 4.18, V_{t1} is the turn-on voltage of the LNPN transistor and the trigger current is I_{t1}. The drain of the nMOS becomes the collector of the LNPN, the source of the nMOS becomes the emitter of the LNPN and the substrate is the base. The turn-on time of the LNPN is defined by the base transit time τ_B, which depends on the gate length L. For a 1 μm channel length, $\tau_B \approx 250\,\text{ps}$ [Krieger89]. Once the LNPN turns on, the drain voltage decreases and a negative resistance region is observed due to

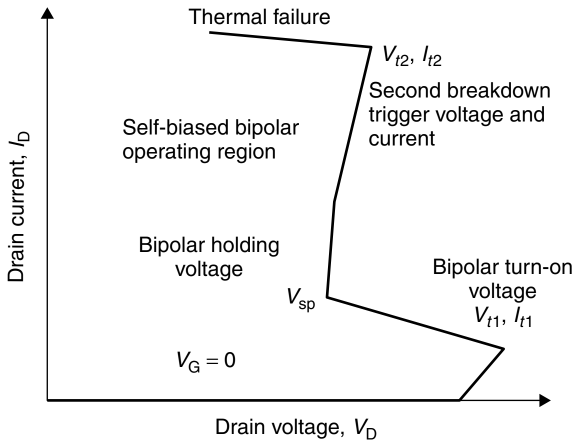


Figure 4.18 High current $I-V$ curve of an nMOS transistor with gate, source, and substrate at zero volts

the availability of more carriers for multiplication until a minimum voltage, V_{sp}, is reached. The $I-V$ curve now shows a positive resistance as further increase in the injected current results in conductivity modulation of the substrate (base) region that reduces the intrinsic substrate resistance. A higher I_{sub} is required to maintain the transistor in the on-condition.

The phenomena associated with the high current effects have been studied analytically [Hsu82][Eitan82][Jankovic91][Pinto-Guedes88] and numerically [Schutz82][Laux87]. In a self-biased LNPN operation, it is important to note that the emitter area of the LNPN is not the same as the diffusion area of the source because only a small portion of the source needs to be forward-biased to operate the transistor. This operation differs from that of an externally biased LNPN where the base voltage is provided at the base contact [Lindmayer67][Verdonckt91]. An understanding of this behavior is important in defining circuit modeling techniques [Amerasekera96] and transistor design and process optimization for ESD performance [Amerasekera96][Gupta98][Bock99].

The substrate current I_{sub} is a function of the avalanche multiplication factor M in the high-field region of the drain. The avalanche generation current at the high-field region due to an incident current I_p is given by

$$I_{gen} = (M - 1) \cdot I_p \tag{4.58}$$

With $V_G = 0$V, the incident current at the drain junction is solely due to thermal generation and minority carrier diffusion. Before the bipolar turns on, $I_{gen} = I_{sub}$. A typical value for I_{sub} for bipolar turn-on in a thin-epi sub-0.5 µm process is $\approx 100\ \mu A\,\mu m^{-1}$, while I_p at room temperature can be as low as 10^{-19}A/ µm. Thus M needs to be close to 10^{15} to provide the required I_{sub} for bipolar action to begin; that is, M must tend to infinity as the drain voltage approaches the avalanche breakdown voltage V_{av}.

A gate voltage of V_G greater than the MOS threshold voltage V_T will result in MOS current I_{DS} between the drain and source of the transistor. I_P will now be much larger and a lower M can sustain the same I_{sub}. Hence the drain voltage at which the LNPN turn-on is initiated reduces as a function of V_G as shown in Figure 4.11. Once the bipolar is triggered the current I_C provides additional current for multiplication, and a lower M can sustain the bipolar on-state. V_D reduces even further until it reaches the sustain voltage V_{sp}. The MOS transistor in this condition is said to be in *snapback*. The value of V_{sp} is dependent on the V_D required to generate enough I_{sub} to maintain the substrate potential at the level required to sustain bipolar action. The condition for snapback is given by Equation 4.49

$$\beta \cdot (M - 1) \geq 1 \tag{4.59}$$

When the LNPN is on, the generation current is

$$I_{gen} = (M - 1) \cdot (I_C + I_{DS}) \tag{4.60}$$

where $(I_C + I_{DS})$ are the sum of the bipolar and MOS currents, and for $V_G = 0$, $I_{DS} = 0$,

$$M = \frac{I_D}{(I_D - I_{gen})} \tag{4.61}$$

Before snapback $I_{gen} = I_{sub}$ and we get

$$M = \frac{I_D}{I_D - I_{sub}} \tag{4.62}$$

and

$$I_{sub} = \left(1 - \frac{1}{M}\right) \cdot I_D \tag{4.63}$$

In snapback,

$$I_{gen} = (I_{sub} + I_b). \tag{4.64}$$

I_b is the base-emitter current of the LNPN. Therefore,

$$I_{sub} = (M - 1) \cdot (I_{DS} + I_C) - I_b \tag{4.65}$$

As I_b required to sustain a given injection current I_D is a function of the current gain β of the LNPN,

$$I_b = I_C/\beta \tag{4.66}$$

and assuming I_{DS} is very much smaller than I_C, we obtain M and β in snapback as;

$$M = \frac{(\beta + 1) \cdot I_D}{\beta \cdot (I_D - I_{sub})} \tag{4.67}$$

and

$$\beta = \frac{1}{(M - 1) - M \cdot \frac{I_{sub}}{I_D}} \tag{4.68}$$

From Section 4.3.2, M is given by

$$M = \frac{1}{1 - A_i \exp(-B_i/V_j)}. \tag{4.69}$$

where A_i and B_i are voltage- and current-independent constants, and V_j is the junction voltage. The dependence of M on V_j is a function of the junction doping concentrations, which can change A_i and B_i. For lower A_i and B_i, M will be lower for the same V_j as shown in the plot of M versus V_j in Figure 4.19. Typical values in a deep submicron technology can range from $5 < A_i < 20$ and $12\,\text{V} < B_i < 20\,\text{V}$ [Ramaswamy97] [Amerasekera99].

Prior to bipolar turn-on, I_{sub} can be written as

$$I_{sub} = I_D \cdot A_i(V_D - V_{DCH})^m \cdot \exp(-B_i/(V_D - V_{DCH})^n) \tag{4.70}$$

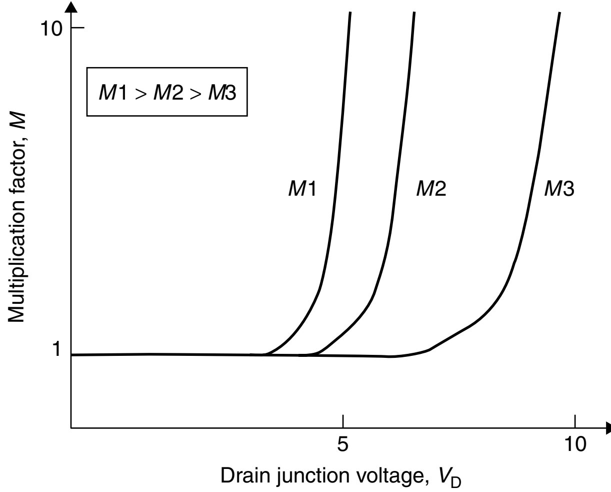


Figure 4.19 Multiplication factor M as a function of the junction voltage V_j with different multiplication constants A_i and B_i. A lower A_i and B_l will result in a lower M at the same V_j

m and n are constants dependent on the drain junction profile. For a device with channel length L, V_{DCH} is modeled as a function of V_G as

$$V_{DCH} = \frac{(V_G - V_T)}{A_{bulk} + ((V_G - V_T)/E_{sat} \cdot L)} \tag{4.71}$$

where E_{sat} is the field at which carriers reach velocity saturation, V_T is the MOS threshold voltage and A_{bulk} is an area dependent constant.

The experimental V_G vs. I_{sub} characteristic before snapback was shown in Figure 4.12. The bell-shaped curve indicates that there is a peak I_{sub} obtained and the reduction in I_{sub} beyond the maximum is attributed to the onset of velocity saturation in the channel.

Before snapback, I_{sub} will have a linear dependence on I_D for a given V_G as shown in Figure 4.20.

The substrate voltage V_{sub} is given by

$$V_{sub} = I_{sub} \cdot R_{sub} \tag{4.72}$$

For $V_{sub} \geq 0.5$V, required to initiate and sustain bipolar action, we get

$$I_{sub} \geq \frac{0.5}{R_{sub}} \tag{4.73}$$

as a condition for bipolar turn-on. If R_{sub} was a constant, then after snapback it would be expected that I_{sub} was a constant too. However, Figure 4.20 shows I_{sub} to continue increasing after snapback, indicating that R_{sub} is not a constant because

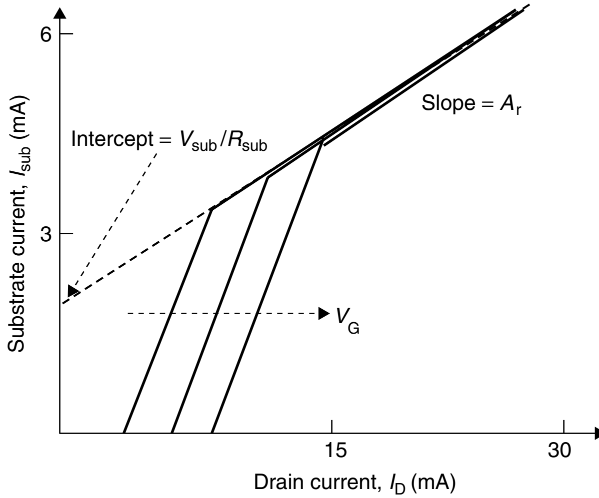


Figure 4.20 Variation of I_{sub} with I_D for a given V_G. The breakpoint occurs when the LNPN is triggered

of conductivity modulation in the substrate at high current levels. Hence in the region after snapback, I_{sub} can be written as

$$I_{sub} = A_r \cdot I_D + V_{sub}/R_{sub} \tag{4.74}$$

A_r is a process-dependent constant.

At high current levels, β decreases as discussed in Section 4.4.1. In this region β dependence on I_D becomes [Amerasekera99]

$$\beta = \frac{A1}{I_D} + A2 \tag{4.75}$$

where $A1$ and $A2$ are process and device-dependent constants. Figure 4.21 shows the dependence of β on I_D [Amerasekera99]. In this case, $\mathrm{A1} = 0.01\,\mathrm{mA}\ \mu\mathrm{m}^{-1}$, and $\mathrm{A2} = 2.6$. The rate of change of β with I_D is determined by A1 at low currents with higher values of $A1$ giving faster change. The limit of the high current β is determined by both $A1$ and $A2$.

The sustained condition for self-biased LNPN operation is a function of the three parameters β, M, and R_{sub}. Since the condition for snapback is

$$\beta \cdot (M - 1) > 1 \tag{4.76}$$

a lower β requires a higher M to sustain snapback. We can plot the dependence of β on M from Equation 4.68 as shown in Figure 4.22. It is seen that for a higher I_D a lower M is required for the same β, but this could be compensated by the β degradation at higher I_D. Furthermore, for low β snapback may not be observed,

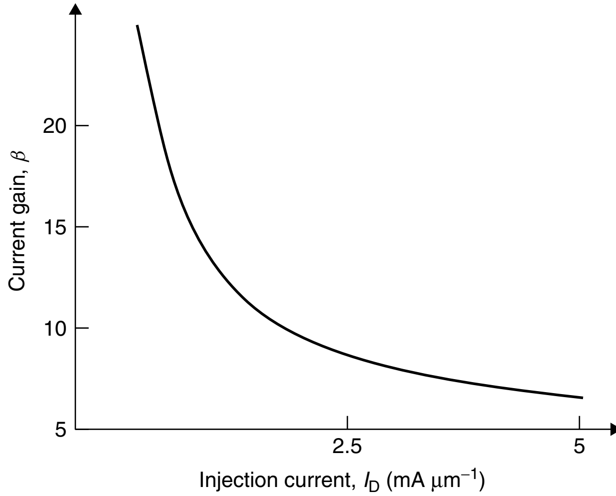


Figure 4.21 Dependence of β on I_D. As I_D increases, β degrades

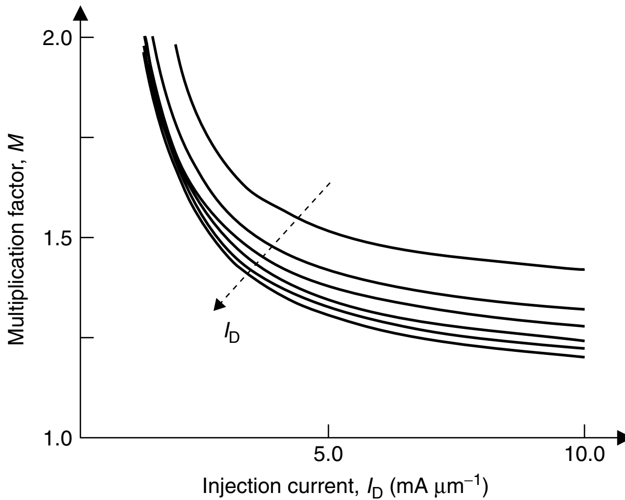


Figure 4.22 Dependence of β on M to sustain snapback for varying I_D. Higher values of I_D require less M to sustain snapback for the same β

even when the base-emitter junction of the LNPN is forward-biased, if the M cannot reduce because of the higher I_{gen} required for the increased I_b. This is similar to the behavior observed in pMOS transistors. In such cases, I_D has to increase to a sufficiently high value to enable a lower M to generate the higher I_b, and only then will V_j drop. In the event that the equilibrium point for sustaining LNPN action

on the $M-\beta$ curve requires a very high M, the LNPN will not turn on before V_j exceeds the collector-base diode failure threshold, or the oxide breakdown voltage BV_{ox} is reached.

Ideally, once the LNPN begins to turn on, the transistor enters a negative resistance region, and the voltage will continue to reduce until other factors require V_j to increase. The conductivity modulation of R_{sub} and the β degradation are effects that will need to be compensated by a higher M and, therefore, an increase in V_j to sustain bipolar action and contribute to a positive on-resistance in the snapback region.

In very short-channel MOS transistors there is some influence of the drain depletion region on the source barrier. The effect of drain induced barrier lowering (DIBL) [Sze81] is to help reduce the source-substrate (emitter-base) barrier to electron injection and is analogous to increasing the substrate potential [Amerasekera94A][Amerasekera94B]. This will be particularly effective if the source barrier lowering occurs deeper in the junction thereby increasing the emitter area and the effective size of the bipolar transistor.

The engine of the high current bipolar action is the avalanche generation at the collector-base junction. It provides the current to raise the substrate potential and forward-bias the base-emitter junction, as well as the base current required to sustain the bipolar action. The application of an external voltage to the substrate will strongly influence the magnitude of I_{gen} required to trigger bipolar action [Amerasekera95]. An external substrate voltage will mean that the I_{gen} required to sustain bipolar action can be reduced since the I_{sub} needed to forward-bias the base-emitter junction is much less, and if the bias voltage is sufficiently high, then the bipolar can be initiated even without avalanche action. This effect has been successfully used in ESD protection circuits for deep submicron CMOS technologies [Amerasekera95] as well as bipolar/BiCMOS technologies [Chen96].

4.6 ELECTROTHERMAL EFFECTS

The previous section dealt with the physics of transistor operation under isothermal conditions where the components were considered to be at a constant temperature. For the most part, the analysis enables us to understand what parameters are critical to transistor behavior and how these need to be optimized for best high current operation. However, the limiting factor in high current operation is the onset of damage owing to thermal breakdown in the transistor. The details of thermal breakdown are discussed in Chapter 10. In this section, we review the thermal effects related to high current bipolar action. For simplicity, we use an nMOS transistor for our analysis.

The high current $I-V$ curve of an nMOS transistor with gate, source, and substrate at 0 V, and current injected into the drain is shown in Figure 4.18. It is seen that after snapback occurs, the transistor is in a low impedance state and the voltage increases very little with current. The on-resistance of the LNPN in this region R_{on} is because of three factors: (1) the increase in M (proportional to V_D) required to

compensate for a reduction in R_{sub} and β at high I_D; (2) the collector (drain) resistance; and (3) high-temperature effects on M, β, R_{sub}, and the drain resistance. As the current increases, the power dissipation causes the temperature in the device to rise and eventually thermal breakdown and silicon melting can occur. At the onset of thermal breakdown, the voltage drops for the second time, and this condition is known as *second breakdown* [Schafft62][Scarlett63]. The voltage and current at which second breakdown takes place is denoted by V_{t2} and I_{t2}, respectively. As I_{t2} is a measure of the maximum high current capability of the transistor, it is commonly used as the figure of merit for process-dependent ESD capability.

The primary heat source driving the thermal effects is the power dissipation at the reverse-biased collector-base junction. I_{t2} is a function of the power density in the junction which at the terminal can be approximated by $V_D \times I_D$. At the junction itself, the power density can be more accurately written as $J \cdot E$, where J is the current density across the junction and E is the electric field. For a given junction, $J \cdot E$ is a maximum at locations where J and E are both high. Hence, the current spreading across the junction will determine the power density. A larger active collector area will improve I_{t2}, as will more current spreading. This is the why the drain (or collector) ballast resistance can have a significant influence on the ESD performance. Even in transistors where the ballast resistance appears to be low, the location of the peak E fields and the current flow lines across the junction are important factors defining the power density.

It has been shown [Amerasekera94B] that at the onset of thermal breakdown, the base region of the bipolar is highly conductivity modulated, and that the point at which the carrier concentration in the drain junction is equal to the background doping concentration (the intrinsic condition) is not where the voltage collapse begins. Instead, the second voltage drop occurs when the thermally generated carriers become significant in comparison to the avalanche generated carriers. That is, when the thermal generation current I_{th} approaches the avalanche generation current I_{gen}, then the bipolar can be sustained with a lower I_{gen} and M and V_j can decrease. Figure 4.23 shows I_{gen} and I_{th} as a function of temperature for an LNPN. As the temperature approaches second breakdown, we see a very rapid increase in I_{th}, and an associated rapid drop in I_{gen}.

Once second breakdown is initiated, the transistor enters a fundamentally unstable negative resistance region [Ridley63][Shaw92]. Current filamentation and local hot spots begin to occur and the temperature in these regions very quickly reaches the semiconductor melt temperature and permanent damage is incurred.

I_{t2} is strongly dependent on the main snapback parameters M, β, and R_{sub} as discussed in Section 4.5. Lower values of M will require more power dissipation to provide the necessary I_{gen} to operate the bipolar. A lower β will require more I_{gen} to support a given injection current and, therefore, a higher M and a higher V_D. If R_{sub} is too low, then I_{gen} needs to be higher to raise the V_{sub} to the level necessary to maintain the bipolar action, which again requires a higher M, and a higher V_D with the associated higher power dissipation.

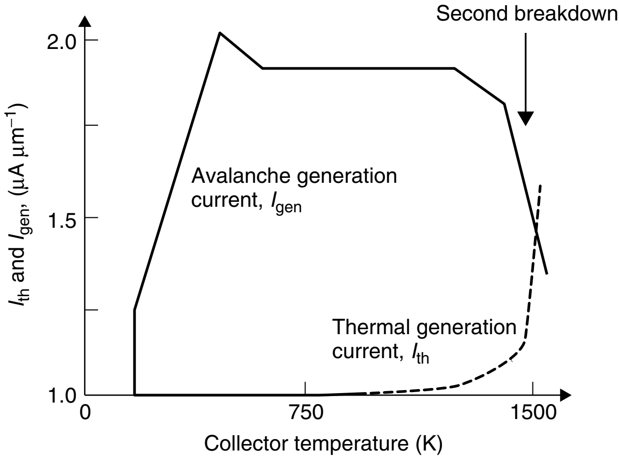


Figure 4.23 The thermal generation current I_{th} and the avalanche generation current I_{gen} as a function of the junction temperature, as observed in electrothermal device simulations. Second breakdown occurs when the current required to support *npn* action is provided by I_{th}, allowing I_{gen} to reduce

Transistor design parameters and operating conditions that tend to reduce M, β and R_{sub} will result in higher I_{t2} and ESD performance. Some typical parameters would be the gate length L, which is related to β of the LNPN. Longer L will mean lower β and possibly lower I_{t2}. Placement of the substrate contact for better current spreading and maximum R_{sub} can help improve I_{t2}. Circuit design techniques that increase V_{sub}, either through additional current injection into the substrate or by raising the local substrate potential will also help to increase I_{t2}.

Techniques that increase V_G above the threshold voltage V_T will provide higher I_{sub} and enable the LNPN to turn on at lower V_j (Figure 4.11). However, for a given V_D as V_G increases, I_{sub} reaches a peak before decreasing (Figure 4.12) and as I_{sub} decreases, an increase in the V_j is required to support snapback. The second breakdown current I_{t2} is seen to track V_j, and decreases initially with V_G, before increasing again as V_G is increased further [Chen98]. Typically, the peak I_{sub} occurs at $V_G \approx V_D/2$, and in snapback it can be expected that I_{t2} will show improvement with V_G until $V_G \approx V_{sp}/2$. For $V_{sp} \approx 5\,\text{V}$, this would mean that I_{t2} would increase until $V_G \approx 2.5\,\text{V}$, and then reduce as V_G is raised beyond that. Note that the increase in I_{t2} with V_G is only observed if the M is low and I_{t2} is low for $V_G < V_T$. If M is high and I_{t2} at $V_G = 0\,\text{V}$ is high to begin with, then very little change in I_{t2} can be expected by increasing V_G. Also, in deep submicron technologies, I_{sub} does not have such a strong peak, and may not reach a maximum even at $V_G = V_D$ in some cases. In these processes, there could be very little dependence of I_{t2} observed as a function of V_G.

In conclusion, the thermal generation current rapidly increases at high temperatures and when it begins to supplement the avalanche generation current to drive the

bipolar action then the second breakdown begins. This negative resistance region is intrinsically unstable and quickly degenerates into current filamentation and local hot spot formation resulting in even higher local temperatures until finally silicon melting occurs with irreversible damage to the device.

4.7 SCR OPERATION

Silicon-controlled rectifiers (SCRs) also known as thyristors are devices that are used extensively in power device applications because of the capability to switch from a very high impedance state to a very low impedance state. For the same reason a properly designed SCR can also be a very efficient ESD protection circuit [Avery83][Chatterjee91B]. A cross section of a simple lateral SCR is shown in Figure 4.24 and essentially consists of a PNPN structure. The p^+ diffusion in the n-well forms the *anode* of the SCR where holes are injected into the n-well. The n^+ diffusion in the p-well forms the *cathode* of the SCR from which electrons are injected into the p-well. The connection to the n-well is made through a n^+ contact in the well, while the p^+ contact in the well is the connection to the p-well.

The SCR may be considered as two bipolar transistors. A *pnp* transistor, $T1$, is formed by the *anode* as emitter, the n-well as base and the p-well as collector. A *npn* transistor, $T2$, is formed by the *cathode* as emitter, the p-well as the base and the n-well as the collector. The SCR may be biased as follows. The n-well is connected to a fixed voltage, V_C, the p-well and the cathode are connected to ground and a voltage V is applied to the *anode*. The $I-V$ curve for the SCR is

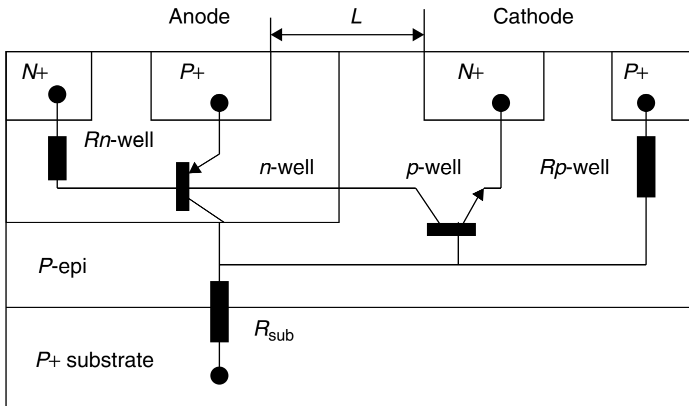


Figure 4.24 Cross section of a lateral SCR in a CMOS process showing the parasitic *pnp* and *npn* transistors. $R_{n-\text{well}}$ is the n-well resistor, $R_{p-\text{well}}$ is the pwell resistor, and R_{epi} is the resistance of the epi layer. The low-resistance (5 mΩ cm) substrate is assumed to be at 0 V

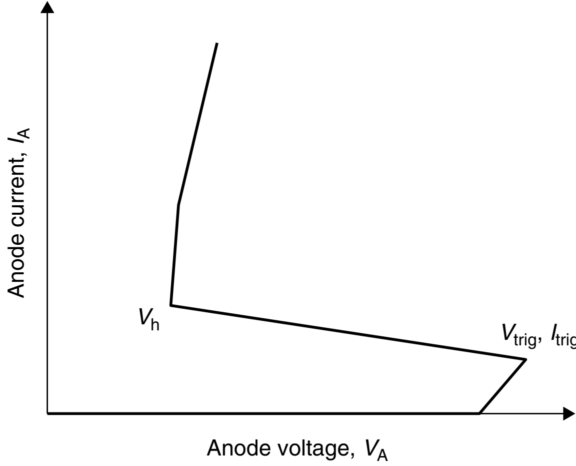


Figure 4.25 High current $I-V$ curve for an SCR showing the trigger voltage and current V_{t1}, I_{t1}, and the holding voltage V_h

shown in Figure 4.25. As V goes above V_C the emitter-base junction of the *pnp* is forward-biased and the *pnp* turns on. The current through the *pnp* flows into the p-well and forward biases the emitter-base junction of the *npn* turning it on. The *npn* current from the n-well to the cathode now supplies the forward-bias for the *pnp* and the voltage at the anode no longer needs to provide the bias for the *pnp* and V begins to decrease resulting in a negative resistance region. The minimum value of V is known as the *holding voltage* V_h defined by the amount of current that the *pnp* needs to supply to forward-bias the *npn* and the base-widths of the lateral *npn* and the lateral *pnp*, which is the anode to cathode spacing L in Figure 4.24.

The two terminal SCR can be approximated by the equivalent circuit shown in Figure 4.26. $R_{n-\text{well}}$ and $R_{p-\text{well}}$ denote the well resistances. that provide the bias to the *pnp* and *npn*, respectively. When the SCR is in the latched mode the requirement that it stays latched is given by the equation [Estreich81]

$$\beta_{npn} \cdot \beta_{pnp} \geq 1 \tag{4.77}$$

β_{npn} and β_{pnp} are the current gains of the *npn* and *pnp*. It must be noted that although the *pnp* and *npn* are shown as discrete transistors, the collector of one is the base of the other. Therefore, it is not possible to use the discrete βs when computing Equation 4.77.

The two important parameters for the SCR are the trigger current, I_{trig}, and V_h. I_{trig} is determined by $R_{p-\text{well}}$, which in turn is determined by the thickness of the epitaxial layer (when present) and the doping of the p-well. V_h is strongly dependent on L, as well as $R_{n-\text{well}}$ and is typically between 2 and 5 V

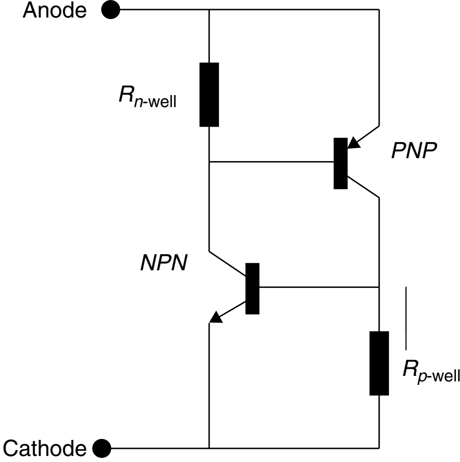


Figure 4.26 Equivalent circuit schematic for the SCR shown in Figure 4.25

in advanced CMOS processes. Once on, the SCR can be modeled as a *p*-i-*n* diode [Herlet66][Chatterjee88][Seitchik87]. The region between the anode and cathode is now fully conductivity modulated, and the on-resistance of the SCR is about 1 Ω making it a low power dissipating device ideal for ESD protection circuits.

When used as an ESD protection circuit, the SCR is connected as a two terminal device, with the anode and *n*-well tied together and the cathode and *p*-well tied together. Triggering now requires avalanche breakdown of the *n*-well to *p* junction. The SCR turns on either when the cathode is forward- biased by the hole current in the *p* region in a manner similar to the triggering of the *npn* in nMOS transistors, or when the *pnp* is turned on by the electron current in the *n*-well. Typically the *npn* gain is an order of magnitude higher than that of the *pnp* at low current levels, and turning the *npn* on is easier to achieve than turning on the *pnp*. The trigger voltage is defined by the avalanche breakdown voltage of the *n*-well to substrate, and the trigger current is the same as for the circuit described in the preceding text.

The avalanche breakdown voltage of the *n*-well to substrate is about 20 V in an advanced CMOS process. In order to make the SCR a good ESD protection circuit the trigger voltage must be reduced. This is accomplished by using an additional n^+ diffusion at the *n*-well edge as shown in Figure 4.27. The breakdown voltage is now reduced to that of the n^+ to substrate, which is about 15 V in a submicron CMOS process. A further reduction of the trigger voltage is achieved by using a gated diode at the *n*-well edge [Chatterjee91B]. The gate of the nMOS transistor that forms the gated diode is connected to the cathode. The trigger voltage for these devices are between 6 and 10 V in submicron processes.

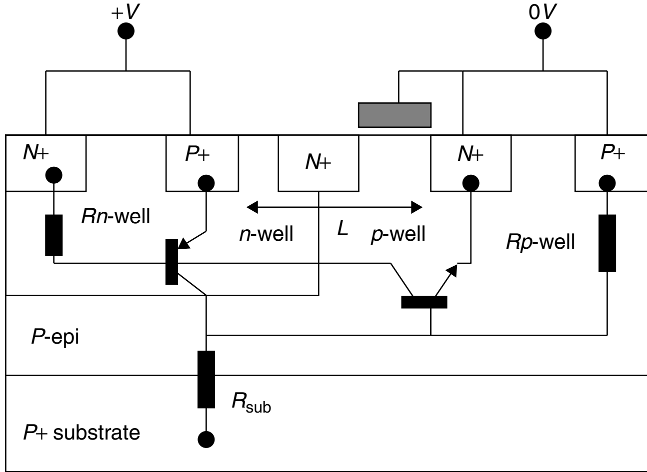


Figure 4.27 Lateral CMOS SCR showing the addition of a $N+$ diffusion at the N-well edge to reduce the SCR trigger voltage

4.8 CONCLUSION

This chapter described the basics of operation of the main components used in ESD protection circuits. Analysis of the high current behavior in these components extended the standard operating regions and showed how the high currents resulted in nonlinear behavior in resistors, diodes, bipolar and MOS transistors. In particular, it was shown that at very high current levels, the injected carriers exceed the background carrier concentrations significantly changing the $I-V$ characteristics of the components.

The operation of bipolar and MOS transistors under high-current ESD conditions was shown to be dependent on the capability of self-biased bipolar operation where avalanche generation current from the collector-base (or drain-substrate) junction provided the current necessary to sustain low impedance bipolar action. In general, the generation current is critical to self-biased bipolar operation, and the parameters M, R_{sub} (or R_{B} for a bipolar junction transistor), and β were parameters that defined I_{gen} and the high-current operation of these devices. The equations governing stable operation in the snapback region have been described and the requirements for good high-current operation have been discussed.

At high temperatures, it was shown that the thermal generation current becomes significant and when it becomes comparable to the avalanche generation current, then a second voltage drop is initiated. The onset of this voltage drop is known as second breakdown and the instability of the negative resistance in this region leads to current filamentation and hot spot formation with eventual permanent damage to the silicon.

Finally, the high-current operation of two terminal SCR devices was presented.

BIBLIOGRAPHY

[Amerasekera92] A. Amerasekera and A. Chatterjee, "An Investigation of BiCMOS ESD Protection Circuit Elements and Applications in Submicron Technologies", in *Proc. 14th EOS/ESD Symposium*, pp. 265–276, 1992.

[Amerasekera93] A. Amerasekera, M.-C. Chang, J. Seitchik, A. Chatterjee, K. Mayaram and J.-H. Chern, "Self-heating effects in basic semiconductor structures", in *IEEE Electr. Device*, **ED-40**, 1836–1844, 1993.

[Amerasekera94A] A. Amerasekera and R. Chapman, "Technology design for high current and ESD robustness in a deep submicron process", in *Elec. Dev. Lett.*, **EDL-15**, 383–385, 1994.

[Amerasekera94B] A. Amerasekera and J. Seitchik, "Electrothermal behavior of deep submicron nMOS transistors under high current snapback (ESD/EOS) conditions", in *Tech. Dig. IEDM*, 455–458, 1994.

[Amerasekera95] A. Amerasekera, C. Duvvury, V. Reddy and M. Rodder, "Substrate triggering and silicide effects on ESD performance and protection circuit design in deep submicron CMOS processes", in *Tech. Dig. IEDM*, 547–550, 1995.

[Amerasekera96] A. Amerasekera, V. McNeil and M. Rodder, "Correlating drain junction scaling, silicide thickness, and lateral *npn* behavior with the ESD/EOS performance of a 0.25 µm CMOS process", in *Tech. Dig. IEDM*, 893–896, 1996.

[Amerasekera99] A. Amerasekera, V. Gupta, K. Vasanth and S. Ramaswamy, "Analysis of snapback behavior on the ESD capability of sub-0.20 µm nMOS", in *Proc. 37th IRPS*, 159–166, 1999.

[Avery83] L.R. Avery, "Using SCRs as Transient Protection Structures in Integrated Circuits", in *Proc. 5th EOS/ESD Symposium*, pp. 177–180, 1983.

[Bock99] K. Bock, B. Keppens, V. De Heyn, G. Groeseneken, L.Y. Ching and A. Naem, "Influence of Gate Length on ESD Performance For Deep Submicron CMOS Technology", in *Proc. 21st EOS/ESD Symposium*, pp. 95–104, 1999.

[Boselli01] G. Boselli, S. Ramaswamy, A. Amerasekera, T. Mouthaan and F. Kuper, "Modeling Substrate Diodes under Ultra-High ESD Injection Conditions", in *Proc. 23rd EOS/ESD Symposium*, pp. 135–144, 2001.

[Caruso74] A. Caruso, P. Spirito and G. Vitale, "Negative resistance induced by avalanche injection in bulk semiconductors", in *IEEE Trans. Electr. Device*, **ED-21**, 578–586, 1974.

[Chatterjee88] A. Chatterjee, J.A. Seitchik, J.-H. Chern, P. Wang and C.-C. Wei, "Direct evidence supporting the premises of a two-dimensional diode model for the parasitic thyristor in CMOS circuits built on thin Epi", in *IEEE Electr. Device Lett.*, **EDL-9**, 509–511, 1988.

[Chatterjee91A] A. Chatterjee, T. Polgreen and A. Amerasekera, "Design and Simulation of a 4 kV ESD Protection Circuit", in *Tech. Dig. IEDM*, 913–916, 1991.

[Chatterjee91B] A. Chatterjee and T. Polgreen, "A low voltage triggering SCR for on-chip ESD protection at output and input pads", in *IEEE Electr. Device Lett.*, **EDL-12**, 21–22, 1991.

[Chen96] J.Z. Chen, X.Y. Zhang, A. Amerasekera and T. Vrotsos, "Design and Layout of a High ESD Performance NPN Structure for Submicron BiMCOS/Bipolar Circuits", in *Proc. 34th IRPS*, pp. 227–232, 1996.

[Chen98] J.Z. Chen, A. Amerasekera and C. Duvvury, "Design Methodology and Optimization of Gate Driven NMOS ESD Protection Circuits in Submicron CMOS Processes", in *IEEE T. Electron Dev.*, **ED-45**, 2448–2456, 1998.

[Corsi93] M. Corsi, R. Nimmo and F. Fattori, "ESD Protection of BiCMOS Integrated Circuits Which Need to Operate in the Harsh Environments of Automotive or Industrial", in *Proc. 15th EOS/ESD Symposium*, pp. 209–214, 1993.

[Crowell66] C. Crowell and S.M. Sze, "Temperature dependence of avalanche multiplication in semiconductors", in *Appl. Phys. Lett.*, **9**, 242–233, 1966.

[Dutton75] R.W. Dutton, "Bipolar transistor modeling of avalanche generation for computer circuit simulation", in *IEEE T. Electron Dev.*, **ED-22**, 334–338, 1975.

[Eitan82] B. Eitan and D. Frohman-Bentchkowsky, "Surface conduction in short-channel MOS devices as a limitation to VLSI scaling", in *IEEE T. Electron Dev.*, **ED-29**, 254–266, 1982.

[Estreich81] D.B. Estreich, *The Physics and Modeling of Latch-Up in CMOS Integrated Circuits*, Ph.D. Thesis, Stanford University, 1981.

[Ghandhi77] S.K. Ghandhi, *Semiconductor Power Devices*, John Wiley & Sons, New York, 1977.

[Grant73] W.N. Grant, "Electron and hole ionization rates in epitaxial silicon at high electric fields", in *J. Solid State Electr.*, **16**, 1189–1203, 1973.

[Gupta98] V. Gupta, A. Amerasekera, S. Ramaswamy and A. Tsao, "ESD-related Process Effects in Mixed Voltage Sub-0.5 µm Technologies", in *Proc. 20th EOS/ESD Symposium*, pp. 161–169, 1998.

[Herlet66] A. Herlet and K. Raithel, "Forward characteristics of thyristors in the fired state", in *J. Solid State Electr.*, **9**, 1089–1105, 1966.

[Hower70] P.L. Hower and V.G.K. Reddi, "Avalanche injection and second breakdown in transistors", in *IEEE T. Electron Dev.*, **ED-17**, 320–335, 1970.

[Hsu82] F.C. Hsu, P.K. Ko, S. Tam, C. Hu and R. Muller, "An analytical breakdown model for short-channel MOSFETs", in *IEEE T. Electron Dev.*, **ED-29**, 1735–1740, 1982.

[Jankovic91] N.D. Jankovic, "Pre-turn-on surce bipolar injection in graded NMOST's", in *IEEE T. Electron Dev.*, **ED-38**, 2527–2530, 1991.

[Khurana66] B.S. Khurana, T. Sugano and H. Yanai, "Thermal breakdown in silicon *p-n* junction devices", in *IEEE T. Electron Dev.*, **ED-13**, 763–770, 1966.

[Krieger89] G. Krieger, "The Dynamics of Electrostatic Discharge Prior to Bipolar Action Related Snapback", in *Proc. 11th EOS/ESD Symposium*, pp. 136–144, 1989.

[Lindmayer67] J. Lindmayer and W. Schneider, "Theory of lateral transistors", in *J. Solid State Electr.*, **10**, 225–234, 1967.

[Miller57] S.L. Miller, "Ionization rates for holes and electrons in silicon", in *Phys. Rev.*, **105**, 1246–1249, 1957.

[Muller86] R.S. Muller and T.I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed., John Wiley & Sons, New York, 1986.

[Okuto75] Y. Okuto and C.R. Crowell, "Threshold energy effect on avalanche breakdown voltage in semiconductor junctions", in *J. Solid State Electr.*, **18**, 161–168, 1975.

[Pinto-Guedes88] M. Pinto-Guedes and P. Chan, "A circuit simulation model for bipolar-induced breakdown in MOSFET", in *IEEE Trans. Comp. Aid. Des.*, **CAD-7**, 289–294, 1988.

[Ramaswamy97] S. Ramaswamy, "A unified substrate current model for weal and strong impact ionization in sub-0.25 um nMOS devices" in *Tech. Dig. IEDM*, 885–888, 1997.

[Reisch92] M. Reisch, "On bistable behavior and open-base breakdown of bipolar transistors in the avalanche regime — modeling and applications", in *IEEE T. Electron Dev.*, **ED-39**, 1398, 1992.

[Ridley63] B.K. Ridley, "Specific negative resistance in solids", in *Proc. Phys. Soc.*, **82**, 954–966, 1963.

[Schafft62] H.A. Schafft and J.C. French, "Second breakdown in transistors", in *IRE Trans. Electr. Device*, **ED-9**, 129–136, 1962.

[Scarlett63] R.M. Scarlett and W. Shockley, "Secondary breakdown and hot spots in power transistors", in *1963 IEEE Int. Conv. Rec. part 3*, 3–13.

[Seitchik87] J. Seitchik, A. Chatterjee and P. Yang, "An analytical model of holding voltage for latch-up in epitaxial CMOS", in *IEEE Electr. Device Lett.*, **EDL-8**, 157–159, 1987.

[Shaw92] M.P. Shaw, V.V. Mitin, E. Schöll and H.L. Grubin, *The Physics of Instabilities in Solid State Electron Devices*, Plenum Publishing, New York, 1992.

[Sze81] S.M. Sze, *Physics of Semiconductor Devices*, 2nd ed., John Wiley & Sons, New York, 1981.

[Verdonckt91] S. Verdonckt-Vandebroek, S.S. Wong, J. Woo and P.K. Ko, "High gain lateral bipolar action in a MOSFET structure", in *IEEE T. Electron Dev.*, **ED-38**, 2487–2496, 1991.

[Yang93] P. Yang and J.-H. Chern, "Design for reliability: the major challenge for VLSI", in *Proc. IEEE*, **81**, 730–744, 1993.

5 ESD Protection Circuit Design Concepts and Strategy

Warren Anderson

According to the ESD testing standards, an ESD event can be delivered between any two pins of an integrated circuit. To adequately protect an IC from damage during an ESD event, an ESD circuit must shunt ESD current between these two pins. While doing so, it must also limit the voltage between the two pins such that devices on the chip do not fail. The generic ESD protection configuration, shown in Figure 5.1 along with a bidirectional I/O circuit, is a standard method to achieve these objectives. The ESD protection can be tailored to meet the operational requirements of the particular I/O while providing ESD immunity to the necessary level.

Of all of the elements in the I/O circuit of Figure 5.1, the output driver nMOS is typically most susceptible to ESD failure, especially for HBM events. The function of the primary ESD clamp is therefore to protect the driver by limiting the pad voltage to a value below which the output driver fails by shunting the majority of the ESD current to a power rail. The series resistor R_s, if present, further limits the current that can flow through the output driver and decreases its drain voltage. The resistor R_{in} and the secondary clamp protect the input receiver's gate oxide, particularly against damage during CDM events. If designed properly for the required ESD withstand level, the clamp network safely consumes the ESD event, protecting the devices in the circuit from damage.

This chapter describes the general principles used to construct ESD protection circuits that not only meet the ESD objectives but also meet the functional objectives of the I/O. The first part of this chapter focuses on the basic concepts that describe the ESD protection ability of the ESD clamps. These concepts can be used to determine the ESD and functional performance of a particular design. The second part of the chapter describes two general classes of ESD

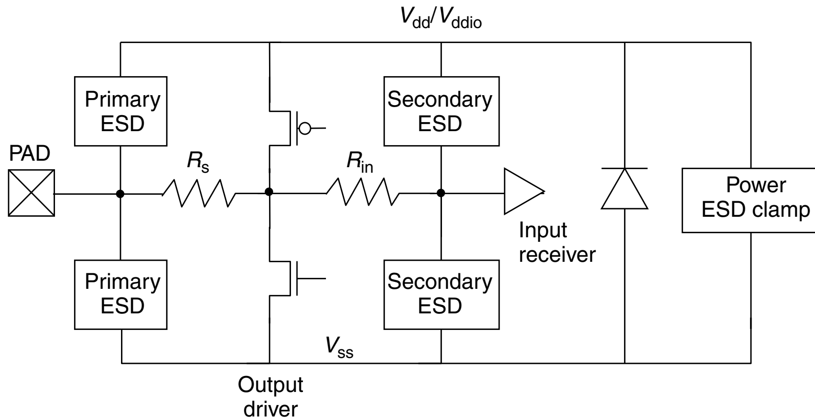


Figure 5.1 The general configuration of the ESD protection in a bidirectional I/O circuit. The primary and secondary protection can be to V_{ss}, to V_{dd}, or to both. The secondary clamp is often a scaled-down version of the primary clamp. The input resistor R_{in} is usually much larger than the output series resistor R_s. The diode from V_{ss} to V_{dd} represents the p-substrate to n-well diode inherent in any CMOS process

protection along with principles for constructing an effective ESD circuit. These strategies can be applied to any number of ESD protection circuit types and configurations.

5.1 THE QUALITIES OF GOOD ESD PROTECTION

The capability of the clamp network can be determined by its performance in four categories: robustness, effectiveness, speed, and transparency. Good ESD protection must function well in all of these areas.

5.1.1 Robustness

Robustness describes the ability of the ESD clamp to handle the ESD current by itself. It is defined as the ESD level at which the clamp, taken on its own, fails. For example, a clamp that can withstand a peak current of 3 A on the HBM timescale has a robustness of 4.5 kV HBM. Robustness is usually, but not always, proportional to the width of the clamp device. Therefore, it is often convenient to measure the breakdown characteristics of a clamp device with TLP or HBM testing and quantify its failure level in milliamperes per micrometer or HBM volts per micrometer. To achieve a particular robustness, one can simply scale the width of the clamp device, assuming that robustness is proportional to device width.

For example, if a particular clamp has a robustness of $10\,\text{mA}\,\mu\text{m}^{-1}$, a total clamp width of $200\,\mu\text{m}$ would be needed to withstand 2 A and thereby achieve a

protection level of 3 kV HBM. In practice, the device should be wider than the HBM target to provide a safety margin for process variations.

5.1.2 Effectiveness

Effectiveness describes the ability of the clamp network to limit the voltage to a safe level such that circuits in parallel with the ESD protection do not fail. If, for example, the clamp alone is robust to 4.5 kV HBM but other elements of the I/O circuit, such as the output driver or some other parasitic path, activate and fail at 2 kV HBM, the ESD clamp is only effective to 2 kV HBM for this particular I/O network. To achieve higher protection levels, the voltage sustained across the ESD clamp must decrease or the turn-on voltage of the failing elements must increase.

Figure 5.2 illustrates several examples of the robustness and effectiveness concepts. The illustrations overlay the $I-V$ characteristics of the ESD clamp and the parallel paths, using a form typical of many types of devices under ESD conditions, namely, a trigger into a nondestructive snapback conduction mechanism at low current and a destructive second breakdown at higher current [Vinson98]. They also assume a desired withstand level of I_{ESD}. As ESD pulses are generally current driven, the current increases rapidly to the I_{ESD} level, with the voltage tracking at the minimum necessary to support the current.

Figure 5.2(a)–(c) shows the characteristics of robust and effective ESD protection networks. In Figure 5.2(a), the ESD protection device triggers and conducts at a lower voltage than the paths in parallel with it. The clamp network is robust because the clamp itself can safely withstand a current of I_{ESD}. Up to its failure current, the clamp limits the voltage to less than the voltage needed for conduction through the parallel paths. Therefore, the clamp network is also effective. Although the parallel paths in Figures 5.2(b) and (c) trigger and conduct below the trigger voltage of the ESD network, they are resistive enough that the ESD network can still trigger and clamp at a voltage below the failure voltage of the parallel paths. Once the ESD network activates, it consumes all of the additional current up to I_{ESD}, limiting the voltage across the parallel paths. Therefore, the protection networks in Figure 5.2(b) and (c) are both robust and effective.

The ESD clamps in Figure 5.2(d)–(f) are robust, but they are ineffective for the particular I/O network in parallel with the clamp. In Figure 5.2(d), the parallel paths trigger and conduct at a lower voltage than the ESD clamp. The parallel paths consume all of the ESD current, failing before the ESD clamp can trigger. In Figure 5.2(e), the ESD clamp triggers and conducts as the ESD current increases. However, the clamp's series resistance is so high that the parallel paths trigger before the ESD current reaches its peak value. Once the parallel paths begin conducting, they consume enough current to fail. Likewise, the ESD protection in Figure 5.2(f) begins conducting below the snapback voltage of the parallel paths, but its series resistance is too high to keep the voltage below that at which the parallel paths fail. A complete effectiveness example is given in [Ting01].

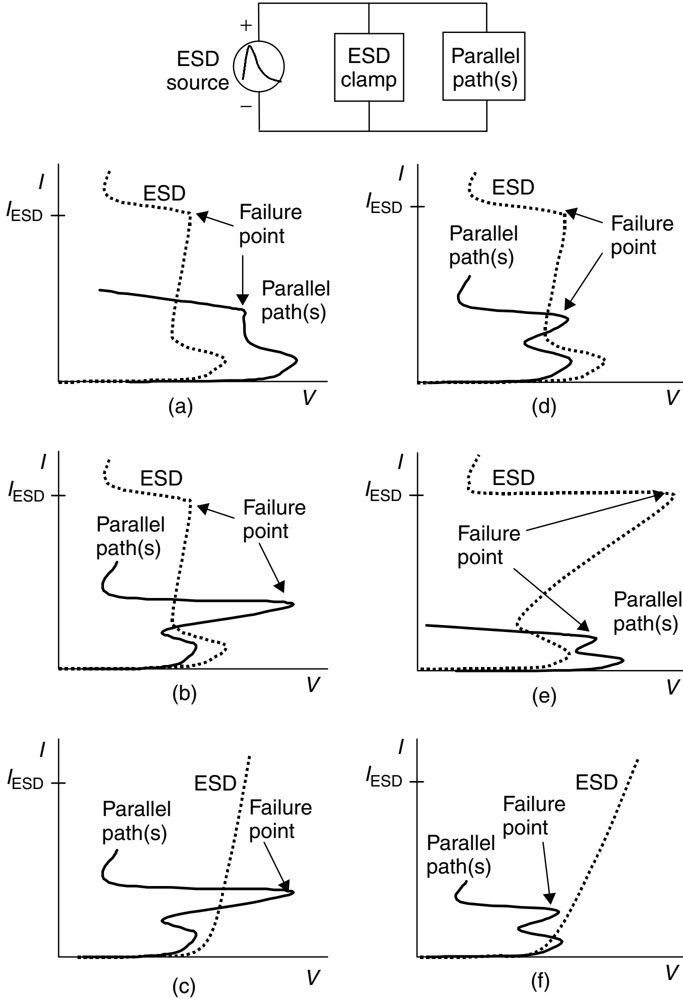


Figure 5.2 Hypothetical examples of (a)–(c) effective and (d)–(f) ineffective ESD protection illustrated through the $I-V$ curves of snapback and nonsnapback devices (After [Thierauf01])

5.1.3 Speed

Even robust and effective ESD networks must activate with enough speed to clamp the ESD event at a safe level. Although this is inherent in nearly all standard protection schemes, some clamps can trigger so slowly that the pad voltage exceeds safe levels for long enough to cause circuits in parallel to fail [Duvvury95][Wu00][Juliano01A]. Speed is of particular importance for CDM events, with their subnanosecond rise times.

5.1.4 Transparency

Transparency, the final quality of good ESD protection, requires that the ESD protection not interfere with the normal operation of the I/O circuit and of the chip itself. This includes the impact of the ESD on I/O parameters and specifications. Examples include:

- Capacitance: the ESD clamp should not have so much capacitance that it violates the loading limits of the I/O signaling specification.
- Leakage: the ESD clamp must not draw excessive current at either high or low input or output levels. A similar requirement usually exists for power supplies.
- Power sequencing: the ESD clamp must be compatible with the normal sequence for applying power.
- Hot swap: when required, the ESD clamp must not interfere with removing a part from a live system.
- Overvoltage conditions: some schemes, such as fail safe and mixed-voltage tolerance, need the ESD to behave a certain way when voltages in excess of the core supply are applied to the I/O.

Transparency becomes particularly important for high-frequency signal pins and low-leakage power supplies.

Section 5.2 shows that, once the characteristics of the ESD and I/O device elements are understood, the concepts of robustness, effectiveness, speed, and transparency can be used to design an ESD protection network and to evaluate its performance.

5.2 ESD PROTECTION DESIGN METHODS

A good ESD protection network must provide a discharge path for all pin combinations and must limit the voltage across any sensitive devices, most particularly the sensitive nMOS output driver. Most ESD solutions rely on shunting charge from an I/O pin to a power supply, from which the charge can be distributed to other I/O pins or supplies. These solutions therefore fall into two general categories: V_{ss}-based ESD protection for ESD approaches that shunt current to the negative supply rail and V_{dd}-based ESD protection for those that shunt current to the positive supply rail. The next section discusses the differences between these two protection methods. The subsequent sections reveal how to quantitatively construct effective ESD protection using either scheme.

5.2.1 V_{ss}-Based Versus V_{dd}-Based Protection

Both methods can be constructed from Figure 5.1 by assuming that ESD protection exists only between the I/O pad and the rail upon which the clamping scheme is

based. For V_{ss}-based protection, the clamp between the I/O pad and V_{ss} is usually a snapback device and there is no clamp to V_{dd}. For V_{dd}-based protection, the clamp is usually a *pnp* emitter-base diode and there is no clamp to V_{ss}. The difference between the methods becomes apparent when examining the ESD discharge path under various pin combinations. Figure 5.3(a) and (b) shows, respectively, the current path for a positive and negative discharge from the I/O pad to V_{ss} for a V_{ss}-based clamp. Figure 5.4 shows the current paths for a V_{dd}-based clamp. Both figures assume that the clamps conduct in the direction from the I/O pad to the V_{dd} and V_{ss} rails only.

Examining the current path for the positive discharge to V_{ss} reveals fundamental differences between the V_{ss}- and V_{dd}-based schemes. For the V_{ss}-based clamp, the

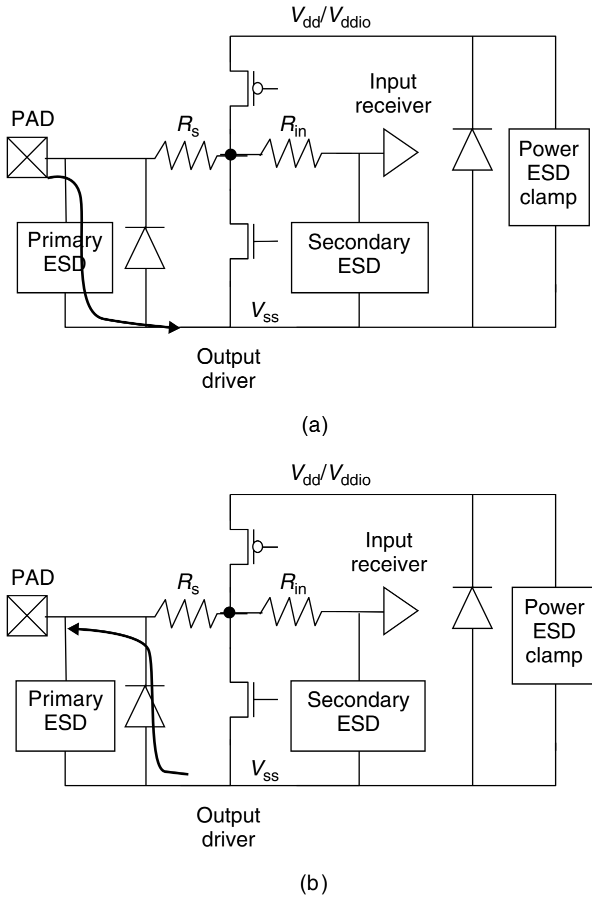


Figure 5.3 Current path for an (a) positive and (b) negative discharge from the I/O pad to V_{SS} for a V_{SS}-based clamp. The arrows indicate the path of current flow, assuming that the protection clamps only conduct in the direction from the I/O pad to the power rails

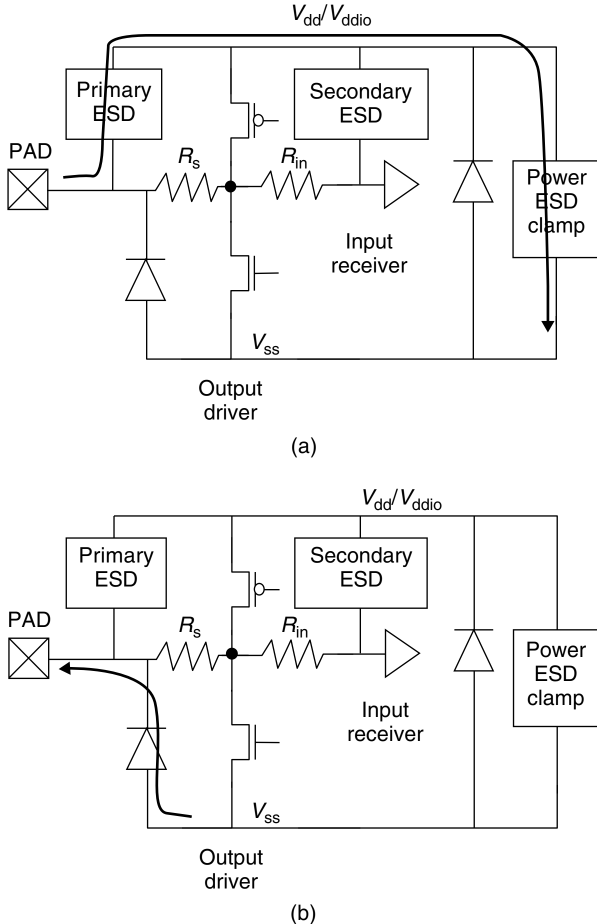


Figure 5.4 Current path for an (a) positive and (b) negative discharge from the I/O pad to V_{ss} for a V_{dd}-based clamp

ESD current flows directly through the clamp. No such direct path exists for the V_{dd}-based clamp, however. The current must flow onto the V_{dd} rail and through the V_{dd} power supply clamp to reach V_{ss}. In both cases, the discharge also appears in parallel across the nMOS output driver and the output series resistor R_s, if present. The objective of ESD protection is to restrict the I/O pad voltage below the failure voltage of device(s) in parallel, in this case the nMOS output driver. Therefore, in the V_{ss}-based scheme, the primary clamp alone must hold the I/O pad voltage below this limit. In the V_{dd}-based scheme, the current discharges through the primary clamp to the V_{dd} rail, then through the V_{dd} supply clamp. Therefore, these two clamps in series must meet the same voltage-limiting criterion to protect the nMOS output driver.

For discharges from I/O to I/O, although not shown here, the same path to V_{ss} usually dominates, particularly for cases where V_{ss} connects to a common p substrate. To get to another I/O pad, the ESD current, after reaching V_{ss}, can forward bias the diode from V_{ss} to the I/O pad. The diode usually exists implicitly as part of the nMOS output driver in most technologies. If V_{ss} connects to a p substrate, any node with an n^+ diffusion in the substrate will form a diode through which the ESD charge can flow. Therefore, the V_{ss} supply can usually distribute the ESD current to any pin at the negative terminal. In most cases, the nMOS output driver is still the most susceptible device in the parallel path. Therefore, common wisdom asserts that if the I/O-to-V_{ss} discharge combination passes an HBM test, the I/O-to-I/O combination will as well.

For the negative discharge from I/O to V_{ss}, the I/O diode conducts the ESD current for both the V_{ss}- and V_{dd}-based approaches. The nMOS driver still appears in parallel, with connections reversed. Provided the I/O pad contacts enough n^+ diffusion area, the I/O diode should have such a small series resistance that the nMOS driver is easily protected.

The final important case is the discharge between the I/O pad and V_{dd}. Figures 5.5 and 5.6 show the current flow for the V_{ss}- and V_{dd}-based approaches, respectively.

The positive discharge paths reveal further differences between the two schemes. For the V_{ss}-based clamp, the current must flow through the primary clamp to V_{ss}, then through a diode to V_{dd}. The diode is implicit in most technologies, for example, the CMOS n-well diffusion to the p substrate. In the V_{dd}-based primary clamp case, the current simply flows from the I/O pad to V_{dd} directly through the clamp. This time, the path in the V_{dd}-based scheme traverses a single clamp, whereas the path in the V_{ss}-based scheme goes through a clamp and a diode. The discharge occurs in parallel across the pMOS output driver, which can generally tolerate a higher drain-source voltage than the nMOS. Therefore, the extra voltage drop from the diode does not usually create a problem for the V_{ss}-based scheme. For the negative discharge, the current path is the same in both cases, from V_{dd} to V_{ss} through the V_{dd} supply clamp, then through a diode to the I/O pad.

This high-level discussion of ESD protection design schemes has brought two concepts to light. First, understanding the discharge path through the ESD protection network reveals which clamps and devices activate under which pin combinations, as well as which devices in parallel require protection. It is the beginning of constructing ESD protection that meets the effectiveness criterion. This discussion progresses in the next section. Second, comparison of the discharge paths allows comparison of different schemes for ESD protection. Of course, it is also possible to design protection using both methods in tandem, albeit with the added cost of increased circuit area and development time.

5.2.2 V_{SS}-Based Circuit Synthesis with a Snapback Device

In the V_{ss}-based ESD protection approach, a primary clamp exists on every I/O pad between the pad and V_{ss}. The clamp itself is usually a snapback device, such

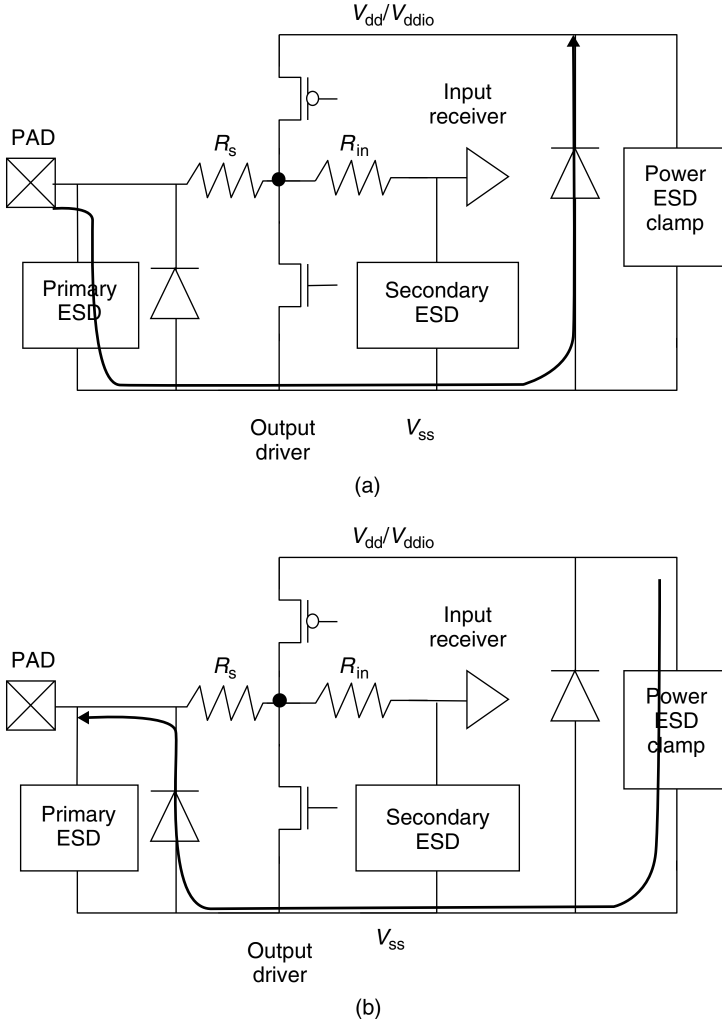


Figure 5.5 Current path for a (a) positive and (b) negative discharge from the I/O pad to V_{dd} for a V_{ss}-based clamp. The arrows indicate the path of the current flow, assuming that the protection clamps only conduct in the direction indicated by the arrows

as an nMOS (Section 6.3), a GCNMOS (Section 6.4), a cascoded nMOS stack (Section 7.5.3), or an SCR (Section 6.6). Figure 5.7(a) shows an example circuit using a grounded-gate nMOS.

To meet the effectiveness criterion, 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 nMOS output driver and should hold the pad voltage below the holding voltage of the nMOS driver, as

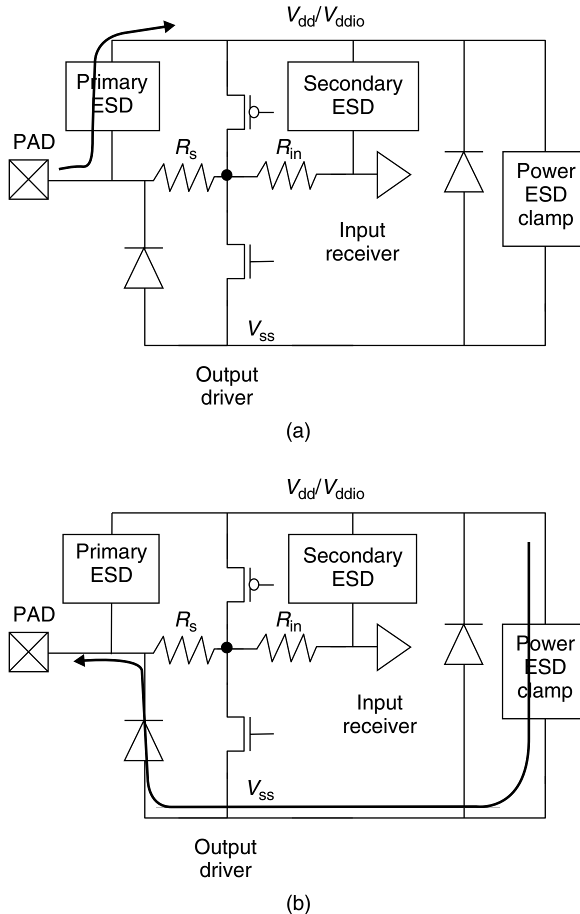


Figure 5.6 Current path for a (a) positive and (b) negative discharge from the I/O pad to V_{dd} for a V_{dd}-based clamp

shown in Figure 5.7(b). In case the clamp accidentally triggers in normal operation, the ideal clamp should also limit the pad voltage to a level slightly higher than V_{op-max}, the voltage seen on the pin during worst-case operating conditions, such as burn-in with I/O reflections.

Unfortunately, no such ideal clamp exists. Real devices offer little ability to so finely tune the clamping voltage. The $I-V$ characteristics of a realistic output driver and ESD clamp are usually closer to one another as both output driver and ESD protection are nMOS devices. However, the two may have important differences. Silicide blocking is usually applied to the grounded-gate nMOS ESD protection device, increasing its series resistance. The gates may behave differently during the ESD event. The output driver and ESD devices may have different

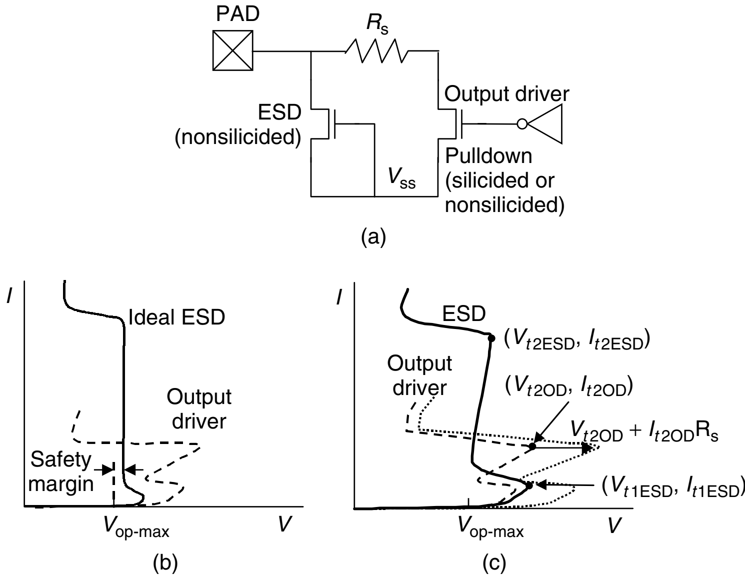


Figure 5.7 The configuration for V_{SS}-based ESD protection using a grounded-gate nMOS snapback device: (a) schematic diagram, assuming there is no pMOS output driver or that it does not participate in the ESD event; (b) desired $I-V$ curve for ideal snapback ESD protection; and (c) actual $I-V$ curves for realistic devices. In (c) the series resistor R_S shifts the output driver failure voltage, as seen at the pad, to the right by $I_{t2OD}R_S$ (After [Thierauf01])

channel lengths, and a resistance R_S can often be added in series with the output driver. All of these features can cause the $I-V$ curves to deviate from one another.

These deviations can be combined to produce effective ESD protection. Effectiveness obligates the ESD device to trigger into snapback and to clamp the pad to a voltage below that at which the output driver path fails. Assuming that the output driver triggers into snapback[1] and can withstand some snapback current, we can add resistance to the output driver circuit branch such that the ESD path triggers at a voltage below the failure voltage of the output driver path:

$$V_{t1ESD} < V_{t2OD} + I_{t2OD}R_S \tag{5.1}$$

Likewise, the ESD clamp should limit the voltage across the output driver path. If the output driver fails at the second breakdown point (V_{t2OD}, I_{t2OD}), the voltage V_{t2ESD} across the ESD device must satisfy

$$V_{t2ESD} < V_{t2OD} + I_{t2OD}R_S \tag{5.2}$$

[1] This is a conservative assumption if V_{t1OD} is less than or close to either V_{t1ESD} or V_{t2ESD}.

where $(V_{t1\text{ESD}}, I_{t1\text{ESD}})$ and $(V_{t2\text{ESD}}, I_{t2\text{ESD}})$ define the trigger point and second breakdown point, respectively, for the ESD clamp. Equations 5.1 and 5.2 can be rewritten to specify a minimum series resistance

$$R_\text{s} > \frac{V_{t\text{maxESD}} - V_{t2\text{OD}}}{I_{t2\text{OD}}} \tag{5.3}$$

where $V_{t\text{maxESD}}$ is the greater of $V_{t1\text{ESD}}$ or $V_{t2\text{ESD}}$. For example, if $V_{t1\text{ESD}} = 11\,\text{V}$, $V_{t2\text{OD}} = 9\,\text{V}$, and $I_{t2\text{OD}} = 200\,\text{mA}$, then $R_\text{s} \geq 10\,\Omega$. Alternatively, if signal integrity requirements specify R_s, Equations 5.1 and 5.2 can be rewritten to specify a minimum output driver I_{t2} and therefore a minimum output driver width.

The resistance requirement is particularly important for nMOS snapback ESD protection in silicide-blocked processes, where the output driver may be fully silicided to minimize its series resistance, causing it to have snapback $I-V$ characteristics significantly different from the ESD device. The resistance R_s may either be added explicitly or be incorporated into the drain junction of the output driver by increasing its contact-to-gate spacing and drawing the silicide block layer around the driver, if necessary. An explicit resistor that conducts enough current during ESD may enter its saturation region (see Section 4.2). This device offers a substantial advantage as the resistor may have a high effective resistance during ESD and a lower resistance during normal operation. During an ESD event, though, the output driver path must safely tolerate enough current to force the resistor into saturation.

5.2.3 V_{SS}-Based Circuit Synthesis with a GCNMOS or GDNMOS

The gate-coupled nMOS (GCNMOS), gate-driven nMOS (GDNMOS), and other advanced techniques drive the gate of the ESD protection device above threshold to ensure that all fingers enter snapback uniformly (see Sections 6.3 and 6.5). They provide the ability to use a silicided, LDD nMOS device for ESD protection, saving the added processing and cost of the silicide block steps. As such, the output driver and the ESD device, for the typical configuration shown in Figure 5.8(a), are nominally identical, with the exception of geometry and the behavior of the gate.

These properties simplify the synthesis of the GCNMOS or GDNMOS ESD protection network. As usual, the ESD protection device should be sized so that it can safely shunt the entire ESD current. The gate modulation circuit should be sized to adequately drive the gate of the protection nMOS. If the ESD device and the output driver pulldown are constructed in the same fashion and have the same source connection, they will have the same snapback $I-V$ characteristics and therefore the same second breakdown voltage V_{t2}, as shown in Figure 5.8(b). Only the behavior of the gate forces the ESD current through the clamp without activating the output driver. Therefore, in many cases, no series resistor is needed. A notable exception is if the source connections are different, for instance, with a

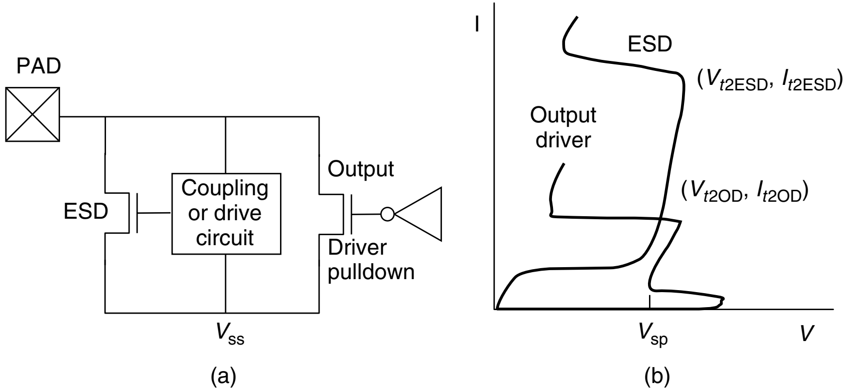


Figure 5.8 The configuration for V_{SS}-based ESD protection with a GCNMOS or GDNMOS clamp: (a) schematic diagram and (b) hypothetical $I-V$ curves for the output driver and the ESD clamp

split supply output driver. The failure voltage of the path through the output driver must be increased, either by added series resistance or by output driver construction, to avoid failures on a discharge between the I/O pad and the output driver's source supply.

5.2.4 V_{SS}-Based Input Protection

The input gate protection, consisting of a series resistor R_{ESD} and the secondary clamp as shown in Figure 5.9(a), limits the voltage across the input receiver's gate oxide.

It functions primarily for CDM discharges, which have a higher current but a shorter duration than HBM events. With the shorter pulse duration, the primary clamp fails at a much higher voltage and current ($V_{t2\mathrm{CDM}}$, $I_{t2\mathrm{CDM}}$) compared to the HBM failure point ($V_{t2\mathrm{HBM}}$, $I_{t2\mathrm{HBM}}$), as shown in Figure 5.9(b). At the current I_{CDM} delivered by a negative CDM pulse, a current I_{p} flows through the primary clamp, which limits the pad to a voltage V_{pad}. A current I_{s} flows through the secondary clamp, which generates a voltage V_{IN} at the input receiver. If the secondary protection clamp exhibits an extrapolated $I = 0$ snapback holding voltage $V_{\mathrm{sp-s}}$ and a linear on resistance $R_{\mathrm{on-s}}$, the current through the secondary clamp is given by

$$I_{\mathrm{s}} = \frac{V_{\mathrm{pad}} - V_{\mathrm{sp-s}}}{R_{\mathrm{ESD}} + R_{\mathrm{on-s}}} \tag{5.4}$$

The voltage across the input receiver's gate oxide is therefore

$$V_{\mathrm{IN}} = \frac{V_{\mathrm{sp-s}} R_{\mathrm{ESD}} + V_{\mathrm{pad}} R_{\mathrm{on-s}}}{R_{\mathrm{ESD}} + R_{\mathrm{on-s}}} \tag{5.5}$$

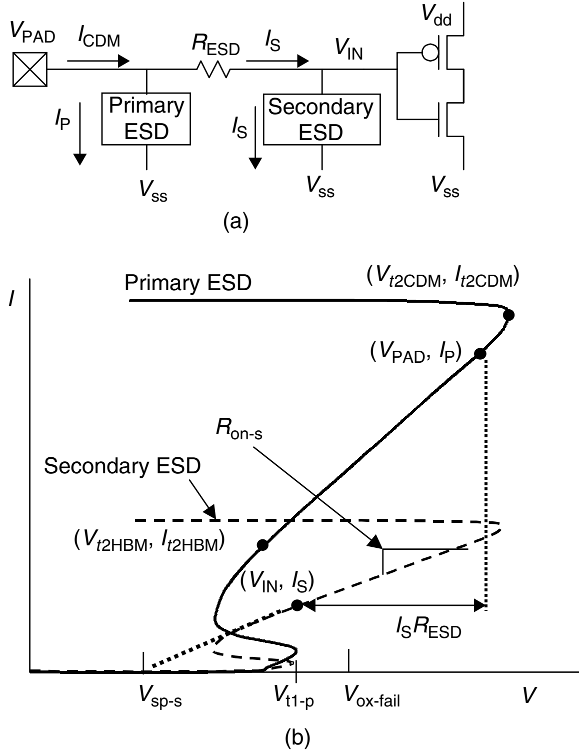


Figure 5.9 The input gate ESD configuration for V_{ss}-based input protection: (a) circuit diagram and (b) CDM timescale $I-V$ characteristics and key operating points of the primary and the secondary clamp devices

which must be less than the oxide breakdown voltage $V_{ox-fail}$ at the duration of the CDM event. In practice, it is difficult, but possible, to model the high-current device behavior [Ramaswamy96B][Gieser96][Wolf99], the short-duration oxide failure [Wu00][Fong89], and the package characteristics that form the CDM pulse [Russ96][Beebe98]. If this data is not available, a good rule of thumb is to set $R_{ESD} \gg R_{on-s}$. A typical implementation will use an R_{ESD} of 100–300 Ω with the R_{ESD} times the width of the secondary clamp of about 1000–1500 Ω-μm for an HBM timescale I_{t2} of 5 mA μm^{-1}. For high-speed I/O, R_{ESD} must not be large enough to cause an appreciable RC delay to the input receiver.

This analysis must also consider the voltage across the input receiver's pMOS, $V_{IN} - V_{dd}$. As a typical CDM event discharges less than about 5 nC during a half cycle, even a modest $V_{dd} - V_{ss}$ decoupling capacitance, greater than 20 nF, will keep V_{dd} within a few hundred millivolts of V_{ss}. For chips with smaller decoupling capacitance, V_{dd} will float, held below V_{ss} only by the V_{ss}-to-V_{dd} diode. An additional 1 V or so may appear across the input receiver's pMOS compared

to the nMOS. Therefore, adding explicit protection to V_{dd} is recommended (see Figure 6.27, e.g.).

The secondary clamp must also safely tolerate its current during a HBM event. At the failure voltage of the primary clamp V_{t2HBM}, the current through the secondary clamp is given by Equation 5.4 with $V_{pad} = V_{t2HBM}$. Furthermore, if the snapback trigger voltage of the secondary clamp V_{t1-s} is less than that of the primary clamp V_{t1-p}, the secondary clamp may snap back before the primary clamp. In this case, the maximum current through the secondary clamp occurs when $V_{pad} = V_{t1-p}$. In both the cases, the current through the secondary clamp must not exceed breakdown: $I_s < I_{t2-s}$.

5.2.5 V_{dd}-Based Circuit Synthesis

In the V_{dd}-based ESD protection approach, a primary clamp exists on every I/O pad between the pad and V_{dd}. The clamp itself is usually the emitter-base diode of a *pnp* or a Darlington-connected *pnp* stack. It conducts in the direction from the pad to the V_{dd} supply only, relying on the power supply clamp to shunt current to V_{ss}. Figure 5.10(a) shows an example circuit using a single *pnp*.

The effectiveness criterion requires that the clamp limit the voltage across the sensitive nMOS output driver during a discharge from the I/O pad to V_{ss}. Although a voltage limit of the nMOS failure voltage V_{t2} can be used if snapback operation is reliable, many advanced CMOS processes have such poor snapback characteristics that the nMOS should not enter snapback at all [Smith99][Miller00]. In this case, the clamp network must limit the voltage below V_{sp}, the nMOS snapback holding voltage. As shown in Figure 5.4(a), the series combination of the primary I/O clamp and the V_{dd} supply clamp determines the voltage across the nMOS output driver during a pad-V_{ss} ESD event. In an ideal ESD design, the primary I/O clamp (the *pnp*) should conduct when the pad voltage rises above V_{dd} and the V_{dd} supply

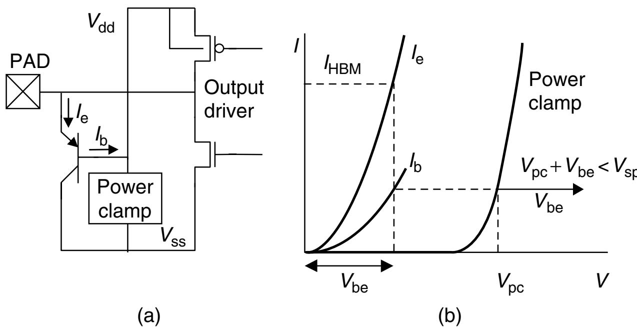


Figure 5.10 The configuration for V_{dd}-based ESD protection using a single *pnp* emitter-base diode: (a) schematic diagram and (b) $I-V$ curves for the *pnp* and for the power supply clamp, which are used to construct the voltage drop across the nMOS output driver on a discharge from the pad to V_{ss} (After [Thierauf01])

clamp should conduct when the V_{dd} supply rail rises slightly above the normal operating voltage. Alternatively, the power supply clamp could activate when it detects a transient characteristic of ESD on the V_{dd} rail. However, real device behavior increases this voltage through several means. The voltage across *pnp* includes the emitter-base diode turn-on voltage of about 0.7 V as well as its series resistance. The power supply clamp and the metal interconnections to the I/O pad are also resistive.

The total voltage across the ESD conduction path can be found either graphically or through SPICE. The graphical process is illustrated in Figure 5.10(b), where the *pnp* and the power clamp's $I-V$ characteristics are overlaid. The ESD current, for example, 2 A for a 3-kV HBM event, determines the voltage V_{be} across the emitter-base junction of the *pnp*. At this emitter current, a base current of $I_b = I_e/(\beta + 1)$ flows out of the base junction into the power clamp, which drops a corresponding voltage V_{pc}. The sum of these two voltages should be less than the nMOS snapback holding voltage V_{sp}.

This same process allows one to determine the appropriate size for the *pnp* device and the power clamp. If the geometry dependence is known for both these devices from the characterization of appropriate test structures [Sematech98][Voldman99], there should be a number of choices that satisfy the constraint

$$V_{be}(I_e) + V_{pc}(I_b) + I_b R_{metal} < V_{sp} \tag{5.6}$$

where R_{metal} is the resistance of the V_{dd} metal bus between the *pnp* base and the power clamp.

As implied in this process, one can trade off the size of the *pnp* and the power supply clamp to achieve the same clamping limit. As one power clamp may serve the power rail for many I/O pads, weighting this trade-off strongly towards a larger power clamp will often result in the smallest overall die area for ESD. To avoid the metal bus resistance contributing significantly to the voltage in the ESD path, it is often advantageous to distribute multiple copies along the bus, as shown in Figure 5.11, rather than simply increasing the size of one [Anderson98B][Torres01].

In many cases, other constraints may influence the size of the *pnp*. The *pnp* and n^+ diffusion diodes can often double as clamps for limiting I/O overshoot and undershoot from reflections in the bus environment. The size of these reflections and the over- and undervoltage limits of the technology may set requirements for the diode size. Implementation specifies for the *pnp* are given in Section 6.9 and for the power clamp in Section 6.10.

5.2.6 V_{dd}-Based Input Protection

The configuration and function of V_{dd}-based input gate protection is similar to V_{ss}-based input protection (see Section 5.2.4). As shown in Figure 5.12(a), the series resistor R_{ESD} and the secondary diode limit the voltage across the input receiver's gate oxide.

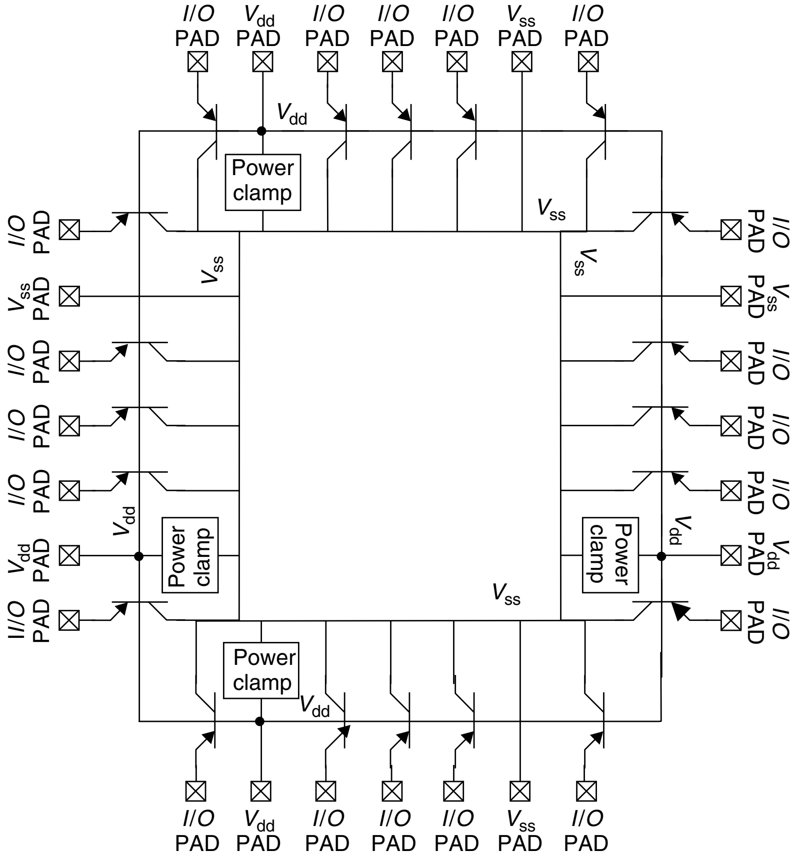
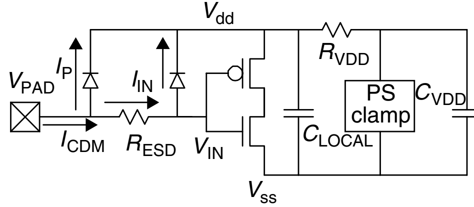


Figure 5.11 Distribution of multiple copies of a power supply clamp around the perimeter of a wire-bond chip will decrease the V_{dd} metal resistance from an I/O pad to the power clamps

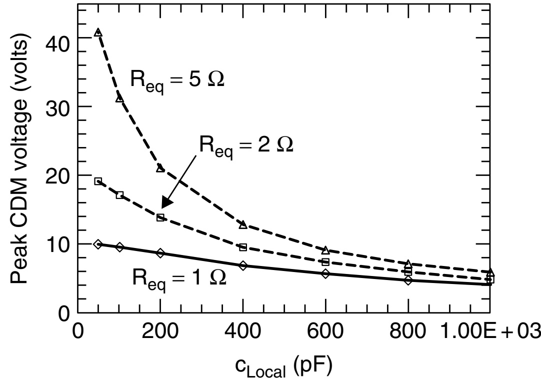
During a negative CDM discharge, the secondary diode clamps V_{IN} a diode drop above V_{dd}, easily protecting the input receiver's pMOS device. As no direct current path to V_{ss} exists, the input receiver's nMOS is much more vulnerable. If C_{local} and C_{VDD} are zero, the voltage across the nMOS is given by

$$V_{IN} - V_{ss} = V_s(I_{IN}) + I_{CDM}R_{VDD} + V_{psclamp}(I_{CDM}) \tag{5.7}$$

where $V_s(I_{IN})$ is the voltage drop across the secondary clamp for a current I_{IN} and $V_{psclamp}(I_{CDM})$ is the voltage drop across the power supply clamp for a current I_{CDM}. This voltage must be less than the breakdown voltage of the gate oxide on the CDM timescale. With no local decoupling capacitance, the entire CDM current flows through the power supply resistance R_{VDD} to the power supply clamp, making a low R_{VDD} critical for avoiding input failure.



(a)



(b)

Figure 5.12 The input gate ESD configuration for V_{dd}-based input protection: (a) circuit diagram and (b) peak voltage using Equation 5.8 for $I_0 = 10\,\text{A}$, $f = 500\,\text{MHz}$, $C_{VDD} = 0$, and various values of equivalent resistance R and C_{local}

Both local and global decoupling capacitance can reduce the peak voltage by shunting the transient current. With $C_{VDD} = 0$, the peak V_{dd} voltage for a peak CDM current I_0 is, by very close approximation,

$$V_{peak} = \frac{I_0 R_{eq}}{\sqrt{1 + \omega^2 R_{eq}^2 C_{local}^2}} \left\{ 1 + \omega R_{eq} C_{local} \times \exp\left[-\frac{1}{R_{eq} C_{local}} \left(\frac{1}{\omega} \arctan\left(\frac{-1}{\omega R_{eq} C_{local}} \right) + \frac{\pi}{\omega} \right) \right] \right\} \tag{5.8}$$

where $\omega = 2\pi f$, f is the frequency of the CDM current oscillation, and R_{eq} is the equivalent resistance of the power supply clamp plus R_{VDD}. Figure 5.12(b) shows V_{peak} as a function of C_{local} for various values of R_{eq}, using a peak CDM current of 10 A and a frequency of 500 MHz. A large local decoupling capacitance can be of substantial benefit. For nonzero values of C_{VDD}, the voltage response is best determined using SPICE.

5.3 SELECTING AN ESD STRATEGY

A wide array of ESD protection options is available. Selecting one option as the foundation for ESD protection requires knowledge of the devices, the process technology, and the circuits to be protected. A test chip provides the best vehicle for assessing the device characteristics, their layout sensitivities, and their geometrical scaling properties [Sematech98][Voldman99]. The test chip should contain not only the ESD protection elements themselves, but also the devices that constitute the output driver and the input receiver. Only when the $I-V$ characteristics of each are compared can the effectiveness of an ESD protection strategy be determined. Geometrical parameters such as device length, width, and contact spacing should all be varied to reveal the layout sensitivity. Variations in width are particularly important to determine how the $I-V$ scales with device size.

In CMOS technologies, the selection decision often reduces to the two methods in most widespread use: either a V_{ss}-based strategy using the nMOS in lateral *npn* snapback, often with supporting circuits to ensure uniform finger triggering, or a V_{dd}-based strategy using the vertical *pnp* and a power supply clamp. Both require managing the process technology to control the behavior of parasitic devices.

Which strategy is better depends on many factors. With an $I_{t2} > 5\,\text{mA}\ \mu\text{m}^{-1}$, the nMOS in snapback can be very area efficient. The SCR is even more so. However, the interaction between nMOS device performance, hot carrier immunity, and ESD robustness demands substantial device engineering to achieve acceptable properties for each. Advanced technologies also call for more elaborate support circuitry to trigger the device uniformly and to maintain an acceptable I_{t2} (see Sections 6.4, 6.5, 7.2, 7.3, and 7.4). With the addition of silicide blocking steps and their modest cost, though, the nMOS forms a robust, reliable, and portable protection device.

On the other hand, V_{dd}-based protection using the *pnp* emitter-base diode is straightforward to model and can be simulated in SPICE. The standard Gummel–Poon model is often adequate for the *pnp*. The power supply clamp often operates in normal MOS channel conduction, which is thoroughly modeled and well controlled. The nMOS snapback model, on the other hand, is more complex, as it must include avalanche multiplication as well as the properties of the lateral *npn* (Section 11.2). Some snapback models even contain thermal effects for accurate prediction of nMOS failure [Diaz95].

Since any protection method must still include a clamp on the V_{dd} supply, V_{dd}-based protection takes advantage of the power supply clamp that must exist with any strategy. The effectiveness of the I/O protection in a V_{dd}-based method is intimately connected with placement of the supply clamps and therefore requires V_{dd} bus resistance modeling. With a V_{ss}-based approach, the current through the nMOS snapback clamp bypasses V_{dd}, flowing directly to V_{ss} whose resistance, as it is often shared with a low-resistance substrate, is generally lower. The V_{ss} resistance is also often common with the nMOS driver's source. The larger area

for the *pnp* implies larger pin capacitance and larger leakage. Although the *pnp* provides I/O overshoot protection, its forward diode to V_{dd} creates power supply sequencing issues. In particular, the *pnp* is incompatible with a fail-safe system, where a voltage can be applied to the pin of a nonpowered IC, unless special steps are taken to float the nonpowered IC's V_{dd} bus [Gauthier01].

5.4 SUMMARY

Good ESD protection must exhibit several properties. It must sink the ESD event without becoming damaged while limiting the voltage across parallel sensitive devices. It must react quickly during an ESD event. It must also meet the functional specifications of the I/O. I/O protection clamps can shunt current to the V_{dd} supply using diode devices, to the V_{ss} supply using snapback devices, or to both. The effectiveness of the ESD protection can be determined from quantitative analysis using transient $I-V$ curves. This analysis reveals not only the requirements for the ESD protection but also the advantages and disadvantages of the V_{dd}- and V_{ss}-based methods.

The next two chapters investigate the construction of the protection devices themselves, exploring how to create robust and effective clamps using various types of protection devices.

BIBLIOGRAPHY

[Anderson98B] W.R. Anderson, J.J. Montanaro and N.J. Howorth, "Cross-referenced ESD protection for power supplies", in *Proceedings of the 20th EOS/ESD Symposium*, pp. 86–95, 1998.

[Beebe98] S. Beebe, "Simulation of complete CMOS I/O circuit response to CDM stress", in *Proceedings of the 20th EOS/ESD Symposium*, pp. 259–270, 1998.

[Diaz95] C.H. Diaz, S.-M. Kang and C. Duvvury, in *Modeling of Electrical Overstress in Integrated Circuits*, Kluwer International, Boston, Mass., 1995.

[Duvvury95] C. Duvvury and A. Amerasekera, "Advanced CMOS protection device trigger mechanisms during CDM", in *Proceedings of the 17th EOS/ESD Symposium*, pp. 162–174, 1995.

[Fong89] Y. Fong and C. Hu, "Internal ESD transients in input protection circuits", in *Proceedings of IRPS*, pp. 77–81, 1989.

[Gauthier01] R. Gauthier, W. Stadler, K. Esmark, P. Riess, A. Salman, M. Muhammad and C. Putnam, "Evaluation of diode-based and nMOS/*Lnpn*-based ESD protection strategies in a triple gate oxide thickness 0.13 μm CMOS logic technology", in *Proceedings of the 23rd EOS/ESD Symposium*, pp. 205–215, 2001.

[Gieser96] H. Gieser and M. Haunschild, "Very fast transmission line pulsing of integrated structures and the charged device model", in *Proceedings of the 18th EOS/ESD Symposium*, pp. 85–94, 1996.

[Juliano01A] P.A. Juliano and E. Rosenbaum, "Accurate wafer-level measurement of ESD protection device turn-on using a modified very fast transmission line pulse system", in *IEEE Transactions on Device and Materials Reliability*, Vol. 1, 2001.

[Miller00] J.W. Miller, M.G. Khazhinsky and J.C. Weldon, "Engineering the cascoded nMOS output buffer for maximum V_{t1}", in *Proceedings of the 22nd EOS/ESD Symposium*, pp. 308–317, 2000.

[Ramaswamy96B] S. Ramaswamy, E. Li, E. Rosenbaum and S. Kang, "Circuit-level simulation of CDM-ESD and EOS in submicron MOS devices", in *Proceedings of the 18th EOS/ESD Symposium*, pp. 316–321, 1996.

[Russ96] C. Russ, K. Verhaege, K. Bock, P. Roussel, G. Groseneken and H. Maes, "A compact model for the grounded-gate nMOS behavior under CDM ESD stress", in *Proceedings of the 18th EOS/ESD Symposium*, pp. 302–315, 1996.

[Sematech98] SEMATECH Technology Transfer Document, *"Test Structures for Benchmarking the Electrostatic Discharge (ESD) Robustness of CMOS Technologies"*, *SEMATECH TT 98013452A-TR*, May, 1998.

[Smith99] J.C. Smith, "An antisnapback circuit technique for inhibiting parasitic bipolar conduction during EOS/ESD events", in *Proceedings of the 21st EOS/ESD Symposium*, pp. 62–69, 1999.

[Thierauf01] S.C. Thierauf and W.R. Anderson, "I/O and ESD circuit design" in A. Chandrakasan, W.J. Bowhill and F. Fox, eds., *Design of High-Performance Microprocessor Circuits*, IEEE Press, pp. 377–396, 2001.

[Ting01] L. Ting, C. Duvvury, O. Trevino, J. Schichl and T. Diep, "Integration of TLP analysis for ESD troubleshooting structures", in *Proceedings of the 23rd EOS/ESD Symposium*, pp. 445–452, 2001.

[Torres01] C.A. Torres, J.W. Miller, M. Stockinger, M.D. Akers, M.G. Khazhinsky and J.C. Weldon, "Modular, portable, and easily simulated ESD protection networks for advanced CMOS technologies", in *Proceedings of the 23rd EOS/ESD Symposium*, pp. 82–95, 2001.

[Vinson98] J.E. Vinson and J.J. Liou, "Electrostatic discharge in semiconductor devices: an overview", in *Proceedings of the IEEE*, 86, pp. 397–418, 1998.

[Voldman99] S. Voldman, W. Anderson, R. Ashton, M. Chaine, C. Duvvury, T. Maloney and E. Worley, "A strategy for characterization and evaluation of ESD robustness of CMOS semiconductor technologies", in *Proceedings of the 21st EOS/ESD Symposium*, pp. 212–224, 1999.

[Wolf99] H. Wolf, H. Gieser and W. Wilkening, "Analyzing the switching behavior of ESD-protection transistors by very fast transmission line pulsing", in *Proceedings of the 21st EOS/ESD Symposium*, pp. 28–37, 1999.

[Wu00] J. Wu, P. Juliano and E. Rosenbaum, "Breakdown and latent damage of ultrathin gate oxides under ESD stress conditions", in *Proceedings of the 22nd EOS/ESD Symposium*, pp. 287–295, 2000.

6 Design and Layout Requirements

Charvaka Duvvury, Warren Anderson

6.1 INTRODUCTION

The requirement for ESD protection circuits to carry currents way beyond the levels for which the elements were initially designed results in regions of high thermal dissipation and high electric fields. As discussed in Chapter 5, a good protection circuit needs to be able to withstand the heating effects, sink the large currents during the ESD event, and not be damaged by the ensuing high electric fields. The capability to meet these requirements is critically dependent on the specific design and layout of the protection circuits and the individual elements as well as the circuits that are being protected. Hence, ESD design is considered to be a very layout-intensive activity. Furthermore, while the design concepts have to change with the technology, the layout techniques must be revised to make them compatible. For example, the same output buffer device will require one type of layout for technologies with no silicided diffusions and a different layout for technologies with no silicided diffusions and a different layout can also vary for grounded substrate logic chip technologies versus floating substrate DRAM chips technologies. For instance, in DRAMs, the proximity effects because of the interaction from neighboring diffusions can lead to failure and force a more conservative implementation of the layout [LeBlanc91].

In this chapter, we present and discuss the design and layout techniques for input/output pins. The focus will be on circuits designed in advanced CMOS processes, but we will also present some typical approaches for bipolar and BiCMOS circuits. Specific examples will be given to illustrate approaches to protection design and layout.

The first requirement for a good protection circuit design methodology is the choice of the appropriate type of protection device that is compatible and/or suitable for the technology. Next, the chip function must be considered, and the operating requirements for the circuit being protected must be compatible with the choice of protection circuit. This includes the operating voltage conditions, capacitance

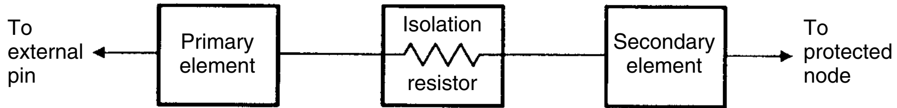


Figure 6.1 Conventional input protection scheme. The primary and the secondary protection devices are separated by an isolation resistor

and resistance loading, and area requirements. Finally, the customer and product engineering requirements for ESD and handling capability must be weighed against the possible impact of the protection circuit on the performance of the IC. For the protection circuit designs described here, each of these aspects is considered in detail.

A common protection circuit schematic is composed of a primary element and a secondary element as shown in Figure 6.1. The primary element will shunt most or all of the current during an ESD event, whereas the secondary element serves to limit the voltage or current at the circuit being protected until the primary device is fully operational. The two elements are isolated by a resistive element. There are several candidates for a primary protection device. It can be a thick field transistor, a silicon controlled rectifier (SCR), an nMOS transistor, or a simple *pn* diode. In this chapter we will discuss some of the commonly used protection devices. In each case, the main features, the advantages, proper methods for layout, and practical limitations will be discussed.

The effectiveness during CDM in most cases is determined by the secondary protection stage working in conjunction with the primary stage. The secondary device can be a small grounded gate MOS transistor or a diode between the pads and the power/ground supplies. The overall design critically depends on the choice of these devices along with proper selection of the resistor element. The resistor element can be polysilicon, n^+ diffusion, p^+ diffusion, or n-well. If a zener diode is to be used as a secondary device in a standard CMOS process, then special process steps are needed to build the diode.

The protection devices mentioned previously are described in the following sections. The effectiveness of the total protection concept is realized only when all of the components are harmonized to work together during an ESD event. To achieve this, a design synthesis approach is needed as illustrated in Section 6.8.

6.2 THICK FIELD DEVICE

The thick oxide or field oxide device (FOD) had been used as a protection circuit element for technologies with feature sizes (defined by the nominal polysilicon gate length) ranging from 3 μm to 1 μm. The standard cross section of this device is shown in Figure 6.2. In effect, the FOD operates as a lateral bipolar transistor, as described in Chapter 4. The spacing from the drain contact to the

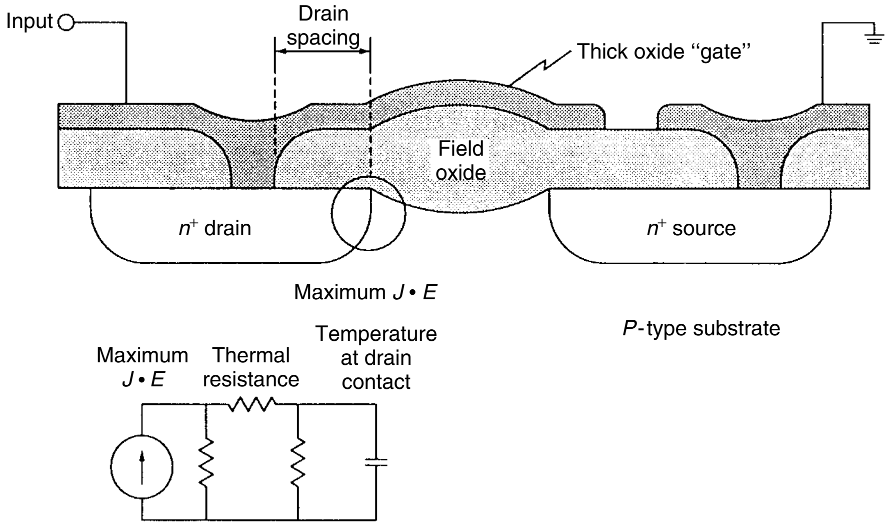


Figure 6.2 Cross section of a thick field-oxide device. The maximum heat occurs because of $J \cdot E$ at the cylindrical junction. The electrical analog is shown in the bottom of the figure

diffusion edge or *drain spacing* is a critical parameter for improved ESD. As indicated in the figure, the maximum heating during ESD occurs at the cylindrical junction where $J \cdot E$ is maximum. The FOD can be laid out in different ways for use as an ESD protection circuit. Three different cross sections of this device are shown in Figure 6.3. In Figure 6.3(a), the device is formed with no gate. As the onset of bipolar action is determined by the n^+ drain diffusion breakdown, a gate is not necessary for *npn* snapback to take place. In Figure 6.3(b), the same device is shown with a metal gate connected to the drain pad. In this case, as the voltage at the drain pad increases, the metal gate is intended to reduce the source (emitter) barrier and turn on the *npn* device at a lower level [Duvvury83]. The lowest trigger voltage can be obtained using a polysilicon gate as shown in Figure 6.3(c). Extensive studies have shown that the metal gate connection has no obvious effect on the ESD failure threshold [Wilson87][DeChiaro86]. There has not been any known data reported on a polysilicon gate structure. In general, the bipolar trigger voltage of the FOD is determined by the n^+ to n^+ spacing as shown by the experimental results in Figure 6.4 [Weston92].

The main design parameters of the FOD as far as ESD is concerned are the channel length (L), the drain contact-to-diffusion spacing toward the channel (DS), and the device width (W). These three parameters are identified in Figure 6.5. First consider the channel length dependence. It has been shown [Palella85] that the Human Body Model (HBM) failure thresholds increase as the channel length is decreased in the 7 μm to 2 μm range. In contrast, DeChiaro

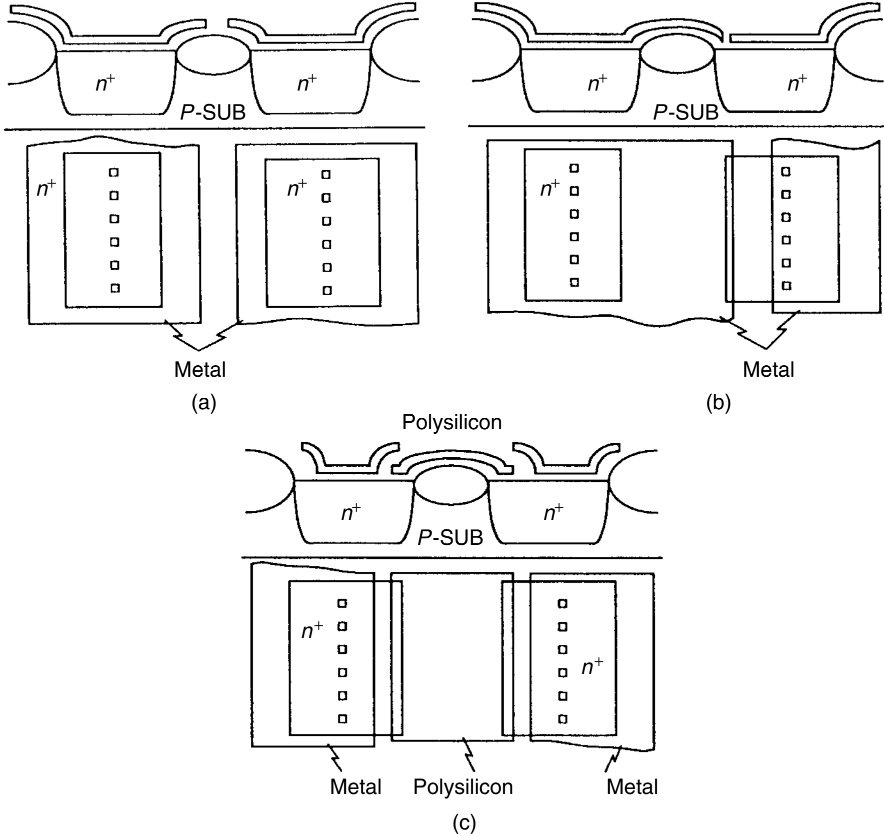


Figure 6.3 Various layout styles for the thick oxide protection device. Structure (a) has no gate; (b) has a metal gate; and (c) has poly gate

[DeChiaro86] reported that the thresholds increase with increasing channel length in the range of 1 μm to 3 μm. Work by Rountree [Rountree85] and Duvvury [Duvvury83] showed that there was little dependence on channel length between 2 μm and 8 μm. These different conclusions are probably caused by the different technologies and structures used in these studies. A detailed investigation of this phenomenon [Wilson87] concluded that all three different channel length dependencies might appear if the data is carefully collected over a wide range of dimensions. Hence, for a given technology, there might very well exist an optimum point close to, but not at, the minimum allowed channel length. As a general rule, for a 2-μm technology the optimum point could be as low as 1 μm or as high as 3 μm. It could be argued that at the minimum point punch through might dominate, leading to some reduction in the failure threshold and at longer lengths the bipolar efficiency goes down also reducing the failure thresholds.

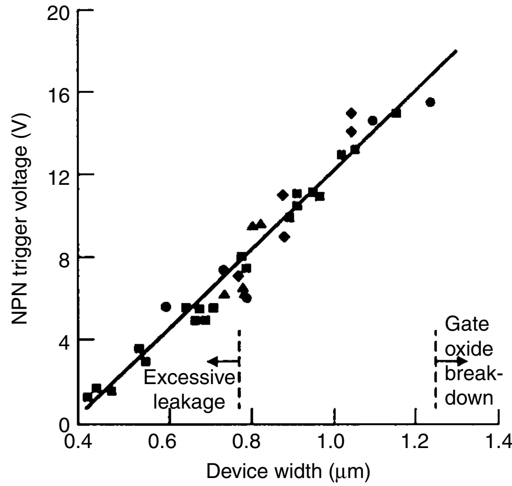


Figure 6.4 The trigger voltage of the thick oxide device as a function of field-oxide width after Weston *et al.* [Weston92]. Note that the 'width' in this figure actually represents the channel length of the device, parameter L in Figure 4.3

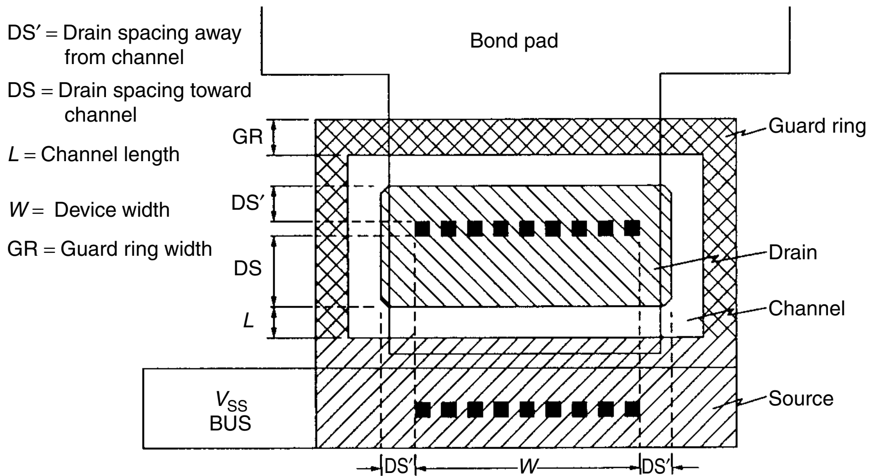


Figure 6.5 Layout for a thick field device with all the critical ESD design parameters

The second parameter to be considered is the drain contact to gate spacing. The original work by Rountree [Rountree85] showed that the impact of this parameter on the ESD failure threshold is very significant. This data is shown in Figure 6.6. As DS is increased the failure level is seen to increase to a certain level and then saturate. It was proposed that the reason for the influence of DS was the

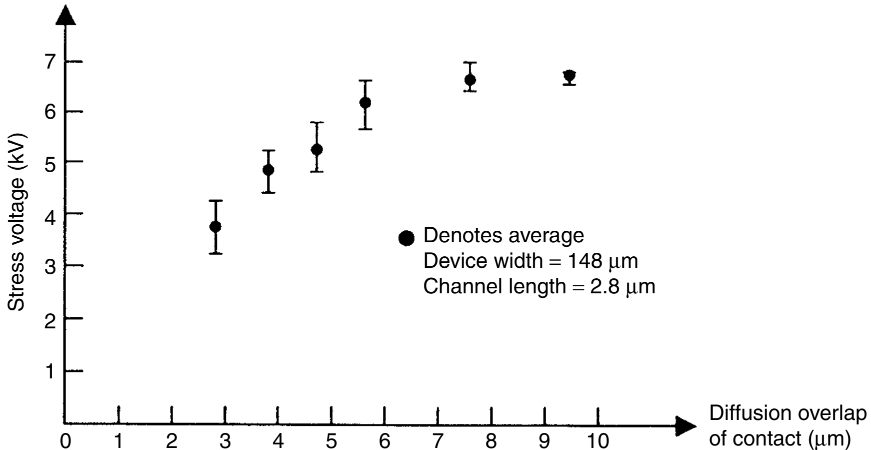


Figure 6.6 Failure voltage for the HBM ESD stress versus the drain contact to diffusion edge spacing, after Rountree *et al.* [Rountree85]

distance between the heat source and the contact. When the contact is close to the diffusion edge, the heat produced at the drain junction isotropically spreads to heat the contact metalization and results in a lower failure voltage. Moving this contact away from the drain edge to an optimum amount improves the failure level. Further increasing DS will only have an incremental effect on ESD performance once the weak failure mode is eliminated. In fact, as DS is increased even further the extra series resistance can degrade the bipolar device performance and lower the protection level. In this extreme case, instead of the avalanche breakdown taking place at the diffusion edge or at the sidewall, it occurs at the bottom wall directly beneath the contact. The bipolar does not trigger, and damage occurs as a result of spiking between the contact and the substrate. A technique where large spacing could be used for its advantage without contact spiking is to place n-well directly under the drain contacts whereby the bottom wall junction avalanche voltage is made substantially higher than the avalanche voltage at the cylindrical sidewall.

The contact spacing phenomenon has been investigated by several other workers (e.g. [McPhee86][Wilson87][Palella85]) and all reached similar conclusions for a non-silicided process. The contact spacing away from the channel (indicated as DS′ in Figure 6.3) was found to have no impact on the failure level.

Although the impact of the contact spacing had been very important for the abrupt junction processes, its effect was found to be weak for lightly doped drain (LDD) junctions [McPhee86][Duvvury85]. As a result, the failure voltage versus DS curve in Figure 6.7 shows a limiting effect at larger DS for the 1 μm LDD process. For a clad silicide process, however, the effect of the drain contact spacing virtually vanishes [McPhee86][Duvvury86][Wilson87]. In all of these works it was clearly shown that the clad silicide drain diffusion reduces the resistance of the drain contact to gate region and eliminates the ballasting that is critical for good ESD

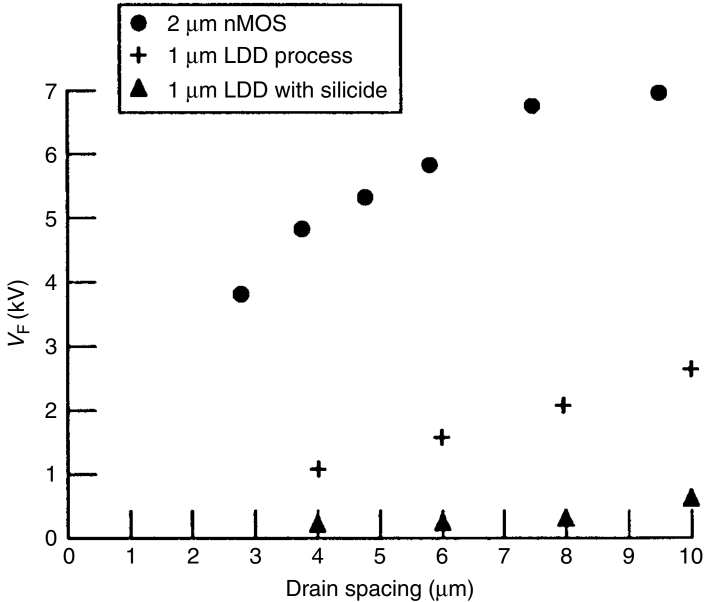


Figure 6.7 Failure voltage for the HBM ESD stress versus the drain contact to diffusion edge spacing for three different process options

levels. This typical result is illustrated in Figure 6.7. Process changes to the silicide can be used to improve ESD levels. For example, a thinner silicided layer with its relatively increased resistance can result in some improvement in ESD performance [Chen88][Duvvury89]. However, the most robust solution is to block the silicide formation in the region between the drain contact and the gate (or diffusion) edge although this increases process complexity; it is an expensive option.

6.3 nMOS TRANSISTORS (FPDs)

nMOS transistors are essentially thin oxide devices as opposed to the FOD discussed in Section 6.2. They are also called *field-plated diodes* (FPD) or *gated-diodes* in the literature because of the effect of the gate on the diode breakdown voltage. It was not commonly used as a primary protection device, in comparison to the FOD in technologies with feature sizes greater than 1 μm. The reason for this was that the FPD requires a relatively larger device size to achieve the equivalent level of protection as an FOD. However, in advanced processes with LDD junctions the FOD performance is limited, and the onset of damage has been observed at between 1 and 2 kV. Hence, nMOS devices are gaining more usage as primary ESD protection devices in advanced CMOS processes.

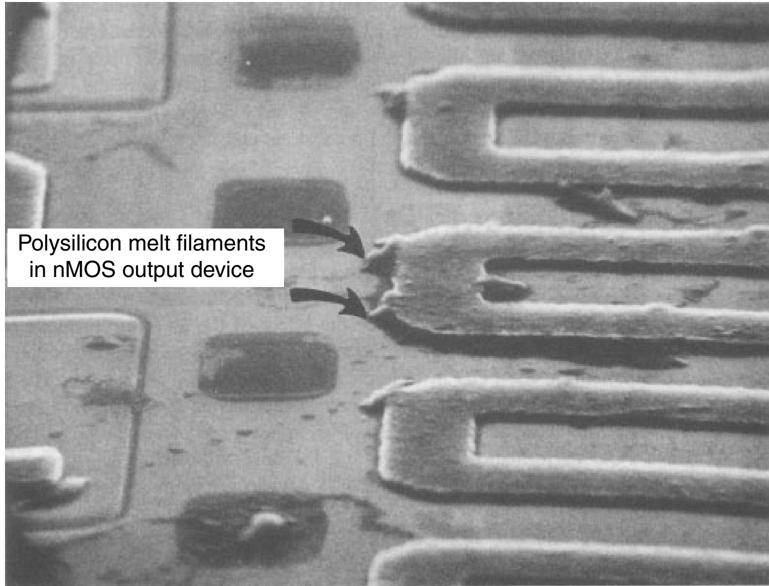


Figure 6.8 Typical ESD failure mode in an nMOS transistor showing gate to drain melt filaments between polysilicon gate and silicon surface

For larger feature size technologies the FOD is better as the bipolar action takes place deeper in the silicon and the peak heating is located further away from the silicon surface. In nMOS devices the peak heating occurs close to the surface that has a poor thermal conductivity. The result is damage to the silicon surface and the formation of melt filaments between the polysilicon gate and the silicon surface. A typical failure site is shown in Figure 6.8.

While the FOD can give as much as 40 V of ESD performance per micrometer of device width for the abrupt junction processes, the performance of the FPD per unit width is relatively lower. A comparison between the two types of devices is shown in Figure 6.9 for nonsilicided technologies. It is seen that with the introduction of LDD junctions at 1.5 μm, intended to improve the hot carrier reliability, a significant drop in the performance occurred for both the FOD and the FPD. But in the submicron range the thin oxide device actually gives a slightly better performance. Analysis has indicated that as technologies approach the 0.10-μm regime the ESD capability of the nMOS device will be better than in the 1-μm technologies [Lin93]. This is attributed to the decreased power dissipation in these devices as the avalanche breakdown voltage and snapback holding voltages are reduced in the scaled technologies. It has been shown [Amerasekera94] that this turn around in the ESD performance of the nMOS device occurs at the 0.8-μm technology node for non-silicided devices and at the 1-μm technology node for silicided devices as shown in Figure 6.10.

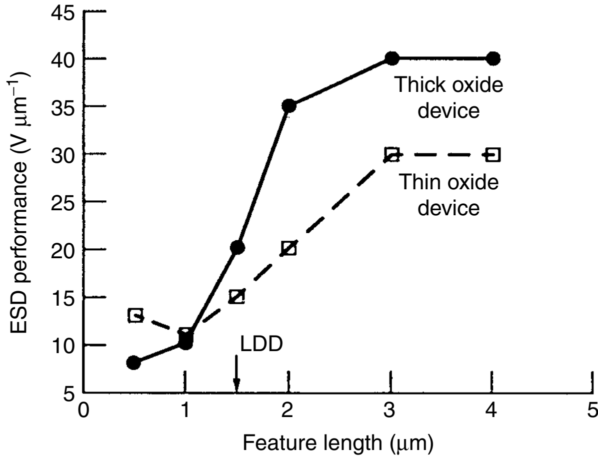


Figure 6.9 The ESD performance (in V μm^{-1}) as a function of feature size for thick oxide and thin oxide devices

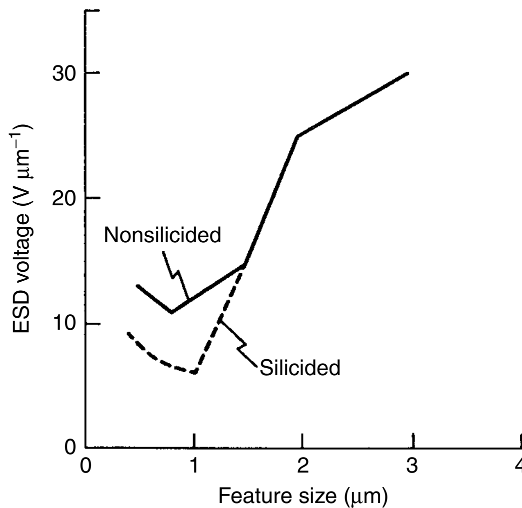


Figure 6.10 The ESD performance (in V μm^{-1}) as a function of feature size for a thin oxide is compared for both silicided and nonsilicided processes

The main design parameters of the nMOS transistor as shown in Figure 6.11 are the transistor channel length (L), the drain contact-to-gate spacing (DCG), and the device width (not shown). The source contact-to-gate spacing (SCG) does not play much of a role and is often kept at its minimum design value. As a large device width is required to obtain a given ESD level, the device is typically

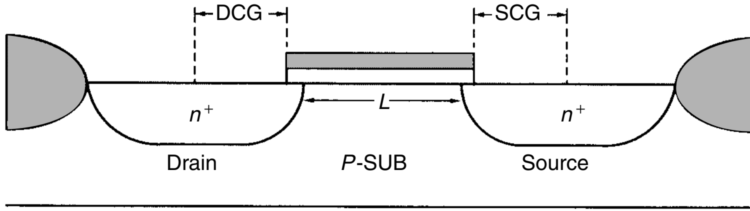


Figure 6.11 An nMOS transistor cross section showing critical ESD design parameters

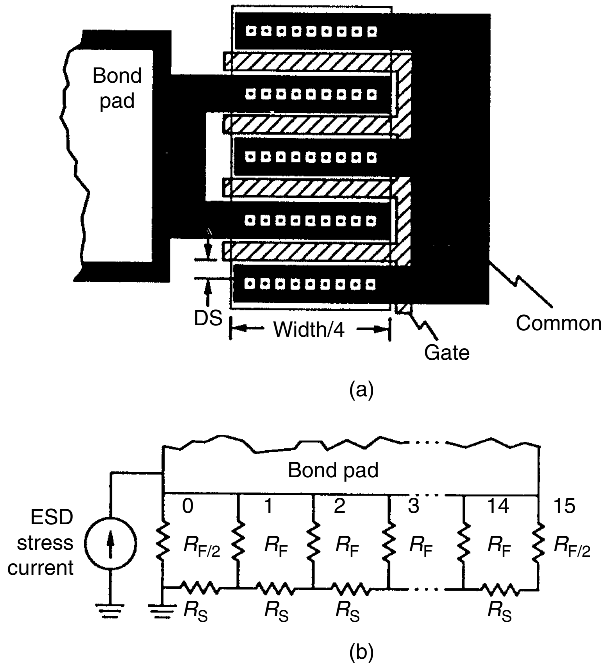


Figure 6.12 The ladder type layout for an nMOS transistor and the equivalent electrical representation of the parasitics

laid out as a finger or ladder structure, as shown in Figure 6.12. This type of layout is often used for output transistors and provides current uniformity between fingers and maximizes the ESD protection level. The parasitic resistance R_F is composed of the metal finger resistance, contact resistance, and diffusion sheet resistance. The parasitic resistance R_S is that associated with the ground bus that connects the fingers of the device. Maintaining a minimum ratio for R_S/R_F is important for obtaining the best possible ESD performance. Minimizing R_S will give more uniform current distribution, as shown in Figure 6.13 [Duvvury88B]. Note that whereas maximum R_F is desirable, the contact resistance itself should

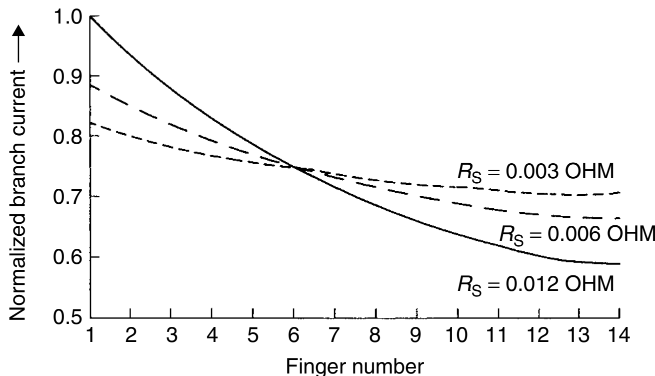


Figure 6.13 Simulated branch currents in the fingers of an nMOS transistor for different resistance values for the source bus

not be increased as damage could then result from excessive contact heating. For the protection device layout shown in Figure 6.12, the nMOS transistor has its gate tied to ground. However, multifinger structures have been found to have inconsistent performance as ESD protection devices because of nonuniformity of the finger turn-on during an ESD event [Chen88][Polgreen89]. In this respect the substrate connection is important. For floating substrate devices such as DRAMs, all the fingers turn-on effectively allowing maximum performance from the device.

The floating substrate nMOS device is essentially a floating base *npn* device. In Chapter 4 we explained how the base resistance influences the *npn* turn-on voltage. The floating-base *npn* has a much lower trigger voltage than a grounded-base device. Thus the difference between the trigger voltage and the on-voltage of the *npn* is small and it is possible to turn on more fingers before the damage threshold of the device is reached.

For logic applications the substrate is tied to ground along with the source, so the uniform turn-on for the device cannot be obtained through modulation of the base resistance. However, as described in another work [Duvvury92A], the gate can be coupled high during the ESD event, allowing MOS current conduction. As described in Chapter 4, this has the effect of lowering the *npn* turn-on voltage, which again increases the number of fingers that are effective during an ESD event. Therefore, the device gate modulation is another important design parameter for the nMOS protection device. This device is described in greater detail in Section 6.4.

The channel length of the thin oxide device does play a role in its ESD protection capability. Similar to the thick oxide device, a minimum channel length is desired for efficient turn on, but the punch through limit and the associated leakage should be avoided. This is especially important when using the output buffer device as the protection device itself. The minimum channel length device is also known to have higher susceptibility to hot carrier stress, which means that the channel length

needs to be optimized between the ESD performance requirement and hot carrier requirement.

The ESD failure threshold of a 200-μm thin oxide device as a function of the drain contact to gate spacing for different technologies is shown in Figure 6.14. As expected, the drain contact spacing has a large effect for nonsilicided processes. It is interesting to note that for the three nonsilicided technologies evaluated in this work, the optimum spacing is approximately 6 μm. For the silicided cases, there is no obvious dependence on this parameter. However, one study [Duvvury90] did find that a minimum spacing gives a better distribution. Even then, the nMOS protection device is still very inefficient in a silicided process.

In contrast to the drain side contact spacing, the source side spacing is not important for grounded substrate technologies. This is because, for a negative voltage applied to the drain, the n^+ to p-substrate diode operates in the forward-biased condition and there is no current flow or heat dissipation in the source diffusion. Keeping the source contact spacing at a minimum is the best approach for minimizing the power dissipation when in the *npn* mode, as well as for reducing the area of the protection device. For floating substrate technologies there is no forward-biased diode for negative applied voltage stress. Under these conditions the *npn* is triggered for both positive and negative stress polarities. As the stress condition

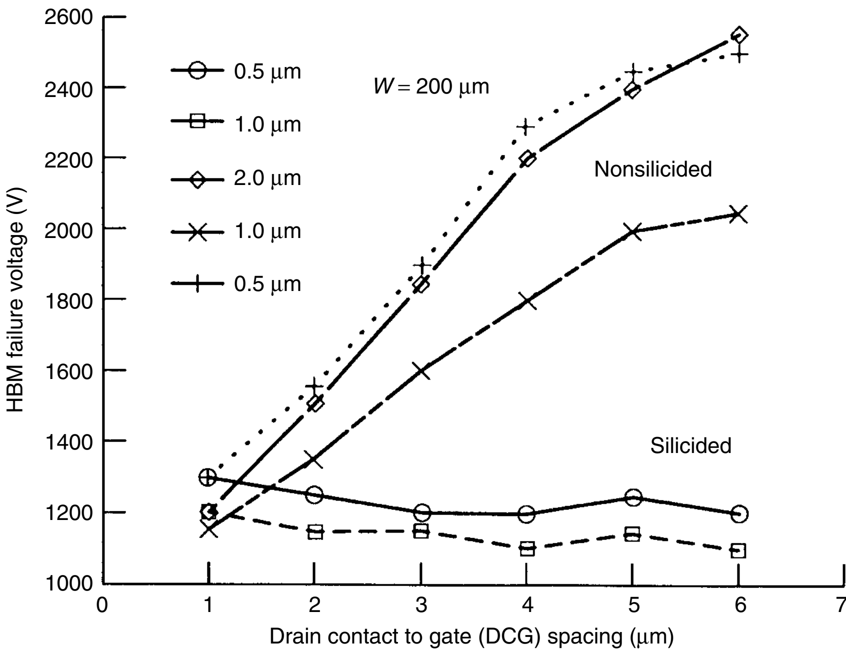


Figure 6.14 The HBM ESD performance in thin oxide transistors as a function of drain contact to gate spacing. The different silicided and nonsilicided processes are as indicated

is symmetrical, the same 6-μm spacing should be maintained at both source and drain diffusions.

In conclusion, the thin oxide device can provide good ESD protection levels. A typical design for nonsilicided or silicide-blocked devices would have a transistor width of 200 μm with close to minimum channel length. The drain contact to gate spacing for a non-silicided process should be about 6 μm for the older 2-μm technologies, to about 1–2 μm for the newer submicron technologies. The reduction in this spacing for the latter technologies is a result of the use of higher diffusion sheet resistance for improved transistor performance. For silicided technologies the drain contact spacing should be minimum unless experimental data indicates otherwise. In general, the spacing on the source should be kept minimum for all technologies. For a multifinger structure the use of more than two or three fingers will not really serve to increase the protection level unless design techniques such as gate coupling or substrate bias are used to improve this efficiency. These methods are discussed in greater detail in the following sections. In addition to the number of fingers, the individual finger length also plays a critical role. For excessively long fingers the voltage dropped within the finger will not make it uniformly robust and the heating tends to localize in the middle of the finger [Scott86]. The advanced technologies with very low resistivity substrates seem to have made this effect even worse for silicided processes [Oh01A]. As a good practice, 40–80-μm finger lengths are recommended for nonsilicided technologies and 20–40 μm for silicided technologies. However, the optimum values should be determined after evaluating I_{t2} for different finger lengths in a given technology. As a final note, the individual finger length dependence seems to be an issue for nMOS transistors only because no similar sensitivity was found for the pMOS.

6.4 GATE-COUPLED nMOS (GCNMOS)

In most applications, the thin oxide device is used as a protection device with its gate grounded. This will always ensure that the protection device, while being robust for ESD protection, will not cause any extra leakage at the pin. However, the thin oxide device can be a more robust protection element if its gate is coupled high during an ESD event. The effect of gate bias on the ESD performance of nMOS devices has been reported [Chen88][Polgreen89][Abderhalden91]. As illustrated in Figure 6.15, if the gate voltage is about 1 V, the nMOS trigger is lowered to less than the onset for avalanche breakdown ($V_{t1'} < V_{t1}$) and is therefore ideal for improved ESD protection. It was also noted that if the gate voltage goes above 5 V, the I_{t2} value (defined in Figure 6.15) decreases to give reduced failure threshold voltage ($I_{t2'} < I_{t2}$). Hence, a gate voltage of between 1 and 2 V is typically needed for best ESD performance.

In a multifinger structure, the gate coupling improves the uniform turn-on of all the fingers. This is not always possible in a grounded gate device as the first *npn* device to turn on discharges all the ESD current potentially preventing the

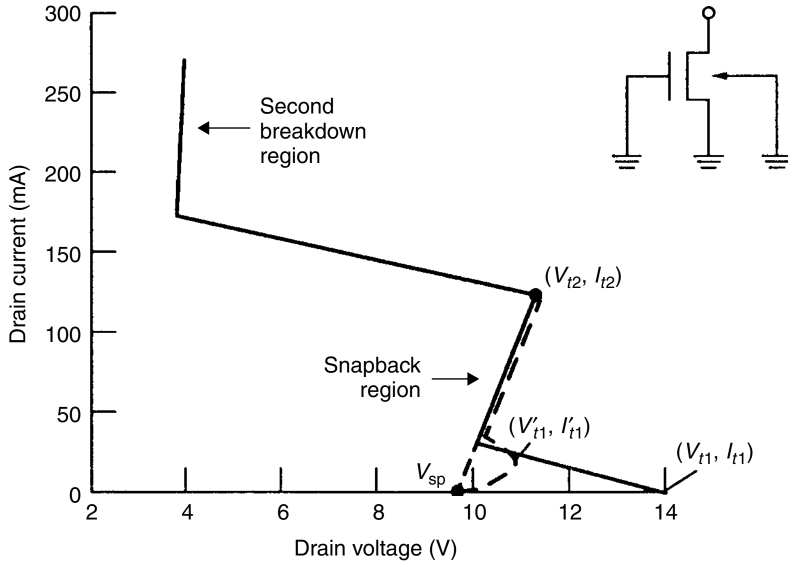


Figure 6.15 The high current $I-V$ curve for an nMOS transistor. The solid line is for the case with the gate grounded and the dashed line is for the case with the gate coupled above V_t

other fingers from turning on. In such a device, each of the other fingers has a chance to turn on as the voltage increases again toward V_{t2}. That is, after one finger begins *npn* conduction and clamps at the snapback voltage, the pad voltage builds up again because of the snapback resistance. When the pad voltage again reaches V_{t1} the next finger turns on, and so on until all the fingers are turned on or the failure current I_{t2} is reached, whichever comes first. Usually I_{t2} (or more accurately V_{t2}) is reached first, and the total number of fingers that actually turn-on varies substantially as shown by the failure distribution in Figure 6.16. Now the effect of gate coupling on the transistor threshold voltage is to lower the avalanche breakdown voltage, V_{av}, as shown in Figure 6.17. This is shown here for four different technologies. Note that in all cases the minimum in V_{av} (same as V_{t1} in Figure 6.15) is reached when the gate is between 1 and 2 V.

There are different techniques to achieve gate coupling during an ESD event. In the structure shown in Figure 6.18, the coupling on the gate is determined by the ratio of the gate–drain overlap capacitance to the thin oxide capacitance [Duvvury92A]. Typically, after the *npn* turns on and clamps the voltage at approximately 8 V, the pad voltage can go as high as 15 V for ESD currents of around 2 A as a result of the bipolar device snapback resistance (about 5 Ω). This can turn on the field oxide device and discharge the gate potential to zero level. Figure 6.19 shows circuit level simulations of these effects using SPICE. The time constant for the gate discharge will depend on the ESD current level and

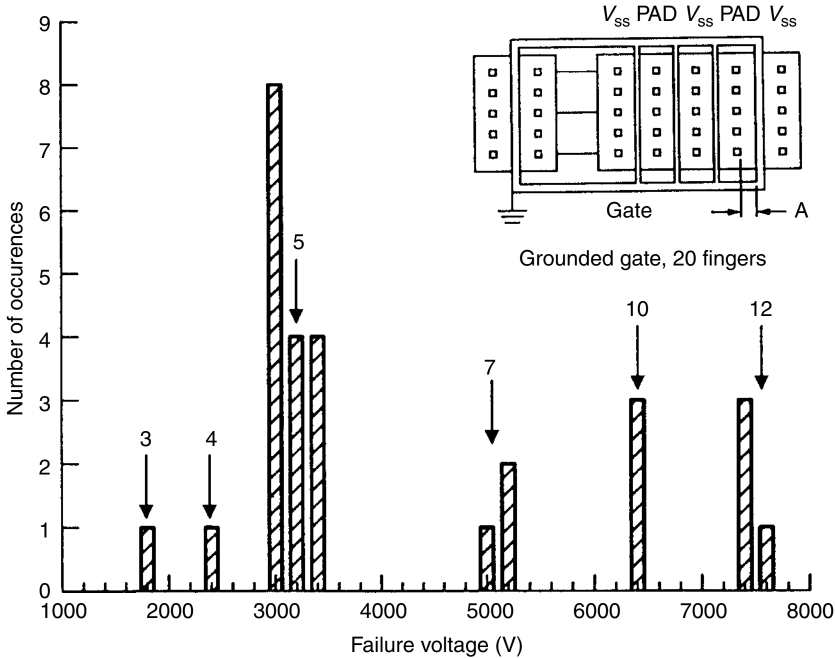


Figure 6.16 The ESD failure distribution for a grounded gate nMOS transistor. The numbers indicated are the number of fingers assumed to be turned on during ESD

the size of the field oxide device width. Usually it is designed so that the gate stays on for a minimum duration of between 5 and 10 ns, corresponding to the rise time of the ESD event. This will allow enough time for all the fingers in the nMOS to turn on. The $I-V$ characteristics of a grounded gate device are compared to the gate-coupled device in Figure 6.20. The arrows in the inset indicate where each nMOS finger goes into *npn* snapback. In contrast, the second breakdown point in the grounded gate devices is reached very soon after turn-on of only two fingers. The ESD failure distribution of the gate-coupled device is shown in Figure 6.21 for a 1-μm silicided technology. The failure threshold voltages show a tight distribution and scale as the device width is increased. The ESD performance for this device in a non-silicided technology is shown in Figure 6.22. Again the width dependence is seen and excellent ESD levels are obtained. Therefore, the gate-coupled device is a robust ESD protection circuit element.

The practical design of the GCNMOS requires that the gate has a good connection to ground during normal circuit operation. The field oxide device at the gate of the nMOS by itself will not bring the gate to ground during normal operation, and unacceptable leakage will result if the gate remains floating. A more robust approach is to connect the gate to ground through a large resistor (approximately

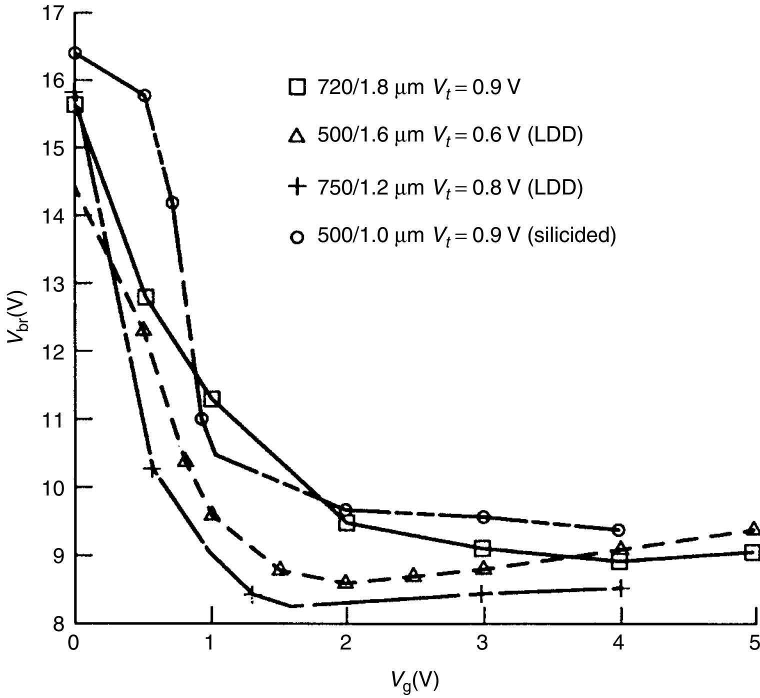


Figure 6.17 The avalanche breakdown voltage as a function of gate bias in an nMOS transistor for various technologies

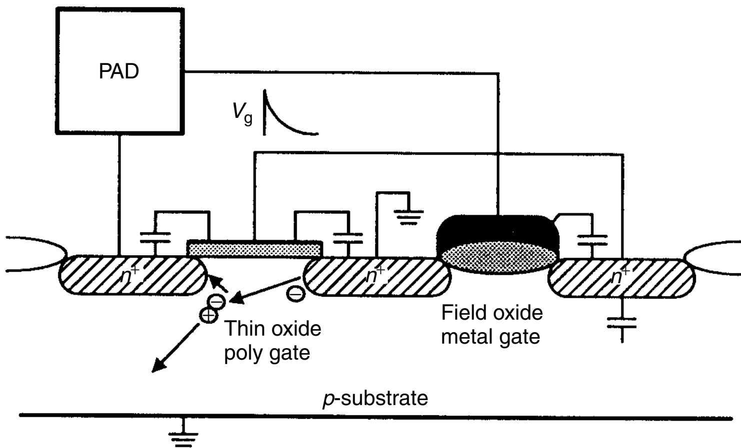


Figure 6.18 Cross section of a gate-coupled nMOS (GCNMOS) where its gate is connected to ground through a thick oxide device. The coupling ratio is shown in the equation in the box

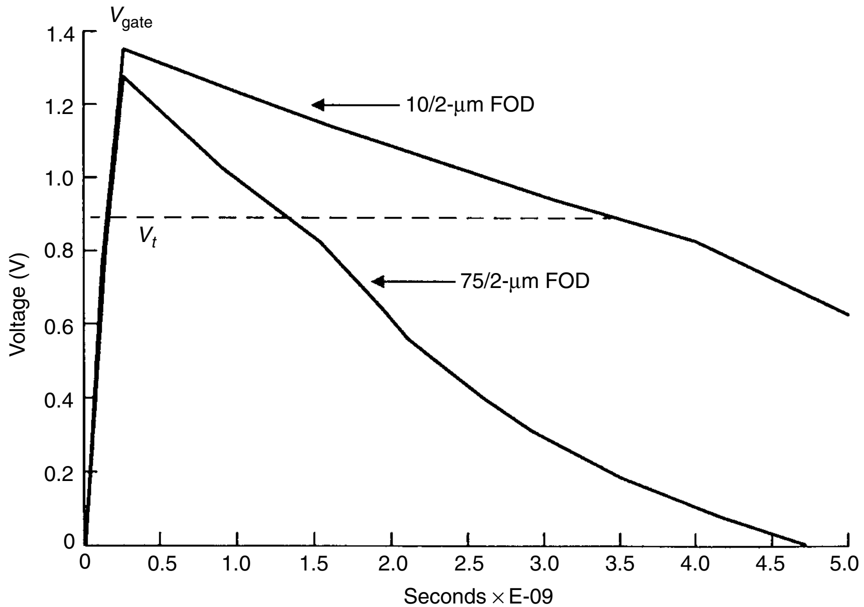


Figure 6.19 Simulated gate transient waveforms for a GCNMOS. The size of the thick field oxide (FOD) has an influence on the gate decay as shown here

10–15 kΩ) so that the gate voltage is discharged by the time constant determined by this resistance and the total capacitance at the gate. Other techniques using external circuitry to raise the gate voltage during an ESD event have also been reported (see Section 6.5).

6.4.1 Gate-Coupled nMOS (GCNMOS) Design

Optimizing the gate-coupling device also takes certain amount of test structure evaluation. This is shown in Figure 6.23 [Chen97]. The reduction in V_{t1} beyond the transistor threshold of approximately 1 V is apparent and in agreement with the data in Figure 6.17. There is a slight increase in V_{t1} at higher gate bias and this is attributed to the reduction in the substrate current after the peak that occurs around 3 V. For efficient multifinger turn on of the nMOS protection device the V_{t1} value needs to be less than the V_{t2} value [Polgreen89]. From the plot of both V_{t1} and V_{t2} versus gate bias in Figure 6.23 it is seen that the optimum gate bias is greater than 1 V. However, the I_{t2} has a roll-off with gate bias, which also needs to be considered. Typically, depending on the technology, in epi or bulk substrates gate bias in excess of 3–5 V will degrade I_{t2}, possibly because of channel heating effects [Oh01B], as shown in Figure 6.24. This phenomenon has not yet been studied in detail. It should be noted that, however, for silicon-on-insulator

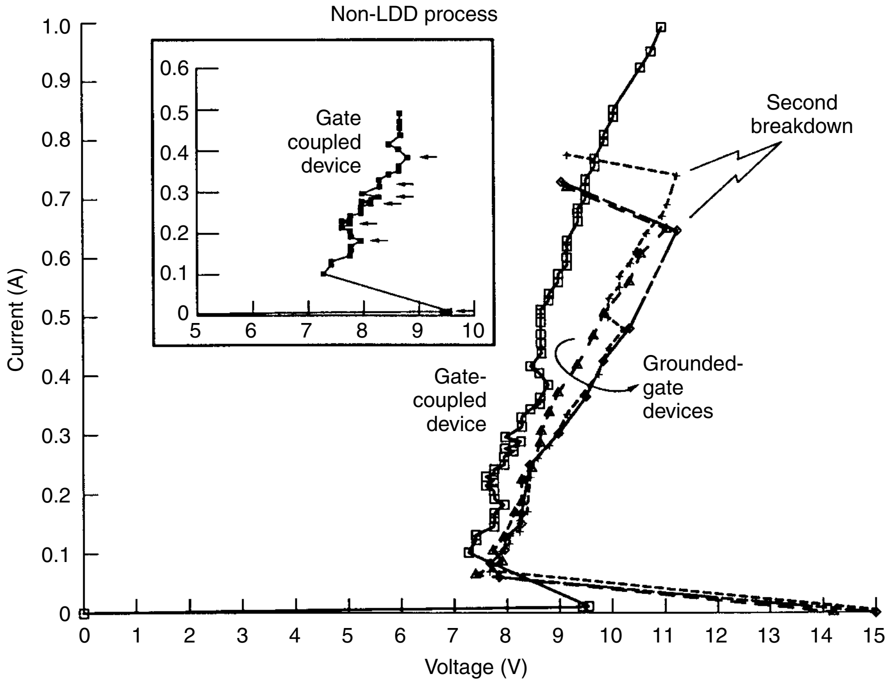


Figure 6.20 The measured $I-V$ curves for a gate-coupled nMOS compared to a grounded gate nMOS. The inset shows the details of individual finger turn on for the gate-coupled device

(SOI) this effect has been observed to be relatively more severe [Duvvury96]. Referring back to Figure 6.23, the optimum gate bias is between 1 and 3 V for this case.

The concept that was shown in Figure 6.18 is not practical for applications as it could lead to leakage with the gate floating. This can be overcome with an RC network as shown in Figure 6.25. Although the gate–drain overlap can give capacitive coupling with the ESD transient, it may not be sufficient and/or would vary with process fluctuations. Therefore, an nMOS capacitor (Cc) can be used with its gate tied to the PAD and the source/drain connected to the gate of the GCNMOS. The resistor could be an n-well resistor, typically in the 10–15 kΩ range. A typical design would have 500 μm wide nMOS with the gate booting capacitor of 20 μm^{-2} and a resistor of 10 k Ω to give a gate transient peak of 2 V and an on-time (defined as above the transistor threshold) of 10 ns. The required on-time generally depends on the technology and has to be determined with test structure characterization.

The GCNMOS has been reported to be robust for nonsilicide or silicide-blocked technologies but it can become erratic with silicide diffusion processes and even completely ineffective if the substrate is of low resistance, as in an epi process.

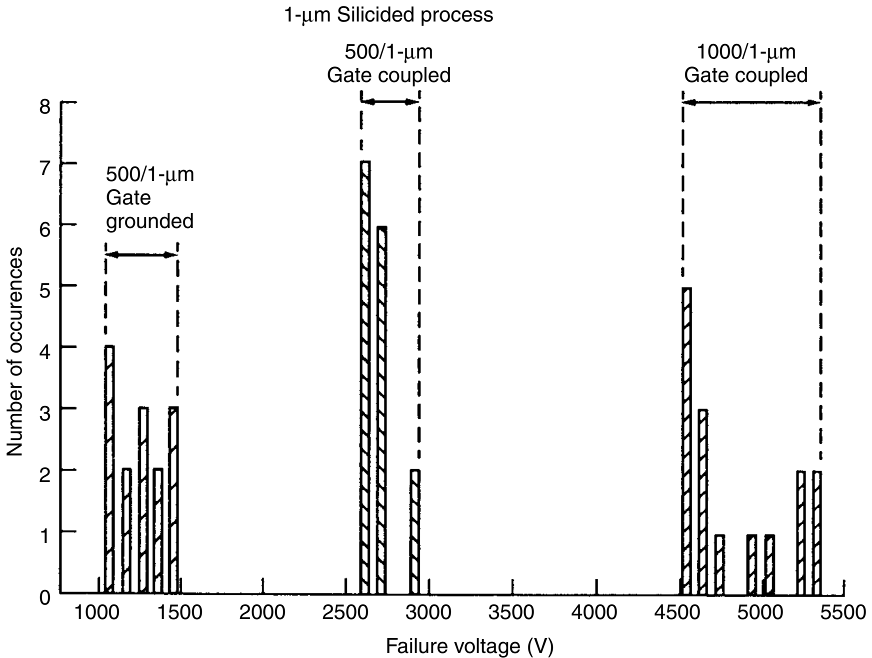


Figure 6.21 The ESD failure distribution for a gate-coupled nMOS transistor for a 1-μm silicided process. The grounded gate case is also shown

Even with gate coupling a certain amount of ballasting is necessary, as reported for a nonsilicide technology [Duvvury92A]. In this case the ballasting was achieved by keeping the drain contact to gate spacing at greater than approximately 3 μm (see Figure 6.26). For a silicided process as this technique cannot be present the GCNMOS may not work effectively. On the other hand, when the substrate resistance is very low the substrate current generated by gate coupling may not be sufficient to keep the multifinger device in bipolar turn-on. This can be overcome by simultaneously applying substrate bias; this effect will be discussed later.

6.4.2 GCNMOS I/O Applications

The GCNMOS can have applications for input, output, and input/output pins. For inputs it can be directly connected to the buffer gates. However, for maximum efficiency and good protection for CDM, the recommended approach is to isolate the protection device with a secondary clamp, as shown in Figure 6.27.

The protection device scheme shown in Figure 6.27 for inputs is effective for HBM as well as CDM requirements. The GCNMOS 2 would be needed close to the input gate to protect effectively against the local transients and for this

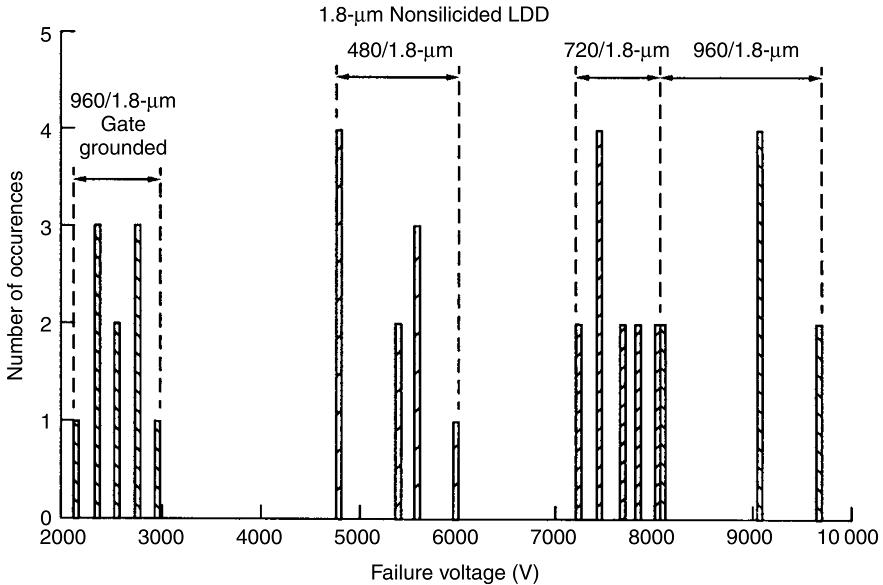


Figure 6.22 The ESD failure distribution for a gate-coupled nMOS transistor for a 1.8-μm nonsilicided process. The grounded gate case is also shown

reason should be connected to the same ground as the input buffer [Maloney88]. The GCNMOS 2 could be about 1/5th the size of GCNMOS 1, with the booting capacitor and gate resistor at the same design sizes. For protecting the pMOS gate a secondary CDM clamp to V_{dd} can be used in case a diode to V_{dd} is not allowed. The two cascoded nMOS devices can be the same as the GCNMOS 2; the isolation resistor is typically about 100 Ω.

In Figure 6.28, the GCNMOS protection application for outputs is shown. The isolation resistor becomes necessary if the output nMOS device width is too small (less than 50 μm). Generally, a resistor is not required if the output nMOS is also designed with the same contact to drain spacing as the GCNMOS 1, but in this case SPICE simulations should be done to ensure the gate coupling on the output device matches with the gate coupling on the GCNMOS. The simulation simply requires that a transient pulse of magnitude V_{t1} with a rise time of 1 ns be applied at the PAD, noting the gate coupling on the GCNMOS and output nMOS gate.

Finally, the GCNMOS concept can also be applied for V_{dd} protection. However, it must be noted that owing to the large capacitance associated with the V_{dd} pin the transient voltage at the pad slows down and discharges the gate to be below the transistor V_t in a short time. In other words, the gate potential could be close to 0 V when the drain potential reaches V_{t1}, making the protection device trigger nonuniform. This can be simply overcome by making the booting capacitor Cc about 10 times larger.

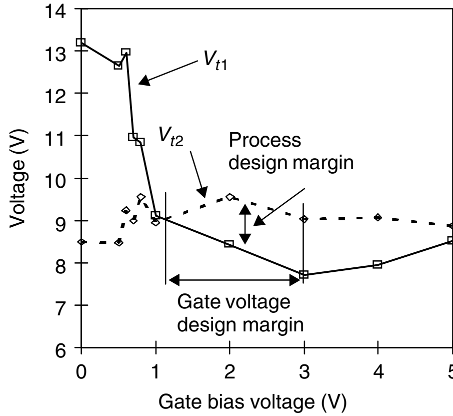


Figure 6.23 The V_{t1} and V_{t2} points of Figure 6.15 are a function of gate bias; for optimum protection design the window shown must be maintained [Chen97]. Reproduced by permission of ESD Association

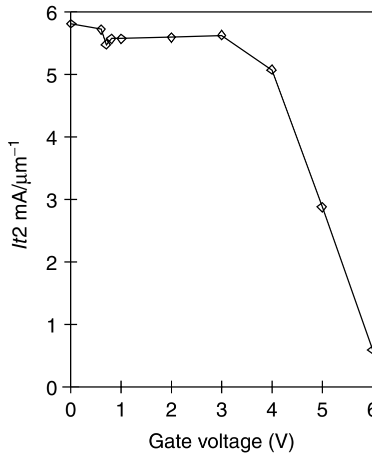


Figure 6.24 The I_{t2} roll-off with gate bias in a 0.6-μm CMOS technology. Note that beyond 4 V the I_{t2} readily degrades, possibly because of channel heating [Chen97]. Reproduced by permission of ESD Association

In all cases of GCNMOS protection applications it is advisable to include the device with its gate resistor and capacitor in the actual circuit performance simulations to note any compatibility issues. For example, in an input application the 0 to V_{dd} transient can momentarily trigger the GCNMOS. The adjustment of the R and C elements can ensure that the gate stays below transistor V_t for 0 to V_{dd} transients but goes above V_t for 0 to V_{t1} transients. Thus the design requires careful optimization.

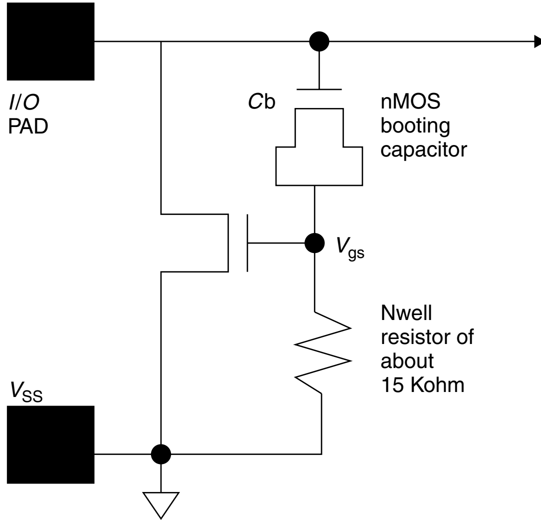


Figure 6.25 Gate-coupled protection nMOS design with R and C elements. The resistor Rg is typically built with n-well and the capacitor is simply an nMOSFET with the gate to PAD and source/drain to the protection nMOS gate

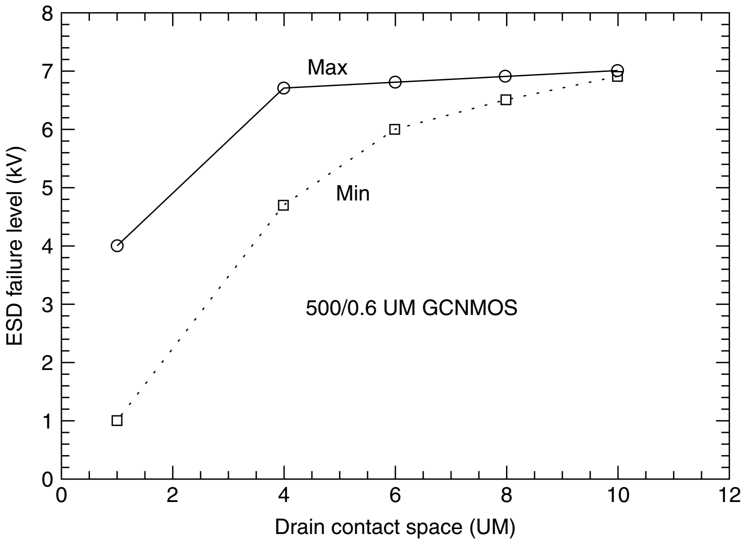


Figure 6.26 The multifinger turn on of a GCNMOS protection device as a function of contact to gate spacing on the drain side for a nonsilicided technology. Note that more uniform trigger resulting in tighter distributions of the failure threshold is seen when this spacing is greater than 4 μm. (After [Duvvury92B], reproduced by permission of ©1992 IEEE)

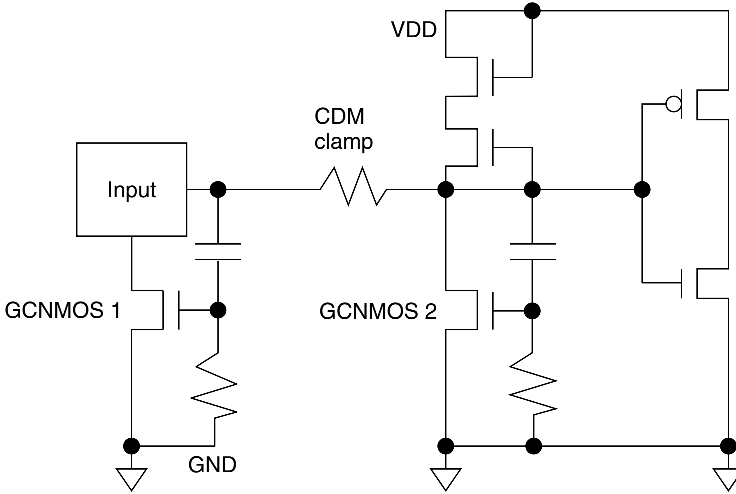


Figure 6.27 Complete input protection with GCNMOS. The primary protection GCNMOS 1 is isolated from the secondary GCNMOS 2, which could be about one-fifth the size of the primary device. For best protection of the nMOS gate, the ground of the GCNMOS 2 should be connected to the same ground as the input buffer ground. To protect the pMOS gate a secondary CDM clamp should also be included as shown

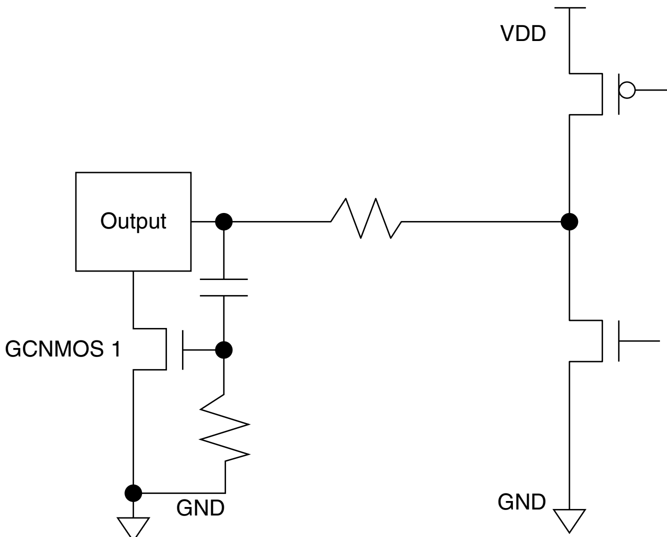


Figure 6.28 GCNMOS protection for output pins. The isolation resistor can be chosen such that the output performance is not degraded

6.5 GATE DRIVEN nMOS (GDNMOS)

For ESD protection using the silicided nMOS transistor in snapback, full triggering requires modulating the gate above threshold during at least the initial phase of the ESD event. Although Section 6.4 demonstrated transient methods for achieving this behavior, steady state biasing techniques can provide the same function. The bias circuit must also set $V_{gs} = 0\,\mathrm{V}$ during normal circuit operation to avoid leakage on the pad. Therefore, the bias circuit must distinguish between two modes: ESD, where V_{gs} should be greater than threshold and less than the I_{t2} roll-off voltage, and normal operation, where $V_{gs} = 0\,\mathrm{V}$.

A number of different gate modulation methods have been described [Anderson97]. Power-supply referenced gate modulation uses the voltage on V_{dd} to distinguish between ESD and normal operation [Krakauer94]. An example is shown in Figure 6.29a. During normal operation, the voltage difference between V_{dd} and V_{ss} causes this circuit to set the bias on the ESD device MESD to V_{ss}. During an ESD event to the I/O pad, V_{dd} floats. The decoupling capacitance between V_{dd} and V_{ss} keeps V_{dd} near V_{ss}. The gate modulation circuit causes the gate of the ESD device MESD to track the I/O pad. If desired, elements can be added to this circuit to reduce the gate bias during ESD [Krakauer94]. The desired behavior of V_{dd} will only occur if several strict precautions are followed. There can be no current paths from the I/O pad to V_{dd}. The decoupling capacitance between V_{dd} and V_{ss} must be much larger than the capacitance between the I/O pad and V_{dd}. The V_{dd} to V_{ss} capacitance must also be large enough to absorb any parasitic transients from the ESD tester [Anderson98B].

Figure 6.29b shows a second example of a gate-driven nMOS circuit [Richier97]. It uses zener diode breakdown to distinguish between ESD and normal operation. During normal operation, the pad stays below the zener breakdown voltage and the gate of the ESD device remains at ground. During ESD, the pad rises above the zener breakdown voltage causing the zener to deliver current to the resistor, raising

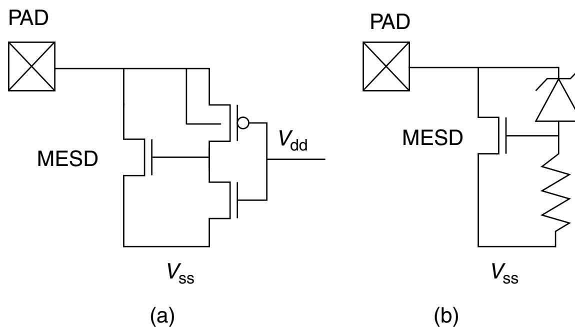


Figure 6.29 Different methods for nMOS gate modulation: (a) V_{dd}-referenced gate modulation and (b) zener-coupled gate modulation (After [Krakauer94][Richier97])

the gate voltage of MESD. This protection method requires a zener breakdown voltage above the high level of I/O signaling but below the nMOS snapback holding voltage V_h. The zener device is not common to many CMOS processes.

6.6 SCR PROTECTION DEVICE

The SCR is the most efficient of all protection devices in terms of ESD performance per unit area. The basic SCR is a *pnpn* device, as shown in Figure 6.30. The device shown in this figure is also referred to as a lateral SCR or LSCR.

The operation of the SCR has been described in Chapter 4. We will briefly repeat the main points here. The adjacent n^+ and p^+ diffusions in the n-well are connected to the input terminal. A vertical *pnp* device is formed with the p-substrate as the collector, n-well as the base, and input p^+ diffusion as the emitter. The n^+ diffusion in the p-well is connected to the ground or substrate bus and forms the emitter of the *npn* transistor. The base of the *npn* is formed by the p-substrate and the collector is the n-well and the n-well contact. During normal circuit operation, CMOS latchup should not be a problem as the emitter and base of the *pnp* are at the same potential. During an ESD stress pulse the collector–base junction of the *npn* goes into avalanche breakdown generating the electron current in the n-well which forward biases the emitter–base junction of the *pnp*. The turn on of the *pnp* occurs in less than 1 ns and this leads to the regenerative *pnpn* action [Rountree88]. Once the SCR is turned on the device is in a low impedance state and the anode to cathode clamping voltage is of the order of 1–2 V in a submicron process. This dramatically reduces the power dissipation and results in an improved ESD performance. The nature of the device operation means that it is not strongly influenced by salicidation, which is a big advantage in advanced CMOS processes. The performance of the device

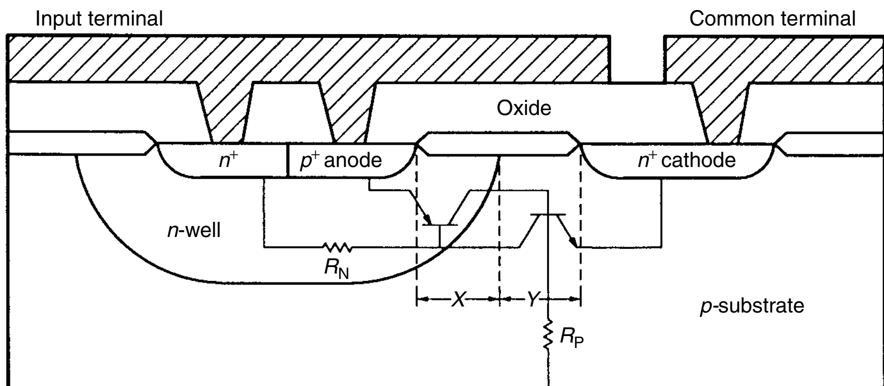


Figure 6.30 Cross section of a lateral SCR device

is higher (60–70 V μm^{-1}) in non-silicided processes than in silicided processes (40–50 V μm^{-1}), but the performance is so high that the difference is not that important. Other key process controlled parameters for successful operation of this device are the holding voltage controlled by the n-well overlap of the p^+ anode (X) and the trigger voltage determined by the p-substrate resistance, R_p in Figure 6.30. Therefore, a thicker epitaxial layer is desirable for better ESD performance. However, a thinner epitaxial layer is required for reducing the CMOS latchup sensitivity in advanced VLSI chips. In such cases the *npn* may not be able to trigger properly because the p-substrate (or base) resistance is so low. Thus there is the need to optimize the choice of epitaxial thickness to trade-off between latchup and ESD performance.

The LSCR trigger level is generally quite high and can vary between 40 and 100 V, depending on the process and design. In an advanced CMOS process the trigger voltage is defined by the n-well to substrate breakdown voltage and is about 50 V. The process parameters that influence the trigger voltage are the n-well and substrate doping levels. The main design parameter influencing the trigger voltage are spacings between the anode and the n-well edge. The trigger level can be lowered by decreasing the critical spacing for the anode (X) and cathode (Y) but this can lead to increased leakage. Figure 6.31 shows the trigger voltage plotted as a function of the SCR spacings. It should be noted that below 3 μm the SCR trigger voltage drops sharply whereas the leakage current increases. To reduce the trigger voltage without significantly impacting the leakage current, the design is modified to include a highly doped region near the surface at the n-well edge. The cross section of the modified LSCR (MLSCR) is shown in Figure 6.32 (inset) along with its $I-V$ breakdown characteristics. Note that for the 2-μm technology the SCR trigger is at approximately 25 V. This trigger voltage can be further reduced to the 12–15 V range by replacing the field oxide of the MLSCR with the thin oxide as shown in Figure 6.33 [Chatterjee91A]. Because of its low trigger level, this last device

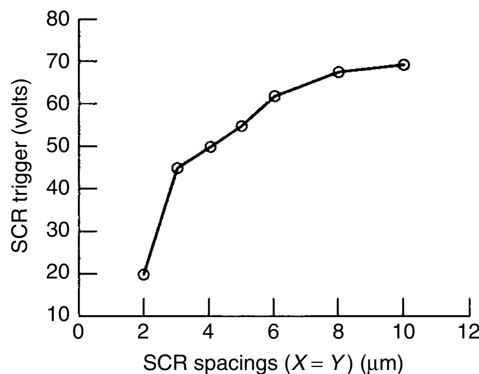


Figure 6.31 The SCR trigger voltage as a function of SCR spacing ($X = Y$)

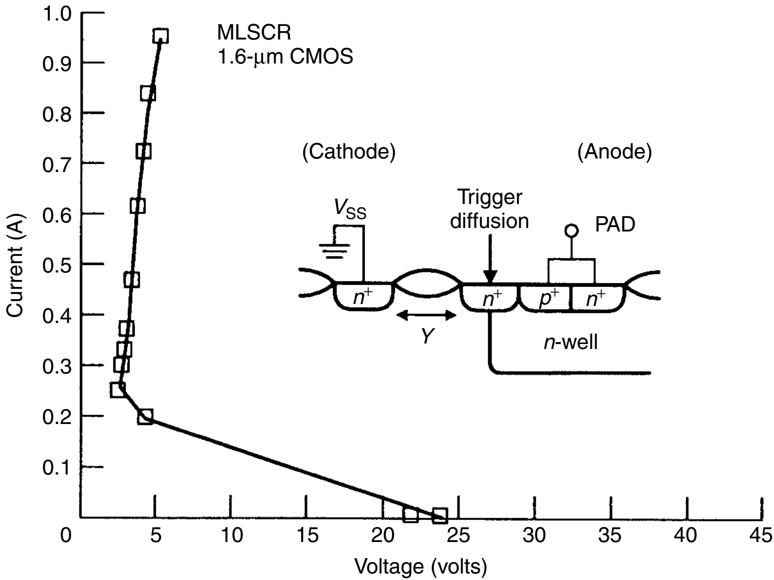


Figure 6.32 $I-V$ characteristics for MLSCR with a 1.6-μm CMOS process

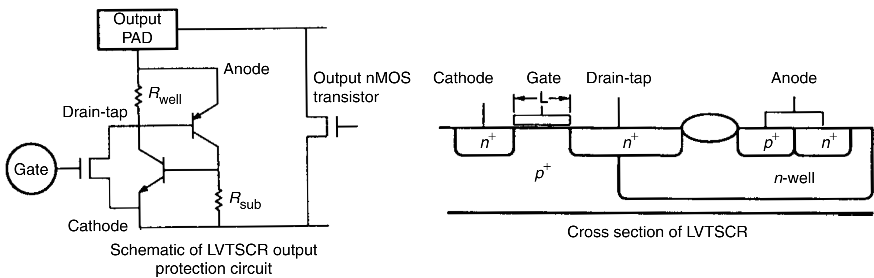


Figure 6.33 Cross section of a low-voltage trigger SCR (LVTSCR). The circuit equivalent is shown on the left

is called a *low-voltage trigger SCR* (LVTSCR). The LVTSCR essentially uses an MOS device in parallel with the SCR. Triggering occurs after the drain junction of the MOS transistor begins avalanching. The avalanche-generated hole current in the *p*-substrate turns on the lateral *npn* and then the vertical *pnp*, followed by eventual regenerative SCR action. The low trigger voltage of the LVTSCR means that it can be used as an ESD protection device for CMOS output buffers. However, one still needs to ensure that the output device does not trigger before the SCR. This can be solved by either making the channel length of the MOS device in the LVTSCR shorter than the output device or by placing an isolation resistor

between the output device and the SCR protection device. The latter option has been successfully implemented in large submicron circuits [Carbajal92], but requires that the performance degradation of the output buffers caused by the addition of the resistor is compensated in the circuit design. An alternative is to reduce the trigger voltage of the SCR further by techniques similar to those used for the GCNMOS. By raising the gate voltage of the LVTSCR the voltage at which the SCR turns on can be significantly reduced [Diaz94]. This technique requires careful tuning to ensure that the trigger circuit does provide the correct triggering for the SCR. For the advanced submicron technologies the MLSCR trigger voltage is around 10–12 V and the LVTSCR trigger voltage is about 8–9 V. However, the MLSCR may not be reliable unless design optimization is properly done. It has some potential future applications in future 5-V tolerant designs as will be discussed in Chapter 7.

The SCR is severely disadvantaged when used in floating substrate technologies. In this case the forward-biased diode is absent when stressing the pad negative with respect to ground. Thus, whereas in the forward direction the protection is provided by the SCR action, the *npn* needs to provide the protection in the reverse direction. To overcome this effect dual SCRs have been proposed which essentially have SCRs connected in parallel providing protection for both positive and negative polarity ESD stress voltages [Ker92].

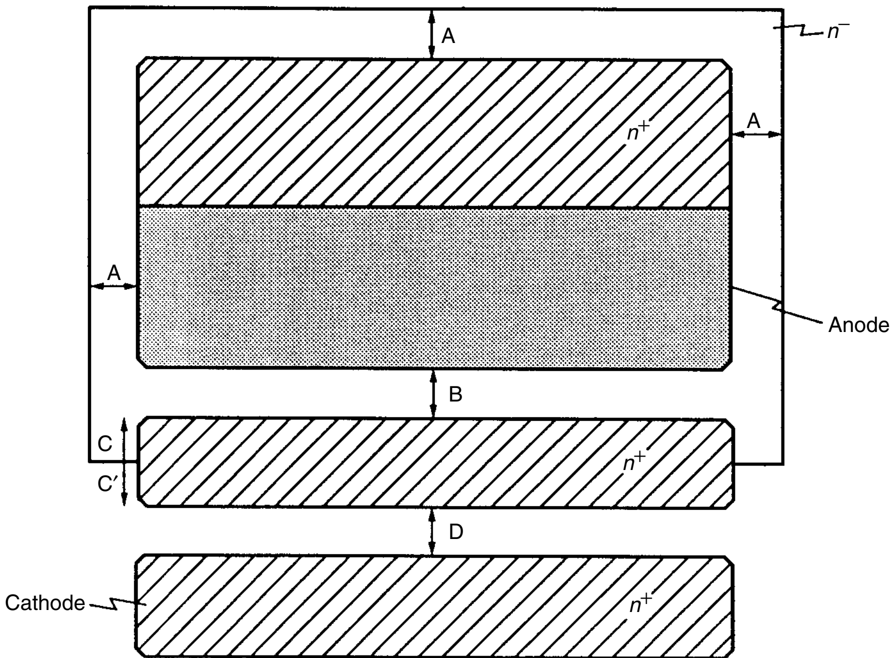


Figure 6.34 Critical layout parameters for an MLSCR

The design and layout of the SCR devices follow a slightly different set of rules from the thin or thick oxide devices. First of all, there is no contact to drain spacing issue and usually all spacings should be at their minimum allowable values. The typical MLSCR layout and the critical layout spacings are noted in Figure 6.34. Spacing A determines the leakage. Spacing B is not very critical and can be collapsed as long as the process is nonsilicided. For a silicided process, B is kept at a minimum to improve the gain of the lateral devices. Spacings C and C′ are process defined parameters and should be kept at a minimum. Spacing D will control the trigger as it is the channel length of the MOS device. For best trigger D is also kept at a minimum. Typical values of these spacings for a 1-μm technology are 4 μm for A, 1 μm for B, 0.5 μm for both C and C′, and 1.6 μm for D. The MLSCR gate can be either metal or polysilicon. These similar spacings also apply to the LVTSCR, except the channel length can be reduced to 1 μm. For the advanced submicron technologies the SCR spacings should be carefully selected as described in the Sematech Document [Voldman99]. Similar to the total protection using the thick oxide device, the SCR protection layout is shown in Figure 6.35 for a nonsilicided process and in Figure 6.36 for a silicided process. Note that the series resistor needs to be much longer for

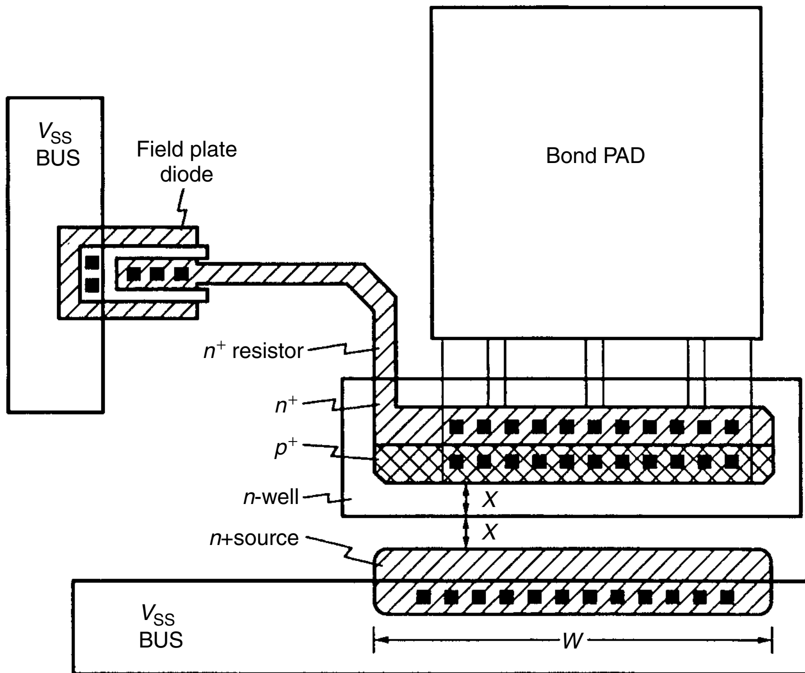


Figure 6.35 Layout example for a input protection scheme using LSCR in a nonsilicided technology

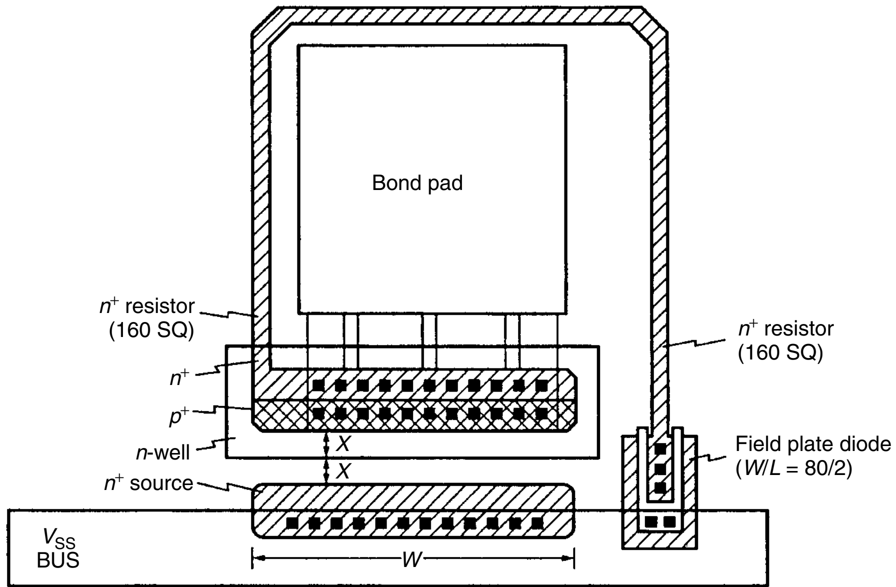


Figure 6.36 Layout example for a input protection scheme using LSCR in a silicided technology

the silicided case because of the low sheet resistivity of the silicided diffusion. A different layout style for the nonsilicided protection scheme is described in Section 6.7.

In this chapter several different protection structures have been described, with design and layout details given for each. However, their most effective protection performance is only determined when they act in a composite scheme as the primary devices. To achieve this, a design synthesis is needed. With this approach the most efficient composite protection scheme not only offers the best possible protection but also minimizes the input transit delay.

6.7 ESD PROTECTION DESIGN SYNTHESIS

In this section a synthesis of an input protection scheme for applications in CMOS technologies is presented. The overall effectiveness of any input protection scheme is determined by the design of the constituent elements of the primary and secondary protection circuits. The SCR, as a primary protection device, was first introduced for bipolar technologies by Avery [Avery83]. For CMOS technologies the LSCR has been shown to be robust, as discussed in Section 6.6. However, the design of the secondary protection to work in conjunction with this LSCR is not straightforward and can lead to failures [Rountree88][Duvvury89]. This is because

the typical SCR trigger level tends to be high (from 12 to greater than 50 V) and the design issues involved for the secondary protection need a clear understanding for effective protection designs using this device. In this section, the details of the secondary protection design are discussed. Moreover, an effective design with a combination of polysilicon resistor and SCR will also be discussed. This latter scheme can be an attractive option in analog circuit applications.

The approach taken here will be to analyze the individual protection elements of an overall input scheme, investigate how these work in conjunction with each other, and establish the ideal total combination. In all cases, pulse testing data is presented to understand the device response under ESD conditions. Failure analysis will be used wherever applicable to demonstrate the protection circuit functions.

We will briefly review the basic input protection scheme first. As seen earlier for the older MOS technologies, an input protection design consists of an FOD operating as a lateral *npn*. This device is combined with a grounded gate nMOS transistor (also known as an FPD) through an isolation resistor to form the total input protection. The scheme is shown in Figure 6.37. During the initial ESD pulse, the pad voltage rises until the voltage across the FPD reaches the junction breakdown voltage. The lateral *npn* associated with the FPD is then turned on and as the current through the device increases, the voltage dropped across the resistor increases the pad voltage accordingly. When the pad voltage reaches the junction breakdown voltage of the FOD, the associated *npn* is triggered and the current is shunted through the FOD. The FOD turns on at about 30 V in a nonsilicided 2-μm process. As discussed previously, the FOD has a higher intrinsic ESD capability than the FPD and if the design of this device is optimized it can yield more than 40 V μm^{-1} of width for the HBM stress [Rountree85]. The FOD has served as an effective protection device, mainly for the 2–3-μm technologies, but is not effective for the submicron technologies and hence is rarely used as a protection device.

It was also observed that, in advanced processes, the thick field device performance can degrade considerably. If LDD junctions are employed for source and drain, the protection per unit device width can degrade to 20 V μm^{-1} although it can be restored to about 30 V μm^{-1} with modifications to the source/drain diffusions [McPhee86]. The inclusion of silicides for the source and drain diffusions will reduce protection levels drastically down to 10 V μm^{-1} with little dependence on process or design parameters [McPhee86]. To overcome this, an SCR protection

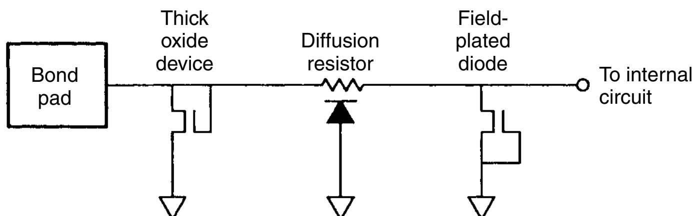


Figure 6.37 Input protection scheme with thick oxide transistor

device can be an attractive option for CMOS processes [Rountree88]. However, an optimum choice of this device, in conjunction with the isolation stage design, is essential for an efficient input protection. These issues are discussed in the following sections. The results given here are mainly for a 1.6-μm CMOS process that does not employ LDD junctions. Nevertheless, the results are equally applicable to other CMOS and BiCMOS processes for VLSI circuits.

6.7.1 SCR Primary Protection

The trigger voltage of the LSCR, V_t, shown in Figure 6.30 is between 40 and 70 V. The high trigger voltage of this makes this very attractive for high voltage protection applications where it is not uncommon to have 40 V SCRs in parallel with the high voltage drain-extended nMOS (DENMOS) or the lateral DMOS (LDMOS). The ESD protection issues for DENMOS and LDMOS are discussed in Chapter 7. As a result of its efficient clamping behavior this device briefly replaced the thick field device as the primary protection device in Figure 6.37. This meant that the secondary stage of the protection circuit, consisting of a diffusion resistor/FPD needed to support the ESD stress until the pad voltage, is high enough to trigger the SCR. When the secondary stage design was not optimal, it led to failure windows as reported by Duvvury [Duvvury89]. This problem can be eliminated if V_t is lowered.

We saw earlier that V_t can be substantially lowered by placing an n^+ diffusion at the n-well boundary, labeled as trigger diffusion in the inset of Figure 6.38. The $I-V$ curves of the MLSCR device are compared to the LSCR in Figure 6.38, also fabricated in a 1.2-μm CMOS LDD process. As expected, V_t is reduced from 50 to 25 V for the MLSCR. Note this that value roughly corresponds to the *npn* breakdown of a thick field device for this older process. When the $I-V$ curves of both LSCR and MLSCR are compared in Figure 6.38, the on-resistance of the MLSCR is relatively larger. Considering that the anode−cathode space is same for both cases, this behavior in the MLSCR could be caused by the additional impedance in the device conduction path introduced by the n^+ at the well boundary. However, the ESD performance of both devices exceeds 6 kV with 100 μm of device width. An additional point for the MLSCR is that its trigger voltage is dominated by the n^+ avalanche threshold, whereas for the LSCR, the n-well to substrate avalanche breakdown voltage is very high. It is usual that punch through will occur before the avalanche breakdown voltage is reached, and hence V_t is determined by the n-well overlap of the anode, denoted as X in Figure 6.30. The variation of V_t with X for the LSCR was reported by Rountree [Rountree88].

6.7.2 Secondary Protection Devices

The secondary protection device needs to be able to carry some current before the primary protection device is triggered. This requirement is applicable only for the SCR protection design schemes where the trigger voltage is relatively higher

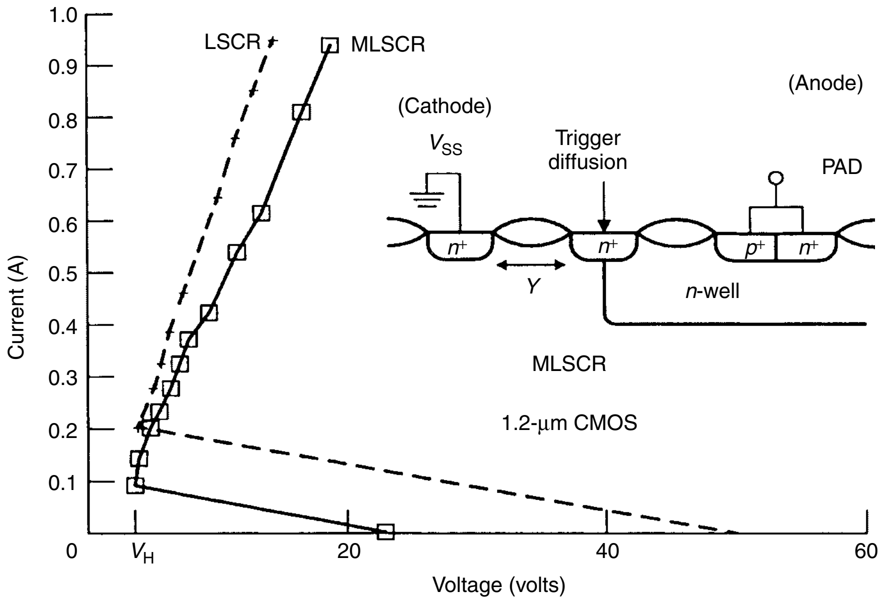


Figure 6.38 $I-V$ characteristics for MLSCR (solid line) compared with $I-V$ characteristics of LSCR without the trigger diffusion (dashed lines), for the 1.2-μm CMOS process with LDD junctions

than the input gate oxide breakdown under ESD conditions. That is, a grounded gate nMOS with a series resistor is needed to serve both as an oxide clamp as well as a breakdown device to support the initial ESD current before the SCR triggers. Therefore, for efficient protection circuit design, the devices used in the secondary protection have to be optimized. In this section, we examine the use of a secondary protection scheme consisting of a resistor and an FPD. The two elements are considered separately and then the combined performance of the two is discussed.

6.7.2.1 Field Plate Diode

The $I-V$ characteristics of a 80/2 μm/μm grounded gate nMOS device are shown in Figure 6.39. The device goes into *npn* snapback after the drain avalanche breakdown at about 14 V. At higher current pulses the device eventually enters the thermal second breakdown region, which eventually leads to failure. The ESD failure threshold level has been correlated to the second breakdown trigger current level [Polgreen89][Amerasekera90]. Considering the 1.5 kΩ for the HBM, the failure current level of 800 mA in the figure, which is obtained with the 150-ns wide constant current pulse, corresponds to about 1200 V HBM failure threshold. This translates to 15 V μm^{-1} for the 1.6-μm nonsilicided process. In a 1-μm

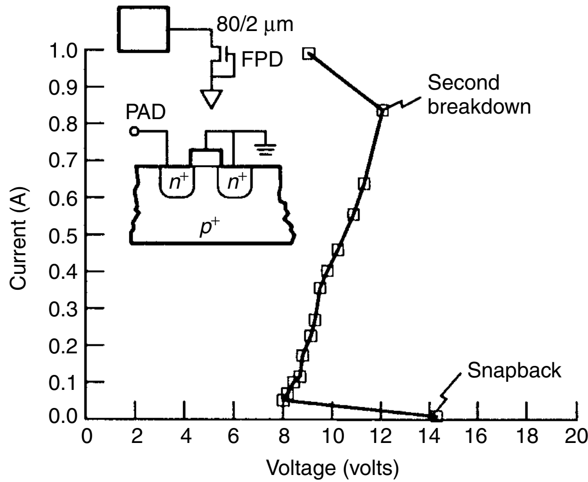


Figure 6.39 $I-V$ characteristics for a 80/2-μm FPD device with 1.6-μm CMOS process

silicided process, this figure of merit can reduce to 4–5 V μm^{-1} [Polgreen89]. Referring back to Figure 6.39, the FPD can be effective in the input protection scheme as long as it is prevented from going into second breakdown. The design of the resistor and the primary SCR should take care of this criterion as discussed below.

6.7.2.2 Isolation Resistor

A diffusion resistor is commonly used for isolation stage protection. However, the diffusion resistor also acts as a parasitic diode to the substrate and may not support the voltage needed to trigger the primary device such as the SCR. That is, the effective value of this resistor can become only a fraction of the designed value. Hence, a large resistor is required to eventually build up the pad voltage for SCR trigger. Moreover, because of the resistor diode breakdown to the substrate, the contacts to the resistor easily get damaged. This situation can be improved by placing an n-well around the resistor at the pad to suppress avalanche of the resistor diode. With this technique, only a minimum resistor will be needed for the overall protection.

A second type of diffused resistor is the n-well resistor [Carbajal92]. The main benefit of this resistor is its current saturation characteristic discussed in Chapter 4. Hence, a small resistance at low current levels can become very high as the current is increased, which is an attractive property to the circuit designer. However, the n-well resistor also has a negative resistance characteristic and will snap back to a low impedance mode at high voltages. The design of the resistor needs to comprehend the snapback problem for effective behavior. The main advantage in using the

n-well resistor is the current saturating phenomenon, which is determined by the width of the well. The snapback voltage is defined by the length of the resistor. Some numbers have been given in Chapter 4.

Polysilicon resistors have also been successfully used as part of the protection circuits [Duvvury83]. However, as these elements are encapsulated in low thermal conductivity oxide they are thermally isolated and the power dissipation causes damage at relatively low ESD levels. As reported by Fukuda [Fukuda88], they can be used to improve the Machine Model performance.

6.7.3 Protection Scheme

The constituent protection elements can be combined to form the total input protection circuit. The secondary protection is first examined before combining the SCR to form the full protection scheme.

The $I-V$ characteristic when the avalanche-suppressed diffusion resistor is combined with the FPD is shown in Figure 6.40. Immediately after the nMOS of the FPD breaks down the current through the resistor causes an increase of the voltage at the pad. For the secondary protection layout (see Figure 6.45) a parasitic thick field device (formed with n^+ diffusion of the resistor, p-substrate, and n^+ connected to V_{SS} of the FPD) turns on when the n^+ to substrate avalanche breakdown voltage is reached. This is one of the disadvantages of using a diffusion resistor, as will be discussed later. In agreement with this assumption the clamping voltage approaches 15 V, typical for the snapback for a thick field npn.

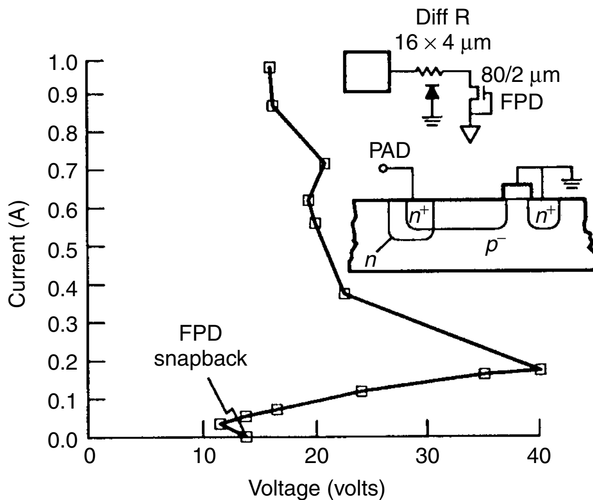


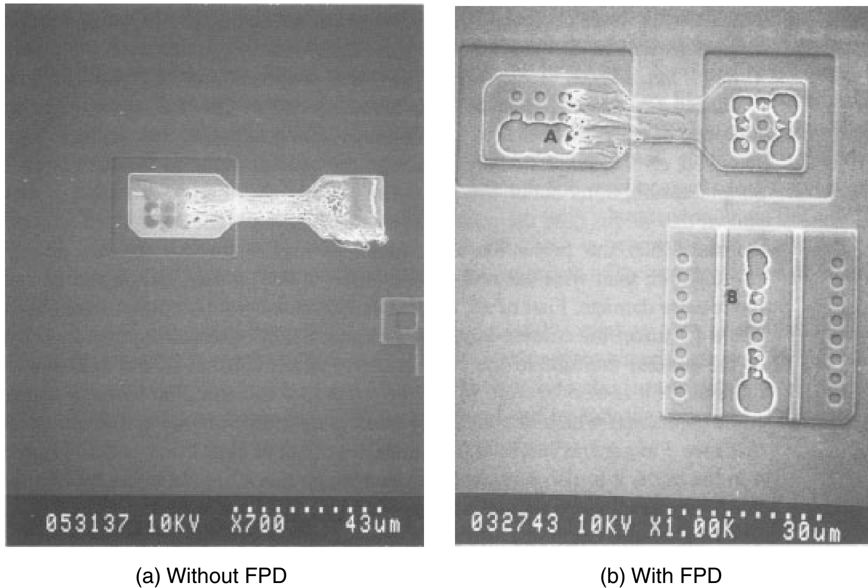
Figure 6.40 $I-V$ characteristics for secondary stage protection with a 16×4-μm diffusion resistor and a 80/2-μm FPD device. The process is 1.6-μm CMOS

During ESD stress, after the *npn* snapback occurs at the FPD, the voltage at the high end of the resistor near the pad increases until a failure occurs. If the FPD were removed, all the current would go through the resistor/diode to the substrate. With simple assumptions about the substrate resistance and diode avalanche, it can be shown that in this case the power dissipation would be higher at the pad side of the resistor. Thus the protection level should be higher for the resistor plus FPD combination than with the resistor/diode alone, if the limiting failure mechanism is the resistor damage. First of all, the HBM failure threshold distributions are shown in Table 6.1 under the column labeled 'V_f (without SCR)'. Comparing case 2 with case 3, the general increase in the protection level with inclusion of the FPD device is quite clear. It is also interesting to note from case 1 that the FPD by itself can offer some protection, which is enhanced by adding the resistor/diode combination. Note that case 3 in general seems to be a cumulative effect of cases 1 and 2. From case 4 in the table, it is also seen that the diode/resistor has a limited effect on improving the failure level as increasing the resistor size does not further enhance it.

These results become clearer when the failure modes are examined. In Figure 6.41(a) the failure mode for the resistor without the FPD is shown. The damage is seen at the end contacts indicating the suppression of avalanche because of the *n*-well. The region outside the *n*-well avalanches and eventually electrical failure occurs when the damage region reaches the substrate. Now if the FPD is added, the damage is shown (Figure 6.41(b)). Here, for an electrical failure, the damage occurs at both points A in the resistor (similar to damage in Figure 6.41(a) and at point B in the FPD. The damage at point B is simply the common gate–drain short. What is more interesting is that the resistor damage is confined in the resistor body and cannot be electrically detected until the short occurs at the FPD. This is an example of how electrical failure detection cannot always reveal the pre-threshold failures. Incidentally, the failure voltage distribution for case 3 corresponds to only resistor damage at the low end and both resistor and FPD damage at the high end. That is, in some cases as noted earlier, if the damage is confined to the resistor only, failure level goes up to 3 kV before the electrical damage is detected.

Table 6.1 Input protection performance with different choices of secondary protection design

Case no.	R (ohms)	W/L of FPD (µm)	V_t (Without SCR) (kV)	V_t (With SCR) (kV)
	(diffusion)			
1	—	80/2	0.9–1.2	0.9
2	150	—	1.2–2.0	>6
3	150	80/2	1.9–3.1	>6
4	240	80/2	2.0–2.5	>6
	(polysilicon)			
5	85	80/2	0.9–1.0	>6
6	150	80/2	0.8–1.0	>6
7	220	80/2	0.7–1.0	>6



(a) Without FPD (b) With FPD

Figure 6.41 Failure sites for secondary protection with diffusion resistor. The damage shown in (a) is for failure in the diffusion resistor without the FPD at a stress level of 2 kV. In (b) the damage locations with the FPD included are shown at a stress level of 3.1 kV. Note that in (b) the damage appears in both (A) the resistor and in (B) the FPD

Instead of the diffusion resistor, a polysilicon resistor may also be used. The $I-V$ curve for a combination of polysilicon resistor and FPD is shown in Figure 6.42. After the FPD snapback the current through the resistor causes an increase in voltage at the pad. Note that the pad voltage build up is continuous without any breakdown. This is expected as no parasitic devices are present and the resistor supports the full voltage. A more detailed analysis of different sized polysilicon resistors is shown in Figure 6.43. At the higher current levels the resistor seems to heat up causing an increase in the resistance, where the onset of heating is indicated by arrows. Note that the heating effect seems to be minimum for the widest resistor. In contrast to this, the heating effect is not observed for the diffusion resistors as shown in Figure 6.44 where much of the current is caused by avalanche breakdown of the diode to substrate. Although there was no evidence that heating in the polysilicon resistors has any adverse effect on its performance, for practical applications the safe region of operation should be kept below the onset of heating. Hence, the ESD protection design can be achieved by keeping the polysilicon resistor operation in the region of low power dissipation. The design variations to characterize polysilicon, diffusion, and n-well resistors are described in the Sematech Document [Voldman99].

The HBM failure voltages for polysilicon resistors are also shown in Table 6.1 (see 'V_f (without SCR)'). Although there is some scatter in the data, it is clear that

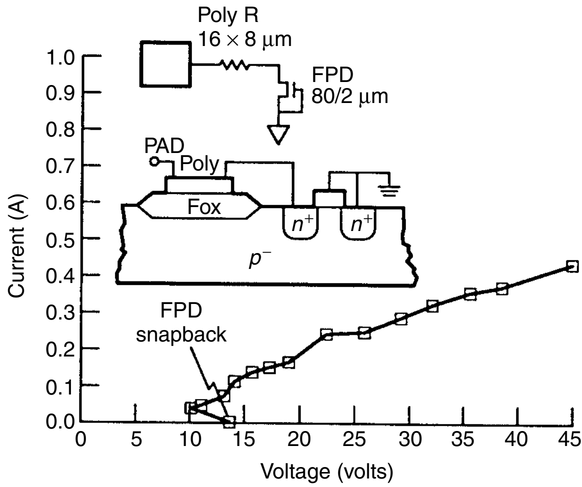


Figure 6.42 $I-V$ characteristics for secondary stage protection with a 16 × 8-μm polysilicon resistor and a 80/2-μm FPD device. The process is 1.6-μm CMOS

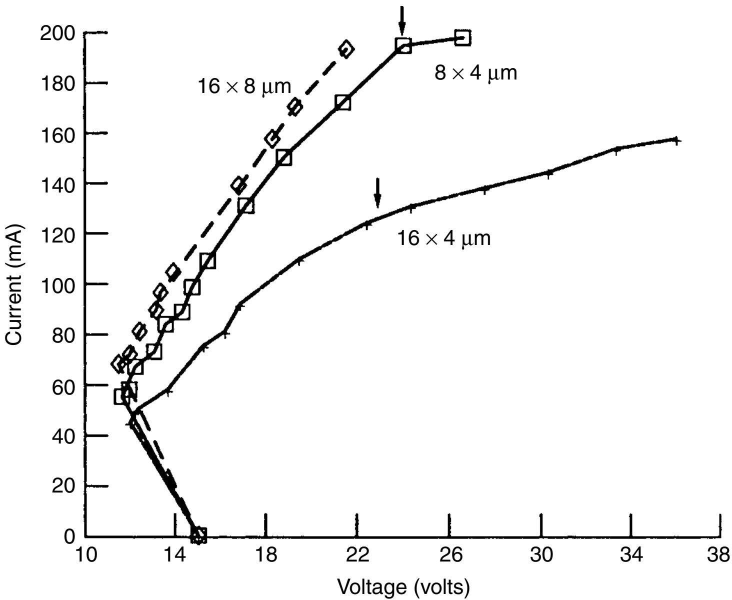


Figure 6.43 $I-V$ characteristics of polysilicon resistors combined with an FPD. The arrows point where heating effects seem to begin

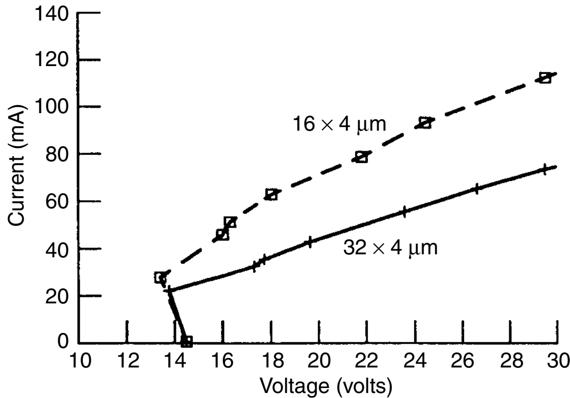


Figure 6.44 $I-V$ characteristics of diffusion resistors combined with an FPD

the minimum failure level increases for the smaller resistor. The differences are not distinct because the failure level of the FPD is very close (see case 1). But it is clear that, unlike for the diffusion resistor, adding the FPD does not improve its failure level. This shows that either the polysilicon resistor or the FPD fails around the same point.

6.8 TOTAL INPUT PROTECTION

The protection elements have been characterized in the previous sections with pulse testing. The ideal combinations of these for an effective input protection are discussed here. Both the diffusion resistor and the polysilicon resistors are considered separately. For the polysilicon resistors the reliability aspects are also examined.

6.8.1 Inputs with Diffusion Resistor

The design with a diffusion resistor becomes effective when the n-well is placed to suppress avalanche. A layout of such a circuit is shown in Figure 6.45. Note that the layout is arbitrary and can be changed to minimize the parasitic device effects. The matrix layout of contacts achieves the optimum performance where the current density through each contact is reduced. The $I-V$ curves for the MLSCR for this particular process of the total protection circuit is shown in Figure 6.33. Compared to the MLSCR of Figure 6.38, the trigger point is about the same but the holding voltage is lower. It should be noted that the *trigger diffusion* in this application is butted against the p^+ anode, but this modification has not been found to have any impact on the SCR operation itself. The

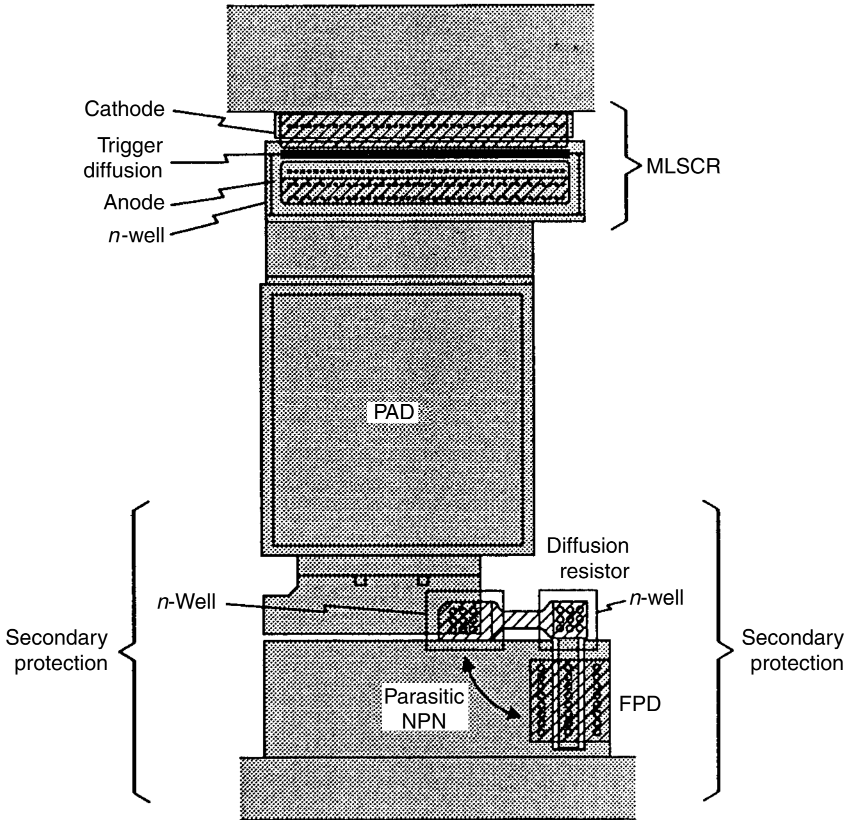


Figure 6.45 Layout of a total protection with diffusion resistor and MLSCR. Note the parasitic *npn* device

overall $I-V$ curves of the protection circuit with a $150\,\Omega$ diffusion resistor are shown in Figure 6.46. When the current level reaches 100 mA the voltage at the high end of the resistor approaches 25 V (see Figure 6.45), which is equal to a 15-V drop across the resistor plus the FPD snapback voltage of 8 V. As expected, this triggers the MLSCR device and the current level abruptly increases by 200 mA.

Design of the secondary element and the resistor essentially follows the following methodology. Assuming that an FPD is used as the secondary element, the maximum current through the FPD must be determined by characterizing these elements for a given process. If the maximum current is I_{t2} (in mA /μm) where I_{t2} is the second breakdown trigger current (Chapter 4), then the maximum allowable current should be guardbanded to about $0.75I_{t2}$. The snapback holding voltage of the FPD, V_{sp} is also determined in the characterization, as is the trigger voltage of the primary device, V_{trig}. The value of the resistor, R, to be used is then

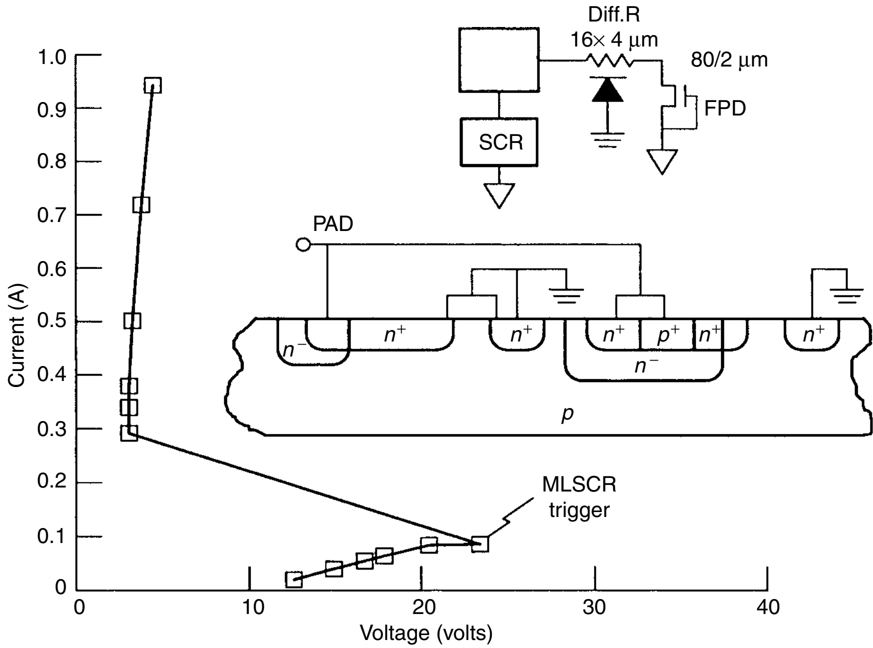


Figure 6.46 $I-V$ Characteristics for total protection with a 16 × 4-μm diffusion resistor, 80/2-μm FPD, and 150-μm wide MLSCR. The process is 1.6-μm CMOS

determined by

$$R \times W \geq \frac{V_{\text{trig}} - V_{\text{sp}}}{0.75 I_{t2}} \tag{6.1}$$

The HBM ESD performance by combining with MLSCR is summarized in Table 6.1 under V_f *(with SCR)*. Considering case 1, it is obvious that the FPD by itself cannot trigger the SCR device as its breakdown voltage is lower. Thus failures occur at low levels corresponding to the FPD protection level. In case 2, the diffusion resistor (with avalanche suppression) alone in parallel with the SCR cannot obviously have an impact on the SCR function. However, for protection of the input gate oxide, an FPD device is also needed. Thus only for case 3, when both the FPD and the resistor are combined, full effective SCR protection for inputs is achieved.

For maximum circuit speed at the input, the isolation resistor needs to be as small as possible. With a regular LSCR of 50 V trigger, this cannot be easily achieved for an n-diffusion resistor. Even with avalanche suppression the full resistor value is not realized and failures occur unless the resistor is made more than 100 Ω. However, when the avalanche-suppressed resistor is used in conjunction with an MLSCR, a minimum necessary value for the resistor can be used. n-well resistors will also have limitations in the maximum voltage across the resistor before snapback occurs.

6.8.2 Inputs with Polysilicon Resistor

The input combination with the polysilicon resistor is shown in Figure 6.47 with the measured $I-V$ curves in Figure 6.48. As before, after the FPD snaps back the pad voltage builds up through the $I \times R$ drop across the resistor to trigger the MLSCR. As long as the SCR trigger occurs below the failure current, the protection scheme would be effective. The curves of Figure 6.48 were measured for the total protection employing the 16×8-μm polysilicon resistor with an effective resistance of 85 Ω as shown in Figure 6.43. The MLSCR is observed to trigger at a low current level of 150 mA (or approximately 250 V HBM). Equation 6.1 is valid for polysilicon resistors as well. An optimum value for the resistor, considering the 15-V drop needed, would be 60 Ω. Values smaller than this would not allow sufficient pad voltage build up and larger than 100 Ω would increase the chances for some latent damage in the secondary protection. It is

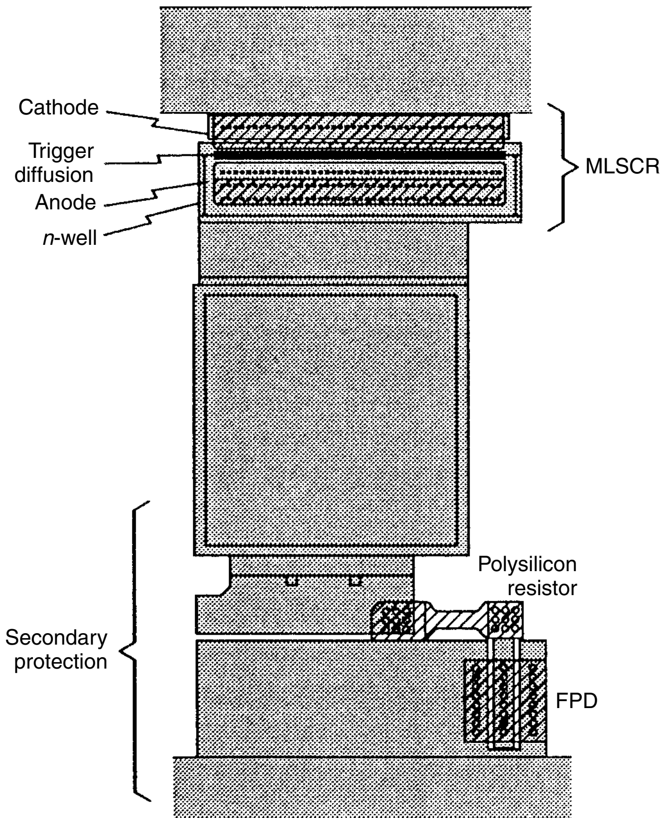


Figure 6.47 Layout of a total protection with polysilicon resistor and MLSCR

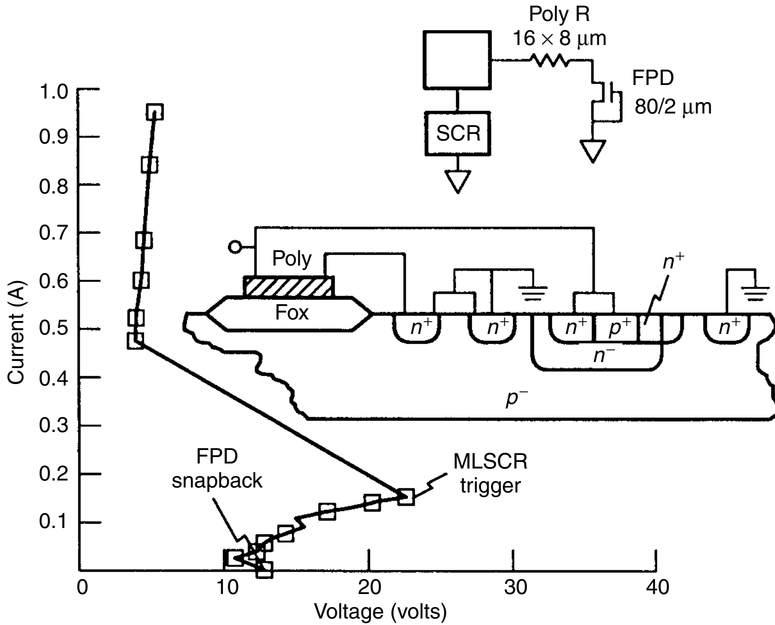


Figure 6.48 $I-V$ Characteristics for total protection with a 16 × 8-μm polysilicon resistor, 80/2-μm FPD, and 150-μm wide MLSCR. The process is 1.6-μm CMOS

interesting to note in Figure 6.48 that following the MLSCR trigger the current level increases by about 330 mA, which is more than for the diffusion resistor case in Figure 6.46. This indicates that after the MLSCR triggers, most of the current flows through this device, removing all of the stress from the polysilicon resistor. This is important for effective and consistent performance of this scheme.

The HBM ESD performance with the polysilicon resistor is summarized in Table 6.1 under V_f *(with SCR)*. The protection for the polysilicon resistors in cases 5–7 all worked well. This is not surprising as the needed voltage drop of 13 V can be achieved below their failure thresholds without the SCR device. A minimum value between 60 and 100 Ω can be easily selected for minimum RC delay effects at the input. A smaller resistance can be used but the width will have to be made greater than 8 μm. In this case, the trade-off between the area and the input delay has to be evaluated. An even lower resistor value can be used if an LVTSCR is combined instead of the MLSCR. This is because the LVTSCR can have a low trigger voltage in the range of 12–16 V depending on the process. This protection design with the MLSCR was also evaluated for negative stress with no failures for greater than 6 kV stress. For this polarity stress, apparently the forward diode in the MLSCR device clamps the voltage and does not allow any significant conduction of the diode from the FPD device. Testing with

the Machine Model has shown that this protection scheme with polysilicon resistor/MLSCR is very effective for this stress also, with a failure threshold greater than 1000 V. This is well in excess of the usually required 200 V for the Machine Model.

6.8.3 Polysilicon Resistor Reliability

The reliability of the polysilicon resistor in a protection circuit would be of great concern, as damage to it may not be electrically detected. To evaluate this, the secondary protection scheme with a polysilicon resistor was stressed at different HBM levels. The stressing was done with 10 pulses in each case to give confidence to the results. The SEM photographs of the stressed devices and the stress levels are shown in Figure 6.49. It is interesting to note that for both 200 and 400 V stress levels there is no physical damage to the polysilicon resistor. At the 600-V level, the first indication of heating is seen at the tapered portion of the resistor. It is possible to improve the layout to reduce this effect. Nevertheless, at

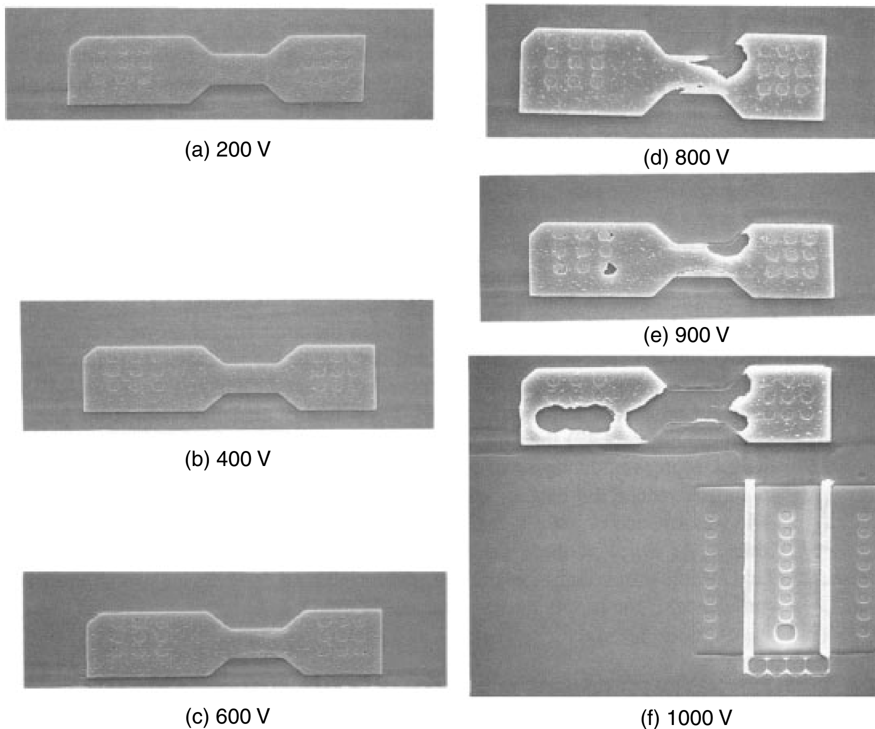


Figure 6.49 Detailed failure analysis of secondary protection with the 8×4-µm polysilicon resistor. The stress conditions for HBM are: (a) 200 V; (b) 400 V; (c) 600 V; (d) 800 V; (e) 900 V; and (f) 1000 V. The initial damage for 600 V stress is indicated by the arrow

higher stress levels the resistor is eventually damaged. What is more interesting is that at 800 and 900 V levels there is obviously severe damage to the resistor, which was not electrically detectable as a continuous polysilicon filament is present. This is another demonstration of the deceptive nature of pure electrical analysis. It is interesting to note here that for the 900-V stress level some contact damage is also present, indicating that the current density through each contact exceeds the failure level. This can also be improved by increasing the number of contacts. At the 1000-V stress level the polysilicon is completely blown giving rise to an open failure. In this experiment in all cases there was no damage to the FPD device. However, in some cases at the 1000-V stress level damage to the polysilicon resistor at the taper corner and to the FPD occurred simultaneously. This is because the failure level of the FPD is also close to this stress level (see Table 6.1, Case 1).

From the analysis of Figure 6.49 it is seen that the total protection with an SCR would require that the polysilicon resistor survive at least 400 V of stress. Referring to Figure 6.48, the SCR triggers at 150 mA or approximately 250 V for the HBM. This clearly indicates that there is a safe margin between the SCR trigger current (at less than 250 V HBM) and the current at which thermal damage to the resistor occurs (at approximately 600 V HBM). The margin could be higher considering that the analysis was done for a 8 × 4-μm resistor which has a higher tendency to heat up (see e.g., Figure 6.43) than the 16 × 8-μm resistor actually used in the protection design. Moreover, the layout of the 16 × 8-μm resistor did not have the taper regions to further reduce the heating effects. Finally, the total protection circuit of Figure 6.48 was stressed 20 times at ±6 kV to note the robustness of the polysilicon resistor. As expected, there was no physical damage to the resistor after this stress. This clearly shows that a polysilicon resistor with an SCR device can be used with confidence, if the proper design synthesis is used.

6.8.4 Selecting an SCR Protection Circuit

In the preceding sections the protection design with an MLSCR device was analyzed. It was shown that once the primary protection is chosen the secondary protection design becomes critical. A scheme with a diffusion resistor can be effective as long as avalanche suppression is employed using an n-well around the n^+ diffusion. Even without n-well the protection works well but requires a larger resistance value for safe design. The higher resistance will have an impact in circuit applications that demand high speed. With the n-well suppression in place the resistance can be much lower for the same ESD performance. Protections with the MLSCR and diffusion resistor, as described here and shown in Figure 6.46, should use a 100 Ω resistor and a 80 μm wide FPD. The channel length can be minimal, although 20% greater than the minimum channel length is recommended for reducing possible leakage effects at the inputs. The

resistor width should be a minimum of 4 µm with the *n*-well overlap of diffusion at 6 µm.

The input protection can also be effective with the use of a polysilicon resistor in conjunction with an MLSCR. There are several advantages in using this approach. First, the parasitic capacitance is significantly reduced to minimize the *RC* delay. Second, polysilicon resistors are not prone to the voltage coefficient like the diffusion resistors are. For example, in voltage divider designs with a series of diffusion resistors the depletion regions will vary, giving rise to nonlinear effects. But this is not the case for polysilicon resistors. In analog circuit applications a polysilicon resistor is preferred because of its better leakage performance at high temperatures. Finally, protection designs with a polysilicon resistor will have the additional advantage of being immune to parasitic device interactions or *proximity effects* as reported for a DRAM protection design [LeBlanc91]. In contrast to these advantages, a polysilicon resistor is highly susceptible to heat damage as it is encapsulated with no available heat sink from the substrate. Thus, there is always a possibility that any latent damage can go undetected. But as shown in this section, careful analysis and design can eliminate such phenomena. The layout implementation should be as in Figure 6.47 but with the taper regions removed. In silicided processes a 4-µm wide polysilicon resistor should be adequate but there has not been any data reported to establish a definite recommendation.

As discussed in detail in this section, the SCR protection device has been very effective, especially for the technologies above 1µm. The introduction of low substrate resistance for latchup immunity combined with the use of shallow trench isolation (STI) for high chip density has made the SCR device ineffective for the submicron technology protection applications. However, for the 0.25-µm technologies the SCR optimization looks promising again, and this issue will be discussed in Chapter 7.

6.9 ESD PROTECTION USING DIODE-BASED DEVICES

For *p*-substrate CMOS technologies, the n^+ in substrate and p^+ in *n*-well parasitic diodes are extremely useful ESD protection devices. They are highly reliable under forward bias, with a forward breakdown current around 50 mA μm^{-1} device width. Their use under reverse bias is not recommended, as other devices break down below the diode avalanche voltage and as diodes in a silicided process fail at small reverse-bias currents. For ESD protection strategies based on other protection elements, diodes still participate in ESD discharges, for instance, in shunting current from V_{ss} to V_{dd} through well diodes. It is important to understand where these diodes are and how they shunt discharge currents for different pin combinations. When necessary, explicit diodes can be added. For a V_{dd}-based primary ESD protection strategy, an understanding of the diode elements is key.

6.9.1 n^+ Diffusion Diodes

Regardless of the primary ESD protection strategy, diodes to V_{ss} are important. As discussed in Section 5.2, V_{ss} usually provides a common point to which all ESD current is eventually shunted. For positive discharges where a non-V_{ss} pad is at the negative test terminal, current must flow from the common V_{ss} point to the negative test terminal, usually through a forward-biased diode. Figures 5.3, 5.4, 5.5, and 5.6 show several such examples. For *p*-substrate technologies with the substrate at V_{ss}, any n^+ diffusion in the substrate will form this diode. For I/O pad protection, this diode can either be drawn explicitly or can be implicit in the drain of the nMOS pull-down.

In *p*-epi processes, the highly doped substrate provides a good conduction path, and the n^+ diffusion diode generally conducts with minimal series resistance. Therefore, a diode layout with minimum width strips and a total length of 50–100 µm can usually limit the voltage across the output driver nMOS to a reasonable level. Long narrow strips conduct best, especially in LOCOS technologies, as forward injection is strongest at the perimeter. In an STI process this may not always be true. Test structure measurements will always reveal which layout works best [Sematech98][Voldman99].

6.9.2 p^+ Diffusion Vertical pnp

A p^+ diffusion in an *n*-well creates the diode that is fundamental to the V_{dd}-based ESD clamping scheme. As shown in Figures 5.4 and 5.6, where they are represented as the primary and secondary ESD devices, the diode conducts from the I/O pad to the power rail of the output driver pull up, which may be either the core supply V_{dd} or an isolated I/O driver supply V_{ddio}. Such devices may also be placed between power supply rails to provide clamping action during ESD events. They may also be stacked, for either the I/O or the power supply application, to allow higher voltage tolerance (see Section 7.5.4).

In *p*-substrate technology, the p^+ diode device, whose cross section is shown in Figure 6.50, is the emitter–base diode of a vertical *pnp* transistor, with the substrate acting as a common collector. Layout of the *pnp* is critical to achieve the minimum clamping voltage during ESD. To minimize the base resistance, the n^+ well plug should surround the p^+ emitter on all sides at the minimum allowed spacing, as shown in Figure 6.51. The p^+ emitter diffusion should be laid out in minimum width strips. This maximizes the perimeter of the well plug base contact facing the emitter, minimizing base resistance. It also maximizes emitter perimeter for perimeter-dominated injection mechanisms.

For *pnp* clamps on output pads with push–pull drivers, the *pnp* can be partially or fully integrated with the pMOS output device through the layout technique shown in Figure 6.52. An n^+ well strap connected to V_{dd} or V_{ddio} is placed between p^+ strips connected to the pad, as shown in Figure 6.53, providing the equivalent

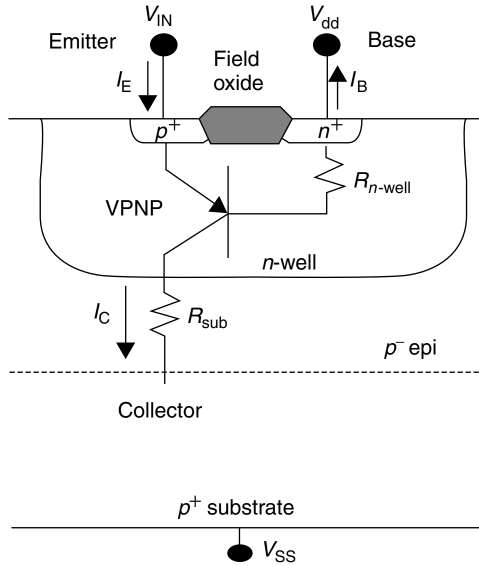


Figure 6.50 A cross section of the vertical *pnp* device for an n-well CMOS process. A diode is present between the p^+ emitter and the n-well (After [Amerasekera95], reproduced by permission of ©1995 IEEE)

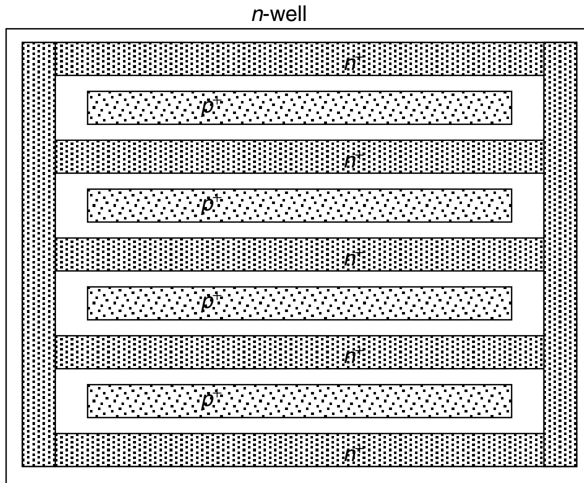


Figure 6.51 Top view of the recommended layout for the vertical *pnp* structure. The p^+ and n^+ diffusions should be separated by the minimum allowed spacing for best ESD performance

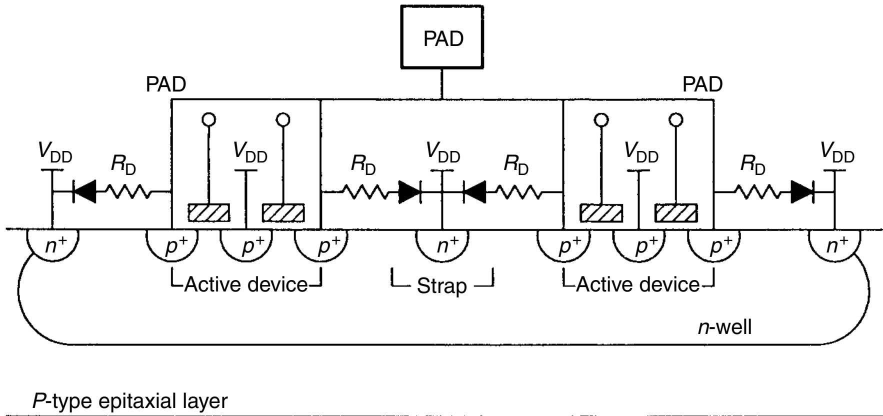


Figure 6.52 Cross section of pMOS transistor with improved lateral diode to V_{dd} by placing additional n^+ straps

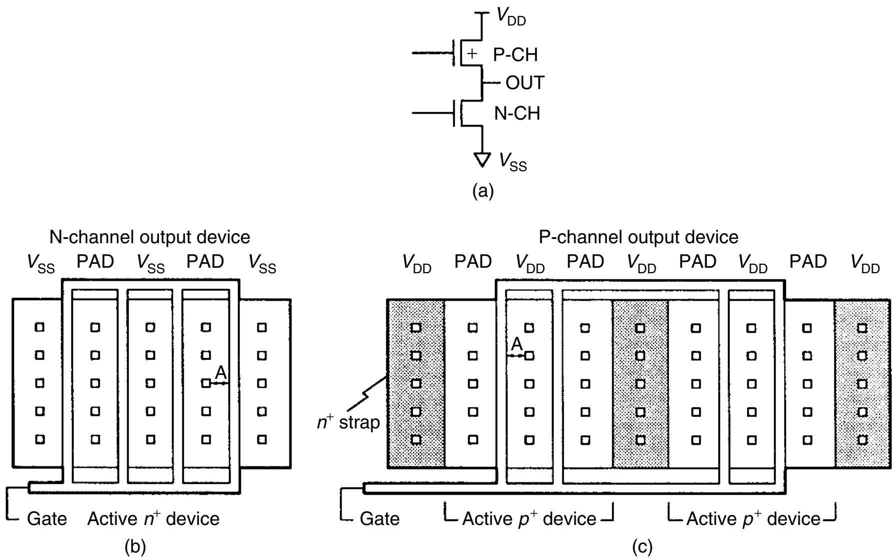


Figure 6.53 Layout of pMOS transistor with improved lateral diode to V_{dd} by placing additional n^+ straps

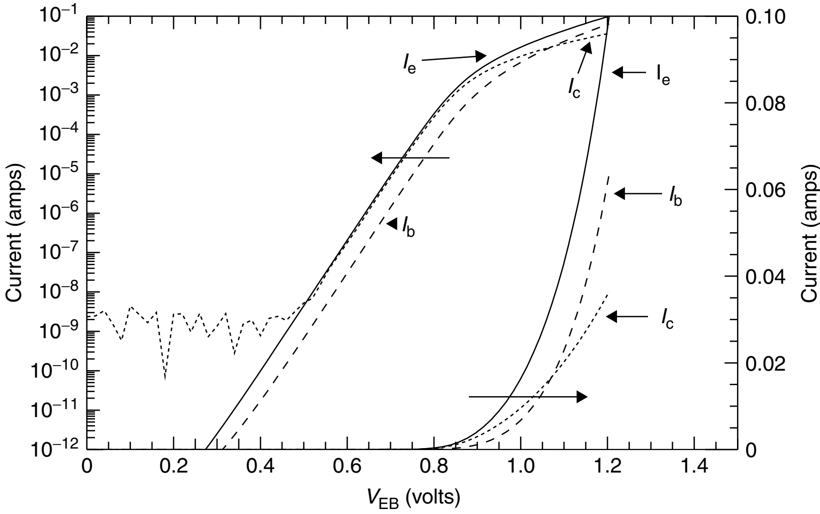


Figure 6.54 The $I-V$ characteristics of a vertical *pnp* device in a *p*-epi CMOS process showing the emitter, collector, and base currents as a function of emitter-to-base voltage. The base and collector are both grounded during the measurement

function of one n^+ well strap in the dedicated *pnp* of Figure 6.50. This technique has been shown to improve the ESD performance of a non-rail-based protection circuit as well [Duvvury88B].

To fully understand the operation of the vertical *pnp*, either in its pure form or as integrated with the output driver, its $I-V$ should be characterized with scaled-down test structures [Sematech98][Voldman99]. It is particularly important to determine how the currents scale as a function of emitter width and number of emitter fingers. Figure 6.54 shows an example $I-V$, measured by ramping V_{EB} with $V_{\text{BC}} = 0$. It exhibits two modes of operation. At $0.4\,\text{V} < V_{\text{eb}} < \approx 0.8\,\text{V}$, the device behavior is ideal, with $I_c = \beta I_b$ and a slope of $60\,\text{mV decade}^{-1}$ at room temperature. For V_{eb} greater than approximately 0.8 V, the device encounters high-level injection and series resistance effects. Although the *pnp* operates in this more complicated region during ESD, SPICE can model it fairly accurately. For a V_{dd}-based clamping scheme, the SPICE model of the *pnp* clamp and the rail clamp can be combined to determine the effectiveness of the full ESD clamping system as discussed in Section 5.2.5.

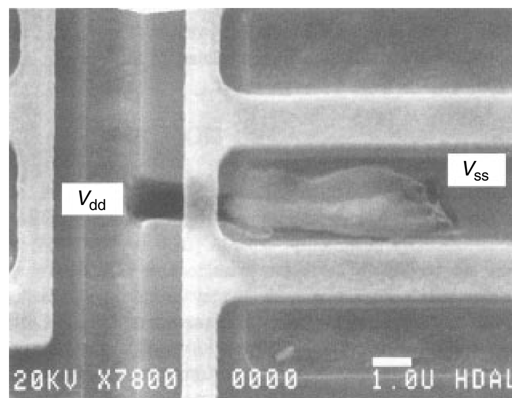
6.10 POWER SUPPLY CLAMPS

Even with effective protection at the pins, many cases of damage phenomena do occur internally in the chip (see, e.g., [Duvvury88A][Maene92][Cook93]). Some of

these can be directly attributed to inadequate protection provided for stress between the power bus lines whereas others may be caused by deficiencies in the layout of the protection circuits. The power bus protection issues and techniques will be discussed in this section.

The MIL-STD testing method requires stressing of all inputs, outputs, and power bus pins (V_{dd}, V_{cc}, V_{ddio}) with respect to the ground pin (V_{ss}). Thus it would seem logical to place a protection circuit between V_{dd} and V_{ss} for direct stress between the two [Palella85][Duvvury87]. The power supply protection must clamp at an adequately low voltage because there are many parasitic devices in the internal chip that may turn on instead of the designated protection circuit. These parasitic devices essentially form the weak links and reduce the power bus protection. For example, in the internal layout of a chip there are many areas where an n^+ diffusion connected to V_{dd} could be close to another n^+ diffusion connected to V_{ss}, such as in the latchup guard rings. These would then form parasitic lateral *npn* devices. Therefore, even if a robust protection circuit is implemented between V_{dd} and V_{ss}, the ESD damage could still occur because of triggering of these parasitic devices which are not designed to sustain high current levels. An example of one such fail site is shown in Figure 6.55. Observe how the molten silicon has tunneled under the polysilicon from the V_{dd} diffusion to the V_{ss} diffusion. To ensure that the protection device between V_{dd} and V_{ss} is effective, the diffusion to diffusion spacing for such internal parasitic thick field devices should be made longer, reducing (preferably eliminating) the effectiveness of the parasitic *npn* transistor.

A wide variety of power supply protection devices and circuits are available. If the I/O protection uses V_{ss}-based protection, such as the grounded-gate nMOS or the gate-modulated nMOS, these protection elements often suit the power rails equally well. The V_{ss}-based clamp affords any nMOS device on the power rail the same protection it provides the nMOS output driver. However, especially with GGNMOS



Damage due to V_{dd}-V_{ss} stress

Figure 6.55 Damage site observed for the V_{dd} to V_{ss} stress

protection, gate interactions or power rail resistance may allow current through the non-ESD path [Duvvury88A][Chaine97]. Special purpose power supplies, particularly ones powering large devices such as clock drivers, can exhibit unexpected failures when snapback clamps are used. Chips containing several power bus lines should be examined for ESD weakness between all bus combinations.

Supply clamps for chips using a V_{dd}-based clamping scheme are subject to an even higher standard. Not only should the supply clamp be robust and effective for direct discharges to the power rail, but it must also restrict the voltage of the *pnp* base during discharges to the I/O pin. Snapback devices for the V_{dd}-based supply clamps are generally not practical. If they were, they would be used as I/O protection elements. Furthermore, a snapback clamp would hold the supply at the snapback voltage. The I/O pad, and therefore the voltage across the nMOS driver, would be at least a diode drop above this, causing such an excessive voltage between the nMOS driver's drain and source that it would fail in most cases. Thus, another type of clamp is needed.

Three categories of non-snapback supply clamps are in widespread use for V_{dd}-based ESD: capacitive clamps, transient-timed clamps, and power-supply referenced clamps.

The capacitive clamp is not really an explicit clamp at all, instead it relies on the on-chip capacitance of the power supply itself to limit the voltage during ESD. If nothing else on the chip sinks current from the supply, a supply with capacitance C_{vdd} will share the charge from the ESD source C_{ESD} such that, after the discharge, the supply will be at a voltage

$$V = \frac{V_{ESD} C_{ESD}}{C_{ESD} + C_{vdd}}, \tag{6.2}$$

where V_{ESD} is the initial voltage across the ESD source capacitance C_{ESD}. For example, a 3-kV HBM pulse ($C_{ESD} = 100\,\text{pF}$) discharged onto a 100-nF supply will result in 3 V across the supply on chip. Some advanced microprocessors have even larger power supply capacitance, reducing the final voltage even further. If the voltage given by Equation 6.2 is low enough, the supply needs no explicit protection. Chips with small, isolated supplies, such as for analog and I/O circuits, still need explicit clamps.

Even if the V_{dd} to V_{ss} capacitance is too small to completely absorb the ESD charge, it can still contribute significantly to the ESD performance of an IC. A direct stress between V_{dd} and V_{ss} will first have to charge up this capacitance, which will slow the rise time of the current pulse and limit the voltage during stress [Duvvury88B]. If the supply uses a snapback protection device, its triggering will only occur after the voltage reaches the required trigger voltage, and by this time the stress current could be well below its peak level. Hence, the chip capacitance will limit the stress in the protection device and increase the ESD levels.

The second type of power supply protection, transient-timed clamps, use the transient rise of the supply along with a timing circuit, usually an *RC* network, to enable a large device that shunts the ESD event on the supply

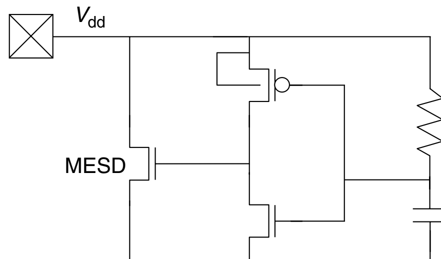


Figure 6.56 Transient power supply clamp using an RC timer, modified from the original description by [Merrill93]

[Merrill93][Worley95]. Figure 6.56 shows one example of such a circuit. The RC timer drives the gate of the nMOS clamping device MESD through an inverter. The device MESD can either operate in snapback or can be sized to handle the ESD current through channel conduction alone, which is usually the case. The number of inverters may be adjusted, together with swapping the position of the resistor and capacitor for each added inverter stage. As the inverters increase the fan-out from the capacitor to the large MESD device, more inverters allow a smaller capacitor. However, more inverters also increase the delay before the gate of the MESD device is fully on, causing an initial transient spike in voltage during ESD. Most implementations use between one and three stages.

An alternative transient-timed clamp design uses a string of stacked *pnp* emitter–base diodes in either by itself [Dabral93][Dabral94] or in series with the MESD device [Maloney95]. The *pnp* current gain β reduces the current going to the MESD device by $1/(\beta + 1)^m$, where m is the number of *pnp* emitter–base diodes in the string. The reduction of β and the decrease in required clamping voltage in advanced CMOS processes make this configuration less effective than it once was. Transient-timed clamp adaptations for high-voltage supplies, that is, those powered to a voltage higher than the nominal process-rated voltage, are discussed in Section 7.5.5.

The power-supply referenced clamp, shown in Figure 6.57 works in the same fashion as power-supply referenced gate modulation, described in Section 6.5. It uses the voltage level on an independent supply, here V_{dd}, to distinguish between ESD events and normal operation. During an ESD event on V_{ddio}, capacitance between the reference supply V_{dd} and V_{ss} keeps the reference supply near V_{ss}, allowing the clamp to turn on. During normal operation, V_{dd} is powered on, keeping the clamp off. For multiple supplies, clamps can be placed on each and cross-referenced to one another as shown in Figure 6.57(b) [Anderson98B]. For any protection scheme using supply-referenced clamps, it is important to ensure that the pin under protection and its reference are truly independent by checking for

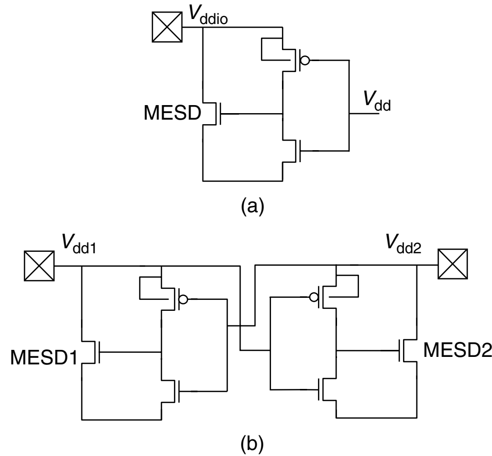


Figure 6.57 Examples of (a) power-supply referenced and (b) cross-referenced power supply clamps (After [Anderson98B])

parasitic leakage paths between them. There should also be enough capacitance on the reference supply to overcome tester-induced coupling [Anderson98B]. This type of clamp also requires that the system bring up the power supplies in a particular sequence during normal chip operation, making the transient-timed clamp a better choice in most cases.

6.11 BIPOLAR AND BiCMOS PROTECTION CIRCUITS

6.11.1 Introduction

The main protection circuit issues for bipolar processes are very similar to those described earlier for CMOS processes. The requirements of the protection circuit in terms of the trigger voltage, high current behavior, capacitive and resistive loading, and area constraints must all be taken into consideration during the design phase. BiCMOS processes that combine both bipolar and CMOS elements on the same chip can have benefits in terms of the available protection elements as well as disadvantages because of the complexity of the circuits to be protected. The present generation of bipolar and BiCMOS circuits are targeted at high-speed applications such as telecom. These applications present additional constraints on the design of ESD protection circuits because of their inability to tolerate any additional capacitance at the I/O pads or resistance in the output transistors themselves. In this section we will briefly look at those protection circuit design issues specifically related to bipolar and BiCMOS circuits.

6.11.2 Protection Circuit Strategies

Large bipolar transistors usually have very good intrinsic ESD protection levels as they are designed to operate as vertical *npn* devices rather than the lateral parasitic *npn* transistors available in MOS processes. Typical ESD capabilities for an advanced bipolar *npn* transistor is about $30\,\mathrm{V\,\mu m^{-1}}$ compared to less than $15\,\mathrm{V\,\mu m^{-1}}$ for the lateral *npn* transistor in a CMOS process. These large *npn* transistors can be made self-protecting provided the basic ESD layout rules for uniform current flow and peak electric field reduction are followed. Some of these rules have already been discussed in Section 6.3 for nMOS transistors; in this section we will discuss specific bipolar-related issues.

BiCMOS processes can have CMOS output buffers which are much weaker ESD elements than *npn* transistors [Amerasekera92][Tandan94]. In advanced BiCMOS processes, therefore, outputs employing nMOS transistors will be the limiting factor in good ESD performance. The focus of BiCMOS ESD protection circuit design should be to protect the nMOS transistors used in the input or output buffers. The available protection circuit elements are the vertical *npn* transistor as well as bipolar SCR structures. The lateral CMOS SCR is made ineffective by the low-resistance buried n^+ diffusion at the bottom of the n-well used for the *npn* collector. It has been found that the bipolar SCR does not gain any ESD performance over the *npn* transistor [Amerasekera92]. The *npn* has the added advantage that it is easier to implement directly into a circuit model such as SPICE for evaluation of both circuit performance as well as ESD performance [Chatterjee91B].

To use the *npn* transistor as a primary protection element for nMOS transistors in the output buffers, it must be ensured that the *npn* triggers before the nMOS transistor that is being protected. It must also have a lower snapback holding voltage and on-resistance than the nMOS. Lowering the snapback trigger voltage is achieved using active elements as triggers as described in [Chatterjee91B][Amerasekera92]. In the event that the *npn* transistor characteristics do not meet the requirements for protecting the nMOS transistor, even with a trigger circuit, nMOS protection design needs to be implemented or the nMOS needs to be made self-protecting as described in previous sections.

6.11.3 Bipolar/BiCMOS Output Protection

The triggering and operation of a vertical *npn* transistor during an ESD event has been described in Chapter 4. It was pointed out that by forward-biasing the emitter–base junction of the *npn* the voltage required to fully turn-on, V_{t1}, the *npn* can be reduced. The circuit schematic shown in Figure 6.58 shows a vertical *npn* transistor $T1$ with its collector connected to the pad. $T2$ is a generic trigger element which turns on during an ESD event and allows current to flow though the resistor R. $T3$ is the nMOS device being protected. When the emitter–base voltage, V_{be}, is sufficiently high that the emitter current can substantially contribute to the avalanche process at the collector–base junction, the *npn* goes into a low

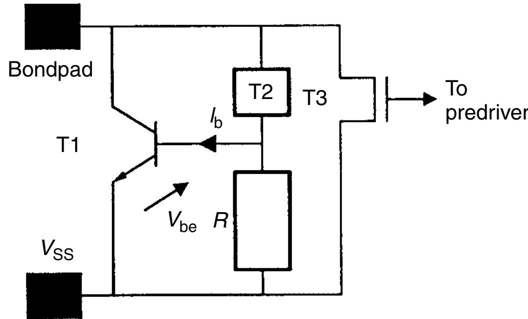


Figure 6.58 Bipolar/BiCMOS output protection scheme using active triggering [Chatterjee91B]

impedance mode shunting the ESD current (see Chapter 4). A rough value of $V_{be} \approx 0.8\,V$ will lower V_{t1} to significantly less than the trigger voltage of the nMOS. Based on this value of V_{be}, if $T3$ carries 1 mA then R needs to be about $1\,k\Omega$.

$T3$ can be an nMOS transistor, a reverse-biased diode or a lateral *pnp* transistor [Chatterjee91B][Amerasekera92][Corsi93]. A low breakdown voltage diode can be easily obtained in a bipolar process by using the emitter–base breakdown voltage of the *npn*, which is typically between 5 and 7 V. ESD levels of greater than 4 kV have been demonstrated in a submicron BiCMOS process using this technique [Chatterjee91B].

In a bipolar output buffer, $T1$ can be the *npn* pull-down in the output buffer itself. Assuming a conservative ESD level of $20\,V\,\mu m^{-1}$ for the *npn*, the minimum emitter length would be 200 µm for a 4-kV performance. If an *npn* is also connected between the pad and V_{dd} the ESD levels will be greatly increased. The ESD performance when the emitter (collector–base grounded) is stressed has not been found to be much less than that obtained with stress applied to the collector (emitter–base grounded). Conservatively, the ESD threshold for the emitter stress is approximately 80% of that for the collector stress.

When emitter-coupled-logic (ECL) buffers are used, the power supply connections are different, but similar designs can be used with adjustments made for the different power supplies (i.e., $V_{cc} = 0\,V$, $V_{ss} < 0\,V$).

6.11.4 Bipolar/BiCMOS Input Protection

Figure 6.59 shows a bipolar protection circuit schematic for protecting a bipolar or MOS input buffer. The circuit design follows the design synthesis described for input protection in Section 6.7 and consists of a primary protection and a secondary protection separated by an isolating resistor. $B1$ is the primary protection using the *npn* transistor. $B2$ and $B3$ provide secondary protection for clamping the voltage

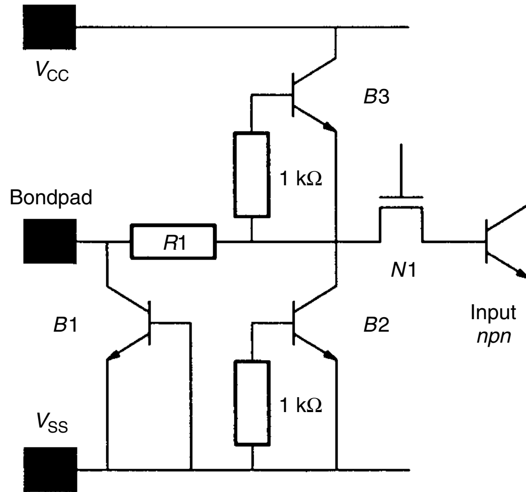


Figure 6.59 Bipolar/BiCMOS input protection scheme

at the input. These are effective for bipolar inputs, but MOS inputs require an FPD clamp $N1$ to protect against gate oxide breakdown. $B3$ provides a low-voltage clamp between the pad and V_{cc}. $R1$ is the isolation resistor. These protection circuits have been shown to be capable of ESD protection levels greater than 4 kV in submicron BiCMOS processes [Amerasekera92].

6.11.5 Layout

Bipolar transistors used in protection circuits must be laid out in straight lines using minimum emitter width. If a suitable base–emitter resistance is used, then the multiple emitter fingers may be used, each of minimum width and separated by the minimum spacing allowed. There is no effect of the contact spacing or the collector to emitter spacing in submicron bipolar processes, and no restrictions need to be placed on these parameters. The nMOS transistors and resistors must be laid out in accordance with the requirements discussed in Sections 6.3 and 6.7.

6.11.6 ESD and Performance Trade-Offs

One of the biggest problems confronting the ESD protection circuit designer in bipolar and BiCMOS circuitry is the requirement for high speed. The ideal protection circuit from the operating perspective will have zero capacitance and zero resistance. This can be achieved to some degree in self-protecting bipolar output buffers, but the need for a minimum length and an emitter–base resistance will affect the speed to some extent.

The areas of ESD protection circuits described here are close to the minimum possible in terms of the number of elements used and the efficiency of the circuit. The easiest solution is to trade-off ESD performance with speed. A 4-kV ESD performance is not always necessary, and the circuit can be fine-tuned to obtain the minimum acceptable ESD level by reducing the size of the protection circuit. As most of the capacitance comes from the vertical region of the collector–substrate junction, reducing the length of the collector and increasing the emitter area will reduce capacitance and improve ESD levels. The emitter area should be increased by adding emitter fingers of the minimum width to avoid emitter crowding effects associated with large area emitters at high current levels. However, this technique will only work if the extrinsic emitter–base resistance is high enough that turn-on of all the emitters is achieved, that is, the current flow is entirely vertical.

6.12 SUMMARY

In this chapter, several protection devices were individually described in detail. Each of these has a limited applicability that is dependent mainly on the process technology. For example, the thick field oxide device worked very well for technologies with feature sizes greater than 1 μm but has not performed as well in advanced processes.

A design synthesis has been shown whereby a total protection scheme can be designed by combining primary protection devices with the secondary protection devices. The SCR device can form a very robust primary protection device for inputs. When its trigger voltage is in the 40–50-V range the secondary protection design becomes critical. But, as shown in this chapter, if the trigger level can be reduced the protection efficiency improves. Both n^+-diffusion and n-well resistors can be used as the isolation elements. Some of the advantages and disadvantages of these resistors have been discussed here.

A protection circuit with low RC delay can also be designed with a polysilicon resistor. Although in the past polysilicon resistors have not been found to be very effective, the modified low trigger SCR described here can make this into an efficient protection scheme. Once a successful protection design is obtained with the polysilicon resistor, it has several advantages over the diffusion resistor, including the improvement in RC delay. Some of these features make it attractive for analog circuit or high-speed circuit applications.

The other concepts introduced include the gate-coupled and gate-driven nMOS protection devices. These can be effective protection devices but require optimization with SPICE modeling. Also, their implementation for I/O circuit applications requires consideration of interaction with the buffer devices.

ESD protection by simply using diodes is becoming an attractive option. An understanding of these diode elements, both n^+ diffusion diodes and p^+/n-well diodes that form parasitic *pnp* devices, were also discussed.

Although the focus is always on I/O protection, the protection for the power pins is equally important to eliminate internal circuit failures. The different options and the strategy of selecting the proper clamp were presented.

Design issues related to bipolar and BiCMOS protection circuits have also been discussed. These circuits use the vertical *npn* transistor as the primary protection device. In BiCMOS circuits, the trigger voltage of the *npn* transistor needs to be reduced to below that of the nMOS transistor for the protection circuit to be effective. Some schemes are discussed for reducing the trigger voltage using active trigger elements. The issues related to the trade-off between ESD performance and circuit speed, an important issue for high-speed circuits, was also briefly addressed.

BIBLIOGRAPHY

[Abderhalden91] J. Abderhalden, *Untersuchugen zur Optimierung von Schutzstrukturen gegen electrostatische Entladungen in integrierten CMOS-Schaltungen*, PhD Thesis, Eidgenossische Technische Hochschule (ETH), Zurich, 1991.

[Amerasekera90] A. Amerasekera, L. Roozendaal, J. Abderhalden, J. Bruines and L. Sevat, “An Analysis of Low Voltage ESD Damage in Advanced CMOS Processes”, in *Proc. 12th EOS/ESD Symposium*, pp. 143–149, 1990.

[Amerasekera92] A. Amerasekera and A. Chatterjee, “An Investigation of BiCMOS ESD Protection Circuit Elements and Applications in Submicron Technologies”, in *Proc. 14th EOS/ESD Symposium*, pp. 265–276, 1992.

[Amerasekera94] A. Amerasekera and C. Duvvury, “The Impact of Technology Scaling on the ESD Robustness and Protection Circuit Design”, in *Proc. 16th EOS/ESD Symposium*, pp. 237–245, 1994.

[Amerasekera95] A. Amerasekera, C. Duvvury, V. Reddy and M. Rodder, “Substrate Triggering and Silicide Effects on ESD Performance and Protection Circuit Design in Deep Submicron CMOS Processes”, in *Tech. Dig. IEDM*, 545, 1995.

[Anderson97] W.R. Anderson, “Circuit and Process Design Considerations for ESD Protection in Advanced CMOS Processes”, *Microelectronics Reliability*, **37**, pp. 1087–1103, 1997.

[Anderson98B] W.R. Anderson, J.J. Montanaro and N.J. Howorth, “Cross-Referenced ESD Protection for Power Supplies”, in *Proc. 20th EOS/ESD Symposium*, pp. 86–95, 1998.

[Avery83] L. Avery, “Using SCR’s as transient protection structures in integrated circuits”, in *Proc. 5th EOS/ESD Symposium*, pp. 177–180, 1983.

[Carbajal92] B. Carbajal, R. Cline and B. Andresen, “A Successful HBM ESD Protection Circuit for Micron and Sub-Micron Level CMOS”, in *Proc. 14th EOS/ESD Symposium*, pp. 234–242, 1992.

[Chatterjee91A] A. Chatterjee and T. Polgreen, “A low-voltage triggering SCR for on-chip protection at output and input pads”, in *Elec. Dev. Lett.*, **EDL-12**, 21–22, 1991.

[Chatterjee91B] A. Chatterjee, T. Polgreen and A. Amerasekera, “Design and simulation of a 4 kV ESD protection circuit for a 0.8 μm BiCMOS process”, in *Tech. Dig. IEDM*, 913–916, 1991.

[Chaine97] M. Chaine, S. Smith and A. Bui, "Unique ESD Failure Mechanisms During Negative to Vcc HBM Tests", in *Proc. 19th EOS/ESD Symposium*, pp. 346–355, 1997.

[Chen88] K-L. Chen, "Effect of interconnect process and snapback voltage on the ESD failure threshold of NMOS transistors", in *Proc. 10th EOS/ESD Symposium*, pp. 212–219, 1988.

[Chen97] J.Z. Chen, A. Amerasekera and C. Duvvury, "Design Methodology for Optimizing Gate Driven ESD Protection Circuits in Submicron CMOS Processes", in *Proc. 19th EOS/ESD Symposium*, pp. 230–239, 1997.

[Cook93] C. Cook and S. Daniel, "Characterization of New Failure Mechanisms Arising From Power-Pin ESD Stressing", in *Proc. 15th EOS/ESD Symposium*, pp. 149–156, 1993.

[Corsi93] M. Corsi, R. Nimmo and F. Fattori, "ESD Protection of BiCMOS Integrated Circuits Which Need to Operate in the Harsh Environments of Automotive or Industrial", in *Proc. 15th EOS/ESD Symposium*, pp. 209–214, 1993.

[Dabral93] S. Dabral and R. Aslett, "Designing On-Chip Power Supply Coupling Diodes for ESD Protection and Noise Immunity", in *Proc. 15th EOS/ESD Symposium*, pp. 239–250, 1993.

[Dabral94] S. Dabral, R. Aslett and T. Maloney, "Core Clamps For Low Voltage Technologies", in *Proc. 16th EOS/ESD Symposium*, pp. 141–148, 1994.

[DeChiaro86] L.F. De Chiaro, "Input ESD Protection Network for Fineline NMOS-Effects of Stressing Waveform and Circuit Layout", in *Proc. 24th IRPS*, pp. 206–214, 1986.

[Diaz94] C. Diaz and G. Motley, "Bi-Modal Triggering For LVSCR ESD Protection Devices", in *Proc. 16th EOS/ESD Symposium*, pp. 106–112, 1994.

[Duvvury83] C. Duvvury and R.N. Rountree and L.S. White, "A summary of most effective electrostatic protection circuits for MOS memories and their observed failure modes", in *Proc. 5th EOS/ESD Symposium*, pp. 181–184, 1983.

[Duvvury85] C. Duvvury, R. Rountree and D. Baglee, A. Hyslop, L. White, "ESD Design Considerations for ULSI", in *Proc. 7th EOS/ESD Symposium*, pp. 45–48, 1985.

[Duvvury86] C. Duvvury, R. McPhee, D. Baglee, and R. Rountree, "ESD Protection Reliability in 1-µm CMOS Circuit Performance", in *Proc. 24th IRPS*, pp. 199–208, 1986.

[Duvvury87] C. Duvvury, R.N. Rountree and R.A. McPhee,"ESD protection: design and layout issues for VLSI circuits", in *IEEE Trans. Ind. Appl.*, **IA-25**, 41–47, 1987.

[Duvvury88A] C. Duvvury, R. Rountree and O. Adams, "Internal chip ESD phenomena beyond the protection circuit", in *IEEE T. Electron Dev.*, **ED-35**, 2133–2139, 1988.

[Duvvury88B] C. Duvvury and R. Rountree, "Output ESD Protection Techniques for Advanced CMOS Processes", in *Proc. 10th EOS/ESD Symposium*, pp. 206–211, 1988.

[Duvvury89] C. Duvvury, T. Taylor, J. Lindgren and S. Kumar, "Input protection design for overall chip reliability", in *Proc. 11th EOS/ESD Symposium*, pp. 190–198, 1989.

[Duvvury90] C. Duvvury, "ESD reliability for advanced CMOS technologies", in *Int. Elec. and Dev. and Mat. Symp.*, 265–272, 1990.

[Duvvury92A] C. Duvvury and C. Diaz, "Dynamic Gate-Coupled NMOS for Efficient Output ESD Protection", in *Proc. 30th IRPS*, pp. 141–150, 1992.

[Duvvury92B] C. Duvvury, C. Diaz and T. Haddock, "Achieving Uniform nMOS Device Power Distribution for Sub-Micron ESD Reliability", in *Tech. Dig. IEDM*, 92, 1992.

[Duvvury96] C. Duvvury, A. Amerasekera, S. Ramaswamy and S. Young, "ESD Design for Deep Submicron SOI Technology", in *Tech. Digest Symposium on VLSI Technology*, 1996.

[Fukuda88] Y. Fukuda and K. Kato, "VLSI ESD Phenomenon and Protection", in *Proc. 10th EOS/ESD Symposium*, pp. 228–234, 1988.

[Hulett81] T.V. Hulett, "On-Chip Protection of NMOS Devices", in *Proc. 3rd EOS/ESD Symposium*, pp. 90–96, 1981.

[Keller81] J.K. Keller, "Protection of MOS Integrated Circuits from Destruction by Electrical Discharge", in *Proc. 3rd EOS/ESD Symposium*, pp. 73–79, 1981.

[Ker92] M. Ker, C. Wu and C. Lee, "A Novel CMOS ESD/EOS Protection Circuit with Full-SCR Structures", in *Proc. 14th EOS/ESD Symposium*, pp. 258–264, 1992.

[Krakauer94] D. Krakauer and K. Mistry, "Circuit Interactions During Electrostatic Discharge", in *Proc. 16th EOS/ESD Symposium*, pp. 113–119, 1994.

[LeBlanc91] J.P. LeBlanc and M.D. Chaine, "Proximity effects of unused output buffers on ESD performance", in *Proc. 29th IRPS*, pp. 327–330, 1991.

[Lin93] D.L. Lin, "ESD Sensitivity and VLSI Technology Trends: Thermal Breakdown and Dielectric Breakdown", in *Proc. 15th EOS/ESD Symposium*, pp. 73–82, 1993.

[Maene92] N. Maene, J. Vandenbroeck and L. van den Bempt, "On Electrostatic Discharge Protections for Inputs, Outputs and Supplies of CMOS Circuits", in *Proc. 14th EOS/ESD Symposium*, pp. 228–233, 1992.

[Maloney88] T. Maloney, "Designing MOS Inputs and Output to avoid Oxide Failures in the Charged Device Model", in *Proc. 10th EOS/ESD Symposium*, pp. 220–227, 1988.

[Maloney95] T.J. Maloney and S. Dabral, "Novel Clamp Circuits for IC Power Supply Protection", in *Proc. 17th EOS/ESD Symposium*, pp. 1–12, 1995. Revised version published in *IEEE Trans. on Components, Packaging, and Manufacturing Technology, Part C*, pp. 150–161, July 1996.

[McPhee86] R. McPhee, C. Duvvury, R. Rountree and H. Domingos, "Thick oxide ESD performance under process variations", in *Proc. 8th EOS/ESD Symposium*, pp. 173–179, 1986.

[Merrill93] R. Merrill and E. Issaq, "ESD Design Methodology" in *Proc. 15th EOS/ESD Symposium*, pp. 233–237, 1993.

[Oh01A] K-H. Oh, C. Duvvury, C. Salling, K. Banerjee and R. Dutton, "Nonuniform Bipolar Conduction in Single Finger NMOS Devices in a 0.13 μm CMOS Technology", in *Proc. IRPS*, pp. 226–234, 2001.

[Oh01B] K-H. Oh, C. Duvvury, K. Banerjee and R. Dutton, "Gate bias induced heating effect and implications for the design of deep submicron ESD protection", in *Tech. Dig. IEDM*, 2001.

[Palella85] A. Palella and H. Domingos, "A Design Methodology for ESD Protection Networks", in *Proc. 7th EOS/ESD Symposium*, pp. 169–174, 1985.

[Polgreen89] T. Polgreen and A. Chatterjee, "Improving the ESD failure threshold of silicided nMOS output transistors by ensuring uniform current flow", in *Proc. 11th EOS/ESD Symposium*, pp. 167–174, 1989.

[Polgreen90] T. Polgreen, A. Chatterjee and R. Cline, *"A Generalized ESD Protection Scheme for Use At All I/O Pads in CMOS Technologies"*, TI Technical Report No. 089044, 1990.

[Richier97] C. Richier, N. Maene, G. Mabboux, R. Bellens, "Study of the ESD Behavior of Different Clamp Configurations in a 0.35 µm CMOS Technology", in *Proc. 19th EOS/ESD Symposium*, pp. 240–245, 1997.

[Rieck89] G. Rieck and R. Manely, "Novel ESD protection for advanced CMOS output drivers", in *Proc. 11th EOS/ESD Symposium*, pp. 182–189, 1989.

[Rountree85] R.N. Rountree and C.L. Hutchins, "NMOS protection circuitry", in *IEEE T. Electron Dev.*, **ED-32,** 910–917, 1985.

[Rountree88] R. Rountree, C. Duvvury, T. Maki, H. Stiegler, "A process-tolerant input protection circuit for advanced CMOS processes", in *Proc. 10th EOS/ESD Symposium*, pp. 201–211, 1988.

[Scott86] D. Scott, G. Giles and J. Hall, "A lumped element model for simulation of ESD failures in silicided devices", in *Proc. 8th EOS/ESD Symposium*, pp. 41–47, 1986.

[Sematech98] SEMATECH Technology Transfer Document, *"Test Structures for Benchmarking the Electrostatic Discharge (ESD) Robustness of CMOS Technologies"*, *SEMATECH TT 98013452A-TR*, May, 1998.

[Tandan94] N. Tandan, G. Conner, "ESD Trigger Circuit", in *Proc. 16th EOS/ESD Symposium*, pp. 120–124, 1994.

[Voldman92] S. Voldman, V. Gross, M.J. Hargrove, J.M. Never, J.A. Slinkman, M.P. O'Boyle, T.S. Scott and J.J. Delecki, "Shallow Trench Isolation Double-Diode Electrostatic Discharge Circuit and Interaction with DRAM Output Circuitry", in *Proc. 14th EOS/ESD Symposium*, pp. 277–288, 1992.

[Voldman99] S. Voldman, W. Anderson, R. Ashton, M. Chaine, C. Duvvury, T. Maloney and E. Worley, "A Strategy for Characterization and Evaluation of ESD Robustness of CMOS Semiconductor Technologies", in *Proc. 21st EOS/ESD Symposium*, pp. 212–224, 1999.

[Weston92] H. Weston, V. Lee and T. Stanik, "A Newly Observed High Frequency Effect On The ESD Protection Utilized In a Gigahertz NMOS Technology", in *Proc. 14th EOS/ESD Symposium*, pp. 95–98, 1992.

[Wilson87] D. Wilson, H. Domingos and M.M.S. Hassan, "Electrical Overstress in nMOS Silicided Devices", in *Proc. 9th EOS/ESD Symposium*, pp. 265–273, 1987.

[Worley95] E. Worley, R. Bupta, B. Jones, R. Kjar, C. Nguyen and M. Tennyson, "Sub-Micron Chip ESD Protection Schemes which Avoid Avalanching Junctions", in *Proc. 17th EOS/ESD Symposium*, pp. 13–20, 1995.

7 Advanced Protection Design

Charvaka Duvvury, Warren Anderson

7.1 INTRODUCTION

In the previous chapter, the design and layout of ESD protection devices were discussed. The basic protection devices included the field oxide *npn*, gate-coupled and gate-driven nMOS devices, and SCRs. While these devices have served well to form efficient protection designs, the advances in both CMOS and BiCMOS technologies towards the deep submicron technologies rendered them to be less robust. In addition to the process technology scaling advances, the demand from the I/O chip designers for smaller area to fit in the given I/O pad pitch, for ESD that tolerates voltages on the I/O greater than the core supply, and for reduced effective input capacitance to meet the high-speed requirements have steadily increased. To meet all of these criteria, newer protection concepts have been introduced during the mid to late 1990s. These new concepts and their implementation in the ESD design are discussed in this chapter.

7.2 PNP-DRIVEN nMOS (PDNMOS)

With low resistivity substrates, the gate-coupled or gate-driven techniques have become less efficient. This has led to the investigation of newer concepts that mainly involve substrate triggering. Some of these are reviewed in this section.

The first device concept is known as PNP-driven nMOS or PDNMOS [Chen97]. The schematic is shown in Figure 7.1.

The PDNMOS is more like a gate-driven nMOS but uses the lateral PNP of the pMOS transistor to drive the gate higher during ESD. That is, the PNP in Figure 7.1 is achieved by using a pMOS transistor with its gate and *n*-well connected to $V_{\rm dd}$, the source to the pad, and the drain to the nMOS protection device gate. With an ESD event, the pMOS either conducts current or the PNP goes into breakdown to pull the nMOS gate up with current driven through its gate resistor.

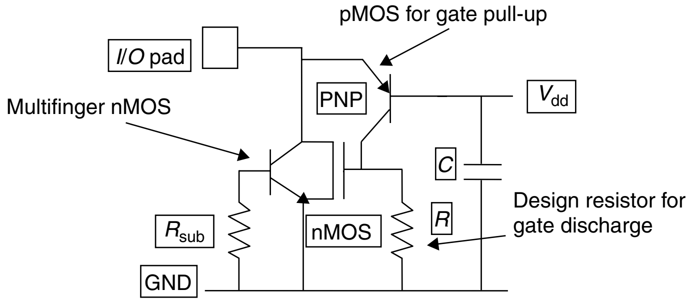


Figure 7.1 Schematic of PNP-driven nMOS or PDNMOS. Reproduced by permission of ESD Association

The size of the pMOS and the nMOS gate resistor can be designed using SPICE simulations. Note that initially the current goes through the lateral diode of the pMOS charging the V_{dd} capacitance and the PNP eventually turns on. Therefore, during the initial stages there will be some vertical PNP current that can charge the substrate. A p^+ guardring can be placed between the pMOS and the nMOS to prevent latchup. Although the vertical PNP effect is not utilized in this concept, it can form as a beneficial effect as will be discussed for some of the other advanced concepts. The advantage of the PDNMOS is that it has been found to be robust even for technologies with low substrate resistance and with silicided diffusions. This device has also been found to work effectively even when the V_{dd} capacitance is low because the vertical PNP effect is not the main mechanism. One distinct disadvantage of this device is that it cannot be used for I/O applications in which fail-safe[1] operation is required. Therefore, the PDNMOS and any other design in which a diode is present between the signal and power supply are known as non-fail-safe designs.

7.3 SUBSTRATE TRIGGERED nMOS (STNMOS)

Another non-fail-safe design that is even more efficient than the PDNMOS is the substrate triggered nMOS or the STNMOS [Amerasekera94]. In this device concept, the vertical PNP current is used to trigger the substrate of the protection nMOS. When a diode is built with diffusions in the n-well, the parasitic vertical PNP can form an important protection device element (see Section 6.9.2).

Figure 7.2 shows the cross section of a lateral diode to V_{dd}, which is simply formed by placing a P^+ diffusion connected to the pad and an n^+ diffusion connected to V_{dd} or the power supply.

[1] Fail-safe is a design feature that allows higher voltages at the signal pin, whereas non-fail-safe limits the signal pin to less than one diode drop above V_{dd}.

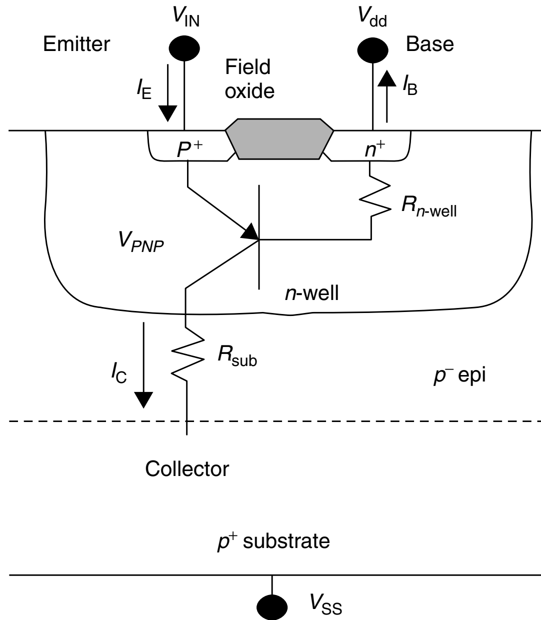


Figure 7.2 Cross section of a lateral diode forming vertical PNP. (After [Amerasekera95], reproduced by permission of ©1995 IEEE)

The spacing between the two diffusions is kept at minimum for the process. Note that this lateral diode is more effective than the parasitic diode that is present in a pMOS transistor since in the latter the channel region effect reduces the diode efficiency. Although the cross section of Figure 7.2 indicates a field oxide between the two diffusions, the same can be implemented with the latest CMOS technologies in which the two diffusions are isolated by shallow trench isolation or STI [Bryan94]. The vertical PNP indicated in the figure plays a key role in pumping the substrate. For most of the technologies 80% of the emitter current goes through the lateral diode and 20% through the vertical PNP. The charging current depends on the β of the transistor and the efficiency in charging the substrate depends on the V_{dd} capacitance. In turn, the substrate potential will depend on the substrate resistance and the placement of the latchup guardrings that consist of P^+ connected to V_{ss}. More details of the PNP modeling are given in the chapter on simulations.

The STNMOS implementation is shown in Figure 7.3 for inputs and outputs. As shown in the Figure 7.3 schematic, the same protection device is used for both without any need for any isolation resistor. Of course, the robustness of this device depends on the layout, which is shown in Figure 7.4. The placement of the nMOS near the diode is critical to taking advantage of the substrate bias from the vertical PNP of the diode. The spacing d shown in Figure 7.4 can determine the parasitic latchup and is determined from test structure evaluation. The methods for effective layout was discussed by Li *et al.* [Li97]. The $I-V$ curves with the

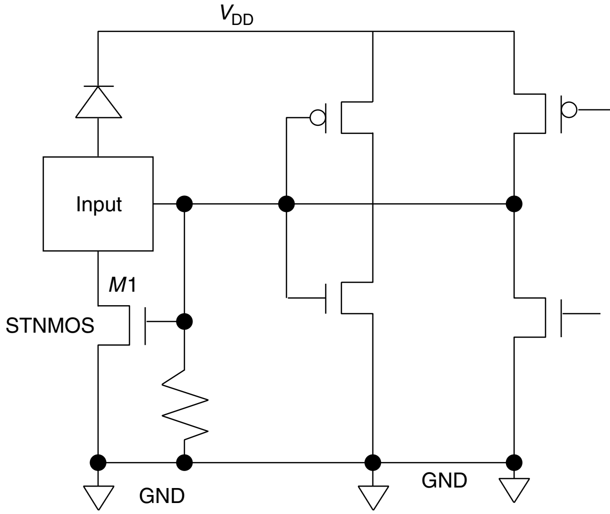


Figure 7.3 The STNMOS used as input/output protection

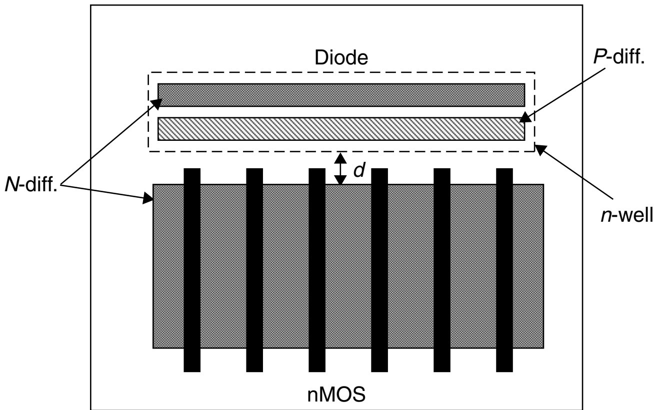


Figure 7.4 The STNMOS layout showing the placement of the diode and the protection nMOS. Note that the spacing *d* is critical to avoid latchup

STNMOS protection are illustrated in Figure 7.5 [Ramaswamy96A]. First, for I/O pin stress positive to V_{dd} the diode provides the clamp. The efficiency of this clamping will depend on the perimeter of the diode that is designed. When the I/O is stressed positive to V_{ss}, the diode to V_{dd} again turns on since V_{dd} would be floating. This charges the V_{dd} capacitance as the vertical PNP pumps current to trigger M1 of Figure 7.4. Notice that M1, which is a multifinger nMOS, clamps the voltage without showing any snapback indicating uniform trigger.

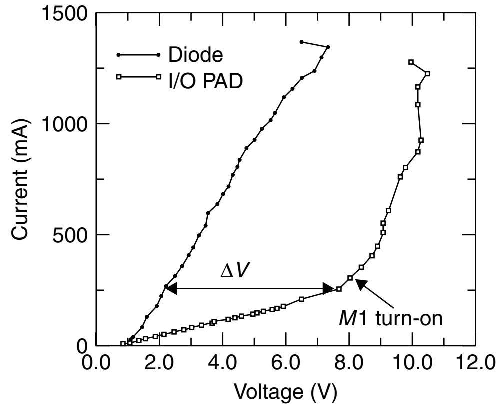


Figure 7.5 The $I-V$ characteristics with the STNMOS protection in Figure 7.4 [Ramaswamy96A]. The dark line is for I/O to V_{dd} stress and the light line is for I/O to V_{ss} stress. (After [Ramaswamy96], reproduced by permission of ESD Association)

7.4 nMOS TRIGGERED nMOS (NTNMOS)

While the STNMOS is a robust protection device with the silicided diffusions, like the PDNMOS, it has its limitations for all applications. For example, the protection device performance to some extent can depend on the V_{dd} capacitance as for smaller chips the lower capacitance would limit the amount of substrate current that could be pumped. Also, it requires a diode path to V_{dd}, which is not always allowed for what is known as 'Fail-safe' applications. Fail-safe, as defined previously, is the requirement for the I/O signal pin to go high when the power pin is low. Furthermore, neither the STNMOS nor the PDNMOS can be useful protection devices for the power pin protection.

The aforementioned limitations can be overcome with the implementation of an nMOS pump that does not require the V_{dd} capacitance effect and is shown in Figure 7.6. It employs the gate-coupled concept where the gate of the protection nMOS and the pump nMOS are coupled. The pump nMOS source is not connected to V_{SS} (GND) but instead is connected directly to a floating substrate P^+ ring that surrounds the protection nMOS. As the pumping is done by an nMOS, this device is called the NTNMOS or the nMOS pumped nMOS.

The layout of the NTNMOS is shown in Figure 7.7 in which the nMOS pump is integrated into the protection nMOS. There are several resistance values that need to be considered for this device to work effectively. First, R1 is the resistance from the body of the protection device to the floating P^+ ring. Next, R2 is the substrate resistance from the floating ring to the nearest latchup P^+ grounded guardring. Finally, RB is substrate resistance from the floating ring to substrate ground. Figure 7.8 shows the cross section of the effective resistances and Figure 7.9 shows the equivalent circuit schematic.

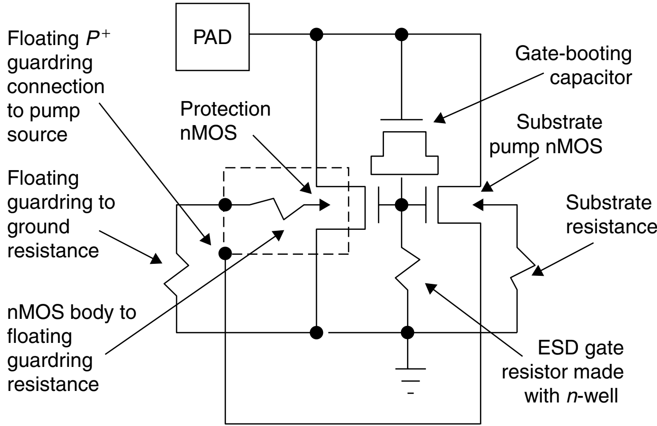


Figure 7.6 The NTNMOS schematic showing the substrate pump nMOS coupled to the same gate as the protection nMOS. Note that the source of the pump device not connected directly to GND but instead to a floating P^+ ring to provide the substrate bias. (After [Duvvury00B], reproduced by permission of ESD Association)

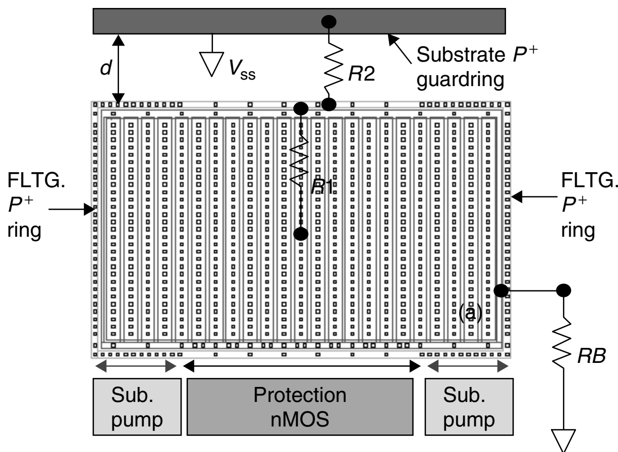


Figure 7.7 The NTNMOS layout showing the integrated pump nMOS into the total protection device. (After [Duvvury00B], reproduced by permission of ESD Association)

To optimize the NTNMOS design, a simple SPICE simulation can be done by ramping the pad from 0 to V_{t1} in 300 ps to represent the ESD event. The simulation results shown in Figure 7.10 represent a more detailed simulation using substrate resistance modeling ESD high current SPICE (see Chapter 11). Notice from Figure 7.10(a) first that the pump current IP and the device substrate current IB split into IR2 and IRB. The pump current IP and substrate current IB add up to effectively bias the substrate to 0.7 V as shown in Figure 7.10(a). The PAD voltage

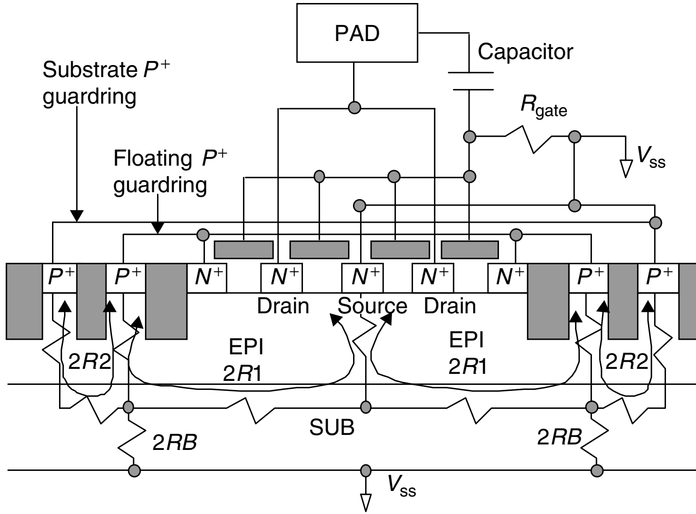


Figure 7.8 The NTNMOS cross section showing the effective resistance values as determined from the floating substrate ring and the latchup guardring to V_{SS}. (After [Duvvury00B], reproduced by permission of ESD Association)

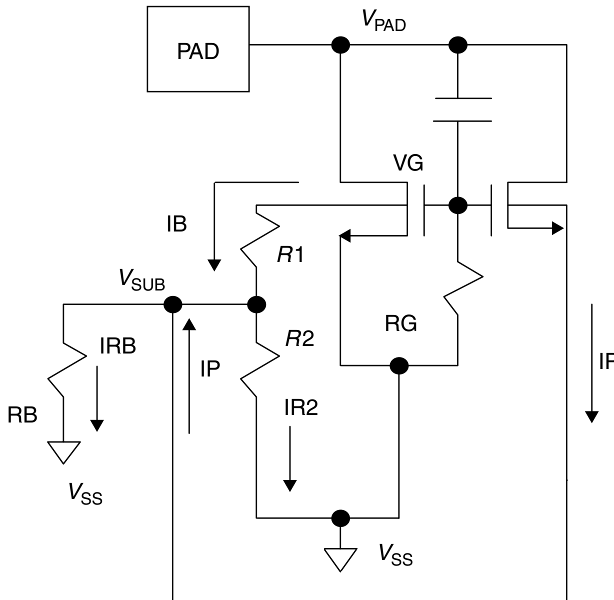
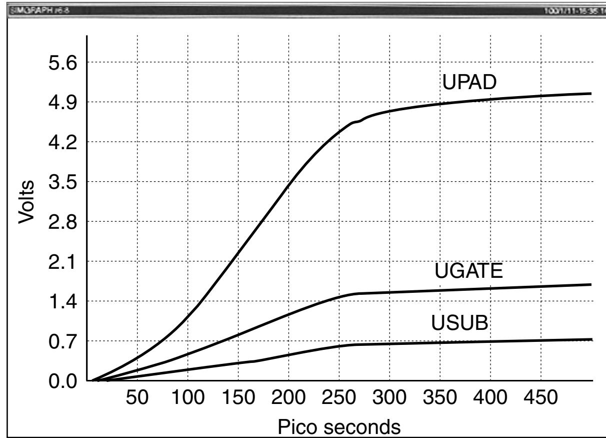
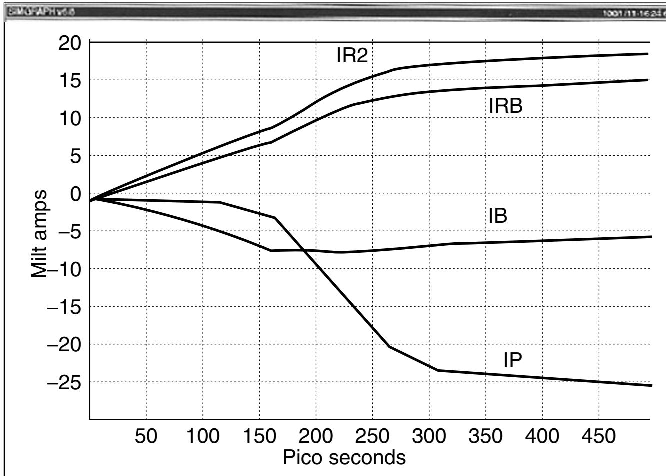


Figure 7.9 The NTNMOS simulation schematic showing the values corresponding to the resistances indicated in Figures 7.7 and 7.8. (After [Duvvury00B], reproduced by permission of ESD Association)



(a)



(b)

Figure 7.10 SPICE simulation results for the NTNMOS with current waveforms in (a), and the voltage waveforms in (b). Note the clamping at VPAD of 5 V in less than 300 ps indicating uniform trigger with snapback. (After [Duvvury00B], reproduced by permission of ESD Association)

clamps at 5 V with no snapback indicating a smooth transition into V_{t1}. Thus, with the substrate at 0.7 V the entire multifinger nMOS would uniformly turn on even for a silicided process.

The implementation of the substrate-biased protection devices seems to offer an additional benefit that has not been realized before. For the very advanced technologies with silicided diffusions and low-resistance substrate, the I_{t2} of the single finger

has been observed to be a sensitive function of the width. Although the I_{t2} degradation with finger width was observed several years ago [Scott86][Polgreen89], its impact on the protection device design was not a critical factor. The I_{t2} degradation is related to current-crowding effects that allows only a portion of the finger to conduct the ESD current. This means that the effective I_{t2} is reduced. The reason for this nonuniform conduction could be related to the intrinsic process defects that lock in the conduction to a local area. Detailed thermal modeling has to be done to clearly understand this effect. However, one interesting result is the effect of substrate bias that seems to reduce the localized conduction to make the design more practical as discussed in the following text.

The width dependence for a 0.13-μm technology has been recently reported to be rather severe as shown in [Oh01A]. From this work by Oh *et al.*, the design implications of this are shown in Figure 7.11. Here, the effective finger width versus the design finger width is plotted. For example, with zero substrate bias a 10-μm finger would effectively conduct only half (at 5 μm) as indicated by the deviation from the ideal line. Thus, using 20- or 30-μm fingers for an nMOS protection device will result in reduced ESD level. As shown in the figure, what is more important is the effect of the substrate bias. Note that with 0.7-V bias the full 10-μm finger is effective and with 1-V bias a 25-μm finger is fully effective. Therefore, a substrate-biased protection device such as the STNMOS or the NTNMOS cannot only take advantage of the multiple finger turn on but at the same time can also enhance the I_{t2} of the individual finger and this is critical for successful design of a substrate biased protection concept in the deep submicron technologies. It should also be pointed out that the effective width is a function of the process parameters such as silicided or nonsilicided, or the effective substrate resistance [Oh01A]. Since the effective I_{t2} is a function of both substrate bias and the finger width it

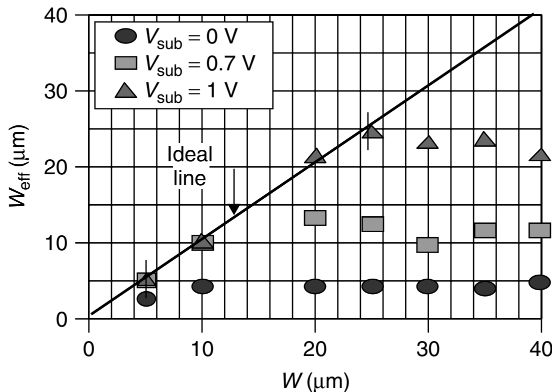


Figure 7.11 Illustration of nonuniform conduction within a single finger of an advanced sub-0.25-μm transistor with silicided diffusions and low-resistance substrate after [Oh01A]. Note that the effective width increases from 5 μm at zero bias to nearly 25 μm at 1 V bias. (After [Oh01], reproduced by permission of ©2001 IEEE)

is now proposed that during process characterization the I_{t2} should be measured at the same conditions as the protection device itself. That is, at 0 V bias for the normal protection concepts, such as the GGNMOS or GCNMOS with the substrate tied to the source, and for the substrate biased protection devices with a substrate bias applied [Salling01].

It was previously discussed in Figure 6.24 that I_{t2} degrades with gate bias. Therefore, some control on the gate potential is essential. An example circuit is discussed by Chen [Chen97] in which a control transistor with its gate tied to V_{dd} is placed in parallel to the gate resistor as shown in Figure 7.12. The pMOS between the PAD and the gate provides the lateral PNP current to bias the gate, and the nMOS across the resistor limits the potential of the protection nMOS. The same concept would apply if the pMOS were replaced with a booting capacitor as in Figure 6.25.

There is another more critical consideration, which is the effective input capacitance for an I/O pin that will influence the normal I/O circuit speed. This would be an issue for any gate-coupled protection device as in Figure 6.25 or for the substrate triggered nMOS as in Figure 7.6. A circuit technique to overcome this effect would be to short out the gate resistor with the V_{dd} powered up as shown in Figure 7.13 [Duvvury00A]. During ESD the gate resistor Rg allows the gate to be coupled but during normal circuit operation with V_{dd} on the resistor is essentially shorted. Although this technique is effective, a problem could occur if V_{dd} is charged up during an ESD pulse. The power up of V_{dd} during ESD can commonly occur through the lateral diode to V_{dd} of the output pMOS or the intentional protection diode to V_{dd} as in the STNMOS (see Figure 7.3). Even if the pMOS back-gate is blocked for an output or if the NTNMOS is used for the protection of an input, the V_{dd} charging could occur because of antenna effects

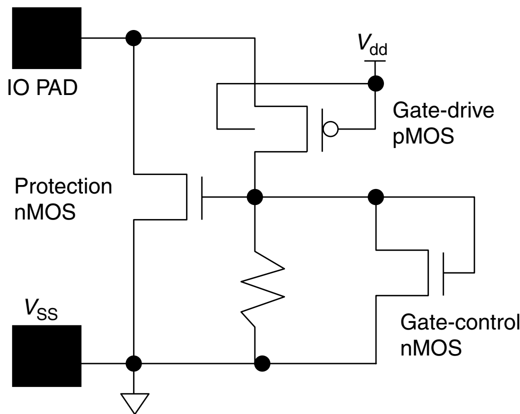


Figure 7.12 Gate control for a gate-coupled protection scheme after Chen *et al.* [Chen97]. The pMOS provides current to drive the gate during ESD and the gate control nMOS limits this gate potential

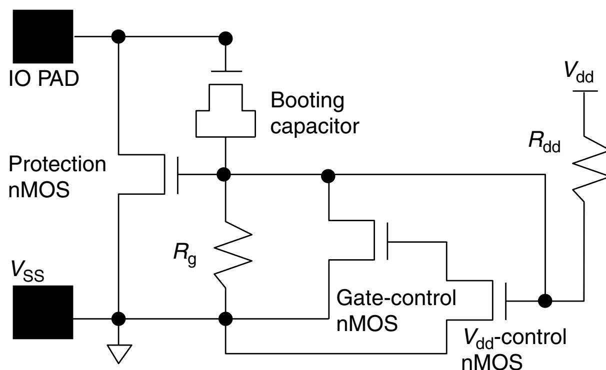


Figure 7.13 Gate control and V_{dd} control for any gate-coupled protection device concept. Note that when V_{dd} is powered up the Gate-control device essentially shorts the gate resistor for lower input capacitance, and during sudden V_{dd} charging the V_{dd}-control nMOS and Rdd provide a delay such that the ESD protection is not disabled

[Anderson98B]. Therefore, an additional control device is needed to ensure that any unintended V_{dd} charging would not disable the protection device. The V_{dd}-control nMOS shown in Figure 7.13 provides this feature in which the V_{dd} potential is delayed through R_{dd} and the V_{dd}-control nMOS. The typical value for R_{dd} is about 15 kΩ. Therefore, this discussion illustrates how both the circuit function and the ESD performance should be simultaneously considered for a compatible protection design.

There are a few other options that should be considered when implementing a protection device into an I/O buffer. During ESD the gate of the protection nMOS can be controlled as discussed in the preceding text. But the gate of the output nMOS must also be considered as the charging up of V_{dd} can influence its gate. This is controlled by the predrivers as shown in Figure 7.14. During the ESD event, when V_{dd} is charged up the gate of the output nMOS could follow, which means that the I_{t2} of the output nMOS is reduced. This could be a problem if the output nMOS is relatively small, such as in a 2-mA buffer, and an even larger consideration if the output ground is separate from the protection device ground as shown in Figure 7.14. For noise considerations, the output nMOS is in some cases connected to a different ground bus called V_{SSO} in the figure. For full ESD tests the I/O pad must be tested with respect to both grounds. As a common practice, diodes are placed between the two grounds as will be discussed later. When the I/O pin is tested with V_{SSO} grounded relatively larger current would go through the output nMOS making it more vulnerable. The same problem could also occur when testing for the CDM [Beebe98]. In such cases, a series resistor between the protection nMOS and the output nMOS must be considered. This need not be a large resistor. As an example, a typical value is 1 kΩ-µm or 10 Ω for a 100-µm wide nMOS buffer device.

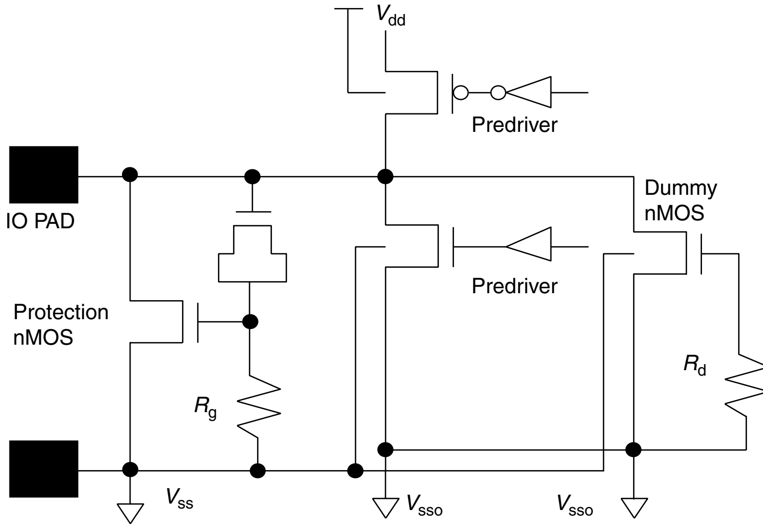


Figure 7.14 Protection nMOS along with the output buffer. Note that the pull-down nMOS ground is connected to V_{SSO}. A Dummy nMOS is added to boost the effective device width of the output buffer

If series resistors cannot be implemented another approach would be to make the output nMOS larger than necessary for circuit operation and tie off the portion not needed through a gate resistor to ground. This is also shown in Figure 7.14 in which the 'Dummy nMOS' approach is used with a gate resistor Rd. When using this technique, a simple SPICE simulation can be done to determine the value of Rd to ensure that the gate coupling on the active output nMOS matches the gate coupling on the Dummy nMOS. The full predriver circuit must be included to ensure that the circuit is correctly represented [Mergens99]. Usually a value less than 1 kΩ for Rd meets this dummy output requirement. The same dummy concept can also be used for a pMOS output [Ting01].

For PAD to V_{dd} stress the lateral diode offers the necessary protection. However, when the V_{dd} pin is stressed with respect to the I/O pad, some sneak current can go through the pull-up pMOS in Figure 7.14 [Ting01]. If the buffer device is small, the PNP of the pMOS can fail in the very advanced technologies. In this case also, in lieu of placing a series resistor between the I/O pad and the output device, a dummy pMOS can be added between with its gate resistively tied to V_{dd}.

The protection device performance and the I/O buffers should be carefully considered when implementing the overall scheme. If a nonsilicided process is used or if a silicide block is used for the protection device, the output nMOS must also follow the same ESD rules as the protection device. That is, the same drain contact-to-gate spacing for both devices. Otherwise, the output nMOS becomes relatively weak unless a series resistor is used.

7.5 ESD FOR MIXED-VOLTAGE I/O

In order to follow constant field scaling requirements, the power supply voltage has decreased as most integrated circuit technologies advance over time. At and before the 1.0-μm node, most CMOS technologies operated with a 5-V supply. Most 0.75- and 0.5-μm processes operated at 3.3 V; 0.35 and 0.25 μm at 2.5 V; 0.18 μm at 1.5 V; and 0.13 μm at 1.2 V. However, integration at the system level often requires I/O communication between parts with different core power supply voltages. Often, for backward compatibility, the I/O must tolerate or drive a voltage that is higher than the technology allows.

Without special design enhancements, mixing voltages on a bidirectional interface using standard driver and receiver circuits, like those in Figure 7.15, can cause several undesired effects. This is because, in general, the power supply voltage rating not only specifies the operating condition of the part, but it also indicates the voltage tolerance for reliable operation at the transistor level: the drain-source voltage limit for hot carrier reliability and the gate-drain or gate-source voltage limit for gate oxide reliability. When the high-voltage chip drives V_{ddhigh} into the low-voltage chip, the input receiver sees excessive gate-source voltage. Excessive voltage also appears across the drain-source of the tri-stated output driver nMOS and the grounded-gate nMOS ESD device. The higher input voltage also forward biases the drain-well junction of the tri-stated output driver pMOS and the ESD diode, as well as turning on the pMOS channel itself.

Therefore, the usual practice is to specify uniform I/O voltage standards that must be observed at both ends of the I/O interface. If the signaling standard requires translating bits into voltages higher or lower than the core power supply, this must be accomplished at the interface between the core logic and the I/O. Many

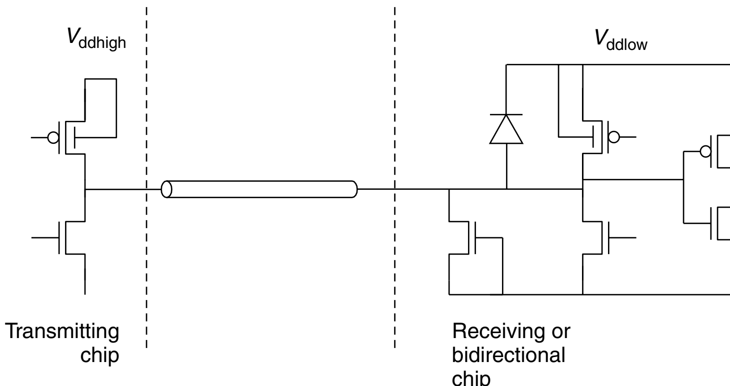


Figure 7.15 A dysfunctional mixed-voltage interface across a bidirectional bus where an IC with a high supply voltage V_{ddhigh} drives signals into an IC with a lower voltage supply V_{ddlow} and vice versa. Several reliability and functionality problems arise in this design, which uses both V_{dd}- and V_{ss}-based ESD protection

circuit solutions can overcome the reliability limitations if the I/O interface voltage is higher than that of the core supply by dividing the higher I/O voltage across multiple transistors.

ESD protection, likewise, must be compatible with the mixed-voltage environment. A number of device and circuit techniques offer ESD solutions. The following sections enumerate each of these, concluding with a guide to selecting the most appropriate ESD solution for different mixed-voltage I/O cases.

7.5.1 The Field Oxide Device

The field oxide device, discussed in Section 6.2, is generally compatible with mixed-voltage I/O, although it is generally not a robust or effective device in today's processes. The field oxide can withstand the high I/O voltage across its gate without degradation. The drain and source diffusions must be separated by a distance larger than the minimum channel length to avoid leakage. A larger FOD channel length increases the snapback holding voltage V_{sp}, one reason that the FOD has diminished effectiveness. Likewise, V_{sp} will increase in processes with shallow trench isolation (STI). Generally, the FOD requires silicide blocking to achieve uniform triggering and reasonable I_{t2} per unit width. Connecting the gate to the I/O pad may lower its V_{t1} and improve triggering uniformity as well.

7.5.2 The Dual Gate Oxide Transistor

Many processes now contain a thicker oxide version of the standard transistor to mitigate I/O voltage tolerance issues. Such a transistor can function well in snapback. As is the case with a standard transistor, the dual gate oxide transistor requires either silicide block for the GGNMOS configuration (see Section 6.3) or gate modulation (see Sections 6.4 and 6.5) for uniform triggering with a fully silicided device. Drain junction engineering may be needed to achieve a reasonable second breakdown current I_{t2}. The dual gate oxide transistor may also be stacked, in a similar way as the standard nMOS described in Section 7.5.3 to tolerate an even higher voltage.

7.5.3 The Stacked nMOS

A common method for tolerating a high voltage across the output driver nMOS pull-down is to form an nMOS stack in a cascode configuration, as shown in Figure 7.16(a). When constructed properly, this device also acts as a lateral *npn* during ESD [Anderson98A][Krakauer98][Amerasekera99]. Figure 7.16(b) shows the normal operation node voltages in the nMOS stack using an example I/O pad voltage of 5 V into an IC with $V_{dd} = 3.3$ V. Independent control of the top and bottom gates allows this device to meet reliability limits during normal circuit operation. The top drain connects to the pad, the bottom source to V_{ss}. The top gate

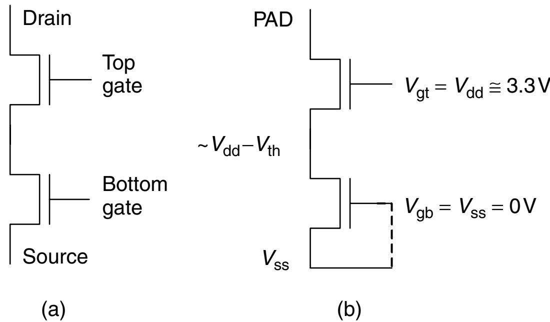


Figure 7.16 Schematic diagram of the nMOS stack with (a) node labels and (b) node voltages for 5 V on the I/O pad with a 3.3-V power supply. Reproduced by permission of the ESD Association

V_{gt} is biased at or near V_{dd} (3.3 V), either directly or through a transistor network. With the pad at 5 V, the center diffusion floats to approximately a threshold drop below V_{dd}, about 2.8 V for $V_{th} = 0.5$ V. To avoid leakage through the structure, the bottom gate V_{gb} must be at V_{ss}, either directly or through a transistor network. None of the voltages across the drain-source, gate-source, and gate-drain terminals of either transistor exceeds 3.3 V. Therefore, the stack operates within dielectric and hot carrier reliability limits.

Under ESD conditions, the stacked device operates in snapback, with the bipolar effect taking place between any combination of the three parasitic *npn* devices shown in Figure 7.17. For minimum spacing between the two poly gates as shown in Figure 7.17(a), the single *npn* between the source of the bottom nMOS (emitter) and drain of the top (collector) dominates. The spacing between these diffusions determines the base width. The center diffusion floats, not participating significantly in the snapback process. Conduction through the series pair of *npn* devices is also

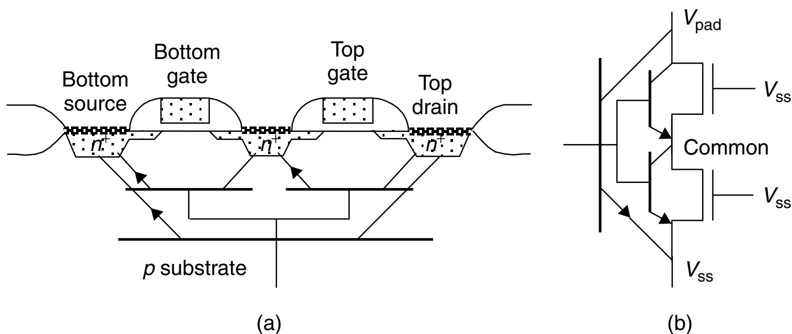


Figure 7.17 Cross section (a) and schematic diagram (b) of the parasitic *npn* devices found within the nMOS stack (After [Anderson98A], reproduced by permission of the ESD Association)

possible, but much less likely for small poly length and separation. Activation of both devices in series requires a collector-emitter breakdown voltage across each, for a breakdown voltage at the top drain of approximately twice that of the single *npn*. For a sufficiently wide spacing, however, avalanching at the collector-base breakdown junction can no longer de-bias the emitter junction of the bottom device and only series conduction is possible [Miller00].

Figure 7.18 shows the TLP $I-V$ characteristics for the nMOS stack in both silicide-blocked and LDD silicided 0.5-μm technologies. The poly length is drawn

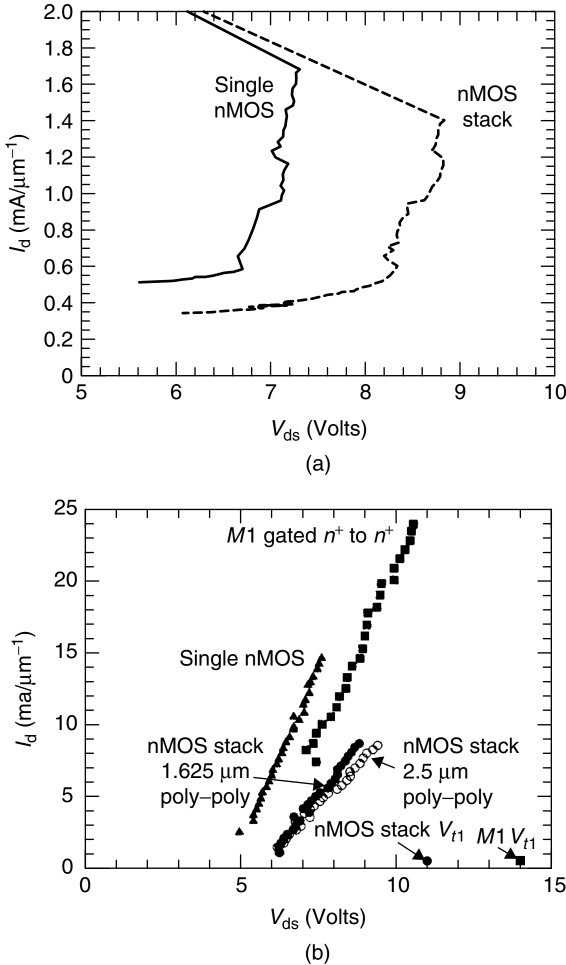


Figure 7.18 The TLP characteristics of the nMOS stack in two 0.5-μm processes (a) with silicide block, and (b) with full silicide. A single nMOS and a metal 1 field oxide device are shown for comparison (After [Anderson98A], reproduced by permission of the ESD Association)

at 0.75 µm for the silicide-blocked and 0.5 µm for the fully silicided devices, with a minimum poly–poly space of 0.625 µm in the stacked device. Compared to the single nMOS, the holding voltage for the stacked nMOS is 1.5 V higher, a result of its approximately 1-µm wider base region. The higher snapback holding voltage of the stacked nMOS also degrades its I_{t2} because the power dissipation is greater. The I_{t2} degradation is consistent with a similar power to failure for both the single and stacked nMOS.

7.5.3.1 nMOS stack failure characteristics

The nMOS stack has two failure modes. In the first failure mode, localized heating near the top drain junction causes current filamentation to occur between the top drain and the bottom source, analogous to second breakdown in a single nMOS. The top drain shorts directly to the bottom source. In the second failure mode, the nMOS stack breaks down between the top drain and the center diffusion with the bottom device still intact. Additional pulses are required to short the bottom device.

The partial failure mode has ramifications for circuit reliability after ESD. As the top gate must be biased around V_{dd} during circuit operational mode, the top device in the stack is on, acting only to limit the voltage from the center diffusion to the bottom gate to reduce dielectric stress. When a short occurs across the top device in the stack and the bottom gate is grounded, the full pad voltage can appear across the bottom device, exceeding dielectric stress limits. Post-ESD testing should ideally contain an I/O shorts test with the top gate held low and the bottom gate high and vice versa. In practice, though, this can be difficult to achieve.

7.5.3.2 The nMOS stack ESD in a silicide-blocked process

ESD-process interactions between fully silicided and silicide-blocked technologies create different ESD design requirements for the nMOS stack in each application. In a silicide-blocked process, I/O protection using the nMOS stack, shown in Figure 7.19, is nearly analogous to the case the of grounded-gate nMOS, but with several added features. In order to divide the high pad voltage across the two devices during normal operation, the top gate connects directly to V_{dd}. Likewise, the bottom gate connects to V_{ss} to keep the stack off. During ESD, the on-chip bypass capacitance on V_{dd} couples the top gate closely to V_{ss}. As with the single nMOS, the ballasting resistance of the nonsilicided drain extensions establishes uniform snapback across the entire device. To achieve uniform current flow through all fingers, layout of the stacked ESD structure should be balanced and symmetrical.

The output driver or any other nMOS device with a channel-connected path between the pad and a power rail is most susceptible to ESD damage. These devices are usually fully silicided, often with a channel length different from the ESD device, causing the $I-V$ characteristics to differ perhaps substantially between the two paths to V_{ss}. Effectiveness requires limiting the voltage across the output driver

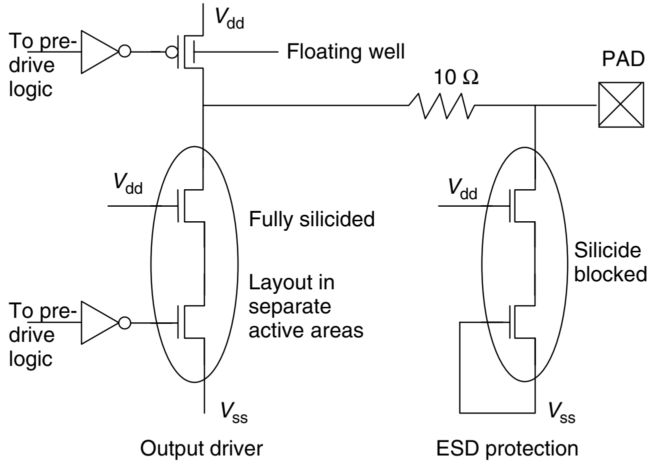


Figure 7.19 I/O ESD protection using the nMOS stack in a silicide-blocked process (After [Anderson98A], reproduced by permission of the ESD Association)

path, which can be accomplished through two means shown in Figure 7.19. First, the 10 Ω series resistor increases the voltage in the output driver path if it begins to conduct during ESD. Second, separating the output driver nMOS pull-down stack into different active areas, as shown in Figure 7.20, increases its breakdown voltage [Anderson98A, Miller00]. As the separation between two nMOS devices increases, the impact ionization current from the top drain is less effective in de-biasing the bottom source. Although unreasonable separation is required for the two devices to breakdown as true *npn*s in series [Miller00], even a small separation can provide enough of an increase to the output driver's breakdown voltage. Larger separation allows the 10-Ω series resistor to be removed, if desired.

7.5.3.3 The nMOS stack ESD in a fully silicided process

In a silicided process technology, the gate of an nMOS ESD device must rise above threshold for all fingers to trigger into snapback [Polgreen89]. The same must hold for the stacked nMOS, but both the top and bottom gate must rise above threshold during ESD to allow the channel to conduct. In addition, during normal circuit operation, the top gate must be driven to V_{dd} and the bottom gate to V_{ss} to properly drop the high voltage on the pad. A circuit that performs both functions is shown in Figure 7.21(a). This circuit is similar to the gate-driven nMOS (Section 6.5), using V_{dd} to distinguish between an ESD event and a normal operation I/O transition through the voltage difference between V_{dd} and V_{ss}. Transistors MET and MEB make up the nMOS stack device. Devices MPS and MNS act as an inverter when the pad is high, grounding node EGB under normal operation and pulling it above V_{ss} during an ESD event. Devices MB1 and MB2 drop the voltage on the pad to

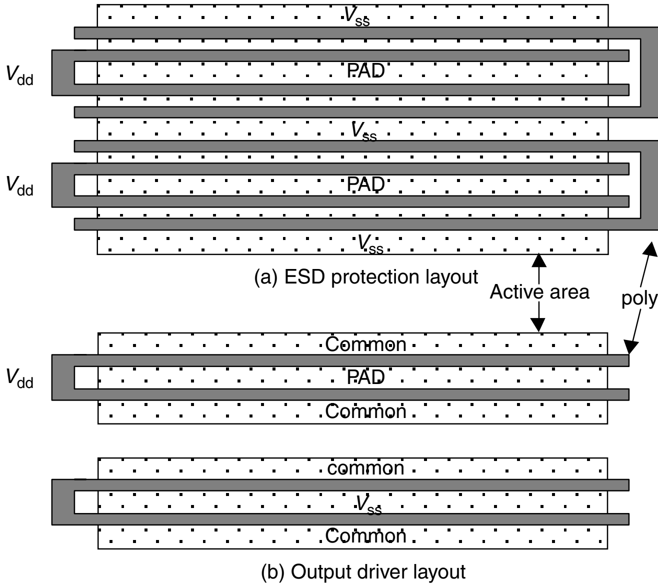


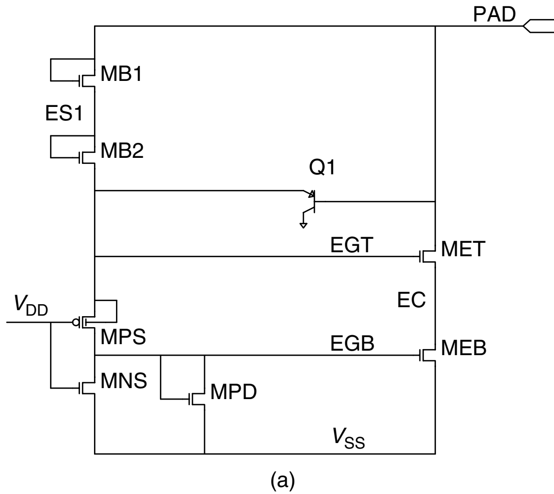
Figure 7.20 Stacked nMOS (a) layout in a single active area and (b) layout in separate active areas. The metal connections between fingers are not shown (After [Anderson98A], reproduced by permission of the ESD Association)

bias EGT safely during normal operation. The ratio of sizes between MB1, MB2, MPS, and MPD determine the voltage on nodes EGT and EGB during ESD. The SPICE simulated response (without nMOS snapback models) during a $1\,\mathrm{V\,ns^{-1}}$ rise on the pad to 5 V in normal operation and to 8 V in an ESD event is shown in Figure 7.21(b).

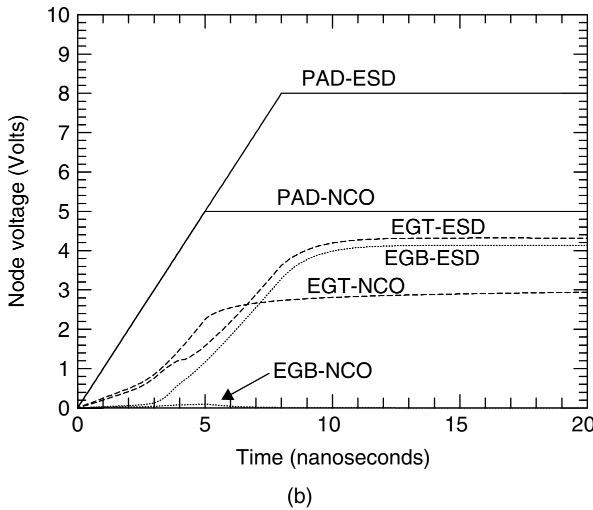
As the stacked nMOS ESD device has a higher clamping voltage than a single nMOS, any I/O design that contains a single nMOS in parallel with the stacked nMOS ESD device will break down and fail through the single nMOS [Anderson98A]. To achieve effective ESD protection, all paths in parallel with the stacked nMOS ESD device must also contain a stacked nMOS. Likewise, the higher clamping voltage makes breakdown through parasitic field devices more likely. Wider spacing between electrically unrelated diffusions in the I/O area will increase the breakdown voltage of such field devices. Separating the nMOS output driver into different active areas, as shown in Figure 7.20(b), is recommended for the design in a fully silicided technology as well.

7.5.4 Stacked Diode Strings

The stacked diode vertical PNP device offers a V_{dd}-based approach to mixed-voltage I/O ESD protection in a CMOS process. The emitter-base diodes of the



(a)



(b)

Figure 7.21 Circuit for controlling the gates of the stacked nMOS during ESD and normal circuit operation and its SPICE simulated response during ESD and I/O switching (NCO) (After [Anderson98A], reproduced by permission of the ESD Association)

vertical PNP can be placed in series to increase the turn-on voltage of the clamp to V_{dd}, allowing the pad to rise multiple diode drops above the V_{dd} level. A cross section and schematic are shown in Figure 7.22.

Several notable effects occur when stacking emitter-base diodes in this fashion. First, the total gain of the transistor stack increases by the power of the number of devices in the string. This is known as the Darlington effect. For a string of m emitter-base diodes, the output base current I_{Bm} is related to the input emitter

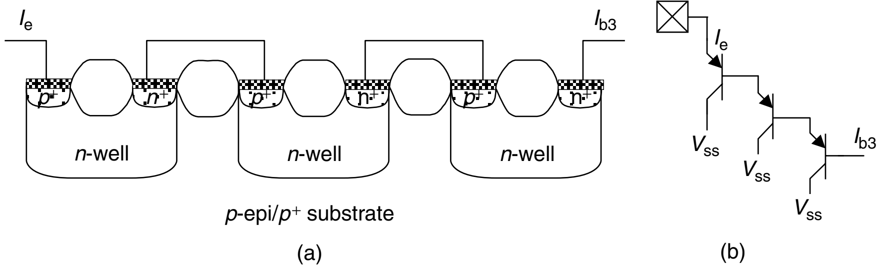


Figure 7.22 A three-high string of PNP emitter-base diodes in cross section (a) and in a circuit diagram (b). The substrate, which is connected to V_{SS}, forms the common collector for each device

current I_E by

$$I_{Bm} = \frac{1}{(\beta + 1)^m} I_E \tag{7.1}$$

where β is the single transistor current gain, which we assume for simplicity is identical for all transistors. The string's gain increase has significant advantages for V_{dd}-based ESD protection circuits, since the fraction of ESD current reaching the power rail is reduced by a factor of $1/(\beta + 1)$ for every transistor in the string.

However, the advantage of the current gain is offset by a larger voltage drop across the emitter-base diode string. For a top emitter current I_E injected into a string of m emitter-base diodes, with an emitter area A_i for transistor i, the voltage drop across the entire string is given by [Voldman95][Dabral93][Dabral94]

$$V_{Em} = m V_{th} \ln \frac{I_E}{A_1 J_{0E}} - V_{th} \frac{m(m-1)}{2} \ln(\beta + 1) + V_{th} \sum_{i=1}^{m} \ln \frac{A_1}{A_i} \tag{7.2}$$

where J_{0E} is the emitter saturation current density of a single transistor ($I_{Ei} = A_i J_{0E} e V_{EBi}/V_{th}$), $V_{th} = n_F kT/q$, n_F is the forward ideality factor, and A_1 is the emitter area of the top transistor. The total voltage across the string consists of three terms: a diode drop for each transistor in the series, a correction term because the bipolar effect reduces the emitter current into each successive transistor in the string, and an area correction term for emitter current density.

The voltage across the emitter-base stack is shown as function of the number of transistors for low- and high-gain devices in Figure 7.23. This relationship shows that the voltage across the stack increases sublinearly with an increasing number of diodes, particularly for higher gain and lower current. In cases where the stack must electrically isolate two nodes during normal chip operation, especially for high-voltage differences, more devices may be necessary than given by a simple m $\times$ 0.6 V calculation. One possible solution is to scale the emitter areas down the stack by $1/(\beta + 1)$ so that each emitter has the same current density, offsetting the second and third terms of Equation 7.2. However, this can increase the clamping voltage during ESD if β rolls off significantly. A second solution is to

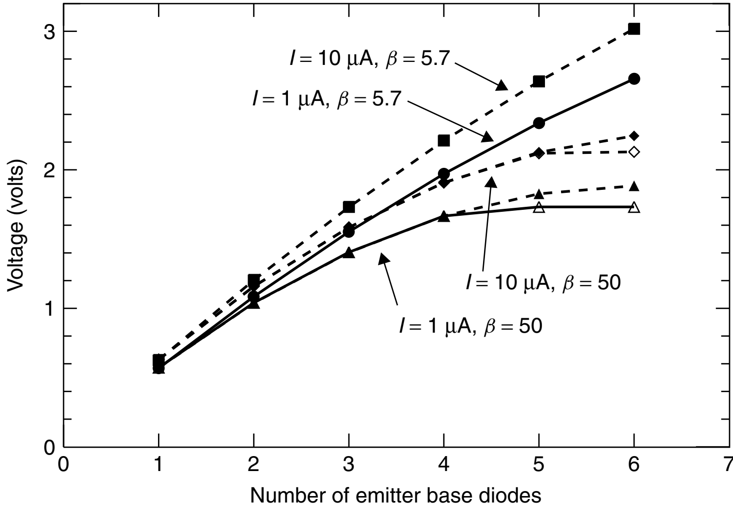


Figure 7.23 The voltage drop across the emitter-base stack as a function of the number of devices for low-gain ($\beta = 5.7$) and high-gain ($\beta = 50$) devices, from Equation 7.2 (solid symbols) and SPICE (open symbols). The deviation between simulation and Equation 7.2 for larger number of diodes occurs because the current at the bottom diodes is so small that $I_E = AJ_{0E}e^{V_{EB}/V_{th}}$ is not a valid approximation. Other parameters: $I_E = 1$ μA (solid lines) and 10 μA (dashed lines), $A_1J_{0C} = 3.1 \times 10^{-16}$ A, $J_{0E} = [(\beta + 1)/\beta]J_{0C}$, all $A_i = A_1$, $T = 25\,°C$

inject a small amount of current into the emitter junctions of the lower transistors through a leakage network, maintaining a uniform emitter current density through the stack [Maloney95].

The temperature dependence of J_{0E} in Equation 7.2 causes the voltage across the stack to decrease with temperature for a given current. Additional transistors may be required in the stack for normal operation at high temperature, but this will increase the voltage drop across the stack during ESD, which takes place at room temperature.

An alternative way to look at these effects is to compare Gummel plots, as shown in Section 7.5.4. Equation 7.2 can be rewritten to give the input emitter current as a function of applied voltage to the top emitter V_{Em}

$$I_E = A_1 J_{0E} \exp\left[\frac{m-1}{2}\ln(\beta+1) - \frac{1}{m}\sum_{i=1}^{m}\ln\frac{A_1}{A_i}\right]\exp\left(\frac{V_{Em}}{mV_{th}}\right). \quad (7.3)$$

The bipolar current gain term and area term now become prefactors, increasing the effective emitter saturation current. At room temperature, the slope of the $\log I - V_{EB}$ plot becomes m × 60 mV decade^{-1} (see Figure 7.25(b)).

A good stacked-diode ESD protection circuit must use the correct number of diodes in the string. One with too few diodes will have high-temperature

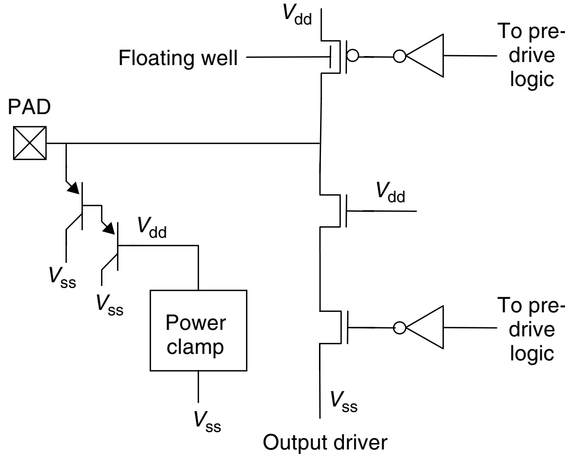
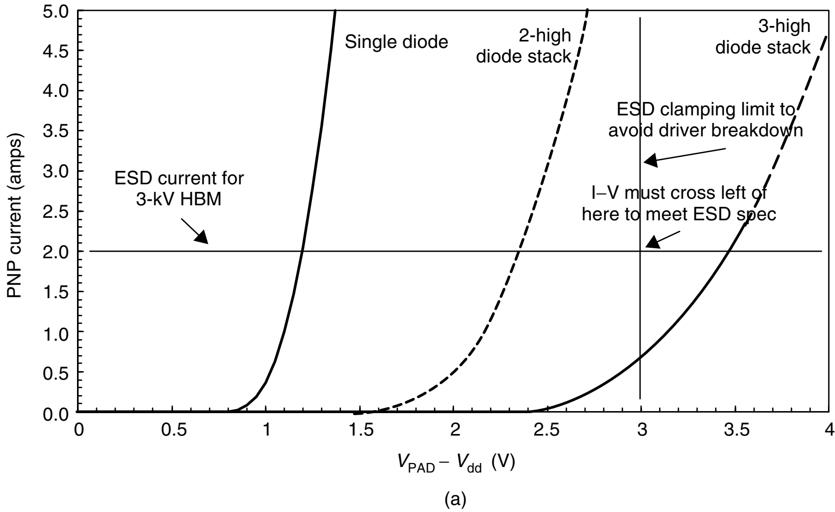


Figure 7.24 ESD protection using a Darlington-connected stack of vertical PNP emitter-base diodes in a high-voltage tolerant I/O circuit

leakage current during normal operation. One with too many produces a large voltage drop during ESD at room temperature, which could cause ESD failure of the output driver. For 5 V-tolerant I/O in a 3.3 V process, strings of five or six diodes were necessary [Voldman94A, Voldman94B, Voldman95]. For 2.2 V-tolerant I/O in a 1.5-V process, two or three are appropriate [Benschneider00]. Such a configuration is shown in Figure 7.24. The clamp is analogous to V_{dd}-based ESD protection with a single diode. It can be evaluated the fashion of Section 5.2.5. In this case, however, ESD current flows to V_{dd} through two emitter-base diodes, dropping nearly twice the voltage as in the single diode case.

Using a 2.2 V-tolerant I/O in a $V_{dd} = 1.5$ V process, Figure 7.25 illustrates how the number of diodes creates a trade-off between ESD clamping voltage and normal operation leakage current. Figure 7.25(a) shows the ESD case, where the pad's ESD current is plotted against the drop across the diode string, $V_{pad} - V_{dd}$, for a one, two, and three diode string at a temperature of 25 °C. The diode string voltage drop under a 3-kV HBM event (2-A current) can be read for each $I-V$ curve. In this case, we require a drop of less than 3 V across the string to avoid output driver breakdown. A string of one or two diodes meets this criterion. A string of three does not. Figure 7.25(b) shows the case of normal operation leakage. The ESD clamp must not draw more than 150 µA. Leakage is highest for a high pad voltage, a low V_{dd}, and maximum die temperature of 100 °C, so we must consider the I/O and power supply voltage tolerance of ±100 mV for each as well. Therefore, the maximum voltage that can appear across the diode string in normal operation is 0.9 V. A string of two and three diodes meets the leakage limit, while a single diode does not. Therefore, a two-diode string is the only choice capable of meeting both the ESD and leakage requirements.



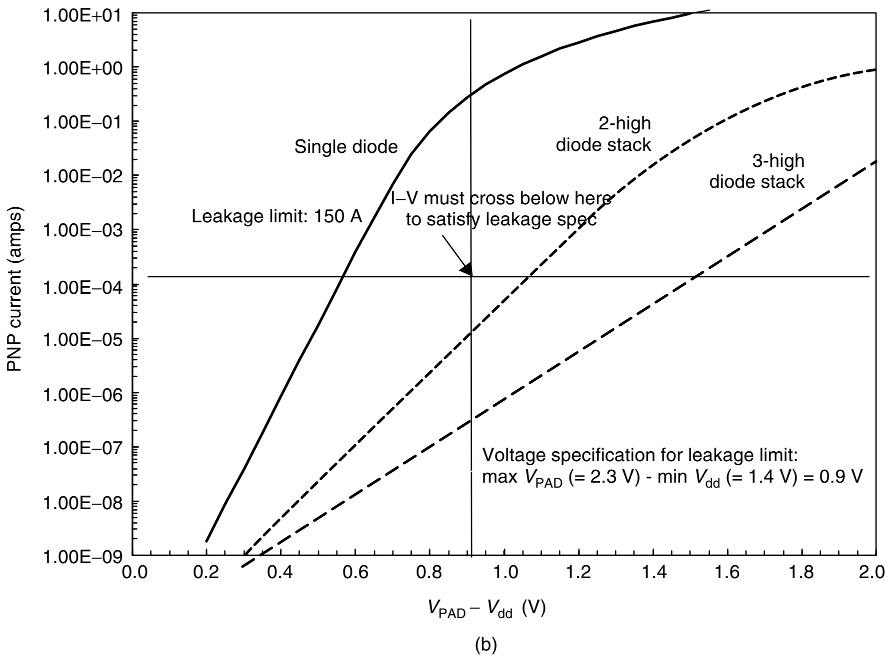


Figure 7.25 The $I-V$ characteristics of a one- two- and three-high stack of PNP emitter-base diodes during (a) ESD and (b) normal operation (After [Benschneider00])

Two further effects, leakage multiplication and capacitance multiplication, deserve particular attention in a stacked-diode ESD design. Since the input emitter current is related to the output base current by $I_E = (\beta + 1)^m I_{Bm}$, a small amount of leakage current, especially in the lower devices or out of the bottom base, can cause a large input emitter current. Capacitance multiplication occurs on stacked PNP clamps tied to switching signals. A small displacement current on one of the lower devices is multiplied by the gain factor and causes a large transient input emitter current, effectively multiplying the pad capacitance. Parallel bias networks [Maloney95] or snubber diodes [Voldman95] can provide an alternate source for both the leakage and the capacitor charging current, reducing the current drawn through the PNP stack, and therefore the current multiplication.

7.5.5 Clamps for High-Voltage Supply Rails

A final ESD option for mixed-voltage I/O is to place a single PNP diode between the pad and a power supply rail powered to the high-voltage I/O level. Such a rail will exist for I/O that must drive the high-voltage level out. On inputs that must only tolerate receiving the high-voltage, or bidirectional pins that must receive the high voltage but can drive only up to the nominal internal V_{dd} level, adding an explicit high-voltage rail for ESD protection purposes is often possible.

In either case, a single diode to a high-voltage supply rail, V_{ddioh}, allows the diode portion of the clamp circuit to function in the same fashion as in non-mixed-voltage I/O. However, the power rail ESD clamp must now tolerate the high voltage during normal operation and meet robustness and effectiveness requirements during ESD. A number of such clamp circuits have been developed. One way to overcome high-voltage limitations is to place a nominal-voltage tolerant RC clamp between V_{ddioh} and V_{dd}. ESD current discharges through this clamp to V_{dd} then through the RC clamp on V_{dd} to V_{ss}, as shown in Figure 7.26. During ESD, the clamp devices MNT and MNB operate in series, with node GT pulled to V_{ddioh} and node GB pulled to V_{dd}. During normal operation, node GT is driven to V_{dd} while node GB is pulled to V_{ss}, turning both nMOS clamp devices off. As long as $V_{ddioh} \leq 2V_{dd}$, which is usually the case, this configuration safely withstands the high voltage. Its disadvantage is that node GB is only driven to a moderate voltage during the ESD event. If node GB could be pulled all the way to V_{ddioh} during ESD, device MNB would drop less voltage and its size could be reduced.

The clamp shown in Figure 7.27 does exactly this. An independent RC timer still activates both the top and bottom clamp devices. The source for the pMOS PB1 driving the bottom gate GB now connects to node GT, which is driven all the way to the V_{ddioh} rail during ESD. Since node GB1 is at V_{ss} during the ESD event, this circuit will pull node GB all the way to the V_{ddioh} rail, maximizing the drive on both clamp devices MNT and MNB. Node EC can now remain low during the ESD event with no adverse effect on the clamp. As the lower half of

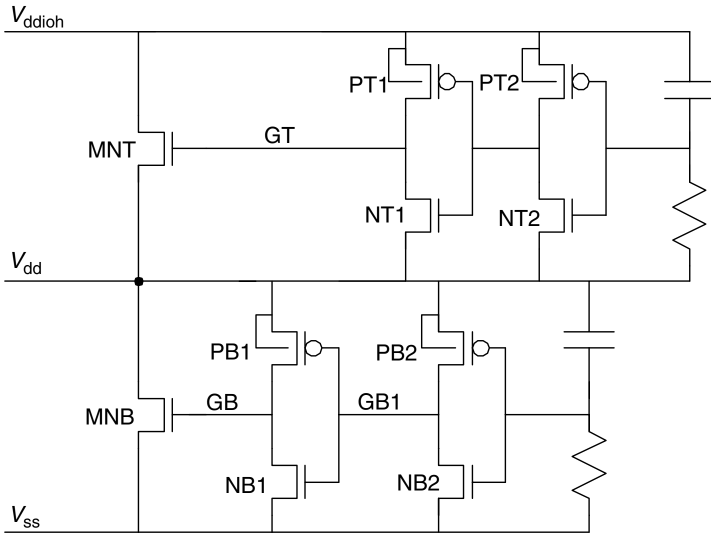


Figure 7.26 An RC clamp for a high-voltage supply rail V_{ddioh}, discharging to the core supply, V_{dd}, then to the ground supply, V_{ss}

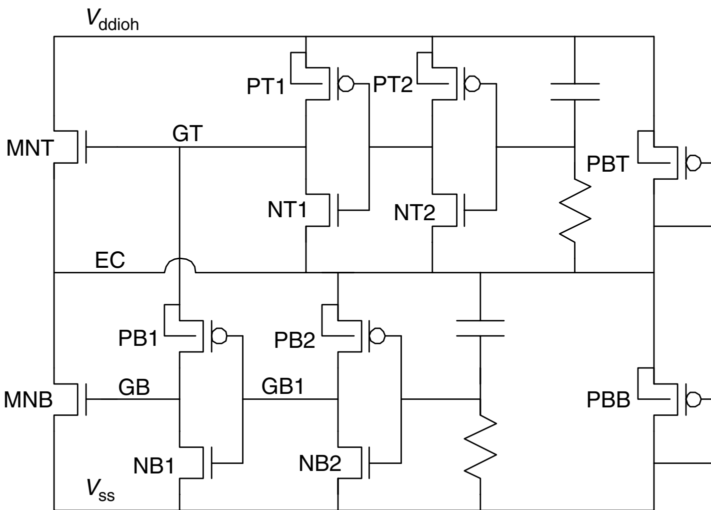


Figure 7.27 An improved RC clamp for a high-voltage supply rail V_{ddioh}, modified from the pMOS circuit in [Maloney99]

this clamp no longer functions on its own, the center node can no longer connect to V_{dd}. Instead, it is set to approximately $V_{ddioh}/2$ through bias devices PBT and PBB. A pMOS alternative with slightly different connections was introduced in [Maloney99].

Another type of high-voltage power supply clamp relies on the PNP diode stack, with a single nMOS pulling down the bottom base during ESD [Maloney95] [Maloney98]. The stack of PNP emitter-base diodes drops the voltage on the high-voltage rail in the same fashion as it does for an I/O (see Section 7.5.4). The stack also decreases the ESD current delivered to the nMOS in the clamp by a factor of $1/(\beta + 1)^m$, where m is the number of diodes in the stack. The gate voltage for the nMOS can be driven from an RC timer. This type of clamp is most useful for supplies at and above 3.3 V.

Regardless of which type of high-voltage power supply clamp is used, its ability to provide effective V_{dd}-based ESD protection can be determined using the principles outlined in Section 5.2.5.

7.5.6 Summary of Mixed-Voltage I/O Options

Mixed-voltage I/O can be divided into two general categories.

(a) High-voltage tolerant I/O requires that the I/O receive a voltage greater than the nominal chip supply voltage V_{dd}. If the pin is bidirectional, the output driver need only drive as high as V_{dd}.

(b) High-voltage drive I/O requires that the I/O receive and drive a voltage greater than V_{dd}. A high-voltage supply V_{ddioh} must be present to accomplish this function.

In both cases, ESD protection on the I/O may use several options: the nMOS stack in snapback, stacked diode strings, snapback through a dual gate oxide transistor, and the FOD. A single diode to a high-voltage rail is straightforward to implement for high-voltage drive I/O, but requires the development of a high-voltage tolerant power supply clamp. This solution is not possible for high-voltage tolerant I/O unless an explicit high-voltage supply rail is added.

7.6 CDM PROTECTION

In the previous chapter and in this chapter, several novel CMOS protection device concepts have been introduced. These protection devices have been known to work well for HBM and their performance to CDM has been tested only in a limited manner. It is important to note that, compared to HBM, the current paths are reversed for CDM discharge [Maloney88]. That is, with the substrate charged and the I/O pin grounded during the test, a positive CDM event is equivalent in current path to a negative HBM pulse, and vice versa. Therefore, for positive

cycle of the CDM pulse the substrate diode essentially provides the protection. But during the negative cycle the speed of the protection device trigger itself becomes critical. One such investigation of the protection device trigger during CDM reported that the SCR device is too slow and subsequently yields low CDM performance [Duvvury95A]. On the other hand, the same study observed that the nMOS (specifically the GCNMOS) is effective for CDM protection. The substrate triggered devices, STNMOS and NTNMOS, both were found also found give good CDM performance [Duvvury00A]. Any of the diode protection schemes are also expected to be generally robust for CDM. Recent work [Kunz01][Russ01] examined the SCR applications for the deep submicron technologies and found that with the much closer spacings the lateral parasitic bipolar devices have become more efficient to successfully enable the SCR during CDM. An effective model to represent the SCR during CDM is important for this device to be applicable in future protection designs. [Juliano01B].

7.7 SOI TECHNOLOGY

Silicon-on-Insulator or SOI is becoming the next important technology of the future because of its advantages of low power consumption. The cross section of an SOI transistor is shown in Figure 7.28. Owing to the buried oxide isolating the substrate the device is symmetrical in nature for both positive and negative stress modes. The consideration for ESD has been dealt by various authors [Verhaege93][Voldman98][Voldman99][Duvvury96]. Because of the floating substrate a gate-coupled nMOS [Duvvury96] protection or a substrate coupled nMOS

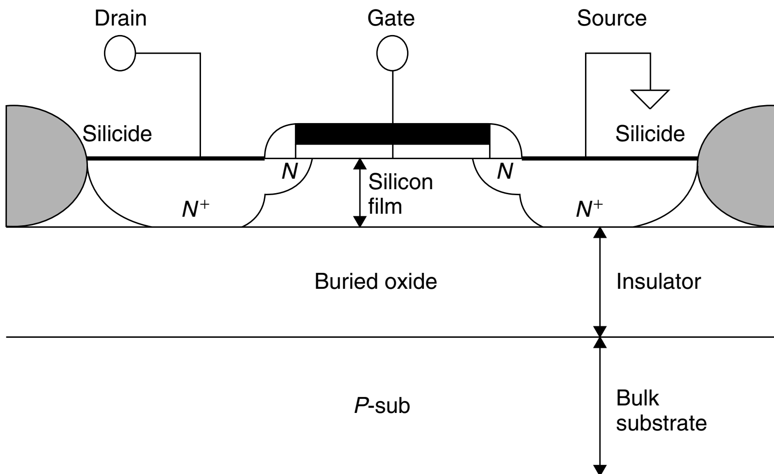


Figure 7.28 Cross section of a Silicon-on-Insulator transistor. With the isolated substrate there is no substrate diode available for negative stress protection

[Voldman98] protection devices are robust. Also, implementing lateral diodes [Voldman99] is a relatively simple approach. In general, good ESD protection can be designed for the SOI technologies at the present time.

7.8 HIGH-VOLTAGE TRANSISTORS

High-voltage transistors are used for Smart Power ICs and Power Logic Chips. There are two types of transistors that are commonly used: Drain-Extended nMOS (DENMOS) and the Lateral DMOS (LDMOS). For both of these transistors the ESD protection design is a challenge because the parasitic bipolar *npn* is difficult to turn on [Duvvury94]. Because of the *n*-well, drain of the DENMOS as shown in Figure 7.29 triggers like a reverse diode and fails at a low ESD voltage with high-power dissipation. The maximum heating occurs at the drain junction with the damage location as shown in Figure 7.29. For relatively high substrate resistance the DENMOS can turn on as an *npn* with gate coupling [Duvvury95B], but this is not practical for modern technologies. The other option is to integrate an SCR device into the DENMOS to achieve excellent protection levels [Duvvury94][Kunz01]. For example, see Figure 7.30 in which an SCR device is combined with the DENMOS. With optimized spacings the SCR can be made to trigger before the DENMOS.

The DENMOS transistors are promising to have applications in the future low voltage CMOS also as the compatibility to the 5-V applications remains important. In such cases the DENMOS will be useful, especially for fail-safe analog circuits. The SCR integrated into the DENMOS again can offer the ESD protection. However, the formation of Shallow Trench Isolation (STI) between the diffusions has resulted in the SCRs from being robust in the advanced technologies [Russ01]. The SCR-DENMOS protection while blocking the STI is also reported [Kunz01] and the cross section is shown in Figure 7.31. The SCR in

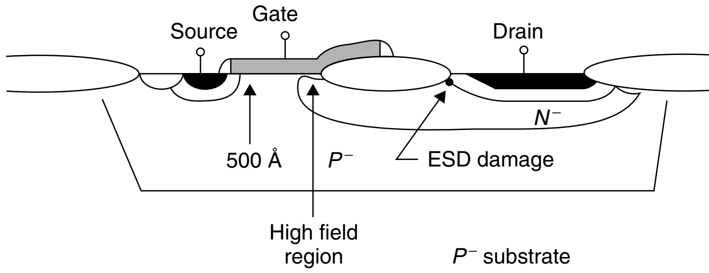


Figure 7.29 Cross section of a Drain-Extended nMOS. Under ESD the damage occurs at the drain junction as the *npn* does not turn on. (After [Duvvury94], reproduced by permission of ©1994 IEEE)

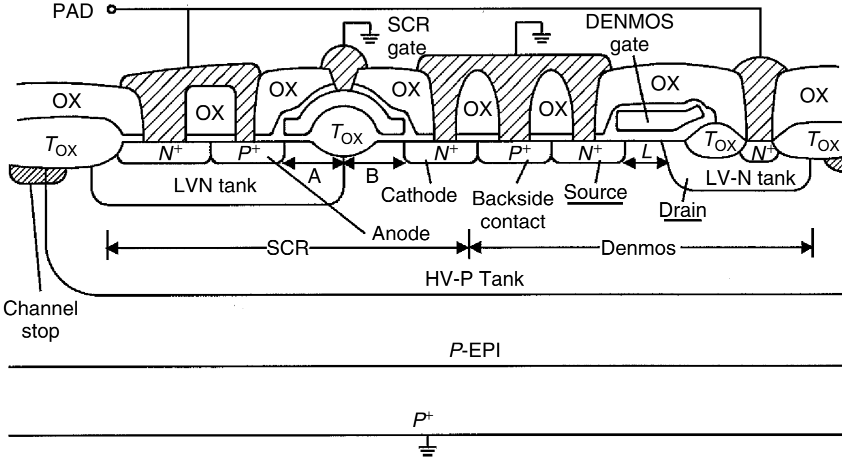


Figure 7.30 Cross section of a Drain-Extended nMOS integrated with an SCR. With this approach the SCR is designed to trigger before the DENMOS to protect it [Duvvury94]. Reproduced by permission of ©1994 IEEE

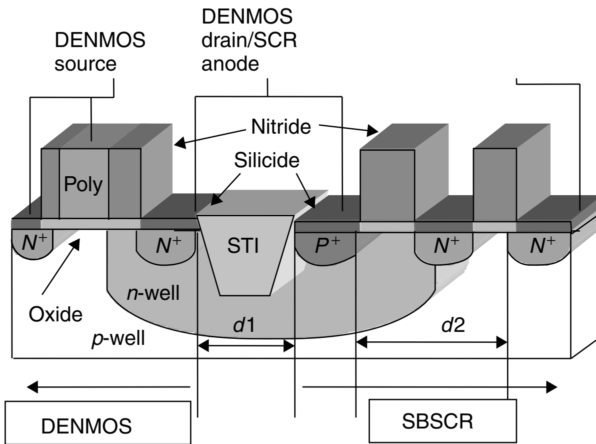


Figure 7.31 Cross section of a Drain-Extended nMOS integrated with an SCR for STI processes of the sub-0.5 µm. Similar to the concept shown in Figure 7.30, this SCR also is designed to trigger before the DENMOS to protect it [Kunz01]. Reproduced by permission of ESD Association

the figure is formed by blocking the STI with a nitride mask, and is known as STI-Blocked SCR or SBSCR. Note that spacings $d1$ and $d2$ can be adjusted to improve the SCR performance. As there is no poly gate on the SCR side it essentially triggers similar to a MLSCR, and with its lower trigger voltage protects the DENMOS.

The LDMOS device also has a large base length to trigger as *npn*. In most of the applications it is a fairly large device of several thousand microns wide and hence its MOSFET conduction can be used to provide high current ESD protection. This can be simply done by using Zener clamps on the gate and drain to allow optimum MOS conduction [Duvvury97]. It is also common to use an SCR protection device in parallel to the LDMOS output as long as the SCR triggers prior to LDMOS breakdown. In such cases the trigger voltage of the SCR is tuned by varying the spacings.

7.9 BiCMOS PROTECTION

The BiCMOS and Bipolar technologies are attractive for both higher voltage applications and high-speed RF applications. For RF applications where a minimum capacitance is desirable for the protection circuit, the SCR could be very useful since it gives relatively higher volts per micron. However, in the bipolar technologies the buried n^+ diffusion can prevent this SCR formation. A device modification as shown in Figure 7.32 can be done to successfully use the SCR [Chen95]. Here, the p^+ anode is the emitter, the n-well is the base, and the p-base of the *npn*

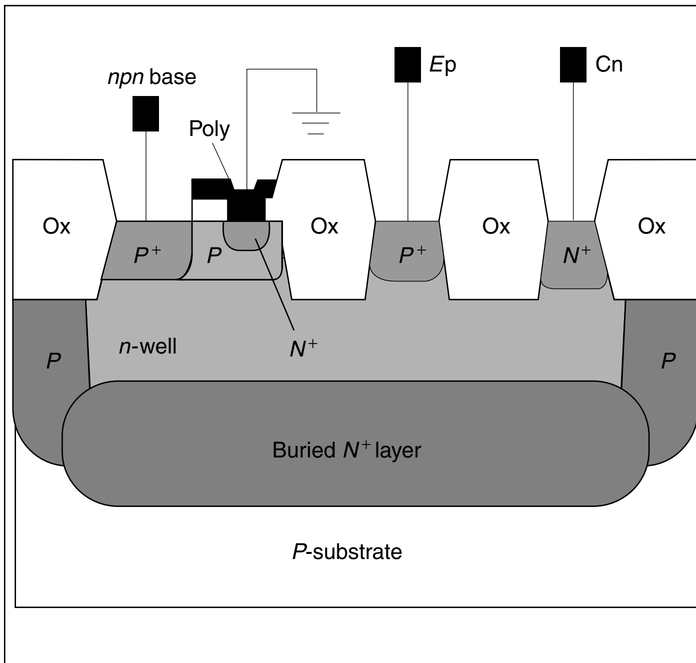


Figure 7.32 Cross section of a Bipolar SCR [Chen95]. An external trigger is applied at the *npn* base to initiate the SCR action. Reproduced by permission of ©1995 IEEE

is the collector. The PNP with the vertical *npn* forms the SCR. Note that the n^+ diffusion contact allows the PNP base-emitter resistor to be increased. With an external trigger (e.g. with a Zener diode) the *npn* base can be triggered to operate the SCR.

For the BiCMOS/Bipolar applications, low capacitance (<0.5 pF) and low series resistance (<0.5 Ω) can be typical requirements. Instead of the BSCR discussed in Figure 7.32, the bipolar *npn* can also be used as the primary protection devices. Of particular importance is the Zener triggered *npn* [Corsi93] where the Zener is placed between the pad the base. Similar to the multfinger nMOS, the multiemitter finger bipolar *npn* also requires a design technique to achieve uniform trigger where equal base resistance is achieved [Chen96].

7.10 RF DESIGNS

Most of the RF designs are done with BiCMOS or GaAs processes. For GaAs good protection is difficult to obtain [Rubalcava86][Diep92] making it sensitive to ESD. RF applications in CMOS are becoming important. A typical specification for a high-speed 2-GHz application circuit would be 200 fF of loading capacitance with no series resistance. For analog operation, the linearity of this capacitance with signal bias should not exceed 20%. Compared with the nMOS and the SCR, the dual-diode approach was established to meet this criterion for 2-kV protection [Richier00]. This comparison is shown in Figure 7.33. The slopes on the lines represent the scaling to reduce capacitance, which will scale down the ESD level. It is clear that the diode option is relatively better.

Finally, the use of heterojunction Si-Ge bipolar transistors (HBT) integrated with the mainstream advanced CMOS promises to offer excellent analog and RF performance along with robust ESD [Voldman00].

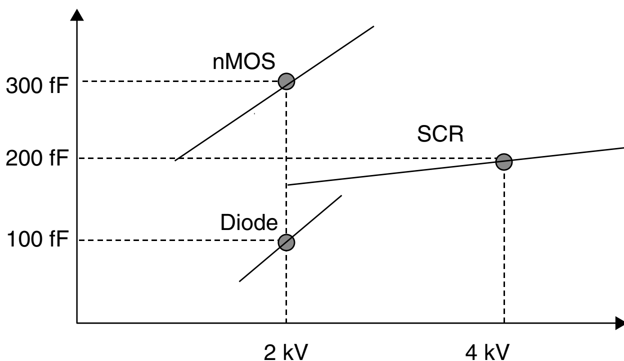


Figure 7.33 Comparison of HBM ESD level versus the different protection options versus the capacitance. The slopes represent the scaling of these devices [Richier00]

7.11 GENERAL I/O PROTECTION SCHEMES

A general I/O protection scheme strategy would involve implementing protection devices for the different current paths as shown in Figure 7.34. Referring to the figure, Clamp1 is the primary protection device for I/O to V_{ss}. For CDM protection, Clamp2 is placed at the input buffer nMOS gate while referencing to the same ground as the input buffer. This is important to effectively clamp the local voltage. Note that the resistor R serves two different purposes: (1) to facilitate the trigger of Clamp1, and (2) protect Clamp2 during CDM events. In protection applications where the STNMOS or the NTNMOS is used this isolation R may not be needed. However, for a CDM event with the substrate (source) charged positive and the I/O pin is ground the current flow would be through both the Clamp1 and Clamp2. As Clamp2 usually is designed to be small to clamp local voltage, the resistor R would protect it during this stress mode. Next, Clamp3 is required to protect the pMOS gate of the input buffer and its design was described in Figure 6.27. Next, Clamp4 is essential to protect the circuits connected between the I/O V_{dd} supply and V_{ss}. Its function also becomes more important when the I/O pin is stressed negative to V_{dd}. During this stress mode the voltage builds up relative to the I/O pad and can cause either ESD failure or even a latchup failure of the buffers and pre-driver circuits (see Figure 8.50).

In the case of outputs, if the buffer nMOS source is referenced to a different ground, V_{sss} as in this example, then diodes should be placed between this ground and V_{ss}. Note that (for the output case) the resistor value for R can be the same as

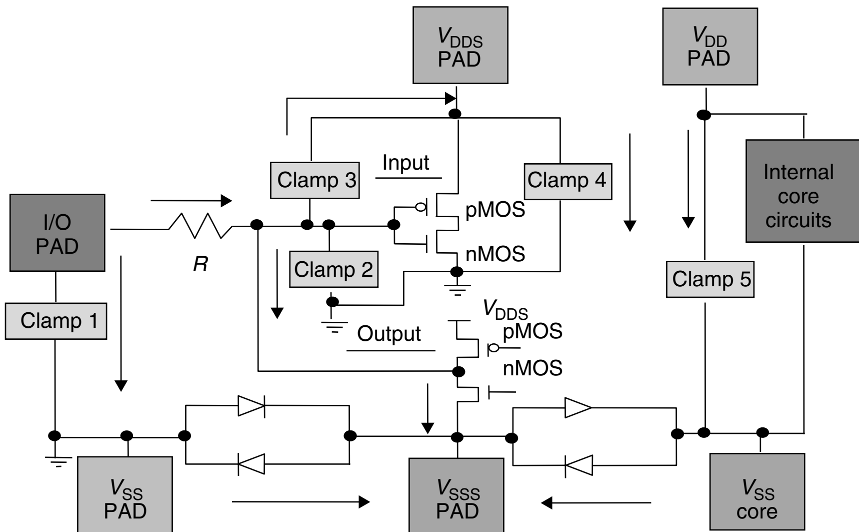


Figure 7.34 A general I/O protection scheme showing the different clamps necessary for HBM and CDM protection

designed for the input buffer or a smaller one to meet the output buffer performance requirements. For the case of STNMOS protection for Clamp1, the resistor may not be necessary but the buffer nMOS must be larger enough to handle the extra current during HBM and CDM.

In mixed-voltage technologies, as the internal core circuits have a different power supply they need a separate protection device shown as Clamp5 in Figure 7.34. If an nMOS is the protection device for Clamp4 or Clamp5, the gate oxide of the protection device must be compatible with the supply voltage. For example, a 3.3-V nMOS cannot protect the internal core circuits on the V_{dd} core supply. Note also that antiparallel diodes are placed to from the V_{ss} Core. Although not shown here, for the internal analog block circuits a separate V_{dd} supply is often used and it needs its own protection device. In general, the circuit designers should be discouraged from using too many different designated power supplies since this could complicate not only the protection design but also the current paths during ESD causing unusual ESD failures.

7.12 DESIGN/LAYOUT ERRORS

Even with the best protection strategy, unexpected ESD failures could occur due to design errors, layout errors, parasitic current paths, and sneak current paths. The design and layout errors can be addressed with thorough peer reviews or by using sophisticated software programs to capture these errors before the design is released for mask generation. One such program called The ESD Checker is described in Figure 7.35 [Sinha98]. The ESD rules are established based on the information from the required protection designs, and good recommended layout practices, such as metal width, contact uniformity, nMOS parameters. The technology information is also important for the program to check the appropriate rules for the sensitivity of the process. A typical program can evaluate the design database to give error reports as well as provide a netlist for simulation purposes.

Parasitic device interaction resulting in unwanted ESD conduction paths and premature ESD failures is well known [Duvvury89][Chaine97][Johnston93]. This occurrence is more common near the I/O pad locations. For example, when the I/O pad is stressed positive to V_{ss}, the protection device trigger could generate local substrate currents to trigger parasitic devices. This is indicated as *npn*-I in Figure 7.36 with the actual damage mode shown on the right side [Sinha98]. Likewise, when the V_{dd} is stressed positive with respect to the I/O, *npn*-II damage could occur. Finally, when the I/O pad is stressed negative with respect to V_{dd} *npn*-III could trigger because under this condition the V_{dd} potential would rise relative to the I/O pin. Ensuring the spacings from the I/O diffusions are a required distance away from the unrelated diffusions could easily eliminate all of these failures. A good rule of thumb would be 2× the minimum allowed for the process although for the bulk substrate technologies this may not always be enough. Test structure

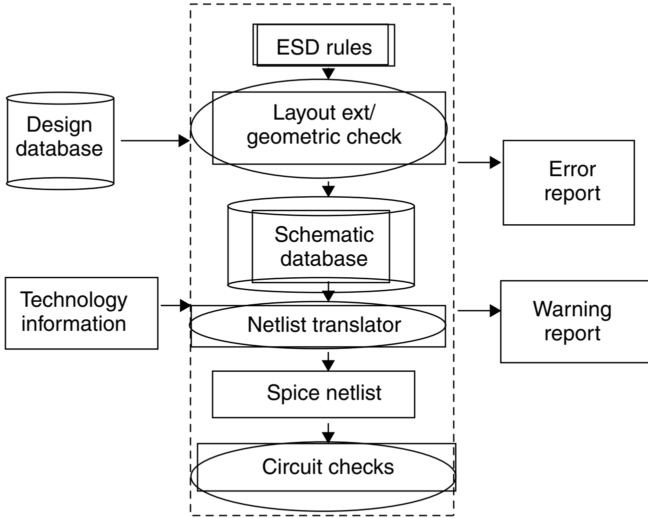


Figure 7.35 The flow diagram of software program to detect ESD design errors [Sinha98]. Reproduced by permission of ESD Association

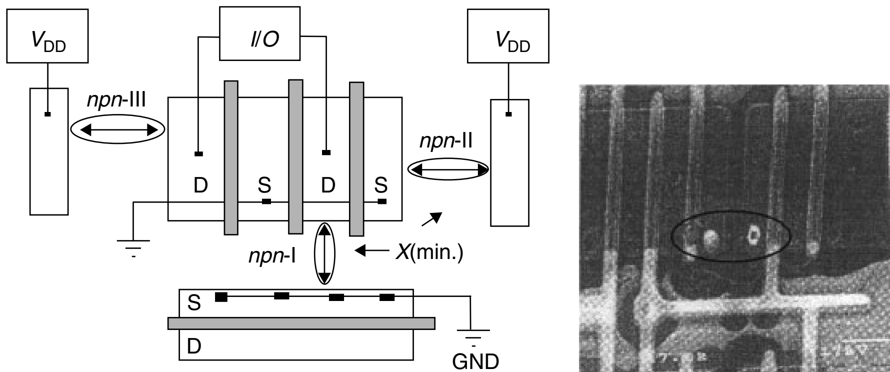


Figure 7.36 Parasitic device interactions near the I/O pad diffusions. The damage picture on the right corresponds to *npn*-III [Sinha98]. Reproduced by permission of ESD Association

analysis is important to establish the clear rules that can be part of the ESD rules in Figure 7.35.

The parasitic device interactions are challenging phenomena that need to be considered for ESD analysis. Once the parasitic device model is established it can be used in the ESD SPICE simulations to predict the failure current paths [Fong89]. Although this is both important and feasible at the I/O pin, it could become more complex for the internal core circuits. Two effective ways to deal with this to minimize core ESD failures is to: (1) place very effective power clamps that have

low holding voltage and low on resistance, and (2) maintain a certain required spacing between diffusions connected to V_{dd} and diffusions connected to V_{ss}. The first option involves selecting the V_{dd} protection device, which could be a large gate-coupled nMOS or a gate-driven nMOS. An SCR is not recommended as it can accidentally trigger during product burn-in tests and cause EOS failures. The second strategy requires defining the spacings between the critical diffusions that are directly connected to the opposite polarity power supplies. As this can have direct impact on the density of the chip layout, a 1.5X spacing for n-diffusions and the process minimum for the p-diffusions could be used. Note that the occurrence of n-diffusions connected to the V_{dd} and V_{ss} supplies, respectively, is rare for the core layouts and hence these can be addressed without much impact on the layout density.

7.13 SUMMARY

In this chapter some of the latest protection device concepts for submicron technologies were summarized. First was the substrate triggered protection devices that efficiently clamp the ESD voltage to protect the internal gate oxides of the inputs or to protect the output buffer without requiring an isolation resistor between the protection device and the buffer devices. Similar to the gate-coupled protection devices discussed in Chapter 6, these devices also need consideration of interaction with the I/O circuit.

The issue of mixed-voltage technologies and the challenges of designing protection devices for voltage tolerant applications were also discussed. The stacked nMOS is a common output device concept but its trigger mechanism as a parasitic bipolar *npn* under ESD is complicated. An effective design requires tuning the various parasitic current paths. Several nMOS stack examples were given. For voltage tolerant protection, stacked diodes are also an attractive option for CMOS I/O. Their design requires an understanding of the multiple series device operation and their leakage behavior.

For smart power ICs, high-voltage transistors such as the drain-extended nMOS are commonly used, but their ESD design requires special considerations as mentioned. There are newer developments such as silicon-on-insulator (SOI) that need attention for future ESD applications. Besides SOI, advanced BiCMOS for RF high-speed applications is becoming important and its ESD design issues were also briefly discussed.

An example of the total protection of an IC chip was illustrated. The chip's overall ESD performance requires comprehending all possible current paths and the placement of various clamps. Finally, the effective ESD performance depends on the careful layout methods to avoid design errors and prevent unexpected parasitic current paths. To efficiently check for these in a complex chip design, software tools will become very important. Once such tool example was discussed.

BIBLIOGRAPHY

[Amerasekera94] A. Amerasekera, C. Duvvury, V. Reddy and M. Rodder, "Substrate triggering and silicide effects on ESD performance and protection circuit design in deep submicron CMOS processes", in *IEDM Tech. Digest*, 547–555, 1995.

[Amerasekera99] A. Amerasekera and R. Salem, "Cascoded-MOS ESD Protection Circuits For Mixed Voltage Chips", U.S. Patent 5,930,094, 1999.

[Anderson98A] W.R. Anderson and D.B. Krakauer, "ESD Protection for Mixed Voltage I/O Using nMOS Transistors Stacked in a Cascode Configuration", in *Proc. 20th EOS/ESD Symposium*, pp. 54–62, 1998.

[Anderson98B] W.R. Anderson, J.J. Montanaro and N.J. Howorth, "Cross-Referenced ESD Protection for Power Supplies", in *Proc. 20th EOS/ESD Symposium*, pp. 86–95, 1998.

[Beebe98] S. Beebe, "Simulation of Complete CMOS I/O Circuit Response to CDM Stress", in *Proc. 20th EOS/ESD Symposium*, pp. 259–270, 1998.

[Benschneider00] B.J. Benschneider, S. Park, R. Allmon, W. Anderson, M. Arneborn, J. Cho, C. Ghoi, J. Clouser, S. Han, R. Hokinson, G. Hwang, D. Jung, J. Kim, J. Krause, J. Kwack, S. Meier, Y. Seok, S. Thierauf and C. Zhou, "A 1 GHz Alpha Microprocessor", in *Proc. IEEE International Solid-State Circuits Conference*, pp. 86, 87, 2000.

[Bryan94] A. Bryan, W. Hänsch and T. Mii, "Characteristics of CMOS device isolation for the ULSI age", in *IEDM Tech. Digest*, 671–674, 1994.

[Chaine97] M. Chaine, S. Smith and A. Bui "Unique ESD Failure Mechanisms During Negative to VCC HBM Stress", in *Proc. 19th EOS/ESD Symposium*, pp. 346–355, 1997.

[Chen95] J. Chen, A. Amerasekera and T. Vrotsos, "Bipolar SCR ESD protection circuit for high speed submicron bipolar/BiCMOS circuits", in *IEDM Tech. Digest*, 337–340, 1995.

[Chen96] J. Chen, X. Zhang, A. Amerasekera and T. Vrotsos, "Design and layout of a high ESD performance *npn* structure for submicron BiCMOS/bipolar circuits", in *IRPS Proc.*, 227–340232, 1996.

[Chen97] J.Z. Chen, A. Amerasekera and C. Duvvury, "Design Methodology for Optimizing Gate Driven ESD Protection Circuits in Submicron CMOS Processes", in *Proc. 19th EOS/ESD Symposium*, pp. 230–239, 1997.

[Corsi93] M. Corsi, R. Nimmo and F. Fattori, "ESD Protection of BiCMOS Integrated Circuits Which Need to Operate in the Harsh Environments of Automotive or Industrial", in *Proc. 15th EOS/ESD Symposium*, pp. 209–214, 1993.

[Dabral93] S. Dabral and R. Aslett, "Designing On-Chip Power Supply Coupling Diodes for ESD Protection and Noise Immunity", in *Proc. 15th EOS/ESD Symposium*, pp. 239–250, 1993.

[Dabral94] S. Dabral, R. Aslett and T. Maloney, "Core Clamps for Low Voltage Technologies", in *Proc. 16th EOS/ESD Symposium*, pp. 141–148, 1994.

[Diep92] T. Diep, S. Phatak and D. Yoo, "PIN Photo-detectors - The ESD Bottleneck in Laser Packages", in *Proc. 14th EOS/ESD Symposium*, pp. 159–167, 1992.

[Duvvury89] C. Duvvury, T. Taylor, J. Lindgren and S. Kumar, "Input protection design for overall chip reliability", in *Proc. 11th EOS/ESD Symposium*, pp. 190–198, 1989.

[Duvvury94] C. Duvvury, J. Rodriguez, C. Jones and M. Smayling, "Device integration for ESD robustness of high voltage power MOSFETs", in *IEDM Digest*, 407–410, 1994.

[Duvvury95A] C. Duvvury and A. Amerasekera, "Advanced CMOS Protection Device Trigger Mechanisms During CDM", in *Proc. 17th EOS/ESD Symposium*, pp. 162–174, 1995.

[Duvvury95B] C. Duvvury, J. Rodriguez, F. Carvajal, A. Young, D. Redwine and M. Smayling, "Efficient *npn* operation in high volage NMOSFET for ESD robustness", in *IEDM Digest*, 345–348, 1995.

[Duvvury96] C. Duvvury, A. Amerasekera, S. Ramaswamy and S. Young, "ESD design for deep submicron SOI technology", in *Tech. Digest Symposium on VLSI Technology*, 1996.

[Duvvury97] C. Duvvury, F. Carvajal, C. Jones and D. Briggs, "Lateral DMOS design for ESD robustness", in *IEDM Digest*, 375–378, 1997.

[Duvvury00A] C. Duvvury, "Substrate Pumped nMOS for ESD Protection Applications", in *Proc. 22nd EOS/ESD Symposium*, pp. 7–17, 2000.

[Duvvury00B] C. Duvvury, S. Ramaswamy, A. Amerasekera, R. Cline, B. Andresen and V. Gupta, "Substrate Pump NMOS for ESD Protection Applications", in *Proc. EOS/ESD Symposium*, 2000.

[Fong89] Y. Fong and C. Hu, "Internal ESD transients in input protection circuits", in *Proc. IRPS*, 77–81, 1989.

[Johnston93] C. Johnston, T. Maloney and S. Qawami, "Two Unusual HBM ESD Failure Mechanisms On A Mature CMOS Process", in *Proc. 15th EOS/ESD Symposium*, pp. 225–232, 1993.

[Juliano01B] P. Juliano and E. Rosenbaum, "A novel SCR macromodel for ESD circuit simulation", in *IEDM Tech. Digest*, 319–322, 2001.

[Krakauer98] D. Krakauer, "ESD Protection Clamp for Mixed Voltage I/O Stages Using nMOS Transistors", U.S. Patent 5,780,897, 1998.

[Kunz01] K. Kunz, C. Duvvury and H. Shichijo, "5-Volt Tolerant Fail-Safe ESD Solutions for a 0.18 um Logic CMOS Process", in *Proc. 23rd EOS/ESD Symposium*, pp. 12–21, 2001.

[Li97] T. Li, C. Tsai, E. Rosenbaum and S. Kang, "Substrate Resistance Modeling", in *Proc. 20th EOS/ESD Symposium*, pp. 281–289, 1997.

[Maloney88] T. Maloney, "Designing MOS Inputs and Output to avoid Oxide Failures in the Charged Device Model", in *Proc. 10th EOS/ESD Symposium*, pp. 220–227, 1988.

[Maloney95] T.J. Maloney and S. Dabral, "Novel Clamp Circuits for IC Power Supply Protection", in *Proc. 18th EOS/ESD Symposium*, pp. 1–12, 1995. Revised version published in *IEEE Trans. on Components, Packaging, and Manufacturing Technology, Part C*, **19**, 150–161, 1996.

[Maloney98] T.J. Maloney, "Designing Power Supply Clamps for Electrostatic Discharge Protection of Integrated Circuits", in *Microelectronics Reliability*, Vol. 38, pp. 1691–1703, 1998.

[Maloney99] T.J. Maloney and W. Kan, "Stacked pMOS Clamps for High Voltage Power Supply Protection", in *Proc. 22nd EOS/ESD Symposium*, pp. 70–77, 1999.

[Mergens99] M. Mergens, W. Wilkening, S. Mettler, H. Wolf and W. Fichtner, "Modular approach of a high current MOS compact model for circuit level ESD simulation", in *Proc. IRPS*, 167–178, 1999.

[Miller00] J.W. Miller, M.G. Khazhinsky and J.C. Weldon, "Engineering the Cascoded nMOS Output Buffer for Maximum V_{t1}", in *Proc. 22nd EOS/ESD Symposium*, pp. 308–317, 2000.

[Oh01] K. Oh, C. Duvvury and C. Salling, "Nonuniform bipolar conduction in single finger nMOS transistors and implications for deep submicron ESD design", in *Proc. IRPS*, 2001.

[Oh01A] K-H. Oh, C. Duvvury, C. Salling, K. Banerjee and R. Dutton, "Nonuniform bipolar conduction in single finger nMOS devices in a 0.13 um CMOS technology", in *Proc. IRPS*, 226–234, 2001.

[Polgreen89] T. Polgreen and A. Chatterjee, "Improving the ESD failure threshold of silicided nMOS output transistors by ensuring uniform current flow", in *Proc. 11th EOS/ESD Symposium*, pp. 167–174, 1989.

[Ramaswamy96] S. Ramaswamy, S. Kang, C. Duvvury and A. Amerasekera, "EOS/ESD Analysis of High-Density Logic Chips", in *Proc. EOS/ESD Symposium*, p. 288, 1996.

[Ramaswamy96A] S. Ramaswamy, S. Kang, C. Duvvury, A. Amerasekera and V. Reddy, "EOS/ESD Analysis of High-Density Logic Chips", in *Proc. 18th EOS/ESD Symposium*, pp. 285–290, 1996.

[Richier00] C. Richier, P. Salome, G. Mabboux, I. Zaza, A. Juge and P. Mortini, "Investigation on Different ESD Protection Strategies Devoted to 3.3 V RF Applications (2 GHz) in a 0.18 um CMOS Process", in *Proc. 22nd EOS/ESD Symposium*, pp. 251–259, 2000.

[Rubalcava86] A.L. Rubalcava, D. Stunkard and W.J. Roesch, "Electrostatic Discharge Effects on Gallium Arsenide Integrated Circuits", in *Proc. 8th EOS/ESD Symposium*, pp. 159–165, 1986.

[Russ01] C. Russ, M. Mergens, J. Armer, C. Jozwiak, G. Kolluri, L. Avery and K. Verhaege, "GGSCRs: GGNMOS Triggered Silicon Controlled Rectifiers for ESD Protection in Deep Submicron CMOS Processes", in *Proc. 23rd EOS/ESD Symposium*, pp. 22–31, 2001.

[Salling01] C. Salling, J. Hu, J. Wu, C. Duvvury, R. Cline and R. Pok, "Development of Substrate Pumped nMOS Protection for a 0.13 um Technology", in *Proc. 23rd EOS/ESD Symposium*, pp. 192–204, 2001.

[Scott86] D. Scott, G. Giles and J. Hall, "A lumped element model for simulation of ESD failures in silicided devices", in *Proc. 8th EOS/ESD Symposium*, pp. 41–47, 1986.

[Sinha98] S. Sinha, H. Swaminathan, G. Kadamati and C. Duvvury, "An automated Tool for Detecting ESD Design Errors", in *Proc. 20th EOS/ESD Symposium*, pp. 208–217, 1998.

[Ting01] L. Ting, C. Duvvury, O. Trevino, J. Schichl and T. Diep, "Integration of TLP Analysis for ESD Troubleshooting Structures", in *Proc. 23rd EOS/ESD Symposium*, pp. 445–452, 2001.

[Verhaege93] K. Verhaege, G. Groesseneken, H. Maes and J. Colinge "The ESD Protection Capability of SOI Snapback NMOSMETs: Mechanisms and Failure Modes", in *Proc. 15th EOS/ESD Symposium*, pp. 215–220, 1993.

[Voldman94A] S. Voldman, "ESD Protection in a Multirail Disconnected Power Grid and Mixed Voltage Interface Environments in a 0.5 and 0.25 μm Channel Length CMOS Technologies", in *Proc. 16th EOS/ESD Symposium*, pp. 125–134, 1994.

[Voldman94B] 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*, 277–280, 1994.

[Voldman95] S.H. Voldman, G. Gerosa, V.P. Gross, N. Dickson, S. Furkay and J. Slinkman, "Analysis of Snubber-Clamped Diode-String Mixed Voltage Interface ESD Protection Network for Advanced Microprocessors", in *Proc. 17th EOS/ESD Symposium*, pp. 43–61, 1995.

[Voldman98] M. Corsi, S. Voldman, F. Assaderaghi, J. Mandelman, L. Hsu and G. Shahidi, “Dynamic threshold body and gate-coupled gated SOI ESD protection networks”, in *J. of Electrostatics*, 239–255, 1998.

[Voldman99] S. Voldman, D. Hui, L. Warriner, D. Young, J. Howard, F. Assaderaghi and G. Shahidi, “Electrostatic Discharge (ESD) Protection in Silicon-on-Insulator (SOI) CMOS Technology with Aluminum and Copper Interconnects in Advanced Microprocessor Semiconductor Chips”, in *Proc. 21st EOS/ESD Symposium*, pp. 105–115, 1999.

[Voldman00] S. Voldman, N. Schmidt, R. Johnson, L. Lanzerotti, C. Brennan, J. Dunn and D. Harame, “Electrostatic Discharge Characterization of Epitaxial-Base Silicon-Germanium Heterojunction Bipolar Transistors”, in *Proc. 22nd EOS/ESD Symposium*, pp. 239–250, 2000.

8 Failure Modes, Reliability Issues, and Case Studies

Charvaka Duvvury, Ajith Amerasekera

8.1 INTRODUCTION

In order to design ICs with good ESD performance it is important that the main design and process parameters, which influence the behavior of the inputs and outputs under ESD conditions, are known. To determine these parameters and understand their significance, it is necessary to determine the electrical and physical failure modes caused by an ESD stress and the relevant physical mechanisms [Rountree85][Duvvury86]. Failure analysis (FA) is essential in the ESD design path. The first part of this chapter will look at some typical failure mechanisms and their electrical signatures in advanced CMOS processes.

In Chapter 5 and Chapter 6, the details of the various primary and secondary protection devices and how they can be combined to form effective protection circuits were presented and discussed. But the total circuit reliability is a function of numerous parameters when the protection devices are included. These will be described in the next part of this chapter with illustrative examples. An effective ESD protection circuit scheme for a full IC requires that the protection circuits also prevent damage to circuitry beyond the input and output buffers. As an example, for a particular stress combination of the input pin and the V_{dd} supply, the current can flow internally in the chip between the V_{dd} and V_{ss} connections leading to possible internal damage. This type of weakness must be detected by analysis of the failure mechanism involved before a solution is identified. These details are presented in the second part of the chapter with illustrative case studies. For the sake of clarity, some of the protection concepts are repeated here while discussing the reliability phenomena.

8.2 FAILURE MODE ANALYSIS

8.2.1 Failure Analysis Techniques

The identification of the failure locations and the type of failure is essential to the process of designing and debugging ESD protection circuits and solving ESD problems in existing circuits. In this section we will discuss briefly the failure analysis tools and techniques for location of the damage site and identifying the failure mode.

The first and easiest failure analysis tool for identification of failure location is Liquid Crystal Analysis. The chip (or wafer) is placed on a temperature-controlled chuck and the temperature is raised to about 50 °C depending on the type of liquid crystal used. The liquid crystal is applied to the chip, and the device is powered up. Figure 8.1 shows an example of a liquid analysis of an output buffer damaged by ESD. The dark region at the bottom right-hand side of the picture indicates the hot spot where the damage is located.

As Figure 8.1 shows, hot spots in the device denoting possible failure locations appear as dark regions detectable visually through a microscope. The dark regions extend over a large area and this method does not reveal the exact location of the failure. However, it allows the FA engineer to identify the area that he needs to focus on in the more detailed analysis that follows. Typically, the liquid crystal method can detect failures on the order of 100 μA or greater, although an experienced engineer can increase the sensitivity to the order of a μA through manipulation of the temperature and the liquid crystal itself. Two major advantages of this technique are its low cost and that it is nondestructive.

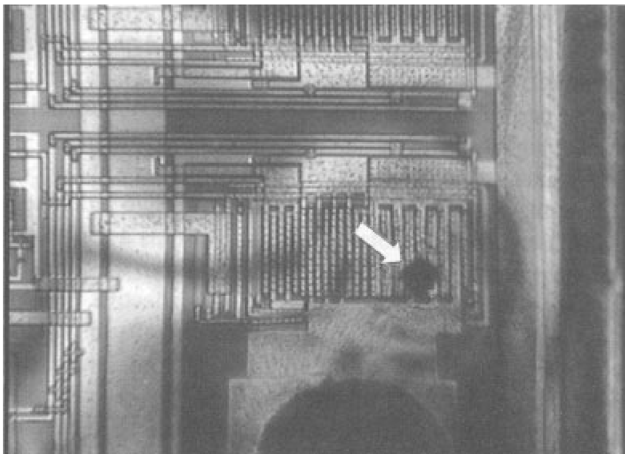


Figure 8.1 Liquid Crystal Analysis of ESD damage in a CMOS output buffer

An FA tool that is gaining increasing popularity is Light Emission Microscopy, also known as Photon Emission Microscopy or EMMI. This technique uses photon detectors to identify regions of excessive light emission which signify high current and reverse-biased breakdown [Wills88][Hannemann90]. The device is placed under a microscope and powered up. The image intensifier in the equipment detects the emitted photons and allows visual observation of the emitted light. Figure 8.2 shows a typical failure location in an input protection circuit identified by photoemission. The arrow points to the light emission from the damaged region. The method allows the failure location to be determined with almost pinpoint accuracy thereby reducing the analysis time.

The peak of the detector sensitivity is at around 3 eV, which makes it ideal for observing photons emitted by reverse-biased junctions. Recent advances in the technology have also made it possible to observe Infrared emission owing to thermal effects. The technique does not require any temperature control, and like the liquid crystal method it is nondestructive. However, the equipment is expensive relative to the liquid crystal. It is also possible to use this technique to observe the real-time high current behavior of an ESD circuit, which is excellent for understanding the basic phenomena involved [Hannemann90][Amerasekera90][Cavone94]. It must be emphasized that this tool is invaluable in the study of ESD phenomena in ICs and in the design debugging process for ICs [Duvvury92B].

Once the failure location has been determined, the device is deprocessed. Deprocessing consists of a series of etches, which remove the levels of passivation, interlevel oxide, and metal until the silicon is exposed. It is possible to deprocess down to the polysilicon gate before submitting the device to Scanning Electron

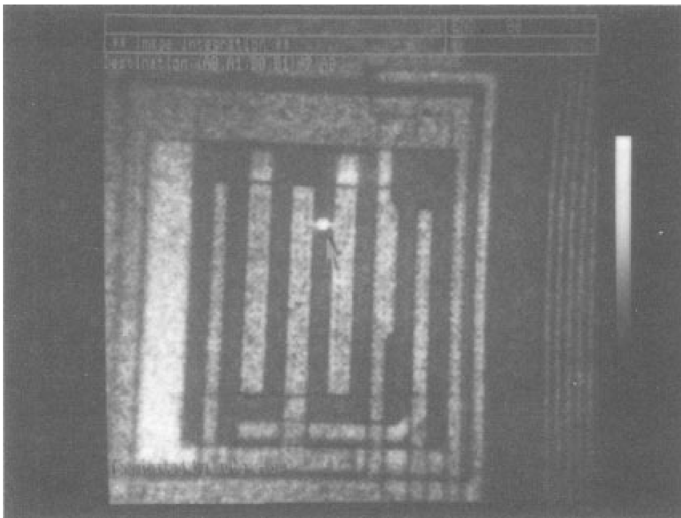


Figure 8.2 Photonemission analysis of ESD damage in a CMOS output buffer

Microscopy (SEM) to examine the damage area. The etch processes tend to remove thermally damaged regions in the silicon leaving the notches or cavities shown in Figures 8.6 and 8.7. Melt regions, which extend between the drain and source of the transistor, can cause breaks in the oxide and even in the polysilicon (Figure 8.9) and should not be mistaken for oxide breakdown itself. Identifying oxide breakdown is more difficult, and requires very sensitive etches with care being taken not to etch through the oxide and remove evidence of damage [Colvin93]. In fact, one of the important issues in deprocessing for ESD failures is that the damage can be very close to the surface of the silicon. Hence, the etch processes must be performed in very small increments to ensure that the damaged region is not etched away.

8.2.2 Electrical Characteristics After Damage

The ESD failure threshold of an IC is determined by monitoring the leakage current, I_{leak}, at the stressed pin. Depending on the failure criterion selected, failure is defined by a change in the leakage current, ΔI_{leak}. It has been shown that ΔI_{leak} in circuits fabricated in an advanced CMOS process can be in one of five categories [Amerasekera92].

(a) No increase in leakage current, $\Delta I_{\text{leak}} = 0$
(b) $\Delta I_{\text{leak}} < 10\ \mu\text{A}$
(c) $1\ \mu\text{A} < \Delta I_{\text{leak}} < 100\ \mu\text{A}$ (note the overlap with 2)
(d) $100\ \mu\text{A} < \Delta I_{\text{leak}} < 1\ \text{mA}$
(e) $\Delta I_{\text{leak}} > 1\ \text{mA}$

In conjunction with detailed failure analysis, it has been shown that the above categories can be related to typical failure modes (see Section 8.2.3). Correlations have been shown between the ESD failure threshold and the median poststress ΔI_{leak} [Amerasekera90]. Figure 8.3 shows the percentage of failed pins as a function of the ESD stress voltage. These pins were each connected to the output buffer of an IC and each data point is based on the results from 80 output pins. It is seen that although 90% of failures occur above 2000 V, a small percentage fail at $\leq$1000 V. The distribution of the median of the poststress leakage current with applied ESD voltage is shown in Figure 8.4.

A direct correlation is observed to the ESD failure distribution. A lower median ΔI_{leak} correlates with lower ESD stress voltages. From the distribution of the poststress electrical characteristic for a given process and design, the main failure mode can be determined.

Device failures due to deviations in the manufacturing process, that is, *freak* failures, can be identified by monitoring the poststress leakage currents. The cause of some of these failures may be self-correcting and not require any further action on the part of the product engineer, while others may indicate that the protection design is too process sensitive. By characterization of the electrical signature associated

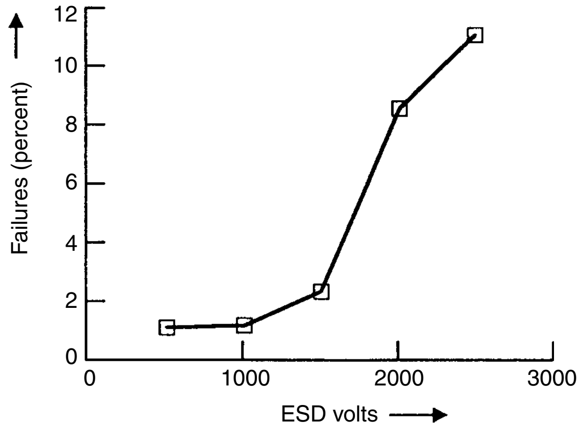


Figure 8.3 Percentage of failed output pins as a function of the applied ESD voltage in a full CMOS IC

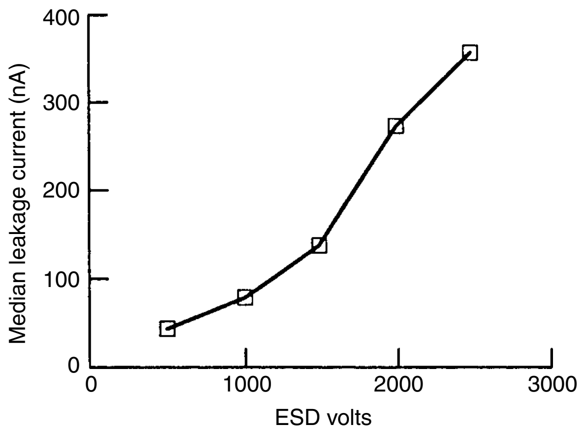


Figure 8.4 Poststress median leakage current as a function of applied ESD stress

with ESD damage, one is, therefore, able to maintain a control on the process-related ESD behavior and improve the capability to solve ESD problems that arise during development and production. An example of this is shown in Figure 8.5. The histogram shows that with standard processing (STD) the main failure mode is that of category 1, while the addition of a $5 \times 10^{14}/\text{cm}^2$ phosphorus implant shifts the main failure mode to category 5 [Amerasekera92]. The respective ESD failure thresholds moved from a minimum of 500 V for the STD devices to >2000 V for the devices with the additional phosphorus implant as expected from the results of Figures 8.3 and 8.4.

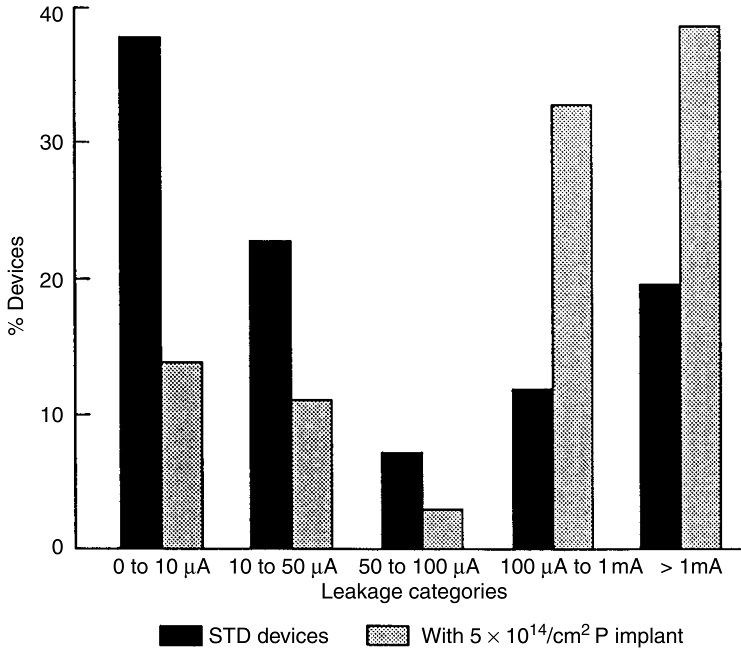


Figure 8.5 Histogram of the distribution of poststress leakage currents for nMOS transistors after ESD stress at 2 kV

8.2.3 Physical Analysis of Failure Modes

Failure mode analysis has shown that ESD-type failures fall into one of the five categories [Amerasekera92].

(a) A common failure mode observed in field oxide devices used as the primary element in input protection circuits is that of holes or notching in the silicon at the field oxide interface as shown in Figure 8.6. The damage is observed at the diffusion connected to the pad and is at a depth related to the bottom of the n^+-diffusion region in the silicon. For this failure mode 100 pA $< \Delta I_{\mathrm{leak}} <$ 10 μA depending on the stress voltage (500 V to 2 kV).

(b) Damage at the diffusion edge is observed in nMOS transistors used in output buffers or input/output protection as indicated by the arrows in Figures 8.7 and 8.8. $\Delta I_{\mathrm{leak}} <$ 10 μA and the ESD stress voltages range from <500 to 2000 V for this failure mode. This failure mode is most often observed in LDD processes.

(c) Large melt regions owing to current filamentation between the n^+ diffusion regions is observed in nMOS transistors (Figure 8.9) and in field oxide devices (Figure 8.10). This is the classic form of ESD damage found in both non-LDD processes and LDD processes. It is the most common failure mode in

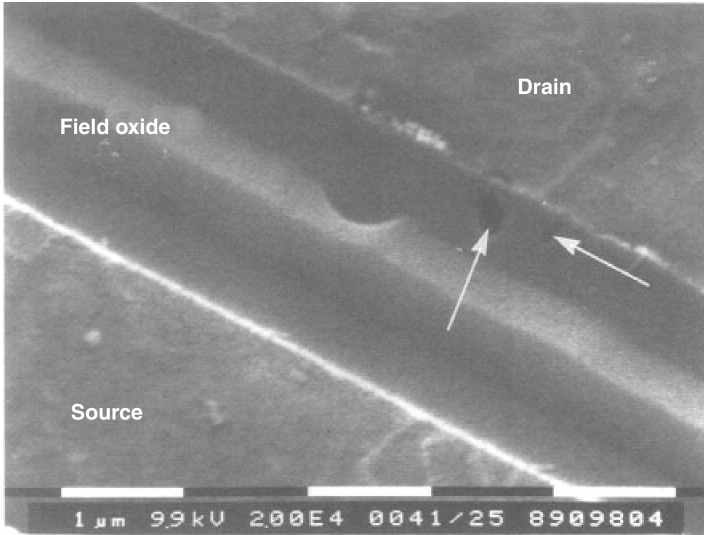


Figure 8.6 SEM photograph showing holes at the silicon to field oxide interface in the drain diffusion of an FOD protection device

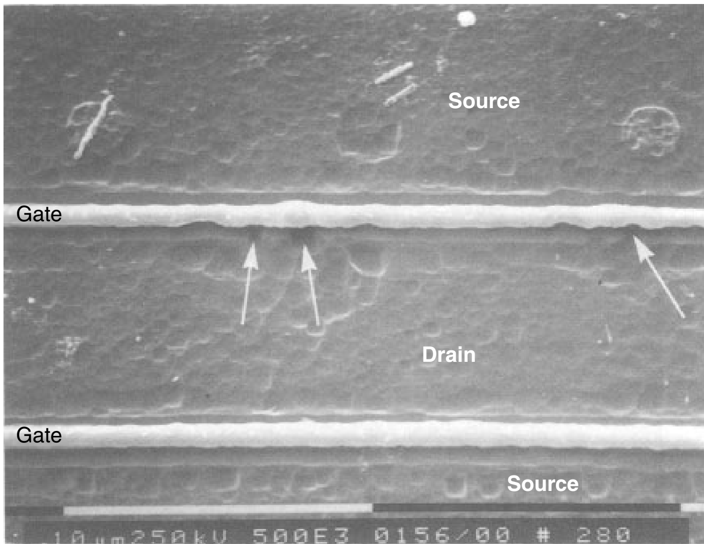


Figure 8.7 SEM photograph of an nMOS transistos in an output buffer showing damage at the drain/gate diffusion edge

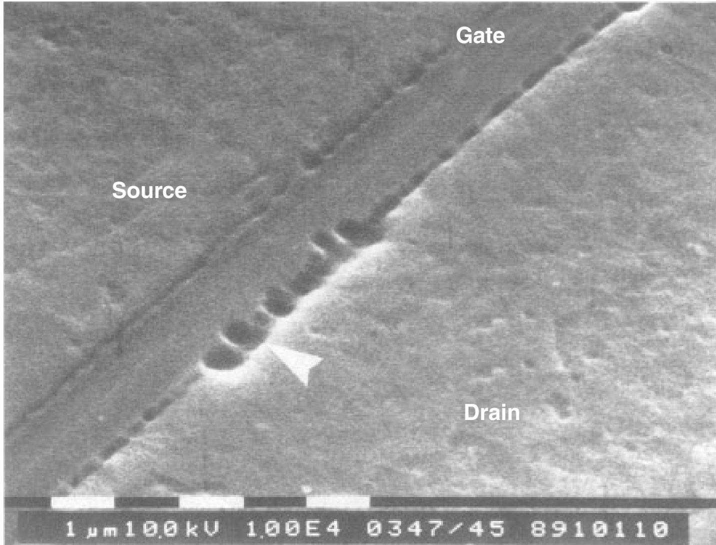


Figure 8.8 SEM photograph of drain/gate diffusion edge damage

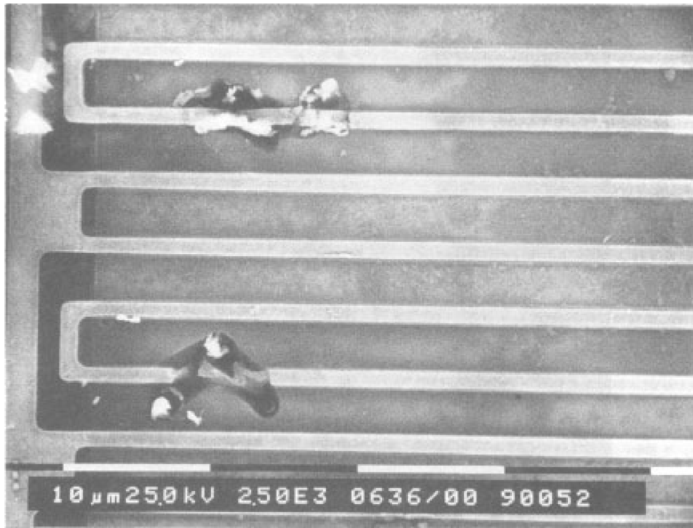


Figure 8.9 SEM photograph of silicon melting due to current filamentation in an nMOS output transistor

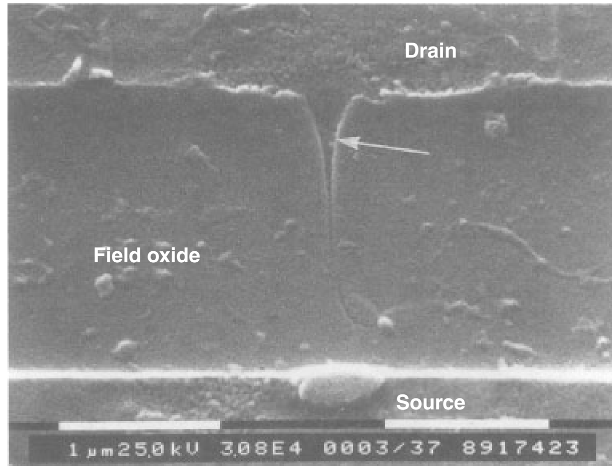


Figure 8.10 SEM photograph of a damage due to a polycrystalline filament in a thick oxide device

silicided processes. ΔI_{leak} can be between 1 μA and 100 μA, and in extreme cases (especially) with silicides) ΔI_{leak} can be >1 mA. The ESD stress levels are very high for this type of damage, usually in excess of 2000 V. However, if the protection or output buffer design is weak then this failure mode can be observed at much lower ESD levels.

(d) Contact spiking in the drain contact region as shown in Figure 8.11 has been observed in both nonsilicided [Rountree85] and silicided [Amerasekera92] processes. ΔI_{leak} is usually >1 mA. ESD stress levels are usually in excess of 2000 V before this failure mode is observed, provided the contact to gate spacing (see Chapter 4) has been properly designed. If contact spiking is observed at lower ESD levels, it indicates that the contact-to-gate spacing is not optimized (it could be too small or too large!).

(e) Gate oxide damage in Figure 8.12 with $\Delta I_{\text{leak}} \approx 1$ mA is usually observed when the protection circuit has not been properly designed especially in technologies with gate oxides <175 Å thick.

In addition, examples of polysilicon filaments at the drain edge of the nMOS transistors have been shown in Chapter 6 (see Figure 6.8). The polysilicon filaments are formed by the migration of polysilicon between the hot junction edge and the gate, aided by the high electric field between the drain and the gate and the temperature gradient (e.g. [Kiefer93]).

Failure modes (a) and (b) are *soft* failure types, in which the leakage current changes with additional stressing. The effect is seen in a decrease in the avalanche breakdown voltage, typically about 1 or 2 V with every additional ESD stress [Amerasekera90][Kuper93]. The poststress leakage currents have also been observed to decrease after a period of time (24 h), as well as when subjected

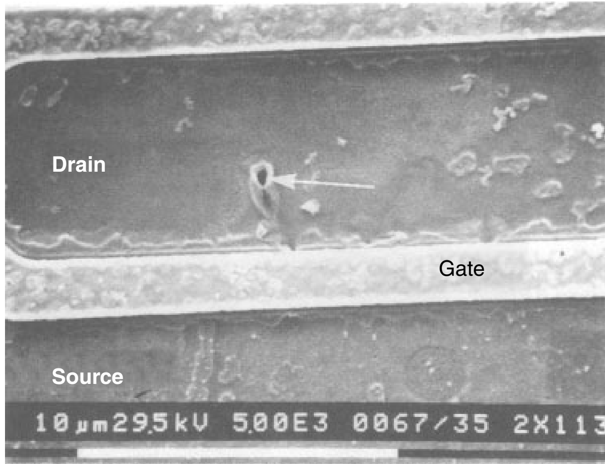


Figure 8.11 SEM photograph of a contact hole in the drain diffusion of an output transistor showing contact spiking

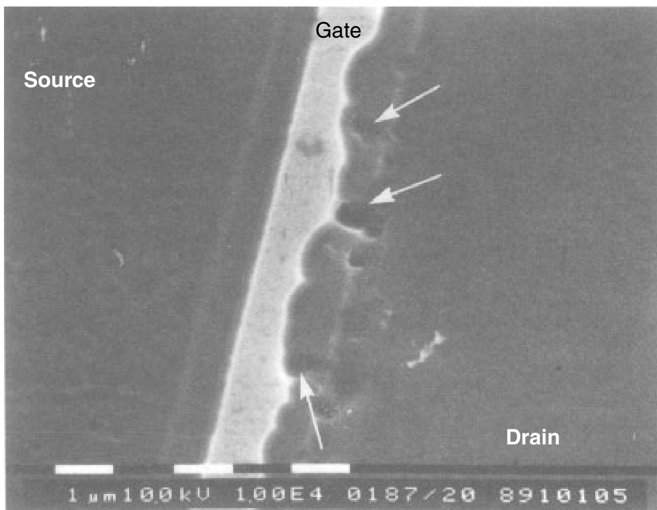


Figure 8.12 SEM photograph of an nMOS transistor showing gate oxide damage

to thermal anneal, but they can never be completely eliminated. This indicates that permanent damage has occurred in the junction rather than the occurrence of charge trapping or an increase in surface states. The size and location of the notches in Figures 8.6 and 8.7 indicates that the damage is in the form of polycrystalline filaments formed thermally at the diffusion edge. Damage locations are usually in the FOD or FPD nMOS devices. pMOS transistors connected in parallel

with the nMOS transistor do not normally show damage because the triggering of the parasitic *npn* is more efficient than that of the parasitic *pnp*. The *npn*, therefore, triggers and shunts all the ESD current and eventually fails, thereby actually protecting the pMOS devices in the circuit. There are conditions under which damage may be observed in the pMOS device as will be discussed later in this chapter.

8.3 RELIABILITY AND PERFORMANCE CONSIDERATIONS

One of the main considerations when designing an ESD protection circuit is to ensure that its incorporation does not alter the intended device performance specifications. For certain pins, the specific function of the pin must also be taken into consideration. In these cases, the input design must be modified so that it will not interfere with the pin applications.

During the qualification of any VLSI chip for production, the high temperature/voltage burn-in test must be successfully met. If the poststress characterization results in either input low (IIL) or input high (IIH) leakage failures, the input protection circuits are usually the cause. The problem is aggravated even more in devices fabricated in advanced process technologies. Such leakage problems are usually associated with nonoptimized input protection circuit designs. A detailed analysis would normally reveal the design parameters that need to be optimized. Under some extraordinary circumstances, the problem may not even be associated with the input design but is caused by an interaction of process and protection circuit layout. In such cases, this type of susceptibility to process variations can be corrected by changing the layout while retaining the original protection design.

Here, the various input protection design considerations for good ESD protection, as well as the issues relating to post-burn-in performance reliability, will be considered. The input protection schemes for advanced CMOS technologies and their operation are discussed. The optimum protection circuit design issues for consistent ESD performance are also examined. The post-burn-in reliability issues include both bakeable and nonbakeable leakage failures. The latter type is associated with the ESD protection circuit and will be discussed here. The input design for pins with special applications will be covered with an example. The post-burn-in bakeable leakage phenomena and the solutions to reduce the leakage are also discussed.

As several different aspects are involved in achieving input protection designs for overall chip reliability, the basic approach in this chapter is to cite the different design experiences and discuss how problems and weaknesses were corrected.

It should be noted here that the protection circuit failures described in this chapter are case studies and actually occurred in product chips. Solving ESD problems involves iterative cycles. For that reason, we have walked through the development of a protection circuit to illustrate the wrong approaches and the corrective actions.

Ineffective protection circuits are deliberately considered to show how failures can occur. The main objective in this chapter is to show how potential ESD failure can occur at the pin or in the internal circuits and the use of failure analysis to identify the causes and make improvements for higher ESD performance and better functional reliability in the IC.

8.4 ADVANCED CMOS INPUT PROTECTION

As mentioned previously (Chapter 6), the thick-field device has served as the primary protection circuit for a few generations of nMOS technologies [Keller81] [Hulett81][Duvvury83]. CMOS is now the mainstream of most VLSI chips and the *n*-channel thick-field device (or FOD) is still effective as long as abrupt junction processes are used. However, for improved reliability and performance, graded junction or lightly doped drain (LDD) transistors with/without silicided diffusions have become important. We also considered the advanced process options and described that the thick-field device is no longer effective and a lateral SCR device is a good option [McPhee86][Rountree88]. In fact, this lateral SCR in a CMOS process can be effective even if LDD junctions and silicided diffusions are not present. An overall protection circuit with the lateral SCR is again first described in the following text.

Consider the cross section of the LSCR repeated in Figure 8.13. The lateral SCR consists of p^+ and n^+ diffused regions in an *n*-well on *p*-substrate connected to the input pad and an n^+ diffused region on *p*-substrate connected to a V_{SS} common terminal. Thus, a *pnp* is formed with the input p^+ as emitter, the *n*-well as base and the *p*-substrate as collector. Similarly, an *npn* is formed with the *n*-well as collector, *p*-substrate as base, and the n^+ connected to the V_{SS} as emitter. During normal operation the *n*-well and the emitter of the *pnp* are tied to the same potential

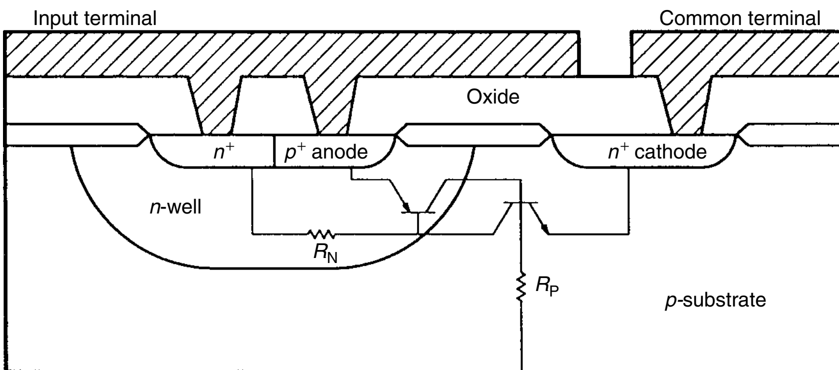


Figure 8.13 Cross section of the LSCR device

and the *pnp* does not turn on. The details of the LSCR trigger and operation during the ESD event have been discussed in Chapter 4. When high current conduction takes place, the total anode-to-cathode potential is only about 1 to 2 V. In general, the holding voltage of the LSCR is determined by the p-substrate resistance R_p and the anode-to-cathode spacing. The switching or trigger voltage of the LSCR is mainly determined by the n-well overlap of the p^+ anode. A typical value for this parameter varies from 3 μm (for 1.0-μm technologies) to 5 μm (for 2.0-μm technologies). The corresponding trigger voltage can vary from 50 to 70 V. Thus, the SCR devices trigger at higher voltages than the thick-field devices used in the previous protection circuits.

Just as in the case of the thick-field primary protection design, a field plated diode (FPD) and a diffusion resistor are needed to provide the overall protection. The schematic for this is shown in Figure 8.14. The resistor can be either p^+ or n^+ diffusion and both are illustrated in the figure. During an ESD event, the FPD initially clamps the voltage and protects the input buffer gate oxide, while the diffusion resistor limits current to the FPD. With the FPD operating in the breakdown mode as a lateral *npn*, the $I \times R$ drop across the resistor increases the voltage at the pad and eventually leads to the triggering of the SCR device. The triggering voltage or triggering current is determined by the LSCR layout and the process. As the SCR devices trigger at a higher voltage than the thick-field devices, a more severe stress is placed on the FPD. In fact, the trigger level or the trigger current is basically determined by the secondary protection. This was originally discussed by Rountree [Rountree88]. A more thorough analysis is presented here.

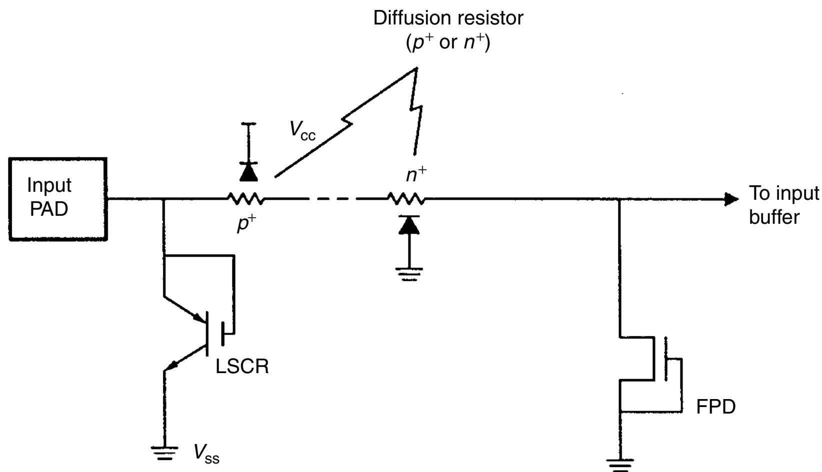


Figure 8.14 Input protection schematic with the LSCR as the primary protection and a field plate diode with a diffusion resistor as the secondary protection. The diffusion can either be n^+ or p^+ as shown

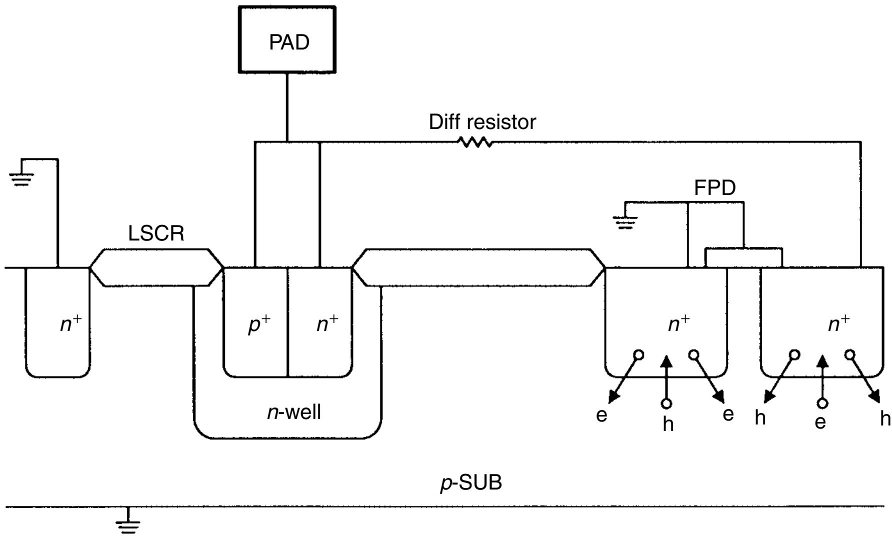


Figure 8.15 Composite input protection circuit using an SCR and an FPD. Holes injected into the substrate due to avalanche breakdown in the drain of the FPD can aid in triggering the SCR device

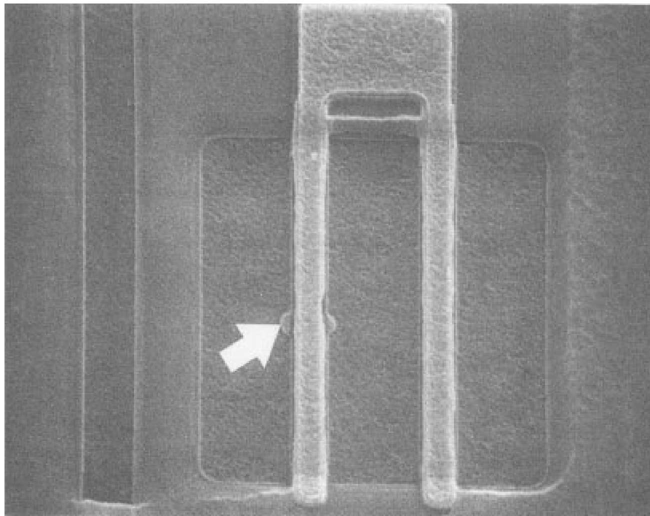


Figure 8.16 Damage to the FPD due to positive polarity ESD stress between pad and V_{SS}. This damage mode of poly gate melt filament shorting to drain occurred since the SCR did not trigger properly

Consider Figure 8.15 in which the cross section of an input protection with the LSCR is shown. When the FPD goes into the breakdown mode, the hole current in the substrate forward-biases the emitter-base junction of the *npn* in the SCR. This causes the SCR to trigger at a lower voltage than if the SCR triggers itself through avalanche breakdown of the n-well to substrate junction. Thus, the interaction with the FPD is important. In inefficient designs, the SCR will not fire properly and often cause failure at the FPD. An example of this is illustrated in Figure 8.16.

8.5 OPTIMIZING THE INPUT PROTECTION SCHEME

The importance of the proper design for secondary protection is illustrated through an example in this section. Detailed analyses of this type are necessary to understand the input protection operation and the correct measures needed to improve its efficiency.

An input protection design with the LSCR is used as an example. The process technology for this structure is 1-μm CMOS with LDD drain/source junctions and silicided diffusions. A cross-section of this protection circuit is shown in Figure 8.17. Note that in this scheme the isolation resistor is of p^+ diffusion. In the same n-well as the resistor, an n^+ diffusion is connected to V_{cc} to form a lateral diode. The diode function is to suppress the voltage levels above V_{cc}, which might occur at the input. During routine testing, this protection failed below 2 kV of stress. It was suspected that the SCR was not triggering and its effectiveness

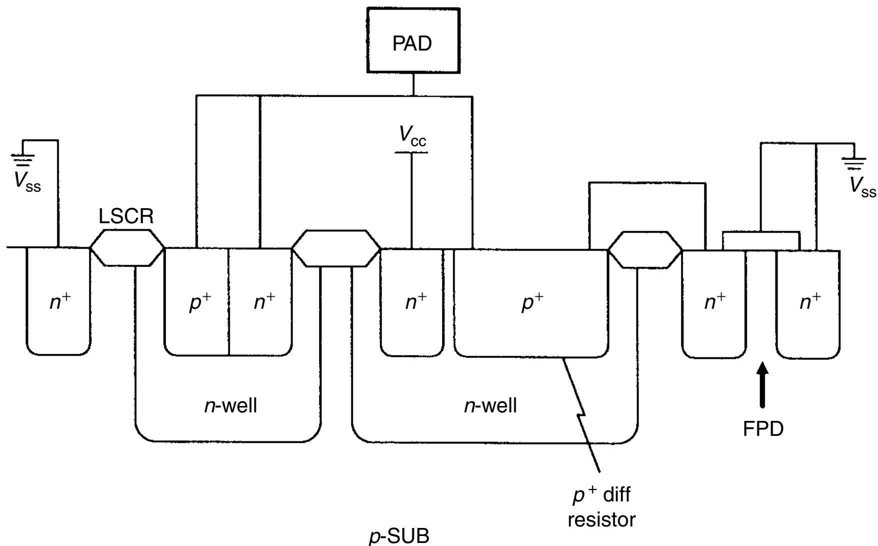


Figure 8.17 Composite input protection scheme with LSCR, p^+ diffusion resistor/diode, and FPD

Table 8.1 Input protection performance for processes with 900 Å and 600 Å titanium

Process	Resistor value	SCR trigger current (mA)	V_f (HBM) $+/V_{ss}$	V_f (HBM) $+/V_{cc}$
Standard (900 Å-Ti)	90 Ω	180	1250 ± 100 V	1250 ± 100 V
Experimental (600 Å-Ti)	180 Ω	150	>6 kV	>6 kV

was tested by disconnecting it from the protection circuit using a laser cutter. The FPD plus resistor combination is then stressed on their own. Step-stressing before and after the laser cut revealed the same average protection level at 1200 V for the Human Body Model (HBM) test with respect to V_{ss}. The results are shown in Table 8.1 and are indicated as standard process.

In a subsequent test when the starting stress level was 3 kV, the same protection circuits with the SCR left intact passed 6 kV! Therefore, these same protection circuits that failed 2 kV passed higher stress levels. To understand these results, a more detailed study was done (with the SCR intact) by applying positive stress with respect to V_{ss} from both the low and high ends. For the low-end stressing, the starting voltage was 100 V with incremental steps of 100 V and for the high-end stressing, the starting voltage was 4 kV with decreasing steps of 100 V. The results are shown in Figure 8.18. Note that the failures occurred between 1100 and 1300 V for the low-end stressing and between 1900 and 2300 V for the high-end stressing. Thus, there is a window between 1100 and 2300 V where the SCR does not trigger and provide protection. Below 1100 V, the protection is provided by the secondary device. A similar analysis is done for stress with respect to V_{cc} and the window in this case was 1900 to 2300 V. These results are shown in Figure 8.19. For successful triggering of the SCR, the minimum protection required from the secondary stage is 2300 V and is consistent from both Figures 8.18 and 8.19. Above 2300 V the SCR triggers and shunts most of the current, thus protecting the FPD/resistor combination. Devices of this type would then show a false failure level for a go/no-go type of stress test depending on the chosen stress level. For example, these might be guaranteed for 4-kV ESD stress but would be susceptible to common 2-kV stress levels with human handling, such as during a burn-in test setup sequence, and end up as input leakage failures. Of course, if the handling stress levels are much higher (>3 kV) there would be no problem.

The failure analysis and the corresponding leakage failures were characterized for these devices with the results shown in Figure 8.18. For the low-end failures the damages were in the FPD, similar to Figure 8.16. The leakage mechanism is IIH and is indicated as IIH1 in Figure 8.20. On the other hand, for the high-end stress failures, the damage was both in the p^+ resistor and the FPD and leads to IIH and IIL leakage failures. These are IIH1, IIH2, and IIL in Figure 8.20. The damage to the p^+ resistor, which was shorted to the n-well, would lead to IIL leakage because during this test, V_{cc} was high and 0 V was applied at the pad. Referring to Figure 8.20, the shorted p^+ resistor (to n-well) can also short to the nearest p^+ substrate contact causing the IIH2 failure. The stress current path and

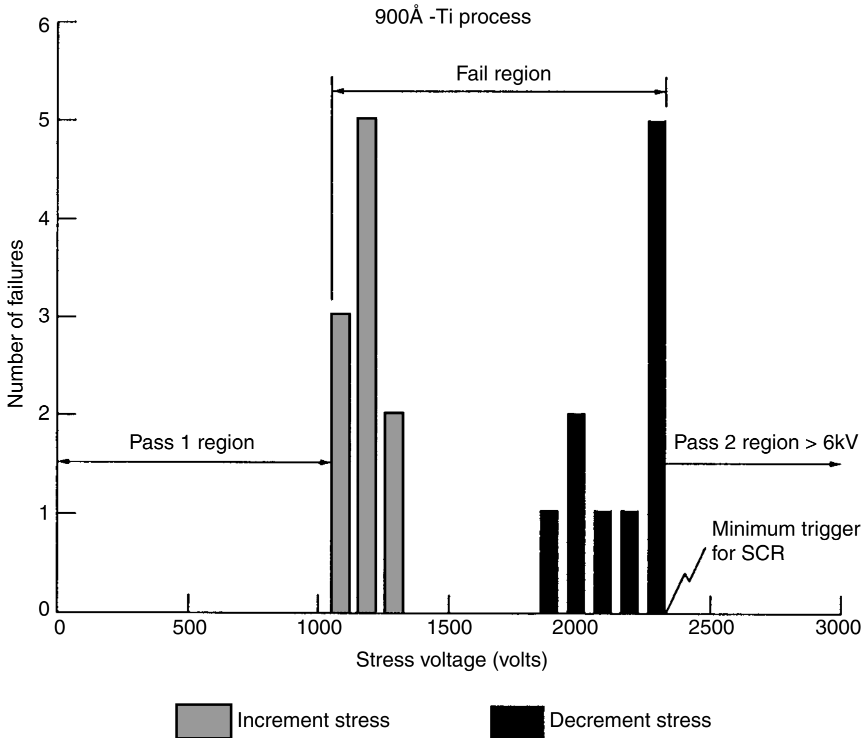


Figure 8.18 Input ESD interim failures for positive ESD stress to V_{SS}. Note that the failures occur in the region where the SCR trigger is critical

the corresponding physical damage are shown in Figure 8.21. Thus, removing this substrate contact from the protection area becomes important. Similar to the V_{SS} stress case, V_{CC} stress failures were also analyzed. As expected, in this case both the low- and high-end failures occurred at the p^+ resistor and resulted in IIL failures. This makes sense since the FPD was not involved for positive voltage stress to V_{CC}. Thus, with the SCR not triggering properly, the p^+ to n-well diode was easily damaged. The actual stress current path and the observed physical damage for the positive voltage stress to V_{CC} are shown in Figure 8.22. Note that once the p^+ to n-well diode was damaged, the stress current shorts the V_{CC} contact to the V_{SS} contact and can additionally lead to the IIH2 leakage shown in Figure 8.17.

The above analysis has shown that potential IIL/IIH leakage failures could occur if the input protection is not effective for all stress voltage levels. As an attempt to improve the protection circuit, the same protection circuits were run with a thinner silicide process. The deposited Titanium (Ti) for this process was reduced from the standard 900 Å to 600 Å. The results showed that there was no window where the SCR did not fire properly. That is, step stress results showed no failures to above 6 kV (Table 8.1). The same is true for decreasing stress steps. This must mean

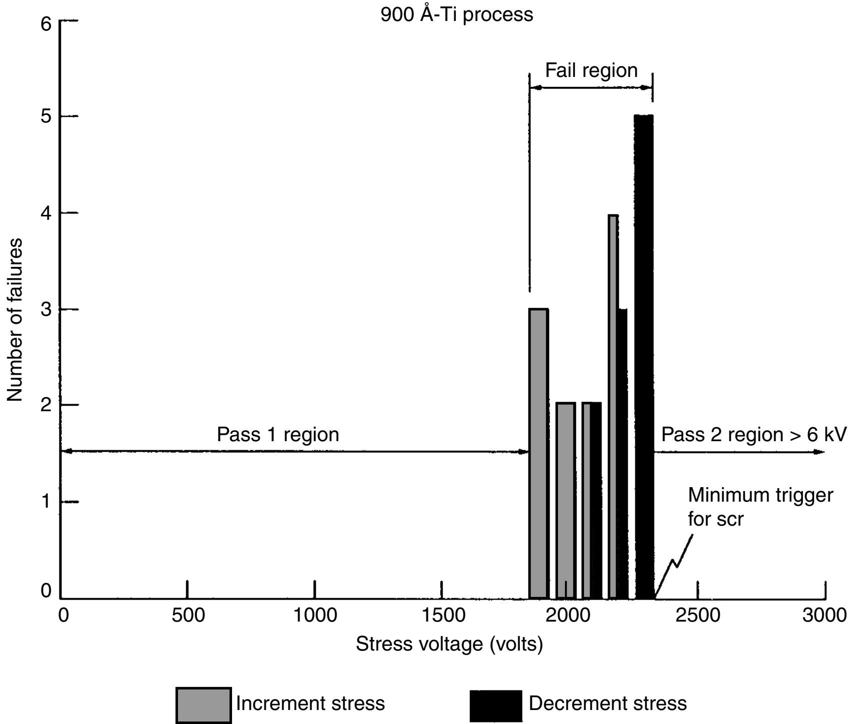


Figure 8.19 Input ESD interim failures for positive ESD stress to V_{CC}. Note that the failures occur in the region where the SCR trigger is critical

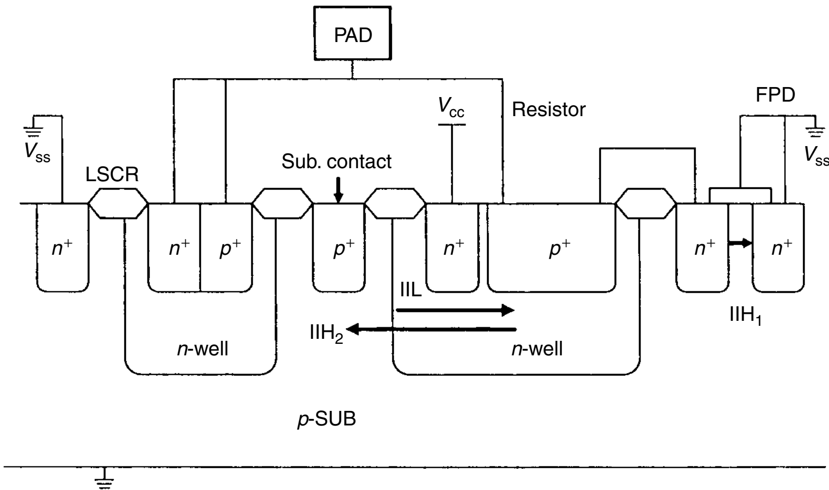


Figure 8.20 Input IIL/IIH leakage mechanisms that are not bakeable

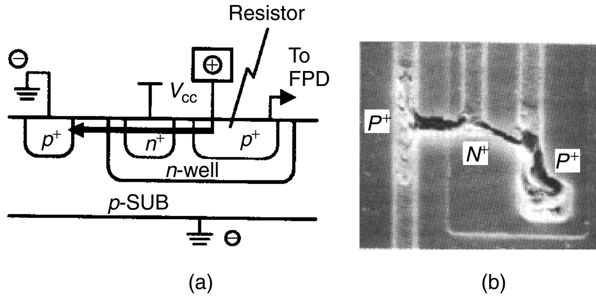


Figure 8.21 Isolation stage damage for low-end stress failure with positive ESD stress to V_{SS}. Note the damage is to the lateral *pnp*

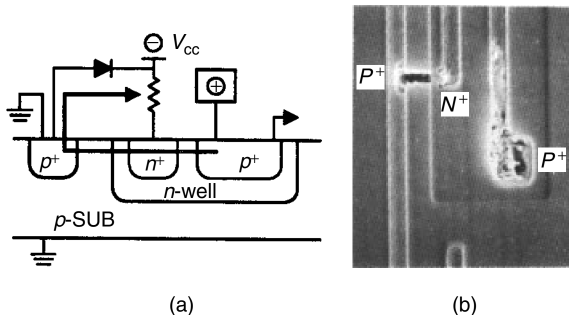


Figure 8.22 Isolation stage failure for low-end stress failure with positive ESD stress to V_{CC}. The damage is to the p^+/n-well diode first and then from the V_{CC} contact to the V_{SS} contact

that the isolation stage performance has improved with the thinner silicide process. (Similar to the V_{SS} stress, V_{CC} stress also shows good performance to >6 kV). This improvement in the isolation stage performance was verified by detaching the SCR using a laser cutter for both 900 Å and 600 Å Ti thickness variations and measuring the failure thresholds of the isolation stage alone. These results are shown in Figure 8.23. Note that with the thinner silicide, there is an apparent improvement in the isolation stage protection. This is because with the thinner silicide the isolation resistor value approximately doubled. The minimum ESD level at which the secondary stage was damaged was determined to be 1100 V from the analysis of Figure 8.18. The higher resistance, therefore, requires less current to raise the pad voltage to the trigger level for the SCR allowing the full protection to be effective at ESD levels <1100 V.

As shown, good input protection levels can be realized in this case by either altering the process or increasing the isolation stage resistor design value to an adequate level. The FPD device width and the resistor value are the main parameters that determine the secondary protection level and consequently the successful operation

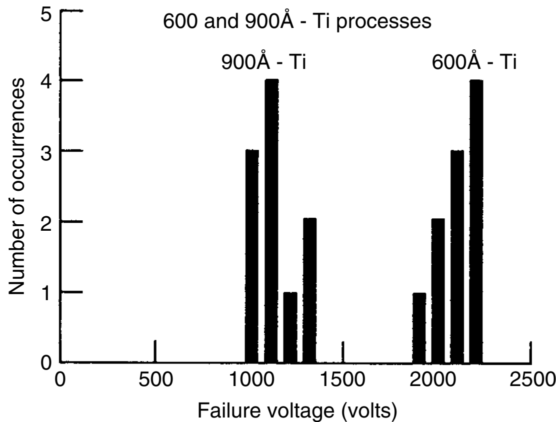


Figure 8.23 Isolation stage performance with the SCR cutoff for 900 Å and 600 Å Titanium processes. The stress was $+/V_{SS}$

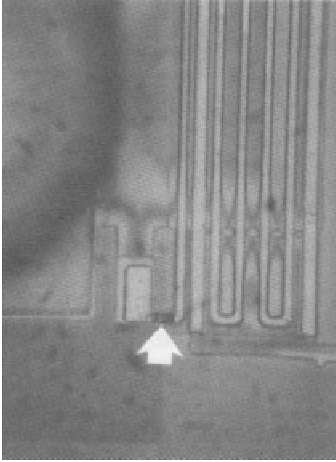
of the SCR. The channel length of the FPD is not found to play any major role and thus a safe minimum length can be chosen without concerns for any subthreshold leakage problems. The FPD channel length, however, can be important for some special application pins. This issue is discussed in Section 8.6.

The protection circuit reliability presented here considered the ineffective triggering with an LSCR. It should be noted that converting the LSCR to an MLSCR, which was described in Chapter 6, could alleviate some of the reliability problems discussed here. This is because the trigger voltage of the MLSCR is lower and is closer to the on-voltage of the lateral *npn* associated with the FPD. Hence, using the MLSCR in the total input protection scheme will provide a more consistent ESD protection circuit.

Some circuit applications such as in automotive or industrial ICs cannot tolerate any accidental triggering of an SCR due to high voltage spikes during normal operation. In these cases the higher triggering LSCR scheme is more desirable but the protection circuit reliability as described here should be carefully evaluated and characterized to ensure optimum ESD protection and reliability.

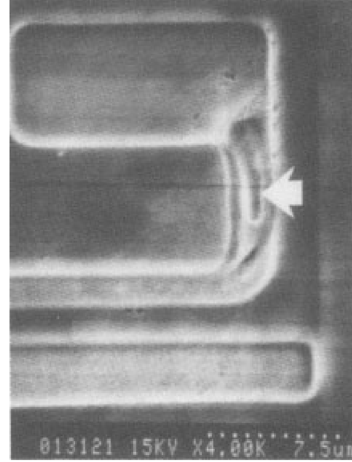
In another example of input protection design, the layout of the resistor itself can become important for chip reliability. During the implementation of the above protection scheme for silicide technologies a long resistor is required. But because of space constrictions at the bond pad, the resistor must be invariably bent to achieve about 200 Ω. This type of layout could also cause unexpected problems as will be discussed in the following text.

The protection level for a circuit with a bent resistor, as measured using a commercial HBM tester yielded a high level (>6 kV) of protection. But the chips had a high rate of burn-in loss with IIL and IIH failures. These were found to be nonbakeable and were most likely a result of protection circuit damage. The failure analysis indicated damage to the p^+ resistor as shown in Figure 8.24(a).



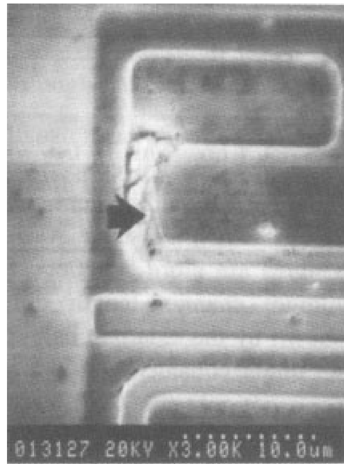
Silicide damage at bent portion of resistor

(a)



Damage internal to resistor

(b)



Damage internal/external to resistor

(c)

Figure 8.24 (a) Damage to the p^+ resistor at the bend in the layout. (b) p^+ resistor damage internal to the resistor, which does not result in IIL or IIH leakages. (c) p^+ resistor damage internal and external to the resistor, which results in both IIL and IIH leakages

Any damage to the p^+ resistor can cause leakage of the reverse-biased p^+/n-well diode (Figure 8.20) for IIL failures, or if the damage extends to the n-well/p-substrate junction it can cause an IIH failure. Hence, these burn-in failures are not surprising. This apparent discrepancy between the tester results and burn-in failures is described in the following text.

Analysis of the circuit revealed that at low ESD stress levels (approximately 1 kV) the current densities in the silicided regions of the bent portions of p^+ resistor were very high. The silicide was heated beyond its eutectic temperature (i.e. it ceased to exist as a conductive layer), which effectively increased the resistance of the diffusion.

The nature of the thermal process is such that no damage was seen outside the resistor (Figure 8.24(b)). Thus, once this self-protection was formed, no degradation of the protection level was seen. But if the starting ESD stress level is 2 or 3 kV, the diode degrades and IIL/IIH failures occur because the damage extends beyond the p^+ resistor surface as shown in Figure 8.24(c). This correlated with the post-burn-in failures where poor handling techniques combined with HBM stress levels of 2 kV or above caused the leakage failures. The solution in this case was first to improve the handling procedures for the ICs during burn-in. A long-term solution was implemented to redesign the protection circuit with a different layout that avoids the bent resistor.

8.6 DESIGNS FOR SPECIAL APPLICATIONS

The design and reliability issues for special applications can be understood by considering actual product chips as they go through different process cycles. For several reliability and functionality factors the process is sometimes changed. In the example given here the process is changed from LDD junctions to abrupt junctions. The reader must keep this in mind while understanding how ESD circuits can cause reliability problems.

The input protection circuits are usually optimized for the best possible performance without any leakage concerns. It is usually recommended that the channel length of the FPD should be kept at a minimum since this will improve the performance of the lateral *npn* transistor. Also, it has been reported [Maloney88] that shorter channel lengths are more beneficial for the Charged Device Model stress. However, a minimum channel length might cause problems for certain pin applications, as shown by the example in the text that follows.

The specific pin under consideration is a high voltage pin (greater than V_{cc}). A high voltage buffer is not only able to differentiate between the normal logic ‘0’ and ‘1’ states, but it is also able to detect a third “HV” (high voltage) level. This feature is used in the example microcontroller chip to place the device into special operating or test modes without dedicating a pin for that purpose. This is extremely useful for in-factory testing or device emulation without compromising the need to minimize the number of input/output (I/O) pins. A typical circuit to

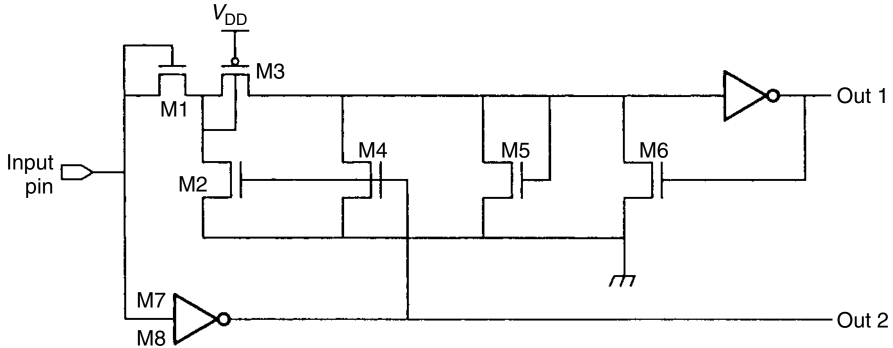


Figure 8.25 Tri-level detection circuit for the HV pin

Table 8.2 Decode logic stages for the tri-level detector circuit

Input pin (volts)	Out 1 (boolean)	Out 2 (boolean)
0	1	1
5	1	0
>7	0	0

achieve the tri-level logic function is shown in Figure 8.25. The circuit consists of special devices capable of handling up to 15 V inputs. This is achieved via double-level polysilicon technology. The voltage level at the input pin is decoded by the circuit into two separate digital bilevel logic levels internal to the device (Out 1 and Out 2 in Figure 8.25). Table 8.2 shows the truth table for the circuit assuming a nominal operating V_{cc} of 5 V.

The ESD protection circuit used for the HV pin is also the same as for the other inputs consisting of an LSCR, FPD, and a diffusion resistor. This is incorporated in a family of microcontroller chips with different processes. During testing with the HV pin, certain chips are found to exhibit a sudden increase of input current. The input waveform and the current through the pin are shown in Figure 8.26. Note that when the voltage reaches 16 V, there is a sudden increase in the current and this decreases as the voltage falls to below 10 V. Emission microscopy analysis can be used to understand this. In this example we have traced the drop in voltage to the FPD.

The $I-V$ characteristics for an FPD device are shown in Figure 8.27. The high current behavior of the FPD has been discussed in detail in Chapter 4. The avalanche breakdown voltage of the drain junction is controlled by the oxide thickness and the substrate doping. Additionally, the parasitic bipolar trigger voltage (or snapback voltage) is controlled by the channel length [Hsu82][Feng86]. The grading of the drain junction also has a significant impact on these voltages. When

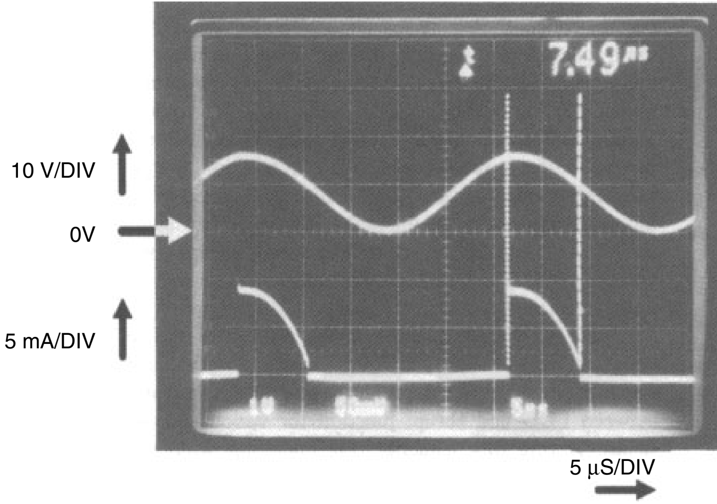


Figure 8.26 The voltage and current waveforms showing the HV pin trigger and snapback phenomena

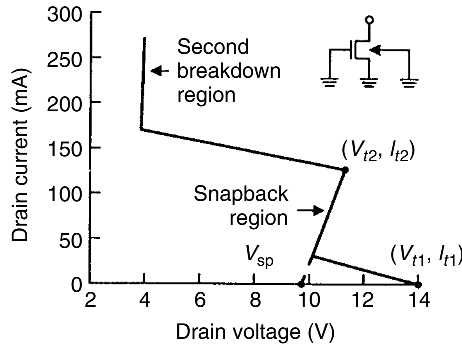


Figure 8.27 $I-V$ characteristics for the field plated diode device

the voltage at the pin exceeds the trigger voltage, the lateral *npn* turns on. Once the bipolar is on, it remains in the low impedance mode until the voltage drops below the snapback holding voltage. Thus, the internal node voltage will never be sufficient to allow the internal buffer to recognize it as HV. In addition, the presence of the high current at extended periods of time could cause other damage within the device. To prevent unintentional turn-on of the parasitic bipolar transistor, the FPD must be designed such that the trigger and snapback holding voltages are above the normal operating voltage level. Voltage overshoot and system noise level can also contribute to the problem of unintentional triggering of the protection circuit. In the case of these microcontroller devices, because

of a constant evolution of the ESD protection circuits, different channel lengths happened to be implemented in each design cycle for the FPD device. At the same time, the process fluctuated between LDD and abrupt junction for various other reliability and functionality issues. During this sequence, the operating window for the HV pin drastically changed from one version to another. This is chronologically recorded in Figure 8.28. The required minimum/maximum levels

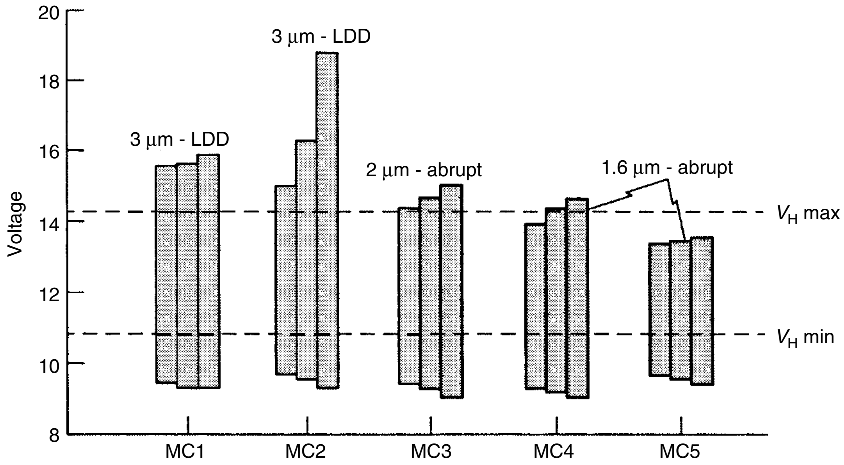


Figure 8.28 HV pin operating window for different processes

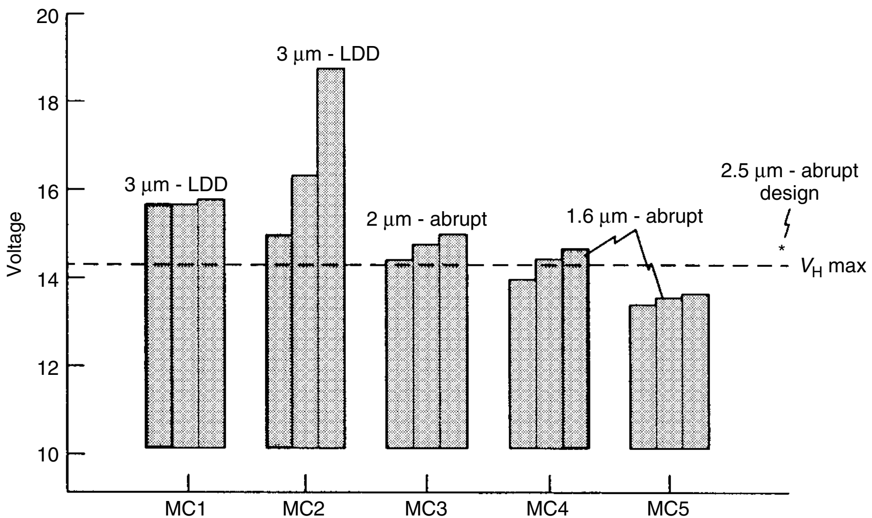


Figure 8.29 FPD trigger voltages for different process options

are also indicated as VH_{min} and VH_{max}. Note that for the 1.6-μm abrupt junction or non-LDD process, the functional failure of the HV pin function (not the ESD performance) is frequent. The maximum is determined by the trigger voltage. The data for this is shown in Figure 8.29. Note again that for the 1.6-μm abrupt junction, the trigger voltage falls below the HV pin maximum. Thus, the FPD channel length needs to be changed so that the trigger voltage is be above VH_{max}. As a result of this analysis, a simple solution that alleviated this problem was to make the FPD channel length 2.5 μm for the HV pin only. The trigger point of this design is shown by "*" in Figure 8.29. As the trigger voltage increases so does the snapback voltage. But this has little consequence on VH_{min} as long as the trigger voltage remains above VH_{max}. Even with the increase in snapback voltage the design showed no measurable reduction in ESD performance for this particular pin.

8.7 PROCESS EFFECTS ON INPUT PROTECTION DESIGN

Even effectively designed input protection schemes could cause unusual leakage problems if their layout interacts with process fluctuations. As discussed in the following example, this could lead to IIL/IIH problems again. Process influences on ESD in general are discussed in more detail in Chapter 9. The leakage problems observed in this case study were post-burn-in IIL/IIH failures that were bakeable and, hence, hard ESD failures could not be suspected. It has been shown that ESD pulses can increase the sensitivity to hot-carrier stress degradation [Aur88A][Aur88B]. We could assume that ESD-induced hot-carrier degradation in the FPD might be the cause for this bakeable leakage. However, this is not a likely cause since the gate is grounded during normal operation, which reduces the chances for hot-carrier injection. Furthermore, leakage at the FPD will not explain why bakeable IIL failures also occurred. In fact, these IIL failures occurred more frequently than IIH failures.

To further investigate this phenomenon, different portions of the input protection circuit were isolated by laser cutting the metal connections. First, the characteristics of a good pin and an IIL leakage pin are compared in Figure 8.30. Note the leakage at 0 V bias at the input. The laser cuts made for the circuit are indicated in Figure 8.31. The corresponding leakage characteristics are shown in Figure 8.32. After Cut 1, the leakage does not improve and thus there is no contribution from the FPD. The slight increase in the leakage after the cut is caused by laser damage. The leakage essentially remains after Cut 2, meaning that the p^+ resistor is not the cause. Finally, after Cut 3 to the V_{cc} line, the IIL leakage disappears. Thus, the leakage originated from the SCR n-well to the V_{cc} n-well (Figure 8.32) and resulted in IIL failure. The cause for this leakage was traced to small levels of positive mobile ions present in the process at the time. During the burn-in test, because both the wells are under temperature/voltage stress, the mobile ions can

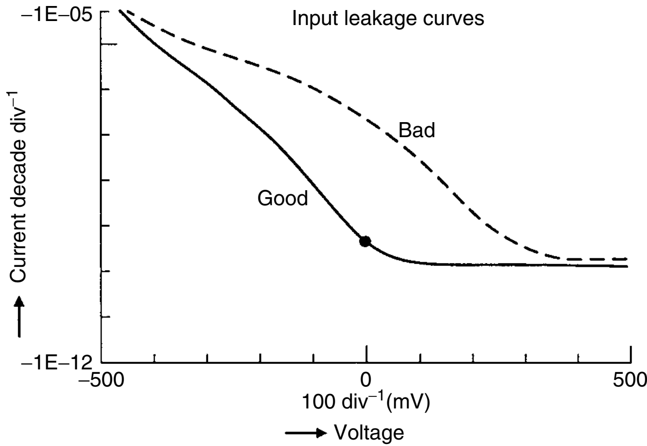


Figure 8.30 Input leakage curves showing good and bad pins. Note the IIL leakage for the bad pin

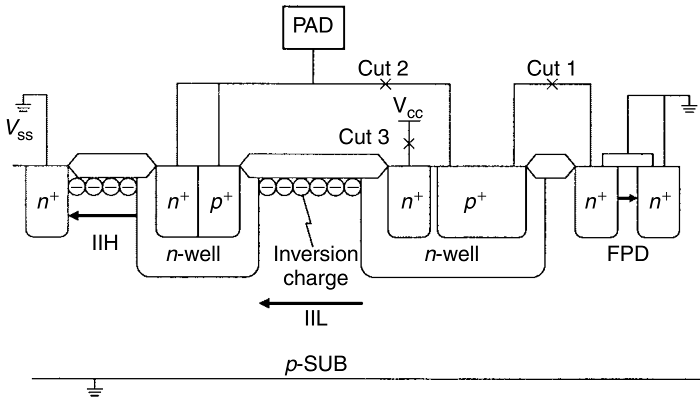


Figure 8.31 Bakeable IIL/IIH leakage mechanisms in an input protection scheme

easily form an inversion layer underneath the field oxide, causing this bakeable IIL leakage. When 0 V is applied to the input pin during this test, the conduction of the channel from the SCR well to the V_{cc} n-well leads to the increase in leakage. In addition, the bakeable IIH leakages are explained by the fact that, during burn-in, the mobile ions could also form an inversion layer between the n-well and the n^+ cathode of the SCR. The IIH bakeable leakages were less frequently observed. This is because during burn-in charge spreading from both wells contributes to IIL, whereas spreading from one well contributes to the IIH failure.

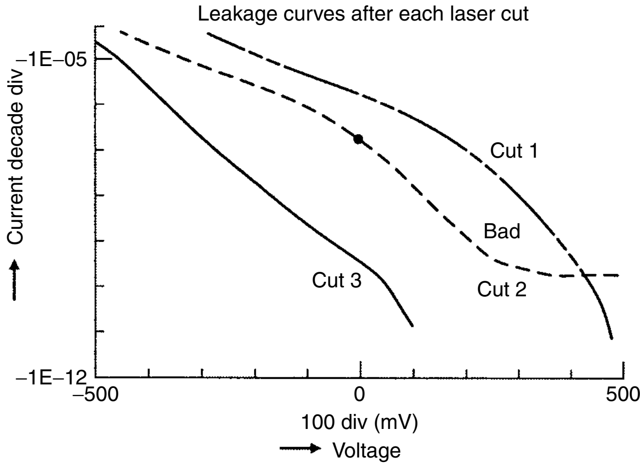


Figure 8.32 Input leakage curves after different laser cuts

The obvious solution to the aforementioned problem is to identify the source of contamination and decrease the mobile ion level with a cleanup of the process. However, to avoid any potential problems caused by small levels of contaminations that are present even in the cleanest of processes, the protection circuit layout could be modified. That is, the SCR n-well is placed on the other side of the pad, greater than 100 μm away. Alternately, only n^+ resistors can be used but at the expense of not having an effective diode to V_{cc} near the bond pad. With the correct design approaches and an improved process, the bakeable leakage phenomena are eliminated.

8.8 TOTAL IC CHIP PROTECTION

The next part of this chapter considers internal chip failures and protection designs. For complete ESD reliability of an IC chip, protection of the I/O pins alone is not sufficient since there may be many other possible sensitive areas on the chip [Krakauer94]. Internal chip failures could occur even if good protection designs are implemented at the I/O pins. In the remaining sections of this chapter, these issues are considered with examples from microcontroller and DRAM chips.

The input/output ESD circuit requirements call for good protection of the pin with respect to both the ground and the power bus pins. Although effective protection can be designed at the pin, many cases of damage phenomena are known to occur internally in the chip beyond the protection circuit. The issue of protection between V_{dd}-V_{ss} will be first discussed. This will be followed by examples of how protection

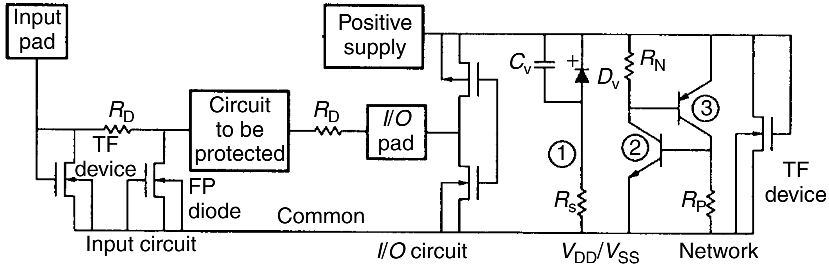


Figure 8.33 Overall CMOS Protection Scheme, showing basic circuit elements such as Thick-Field (TF) Devices and Field Plate (FP) diode

circuit performance can be sensitive to internal chip layout, independent of its effective design. Several illustrative case studies will be reported to emphasize the internal chip ESD phenomena and their adverse effects.

An effective protection between the power bus lines is often overlooked although this is equally important for overall ESD immunity [Maene92]. A comprehensive method of testing requires not only stressing of every pin with respect to V_{dd} (V_{cc}) and V_{ss} but also between inputs and outputs. This would lead to very complex internal ESD stress currents, which must be carefully analyzed in order not to compromise the circuit reliability. Consider the overall protection scheme for a CMOS circuit chip that is shown in Figure 8.33. As seen from this diagram, there are several parasitic devices involved in the stress current path for stressing between the different pin combinations of the MIL-STD.

Internal chip ESD damage could result due to direct stress applied between V_{dd} and V_{ss} pins. Some illustrative examples of this and the possible solutions are discussed in Section 8.10. Generally, current flow through the internal circuitry can take place when outputs/inputs are stressed with respect to V_{dd} or V_{ss}. Hence, it is important to consider the issues related to the internal circuits and layout, or the overall ESD performance can suffer.

8.9 POWER BUS PROTECTION

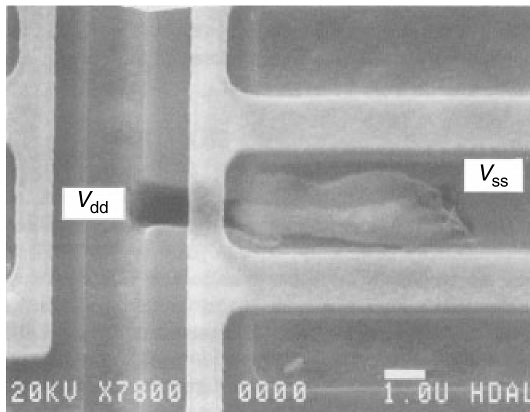
Even with effective protection at the pins, many cases of damage phenomena do occur internally in the chip (see e.g. [Duvvury88A][Maene92][Cook93]). Some of these can be directly attributed to inadequate protection provided for stress between the power bus lines while others may be due to deficiencies in the layout of the protection circuits. The power bus protection issues and techniques will be discussed in this section.

The MIL-STD testing method requires stressing between all inputs, outputs, and power bus pins (V_{dd}, V_{cc}) with respect to the ground pin (V_{ss}). Thus it would seem logical to place a protection circuit between V_{dd} and V_{ss} for direct stress between the two bus lines [Palella85][Duvvury87B].

In general, an MOS thick-field device between V_{dd} and V_{ss} is adequate for either positive or negative stress (e.g. [Merrill93]). For positive stress on V_{dd} with respect to V_{ss}, the lateral *npn* in breakdown would form the effective protection; for negative stress on V_{dd} with respect to V_{ss}, the forward-biased diode would turn on provided the substrate is also at V_{ss}. If the substrate connection is not the same as the V_{ss} connection during the ESD stress then the lateral *npn* transistor will be required to trigger and carry the stress current. In large microcomputer chips several power bus lines are often used. For these chips, protection circuits should be used between all bus combinations.

The V_{dd} to V_{ss} capacitance can contribute significantly to the ESD performance of an IC. In large circuits, this capacitance can be on the order of 10 nF. A direct stress between V_{dd} and V_{ss} will first have to charge up this capacitance, which will slow the risetime of the current pulse and limit the voltage during stress [Duvvury88B]. Triggering of the protection device only occurs after the voltage reaches the required trigger voltage and by this time the stress current could be well below its peak level. Hence, the chip capacitance will limit the stress in the protection device and increase the ESD levels.

As mentioned earlier, a comprehensive ESD testing methodology includes stress between the input pins and V_{dd}. One such case is considered next. Shown in Figure 8.35 is an equivalent circuit schematic of input protection. The thick-field device is shown as a bipolar device for positive stress and as a diode for negative stress. There is no dedicated protection element directly between the pad and V_{dd}. The elements Rv and Cv represent the parasitic resistance and capacitance between V_{dd} and V_{ss}. When the pad is stressed with respect to V_{dd}, the stress current would eventually flow between V_{ss} and V_{dd}. This could lead to damage internally in the chip.



Damage due to V_{dd}-V_{ss} stress

Figure 8.34 Damage site observed for V_{dd}-V_{ss} stress

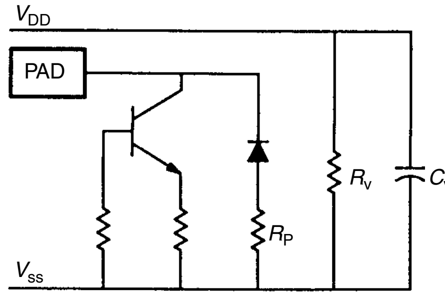


Figure 8.35 Equivalent component schematic of input protection

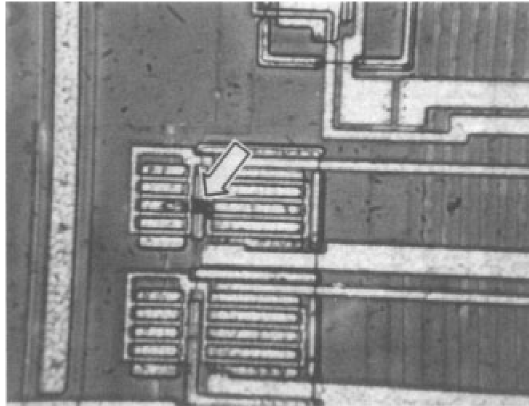
8.10 INTERNAL CHIP ESD DAMAGE

In Section 8.9, ESD damage due to direct stress between the V_{dd} and V_{ss} pins was discussed. An example was given where the damage could be directly attributed to V_{dd} and V_{ss} diffusions. However, for applied stress between the power bus pins, some subtle damage phenomena could also result. Furthermore, as mentioned above, for stress between input or output pins and V_{dd}, internal damage could occur, again due to the stress current path between V_{dd} and V_{ss}. Some examples of both cases will be given in the next two subsections.

8.10.1 V_{dd}-V_{ss} Stress Current Damage

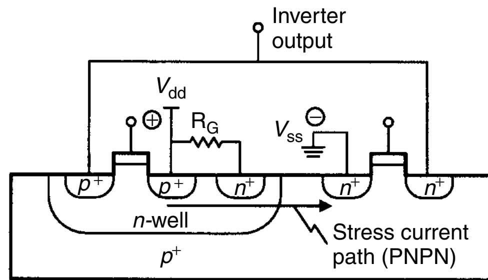
Internal damage due to stress current between V_{dd} and V_{ss} is often difficult to identify if it is not obvious as was shown in Figure 8.34. Liquid crystal analysis can be used to locate these failure sites.

In a CMOS chip the damage site observed through liquid crystal analysis is shown in Figure 8.36(a). In this case a positive ESD stress is applied to the V_{dd} pin with respect to the V_{ss} pin. This damage site is identified as a CMOS inverter in the internal circuitry. As V_{dd} is stressed positive with respect to V_{ss}, the stress current passes through a parasitic *pnpn* device, as illustrated in Figure 8.36(b), and caused the failure. The parasitic *pnpn* device is formed by the p^+ source diffusion of the pMOS transistor connected to V_{dd}, the n-well, the p^+ substrate, and the n^+ source diffusion connected to V_{ss} of the nMOS transistor. Referring to Figure 8.36(a), it should be noted that the damage extends from p^+ to n^+ through the n^+ guardring contact. One way to reduce susceptibility to this problem would be to increase the resistance (R_G) in the current path. An increase in R_G will reduce the holding voltage of the parasitic SCR resulting in less power dissipation and heat generation. This could be achieved by identifying the inverter fail sites and removing the guardring contacts in the direct current path. However, the latchup performance of the CMOS chip must not be affected by this approach. For this reason only the guardring contact facing the nMOS device are



Damage in inverter for V_{dd} to V_{ss} positive stress

(a)



V_{dd} to V_{ss} stress

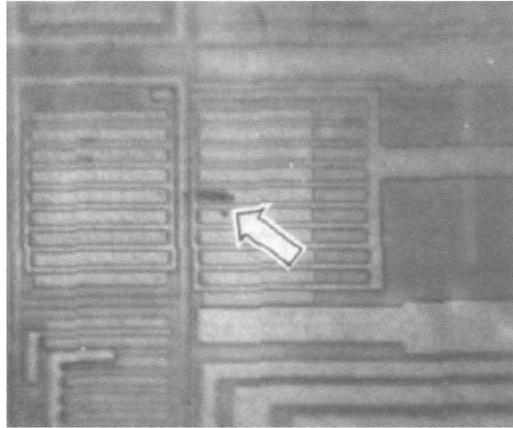
(b)

Figure 8.36 Damage site observed in an output buffer for positive stress on V_{dd} with respect to V_{ss}. The p-channel device is on the left side of the schematic

removed. Test structure analysis has shown a marked improvement in the ESD performance of individual buffer devices without the direct path guardring contacts (4 kV for Human Body Model stress), as compared to those with complete guardring contacts (600 V for Human Body Model stress). In both structures the same failure mechanism occurred but the failure voltage was higher in the first structure.

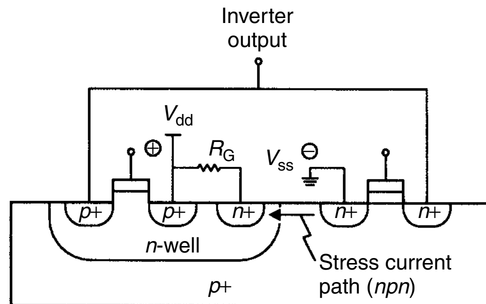
The example cited here is for the case of an advanced CMOS process involving silicided diffusions. As it has been reported earlier that silicided diffusions are more susceptible to heat damage due to ESD [McPhee86][Duvvury86], it is suspected that this situation may not be so severe for nonsilicided cases.

For positive stress applied to V_{ss} with respect to V_{dd}, or negative stress applied to V_{dd} with respect to V_{ss}, the situation is different. The internal failure was again seen in an inverter, as shown in Figure 8.37(a). However, the damage does not



Damage in inverter for V_{dd} to V_{ss} negative stress

(a)



V_{ss} to V_{dd} stress

(b)

Figure 8.37 Damage site observed in an output buffer for negative stress on V_{dd} with respect to V_{ss}. The p-channel device is on the left side of the schematic

extend to the p^+ drain of the pMOS device. This can be explained by considering the stress current path as shown in Figure 8.37(b). With positive stress on V_{ss}, the stress current path is through a lateral *npn*, formed with the n^+ source diffusion of the nMOS device, p^+ substrate, and the n^+ guardring of the pMOS device. As noted previously, this would be more severe for chips with silicided diffusions. In the case of devices with substrate bias generators, such as DRAM's, there is no effective low impedance current path to offer protection for this stress condition. To divert the stress current path, a protection diode may be employed between V_{dd} and V_{ss} as shown in Figure 8.38. Such an approach would be effective only for positive stress on V_{ss} with respect to V_{dd}. This type of protection could be distributed throughout the chip to attain overall protection for V_{ss} to V_{dd} stress. A chip with grounded substrate may not necessarily need this additional diode

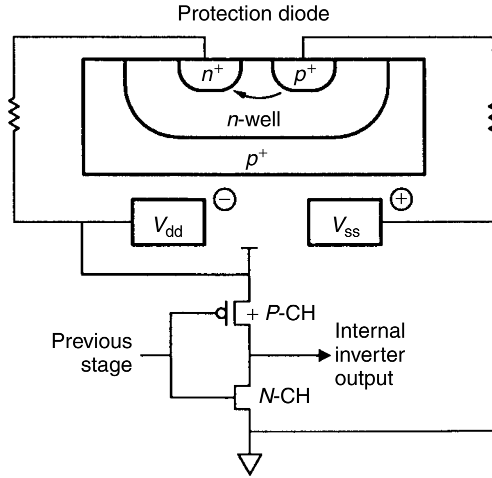


Figure 8.38 Damage site observed for positive stress on V_{ss} with respect to V_{dd}. The p-channel device is on the left side

because the $n^+ V_{dd}$ to p-substrate diode is intrinsic to the circuit. ICs fabricated in advanced CMOS processes employing the above protection techniques in the internal chip layout were found to be virtually free of the damage phenomena shown in Figures 8.36 and 8.37.

Internal chip layouts for improving circuit performance can also cause unexpected ESD failures. Consider the internal circuit schematic shown in Figure 8.39. The polysilicon-1 layer forms the gates of the internal transistors. The polysilicon-2 layer connected to V_{ss} is used as a means to decouple the noise on the gates of these transistors. However, this technique is found to have an adverse impact on the ESD performance. By testing the chip with positive ESD pulses between V_{dd}

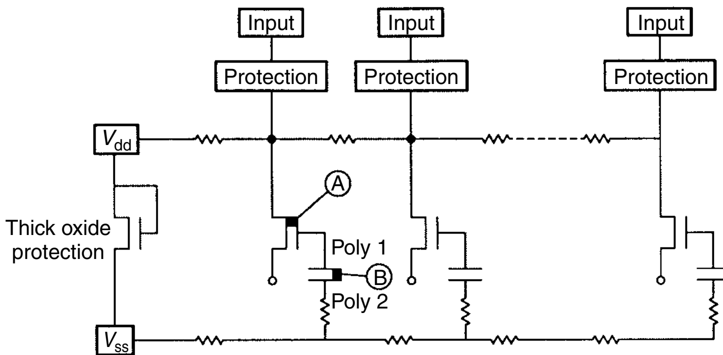


Figure 8.39 Internal circuit damage caused by positive stress between V_{dd} and V_{ss}

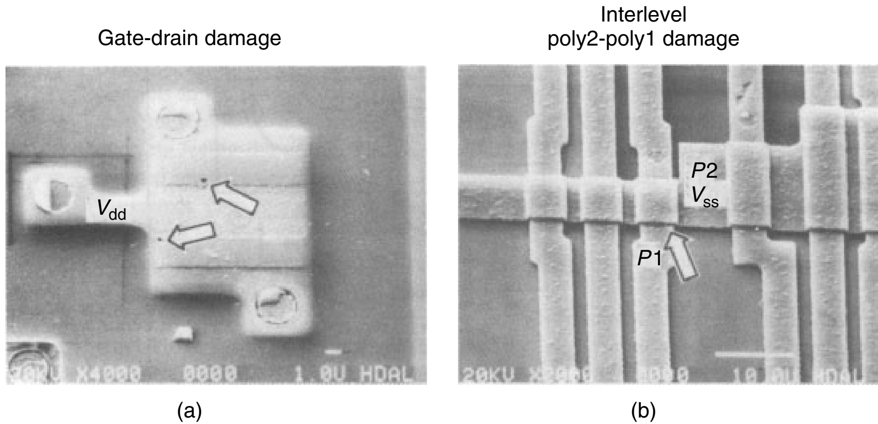


Figure 8.40 (a) Gate-drain damage at location A in Figure 7. (b) Interlevel oxide damage at location B in Figure 7

and V_{ss} it was found that failures occurred with only 2 kV of stress according to the Human Body Model Test. Additionally, it was discovered that after $V_{dd} - V_{ss}$ stress the input pins shown in Figure 8.39 were shorted to V_{ss}. Although a large thick-field device existed between V_{dd} and V_{ss} (see Figure 8.39), it was found to be ineffective in this case. This was because the layout resulted in stress current path to ground. The internal damage occurred as a gate-drain short at location '*A*' and an interlevel oxide rupture at location '*B*', as indicated in the figure. These fail modes are illustrated in Figures 8.40(a) and 8.40(b), respectively. An obvious modification that improves the ESD performance for $V_{dd} - V_{ss}$ stress is to eliminate such stress current paths to ground by removing the noise decoupling polysilicon-2 layer at these locations.

8.10.2 Output to V_{dd} Stress

It was mentioned previously that applying ESD stress between input pad and V_{dd} can cause current to flow between V_{dd} and V_{ss} and lead to damage in the internal parasitic devices. Similarly, internal parasitic device damage could also occur when a CMOS output buffer is stressed with respect to V_{dd}. It was reported by Duvvury [Duvvury87A] that good ESD performance can be obtained for the case of CMOS output buffers by optimizing the layout of the pMOS device to improve the *pn* diode between the pad and V_{dd} (as shown in Figure 6.53). The layout minimizes the resistance of the diode formed by the p^+ diffusion and *n*-well contact. However, in the limiting case for positive stress with respect to V_{dd} excessive current could circumvent the pMOS device path and take an alternative path as shown in Figure 8.41. This will trigger the lateral *npn* (formed with n^+ drain, *p*-substrate, and n^+ source) as well as other series *npn* devices between V_{ss} and

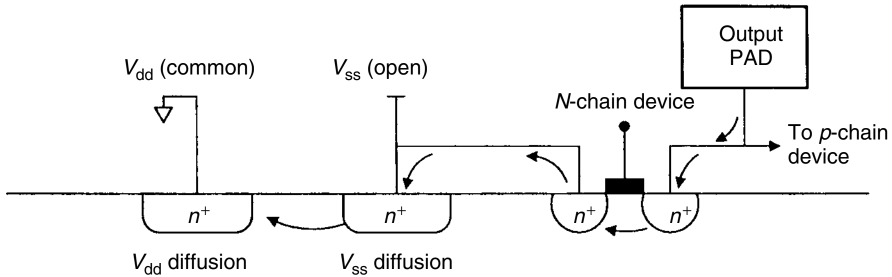


Figure 8.41 Circuit schematic of an output buffer showing current path during a V_{dd}-V_{ss} stress

V_{dd}, as indicated in the figure. If an advanced CMOS process is used with silicided diffusions, the nMOS output device will likely to show damage. Since complicated V_{dd}-V_{ss} current paths are involved, the damage may also occur in one of the many parasitic *npn* devices in the internal circuitry. Therefore a careful consideration of this phenomenon must also be made in optimizing the internal chip layout.

8.11 STRESS DEPENDENT ESD BEHAVIOR

We saw earlier that a window of ESD failures can occur for ineffective protection designs. This phenomenon could also occur due to internal current paths. In most qualification procedures for ESD performance, the chips are usually stressed at a single ESD voltage level given by the *specification* and if it passes this test it is considered to have qualified. This is known as a “go/no-go” method. For example, if the chip passes a stress level of 4 kV (according to the Human Body Model), it is assumed to pass 2 kV also. On the other hand, if it fails a stress level of 2 kV, it is assumed that it will not pass 4 kV. It will be shown here that this may not necessarily be the case if there are interactions with internal parasitic devices.

In Figure 8.42 the overall ESD performance of a VLSI chip is shown as a function of ESD stress level. These devices were stressed according to the Machine Model with all pins positive to V_{ss} and negative to V_{dd}. On the y-axis the percentage number of devices passing full functional test after ESD stress are shown and the x-axis gives the applied ESD stress level. There is clearly a window where the ESD performance of this device is weak. In this particular case, the failing devices are found to have a drastic increase in the V_{dd} to V_{ss} current, I_{dd}, after ESD stress. A more thorough investigation led to the conclusion that the I_{dd} current increase in the failed devices occurred only when the stress applied was negative with respect to V_{dd}. After more detailed data the particular I/O pins that are causing this phenomenon were isolated.

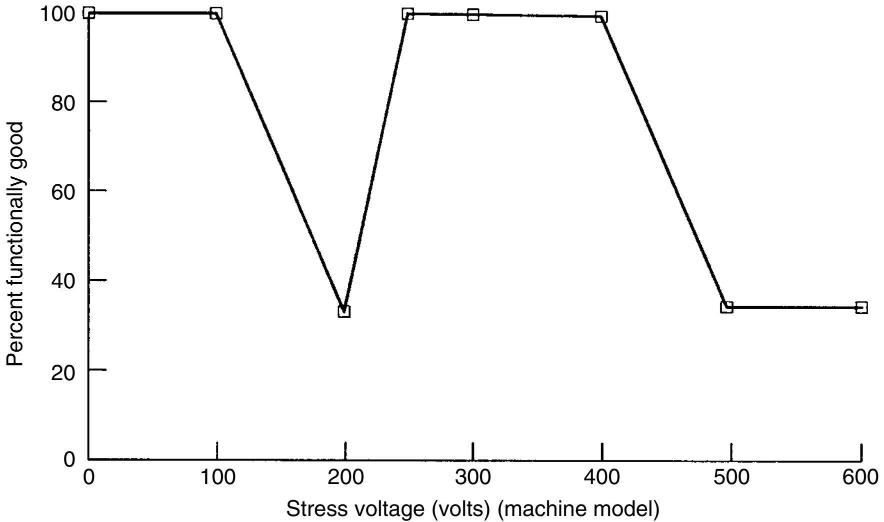


Figure 8.42 Functionally good device post-ESD stress vs. stress voltage

Although the results shown in Figure 8.43 are for the Human Body Model stress, it is still seen that when these I/O pins are stressed negative with respect to V_{dd}, an I_{dd} failure window, similar to the one in Figure 8.42, does exist. Liquid crystal analysis in the case of failed devices revealed the damage sites to be in the internal clock buffers.

The damaged clock buffer, as shown in Figure 8.44, is CMOS. With a negative stress to V_{dd}, the parasitic n^+ drain to substrate diode in the protection circuit becomes forward-biased and the stress current has to flow between V_{dd} and V_{ss}, as indicated in the figure. At relatively low stress levels (1.5 kV for the Human Body Model or 150 V for the Machine Model) the stress current caused damage in the parasitic *pnpn* (SCR) devices in the clock buffers. The stress current path responsible for this damage is highly layout dependent as shown in Figure 8.45. Note that the V_{dd}-V_{ss} protection circuit indicated in the figure could not have prevented the failure. The damage was identified through liquid analysis as nMOS gate-drain short in one or two of the clock buffers as indicated in Figure 8.46. This short apparently caused self-biasing of the damaged buffer and manifested itself as increased I_{dd} current (see Figure 8.47). It is hypothesized that at lower stress levels the parasitic SCR in only one of the clock buffers was triggered. Therefore, this acted as a parasitic ESD protection circuit with a protection level of 1.5 to 2 kV. At higher stress levels it is likely that all five of the clock buffer SCRs were triggered. This would explain why no I_{dd} current increase or any damage in the clock buffers was observed at stress of >2 kV. In fact, with all five SCRs triggered no I_{dd} failures would be observed up to >6 kV as shown in Figure 8.43. However, at these high stress levels the ESD protection circuits begin to fail and

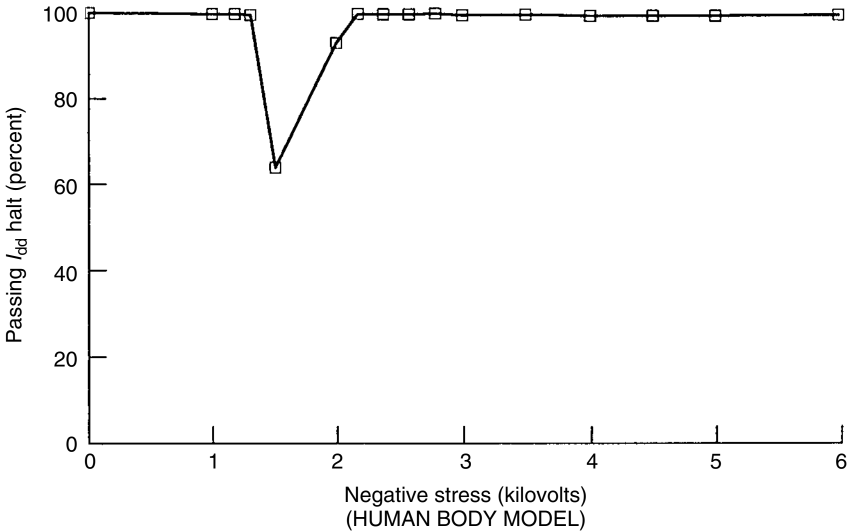


Figure 8.43 Devices passing I_{dd} leakage test post-ESD stress vs. stress voltage. Applied stress was negative with respect to V_{dd}

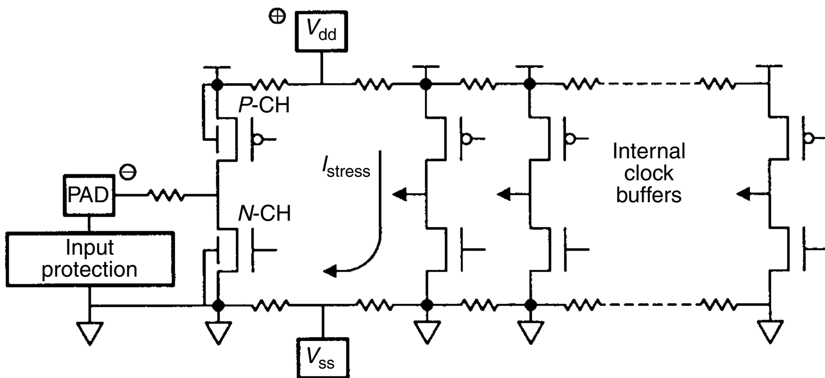


Figure 8.44 Layout schematic showing current path for negative stress between an I/O pin and V_{dd}

the chip functionality is affected as shown in Figure 8.42 for the Machine Model. One possible solution to eliminate this failure would be to increase the individual clock buffer sizes to improve the parasitic SCR protection.

The type of stress dependent ESD behavior discussed in the preceding text is unusual but not uncommon. Effectively designed input protection circuits tend to perform very well at 4 or 6 kV but often cause pin leakage at stress levels of 2 kV

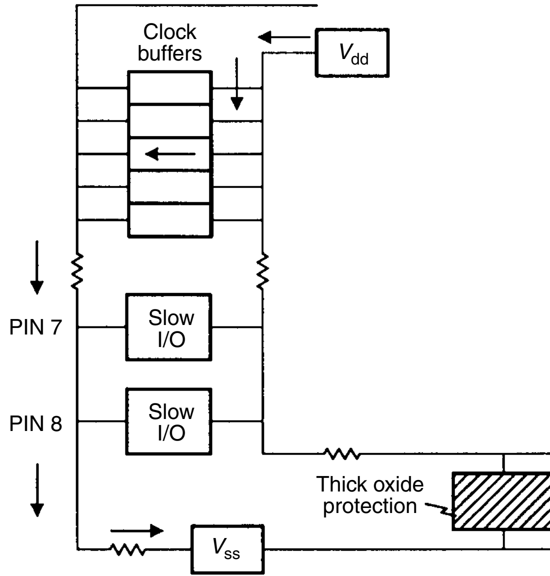


Figure 8.45 Block diagram of ESD protection, I/O buffers and internal clocks

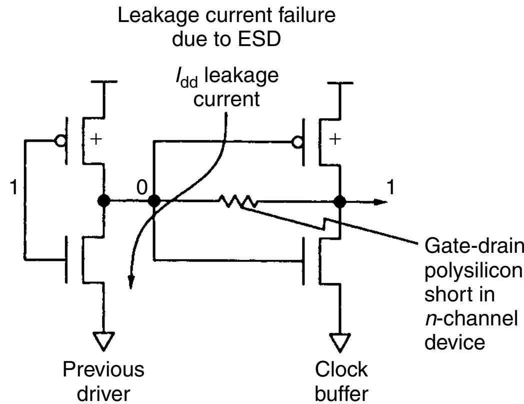


Figure 8.46 Damage in internal clock buffer with gate-drain short

or below. Failure windows of this type point out the importance of step-stressing the device pins to establish the overall ESD immunity of a circuit chip.

The V_{dd}-V_{ss} protection design requirements and how these can lead to internal chip ESD damage phenomena were discussed here. Implementing thick-field oxide devices between the V_{dd} and V_{ss} power bus lines could offer protection against stress between the two but only as long as the weak internal parasitic

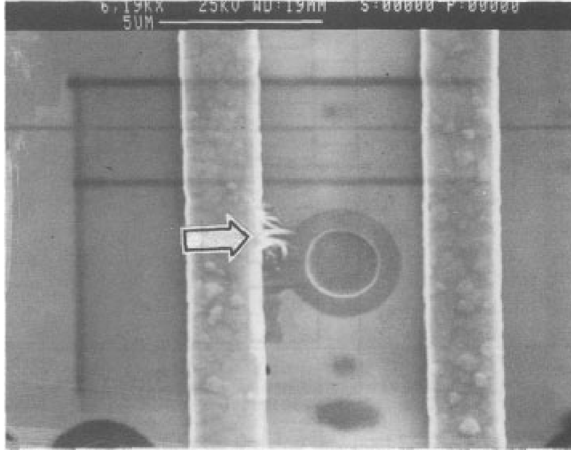


Figure 8.47 Circuit schematic of clock buffer showing I_{dd} leakage current path

devices are eliminated. An example was shown where n^+ diffusions connected to V_{dd} and V_{ss} respectively should not be too close to each other in the internal chip layout.

8.12 FAILURE MODE CASE STUDIES

For any optimized ESD protection scheme, the failure is expected to occur in the protection device. However, for different reasons the failure modes can be seen either in the protected gate of an input buffer or in the output buffer transistors. When the internal pull-down transistor is not effectively protected, for Pad to GND HBM stress damage can occur in the nMOS transistor as shown in Figure 8.48. Incidentally, for this silicided process notice the symmetrical melt filaments at the drain and source. This is typical for a silicided process since the melting of silicide is initiated at the sidewall edges. In this case, the internal output damage can be prevented by either making the output device relatively larger or by introducing an isolation resistor between the output and the protection device at the pad. Similar damage phenomena are also seen for CDM when the output device is not well protected.

When the IO pad is stressed negative to V_{dd}, there are several current paths that can create damage. This will, to some extent, depend on the efficiency of the V_{dd} protection device. If, for example, the V_{dd} pad is very close to the IO pad, some sneak current can go through the pMOS pull-up transistor for the advanced deep submicron technologies [Ting01]. In this case, the pMOS triggers as a lateral PNP, as shown in Figure 8.49. Note again that for silicided processes the PNP also shows symmetrical melt filaments. For both the

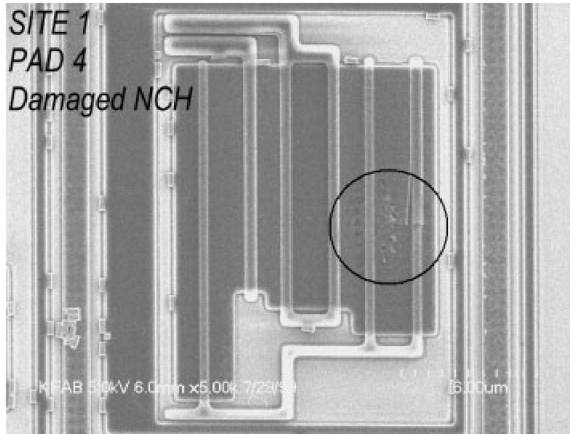


Figure 8.48 The damage to the output nMOS transistor for HBM stress. Notice the symmetrical melt filaments indicating a silicided process

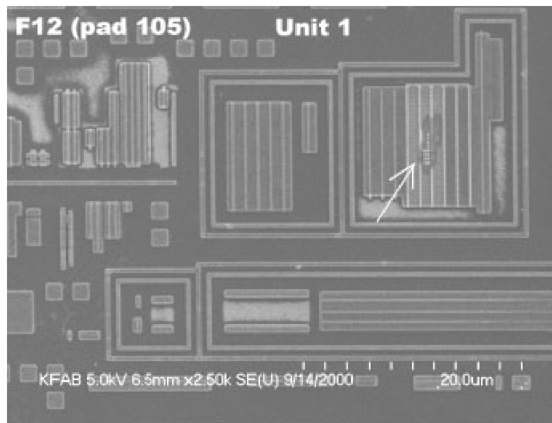


Figure 8.49 The damage to the output pMOS transistor for HBM stress. Notice the symmetrical melt filaments indicating a silicided process

nMOS and pMOS the ESD performance can be improved by making the buffer devices larger by adding dummy sections where the unused portion are tied to respective power supplies (gate of nMOS to GND, and gate of pMOS to V_{dd}). Of course these are effective only if the dummy device gates are tied through a resistor such that under ESD the coupling on the gate of the buffer device matches the gate of the active device. This technique was discussed in Figure 7.14.

For IO to V_{dd} negative stress the damage could also occur internal to the circuits connected to V_{dd} or could result in a latchup like failure. For this stress combination

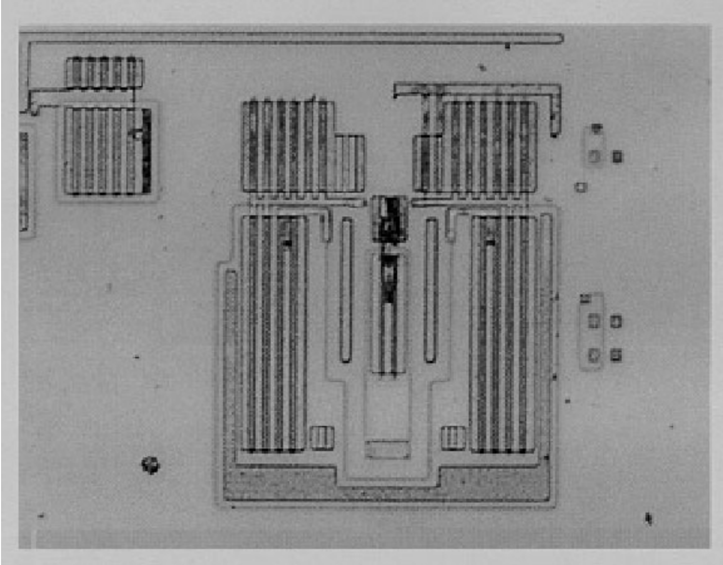


Figure 8.50 The damage to the input buffer appearing as latchup for negative polarity stress between IO and V_{dd}

the voltage at the V_{dd} pad builds up, and if the power supply protection is not efficient or if there is too much bus resistance in the path, latchup in the input buffer can occur as shown in Figure 8.50. Note that an interesting aspect of this is that the buffer might pass the standard latchup test but still fail as a latchup site during ESD stress.

Instead of the input buffer damage, as shown in Figure 8.50, it could also occur in a clock driver circuit connected very close to the V_{dd} pad. One example was previously shown in Figure8.47. Recent studies also indicated that such damage is common if the ground bus resistance of the V_{dd} protection device is not optimized [Chaine97][Puvvada98].

Even if an effective protection device is placed at the input pad the voltage buildup at the inputs side of a pass gate transistor must be considered. When a secondary clamp is not placed as recommended in Figure 6.27, the input gate oxide can be blown for CDM or the input gate poly can be damaged for HBM or MM as shown in Figure 8.51.

Finally, ESD failures are sometimes likely to occur due to process defects. Figure 8.52 shows an example where holes are seen in the poly gates of the nMOS protection device.

Although this damage became apparent after HBM stress, a careful examination indicates that there are no drain-source melt filaments but rather a rupture to the gates at multiple locations. Therefore careful consideration of process anomalies is always important during ESD evaluation.

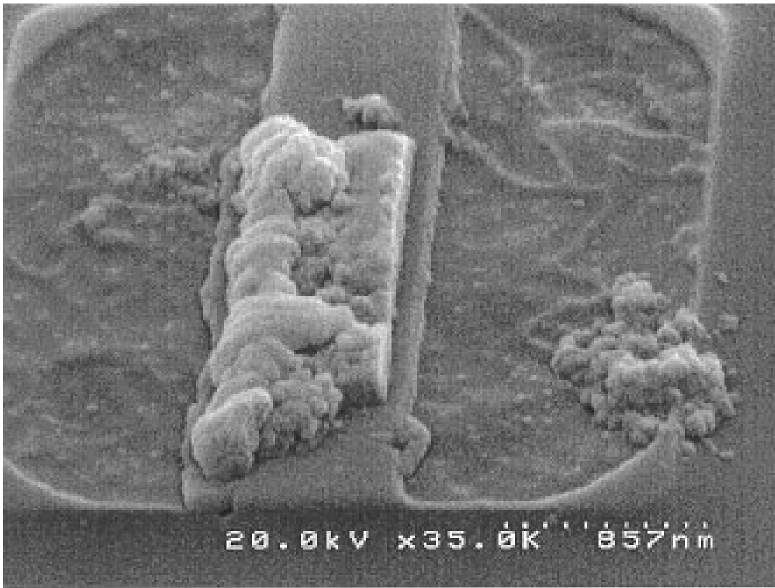


Figure 8.51 The damage to the input buffer gate during MM. Notice the severe heating of the poly gate

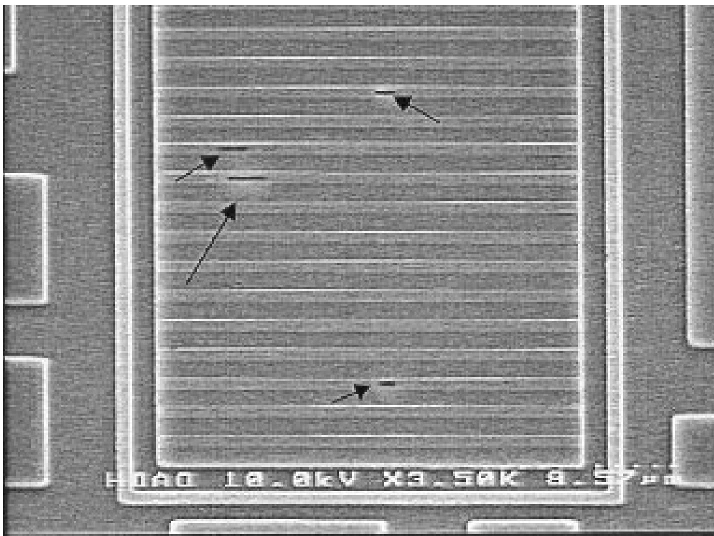


Figure 8.52 The damage to multi-finger nMOS is related to a process-induced damage

8.13 SUMMARY

Failure analysis techniques have been discussed in this chapter together and some of the principal failure modes and their associated electrical signatures have been presented. Liquid crystal analysis and photon emission microscopy are extremely useful tools for the identification of failure locations. In addition, photon emission is an effective tool for the analysis of the operation of protection elements under high current conditions and in the debugging of weak designs and layouts of protection circuits. Failure modes can extend from *soft* to *severe* with poststress leakage currents ranging from 1 nA to 1 mA and more in some cases. The most common failure mode in a properly designed circuit was that due to silicon melting as a result of current localization. In advanced CMOS circuits it was also shown that damage at the diffusion edge can be the dominant failure mechanism with leakage currents between 1 nA and 10 μA.

The effects of ESD stress between various pin combinations (such as output to V_{dd}) can cause ESD stress current to flow between V_{dd} and V_{ss} and cause internal ESD damage. It has been shown in this chapter that careful analysis of these issues are needed and proper layout techniques must be employed to ensure the good overall ESD performance in an IC.

Results from the analysis of a specific chip were reported. These show that when specific I/O pins are stressed negative with respect to V_{dd}, I_{dd} increases after a low level ESD stress (1.5 kV) but not after a higher stress level of 4 kV. This was explained by the damage caused by the stress current through the large internal CMOS clock buffers on the chip, which caused them to act as parasitic SCR devices. For lower stress levels only one clock buffer SCR was triggered but at higher stress levels all five of the clock buffers triggered to provide high levels of ESD protection. It was pointed out that the possible existence of such stress dependent ESD performance should be determined by step-stressing the devices and testing for functionality.

The actual case studies presented in this chapter clearly illustrated the importance of internal ESD phenomena. In the current advanced technology VLSI chips and in the next generation of very high density ICs the protection might very well be dictated by internal $V_{dd}-V_{ss}$ stress currents either as a result of direct stress between the two or as a result of indirect stress such as when I/O pins are stressed with respect to V_{dd}. A good protection scheme would, of course, overcome this problem as well as provide robust protection circuits at the pins. Aside from the obviously visible internal ESD damage, the $V_{dd} - V_{ss}$ stress currents might possibly cause latent failures in the internal circuitry. Other reliability stress analyses are necessary to thoroughly understand these effects. For example, it has been shown that ESD stress can increase the susceptibility of nMOS devices to hot-carrier degradation [Aur88A][Aur88B]. The significance of such issues in the next generation of ICs will need more detailed study especially with regard to the effects of latent ESD effects.

So far in this chapter the various input protection design issues for good ESD protection as well as for post-burn-in reliability have been discussed. A protection scheme for the current technologies, consisting of a lateral SCR for the primary protection with a diffusion resistor and a field plated diode for the secondary protection, can be made effective and reliable only through a careful design and layout of the isolation stage. As shown here, without this approach, false levels of protection can be assumed. These, in turn, can lead to post-burn-in stress test losses.

The leakage failures after burn-in could be either input low (IIL) or input high (IIH). Both of these are associated with the isolation stage design. The nonbakeable or hard failures are caused by either damage to the field plated diode or, in the case of p^+ resistor/V_{cc} diode, to the parasitic devices formed. In either case, the protection inefficiency as well as the burn-in loss can be avoided by improving the isolation resistor value to an adequate level.

The bakeable IIL/IIH failures have been traced to the interaction of the ESD protection layout with fluctuations of the positive mobile ions in a process. Simple ESD design layout modifications can alleviate these.

Finally, the requirements for special high voltage application pins have also been discussed. In these cases the field plated diode channel length must be optimized to achieve the best possible ESD performance without disturbing the pin function. A study of the channel length dependence of this device will lead to the correct special design.

The input protection schemes are not only important for ESD but can also have an impact on the total reliability of the chip. Besides the leakage failures discussed here, latchup susceptibility or latent damage could also become more important. As even more advances are made in the process technologies, the protection schemes may have to be constantly revised in light of these other reliability issues.

The internal chip ESD damage could also be associated with stress between V_{dd} and V_{ss} in an indirect manner. For example, in the case of advanced CMOS processes it was shown that for positive stress between V_{dd} and V_{ss} the damage occurred in an inverter circuit with a *pnpn* stress current path. This can be eliminated by removing the n-well guardring contacts in the direct current path. On the other hand, for negative stress between V_{dd} and V_{ss} the damage was also shown to be in an inverter circuit but through a lateral *npn* stress current path. The latter would require placing a diode between V_{dd} and V_{ss}. These two internal chip layout considerations were found to improve the protection between the power bus lines. Careful analysis of the chip layout to remove all stress current paths to ground is necessary to improve $V_{dd} - V_{ss}$ protection.

BIBLIOGRAPHY

[Abderhalden91] J. Abderhalden

[Adler78] M. Adler, "Accurate calculations of the forward drop and power dissipation in thyristors", *IEEE T. Electron Dev.*, **ED-25**, 16–22, 1978.

[Agatsuma65] T. Agatsuma, "Turnover phenomenon in NνN Si devices and second breakdown in transistors", *IEEE T. Electron Dev.*, **ED-13**, 748–755, 1966.

[Alexander78] D.R. Alexander, "Electrical overstress failure modeling for bipolar semiconductor components", *IEEE Trans. Comp. Hyb. Man. Tech.*, **CHMT-1**, 345–353, 1978.

[Amerasekera86] A. Amerasekera and D. Campbell, "ESD Pulse and Continuous Voltage Breakdown in MOS Capacitor Structures", in *Proceedings of the 8th EOS/ESD Symposium*, pp. 208–213, 1986.

[Amerasekera90] E.A. Amerasekera, L.J. van Roozendaal, J. Abderhalden, J.J.P. Bruines and L. Sevat, "An analysis of low voltage ESD damage in advanced CMOS processes", in *Proceedings of the 12th EOS/ESD Symposium*, pp. 143–150, 1990.

[Amerasekera91] A. Amerasekera, L. van Roozendaal, J. Bruines and F. Kuper, "Characterization and modeling of second breakdown in nMOST's for the extraction of ESD-related process and design parameters", *IEEE T. Electron Dev.*, **ED-38**, 2161–2168, 1991.

[Amerasekera92] E.A. Amerasekera, W. van den Abeelen, L.J. van Roozendaal, M. Hannemann and P.J. Schofield, "ESD failure modes: Characteristics, mechanisms and process influences", *IEEE T. Electron Dev.*, **ED-39**, 2, 1992.

[Amerasekera93A] A. Amerasekera, A. Chatterjee and M.-C. Chang, "Prediction of ESD Robustness of a Process Using 2-D Device Simulations", in *Proceedings of the 31st Int. Rel. Phys. Symp.*, pp. 161–167, 1993.

[Amerasekera93B] A. Amerasekera, M.-C. Chang, J. Seitchik, A. Chatterjee, K. Mayaram and J.-H. Chern, "Self-heating effects in basic semiconductor structures", *IEEE Elec. Dev.*, **ED-40**, 1836–1844, 1993.

[Amerasekera94A] A. Amerasekera and R. Chapman, "Technology design for high current and ESD robustness in a deep submicron process", *Elec. Dev. Lett.*, **15**, 1994.

[Amerasekera94B] A. Amerasekera and C. Duvvury, "The Impact of technology Scaling on ESD Robustness and Protection Circuit Design", in *Proceedings of the 16th EOS/ESD Symposium*, 1994.

[Amerasekera94C] A. Amerasekera and J.A. Seitchik, "Electrothermal behavior of nMOS transistors under high current snapback (ESD/EOS) conditions", in *Tech. Dig. IEDM*, 1994.

[Anderson87] W.T. Anderson, "Electrostatic Discharge (ESD) Thresholds for GaAs Microwave Power FET Devices", *Proceedings of the 9th EOS/ESD Symposium*, pp. 205–207, 1987.

[Ash83] M.S. Ash, "Semiconductor Junction non-linear failure power thresholds: Wunsch-Bell revisited", *Proceedings of the 5th EOS/ESD Symposium*, pp. 122–127, 1983.

[Aur88A] S. Aur, A. Chatterjee and T. Polgreen, "Hot electron reliability and ESD latent damage", *IEEE T. Electron Dev.*, **ED-35**, 2189–2193, 1988.

[Aur88B] S. Aur, A. Chatterjee and T. Polgreen, "Hot electron reliability and ESD latent damage", in *Proc 24th IRPS*, 100–105, 1986.

[Avery84] L. Avery, "Using SCR's as transient protection structures in integrated circuits", *Proceedings of the 6th EOS/ESD Symposium*, pp. 177–186, 1984.

[Avery87] L. Avery, "Charged device model testing: Trying to duplicate reality", *Proceedings of the 9th EOS/ESD Symposium*, pp. 88–92, 1987.

[Bardeen50] J. Bardeen and W. Shockley, "Deformation potentials and mobilities in nonpolar crystals", *Phys. Rev.*, **80**, 72–76, 1950.

[Barnett66] A.M. Barnett and A.G. Milnes, “Filamentary injection in semi-insulating silicon”, *J. Appl. Phys.*, **37**, 4215–4223, 1966.

[Beltman90] R.A.M. Beltman, H. van der Vlist and A.J. Mouthaan, “Simulation of thermal runaway during ESD events”, in *Proceedings of the 12th EOS/ESD Symposium*, pp. 157–161, 1990.

[Bowers68] H.C. Bowers, “Space-charge-induced negative resistance in avalanche diodes”, *IEEE T. Electron Dev.*, **ED-15**, 343–350, 1968.

[Bridgewood85] M. Bridgewood and R. Kelly, “Modeling the Effects of Narrow Impulsive Overstress on Capacitive Test Structures”, in *Proceedings of the 7th EOS/ESD Symposium*, pp. 84–91, 1985.

[Brown72] W.D. Brown, “Semiconductor device degradation by high amplitude current pulses”, *IEEE T. Nucl. Sci.*, **NS-19**, 68–75, Dec.1972.

[Burgess60] R. Burgess, “Negative resistance in semiconductor devices”, *Can. J. Phy.*, **38**, 369–375, 1960.

[Carslaw59] H.S. Carslaw and J.C. Jaeger, *Conduction of Heat in Solids*, 2nd edn., Oxford University Press, London, 1959.

[Caruso74] A. Caruso, P. Spirito and G. Vitale, “Negative resistance induced by avalanche injection in bulk semiconductors”, *IEEE T. Electron Dev.*, **ED-21**, 578–586, 1974.

[Caughey67] D.M. Caughey and R.E. Thomas, *Proc. IEEE*, **55**, 2192–2193, 1967.

[Cavone94] M. Cavone, M. Nuschietiello, R. Rivoir and M. Stucchi, “A Method For Characterization And Evaluation Of ESD Protection Structures And Networks”, in *Proceedings of the 16th EOS/ESD Symposium*, pp. 292–300, 1994.

[Chaine92] M. Chaine, S. Desai, C. Dunn, D. Dolby, W. Holland and T. Pekny, “ESD Improvement Using Low Concentrations of Arsenic Implantation in CMOS Output Buffers”, in *Proceedings of the 14th EOS/ESD Symposium*, pp. 136–142, 1992.

[Chaine97] M. Chaine, S. Smith and A. Bui, “Unique ESD Failure Mechanisms During Negative to VCC HBM Stress”, in *Proceedings of the 19th EOS/ESD Symposium*, pp. 346–355, 1997.

[Chang88] P.-H. Chang, R. Hawkins, T. Bonifield and L. Melton, “Aluminum spiking at contact windows in Al/Ti-W/Si”, *App. Phys. Lett.*, **52**, 272–274, 1988.

[Chatt91] A. Chatterjee and T. Polgreen, “A low-voltage triggering SCR for on-chip ESD protection at output and input pads”, *IEEE Elec. Dev. Lett.*, **EDL-12**, 21–22, 1991.

[Chatterjee91] A. Chatterjee, T. Polgreen, A. Amerasekera, “Design and simulation of a 4 kV ESD protection circuit for 0.8 μm BiCMOS process”, *Tech. Dig. IEDM*, 913–916, 1991.

[Chen71] H.C. Chen, W.M. Portnoy and D.K. Ferry, “Doping dependence of second breakdown in a *p-n* junction”, *Sol. St. Elec.*, **14**, 747–751, 1971.

[Chen86] K.L. Chen, G. Giles and D. Scott, “Electrostatic discharge protection for 1-μm CMOS devices and circuits”, *Tech. Dig. IEDM*, 484–487, 1986.

[Chen88] K-L.Chen, “Effect of interconnect process and snapback voltage on the ESD failure threshold of nMOS transistors”, in *Proceedings of the 10th EOS/ESD Symposium*, pp. 212–219, 1988.

[Colvin93] J. Colvin, “The Identification and Analysis of Latent ESD Damage on CMOS Input Gates”, in *Proceedings of the 15th EOS/ESD Symposium*, pp. 109–116, 1993.

[Cook93] C. Cook and S. Daniel, "Characterization of New Failure Mechanisms Arising From Power-Pin ESD Stressing", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 149–156, 1993.

[Crowell66] C. Crowell and S.M. Sze, "Temperature dependence of avalanche multiplication in semiconductors", *Appl. Phys. Lett.*, **9**, 242–233, 1966.

[Dabral93] S. Dabral and R. Aslett, "Designing On-Chip Power Supply Coupling Diodes for ESD Protection and Noise Immunity", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 239–250, 1993.

[Daniel90] S. Daniel and G. Krieger, "Process and Design Optimization for Advanced CMOS I/O ESD Protection Devices", in *Proceedings of the 12th EOS/ESD Symp.*, pp. 206–213, 1990.

[DeChiaro81] L.F. De Chiaro, "Electrothermomigration in nMOS LSI devices", *Proc. 19th IRPS*, 212–217, 1981.

[Diaz93A] C. Diaz, C. Duvvury and S.-M. Kang, "Studies of EOS Susceptibility in 0.6 µm nMOS ED I/O Protection Structures", in *Proceedings of the 15th EOS/ESD Symp.*, pp. 83–91, 1993.

[Diaz93B] C. Diaz, C. Duvvury and S.-M. Kang, "Electrothermal simulation of electrical overstress in advanced nMOS ESD I/O protection circuits", in *Tech. Dig. IEDM*, 899–902, 1993.

[Duvv86] C. Duvvury, R. McPhee, D. Baglee and R. Rountree, "ESD protection reliability in 1 µm CMOS technologies", *Proc. 24th IRPS*, 199–205, 1986.

[Duvv91] C. Duvvury and R. Rountree, "A Synthesis of ESD input protection scheme", in *Proceedings of the 13th EOS/ESD Symposium*, 1991.

[DuvvIRPS88] C. Duvvury, R. Rountree and O. Adams, "Internal chip ESD phenomena beyond the protection circuit", *IEEE T. Electron Dev.*, **ED-35**, 2133–2139, 1988.

[Duvvury83] C. Duvvury, R.N. Rountree and L.S. White, "A summary of most effective electrostatic protection circuits for MOS memories and their observed failure modes", in *Proceedings of the 5th EOS/ESD Symposium*, pp. 181–184, 1983.

[Duvvury85] C. Duvvury, R. Rountree, D. Baglee, A. Hyslop and L. White, "ESD Design Considerations for ULSI", in *Proceedings of the 7th EOS/ESD Symposium*, 45–48, 1985.

[Duvvury86] C. Duvvury, R. McPhee, D. Baglee and R. Rountree, "ESD protection reliability in 1-µm CMOS circuit performance", in *Proc 24th IRPS*, 199–208, 1986.

[Duvvury87A] C. Duvvury, R.N. Rountree, Y. Fong and R.A. McPhee, "ESD phenomena and protection issues in CMOS output buffers", in *Proc 25th IRPS*, 174–180, 1987.

[Duvvury87B] C. Duvvury, R.N. Rountree and R.A. McPhee, "ESD protection: design and layout issues for VLSI circuits", *IEEE T. Ind. Appl.*, **IA-25**, 41–47, 1987.

[Duvvury88A] C. Duvvury, R. Rountree and O. Adams, "Internal chip ESD phenomena beyond the protection circuit", *IEEE T. Electron Dev.*, **ED-35**, 2133–2139, 1988.

[Duvvury88B] C. Duvvury and R. Rountree, "Output ESD Protection Techniques for Advanced CMOS Processes", in *Proceedings of the 10th EOS/ESD Symposium*, pp. 206–211, 1988.

[Duvvury89] C. Duvvury, R. Rountree, H. Stiegler, T. Polgreen and D. Corum, "ESD phenomenon in graded junction devices", in *Proc 27th IRPS*, 71–76, 1989.

[Duvvury92A] C. Duvvury and C. Diaz, "Dynamic gate-coupled nMOS for efficient output ESD protection", in *Proc. 30th IRPS*, 141–150, 1992.

[Duvvury92B] C. Duvvury, C. Diaz and T. Haddock, "Achieving uniform nMOS power distribution for improving ESD/EOS reliability", in *Tech. Dig. IEDM*, 131–134, 1992.

[Dwyer90] V.M. Dwyer, A.J. Franklin and D.S. Campbell, "Thermal failure in semiconductor devices", *Sol. St. Elec.*, **33**, 553–560, 1990.

[English66] A.C. English, "Physical investigation of the mesoplasma in silicon", *IEEE T. Electron Dev.*, **ED-13**, 662–667, 1966.

[Feng86] W.-S. Feng, T.Y. Chan and C. Hu, "MOSFET drain breakdown voltage", in *Elec. Dev. Lett.*, **EDL-7**, 449–452, 1986.

[Fong87] Y.-P. Fong and C. Hu, "The Effects of High Electric Field Transients on Thin Gate Oxide MOSFETs", in *Proceedings of the 9th EOS/ESD Symposium*, pp. 252–257, 1987.

[Franklin190] A.J. Franklin, *"Electrical overstress in GaAs MESFET structures"*, PhD Thesis, Loughborough University, U.K., 1990.

[Geballe55] T.H. Geballe and G.W. Hull, "Seebeck effect in silicon", *Phys. Rev.*, **98**, 940–947, 1955.

[Ghandhi77] S.K. Ghandhi, *"Semiconductor Power Devices"*, John Wiley & Sons, New York, 1977.

[Grant73] W.N. Grant, "Electron and hole ionization rates in epitaxial silicon at high electric fields", *Sol. St. Elec.*, **16**, 1189–1203, 1973.

[Hannemann90] M. Hannemann and A. Amerasekera, "Photon emission as a tool for ESD failure localization and as a techniques for studying ESD phenomena", in *Proc. 1st European Symp. on Rel. and Failure Anal.*, 77–84, 1990.

[Hannemann90] M. Hannemann and E.A. Amerasekera, "Photon emission as a tool for ESD failure localisation and as a technique for studying ESD phenomenon", *Proceedings of the 1st Eur. Symp. on Semiconductor Rel. (ESREF)*, pp. 77–83, 1990.

[Hower80] P.L. Hower and V.G.K. Reddi, "Avalanche injection and second breakdown in transistors", *IEEE Trans. Elec. Dec.*, **ED-17**, 320–335, 1970.

[Hsu82] F.C. Hsu, P.K. Ko, S. Tam, C. Hu and R. Muller, "An analytical breakdown model for short-channel MOSFETs", *IEEE T. Electron Dev.*, **ED-29**, 1735–1740, 1982.

[Hulett81] T.V. Hulett, "On-Chip Protection of nMOS Devices", in *Proceedings of the 3rd EOS/ESD Symposium*, pp. 90–96, 1981.

[Ishizuka94] H. Ishizuka, K. Okuyama and K. Kubota, "Photon emission study of ESD protection devices under second breakdown conditions", in *Proc 32nd IRPS*, 286–291, 1994.

[Jacoboni77] C. Jacoboni, C. Canali, G. Ottaviani and A. Quaranta, "A review of some charge transport properties of silicon", *Sol. St. Elec.*, **20**, 77–89, 1977.

[Jaffe90] M. Jaffe and P.E. Cottrell, "Electrostatic discharge protection in a 4-Mbit DRAM", in *Proceedings of the 12th EOS/ESD Symposium*, pp. 218–223, 1990.

[Keller81] J.K. Keller, "Protection of MOS Integrated Circuits from Destruction by Electrical Discharge", in *Proceedings of the 3rd EOS/ESD Symposium*, pp. 73–79, 1981.

[Khurana66] B.S. Khurana, T. Sugano and H. Yanai, "Thermal breakdown in silicon *p-n* junction devices", *IEEE T. Electron Dev.*, **ED-13**, 763–770, 1966.

[Khurana85] N. Khurana, T. Maloney and W. Yeh, "ESD on CHMOS devices — equivalent circuits, physical models and failure mechanisms", *Proc. 23rd IRPS*, 212–223, 1985.

[Khurana85] N. Khurana, T. Maloney and W. Yeh, "ESD on CHMOS devices — equivalent circuits, physical models and failure mechanisms", *Proc. 23rd IRPS*, 212–223, 1985.

[Kiefer93] S. Kiefer, R. Milburn and K. Rackley, "EOS Induced Polysilicon Migration in VLSI Gate Arrays", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 123–128, 1993.

[Knight80] E.R. Knight and P.P. Budenstein, "Effect of junction spikes and doping level on the second breakdown susceptibility of silicon-on-sapphire diodes", in *Proceedings of the 2nd EOS/ESD Symposium*, pp. 122–129, 1980.

[Koyanagi77] K. Koyanagi, K. Hane and T. Suzuki, "Boundary conditions between current mode and thermal mode second breakdown in epitaxial planar transistors", *IEEE T. Electron Dev.*, **ED-24**, 672–678, 1977.

[Krabbenborg91] B. Krabbenborg, R. Beltman, P. Wolbert and T. Mouthan, "Physics of Electro-Thermal Effects in ESD Protection Devices", in *Proceedings of the 13th EOS/ESD Symp.*, pp. 98–103, 1991.

[Krakauer92] D. Krakauer and K. Mistry, "ESD Protection in a 3.3 V Sub-Micron Silicided CMOS Technology", in *Proceedings of the 14th EOS/ESD Symp.*, pp. 250–257, 1992.

[Krakauer94] D. Krakauer and K. Mistry, "Circuit Interactions During Electrostatic Discharge", in *Proceedings of the 16th EOS/ESD Symposium*, 1994.

[Krieger87] G. Krieger, "Thermal response of integrated circuit input devices to an electrostatic energy pulse", *IEEE T. Electron Dev.*, **ED-34**, 877–882, 1987.

[Kuper93] F. Kuper, J. Bruines and J. Luchies, "Suppression of Soft failures in a Submicron CMOS Process", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 117–122, 1993.

[Kurimoto94] K. Kurimoto, K. Yamashita, I. Miyanaga, A. Hori and S. Odanaka, "An electrothermal circuit simulation using an equivalent thermal network for electrostatic discharge", in *Proc. VLSI Symp. on Tech.*, 127–128, 1994.

[Lau82] C.K. Lau, Y.C. See, D. Scott, J. Bridges, S. Perna and R.D. Davies, "Titanium disilicide self-aligned source/drain and gate CMOS technology", *Tech. Dig. IEDM*, 714–717, 1982.

[Laux87] S.E. Laux and F.H. Gaensslen, "A study of channel avalanche breakdown in scaled n-MOSFETs", *IEEE T. Electron Dev.*, **ED-34**, 1066–1073, 1987.

[Leblanc91] J.P. LeBlanc and M.D. Chaine, "Proximity effects of unused output buffers on ESD performance", *Proc. 29th IRPS*, 327–330, 1991.

[Lin86] C.M. Lin, K. Chi, L. Richardson, R. Simcoe, J. Averill, L. Brooke and D. Nelsen, "Design and Test Results For a Robust CMOS VLSI Input Protection Network", in *Proceedings of the 8th EOS/ESD Symposium*, pp. 188–192, 1988.

[Lin90] D.L. Lin, "Thermal breakdown of VLSI by ESD pulses", *Proc. 29th IRPS*, 281–287, 1990.

[Maene92] N. Maene, J. Vandenbroeck and L. van den Bempt, "On Electrostatic Discharge Protections for Inputs, Outputs and Supplies of CMOS Circuits", in *Proceedings of the 14th EOS/ESD Symposium*, pp. 228–233, 1992.

[Maloney85] T. Maloney and N. Khurana, "Transmission Line Pulsing Techniques for Circuit Modeling of ESD Phenomena", *Proceedings of the 8th EOS/ESD Symposium*, pp. 49–54, 1985.

[Maloney86] T. Maloney, "Contact Injection: A Major Cause of ESD Failure in Integrated Circuits", in *Proceedings of the 8th EOS/ESD Symposium*, pp. 166–172, 1986.

[Maloney88] T.J. Maloney, "Designing MOS inputs and outputs to avoid oxide failure in the charged device model", in *Proceedings of the 10th EOS/ESD Symposium*, pp. 220–227, 1988.

[Mathews80] D. Mathews, "Some design criteria for avoiding second breakdown in bipolar devices", in *Proceedings of the 2nd EOS/ESD Symposium*, pp. 117–121, 1980.

[Mayaram91] K. Mayaram, J.-H. Chern, L. Arledge and P. Yang, "Electrothermal simulation tools for analysis and design of ESD protection devices", in *Tech. Dig. IEDM*, 909–912, 1991.

[McPhee86] R. McPhee, C. Duvvury and R.N. Rountree, "Thick Oxide Device ESD Performance Under Process Variations", in *Proceedings of the 8th EOS/ESD Symposium*, pp. 173–181, 1986.

[MEDICI92] *MEDICI: 2-Dimensional Device Simulation Program*, Technology Modeling Associates Inc., Palo Alto, California, 1992.

[Melchior64] H. Melchior and M.J.O. Strutt, "Secondary breakdown in transistors", *Proc. IEEE (Correspondence)*, **52**, 439–440, 1964.

[Merrill93] R. Merrill and E. Issaq, "ESD Design Methodology", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 233–237, 1993.

[Minear78] R.L. Minear and G.A. Dodson, *"The phantom emitter - an ESD resistant bipolar transistor design and its applications to linear integrated circuits"*, pp. 188–192, 1978.

[Nienhuis66] R.J. Nienhuis, "Second breakdown in the forward and reverse base current region", *IEEE T. Electron Dev.*, **ED-13**, 655–662, 1966.

[Ohtani90] S. Ohtani and M. Yoshida, "Model of Leakage Current in LDD Output MOSFET Due to Low-Level ESD Stress", in *Proceedings of the 12th EOS/ESD Symposium*, pp. 177–181, 1990.

[Okuto75] Y. Okuto and C.R. Crowell, "Threshold energy effect on avalanche breakdown voltage in semiconductor junctions", *Sol. St. Elec.*, **18**, 161–168, 1975.

[Orvis83] W.J. Orvis, C.F. McConaghy, J.H. Yee, G.H. Khanaka, L.C. Martin and D.H. Lair, "Modeling and testing for second breakdown phenomena", in *Proceedings of the 5th EOS/ESD Symposium*, pp. 108–117, 1983.

[Orvis85] W.J. Orvis, C.F. McConaghy, J.H. Yee, G.H. Khanaka, L.C. Martin and D.L. Lair, "Modeling and testing for second breakdown phenomena", *Proceedings of the 7th EOS/ESD Symposium*, pp. 108–117, 1985.

[Palella85] A. Palella and H. Domingos, "A Design Methodology for ESD Protection Networks", in *Proceedings of the 7th EOS/ESD Symposium*, pp. 169–174, 1985.

[Pierce84] D.G. Pierce, "Modeling metallization burn-out of integrated circuits", *Proceedings of the 6th EOS/ESD Symposium*, pp. 56–61, 1984.

[Pierce85] D.G. Pierce, "Electro-Thermomigration as an Electrical Overstress Failure Mechanism", in *Proceedings of the 5th EOS/ESD Symposium*, pp. 67–76, 1987.

[Pierce85] D.G. Pierce, "Electrothermomigration as an electrical overstress failure mechanism", *Proceedings of the 7th EOS/ESD Symposium*, pp. 67–76, 1985.

[Pierce88] D.G. Pierce, W. Shiley, B.D. Mulcahy, K.E. Wagner and M. Wunder, "Electrical overstress testing of a 256K UVEPROM to rectangular and double exponential pulses", in *Proceedings of the 10th EOS/ESD Symposium*, pp. 137–146, 1988.

[Polgreen89] T. Polgreen and A. Chatterjee, "Improving the ESD failure threshold of silicided nMOS output transistors by ensuring uniform current flow", in *Proceedings of the 11th EOS/ESD Symposium*, pp. 167–174, 1989.

[Popescu70] C. Popescu, "The thermal runaway mechanism of second breakdown phenomenon", *Sol. St. Elec.*, **13**, 887–901, 1970.

[Puvvada98] V. Puvvada and C. Duvvury, "A Simulation Study of HBM Failure in an Internal Clock Buffer and the Design Issues for Efficient Power Pin Protection Strategy", in *Proceedings of the 20th EOS/ESD Symposium*, pp. 104–110, 1998.

[Rafferty93] C.S. Rafferty, H.-H. Vuong, S.A. Eshraghi, M. Giles, M. Pinto and S. Hillenius, "Explanation of reverse short-channel effect by defect gradients", in *Tech. Dig. IEDM*, 311–314, 1993.

[Renn89] R.G. Renninger, M.-C. Jon, D.L. Lin, T. Diep and T.L. Welsher, "A field-induced charged device model simulator", *Proceedings of the 11th EOS/ESD Symposium*, pp. 59–71, 1989.

[Ridley63] B.K. Ridley, "Specific negative resistance in solids", *Proceedings of the Phys. Soc*, **82**, 954–966, 1963.

[Rieck89] G. Rieck and R. Manely, "Novel ESD protection for advanced CMOS output drivers", *Proceedings of the 11th EOS/ESD Symposium*, pp. 182–189, 1989.

[Rooz90] L.J. van Roozendaal, E.A. Amerasekera, P. Bos, W. Baelde, F. Bontekoe, P. Kersten, E. Korma, P. Rommers, P. Krys, U. Weber and P. Ashby, "Standard ESD testing", *Proceedings of the 12th EOS/ESD Symposium*, pp. 119–130, 1990.

[Roozendaal90] L.J. van Roozendaal, E.A. Amerasekera, P. Bos, W. Baelde, F. Bontekoe, P. Kersten, E. Korma, P. Rommers, P. Krys, U. Weber and P. Ashby, "Standard ESD testing", in *Proceedings of the 12th EOS/ESD Symposium*, pp. 119–130, 1990.

[Roun88] R. Rountree, C. Duvvury, T. Maki and H. Stiegler, "A process-tolerant input protection circuit for advanced CMOS processes", *Proceedings of the 10th EOS/ESD Symposium*, pp. 201–211, 1988.

[Rountree85] R.N. Rountree and C.L. Hutchins, "nMOS protection circuitry", *IEEE T. Electron Dev.*, **ED-32**, 910–917, 1985.

[Rountree88] R. Rountree, "ESD protection for submicron CMOS circuits issues and solutions", *Tech. Dig. IEDM*, 580–583, 1988.

[Rountree88] R.N. Rountree, C. Duvvury, T. Maki, and H. Stiegler, "A process tolerant input protection circuit for advanced CMOS processes", in *Proceedings of the 10th EOS/ESD Symposium*, pp. 201–205, 1988.

[Runyan65] W.B. Runyan, *Silicon Semiconductor Technology*, McGraw-Hill, New York, 1965.

[Scarlett63A] R.M. Scarlett, W. Shockley and R.H. Haitz, "Thermal instabilities and hot spots in junction transistors", in *Physics of Failure in Electronics*, ed: M.F. Goldberg and J. Vaccaro, Spartan Books, Baltimore, Md., pp. 194–203, 1963.

[Scarlett63B] R.M. Scarlett and W. Shockley, "Secondary breakdown and hot spots in power transistors", *IEEE Int. Conv. Rec. part 3*, 3–13, 1963.

[Schafft62] H.A. Schafft and J.C. French, "Second breakdown in transistors", *IRE Trans. Elec. Dev.*, **ED-9**, 129–136, 1962.

[Schafft66] H.A. Schafft and J.C. French, "A survey of second breakdown", *IEEE T. Electron Dev.*, **ED-13**, 613–618, 1966.

[Schroen66] W. Schroen and R.M. Scarlett, "Second breakdown in simplified transistor structures and diodes", *IEEE T. Electron Dev.*, **ED-13**, 619–626, 1966.

[Schutz82] A. Schutz, S. Selberherr and H. Potzl, "Analysis of breakdown in MOSFETs", *IEEE Comp. Aid. Des.*, **CAD-1**, 77–84, 1982.

[Scott86] D. Scott, J. Hall and G. Giles, "A Lumped Element Model for Simulation of ESD Failures in Silicided Devices", in *Proceedings of the 8th EOS/ESD Symposium*, pp. 41–47, 1986.

[Scott87] D. Scott, R. Chapman, C.-C. Wei, S.S. Mahant-Shetti, R. Haken and T. Holloway, "Titanium disilicide contact resistivity and its impact on 1-µm CMOS circuit performance", *IEEE T. Electron Dev.*, **ED-34**, 562–574, 1987.

[Selberherr84] S. Selberherr, *Analysis and simulation of semiconductor devices*, Springer-Verlag, New York, 1984.

[Shabde84] S.N. Shabde, G. Simmons, A. Baluni and D. Back, "Snapback induced gate dielectric breakdown in graded junction MOS structures", *Proceedings of the 22nd IRPS*, 165–168, 1984.

[Shaw92] M.P. Shaw, V.V. Mitin, E. Schöll and H.L. Grubin, *The Physics of Instabilities in Solid State Electron Devices*, Plenum, New York, 1992.

[Smith73] W.B. Smith, D.H. Pontius and P.P. Budenstein, "Second breakdown and damage in junction devices", *IEEE T. Electron Dev.*, **ED-20**, 731–744, 1973.

[Smith73] W.B. Smith, D.H. Pontius and P.P. Budenstein, "Second breakdown and damage in junction devices", *IEEE T. Electron Dev.*, **ED-20**, 731–744, 1973.

[Speakman74] T.S. Speakman, "A model for the failure of bipolar silicon integrated circuits subjected to electrostatic discharge", in *Proceedings of the 12th IRPS*, 60–69, 1974.

[Steele62] M.C. Steele, K. Ando and M.A. Lampert, "Avalanche breakdown double injection induced negative resistance of semiconductors", *J. Phys. Soc. Jpn.*, **17**, 1729–1736, 1962.

[Stratton72] R. Stratton, "Semiconductor current-flow equations (diffusion and degeneracy)", *IEEE T. Electron Dev.*, **ED-19**, 1288–1300, 1972.

[Strauss87] M. Strauss, D. Lin and T. Welsher, "Variations in Failure Modes and Cumulative Effects Produced by Commercial Human Body Model ESD Simulators", in *Proceedings of the 9th EOS/ESD Symposium*, pp. 59–63, 1987.

[Sze81] S.M. Sze, *Physics of Semiconductor Devices*, John Wiley & Sons, New York, 1981.

[Tasca70] D.M. Tasca, "Pulse power failure modes in semiconductors", *IEEE T. Nucl. Sci.*, **NS-17**, 364–372, Dec. 1970.

[Tasca72] D.M. Tasca, J.C. Peden and J. Miletta, "Non-destructive screening for thermal second breakdown", *IEEE T. Nucl. Sci.*, **NS-19**, 57–67, 1972.

[Tauc57] J. Tauc and A. Abraham, "Thermal breakdown in silicon *p-n* junctions", *Phys. Rev.*, **108**, 936–937, 1957.

[Thornton58] C.G. Thornton and C.D. Simmons, "A new high current mode of transistor operation", *IRE Trans. Elec. Dev.*, **ED-5**, 6–10, 1958.

[Ting01] L. Ting, C. Duvvury, O. Trevino, J. Schichl, and T. Diep, "Integration of TLP Analysis for ESD Troubleshooting Structures", in *Proceedings of the 23rd EOS/ESD Symposium*, 2001.

[Toyabe78] T. Toyabe, K. Yamaguchi, S. Asai and M. Mock, “A two-dimensional avalanche breakdown model of submicron MOSFETs”, *IEEE T. Electron Dev.*, **ED-25**, 825–833, 1978.

[Verhaege93] K. Verhaege, P. Roussel, G. Groensenecken, H. Maes, H. Gieser, C. Russ, P. Egger, X. Guggenmos and F. Kuper, “Analysis of HBM ESD Testers and Specifications Using 4th Order Lumped Element Model”, in *Proceedings of the 15th EOS/ESD Symposium*, pp. 129–138, 1993.

[Verhaege94] K. Verhaege, G. Groeseneken, H. Maes, P. Egger and H. Gieser, “Influence of Tester, Test Method and Device Type on CDM ESD Testing” in *Proceedings of the 16th EOS/ESD Symposium*, 1994.

[Voldman92] S. Voldman, V. Gross, M.J. Hargrove, J.M. Never, J.A. Slinkman, M.P. O’Boyle, T.S. Scott and J.J. Delecki, “Shallow Trench Isolation Double-Diode Electrostatic Discharge Circuit and Interaction with DRAM Output Circuitry”, in *Proceedings of the 14th EOS/ESD Symposium*, pp. 277–288, 1992.

[Voldman93] S. Voldman and V. Gross, “Scaling Optimization and Design Considerations of Electrostatic Discharge Protection Circuits in CMOS Technology”, in *Proceedings of the 15th EOS/ESD Symposium*, pp. 251–260, 1993.

[Voldman94] S. Voldman *et al.*, “ESD Protection in a Multi-Rail Disconnected Power Grid and Mixed Voltage Interface Environments in 0.5 and 0.25 μm Channel Length CMOS Technologies”, in *Proceedings of the 16th EOS/ESD Symposium*, 1993.

[Ward2] A.L. Ward, “Studies of second breakdown in silicon diodes”, *IEEE T. Parts Hyb. Pac.*, PHP-13, **4**, 361–368, 1977.

[Ward76] A.L. Ward, “An electrothermal model of second breakdown”, *IEEE T. Nucl. Sci.*, **NS-23**, 1679–1684, 1976.

[Ward77] A.L. Ward, “Studies of second breakdown in silicon diodes”, *IEEE T. Parts. Hyb. Pac.*, **PHP-13**, 361–368, Dec, 1977.

[Watchutka90] G. Wachutka, “Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling”, *IEEE Trans. Comp. Aid. Des.*, **CAD-9**, 1141–1149, 1990.

[Wei92] Y. Wei, Y. Loh and C. Wang, “MOSFET Drain Engineering for ESD Performance”, in *Proceedings of the 14th EOS/ESD Symposium*, pp. 143–148, 1992.

[Wills88] K. Wills, C. Duvvury and O. Adams, “Photon Emission Testing for ESD Failures, Advantages and Limitations”, in *Proceedings of the 10th EOS/ESD Symposium*, pp. 53–61, 1988.

[Wilson87] D. Wilson, H. Domingos and M.M.S. Hassan, “Electrical Overstress in nMOS Silicided Devices”, in *Proceedings of the 9th EOS/ESD Symposium*, pp. 265–273, 1987.

[Wunsch68] D.C. Wunsch and R.R. Bell, “Determination of threshold failure levels of semiconductor diodes and transistors due to pulse voltages”, *IEEE T. Nucl. Sci.*, **NS-15**, 244, Dec, 1968.

[Yee82] J. Yee, W. Orvis, L. Martin and J. Peterson, “Modeling of Current and Thermal Mode Second Breakdown Phenomena”, in *Proceedings of the 4th EOS/ESD Symposium*, pp. 76–81, 1982.

9 Influence of Processing on ESD

Ajith Amerasekera

9.1 INTRODUCTION

ESD protection circuits are required to have high impedance during normal operation of the chip, and a very low impedance during an ESD event. During the low impedance phase, the protection circuit elements must shunt >1 A of current, for a duration of ≈ 100 ns. Such conditions are outside the typical operating regions of most circuit elements in a submicron CMOS or bipolar process; however, their behavior at these current levels will determine the ESD capability of the protection circuit. Nonlinearity of the operation at high currents and temperatures applies to all protection circuit elements and this behavior is strongly process and structure dependent. The high electric fields, current densities, and temperatures result in a strong sensitivity to the structure of the circuit element and the process technology. Hence, it would be observed that a circuit that functions well in one technology may have poor performance in another technology, or in different fabrication facilities. Process dependencies are especially true when translating ESD protection circuits through technology shrinks. Hence, an understanding of the key process parameters influencing ESD is essential to the development of ESD protection circuit design methodology [Gupta98][Amerasekera00].

Most ESD protection circuits depend on the action of various parasitic elements to provide the necessary current shunting and voltage clamping. Even in circuits with current paths defined by nonparasitic elements, such as *pn* diodes or large area MOS transistors, parasitic elements in the chip beyond the protection circuit will eventually trigger and influence the ESD behavior. Figure 9.1 shows a cross section of an nMOS transistor and the associated parasitic lateral *npn* (LNPN) device. The collector junction is reverse-biased and the base current, to sustain the LNPN in the on-state, is provided by avalanche generation in the reverse-biased collector (drain) junction. As discussed in Section 4.6, the temperature rise in the junction and the onset of thermal breakdown are dependent on the power density, $J \cdot E$ (W/cm^3), in the junction. The boundary

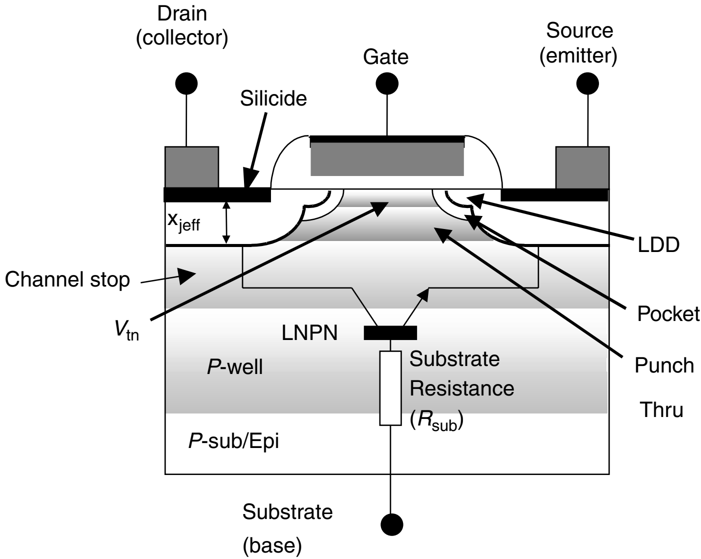


Figure 9.1 Cross section of an nMOS transistor showing the process implant levels and the parasitic bipolar transistor

condition for damage is the eventual collapse of the current in a localized region leading to the formation of melt filaments. Both the current density, J, and the electric field, E, are functions of the doping profiles of the junctions and the substrate, and the properties of the *npn* and the layout of the circuit [Shabde84][Duvvury89][Ohtani90][Amerasekera90][Amerasekera96C][Chen98].

The ballast resistance between the contact and the collector junction where the heating takes place can influence the onset of current localization and failure. Hence, the contact resistance and the source/drain sheet resistance are important process parameters. Advanced CMOS technologies with silicide-clad source and drain diffusions (Figure 9.1) have lower drain resistance and better speed of the transistors [Lau82]. Unfortunately, this also results in a drastic reduction in the ESD thresholds of previously high performing protection circuits elements in non-silicided technologies [Duvvury85][McPhee86][Chen86].

Resistance has an important part in the voltage clamping limits of ESD protection circuits. Resistors are either dedicated as part of the protection circuit to ensure appropriate sequencing between primary and secondary protection circuits, or can be part of the metal interconnect, substrate, *n*-well, or other indirect current path. Changes in the dedicated or parasitic resistance can lead to the voltage clamp limit being reached at a lower ESD current level, or a circuit element reaching its current limit before the main protection element is fully operational, thus resulting in ESD failure. For example, in a *pn* diode protection element, the resistance of the *pn* diode determines the maximum current (and thus the

ESD voltage) that the circuit can handle before the voltage clamp limit is reached [Voldman93][Dabral93][Voldman94][Dabral98].

The main current path during the ESD event is through the circuits (e.g., output buffers) connected to the bond pads. In a CMOS technology, these are the nMOS and pMOS transistors, and because of the characteristics of the associated lateral bipolar transistor, the nMOS transistor tends to be the preferred path for the ESD current, whether intended or not. As a consequence, the nMOS has been the workhorse of ESD protection circuits for many years. Protection circuits that do not use the nMOS as the primary protection circuit have to find means to limit its exposure to the ESD current through circuit design approaches [Dabral98][Smith99][Maloney99].

In this chapter, the phenomena and mechanisms involved in each of these effects will be presented, with emphasis on the nMOS transistor, because that tends to be the limiting factor in the ESD capability of most ICs.

9.2 HIGH CURRENT BEHAVIOR

9.2.1 nMOS Transistor

The high current behavior of an nMOS transistor (Figure 9.2), with gate voltage $V_g = 0\,\text{V}$, shows the triggering of the LNPN after drain junction avalanche begins at a drain voltage $V_d = V_{t1}$. The drain voltage then drop to the *snapback* holding voltage V_{sp}, at which the LNPN operates in the self-biased mode, that is, the substrate current I_{sub} required to raise the substrate potential V_{sub} and forward-bias the emitter-base junction, as well as the base current I_b, are provided by the

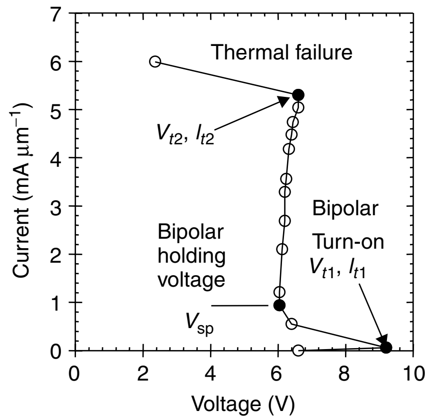


Figure 9.2 High current $I-V$ curve of an nMOS transistor with gate, source, and substrate voltages at 0 V. The bipolar trigger voltage and current are at V_{t1} and I_{t1}, the bipolar holding voltage is at V_{sp}, and the failure threshold occurs at V_{t2}, I_{t2}

avalanche generation current I_{gen} at the collector-base junction (see Section 4), where

$$I_{gen} = I_{sub} + I_b. \quad (9.1)$$

The injected ESD current I_D is conducted through the MOS in the form of the MOS current I_{DS} and the bipolar current I_C. At high injection conditions, even with $V_g > 0\,V$, I_{DS} is small compared to I_C. As the injection current is increased, the LNPN clamps the voltage until the heating at the drain-substrate junction (i.e., the collector-base junction of the LNPN) causes thermal or *second breakdown* to occur. The current at which second breakdown occurs is given by I_{t2} as opposed to I_{t1}, which is when the avalanche multiplication triggers snapback and the LNPN turns on. I_{t2} is a measure of the ESD capability of the process and is used extensively as a measure of the high current capability of a transistor. It has been shown experimentally that the current gain of the LNPN (β), the drain-substrate avalanche multiplication factor (M), and the p-substrate (or p-well) resistance (R_{sub}), can be directly related to I_{t2} [Amerasekera96C][Gupta98]. In order to understand the process-related effects on the ESD performance of the parasitic bipolar transistors, LNPN or LPNP, we must know how the process influences these three key parameters M, β, and R_{sub}.

M varies as the drain-substrate electric field. The higher the electric field, the larger M will be. β is a function of the gate length L of the MOS transistor (base width of the LNPN), the source depth and recombination (emitter efficiency of the bipolar), and the doping in the channel and substrate region (base doping). R_{sub} is a function of the p-well doping, the substrate background doping and the spreading resistance to the substrate connection on-chip.

Lower M, β and R_{sub} result in a lower ESD performance for the following reasons [Amerasekera96C]:

- A lower M means that a higher junction voltage is needed to support the I_{gen} required to initiate and sustain parasitic bipolar action. A higher junction voltage means higher power dissipation in snapback and also a higher ESD current before snapback begins.
- A lower R_{sub} means that a higher I_{gen} is needed to generate the V_{sub} required to initiate and sustain bipolar action. Once again, this results in a higher junction voltage and higher power for the same injection current.
- A lower β means that a higher base current is needed to initiate and sustain bipolar action, which in turn demands a higher I_{gen}, a higher junction voltage and higher junction temperatures.

9.2.2 pMOS Transistors

The pMOS transistor (Figure 9.3) has a comparatively weaker parasitic lateral bipolar transistor associated with it. The LPNP is harder to turn on and has a higher holding voltage than the LNPN in the same technology. Hence, the parasitic LPNP

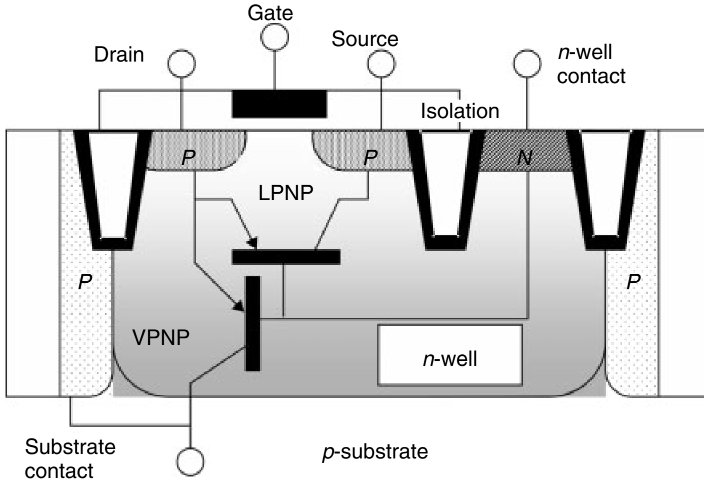


Figure 9.3 Cross section of the pMOS transistor showing the lateral and vertical bipolar transistors. During normal operation in an output buffer, the *n*-well and source will be connected to the power supply, while the drain would be connected to the bondpad

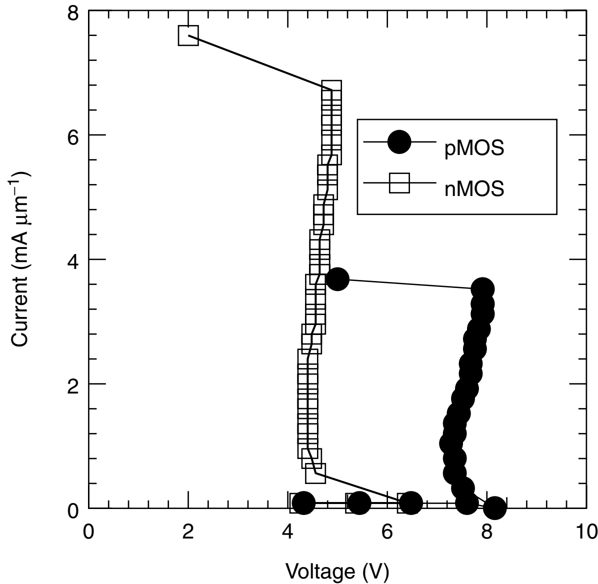


Figure 9.4 High current $I-V$ curve of a pMOS transistor and comparison to an nMOS with the same gate length and gate width in a 0.18 µm CMOS process. The lateral *pnp* has a higher bipolar holding voltage and a lower failure threshold than the lateral *npn*

does not carry the bulk of the ESD current unless the protection circuit is specifically designed to make that happen [Maloney99].

The high current behavior of a pMOS transistor in a 0.18 μm process is shown in Figure 9.4, compared to that of an nMOS. Note that the parasitic bipolar transistor does turn-on, but that the voltage drop in snapback is significantly less than for the LNPN. The reason for this is that the LPNP has a low β and the I_{gen} and junction voltage need to be higher to sustain the bipolar condition. A reduction in I_{t2} of about 30% is observed in the pMOS when compared to the nMOS in the 0.18 μm process.

In a conventional CMOS technology, the pMOS also has a vertical *pnp* associated with it, formed with the drain of the pMOS as the emitter, the *n*-well as the base, and the *p*-substrate as the collector (Figure 9.3). Both the LPNP and the VPNP can be triggered in the nonavalanche condition when the voltage on the drain (emitter) of the *pnp* goes positive compared to the *n*-well (base) voltage. This would happen for a positive strike between the signal pads and the power supplies. The nonavalanching operation of the parasitic *pnp*s have been successfully used in protection circuits [Voldman92][Dabral93][Dabral98].

9.3 CROSS SECTION OF A MOS TRANSISTOR

The implant levels in a deep submicron nMOS transistor are shown in Figure 9.1. The purpose of the LDD is to reduce the electric field at the drain junction during normal operating conditions and improve the robustness to channel hot carrier degradation of the MOS. The lightly doped drain (LDD) region is self-aligned to the polysilicon gate edge and implanted using the polysilicon gate to define the junctions. Arsenic or phosphorus can be the dopant species in an nMOS, while boron is the dopant species in pMOS LDD. In a typical LDD structure, doses range from $10^{13}/\text{cm}^2$ to $10^{14}/\text{cm}^2$. The LDD region is relatively shallow, being anywhere from 5-nm to 50-nm deep, compared to the more highly doped drain/source regions that form the main part of the transistor.

The drain/source regions are formed by arsenic (nMOS) or boron (pMOS). Junction depths are between 0.10 μm and 0.20 μm. The main source/drain implant in an nMOS usually consists of arsenic with doses in the $2 \times 10^{15}/\text{cm}^2$ to $5 \times 10^{15}/\text{cm}^2$ range. In order to ensure that the LDD region is not overdoped by the drain/source implant, the implant is self-aligned to a spacer deposited against the gate. The spacer can be between 0.05 μm and 0.25 μm in thickness depending on the feature size of the technology. Both silicon dioxide and silicon nitride spacers are used in advanced technologies.

At technology nodes >1 μm, phosphorus-only drain-source junctions with different dopant gradings were also used to eliminate the need for the sidewall oxide but still achieve channel hot carrier robustness. Another non-LDD approach is the *double-diffused* junction with both arsenic and phosphorus drain-source implants. However, these were too deep to support true submicron transistor operation and

are no longer common. These types of transistors were the source of a lot of the early work on process sensitivities of ESD and the results published from that work helped to develop the understanding that was used later in developing deep submicron ESD robustness.

The surface of the silicon under the polysilicon gate is the channel, and the threshold (turn-on) voltage of the MOS transistor is adjusted by means of a threshold voltage implant V_{tn}. The V_{tn} implant is a higher doping level of the same type (p or n) as the background doping of the channel region. The punch-through implant goes below the V_{tn} region and raises the substrate doping level to prevent the drain depletion region from extending across to the source region. Another implant now common in deep submicron transistors is the *pocket* or *halo* implant that surrounds the LDD implant region. This is required for very short channel transistors, to prevent the drain-induced barrier lowering (DIBL) effect that limits the transistor scaling. The pocket implant is a higher dose of the background (channel) species, that is, boron for nMOS transistors.

Below the drain/source junctions, we have the channel stop implant, which is intended to improve the isolation between adjacent transistors. Again, it is a higher dose version of the background doping, that is, boron in an nMOS transistor. The next level down is the well implant doping. In deep submicron technologies, a retrograde well is used whereby the peak doping is about 1 μm from the surface. This has the purpose of reducing the spacing between adjacent transistors, and lowering the substrate resistance underneath the transistors for the improvement of latchup robustness without affecting the dopings nearer the surface that are critical to transistor operation.

The transistor structure is placed in a p-type substrate region that is either a low-doped *epitaxial* layer on top of a highly doped substrate, or directly into a more lightly doped *bulk* substrate. Epitaxial layers are a few microns thick, and have benefits of lower substrate resistance (because of the very highly doped substrates), but are more expensive than using bulk substrates. Bulk substrates have the disadvantage of being more resistive and sensitive to latchup, but the use of retrograde wells has helped to alleviate that problem.

9.4 DRAIN-SOURCE IMPLANT EFFECTS

9.4.1 Background

In technologies with feature sizes greater than 2 μm, highly doped abrupt junctions were formed without LDD regions. These junctions were on the order of 0.5 μm deep and had high doping concentrations. Typically, avalanche breakdown occurred uniformly through the junction depth, and the parasitic bipolar action utilized the entire junction sidewall. While not being equivalent in high current performance to an epitaxial bipolar transistor, these MOS devices had high ESD capability. The first indications of the importance of the source/drain junction profiles to

ESD performance came in sub-2 μm technologies with the introduction of graded junctions and LDD regions in an attempt to achieve aggressive technology scaling and to reduce the channel hot carrier sensitivity [Shabde84][Duvvury86].

Early analysis on the effect of drain/source junction grading on high current robustness and ESD thresholds, showed that as the phosphorus implant dose was increased from 0 to $4 \times 10^{13}/cm^2$, the current required to cause breakdown decreased [Shabde84]. The ESD performance of a given protection circuit was lower. Figure 9.5 shows the effect of the introduction of LDD and silicides on ESD performance for 2 μm and 1 μm technologies [Duvvury86]. The 2 μm process had abrupt junctions while the 1 μm processes had LDD junctions. In addition, the effect of silicide vs. nonsilicide is shown for the 1 μm process. The difference in performance between the LDD and the abrupt junction devices was ∼1.5.

In a later paper, the effect of the As drain-source implants on the ESD performance were also shown, indicating that the junction grading had a negative impact on ESD [Duvvury89]. Results showed that more graded junctions using a phosphorus-only implant with a dose of $5 \times 10^{14}/cm^2$ had an average ESD level of about $10\,V\mu m^{-1}$, compared to almost $15\,V\mu m^{-1}$ for more abrupt junctions with both phosphorus and a $2 \times 10^{15}/cm^2$ As implant. These early devices did not use a separate LDD implant, trying to achieve the profiles necessary for reducing hot carrier effects purely from grading the drain-source implant. Devices with higher phosphorus doping levels of $1.2 \times 10^{15}/cm^2$ and no arsenic had higher

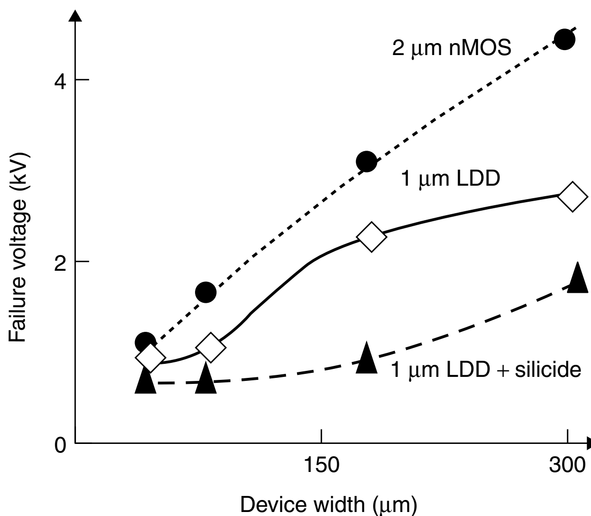


Figure 9.5 Effect of LDD and silicided diffusions on ESD performance. The 2 μm process did not have LDD or silicides, while the 1 μm process with LDD and no silicide shows more ESD robustness than the same transistor with LDD and silicide (After [Duvvury86])

$V\mu m^{-1}$ levels but were still not as good as that of devices with the arsenic implant.

9.4.2 Submicron Technologies

Results comparing double-diffused drain structures, where the arsenic and phosphorus are both implanted into the drain-source area, with LDD structures and graded junction devices for a 1 μm process [Chen88] showed that double-diffused transistors are better by a factor of almost three than equivalent LDD devices.

The variation of ESD performance with the phosphorus LDD dose in general seems to indicate that, for a nonsilicided 0.5 μm process, as the phosphorus dose is increased from $10^{13}/cm^2$ to $4 \times 10^{13}/cm^2$ the ESD performance increases as shown in Figure 9.6 [Ishizuka94]. However, earlier work [Shabde84] showed that in the same range the ESD performance decreased. Similar trends were observed by Ohtani *et al* [Ohtani90] for phosphorus doses from $4 \times 10^{12}/cm^2$ to $5 \times 10^{14}/cm^2$. The $4 \times 10^{12}/cm^2$ dose and the $5 \times 10^{14}/cm^2$ dose were shown to give high ESD thresholds while an intermediate $5 \times 10^{13}/cm^2$ dose gave a low ESD threshold. It has been shown [Ishizuka94] that devices with a phosphorus LDD dose are much weaker than those using the same arsenic LDD dose. Higher phosphorus LDD doses do indeed improve the ESD performance of these devices.

The difference in the results between these studies may be explained by the influence of the p-well doping profile, which would not be the same for the different processes. A higher p-well concentration will result in a lower avalanche breakdown voltage for the same LDD dose, and will certainly reduce the gain of the LNPN transistor. A nonuniform p-well profile will favor current paths, which

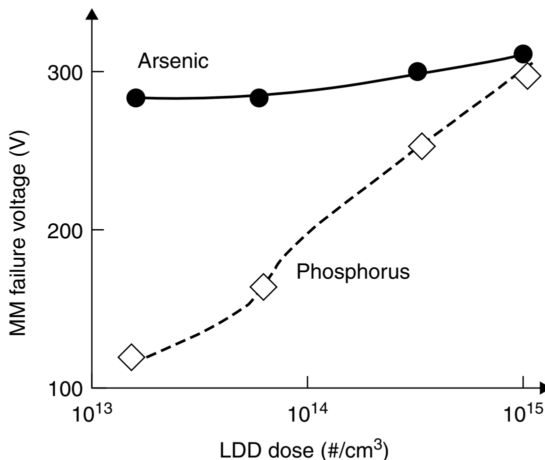


Figure 9.6 Machine Model (MM) ESD failure voltage as a function of As LDD and P LDD doses (After [Ishizuka94])

may be different to those of a uniform p-well. It is not easy to identify the exact p-well profile in the active transistor regions, especially in submicron technologies [Rafferty93], but it could be responsible for some of the variations observed in the reported results. It is important to observe that even though the drain-source implant experiments reported in the literature may look similar, the exact structure and the process conditions need to be fully comprehended to be able to compare results. Also of significance to note is that process changes in parts of the device could reverse trends due to processing elsewhere in the device. Thus, it may be possible to find the sweet spot for transistor performance and ESD robustness in a given transistor design. The M, β, R_{sub} relationship in Section 9.3 describes these trade-offs.

In addition to the work on LDDs, there have been results reported on the effect of the arsenic drain-source implant dose on ESD [Chaine92]. These results need to be compared with the work on graded junctions discussed earlier [Chen88][Duvvury89]. As the drain-source arsenic dose is reduced, the resistance of the drain increases providing more ballast for the drain junction. Thus, current localization is inhibited and higher second breakdown thresholds are achieved.

One approach to improving the ESD performance of devices with LDD has been to use the double-diffused drain (DDD) where an additional high phosphorus dose $5 \times 10^{14}/\text{cm}^2$ to $10^{15}/\text{cm}^2$ is implanted into the drain/source region. The intention of the DDD implant is to make the junction deeper as well as to overdope the lightly doped region, thus creating a drain profile similar to those of the abrupt junction technologies [Daniel90][Amerasekera90][Wei92]. Results have shown that the additional (ESD) implant can improve ESD performance by more than a factor of two, even in a fully silicided process [Amerasekera90][Amerasekera91]. As in the LDD case, the optimum dose and energy for the ESD implant will vary depending on differences in the p-well doping concentrations and the p-well doping profiles. A higher p-well concentration requires a higher ESD implant dose than a lower p-well concentration because of the reduced depletion region width and its impact on the turn-on of the LNPN transistor.

The problem with the ESD implant approach is that the additional phosphorus implant tends to significantly change transistor performance, especially the short channel effects and hot carrier reliability. Hence, the implementation requires an additional mask to block the ESD implant from the critical transistors on the chip and adds to process complexity and cost. It is best to try to integrate ESD robustness into the technology as a whole by optimizing the drain/source and LDD doses for transistor performance and the whole structure for ESD rather than use special process steps [Amerasekera94A][Gupta98][Bock99].

In deep submicron technologies, the implementation of the pocket implant (Figure 9.1) affects the ESD performance. While it was initially thought that the pocket implant would benefit the ESD performance because it would increase M, a number of recent studies have shown that the pocket implant had a negative effect as shown in Figure 9.7 [Gupta98][Consiglio95]. This is caused by the reduction

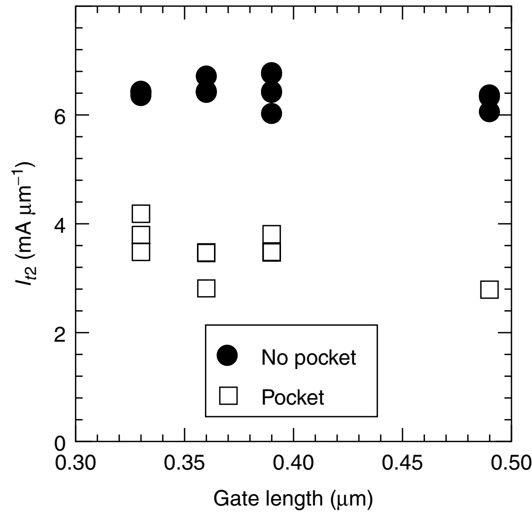


Figure 9.7 Effect of pocket (halo) implant on ESD performance of nMOS transistors of varying gate length in a deep submicron technology

in the β of the LNPN due to the higher p^+ in the channel region, that more than compensates for the increase in M.

9.4.3 Summary of effect of drain-source implants on ESD

The effect of the drain-source implants on ESD performance can be summarized in the following:

(a) It is observed that deeper junctions with higher doping levels have better ESD performance. Relating this to the three parameters M, β and R_{sub}, the reason for deeper junctions being better is that they have improved bipolar performance due to a higher β, and increased robustness to current localization and thermal runaway [Amerasekera96C][Gupta98][Noterman99]. In general, this is true for both LNPN and VPNP transistors.

(b) More abrupt drain junctions improve ESD capability. This is related to the M parameter. A more abrupt junction has a higher electric field for the same junction voltage. Since M is directly related to the electric field, a higher multiplication factor is obtained at this voltage, which provides more I_{sub} for self-biasing the LNPN as well as the sustaining base current for the same power dissipation [Gupta98].

(c) Higher channel doping can reduce ESD capability through lower β, although M may be increased. This is observed for pocket implants in deep submicron technologies and indicates that there is an optimum transistor design possible with M, β and R_{sub} [Amerasekera99].

9.5 *p*-WELL EFFECTS

The *p*-well resistance and doping concentration influences both the R_{sub} and the β of the LNPN. Increasing the *p*-well doping concentration can have a strong negative impact on ESD performance as shown in Figure 9.8, especially if there is little margin left within the M and β dependencies [Gupta98][Amerasekera00][Bock97]. Furthermore, the *p*-well resistance will influence the layout requirements for protection circuits, and the design of the protection circuit. Sensitivities to *p*-well resistance can lead to variable performance in the ESD protection circuit, if not fully comprehended in the design and implementation of the protection strategy.

In Figure 9.8, it is seen that the I_{t2} is lower for the higher well doping, due to the reduction in R_{sub}. Also, the sensitivity of I_{t2} to increased gate length L is increased with the higher doping. Since longer L means lower β, the data indicates that as the LNPN gain becomes marginal, the *p*-well resistance becomes the dominant parameter in the ESD performance of the devices.

Until the 0.50-μm CMOS technology node, the wells were mostly diffused. The resultant doping profile was higher at the surface and lower at the bottom of the well. In sub-0.25-μm technologies, the wells are implanted and the profile is now *retrograded*. That is, the peak doping occurs about 1 μm from the surface, where it has the most benefit for transistor isolation, and latchup. The use of retrograde wells has been shown to benefit ESD capability in comparison to diffused wells [Bock97][Gupta98], because the lower doping at the surface improves the LNPN gain and provides a higher local substrate resistance close to the junctions.

The *p*-well is typically designed to optimize transistor performance and isolation, together with the latchup immunity of the circuits. Both of these can run counter to

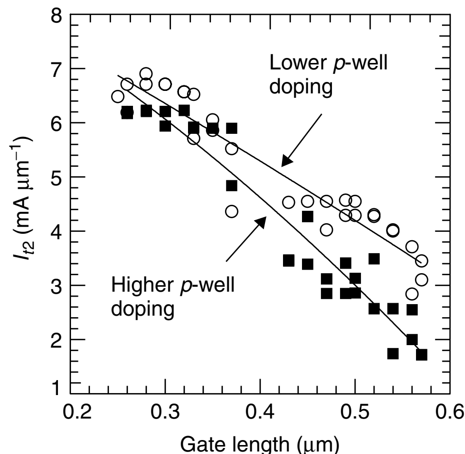


Figure 9.8 ESD performance as a function of nMOS gate length for two different *p*-well doping concentrations

the needs for best ESD performance. Higher p-well resistance is generally better for ESD performance in nMOS (LNPN) based protection circuits, but will lower latchup immunity. In addition, protection circuit strategies that are based on the vertical *pnp* or pMOS approaches would benefit from lower p-well resistance. The choice of protection circuit strategy, as well as the technology requirements must be considered in determining what the p-well design needs are for ESD performance.

9.6 n-WELL EFFECTS

n-Well dopant concentrations and depths have a large impact on protection circuits using dual-diode schemes [Voldman92][Voldman93]. Figure 9.9 shows the ESD performance as a function of the n-well sheet resistance for a dual-diode protection circuit [Voldman92]. As the sheet resistance is decreased from 1100 Ω/ □ to 330 Ω/ □, the ESD performance increases from 2.5 kV to nearly 7 kV. However, higher n-well doping will have an effect on junction capacitance and influence circuit speed. The n-well design must, therefore, be optimized for both ESD performance and transistor/circuit performance.

The vertical *pnp* is inherent in diode protection structures (Figure 9.10), and has also been used as a trigger element for substrate-triggered LNPN approaches [Voldman92][Dabral98][Amerasekera95][Duvvury00]. The main parameter influencing the VPNP operation is the n-well doping concentration, which directly impacts the current gain β. Higher peak doping concentrations in the n-well will reduce β. Shallower wells on the other hand will increase β, but these need to be considered against the higher doping associated with shallower wells. The β

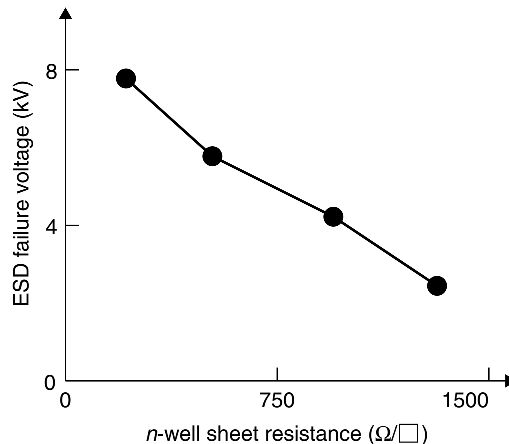


Figure 9.9 ESD performance of a dual-diode protection circuit as a function of the n-well sheet resistance (After [Voldman92])

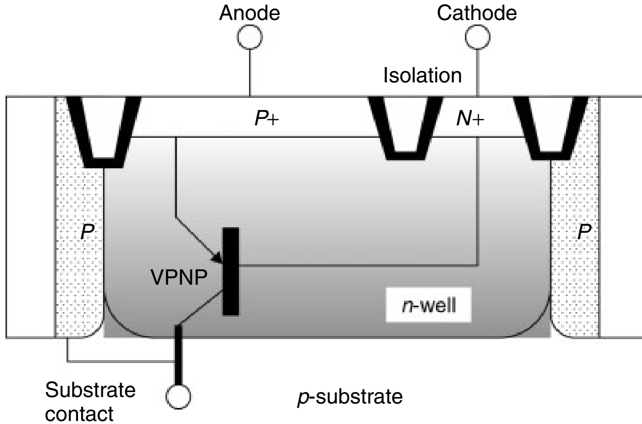


Figure 9.10 Cross section of a *pn* junction diode in an *n*-well showing the vertical PNP structure

roll-off at high current levels is also an important factor that is influenced by the emitter and base engineering of the VPNP.

In SCR protection circuits, increasing the *n*-well doping will have a negative effect on the ability to trigger the SCR and will also increase the SCR holding voltage. This could make SCR protection circuits less effective and the ESD performance of the SCR protection circuit will be affected, particular with regard to the low-level protection capability [Duvvury88]. Indeed, the introduction of retrograde wells with higher peak *n*-well doping concentrations has resulted in SCR-type protection circuits no longer being a good option in sub-0.5 μm technologies.

9.7 EPITAXIAL LAYERS AND SUBSTRATES

Most advanced technologies are based on p^- substrate starting material and the focus here will be on *p*-substrates. Two types of substrates can be used. The first is the epitaxial substrate where a lightly doped *p*-layer is grown on a highly doped p^+ substrate. The second type is a uniformly doped *p*-type substrate known as *bulk*, with typical resistivities used in manufacturing ranging from 1 Ω-cm to 10 Ω-cm.

The resistivity of the epitaxial layer (*epi*) can be about 10 Ω-cm, compared to about 0.01 Ω-cm for the underlying p^+ substrate. The doping of the epitaxial layer and its thickness will define the substrate resistance. It is seen in Figure 9.11 [Amerasekera00][Gupta98], that the ESD performance of an individual nMOS transistor can be increased substantially with thicker epi. The major effect of the substrate resistance is in the turn-on of the parasitic bipolar transistor in the nMOS protection circuits [Toyabe78][Schutz82][Hsu82][Laux87]. However, depending on

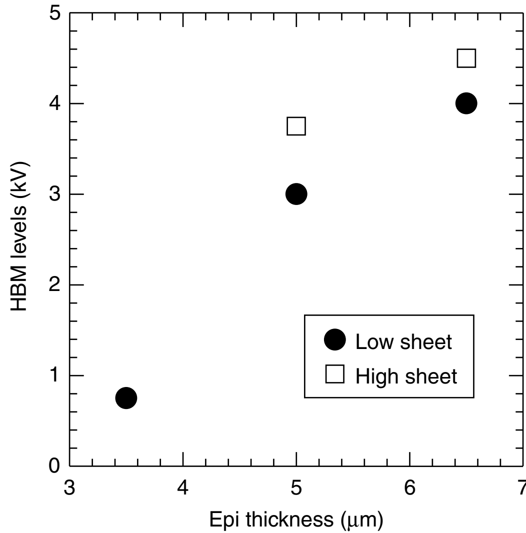


Figure 9.11 ESD performance of an nMOS device as a function of epitaxial thickness for different p-well doping concentrations

the doping of the underlying p^+ substrate, the p-well doping begins to dominate and the ESD performance will not benefit any further. The resistivity of the epi layer will also influence the ESD performance. As seen in Figure 9.11, the difference in epi doping in this case, can provide nearly 30% higher ESD capability. The extent of the influence of the epi and starting material on LNPN devices will depend on how strongly the ESD capability is defined by the R_{sub} compared to the M and β of the LNPN [Gupta98][Amerasekera99].

In dual-diode protection circuits, Figure 9.12 shows about a 50% improvement in ESD performance as the epitaxial thickness is increased from ≈ 2 μm to ≈ 3 μm [Voldman93]. For thinner epi thicknesses, the p^+ substrate dopant compensates the n-well doping and changes the n-well sheet resistance. As the n-well resistance drops, the HBM performance increases significantly. The problem would be that as the thickness is increased, the diode resistance could also increase depending on the layout. If there are not enough contacts to the p-substrate, the diode resistance will be dominated by the resistance of the epi layer, which will lower the effectiveness of the diode-type protection. Note that the extent of the dependence will also depend on the n-well and p-well designs and will vary between different process technologies.

Reducing the epitaxial thickness has the biggest impact on SCR protection circuits. Since one of the primary reasons for using epitaxial layers is to reduce the latchup sensitivity, this is not surprising. The effect comes from the need for a higher trigger current for the SCR, which will mean that the secondary protection circuit elements will be stressed at higher levels and may fail before the SCR

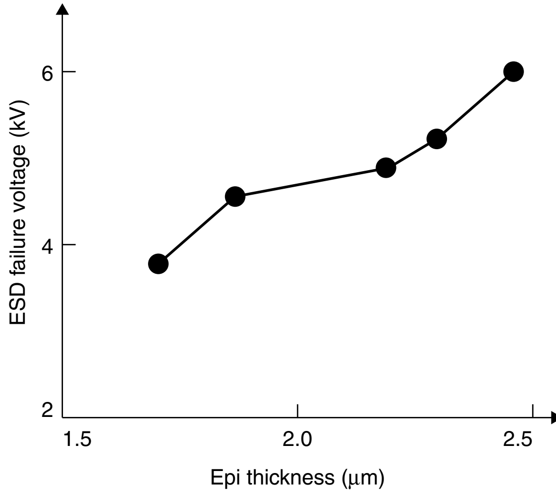


Figure 9.12 ESD performance of a dual-diode protection circuit as a function of epitaxial thickness (After [Voldman92])

triggers [Duvvury88]. In the low voltage SCR, the built-in nMOS transistor will fail if the SCR does not trigger before the current reaches the second breakdown trigger current level for the nMOS. Thinner epi also increases the holding voltage for the SCR (once it triggers). These results indicate that a protection circuit, which provided adequate ESD protection in a thicker epi process may result in poor performance when the epitaxial thickness is reduced unless the design is optimized to accommodate the impact of the epitaxial thickness.

Bulk silicon shows similar dependencies in terms of the resistivity. Higher resistivity has better performance, until the *p*-well sheet begins to dominate.

While it is clear that higher R_{sub} improves the ESD performance of LNPN based protection circuits, it can adversely effect the behavior of the intrinsic *np* substrate diode in the ESD current path. In thin epi processes with relatively low substrate resistance, the p^+ substrate is essentially a low-resistance anode for these diodes and their ESD performance is excellent, with little need to specifically design for it. However, in thicker epi material or in bulk material with higher resistivity, the spreading resistance of the substrate can be significant. Hence, the placement of topside substrate contacts and design/layout requirements for the *np* diode become important if the ESD performance for a negative ESD voltage on the pad with respect to the VSS/substrate is not to be a limiting factor [Worley00]. The implication here is that when migrating protection schemes between technologies, it is necessary to consider all possible ESD stress combinations, and ensure that the process differences will not result in the degradation of what was previously considered to be a robust ESD current path.

9.8 GATE OXIDES

9.8.1 Oxide Thickness and Technology Scaling

Continued scaling of technologies has resulted in an almost 10× reduction in the gate oxide thickness in the period 1990 to 2000. The effect of this oxide thickness scaling on ESD performance is of concern because of the high voltages on the IC during the ESD event. One objective of an ESD protection circuit is to clamp the voltage at an input gate or output buffer, so that the gate oxide breakdown voltage (BV_{ox}) is not exceeded. Under pulsed conditions oxides can withstand higher electric fields than under steady state DC conditions [Bridgewood85][Amerasekera86][Fong87]. Therefore, while a 100 Å gate oxide may have a $BV_{ox} \approx 10$ V under DC conditions, the pulsed breakdown may be as high as 20 V. In general, the drain junction avalanche breakdown voltage, BV_{av}, is lower than BV_{ox} as shown in Figure 9.13 for a 0.25 μm process [Amerasekera99], and a concern is that as the oxide thickness decreases without additional drain-source engineering, BV_{av} can become greater than BV_{ox} and gate oxide rupture will occur during ESD.

Most output circuits will see some snapback action during an ESD event, and the pad voltage drops to allow suitable margins between BV_{av} and BV_{ox} even at current levels of 2 A and greater. However, some processing options can decrease the efficiency of the snapback action. For example, lower epitaxial thickness t_{epi} can raise the trigger voltages and currents of the SCRs used as protection circuits

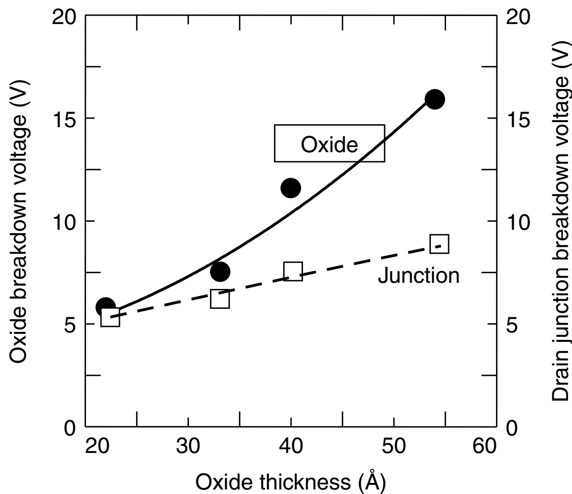


Figure 9.13 Oxide breakdown voltage and drain junction breakdown voltage as a function of oxide thickness in different process technologies. The transistors were optimized for performance in each process

to levels greater than BV_{ox}. Thinner epitaxial layers also make it harder for snapback action to take place in nMOS transistors, and in some cases can eliminate snapback completely. In such cases, without suitable protection design techniques, BV_{ox} can be reached at low ESD levels and oxide breakdown occurs during an ESD event. There are protection strategies that specifically employ antisnapback design techniques to reduce the sensitivity to second breakdown [Smith99][Miller00]. In these circuits, the voltage across the gate oxide of the output or input transistor could reach avalanche breakdown depending on the design of the protection circuit.

Experimental data comparing the behavior of transistors with different gate oxide thicknesses ranging from 12 nm to 3.5 nm have shown that there is no significant effect on the ESD performance for these nMOS devices with thinner gate oxides [Gupta98]. Figure 9.14 shows the I_{t2} for 8 nm and 12 nm showing no difference for the two oxide thicknesses. Figure 9.13 shows the BV_{ox} for gate oxide thicknesses from 5.5 nm down to 2.5 nm, for 200 ns pulse durations. The electric field for oxide breakdown under these conditions is about 20 MV/cm. It is clear that the maximum tolerable voltage is decreasing rapidly as the oxide thickness gets smaller. The sensitivity of very thin oxide transistors to gate oxide breakdown will depend on how long the oxide has to sustain that stress since the maximum electric field is inversely related to the duration of the stress [Fong87][Wu00]. In the case of LNPN turn-on, the sensitivity to gate oxide breakdown at sub 2.5 nm gate oxide thicknesses will depend on how quickly the bipolar can trigger, and the maximum

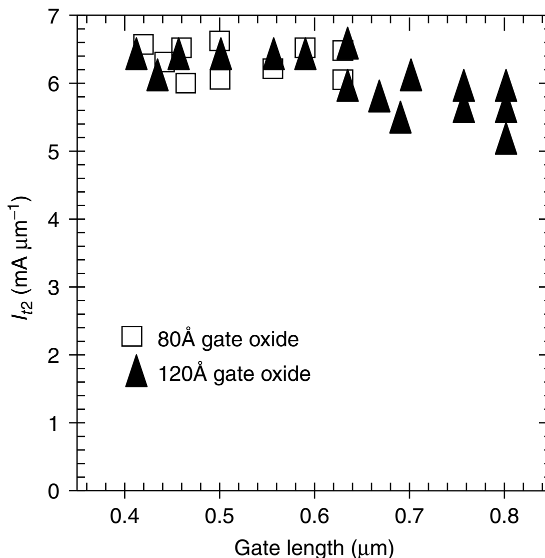


Figure 9.14 I_{t2} as a function of gate length with two different oxide thickness in the same process shows no effect of oxide thickness

drain voltage during this process [Amerasekera99][Wu00]. The factors influencing this would be the M, β and R_{sub} properties of the LNPN, and whether the protection circuit design can turn on the LNPN at lower drain voltages using either gate driven or substrate driven techniques [Chen98][Duvvury00]. Once the LNPN reaches the relatively lower voltage snapback region, there is considerably less sensitivity to gate oxide breakdown [Amerasekera99].

9.8.2 Charged Device Model (CDM) Effects

Another cause of oxide breakdown is when the stress pulse has a very high current level for very short durations, as in the case of the Charged Device Model (CDM) test method [Duvvury96A]. Current levels on the order of 10 A are typical for a 1500 V CDM stress level according to the present industry standard. Under these conditions on-resistance of the protection device in snapback may not be able to prevent the voltage across the oxide from reaching BV_{ox} and oxide breakdown will be observed [Fukuda86]. The reason that oxide breakdown occurs before thermal damage is because the duration of the pulse is less than 1 ns, which does not allow the temperature in the device to reach levels required for thermal breakdown before the voltage reaches BV_{ox}. It is also possible that the pulse duration is less than the time required for the devices to turn on. However, most parasitic devices in advanced CMOS processes trigger at around 100 ps, and this is not necessarily a primary concern. A suitable clamping device needs to be selected, which has a low enough on-resistance that the voltage does not reach BV_{ox} at current levels close to 10 A and durations of $\sim$1 ns.

9.9 SILICIDES

In silicided processes, the source and drain diffusions are clad in metal silicide with the intention of reducing the contact resistance [Lau82]. The first silicides were formed with titanium, giving titanium silicide, $TiSi_2$, while tungsten silicide WSi_2 and cobalt silicide $CoSi_2$ are also used in CMOS processes. The deposition of the refractory metal on the diffusions is self-aligned to the polysilicon gate, and these self-aligned silicided diffusions are also known as *silicides*. The introduction of silicided drain/source diffusions provided the next big discontinuity in ESD performance after the LDD, at about the 1 µm technology node. A dramatic reduction in the ESD performance of protection circuits, which functioned well in nonsilicided processes was observed [Duvvury86][Scott86][Wilson87][Chen88]. Figure 9.5 shows the effect of silicides on the ESD performance of nMOS devices in a 1 µm process. Silicided devices were between 3× and 6× worse than the nonsilicided devices. The main cause of the problem was identified to be the silicide cladding of the source/drain diffusions and not the silicide on the polysilicon gates.

From the ESD viewpoint, the primary effect of the silicide cladding on the diffusions is to bring the contact closer to the gate and the diffusion edge. The consequence is that under high current conditions, the ballasting resistance between the contact and the hot spot is reduced. Hence, once a hot spot is initiated at the diffusion edge, there is very little resistance to prevent current localization through the hot spot. When the temperature at the silicide contact reaches 1000 °C, the silicide begins to decompose, interact with the silicon, or both, in a similar manner to that of aluminum at the eutectic temperature. The higher critical temperature indicates that damage to the silicide itself is not the principle cause of failure.

The mechanisms involved in this degradation have been the source of many studies [Scott86][Scott87][Rountree88][Amerasekera95][Amerasekera96C] [Noterman99]. The contact transfer length, L_c, characterizes the distance over which the current moves from the silicide into the diffusion [Scott87]. It is possible to improve the current distribution in the silicide by changing L_c. Higher values of L_c force the current to flow more in the diffusion, and makes for more uniform current densities at the silicide edge. By moving the current into the diffusion earlier, the effective spreading resistance of the contact is increased, thus improving the ballasting effect of the drain. Processing techniques that increase the silicide sheet resistance while reducing the silicide to diffusion contact resistance forces the current into the diffusion earlier and increases the spreading resistance. Additionally, increasing the junction depth will also increase the spreading resistance of the diffusion region thereby increasing the ESD robustness.

A second effect of silicides is to reduce the effective junction depth, which in turn affects the current gain β and bipolar performance of the LNPN [Chen88][Amerasekera96C]. The actual effect of the silicide processing will vary for different processes since the doping concentrations of the source and drain diffusions and the anneal times will change the silicide thickness and the effective resistance [Scott87]. For example, in a 1-µm process the ESD performance was found to be best for a 600 Å starting Ti thickness, which resulted in $\approx$ 475 Å of $TiSi_2$ assuming 75% of the Ti is converted into $TiSi_2$. Figure 9.15 shows the dependence of HBM ESD thresholds on the $TiSi_2$ thickness for nMOS transistors. For a Ti thickness of 600 Å the ESD voltage was 4.5 kV. In comparison, for an 800 Å starting Ti thickness the ESD pass voltage was 4 kV and a 1000 Å starting thickness had an ESD pass threshold of 1.5 kV. The cause of the decrease in ESD threshold was related to the unsilicided silicon junction depth as shown in Figure 9.16 [Chen88]. The unsilicided junction depth is calculated by subtracting the amount of silicon consumed during the silicidation process from the original junction depth, x_j. A minimum silicon junction depth was required to ensure good ESD performance. Shallower junctions will reduce the current spreading at the collector of the bipolar, while the reduced emitter depth will lower the β through increased emitter recombination. While thicker silicide also means a lower contact resistance and less ballasting in the drain leading to lower ESD performance, it has been shown that even if only the source was silicided,

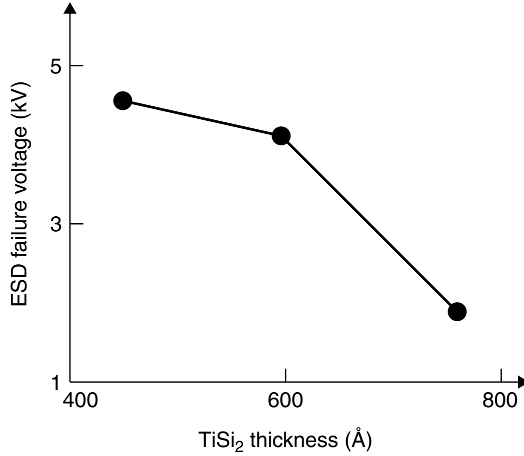


Figure 9.15 Effect of $TiSi_2$ silicide thickness on the ESD performance for nMOS transistors in the same process (After [Chen88]). Thicker silicides result in lower ESD performance for the same transistor

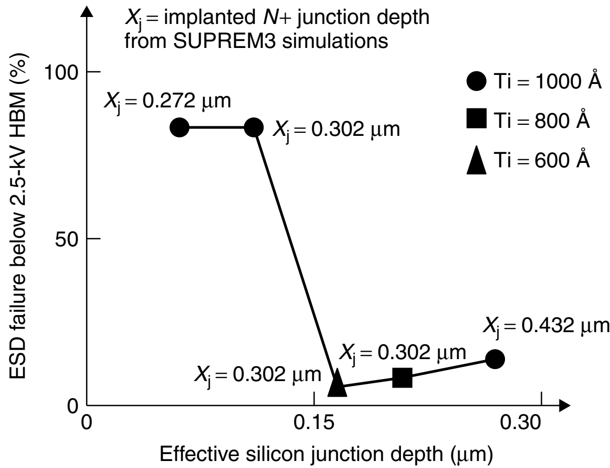


Figure 9.16 Percentage of devices that fail below 2.5 kV HBM as a function of the effective silicon junction depth (After [Chen88]). It is seen that deeper junctions have better performance, and thinner silicides also have better performance

the device has lower high current performance due to the emitter degradation [Amerasekera96C].

As technologies scale and effective junction depths are reduced, the effects of salicidation on junction depths and the contact resistance become more critical to the ESD capability of the parasitic bipolar devices. At sub-0.20-µm gate

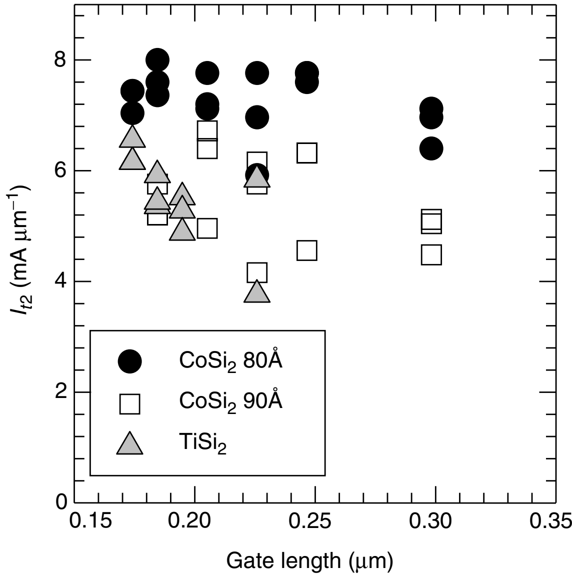


Figure 9.17 Effect of $CoSi_2$ thickness on I_{t2} and comparison to $TiSi_2$ as a function of nMOS gate length. Thinner $CoSi_2$ has higher I_{t2} and less gate length sensitivity

lengths, $CoSi_2$ is the preferred silicide compared to $TiSi_2$ at longer gate lengths. Figure 9.17 shows the effect of thicker cobalt deposition, and hence thicker $CoSi_2$ on the I_{t2} [Amerasekera99]. The figure also shows that $CoSi_2$ can be nearly $2\times$ better for ESD than $TiSi_2$ for the same final silicide thickness, especially for longer gate lengths. This has been shown to be due to a better β for the LNPN in the $CoSi_2$ process compared to $TiSi_2$. The optimization of the silicide thickness, using thinner $CoSi_2$ has been demonstrated to enable good I_{t2} to be obtained at sub-0.18 µm technology nodes without additional process steps or mask levels.

A very effective solution to the silicide problem is to block the silicide from the diffusion areas close to the gate edge. Ideally, the nonsilicided area (silicide block) must be at least 1 µm to take advantage of the increased drain-contact resistance. Silicide block of ≤ 1 µm has been shown to be ineffective in some instances because it does not provide sufficient ballast resistance to prevent current localization as the device approaches second breakdown [Amerasekera92A]. However, with narrower gate lengths and shallower junctions, the optimum distance between the silicide edge and the gate edge can be <1 µm depending on the spreading resistance and β of the bipolar transistor. A second process approach to this problem that has been effective is to increase the junction depth using an additional drain-source implant [Amerasekera91][Daniel90]. The drawback with these solutions is that they require extra masks, as well as increasing the number of process steps and complexity making them not very desirable. Both the silicide blocking mask and

the deeper implant will change the performance of the transistor and can result in an increase in layout area to take care of the transistor behavior and additional mask levels.

9.10 CONTACTS

In technologies with feature sizes >1 μm, the contacts were considered to be the weakest link in the protection circuit since the most extensive damage was usually observed at the contacts [DeChiaro81][Maloney86][Duvvury86][Strauss87]. It was first proposed that contact damage was the result of electrothermomigration [DeChiaro81], which was metal migration into the silicon as a result of a combination of the temperature gradient and the electric field between the contact and the hot spot. While this explanation is still applied to qualitatively explain these failure modes, a detailed analysis [Pierce85] showed that electrothermomigration was not physically possible under ESD conditions, given the time durations of the ESD event. He proposed that metalization burnout and contact damage may actually be a secondary failure mechanism following second breakdown at the junction. The current filamentation, which follows second breakdown leads to very high local temperatures and eventual melting of the contact.

The aluminum-silicon eutectic temperature is about 500 °C, and when the contact reaches this temperature significant metal migration into the contact will occur. It is possible that when the junction temperature is about 1000 °C the thermal gradient between the junction edge and the contact will result in temperatures approaching 500 °C at the contact. This explanation of the contact damage mechanism led to design rules specifying the minimum distance between the contact and the junction edge.

In submicron technologies, contacts usually have barrier metals such as titanium-tungsten (TiW) to prevent contact migration under normal operating conditions. TiW has a much higher eutectic temperature (>1000 °C) closer to the melt temperature of silicon, and hence contact spiking observed in older processes is hardly seen in submicron technologies. In the event that contact spiking is observed in a technology with barrier metals in the contacts, it is probably related to poor step coverage of the TiW layer in the contact [Amerasekera92A]. As an example, the WAl_{12} alloy formed at the top barrier interface can extend all the way to the silicon and Al-Si interdiffusion is then possible during normal processing [Chang88]. With the aluminum now in contact with the silicon, the possibility of contact failure during the ESD event increases since after second breakdown and current localization occurs, the high temperatures can result in the aluminum forming a spike through the diffusion. The use of the tungsten plug in deep submicron technologies has virtually eliminated any possibility of the aluminum coming into contact with the silicon, making contact integrity a nonissue as far as ESD is concerned.

9.11 INTERCONNECT AND METALIZATION

9.11.1 Metal Lines

As technologies and chips scale, and design rules are squeezed, the metal interconnects and vias have to carry higher currents and be less resistive in order to sustain the same ESD levels. Any resistance drop in the metal will reduce the effectiveness of the ESD protection circuit whether it is LNPN or diode-based.

Characterization of the current carrying capability of the interconnect is important in the design and layout of the ESD protection circuit. The maximum current that a given width of metal can sustain is determined by the thermal failure threshold. It has been shown that the thermal failure threshold is a function of the interconnect material as well as the surrounding dielectric [Banerjee96A][Banerjee97A]. It is important to note that if a metal line reaches its melting temperature but does not actually go open circuit due to the constraints of the surrounding dielectric, the electromigration behavior of the metal can be strongly degraded because of the quenching effect of the solid-liquid-solid transition that takes place.

Copper interconnect systems have better ESD current carrying capability than aluminum, and allow more aggressive linewidths to be used in the deep submicron technologies. Figure 9.18 shows the high current behavior of two identical metal lines, one with aluminum and one with copper [Amerasekera00]. For the same

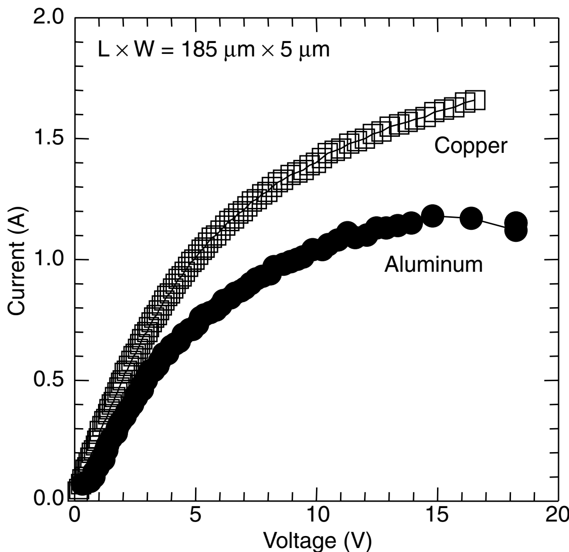


Figure 9.18 High current behavior of same geometry aluminum and copper metal lines. Note the nonlinearity of the resistance at high current levels due to self-heating

geometry, the copper line has a more linear behavior at high currents, and has a higher failure threshold by almost 50%. It is significant to note that the metal resistance is not constant beyond a certain current density, and this nonlinearity will impact ESD protection circuits if it is not taken into consideration in the design.

9.11.2 Interlevel Dielectric

The thermal failure threshold of the metal line is a function of the thermal properties of the metal as well as the capability of surrounding dielectric to dissipate the heat [Banerjee96B]. The thermal conductivity of the surrounding dielectric and proximity to the silicon surface and adjacent metal lines will influence the high current failure thresholds of the metal line. In deep submicron technologies, interlevel dielectrics with lower dielectric constants and hence lower thermal conductivity, will result in lower ESD failure thresholds for the same metal line.

9.11.3 Vias

In general, vias do not show a process-related concern for ESD provided the design of the ESD protection circuit includes a large number of them. Again, it is necessary to characterize the vias and provide the design rules needed to remove them from the critical path in the ESD protection circuit performance [Banerjee97B].

9.12 GATE LENGTH DEPENDENCIES

The MOS gate length L is the base width of the LNPN. The β of the LNPN is inversely proportional to L, and longer L will reduce β and can result in a lower ESD performance [Amerasekera94A][Amerasekera94C]. Figure 9.19 shows I_{t2} as a function of L for 1.8 V, 2.5 V, 3.3 V, and 5 V nMOS transistors in three different production technologies with feature sizes from 0.18 μm to 0.50 μm. This trend has been generally observed across all technology nodes. It has been shown that it is possible to compensate for the reduction in β by increasing one of the other two factors, that is, M or R_{sub}, to sustain good ESD performance [Gupta98][Amerasekera99].

Figure 9.20(a), (b), and (c), show the variation of I_{t2}, M, β and R_{sub} with L for a 0.25 μm technology. It is seen that the reduction in I_{t2} is tracked by a reduction in β as well as a decrease in R_{sub}. R_{sub} is inversely dependent on L due to the lower spreading resistance seen by the substrate current for longer L. A lower β means that more hole current is required to sustain the LNPN action and the M increases. The power dissipation in the junction given by $\approx V_D \times I_D$ increases, for the same injection current I_D, and this can result in a lower I_{t2} depending on the heat sinking capability. Similarly, a lower R_{sub} will also require more hole

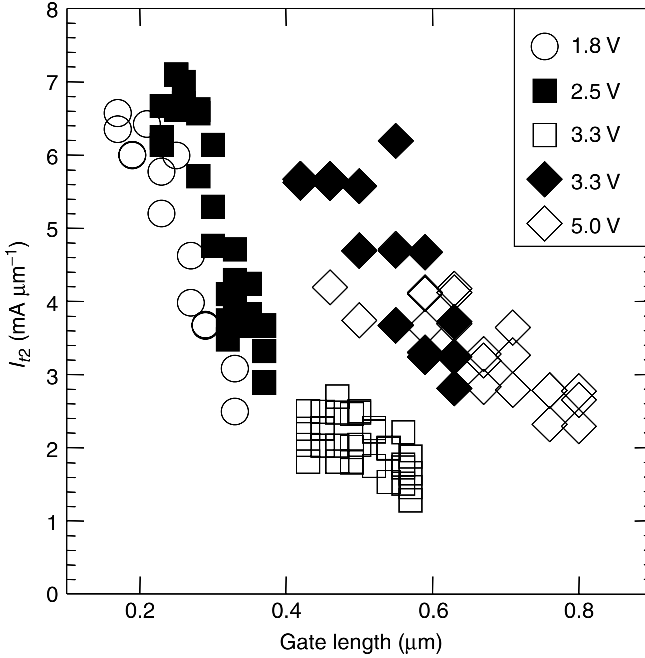


Figure 9.19 I_{t2} as a function of gate length for five different transistors in three different deep submicron technologies

generation (higher I_{sub}) with corresponding higher power dissipation and tendency to lower I_{t2} [Gupta98]).

While in general the inverse dependence of I_{t2} on gate length holds true and has been confirmed in different technologies across the industry, there have been technologies where I_{t2} has been observed to increase with longer L [Bock97]. This indicates that β alone does drive the dependence on L, but the other factors such as R_{sub}, M, and the current flow in the LNPN need to be considered as well. As technologies scale beyond 0.25 μm, and process features such as pocket implants, retrograde wells, ultrashallow junctions and very small silicided contact-gate spacings become standard, the straightforward bipolar dependence will no longer dominate, and all the factors in the turn-on and sustaining of the bipolar need to be included in the analysis [Amerasekera99].

The interaction between the drain depletion region and the source determines whether the conduction mechanism takes place at the surface or below it. While Figure 9.19 shows a monotonically decreasing I_{t2} with L, there are many technologies where I_{t2} shows no dependence on L for shorter gate lengths, and then rapidly decreases when a critical gate length L_{crit} is exceeded [Amerasekera94B][Amerasekera99]. Figure 9.21 shows the I_{t2} variation for 1.8 V and 3.3 V nMOS transistors in two 0.18 μm technologies. The solid symbols show

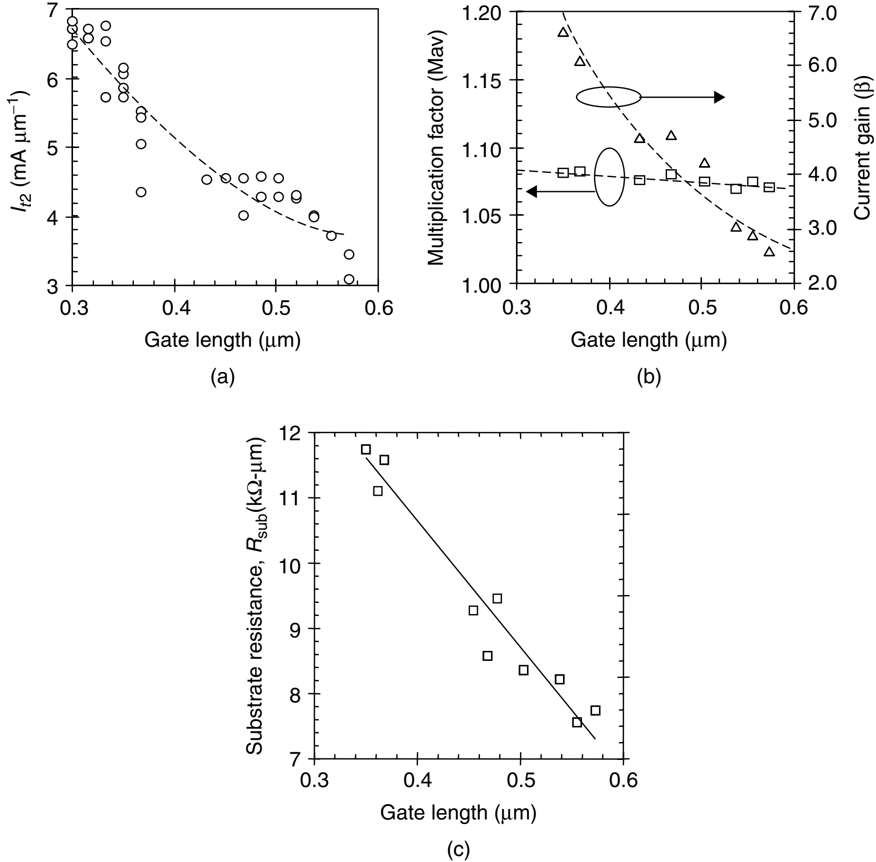


Figure 9.20 Analysis of the effect of transistor parameters on I_{t2} for an nMOS transistor in a 0.18 μm process. (a) I_{t2} as a function of gate length (b) M and β as a function of gate length (c) R_{sub} as a function of gate length

very little dependence on L, while the open symbols show a strong dependence on L. Clearly, the solid symbols imply a more robust technology and is preferred for ESD performance. The influence on processing on the dependence of I_{t2} on L is shown in Figure 9.22, where two 0.35-μm processes are compared. The arsenic-phosphorus LDD shows almost no L dependence over the range 0.4 μm to 0.8 μm, while the phosphorus-only LDD shows a 3× reduction of I_{t2} over the same range. As discussed in Section 9.4, phosphorus-only LDD result in a more graded junction than those with arsenic in them and, therefore, require higher junction voltage to sustain the LNPN action. The value of L_{crit} and the steepness of the I_{t2} roll-off are important process parameters for ESD robustness. If L_{crit} occurs close to the design length, then there will be significant sensitivity to process variations that cause small changes in L. If the I_{t2} roll-off is steep, then a small change in L

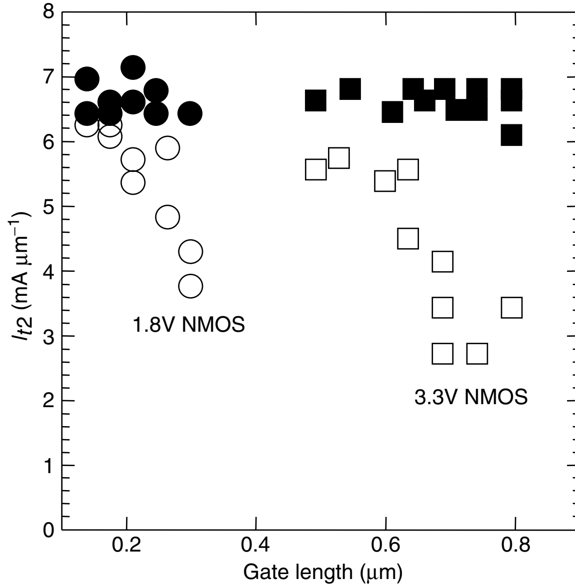


Figure 9.21 Dependence of I_{t2} on L for two different 0.25 μm processes. One shows a strong L dependence while the other has almost no L dependence

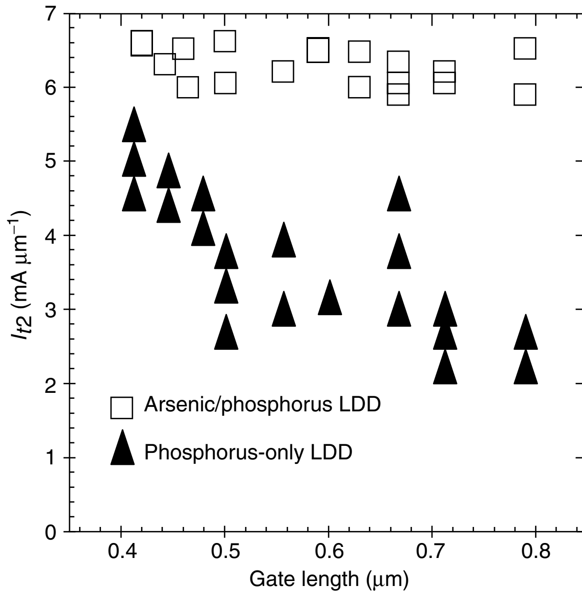


Figure 9.22 The effect of LDD implant type on the dependence of I_{t2} on L

can result in a drastic reduction in I_{t2} and the ESD performance of a protection structure.

Gate voltage or substrate bias in the MOS structure during the ESD event can change these dependencies by providing additional substrate or base current to the bipolar transistor to aid the turn-on and sustain conditions [Amerasekera95]. Protection circuit designs try to provide additional gate or substrate bias circuitry to improve the robustness of the protection circuit by lowering the design sensitivity to gate length or substrate contact, and increasing the process margins to drain-source doping (M, β) and well formation (R_{sub}).

9.13 SILICON-ON-INSULATOR (SOI)

9.13.1 Self-Heating Issues

At the sub-0.20-μm regime, SOI is being considered as a high-speed and low-power solution for high performance ICs. A cross section of an nMOS transistor built-in an SOI technology is shown in Figure 9.23. The ESD performance of SOI devices is one major concern because protection circuits have been observed to have worse ESD performance than those fabricated in standard bulk CMOS [Verhaege93A][Chan94]. The silicon film is between 50 nm and 100 nm thick, and the drain and source regions can reach all the way to the underlying silicon-dioxide layer. The main reason for the lower ESD capability of SOI is that the heat dissipation region in the silicon is surrounded by SiO_2.

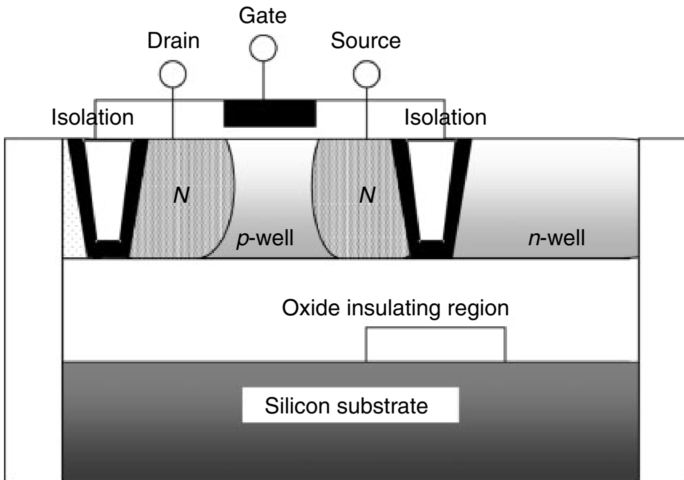


Figure 9.23 Cross section of nMOS transistor in SOI technology. Note that the *p*-well is isolated giving rise to the floating body effect

In bulk devices most of the thermal dissipation takes place through the silicon substrate, which has a reasonably low thermal resistance. In contrast to the silicon substrates in bulk devices, the buried SiO_2 has a very low thermal conductivity, which results in higher temperatures for the same injected current.

9.13.2 ESD Performance

It has been shown that ESD levels of 10 V/µm can be achieved in this SOI nMOS transistors, which are comparable to silicided bulk nMOS transistors [Verhaege93B]. The ESD capability is a function of the channel length, and decreases with increasing channel length as has also been observed in bulk devices. Two important features in SOI devices enable high ESD levels. The *floating body* effect and the *double snapback* phenomenon [Verhaege93A].

Double snapback occurs because of the presence of two parasitic bipolar paths in SOI nMOS devices. The first snapback is associated with the top interface while the second snapback takes place when the potential in the depleted region deeper in the film forms a second bipolar path.

The floating body effect has a great impact on SOI devices with thin silicon films [Duvvury96B][Koizumi00]. Essentially the R_{sub} of the SOI transistor is very high because of the thin silicon film and in some cases would even be floating depending on the type of structure used. The amount of avalanche generated current required to turn-on the bipolar is much lower than for a bulk device, and the voltage V_{t1} for LNPN turn-on is low, so the snapback holding voltage V_{sp} is also low. Figure 9.24 shows the $I-V$ curve for an SOI transistor with grounded gate,

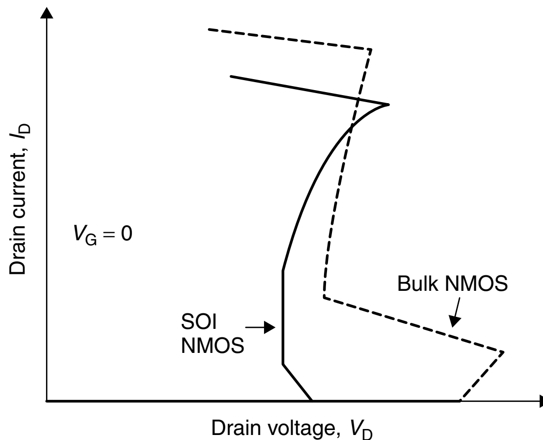


Figure 9.24 High current $I-V$ curve for an SOI nMOS compared to that of an nMOS in a bulk process. The SOI nMOS has a lower bipolar trigger voltage and lower holding voltage, but shows a higher on-resistance due to self-heating and a lower failure threshold

compared to an equivalent bulk transistor also with grounded gate. The low trigger voltage for LNPN turn-on makes it easier to uniformly trigger the entire width of the ESD protection device across multiple fingers. LNPN turn-on due to the floating base effect has been shown to be related to the threshold voltage of the transistor [Koizumi00]. Diode-based protection structures using MOS devices also benefit from a lower threshold voltage, and circuit design techniques may be used to reduce the threshold voltage during an ESD event [Voldman98][Voldman00A]. A lower threshold voltage due to lower channel doping will lead to easier LNPN turn-on and better ESD performance. Although the I_{t2} is lower than in bulk devices because of the poor thermal dissipation in SOI, the LNPN in the SOI process enables good ESD protection to be achieved with not much more effort than for bulk [Duvvury96B]. As in bulk processes, silicides can cause nonuniform turn-on in multifinger structures and a lower ESD performance [Koizumi00]. Note that the parasitic bipolar turn-on due to the floating base region is a problem for standard operation of the nMOS because it results in loss of gate control in the nMOS and *kinks* in the $I_D V_D$ curve. It is a feature that is preferably suppressed for normal operation.

Double snapback occurs because of the presence of two parasitic bipolar paths in SOI nMOS devices. The first snapback is associated with the top interface while the second snapback takes place when the potential in the depleted region deeper in the film forms a second bipolar path.

9.14 BIPOLAR TRANSISTORS

9.14.1 Important Issues

Bipolar transistors have the advantage over CMOS in that the *npn* is a vertical device with majority of power dissipation across a large area junction located beneath the surface of the silicon, compared to the lateral *npn* operating across a very small, shallow junction at or close to the silicon surface. Figure 9.25 shows the cross section of a *npn* bipolar transistor. The main features are the buried N^+ collector region, and the deep N^+ contact pillar to the buried contact region. The P^+ base is extremely shallow, and the emitter is a thin polysilicon film deposited on the base region. In general, the ESD performance of bipolar technologies has been significantly better than for CMOS technologies, and there has not been much attention paid to the important mechanisms as technologies are scaled and process complexities increase. Two of the main process issues are the capability of the process to sustain the vertical bipolar action, which is critical to the intrinsic ESD robustness of bipolar transistors, and the capability to build multifinger structures and have uniform turn-on. The first case is exacerbated by the need to reach higher operating frequencies, demanding lower capacitance, which in turn drives down the size of the input circuit and the ESD protection circuit. It is important, therefore, that bipolar processes continue to maintain, and even increase, their intrinsic ESD robustness through technology scaling.

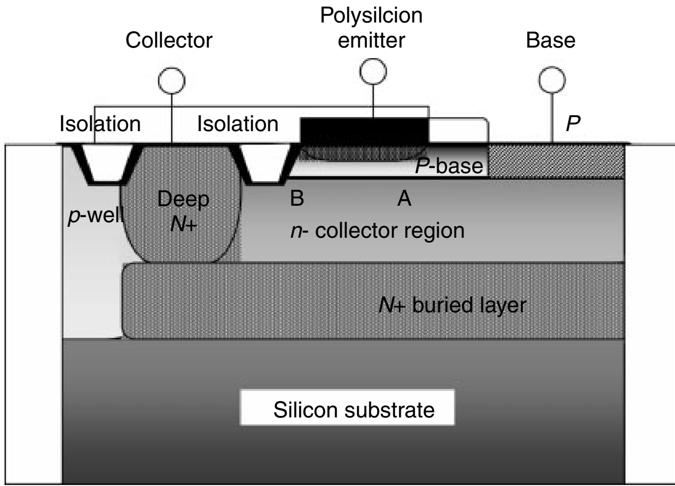


Figure 9.25 Cross section of a bipolar junction transistor *npn*. The self-biased transistor is vertical when triggered by junction breakdown at A. However, if junction breakdown occurs at B, the self-biased transistor will become lateral and have lower ESD performance

9.14.2 Critical Parameters

As technologies shrink, the base region of the transistor has got shallower. Consequently, the peak electric field prior to triggering the bipolar under high current self-biased conditions moves from the bottom of the base directly below the emitter indicate by A in Figure 9.25, to the edge of the base at the deep N^+ pillar close to the surface indicated by B in Figure 9.25. This results in a drastic reduction in the high current capability of the transistor and a rapid decrease in ESD performance. An important parameter in the operation of the bipolar transistor is the BV_{ceo} or the collector-emitter breakdown voltage with base open. BV_{ceo} is a direct function of the base width and the doping in the base. The Early voltage V_{A}, which defines the linearity of the transistor behavior in the on-state is another critical parameter. In order to maintain a high BV_{ceo} and a high V_{A} without reducing the bipolar gain, an additional N^+ implant can be made into the low-doped N^- epitaxial collector region. This implant has the benefit of lowering the bulk collector-base avalanche voltage and improving the vertical conduction path over the lateral path. The parasitic bipolar turn-on voltage is proportional to the N^- well doping concentration reducing from $\approx 14\,\mathrm{V}$ to $\approx 9\,\mathrm{V}$ as the concentration increases from $\approx 5 \times 10^{15}/\mathrm{cm}^3$ to $10^{17}/\mathrm{cm}^3$ [Amerasekera92B]. In general, it is important to note that the critical parameter in advanced high frequency bipolar transistors is the collector doping, since the base and emitter have very little room for process variations. Any process step that enhances the vertical bipolar turn-on, and ensures that very little emitter debiasing occurs at high (ESD-level) current densities, will be beneficial to ESD performance.

At sub-0.25-µm geometries, the use of double-poly processes where the emitter and base were both made of polysilicon caused a reduction in ESD performance. The reason for this is that the emitter is typically surrounded by the base poly region, and the collector-base breakdown and hole current flow can occur without forward-biasing the emitter-base junction. Increase the extrinsic base resistance can significantly increase the ESD performance for these structures.

Silicides have no obvious influence on the ESD performance of bipolar transistors. Similarly, the substrate doping and starting material has little impact on ESD.

Trench isolation increases the thermal resistivity of the *npn* structure, and can make it more sensitive to failure at high currents. Results for SiGe transistors show about a 2× improvement without trench isolation [Voldman00B].

SiGe bipolar transistors are being considered as the main component in building very high frequency ICs on the order of 10 GHz and more [Konig98]. These transistors have a SiGe P^+ base region, which has significant benefits to high frequency operation [Harame95]. It has been shown that SiGe transistors have about a 30% improvement in ESD capability over equivalent Si bipolar transistors, due to the improvement in current gain β, lower base transit time, lower emitter transit time, and lower intrinsic base resistance [Voldman00B]. The difference increases to nearly 200% for shorter stress pulse widths as would be expected for a higher performance transistor.

9.15 DIODES

The effect of processing on nonavalanching diodes used in protection circuits were discussed in the sections on well and substrate effects above (Section 9.5, 9.6, and 9.7). The forward-biased on-resistance of these diodes is the critical parameter, and this is a function of the n-well doping for the *pn* diode in n-well, and the p-well/substrate doping for the *np* diode in substrate. Lower well and substrate resistance will be beneficial to these diodes, however, it is possible to compensate for higher resistance by increasing the size of the diode. Thus, epitaxial substrates have the lowest *np* diode on-resistance, since the highly doped substrate forms an excellent large area contact region for the anode. Bulk substrates will have a higher on-resistance, and because the contact region will be topside, the current flow will depend on the location and size of the anode.

The *pn* diode in n-well has the added vertical *pnp* to the substrate associated with it. If the substrate is grounded during the ESD event, or held at ground potential due to large chip capacitance, the vertical *pnp* will conduct current and help to lower the on-resistance. The benefit of the vertical *pnp* will depend on the gain, which in turn is dependant on the n-well doping concentration (base of the *pnp*), and the depth of the n-well (width of the base). The emitter of the vertical *pnp* is the P^+ diffusion in the n-well, and the emitter efficiency will be influenced by the depth of this diffusion. A deeper emitter has higher gain. Silicides will degrade

the emitter efficiency and reduce the gain of the vertical bipolar. β roll-off at high current is an important factor under ESD conditions, and this is influenced by the emitter and base design. The trade-off between the on-resistance of the emitter-base diode (anode-cathode of the *pn* diode), and the parameters that influence the vertical *pnp* gain need to be considered in process optimization and design/layout rule generation.

9.16 RESISTORS

9.16.1 *n*-well Resistors

The two common resistor types used in ESD protection circuits are the *n*-well resistor and the polysilicon resistor. The *n*-well resistor has a nonlinear $I-V$ characteristic that is a strong function of the *n*-well doping concentration. Figure 9.26 shows the $I-V$ curves of a family of *n*-well resistors of constant width W, and varying head-head spacing L. It is seen that the resistance varies with increasing L, but the current at which the saturation begins is nearly constant with L. The saturation current level is a function of the *n*-well doping, with higher doping giving higher saturation currents. An important feature is the snapback of the resistor when the electric field in the *n*-well reaches avalanche

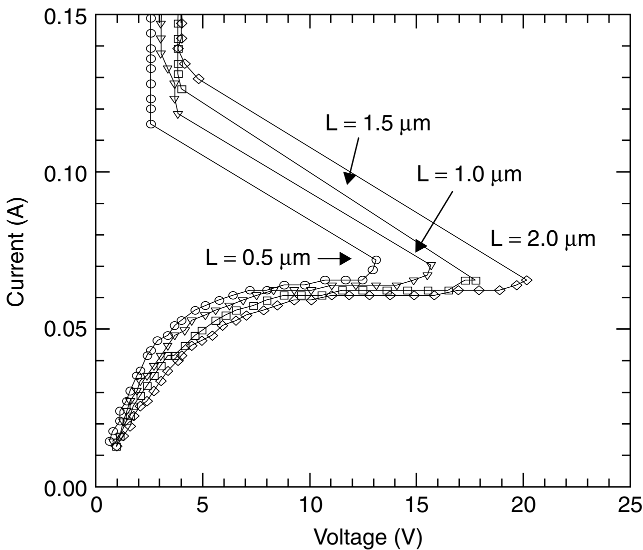


Figure 9.26 High current $I-V$ curves for *n*-Well resistors with different lengths but the same width. The saturation current is constant since it is a function of width, but the breakdown voltage is lower for shorter length structures

generation levels. The snapback voltage marks the limit of useful operation for the resistor, and is a function of L and the n-well doping concentration. When used in protection circuits, changes in the n-well doping concentration can cause big changes in the ESD performance as well as the operation of the output buffers.

9.16.2 Polysilicon Resistors

Polysilicon resistors are generally used without silicide cladding to increase their resistivity and provide better area efficiency. The current carrying limits of the polysilicon resistor are defined by the resistivity and width of the resistor. The wider the resistor, the higher its thermal capability. In the case that the silicide is not removed, the polysilicon resistor will have a lower failure threshold determined by the silicide and polysilicon thickness [Banerjee98].

9.16.3 Other Resistor Types

Diffusion resistors, either n-type of p-type, have been used both cladded and non-cladded in ESD protection circuits. Main limitation is the low junction breakdown voltage, compared to the n-well resistor. The high doping concentration in these resistors results in a more linear $I-V$ curve, but the problems with heat dissipation require design rules similar to those for the polysilicon resistor. Silicided diffusion resistors are more robust than their polysilicon counterparts due to better heat sinking of the silicon substrate [Banerjee98].

Buried resistors are used in high performance output buffers where precision resistance is needed across process and temperature [Sanchez99]. These resistors have better thermal capability, but the junction breakdown voltage will be process dependent.

9.17 RELIABILITY TRADE-OFFS

While optimizing for ESD performance, it is critical to realize that most process improvements that improve ESD will adversely affect other important reliability parameters. The most significant trade-off is the effect on hot carriers [Aur86][Groeseneken00]. Since the ESD performance is degraded by the presence of the LDD while hot carrier is improved by it, there is clearly an optimum point that needs to be found where ESD performance and hot carriers are both acceptable.

Similarly, ESD performance is improved with thicker epitaxial layers and more resistive substrates. However, latchup robustness is reduced by the same parameters. Again, it is essential to find an acceptable point for both ESD and latchup robustness in choice of epi thickness, well doping, and substrate doping.

A properly designed protection circuit should ensure that gate oxide breakdown does not take place, but during the ESD strike the voltage across the gate oxide will be higher than in typical operation. There is concern that in very thin gate oxides, the stress induced leakage current (SILC) at elevated gate oxide voltage levels can cause degradation in the oxide reliability after the ESD event [Wu00].

Interconnect reliability is also affected by ESD, and a reduction in electromigration robustness is possible if the metal melts during ESD stress. The quenching effect of the melting and cooling will change the grain properties of the interconnect and make it more sensitive to electromigration.

It is important to ensure that these limits are understood and taken into account during process development and protection circuit design. While the process must be optimized for ESD, it is not possible to maximize the process for ESD.

9.18 SUMMARY

The high current performance of the primary elements used in ESD protection circuits are very sensitive to the process. This chapter has described the important process and device parameters and their dependencies.

Drain-source engineering

The ESD performance of MOS-type elements operating in high current mode is dependent on the interaction between the multiplication factor M, the parasitic bipolar current gain β, and the substrate resistance R_{sub}. Higher values of M, β and R_{sub} are better for ESD. Process variations that achieve that will improve ESD performance. Specifically, more abrupt junctions (e.g., arsenic) will have better ESD performance than graded junctions (e.g., phosphorus). Deeper junctions will improve ESD. Lower-doped junctions (LDDs) will reduce ESD performance. Pocket (halo) implants can reduce ESD performance depending on the doping, implant energy, and implant angle, which determines if it will adversely affect M, or β.

Wells and substrates

p-well doping is directly related to the substrate spreading resistance and R_{sub}. Lower p-well doping means higher R_{sub} and is better for ESD. Similarly, for epitaxial layers and starting material type for p-type substrates. In general, non-epi material is better for ESD than epi material, and where epi is used, thicker epi gives better ESD performance. Retrograde implanted p-wells have been shown to generally improve ESD performance compared to the diffused p-wells.

n-well doping concentration and the doping profile are important to circuits that use diode protection or rely on the parasitic bipolar, lateral or vertical, of the pMOS

transistor. Higher n-well doping will result in lower β for the vertical *pnp* structures and can reduce ESD capability. Shallower wells will improve β but increase the n-well resistance which is not good for diode-type structures.

Gate oxides

In deep submicron technologies, very thin gate oxides are a concern for ESD. As long as the voltage across these oxides is limited to less than the oxide breakdown voltage by the turn-on of the protection circuit element gate oxide failure should not be dominant in these circuits. In protection circuits using snapback-type circuit elements, the essential trade-off is between the oxide breakdown voltage and the LNPN turn-on capability.

Silicides and contacts

Silicides have a significant impact on ESD performance through reduction in the drain and source spreading resistances, as well as degrading the parasitic bipolar action. Thinner silicide layers are better than thicker layers, and $CoSi_2$ is better than $TiSi_2$. Deeper junctions have better performance for the same silicide thickness. Optimization of the silicide for ESD performance is possible while maintaining transistor performance. Alternatively, an extra mask level can be used to block the silicide from the protection circuit regions.

The use of tungsten plugs in submicron technologies has almost eliminated problems related to contact melting and spiking.

Interconnect

At high currents, self-heating makes the resistance of the metal interconnects nonlinear. This needs to be included in analysis of metal interconnect behavior and design rules for protection circuit design. At high current, metal interconnect will eventually melt and vaporize. Copper has lower resistance than aluminum and for the same geometry will have better performance by nearly 50%. In cases where metal has melted and reformed without a resulting open circuit, the electromigration performance will still be strongly degraded.

The thermal properties and geometries of the interlevel dielectric needs to be given serious consideration as well. Low-K material with higher thermal resistivity can negate the beneficial properties of the copper metalization.

Gate length

The gate length of the MOS device in protection circuits and output transistors has a big influence on ESD capability. Typically longer gate lengths result in lower ESD performance due to the lower β of the parasitic bipolar transistor. However, at the sub-0.25 µm technology node, some reports have shown increasing ESD

capability with longer gate lengths, indicating that the β itself is no longer the dominant parameter and the well and the drain-source engineering need to be considered in the design optimization.

SOI

SOI devices have high thermal resistance that results in lower ESD performance than equivalent bulk circuits. However, the floating body effect (substrate is not grounded) helps uniform turn on of the parasitic bipolar and can provide ESD performance comparable to that of bulk material.

Bipolar transistors

Bipolar transistors have a vertical parasitic bipolar action that is much more uniform and robust than the lateral action in MOS devices. However, advanced bipolar transistor structures can have lateral paths that are more likely to be triggered and result in low failure thresholds. These can be suppressed using implant options that will also be beneficial to transistor performance.

Diodes

The forward resistance in nonavalanching diode structures is a function of the well and/or substrate doping. *np* diodes in epitaxial substrates have better on-resistance and current spreading at high current levels than bulk substrates. For the *pn* diodes in n-wells, the on-resistance is a function of the anode-cathode spacing, the n-well doping concentration as well as the factors that affect the gain of the vertical *pnp* transistor.

Resistors

n-well resistors have a nonlinearity that is a strong function of the n-well doping. In addition the upper operating voltage level is determined by the snapback of the resistor, which is dependent on the doping concentration and the head-head spacing of the resistor. The doping concentration also determines the saturation current, which is important in the application of these resistors in ESD protection circuits.

Polysilicon resistors depend on the thickness of the polysilicon and proximity to silicon substrate to enhance their high current capability, by improving the heat sinking.

Diffusion resistors are strongly affected by process through the junction breakdown voltage. They do have better heat sinking than poly resistors, and when implemented as buried resistor structures can provide robust precision resistors for high performance output drivers.

Reliability trade-offs

Process optimization needs to consider effects on other reliability issues. In general, process changes to improve ESD will result in degraded hot carrier performance, possibly lower gate oxide reliability and an increased sensitivity to latchup. In addition, the electromigration performance can be reduced after an ESD event, even if no obvious metal damage has been incurred.

BIBLIOGRAPHY

[Amerasekera86] A. Amerasekera and D. Campbell, "ESD Pulse and Continuous Voltage Breakdown in MOS Capacitor Structures", in *Proc. 8th EOS/ESD Symposium*, pp. 208–213, 1986.

[Amerasekera90] A. Amerasekera, L.J. van Roozendaal, J. Abderhalden, J.J.P. Bruines and L. Sevat, "An analysis of low voltage ESD damage in advanced CMOS processes", in *Proc. 12th EOS/ESD Symposium*, pp. 143–150, 1990.

[Amerasekera91] A. Amerasekera, L. van Roozendaal, J. Bruines and F. Kuper, "Characterization and modeling of second breakdown in NMOST's for the extraction of ESD-related process and design parameters", *IEEE T. Electron Dev.*, **ED-38**, 2161–2168, 1991.

[Amerasekera94A] A. Amerasekera and R. Chapman, "Technology design for high current and ESD robustness in a deep submicron process", *Electron Devic. Lett.*, **15**, 1994.

[Amerasekera94B] A. Amerasekera and C. Duvvury, "The Impact of Technology Scaling on ESD Robustness and Protection Circuit Design", in *Proc. 16th EOS/ESD Symposium*, 1994.

[Amerasekera94C] A. Amerasekera and J.A. Seitchik, "Electrothermal behavior of nMOS transistors under high current snapback (ESD/EOS) conditions", *Tech. Dig. IEDM*, 455–458, 1994.

[Amerasekera92A] E.A. Amerasekera, W. van den Abeelen, L.J. van Roozendaal, M. Hannemann and P.J. Schofield, "ESD failure modes: characteristics, mechanisms and process influences", *IEEE T. Electron Dev.*, **ED-39**, 2, 1992.

[Amerasekera92B] A. Amerasekera and A. Chatterjee, "An Investigation of BiCMOS ESD Protection Circuit Elements and Applications in Submicron Technologies", in *Proc. 14th EOS/ESD Symposium*, pp. 265–276, 1992.

[Amerasekera95] A. Amerasekera, C. Duvvury, V. Reddy and M. Rodder, "Substrate triggering and silicide effects on ESD performance and protection circuit design in deep submicron CMOS processes", *Tech. Dig. IEDM*, 547–550, 1995.

[Amerasekera96C] A. Amerasekera, V. McNeil and M. Rodder, "Correlating Drain Junction Scaling, Silicide Thickness, and lateral *npn* Behavior with the ESD/EOS Performance of a 0.25 μm CMOS Process", *Tech. Dig. IEDM*, 893–896, 1996.

[Amerasekera99] A. Amerasekera, V. Gupta, K. Vasanth and S. Ramaswamy, "Analysis of Snapback Behavior on the ESD Capability of Sub-0.20 μm nMOS", in *Proc. 37th IRPS*, 159–166, 1999.

[Amerasekera00] A. Amerasekera and V. Gupta, "Process-Related Effects on ESD Performance in Deep Submicron Technologies", in *Proc. SEMICON Korea*, 1–9, 2000.

[Aur86] S. Aur, A. Chatterjee and T. Polgreen, "Hot electron reliability and ESD latent damage", *IEEE T. Electron Dev.*, **ED-35**, 2189–2193, 1986.

[Banerjee96A] K. Banerjee, A. Amerasekera, G. Dixit and C. Hu, "Characterization of VLSI Circuit Interconnect Heating And Failure Under ESD Conditions", in *Proc. 34th IRPS*, 237–245, 1996.

[Banerjee96B] K. Banerjee, A. Amerasekera, G. Dixit and C. Hu, "The effect of interconnect scaling and low-k dielectric on the thermal characteristics of the IC metal" *Tech. Dig. IEDM*, 65–68, 1996.

[Banerjee97A] K. Banerjee, A. Amerasekera, N. Cheung and C. Hu, "High current failure model for VLSI interconnects under short-pulse conditions", *IEEE Electron Devic. Lett.*, 405–407, 1997.

[Banerjee97B] K. Banerjee, A. Amerasekera, G. Dixit, N. Cheung and C. Hu, "Characterization of contact and via failure under short duration high current stress", in *Proc. 35th IRPS*, 216–220, 1997.

[Banerjee98] K. Banerjee, A. Amerasekera, J. Kittl and C. Hu, "High current effects in silicide films for sub-0.25 μm VLSI technologies", in *Proc. 36th IRPS*, 284–292, 1998.

[Bock97] K. Bock, C. Russ, G. Badenes, G. Groeseneken and L. Deferm, "Influence of Well Profile and Gate Length on the ESD Performance of a Fully Silicided 0.25 μm CMOS Technology", in *Proc. 19th EOS/ESD Symposium*, pp. 308–315, 1997.

[Bock99] K. Bock, B. Keppens, V. De Heyn, G. Groeseneken, L.Y. Ching and A. Naem, "Influence of Gate Length on ESD Performance For Deep Submicron CMOS Technology", in *Proc. 21st EOS/ESD Symposium*, pp. 95–104, 1999.

[Bridgewood85] M. Bridgewood and R. Kelly, "Modeling the Effects of Narrow Impulsive Overstress on Capacitive Test Structures", in *Proc. 7th EOS/ESD Symposium*, pp. 84–91, 1985.

[Chaine92] M. Chaine, S. Desai, C. Dunn, D. Dolby, W. Holland and T. Pekny, "ESD Improvement Using Low Concentrations of Arsenic Implantation in CMOS Output Buffers", in *Proc. 14th EOS/ESD Symposium*, pp. 136–142, 1992.

[Chan94] M. Chan, S.S. Yuen, Z.-J. Ma, K.Y. Hui, P.K. Ko and C. Hu, "Comparison of ESD protection capability of SOI and bulk CMOS output buffers", in *Proc. 32nd IRPS*, 292–298, 1994.

[Chang88] P.-H. Chang, R. Hawkins, T. Bonifield and L. Melton, "Aluminum spiking at contact windows in Al/Ti-W/Si", *Appl. Phys Lett.*, **52**, 272–274, 1988.

[Chen86] K.L. Chen, G. Giles and D. Scott, "Electrostatic discharge protection for 1-μm CMOS devices and circuits", *Tech. Dig. IEDM*, 484–487, 1986.

[Chen88] K-L. Chen, "Effect of interconnect process and snapback voltage on the ESD failure threshold of nMOS transistors", in *Proc. 10th EOS/ESD Symposium*, pp. 212–219, 1988.

[Chen96] J.Z. Chen, X.Y. Zhang, A. Amerasekera and T. Vrotsos, Design and layout of a high ESD performance NPN structure for submicron BiCMOS/bipolar circuits, in *Proc. 34th IRPS*, 227–232, 1996.

[Chen98] J.Z. Chen, A. Amerasekera and C. Duvvury, "Design methodology and optimization of gate driven nMOS ESD protection circuits in submicron CMOS processes", *IEEE T. Electron Dev.*, **ED-45**, 2448–2456, 1998.

[Consiglio95] R. Consiglio and T.Y. Huang, "ESD Reliability Impact of p^+ Pocket Implant on Double Implanted NLDD MOSFETs", in *Proc. 17th EOS/ESD Symposium*, pp. 199–206, 1995.

[Dabral93] S. Dabral and R. Aslett, "Designing On-Chip Power Supply Coupling Diodes for ESD Protection and Noise Immunity", in *Proc. 15th EOS/ESD Symposium*, pp. 239–250, 1993.

[Dabral98] S. Dabral and T. Maloney, *Basic ESD and I/O Design*, John Wiley & Sons New York, 1998.

[Daniel90] S. Daniel and G. Krieger, "Process and Design Optimization for Advanced CMOS I/O ESD Protection Devices", in *Proc. 12th EOS/ESD Symp.*, pp. 206–213, 1990.

[DeChiaro81] L.F. De Chiaro, "Electrothermomigration in nMOS LSI devices", in *Proc. 19th IRPS*, 212–217, 1981.

[Duvvury85] C. Duvvury, R. Rountree, D. Baglee, A. Hyslop and L. White, "ESD Design Considerations for ULSI", in *Proc. 7th EOS/ESD Symposium*, pp. 45–48, 1985.

[Duvvury86] C. Duvvury, R. McPhee, D. Baglee and R. Rountree, "ESD protection reliability in 1-μm CMOS circuit performance", in *Proc 24th IRPS*, 199–208, 1986.

[Duvvury88] C. Duvvury, R. Rountree and O. Adams, "Internal chip ESD phenomena beyond the protection circuit", in *Proc. 26th IRPS*, 19–25, 1988.

[Duvvury89] C. Duvvury, R. Rountree, H. Stiegler, T. Polgreen and D. Corum, "ESD phenomenon in graded junction devices", in *Proc 27th IRPS*, 71–76, 1989.

[Duvvury96A] C. Duvvury and A. Amerasekera, "Advanced CMOS protection device trigger mechanisms during CDM", *IEEE T. Compon. Pack. T.*, Part C, **CPMT-19**, 169–177, 1996.

[Duvvury96B] C. Duvvury, A. Amerasekera, K. Joyner, S. Ramaswamy and S. Young, "ESD Design For Deep Submicron SOI Technology", in *Proc. VLSI Symposium on Tech.*, pp. 194–195, 1996.

[Duvvury00] C. Duvvury, S. Ramaswamy, A. Amerasekera, R. Cline, B. Andresen and V. Gupta, "Substrate Pump nMOS for ESD Protection Applications", in *Proc. 22nd EOS/ESD Symposium*, pp. 7–17, 2000.

[Eitan82] B. Eitan and D. Frohman-Bentchkowsky, "Surface conduction in short-channel MOS devices as a limitation to VLSI scaling", *IEEE T. Electron Dev.*, **ED-29**, 254–266, 1982.

[Fong87] Y.-P. Fong and C. Hu, "The Effects of High Electric Field Transients on Thin Gate Oxide MOSFETs", in *Proc. 9th EOS/ESD Symposium*, pp. 252–257, 1987.

[Fukuda86] Y. Fukuda, S. Ishiguro and M. Takahara, "ESD protection network evaluation by HBM and CPM (Charged Package Method)", in *Proceedings of the 8th EOS/ESD Symposium*, pp. 193–199, 1986.

[Groeseneken00] G. Groeseneken, "Hot Carrier Degradation and ESD in Submicron CMOS Technologies: How Do They Interact?", in *Proc. 22nd EOS/ESD Symposium*, pp. 276–286, 2000.

[Gupta98] V. Gupta, A. Amerasekera, S. Ramaswamy and A. Tsao, "ESD-related Process Effects in Mixed Voltage Sub-0.5 μm Technologies", in *Proc. 20th EOS/ESD Symposium*, pp. 161–169, 1998.

[Harame95] D. Harame, J. Comfort, J. Cressler, E. Crabbe, J.Y-C. Sun and B. Meyerson, "Si/SiGe epitaxial base transistors — Part 1: materials, physics and circuits", *IEEE T. Electron Dev.*, **ED-42**, 455–468, 1995.

[Hsu82] F.C. Hsu, P.K. Ko, S. Tam, C. Hu and R. Muller, "An analytical breakdown model for short-channel MOSFETs", *IEEE T. Electron Dev.*, **ED-29**, 1735–1740, 1982.

[Ishizuka94] H. Ishizuka, K. Okuyama and K. Kubota, "Photon emission study of ESD protection devices under second breakdown conditions", in *Proc. 32nd IRPS*, 286–291, 1994.

[Jankovic91] N.D. Jankovic, "Pre-turn-on surce bipolar injection in graded NMOST's", *IEEE T. Electron Dev.*, **ED-38**, 2527–2530, 1991.

[Koizumi00] H. Koizumi, Y. Komine, T. Ohtomo and M. Shimaya, "ESD Protection in Fully-Depleted CMOS/SIMOX With a Tungsten Clad Source/Drain", in *Proc. 22nd EOS/ESD Symposium*, pp. 260–265, 2000.

[Konig98] U. Konig, "SiGe and GaAs as competitive technologies for RF-applications", *Tech. Dig. BCTM*, 87–92, 1998.

[Lau82] C.K. Lau, Y.C. See, D. Scott, J. Bridges, S. Perna and R.D. Davies, "Titanium disilicide self-aligned source/drain and gate CMOS technology", *Tech. Dig. IEDM*, 714–717, 1982.

[Laux87] S.E. Laux and F.H. Gaensslen, "A study of channel avalanche breakdown in scaled n-MOSFETs", *IEEE T. Electron Dev.*, **ED-34**, 1066–1073, 1987.

[Maloney86] T. Maloney, "Contact Injection: A Major Cause of ESD Failure in Integrated Circuits", in *Proc. 8th EOS/ESD Symposium*, pp. 166–172, 1986.

[Maloney99] T. Maloney and W. Kan, "Stacked pMOS Clamps For High Voltage Power Supply Protection", in *Proc. 21st EOS/ESD Symposium*, pp. 70–77, 1999.

[McPhee86] R. McPhee, C. Duvvury and R.N. Rountree, "Thick Oxide Device ESD Performance Under Process Variations", in *Proc. 8th EOS/ESD Symposium*, pp. 173–181, 1986.

[Miller00] J. Miller, M. Khazhinsky and J. Weldon, "Engineering the Cascoded nMOS Output Buffer For Maximum V_{t1}", in *Proc. 22nd EOS/ESD Symposium*, pp. 308–317, 2000.

[Noterman99] G. Notermans, A. Heringa, M. van Dort, S. Jansen and F. Kuper, "The effect of silicide on ESD performance", in *Proc. 37th IRPS*, 154–158, 1999.

[Ohtani90] S. Ohtani and M. Yoshida, "Model of Leakage Current in LDD Output MOSFET Due to Low-Level ESD Stress", in *Proc. 12th EOS/ESD Symposium*, pp. 177–181, 1990.

[Pierce85] D.G. Pierce, "Electro-Thermomigration as an Electrical Overstress Failure Mechanism", in *Proc. 5th EOS/ESD Symposium*, pp. 67–76, 1987.

[Pinto-Guedes88] M. Pinto-Guedes and P. Chan, "A circuit simulation model for bipolar-induced breakdown in MOSFET", *IEEE Trans. Comp. Aid. Des.*, **CAD-7**, 289–294, 1988.

[Rafferty93] C.S. Rafferty, H.-H. Vuong, S.A. Eshraghi, M. Giles, M. Pinto and S. Hillenius, "Explanation of reverse short-channel effect by defect gradients", *Tech. Dig. IEDM*, 311–314, 1993.

[Rountree88] R. Rountree, "ESD protection for submicron CMOS circuits issues and solutions", *Tech. Dig. IEDM*, 580–583, 1988.

[Sanchez99] H. Sanchez, J. Siegel, C. Nicoletta, J. Alvarez and J. Nissen, "A versatile 3.3 V/2.5 V/1.8 V CMOS I/O driver built in a 0.2 μm 3.5 nm Tox 1.8 V technology", *Tech. Dig. ISSCC*, 276–277, 1999.

[Schutz82] A. Schutz, S. Selberherr and H. Potzl, "Analysis of breakdown in MOSFETs", *IEEE T. Comput. Aid. D.*, **CAD-1**, 77–84, 1982.

[Scott86] D. Scott, J. Hall and G. Giles, "A Lumped Element Model for Simulation of ESD Failures in Silicided Devices", in *Proc. 8th EOS/ESD Symposium*, pp. 41–47, 1986.

[Scott87] D. Scott, R. Chapman, C.-C. Wei, S.S. Mahant-Shetti, R. Haken and T. Holloway, "Titanium disilicide contact resistivity and its impact on 1-μm CMOS circuit performance", *IEEE T. Electron Dev.*, **ED-34**, 562–574, 1987.

[Shabde84] S.N. Shabde, G. Simmons, A. Baluni and D. Back, "Snapback induced gate dielectric breakdown in graded junction MOS structures", *Proc. 22nd IRPS*, 165–168, 1984.

[Smith99] J. Smith, "Anti-Snapback Circuit Techniques For Inhibiting Parasitic Bipolar Conduction", in *Proc. 21st EOS/ESD Symposium*, pp. 62–69, 1999.

[Strauss87] M. Strauss, D. Lin and T. Welsher, "Variations in Failure Modes and Cumulative Effects Produced by Commercial Human Body Model ESD Simulators", in *Proc. 9th EOS/ESD Symposium*, pp. 59–63, 1987.

[Toyabe78] T. Toyabe, K. Yamaguchi, S. Asai and M. Mock, "A two-dimensional avalanche breakdown model of submicron MOSFETs", *IEEE T. Electron Dev.*, **ED-25**, 825–833, 1978.

[Verhaege93A] K. Verhaege, G. Groeseneken, J.-P. Collinge and H.E. Maes, "Double snapback in SOI nMOSFETs and its application for SOI ESD protection", *Elec. Dev. Lett.*, **EDL-14**, 326–328, 1993.

[Verhaege93B] K. Verhaege, G. Groeseneken, J.-P. Collinge and H.E. Maes, "The ESD Protection Capability of SOI Snapback nMOS FETs: Mechanisms And Failure Modes", in *Proc. 15th EOS/ESD Symposium*, pp. 215–219, 1993.

[Voldman92] S. Voldman, V. Gross, M.J. Hargrove, J.M. Never, J.A. Slinkman, M.P. O'Boyle, T.S. Scott and J.J. Delecki, "Shallow Trench Isolation Double-Diode Electrostatic Discharge Circuit and Interaction with DRAM Output Circuitry", in *Proc. 14th EOS/ESD Symposium*, pp. 277–288, 1992.

[Voldman93] S. Voldman and V. Gross, "Scaling Optimization and Design Considerations of Electrostatic Discharge Protection Circuits in CMOS Technology", in *Proc. 15th EOS/ESD Symposium*, pp. 251–260, 1993.

[Voldman94] S. Voldman *et al.*, "ESD Protection in a Multi-Rail Disconnected Power Grid and Mixed Voltage Interface Environments in 0.5 and 0.25 μm Channel Length CMOS Technologies", in *Proc. 16th EOS/ESD Symposium*, 1993.

[Voldman98] S. Voldman, F. Assaderaghi, J. Mandelman, L. Hsu and G. Shahidi, "Dynamic threshold body and gate-coupled SOI ESD protection networks", in *J. Electrostat.*, 239–255, 1998.

[Voldman00A] S. Voldman, D. Hui, D. Young, R. Williams, D. Dreps, J. Howard, M. Sherony, F. Assaderaghi and G. Shahidi, "Silicon-On-Insulator Dynamic Threshold ESD Networks and Active Clamp Circuitry", in *Proc. 22nd EOS/ESD Symposium*, pp. 29–40, 2000.

[Voldman00B] S. Voldman, P. Juliano, N. Schmidt, R. Johnson, L. Lanzerotti, A. Joseph, C. Brennan, J. Dunn, D. Harame, E. Rosenbaum and B. Meyerson, "Electrostatic Discharge Characterization of Epitaxial-Base Silicon-Germanium Heterojunction Bipolar Transistors", in *Proc. 22nd EOS/ESD Symposium*, pp. 239–250, 2000.

[Wei92] Y. Wei, Y. Loh and C. Wang, "MOSFET Drain Engineering for ESD Performance", in *Proc. 14th EOS/ESD Symposium*, pp. 143–148, 1992.

[Wilson87] D. Wilson, H. Domingos and M.M.S. Hassan, "Electrical Overstress in nMOS Silicided Devices", in *Proc. 9th EOS/ESD Symposium*, pp. 265–273, 1987.

[Worley00] E. Worley, A. Salem and Y. Sittampalam, "High Current Characteristics Of Devices In A 0.18 μm CMOS Process", in *Proc. 22nd EOS/ESD Symposium*, pp. 296–307, 2000.

[Wu00] E.Y. Wu, J. Aitken, E. Nowak, A. Vayshenker, P. Varekamp, G. Hueckel, J. McKenna, D. Harmon, L.-K. Han, C. Montrose and R. Dufresne, "Voltage-dependent voltage-acceleration of oxide breakdown for ultra-thin oxides", *Tech. Dig. IEDM*, 541–544, 2000.

10 Device Modeling of High Current Effects

Ajith Amerasekera

10.1 INTRODUCTION

The development of protection circuits for new technologies, or the application of existing protection circuits to new applications requires a number of design iterations and testing. Experienced ESD protection circuit designers can require three design iterations before the specified ESD level can be guaranteed. Hence, the cycle times for the characterization of the ESD robustness of a technology can exceed the total available time for the introduction of a new technology. Furthermore, the need to experimentally optimize the ESD performance of a protection circuit in a specific chip increases the product development cycle time. Over the years, IC manufacturers have begun to integrate ESD characterization into the early development phase of new technologies [Daniel90][Krakauer92][Amerasekera94A]. At the same time, IC designers have considered the ESD protection circuit requirement in parallel with input and output circuit design and area allocations [Duvvury88][Polgreen89].

The purpose of modeling the behavior of devices and circuits under ESD conditions is to reduce the iterations and cycle times involved in achieving successful ESD performance. Even simple modeling allows technology and circuit designers to rapidly evaluate the ESD robustness of their designs and to reduce the number of iterations required to obtain good protection circuits [Chatterjee91]. Modeling also enables ESD robustness to be designed into a new technology [Amerasekera93A], thereby ensuring that circuit designers have a solid foundation upon which to develop good ESD protection circuits. By rapidly exploring the technology design space, simulation techniques can enable the technology designer to optimize for ESD robustness as well as device performance.

In Chapter 4, the physics of devices under high current and high voltage conditions was reviewed. In this chapter, we will discuss modeling the circuit elements under ESD conditions with thermal effects, and the approaches used. ESD failure is the result of thermal damage in the semiconductor and most ESD modeling techniques are based on an analysis of the heat dissipated

in the semiconductor element during a high current stress event [Krieger87] [Pierce88][Lin90][Dwyer90][Amerasekera91][Diaz93A][Diaz93B]. More detailed models have coupled the electrical behavior of the elements with the thermal behavior in electrothermal models. A simple approach is to use lumped circuit elements to describe the thermal resistances and capacitances and implement the full model in a circuit simulator [Scott86][Beltman90][Kurimoto94]. A more complex technique is to couple the full heat equation with the device equations in a complete electrothermal model of the element. Such models have been used for some time in power bipolar analysis and thyristor development [Gaur76][Adler78]. The more recent advances in computing have enabled these models to be applied to ESD type high currents in semiconductor circuit elements [Krabbenborg91][Mayaram91][Amerasekera93B].

The simulation of circuit behavior during ESD events is discussed in detail in Chapter 11. The simplest approach to date is the use of circuit simulation tools such as SPICE with the maximum current and voltage of the individual circuit elements used as the boundary conditions. Advances in mixed-mode simulators, which couple device modeling with circuit modeling have shown that it is possible to estimate the operating thresholds for ESD protection circuits [Duvvury92A][Diaz93B].

10.2 THE PHYSICS OF ESD DAMAGE

The high current behavior of semiconductor elements used in ESD protection circuits has been presented in Chapter 4. The circuits, which are most commonly used are the lateral *npn* in MOS technologies (Figure 10.1(a)), and the vertical *npn* in bipolar and BiCMOS technologies (Figure 10.1(b)). Both these elements operate in a two-terminal mode during an ESD stress, relying on the avalanche generation of holes to provide the base current required to turn them on. Once on, the *npn* conducts current between the collector and the emitter with an on-resistance of a few ohms and a holding voltage of between 5 V and 10 V, depending on the technology. Most of the holding voltage is dropped across the collector-base junction and is required to maintain the avalanche generation of holes, which provides the *extrinsic* base current to maintain the *npn* in the *on* state. Hence, almost all the power in the high current operating mode can be assumed to be generated in the collector-base junction. It follows that a lower collector-base voltage will result in a lower power dissipation in the device and will result in a higher current capability.

Note: The extrinsic base current is the current that flows in the extrinsic base resistance R_{sub} as opposed to the intrinsic base current, which flows across the forward-biased emitter base junction. This difference is important in understanding the self-biasing mechanism of lateral *npn* transistor action.

When a device is subjected to high current injection, the power dissipated in the device results in an increase in the internal temperature. At some point, given that there is no electrothermal effect on the device parameters, the internal temperature will approach that of the melting point of the semiconductor material (1685 K for

(a)

(b)

Figure 10.1 (a) Cross section of an nMOS transistor in a CMOS technology showing the lateral *npn* transistor and the current flow. (b) Cross section of a vertical *npn* transistor in a bipolar technology showing the direction of current flow

silicon). A change of phase will then take place and the device properties will be irreversibly altered. However, it is well known that thermal breakdown occurs at temperatures below the melting point [Tauc57], [Scarlett63][Melchior64]. This breakdown has been termed *second breakdown*, distinguishing it from the avalanche breakdown, which occurs at lower injection currents. Second breakdown results in a collapse in the voltage across the device, and is significant in that the electrical properties of the device such as off-state current and the $I-V$ curves are observed to be severely changed after second breakdown. Therefore, the onset of second breakdown is the damage threshold of the device. Since second breakdown is thermally initiated, models for second breakdown focus on the temperature rise in the device due to the power dissipated. The high current $I-V$ curve of an nMOS transistor with current injection into the drain and the source, gate, and substrate at zero volts is shown in Figure 10.2. The triggering of the bipolar transistor occurs at drain voltage of V_{t1}, and injection current of I_{t1}. The drain voltage then reduces to the bipolar holding voltage V_{sp}, also known as the snapback voltage. The device is now in a low impedance condition and can handle much higher injection current, until junction heating results in second breakdown at voltage of V_{t2} and current of I_{t2}. I_{t2} is a direct indicator of the ESD capability of the device and is used as the figure-of-merit for ESD robustness of a component used in ESD protection circuits.

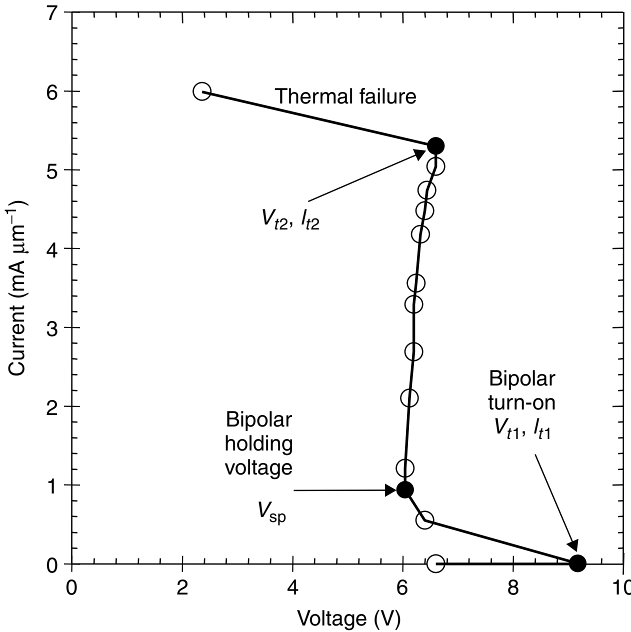


Figure 10.2 High current $I-V$ curve of an nMOS device showing the bipolar triggering and thermal failure points

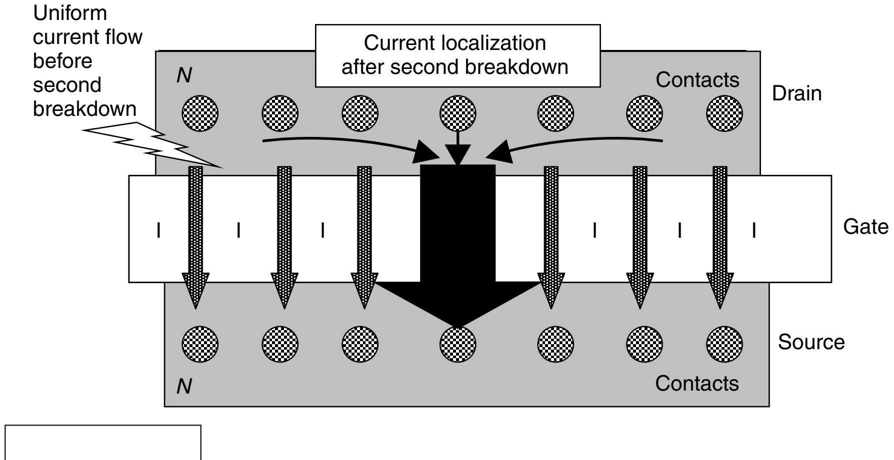


Figure 10.3 Top view of an nMOS transistor schematically depicting current localization and filamentation at the point of second breakdown. Initially current flows uniformly in the silicon between drain and source. After current localization takes place, all the current flows through a smaller region of the silicon

Once the second breakdown threshold is reached, the device goes into a negative differential resistance (NDR) mode. In the NDR mode, the voltage and current are unstable [Ridley63][Shaw92], and the voltage will decrease until a stable bias condition is reached. It has been shown [Ridley63][Khurana66][Shaw92] that the stable bias condition is in fact a microplasma formed by the constriction of the current into a filament. Figure 10.3 shows the formation of the filament as the current goes being uniformly distributed from contact to contact across the drain-source, to being localized. Now current from many drain contacts will be channeled through a small region of the drain-source. When this occurs, the current density in the filament is very high, and the melt temperature of silicon is reached rapidly in the localized region. The formation of a molten filament following second breakdown has been used to explain the failure modes observed in devices, which undergo second breakdown [Khurana66][Nienhuis66][Brown72][Smith73][Amerasekera92] including junction edge damage and contact burn-out. By understanding the mechanisms of the thermal breakdown phenomena, it is possible to model the ESD behavior of these structures as well as to design more robust ESD structures.

10.3 THERMAL (“SECOND”) BREAKDOWN

The conditions for thermal or *second* breakdown are defined by the internal temperature of the device at which the voltage begins to collapse. Accurate measurements of the temperature rise in the silicon during the ESD event are still not possible, and makes it difficult to verify the various theories regarding the

precise mechanisms governing the onset of second breakdown. The advent of electrothermal device simulators has enabled more insight to be gained into the device behavior at high temperatures and has resulted in more understanding of the effects leading to second breakdown.

Tauc and Abraham [Tauc57] first suggested that thermal generation of carriers could be responsible for permanent damage seen in germanium *pn* junctions operated in the breakdown region. They also observed that the decrease in voltage caused by thermal breakdown was accompanied by a current constriction. Later Schafft and French [Schafft62][Schafft66] showed that the high current breakdown could be related to thermal effects based on their experiments on the relationship between the energy required for breakdown and the time to breakdown. The first criterion for thermal breakdown can be attributed to Melchior and Strutt [Melchior64] who concluded on the basis of experiments on *npn* transistors that the breakdown was triggered when the collector-base junction reached its intrinsic temperature, T_i. T_i was defined as the temperature at which the intrinsic carrier concentration, n_i, was equal to the background doping concentration given by N_d for *n*-type material. Later experimental work [Chen71] and simulations [Ward76], [Orvis83] were not able to confirm, or refute, that thermal breakdown was definitely initiated at T_i. In the absence of a suitable alternative, T_i has been used as a fitting parameter to obtain approximate solutions for the power required for thermal breakdown [Popescu70][Smith73][Ghandi78][Alexander78].

In order to understand the concept of T_i and its relationship to thermal breakdown in more depth, we should consider the variation of the silicon resistivity with temperature for different doping concentrations as shown in Figure 10.4 [Runyan65].

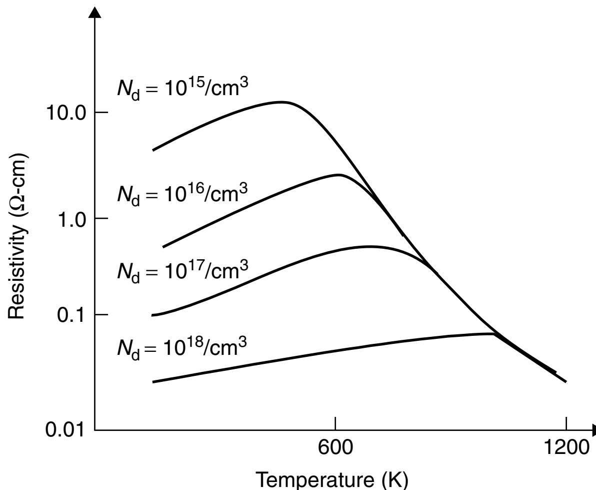


Figure 10.4 Resistivity as a function of temperature for different doping concentrations, showing the positive and negative temperature coefficients as the temperature is increased. Lower doped silicon has a lower critical temperature before the coefficient becomes negative

As the temperature increases, the resistivity first increases as the mobility decreases. At a critical temperature given by T_c, the resistivity reaches a maxima and then decreases with increasing temperature. The maxima occurs when the effect of the increase in n_i on the resistivity begins to dominate over the decrease in mobility. The conductance for an n-type material with a doping concentration N_d is given by,

$$\sigma = N_d \times q \times \mu \tag{10.1}$$

and n_i is given by

$$n_i = A \cdot T^{\frac{3}{2}} \cdot \exp\left(-\frac{E_g}{2kT}\right) \tag{10.2}$$

The temperature dependence of mobility is theoretically determined from phonon scattering, which gives [Bardeen50],

$$\mu \sim T^{-\frac{3}{2}} \tag{10.3}$$

which has been experimentally shown to hold for higher impurity concentrations $N_d \approx 2 \times 10^{17}/\text{cm}^3$ [Jacoboni77][Sze81]. At lower doping concentrations other scattering mechanisms result in an increased temperature dependence, which has been determined empirically for n-type silicon [Jacoboni77],

$$\mu \sim T^{-2.4} \tag{10.4}$$

The transition from a positive temperature coefficient of resistance to a negative temperature coefficient occurs when $n_i = N_d$. For higher temperatures, $n_i > N_d$, and N_d in Equation 10.1 is replaced by the temperature-dependent n_i. Hence, for a purely resistive material with no other carrier generation or injection mechanisms $T_c = T_i$ can be taken to be the temperature at which the $I-V$ curve shows NDR characteristics.

In reverse-biased semiconductor junctions, there are additional mechanisms, which add complexity to the analysis described above. It has been argued [Smith73][Alexander78] that since the potential barrier at the junction is a function of n_i, as n_i, increases the potential barrier decreases. When $n_i = n = p$, the potential barrier is reduced to zero and the device is in thermal breakdown. The basis for this argument is as follows. The quasi-fermi levels are given by ϕ_n and ϕ_p for electrons and holes respectively. For a pn junction, the pn product in the depletion layer is

$$pn = n_i^2 \exp\left(\frac{q(\phi_p - \phi_n)}{kT}\right) \tag{10.5}$$

The potential barrier is given by $(\phi_p - \phi_n)$, which is less than zero for a reverse-biased junction. When the potential barrier vanishes, then $\phi_p = \phi_n$, and $pn = n_i^2$, giving $p = n = n_i$. Again the assumption is that all generation is thermal, and that there is no avalanching taking place, which makes this analysis only applicable at low reverse-biased voltages well below that required to cause

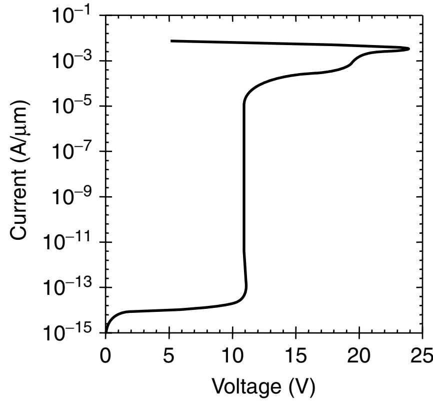


Figure 10.5 High current $I-V$ curve for a *pn* diode in reverse bias

avalanche breakdown. At high current levels, the junction is in avalanche breakdown, and simulations have shown that the avalanche generated carriers result in n_i^2 actually being much greater than pn at the second breakdown threshold [Ward77][Yee82][Orvis83][Mayaram91][Amerasekera93B]. Thus the use of $T = T_i$ as the condition for second breakdown can only be regarded as an approximation for ESD type second breakdown.

The conditions for second breakdown can be more generally defined by considering the terminal current and voltage across a device [Burgess60][Shaw92][Amerasekera93B]. The $I-V$ curve for a typical reverse-biased *pn* junction is shown in Figure 10.5. At the second breakdown point, the voltage reaches a maxima, and as the current increases further, the voltage decreases. In general, the current is given by

$$I = I(V, T) \tag{10.6}$$

and, therefore,

$$\frac{dI}{dV} = \left.\frac{\partial I}{\partial V}\right|_T + \left.\frac{\partial I}{\partial T}\right|_V \cdot \frac{dT}{dV} \tag{10.7}$$

Now the temperature can be determined from the heating in the device, and so $T = T(P)$, where $P = I \times V$. Hence,

$$\left.\frac{\partial T}{\partial V}\right|_I = I \cdot \frac{dT}{dP} \tag{10.8}$$

and

$$\left.\frac{\partial T}{\partial I}\right|_V = V \cdot \frac{dT}{dP} \tag{10.9}$$

An analysis of thermistor devices [Burgess60], showed that Equation 10.7 to Equation 10.9 could be rearranged to give

$$\frac{dV}{dI} = \frac{V}{I} \cdot \frac{1 - y}{x + y} \tag{10.10}$$

where

$$x = R \cdot \left. \frac{\partial I}{\partial V} \right|_T \tag{10.11}$$

and

$$y = V \cdot \frac{dT}{dP} \cdot \left. \frac{\partial I}{\partial T} \right|_V \tag{10.12}$$

At second breakdown, $dV/dI = 0$, and $y = 1$. The condition for second breakdown is then given by,

$$\left. \frac{\partial I}{\partial T} \right|_V \cdot \frac{dT}{dP} = 1 \tag{10.13}$$

A similar condition for second breakdown was presented and explained in terms of the electrothermal effects in the device [Amerasekera93B]. Essentially, second breakdown occurs when the heat produced by an increase in current, ΔI, generates the exact amount of minority carriers required to support the increase in current without raising the electric field. As the thermal heating increases, the heat generated by ΔI generates more carriers than needed to support the increase in I, and the voltage decreases leading to the familiar *snapback* phenomenon. It was shown that this condition was satisfied at second breakdown using electrothermal simulations of $p^+/n/n^+$ avalanche diodes.

However, the above condition is not easily applicable as a predictor, because of the nonlinearities of the equations governing avalanche and thermal generation currents. It is possible to make assumptions for the power dissipation and the conductivity [Shaw92], but those do not really solve the problem of predictive modeling.

A first-order model of the high current performance of an ESD protection circuit element can best be achieved by considering the maximum power that the element is capable of withstanding before damage occurs. This would enable the maximum current that can be passed through the element to be determined, which can subsequently be translated into an ESD failure threshold.

The boundary condition used in the development of the model is the damage threshold under high current injection. Recent works [Krieger87][Pierce88][Lin90] took this boundary condition to be the temperature at which the melting temperature of silicon, 1412 °C, is reached or, in the case of aluminum contacts, the eutectic temperature for Al-Si interdiffusion (550 °C). More recently still, it was shown that device failure was related to second breakdown at the stressed junction [Polgreen89][Amerasekera90]. However, since there is no real way of determining the actual device temperature at the onset of failure (although a failed device has obviously seen >1412 °C), it is necessary to extract the dependence of the

failure power, P_f, as a function of stress time for failure, t_f, using experimental data [Pierce88][Dwyer90][Amerasekera91]. While this approach enables the ESD sensitivity of some design and process related parameters to be determined, it is not directly suitable for predictive modeling of ESD capability of a technology of device structure.

10.4 ANALYTICAL MODELS USING THE HEAT EQUATION

An analytical approach to ESD modeling starts with the solution of the heat conduction equation for the device geometry under consideration.

$$\frac{\partial T}{\partial t} - D \cdot \nabla^2(T) = \frac{q(t)}{\rho \cdot C_p} \tag{10.14}$$

ρ is the density of the semiconductor in g cm^{-3}, C_p is the specific heat in J g^{-1}-K, $D = K/(\rho C_p)$ is the thermal diffusivity in cm^2/s, and K is the thermal conductivity in W cm^{-1}-K. $q(t)$ is the rate of heating per unit volume of the heat source.

The standard solution to the heat equation for a variety of sources and boundary conditions has been well documented [Carslaw59]. In modeling electrical overstress and ESD, a simple thermal model for the power as a function of the time is given by [Wunsch68]

$$\frac{P}{A} = \sqrt{\pi K \rho C_p} \cdot (T - T_0) \cdot t^{\frac{1}{2}} \tag{10.15}$$

where P is the input power and A is the area of the heat source, which is effectively the junction area. T is the temperature in the device and T_0 is the initial temperature. The great advantage of the equation is its simplicity and ease of use. The power dependence of time easily translates to a straight line on a log-log plot, and the coefficients can be lumped to give

$$P = B \cdot t^{\frac{1}{2}} \tag{10.16}$$

where

$$B = A\sqrt{\pi K \rho C_p}(T - T_0) \tag{10.17}$$

is obtained from a set of experimental results and then used to determine the effect of a given transient stress on the failure threshold. The simplicity of the equation has ensured its wide use in the analysis of electrical overstress and ESD failure data [Speakman74][Mathews80][Pierce88][Diaz92].

A more detailed analysis of the heat equation leads to a solution of the form described by [Tasca70][Tasca72]

$$P = \left(\frac{A}{t} + B \cdot t^{\frac{1}{2}} + C\right)(T - T_0) \tag{10.18}$$

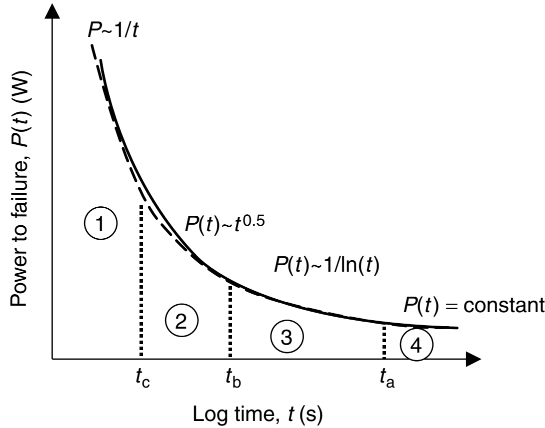


Figure 10.6 Dependence of power-to-failure P(t) on time for a given failure temperature. As t is decreased, the power required to reach that temperature increases. The analytical solutions to the heat equation show four regions in the curve as marked

The above equation shows that the time dependence of the power dissipation is determined by the time period of the applied pulse. For very short duration pulses, there is very little heat flow out of the defect region, and the time dependence follows an adiabatic or $1/t$ dependence. At longer stress times, the Wunsch-Bell equation is followed, and the third term in the parentheses simply describes the steady state condition. The analysis was further developed equation further to show that there are actually four time-dependent regions [Dwyer90]. In addition to the three described by Tasca, it was shown that in the period immediately before steady state was reached, the dependence has a $1/\ln(t)$ form. The four regions are shown in the curve in Figure 10.6.

The thermal diffusion length $L = \sqrt{Dt}$, is on the order of $10\,\mu\text{m}$ for a pulse duration of 1 μs. This is much smaller than the distance between the protection circuit and the edge of the chip or wafer. We can assume that the heat is being radiated into an infinite medium, since as either x or y tends to infinity, T tends to ambient. Similarly, in the z direction, for $z > 0$ (vertical depth) the silicon thickness is much greater than the diffusion length, and we may assume an infinite medium for $z > 0$. Therefore, the problem can be reduced to one of a heat source of a given geometry in a semi-infinite medium $-\infty < x < \infty$, $-\infty < y < \infty$ and $z > 0$. Usually, the region $z < 0$ is covered by oxide and may be assumed to be a poor heat conductor. The boundary $z = 0$ can, therefore, be considered to be reflective and by the method of images, the heat source is mirrored in the region $z < 0$ [Carslaw59][Krieger87]. In the direction $z < 0$ the structure may not be thermally insulating because of the presence of the polycrystalline silicon gate and associated metalization, which lies just above the junction close to the location of the heat source. This may have some effect on the analytical models described below, but the discrepancy is not expected to be significant.

The general solution for this problem can be found using the Green's function method [Carslaw59] whereby the solution to the equation

$$\frac{\partial G}{\partial t} - D \cdot \nabla^2(G) = \delta(r - r') \cdot \delta(t - t') \tag{10.19}$$

is given by

$$T(r, t, r', t') = \int_0^t dt' \int_{\mathrm{r}} q(r', t') \cdot G(r, t, r', t') d^3 r' \tag{10.20}$$

$G(r, t, r't')$ is the Green's function for the given heat source. In $3 - D$ rectangular coordinates the Green's function is,

$$G(r, t, r', t') = \frac{1}{\{4\pi D(t - t')\}^{\frac{3}{2}}} \cdot \exp\left(-\frac{(r - r')^2}{4D(t - t')}\right) \tag{10.21}$$

where,

$$(r - r')^2 = (x - x')^2 + (y - y')^2 + (z - z')^2 \tag{10.22}$$

For a rectangular heat source $-\frac{a}{2} < x < \frac{a}{2}, -\frac{b}{2} < y < \frac{b}{2}$, and $0 < z < \frac{c}{2}$

$$G(x, a, t) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{\frac{a}{2} + x}{\sqrt{4Dt'}}\right) + erf\left(\frac{\frac{a}{2} - x}{\sqrt{4Dt'}}\right)\right] \tag{10.23}$$

$$G(y, b, t) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{\frac{b}{2} + y}{\sqrt{4Dt'}}\right) + erf\left(\frac{\frac{b}{2} - y}{\sqrt{4Dt'}}\right)\right] \tag{10.24}$$

$$G(z, c, t) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{z}{\sqrt{4Dt'}}\right) + erf\left(\frac{\frac{c}{2} - z}{\sqrt{4Dt'}}\right)\right] \tag{10.25}$$

The heat flow $q(t')$ is given by $2 \cdot P(t')/(a \cdot b \cdot c/2)$ where $P(t')$ is the input power in Watts, and the factor of 2 accounts for the reflective boundary condition at $z = 0$.

Using the above equations, the temperature at a point r at time t is given by

$$T(r, t) = T_0 + \frac{P_0}{\rho C_p abc} \int_0^t G(x, a, t') \cdot G(y, b, t') \cdot G(z, c, t') dt' \tag{10.26}$$

where the power $P(t')$ is considered to be time-independent and equal to P_0. Hence, the peak temperature at the center of the heat source $r = 0$ is

$$T(0, t) = T_0 + \frac{P_0}{\rho C_p abc} \int_0^t \mathrm{erf}\left(\frac{a}{4\sqrt{Dt'}}\right) \cdot \mathrm{erf}\left(\frac{b}{4\sqrt{Dt'}}\right) \cdot \mathrm{erf}\left(\frac{c}{4\sqrt{Dt'}}\right) dt' \tag{10.27}$$

The diffusion times related to a, b, and c, are defined as

$$t_a = \frac{a^2}{4\pi D} \tag{10.28}$$

$$t_b = \frac{b}{4\pi D} \tag{10.29}$$

$$t_c = \frac{c^2}{4\pi D} \tag{10.30}$$

and are the times required for the device to reach thermal equilibrium in the x, y, and z directions of the device. If $a > b > c$, then after a time t_a, the device has reached its steady state temperature and $dT/dt = 0$.

The error functions can be approximated using [Dwyer90]

$$\mathrm{erf}(x) \approx \begin{cases} \frac{2x}{\sqrt{\pi}}; & \text{if } x \leq \sqrt{\frac{\pi}{2}} \\ 1; & \text{if } x \geq \sqrt{\frac{\pi}{2}} \end{cases} \tag{10.31}$$

to obtain an estimate of the value of the integral in Equation 10.27. The approximations in Equation 10.28 and Equation 10.30 are excellent except in the range $0.22 < x < 1.82$, and are considered to be reasonable in this range. Using these approximations, the four time-dependent regions of the power vs. time curve in Figure 10.6 are given by

$$P(t) = \frac{\rho C_p abc \cdot (T - T_0)}{t}; \quad 0 < t < t_c \tag{10.32}$$

$$P(t) = \frac{ab\sqrt{\pi K \rho C_p} \cdot (T - T_0)}{\sqrt{t} - \frac{\sqrt{t_c}}{2}}; \quad t_c < t < t_b \tag{10.33}$$

$$P(t) = \frac{4\pi K a \cdot (T - T_0)}{\ln\left(\frac{t}{t_b}\right) - 2 - \frac{c}{b}}; \quad t_b < t < t_a \tag{10.34}$$

$$P(t) = \frac{2\pi K a \cdot (T - T_0)}{\ln\left(\frac{a}{b}\right) + 2 - \frac{c}{2b} - \sqrt{\frac{t_a}{t}}}; \quad t_a < t \tag{10.35}$$

Equation 10.35 is the steady state equation when t tends to infinity. Equation 10.33 reduces to the Wunsch-Bell equation given in Equation 10.14 for small t_c; that is, when the heat source is effectively a two dimensional plane. It is possible to link the relationship between the different regions and the equations obtained by Wunsch and Bell [Wunsch68], Tasca [Tasca70] and Arkhipov *et al.* [Arkhipov83] in terms of the relative dimensions of a, b and c. If $a = b = c$, then the equations reduce to that of the adiabatic ($1/t$) dependence and the steady state condition in Equation 10.32 and Equation 10.35 respectively. The infinite cylinder described by Arkhipov *et al*, has $a = \infty$ and $b = c$, and the thermal behavior is described by Equation 10.32 and Equation 10.34. Similarly, other

forms of heat sources can be approximated by combinations of Equation 10.32 to Equation 10.35. Experimentally, the power vs. time curve can be used to obtain the actual dimensions of the heat source by determining the time at which the transition is made from one region to the next [Dwyer90].

While the above analysis allows the entire range of electrical overstress related thermal failure in semiconductor structures to be studied, the ESD conditions actually require a much more limited approach. The ESD pulse duration is on the order of 100 ns to 200 ns, and can be studied using a constant current pulse with pulse widths in that range. The dimensions of the devices being stressed indicate that the time to failure typically lies in the interval $t_b < t < t_a$ [Dwyer90][Amerasekera91]. Therefore, one need only apply Equation 10.34 to the analysis of the behavior of ESD protection circuits. Using a combination of experimental and analytical methods, the process and design dimensions influencing ESD can be extracted and used to improve the ESD performance [Pierce88][Amerasekera91]. However, this is an interactive process performed during or after technology development in order to characterize the technology, and does not achieve the requirements of a predictive modeling technique.

10.5 ELECTROTHERMAL DEVICE SIMULATIONS

Electrothermal simulations with full coupling between the electrical and thermal equations are important to accurately describe the behavior of the device in the high current region close to second breakdown. In Section 10.3 it was shown that the onset of second breakdown was dependent on the rate of change of the avalanche and thermal generation currents with temperature. Hence, the coupling between the temperature and the current densities, impact ionization coefficients, mobilities, and electric fields are important in simulating ESD phenomenon in devices. Full coupling requires that the correct forms of the current flow equations in the presence of thermal gradients are used [Watchutka90][Selberherr84]. Following the method of Stratton [Stratton72], the current densities for electrons and holes, $\vec{J}_n$ and $\vec{J}_p$ are written as

$$\vec{J}_n = qn\mu_n \cdot \vec{E} + qD_n \cdot \vec{\nabla} n + qnD_n^T \cdot \vec{\nabla} T \tag{10.36}$$

$$\vec{J}_p = qp\mu_p \cdot \vec{E} - qD_p \cdot \vec{\nabla} p - qpD_p^T \cdot \vec{\nabla} T \tag{10.37}$$

where D_n and D_p are the diffusion constants for electrons and holes, and $D_n^T \approx D_n/2T$ and $D_p^T \approx D_p/2T$ are the thermal diffusion constants for electrons and holes [Stratton72], n and p are the electron and hole concentrations and μ_n and μ_p are the electron and hole mobilities. The thermal gradients denoted by $\vec{\nabla} T$ account for the additional driving force of temperature on the current, which is the Seebeck effect [Geballe55].

These equations have been subsequently incorporated into rigorous treatments of heat generation and conduction in semiconductor devices [Alwin77][Adler78]

[Chryssafis79][Watchutka90]. However, they increase the complexity of the simulations, resulting in longer computation times and increased difficulties in obtaining converged solutions. Therefore, many simulations for second breakdown [Gaur76][Ward76][Orvis83] used isothermal diffusion coefficients. Such simulations, while limited, improve the understanding of the phenomena involved in second breakdown by providing a qualitative insight into the phenomena involved. In particular, simulations were able to show that second breakdown was indeed thermally initiated in the time durations being considered. [Ward76][Koyanagi77] [Orvis83]. The question of whether electrical or thermal effects were responsible for second breakdown had been asked for many years, based on work on snapback in avalanche diodes and *npn* transistors [Steele62][Grutchfield66][Roman70] [Caruso74][Hower80].

The heat source, H (Wcm^{-3}), was originally considered simply from the Joule heating term $J \cdot E$ [Ward76]. However, the heat gained by the lattice through recombination was shown later to play an important part in the thermal process [Adler78]. A form of the heat source is given by [Watchutka90]

$$H = \vec{J} \cdot \vec{E} + (R - G) \cdot (\varepsilon_g + 3kT) - \frac{\vec{J}}{q} \cdot \left(\frac{3}{2} k(T) \cdot \vec{\nabla} T + \frac{1}{2} \nabla \varepsilon_g \right) \tag{10.38}$$

$\vec{J}$ is the total current density, $\vec{E}$ is the electric field, ε_g is the energy gap, T is the temperature, R is the recombination rate, and G is the generation rate. The first term on the right-hand side is the standard Joule heating term, the second term accounts for lattice heating due to recombination/generation, the third term brings in the Thompson effect due to the heating (or cooling) that takes place when carriers traverse a region with spatially varying thermoelectric power, P_n or P_p [Callen60][Sze81], caused by large temperature gradients. In reverse-biased junctions Joule heating will be most dominant, with some influence of the Thompson heating in regions with large temperature gradients. In forward-biased junctions, Joule heating provides the initial temperature rise, while the recombination term plays a large role in the heating effect at higher temperatures [Adler78][Krabbenborg91].

The temperature-dependent thermal conductivity $k(T)$ is given by [Selberherr84]

$$k(T) = \frac{1}{0.03 + 1.56 \times 10^{-3} T + 1.65 \times 10^{-6} T^2} \tag{10.39}$$

Mobility is modeled using the Caughey-Thomas empirical mobility model [Caughey67] for the concentration dependent zero-bias mobilities μ_{0n} and μ_{0p} in cm^2/V-s for electrons and holes.

$$\mu_{0n,0p} = \mu_{n,p}^{min} + \frac{\mu_{n,p}^{L} - \mu_{n,p}^{min}}{1 + \left[\frac{(T/300)^{-3.8}}{N/N_{\mathrm{ref}}} \right]^{\beta_{n,p}}} \tag{10.40}$$

$\mu_{n,p}^{L}$ is the lattice mobility, which is temperature dependent, $\mu_{n,p}^{min}$ is the coefficient for ionized impurity scattering, N is the local total impurity concentration and N_{ref}

is a constant [Selberherr84]. $\beta_{n,p}$ is an empirically determined constant. The zero-bias mobility is then used to determine the mobility at high electric fields using the equation

$$\mu_{n,p} = \frac{\mu_{0n,0p}}{1 + \frac{\mu_{0n,0p} \cdot E}{v_{\text{sat}}}} \tag{10.41}$$

v_{sat} is the temperature-dependent saturation velocity [Selberherr84],

$$v_{\text{sat}} = \frac{2.4 \times 10^7}{1 + 0.8 \cdot \exp\left(\frac{T}{600}\right)} \tag{10.42}$$

The impact ionization coefficients need to be temperature dependent, and there have been many empirical forms used in the literature [Chynoweth58][Crowell66] [Overstraeten70][Okuto75][Ward76]. An empirically determined form for the impact ionization coefficients for electrons and holes $\alpha_{n,p}$/cm, as a function of temperature [Okuto75] is,

$$\alpha_{n,p} = A_{n,p} \cdot \{1 + C_{n,p} \cdot 10^{-4}(T - 300)\} \cdot E \cdot \exp\left(\frac{-B_{n,p}^2 \cdot [1 + D_{n,p} \cdot (T - 300)]^2}{E^2}\right) \tag{10.43}$$

$A_n = 0.426/\text{V}$, $A_p = 0.243/\text{V}$, $B_n = 4.81 \times 10^5\,\text{V cm}^{-1}$, $B_p = 6.53 \times 10^5\,\text{V cm}^{-1}$, $C_n = 3.05 \times 10^{-4}$, $C_p = 5.35 \times 10^{-4}$, $D_n = 6.86 \times 10^{-4}$, and $D_p = 5.87 \times 10^{-5}$ are the coefficients for electrons and holes; E is in Vcm^{-1}. The above equations can be simplified to obtain $\alpha_{n,p}$ by reducing to,

$$\alpha_{n,p} = C_1 \cdot \exp(-C_2/E) \tag{10.44}$$

where $C_2 = E_g/q\lambda$ and λ is the mean free path of the carrier [Crowell66]

$$\lambda = \lambda_0 \tanh(E_{r0}/2kT) \tag{10.45}$$

$\lambda_0 \approx 50\,\text{Å}$ and $E_{r0} = 50\,\text{meV}$ are λ and the optical phonon energy, E_r, at 0 K. A reasonable fit to experimental results has been obtained by using [Grant73]

$$\alpha_{n,p} = C_1 \cdot \exp(-b(T)/E) \tag{10.46}$$

where the coefficient C_1 remains constant as a function of temperature and the major variation with temperature is assumed to occur in the exponent. $b(T)$ is assumed to have a linear dependence on temperature and for electrons $db_e/dT \approx 1.3 \times 10^3\,\text{cm V}^{-1}$-K, while for holes $db_h/dT \approx 1.1 \times 10^3\,\text{cm V}^{-1}$-K [Grant73]. The simplified forms can be used in the development of analytical models (e.g., [Abderhalden91]), but it is best to use the complete form in a full electrothermal model.

The temperature rise in the structure being simulated is a very strong function of the ambient conditions and, therefore, strongly dependent on the thermal boundary conditions [Amerasekera93B][Yang93]. The analytical models described in the

previous section assumed that heat was dissipated into a semi-infinite medium. It was shown that, considering the dimensions of the silicon, these approximations are reasonable. However, in an electrothermal simulator, the implementation of a large volume of silicon surrounding the device leads to drastic increases in the computation time. Hence, it is prudent to use thermal boundary conditions closely resembling the actual conditions in the simulations.

The thermal boundary conditions at the contacts are implemented using a lumped thermal resistance, $R_{th}\,\mathrm{K\,W^{-1}}$ [Yang93]. The temperature at the contact, T_{cont} is determined from

$$T_{cont} - T_{amb} = R_{th} \cdot \Lambda \tag{10.47}$$

Λ is the heat flux through the contact obtained from Equation 10.14. T_{amb} is the ambient temperature, and $T_{cont} = T_{amb}$ for $R_{th} = 0\,\mathrm{K\,W^{-1}}$. At the other boundaries, distributed thermal resistances should be used, allowing each boundary node to be contacted to a thermal resistance. In addition, to account for the heat capacity of the surrounding silicon, distributed thermal capacitances, C_{th}, need to be included at the boundaries. As would be expected, C_{th} plays a significant role in transient thermal simulations.

It should be noted that none of these models have been experimentally verified above temperatures of 1000 K. It is possible that at higher temperatures the behavior of these parameters will change. However, it has been observed that simulations of simple semiconductor structures using these models with internal device temperatures $>$1000 K show a reasonable agreement with the experimentally observed behavior as indicated in Figure 10.7 for an n-type resistor [Amerasekera93B]. The reduction in current after saturation is a function of the thermal resistance used at the contacts. These simulations indicate that the thermal resistance is too high, and that results in a higher internal temperature and higher resistance compared to the experimental result. This curve is shown as an example of how the thermal parameters will influence the simulation results. Changing (tuning) the thermal resistance will produce very close matching between experiment and simulation. The temperature at which thermal snapback occurs is 1100 K for the simulated I–V curve.

The I–V curve of a $p^+/n/n^+$ diode is shown in Figure 10.8 as a function of the contact thermal resistance, R_{th}. Curve (1) has $R_{th} = 0\,\mathrm{K\,W^{-1}}$, Curve (2) has $R_{th} = 10^4\,\mathrm{K\,W^{-1}}$, and Curve (3) has $R_{th} = 10^6\,\mathrm{K\,W^{-1}}$. It is seen that R_{th} has a large influence on both the second breakdown voltage and the current. Such simulations have shown [Amerasekera93B] that the onset of second breakdown is the result of conductivity modulation taking place in the device. Conductivity modulation occurs when the number of generated holes become large enough to support part of the current and, therefore, the electric field does not need to increase further to support an increase in the current. Typically, the onset of conductivity modulation occurs when $p \approx 0.2n$ [Amerasekera93B]. In the resistor structures, the simulations showed that the decrease in resistivity with increased temperature alone is not responsible for the collapse in voltage. It is essential that there is significant hole generation to enable the conductivity modulation mechanism to be initiated for second breakdown to occur.

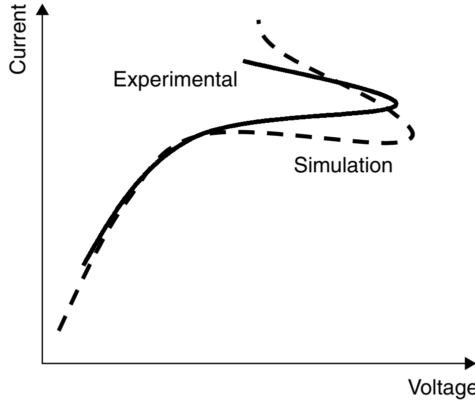


Figure 10.7 High current $I-V$ curve of an N-type resistor comparing experiment to electrothermal device simulations

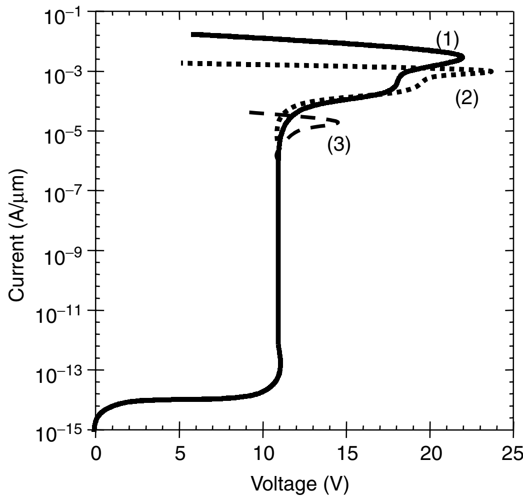


Figure 10.8 $I-V$ curves for a $P+/N/N+$ diode from electrothermal simulations showing the effect of the contact thermal resistance, R_{th}. Curve (1) has $R_{th} = 0\,\mathrm{K\,W^{-1}}$; Curve (2) has $R_{th} = 10^4\,\mathrm{K\,W^{-1}}$; Curve (3) has $R_{th} = 10^6\,\mathrm{K\,W^{-1}}$

The voltage as a function of time obtained from electrothermal simulations of n-channel MOS transistors under constant current pulsed conditions is shown in Figure 10.9. The gate, substrate, and source are at zero volts. At V_{t1} the parasitic *npn* begins to turn on. The *npn* is fully on at V_{sp}, which is the snapback holding voltage. The temperature begins to rise, which causes the avalanche multiplication

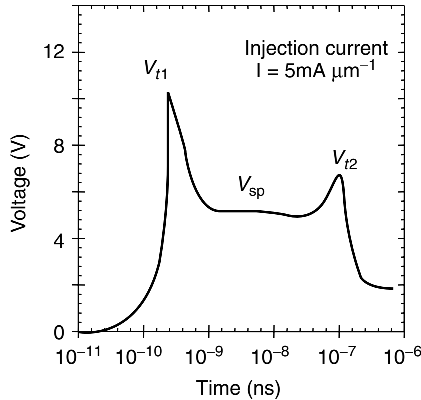


Figure 10.9 Electrothermal device simulation results showing the variation of drain voltage with time for a constant injection current in an nMOS with gate, source, and substrate at zero volts. The bipolar trigger voltage V_{t1}, snapback V_{sp} and second breakdown V_{t2} are clearly seen

to reduce, thereby requiring a higher drain voltage to sustain the *npn* in the on-state. At V_{t2} the temperature is $\sim$1300 K and the thermally generated holes can now provide the base current reducing the need for avalanche generated carriers and the voltage begins to decrease rapidly. The temperature is a function of the thermal boundary conditions, and this simulation used $R_{th} = 10^6\,\mathrm{K\,W^{-1}}$ in order to reduce internal temperature gradients and speed up the computation. Actual values of R_{th} are closer to $10^4\,\mathrm{K\,W^{-1}}$ [Amerasekera93B][Diaz93A]. The avalanche and thermal currents in the device as a function of temperature are shown in Figure 10.10. Second breakdown is observed to take place when the thermally generated current, I_{th}, is large enough to augment the avalanche generated current, I_{av}, in supporting the lateral *npn* in the on-state. As the temperature increases, I_{av} decreases due to the decrease in the impact ionization coefficient, while I_{th} increases. When I_{th} becomes a sufficiently large fraction of the base current, then I_{av} is no longer required to increase to support the transistor action and the voltage eventually begins to decrease. The resultant negative resistance region results in instabilities and the onset of thermal second breakdown [Amerasekera94B].

10.6 CONCLUSION

Analytical thermal modeling is a very useful tool for the analysis of experimental data. Analytical models enable the evaluation of the influence of process and design parameters on the high current behavior of these structures. They also provide a means by which quick sanity checks can be made on the quality of the data being collected and the relationship to numerical device simulation results.

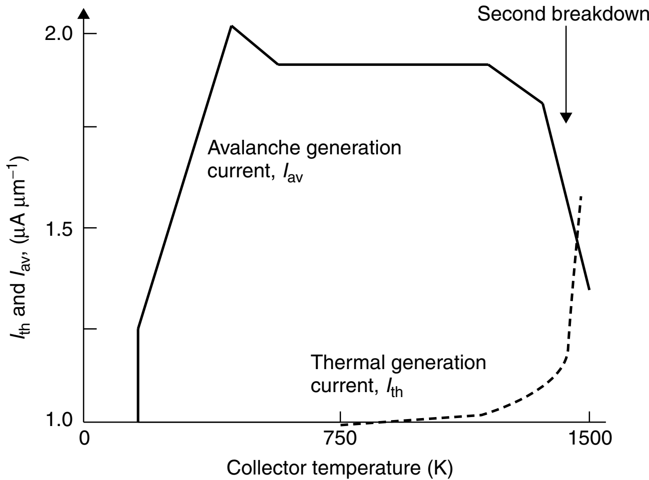


Figure 10.10 The thermal generation current I_{th} and the avalanche generation current I_{av} as a function of the junction temperature. Results are obtained from electrothermal device simulations for an nMOS. Second breakdown occurs when the current required to support *npn* action is provided by I_{th}, allowing I_{av} to reduce

Numerical device simulations provide an excellent means by which to gain understanding of the high current phenomena. The insight gained from the analysis of these simulations will enable improvements to be made for high current performance, as well as the development of new protection circuits using these elements. With these caveats in mind, the use of device simulations is strongly recommended in the study and evaluation of ESD robustness in semiconductor devices.

BIBLIOGRAPHY

[Abderhalden91] J. Abderhalden, *Untersuchugen zur Optimierung von Schutzstrukturen gegen electrostatische Entladungen in integrierten CMOS-Schaltungen*, PhD Thesis, Eidgenossische Technische Hochschule (ETH), Zurich, 1991.

[Adler78] M. Adler, "Accurate calculations of the forward drop and power dissipation in thyristors" *IEEE T. Electron Dev.*, **ED-25**, 16–22, 1978.

[Alexander78] D.R. Alexander, "Electrical overstress failure modeling for bipolar semiconductor components", *IEEE Trans. Comp. Hyb. Man. Tech.*, **CHMT-1**, 345–353, 1978.

[Alwin77] V.S. Alwin, D. Navon and L. Turgeon, "Time-dependent carrier flow in a transistor structure under nonisothermal conditions", *IEEE T. Electron Dev.*, **ED-24**, 1297–1303, 1977.

[Amerasekera90] E.A. Amerasekera, L.J. van Roozendaal, J. Abderhalden, J.J.P. Bruines and L. Sevat, "An analysis of low voltage ESD damage in advanced CMOS processes", in *Proc. 12th EOS/ESD Symposium*, pp. 143–150, 1990.

[Amerasekera91] A. Amerasekera, L. van Roozendaal, J. Bruines and F. Kuper, "Characterization and modeling of second breakdown in NMOST's for the extraction of ESD-related process and design parameters", *IEEE T. Electron Dev.*, **ED-38**, 2161–2168, 1991.

[Amerasekera92] E.A. Amerasekera, W. van den Abeelen, L.J. van Roozendaal, M. Hannemann and P.J. Schofield, "ESD failure modes: characteristics, mechanisms and process influences", *IEEE T. Electron Dev.*, **ED-39**, 2, 1992.

[Amerasekera93A] A. Amerasekera, A. Chatterjee and M.-C. Chang, "Prediction of ESD Robustness of a Process Using 2-D Device Simulations", in *Proc. 31st Int. Rel. Phys. Symp.*, pp. 161–167, 1993.

[Amerasekera93B] A. Amerasekera, M.-C. Chang, J. Seitchik, A. Chatterjee, K. Mayaram and J.-H. Chern, "Self-heating effects in basic semiconductor structures", *IEEE T. Electron Dev.*, **ED-40**, 1836–1844, 1993.

[Amerasekera94A] A. Amerasekera and R. Chapman, "Technology design for high current and ESD robustness in a deep submicron process", *Elec. Dev. Lett.*, **15**, 1994.

[Amerasekera94B] A. Amerasekera and J. Seitchik, "Electrothermal behavior of deep submicron nmos transistors under high current snapback (ESD/EOS) conditions", in *Tech. Dig. IEDM*, 455–458, 1994.

[Arkhipov83] V.I. Arkhipov, E.R. Astvatsaturyan, V.I. Godovitsyn and A.I. Rudenko, *Int. J. Electron.*, **55**, 395–401, 1983.

[Bardeen50] J. Bardeen and W. Shockley, "Deformation potentials and mobilities in nonpolar crystals," *Phys. Rev.*, **80**, 72–76, 1950.

[Beltman90] R.A.M. Beltman, H. van der Vlist and A.J. Mouthaan, "Simulation of Thermal Runaway During ESD Events", in *Proc. 12th EOS/ESD Symposium*, pp. 157–161, 1990.

[Burgess60] R. Burgess, "Negative resistance in semiconductor devices", *Can. J. Phy.*, **38**, 369–375, 1960.

[Callen60] H.B. Callen, *Thermodynamics*, John Wiley & Sons, New York, 1960.

[Carslaw59] H.S. Carslaw and J.C. Jaeger, *Conduction of Heat in Solids*, 2nd edn., Oxford University Press, London, 1959.

[Caughey67] D.M. Caughey and R.E. Thomas, "Carrier mobilities in silicon empirically related to doping and field", *Proc. IEEE*, **55**, 2192–2193, 1967.

[Caruso74] A. Caruso, P. Spirito and G. Vitale, "Negative resistance induced by avalanche injection in bulk semiconductors", in *IEEE T. Electron Dev.*, **ED-21**, 578–586, 1974.

[Chatterjee91] A. Chatterjee, T. Polgreen and A. Amerasekera, "Design and Simulation of a 4 kV ESD Protection Circuit for 0.8 μm BiCMOS process", in *Tech. Dig. IEDM*, 913–916, 1991.

[Chen71] H.C. Chen, W.M. Portnoy and D.K. Ferry, "Doping dependence of second breakdown in a $p-n$ junction", *Sol. St. Elec.*, **14**, 747–751, 1971.

[Chryssafis79] A. Chryssafis and W. Love, "A computer-aided analysis of one-dimensional thermal transients in *npn* power transistors", *Sol. St. Elec.*, **22**, 249–256, 1979.

[Chynoweth58] A. Chynoweth, "Uniform silicon $p-n$ junctions II, ionization rates for electrons", *J. Appl. Phys.*, **31**, 1161–1165, 1958.

[Crowell66] C. Crowell, S.M. Sze, "Temperature dependence of avalanche multiplication in semiconductors", *Appl. Phys. Lett.*, **9**, 233–242, 1966.

[Daniel90] S. Daniel and G. Krieger, "Process and Design Optimization for Advanced CMOS I/O ESD Protection Devices", in Proceedings of the 12th EOS/ESD Symposium, pp. 206–213, 1990.

[Diaz92] C. Diaz, S. Kang, C. Duvvury and L. Wagner, "Electrical overstress (EOS) power profiles: a guideline to qualify EOS hardness of semiconductor devices", *Proc. 14th EOS/ESD Symposium*, 88–94, 1992.

[Diaz93A] C. Diaz, C. Duvvury and S.-M. Kang, "Studies of EOS Susceptibility in 0.6 μm nMOS ED I/O Protection Structures", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 83–91, 1993.

[Diaz93B] C. Diaz, C. Duvvury and S.-M. Kang, "Electrothermal simulation of electrical overstress in advanced nMOS ESD I/O protection circuits", in *Tech. Dig. IEDM*, 899–902, 1993.

[Dwyer90] V.M. Dwyer, A.J. Franklin and D.S. Campbell, "Thermal failure in semiconductor devices", *Sol. St. Elec.*, **33**, 553–560, 1990.

[Duvvury88] C. Duvvury, R. Rountree and O. Adams, "Internal chip ESD phenomena beyond the protection circuit", in *Proc. 26th IRPS*, 19–25, 1988.

[Brown72] W.D. Brown, "Semiconductor device degradation by high amplitude current pulses", *IEEE T. Nucl. Sci.*, **NS-19**, 68–75, Dec. 1972.

[Duvvury92A] C. Duvvury and C. Diaz, "Dynamic gate-coupled nMOS for efficient output ESD protection", in *Proc. 30th IRPS*, 141–150, 1992.

[Gaur76] S. Gaur, D. Navon and R. Teerlinck, "Transistor design and thermal stability", *IEEE T. Electron Dev.*, **ED-20**, 527–534, 1976.

[Ghandi78] S.K. Ghandhi, *Semiconductor Power Devices*, John Wiley & Sons, New York, 1977.

[Geballe55] T.H. Geballe and G.W. Hull, "Seebeck effect in silicon", *Phys. Rev.*, **98**, 940–947, 1955.

[Grant73] W.N. Grant, "Electron and hole ionization rates in epitaxial silicon at high electric fields", *Sol. St. Elec.*, **16**, 1189–1203, 1973.

[Grutchfield66] H.B. Grutchfield and T.J. Moutoux, "Current mode second breakdown in epitaxial planar transistors", *IEEE T. Electron Dev.*, **ED-13**, 743, 1966.

[Hower80] P.L. Hower and V.G.K. Reddi, "Avalanche injection and second breakdown in transistors", *IEEE Trans. Elec. Dec.*, **ED-17**, 320–335, 1970.

[Jacoboni77] C. Jacoboni, C. Canali, G. Ottaviani and A. Quaranta, "A review of some charge transport properties of silicon", *Sol. St. Elec.*, **20**, 77–89, 1977.

[Koyanagi77] K. Koyanagi, K. Hane and T. Suzuki, "Boundary conditions between current mode and thermal mode second breakdown in epitaxial planar transistors", *IEEE T. Electron Dev.*, **ED-24**, 672–678, 1977.

[Krabbenborg91] B. Krabbenborg, R. Beltman, P. Wolbert and T. Mouthan, "Physics of Electro-Thermal Effects in ESD Protection Devices", in *Proceedings of the* 13th *EOS/ESD Symposium*, pp. 98–103, 1991.

[Krakauer92] D. Krakauer and K. Mistry, "ESD Protection in a 3.3 V Sub-Micron Silicided CMOS Technology", in *Proceedings of the* 14th *EOS/ESD Symposium*, pp. 250–257, 1992.

[Krieger87] G. Krieger, "Thermal response of integrated circuit input devices to an electrostatic energy pulse", *IEEE T. Electron Dev.*, **ED-34**, 877–882, 1987.

[Kurimoto94] K. Kurimoto, K. Yamashita, I. Miyanaga, A. Hori and S. Odanaka, "An Electrothermal Circuit Simulation Using an Equivalent Thermal Network for Electrostatic Discharge", in *Proc. VLSI Symp. on Tech.*, 127–128, 1994.

[Khurana66] B.S. Khurana, T. Sugano and H. Yanai, "Thermal breakdown in silicon $p-n$ junction devices", *IEEE T. Electron Dev.*, **ED-13**, 763–770, 1966.

[Lin90] D.L. Lin, "Thermal breakdown of VLSI by ESD pulses", in *Proc.* 29th *IRPS*, 281–287, 1990.

[Mathews80] D. Mathews, "Some Design Criteria for Avoiding Second Breakdown in Bipolar Devices", in *Proceedings of the 2nd EOS/ESD Symposium*, pp. 117–121, 1980.

[Mayaram91] K. Mayaram, J.-H. Chern, L. Arledge and P. Yang, "Electrothermal Simulation Tools for Analysis and Design of ESD Protection Devices", in *Tech. Dig. IEDM*, 909–912, 1991.

[Melchior64] H. Melchior and M.J.O. Strutt, "Secondary breakdown in transistors", *Proc. IEEE (Correspondence)*, **52**, 439–440, 1964.

[Miller57] S.L. Miller, "Ionization rates for holes and electrons in silicon", *Phys. Rev.*, **105**, 1246–1249, 1957.

[Muller86] R.S. Muller and T.I. Kamins, *Device Electronics For Integrated Circuits*, 2nd edn.,John Wiley & Sons, New York, 1986.

[Nienhuis66] R.J. Nienhuis, "Second breakdown in the forward and reverse base current region", *IEEE T. Electron Dev.*, **ED-13**, 655–662, 1966.

[Okuto75] Y. Okuto and C.R. Crowell, "Threshold energy effect on avalanche breakdown voltage in semiconductor junctions", *Sol. St. Elec.*, **18**, 161–168, 1975.

[Overstraeten70] R. van Overstraeten and H. de Man, "Measurement of the ionization rates in diffused silicon $p-n$ junctions", *Sol. St. Elec.*, **13**, 583–608, 1970.

[Popescu70] C. Popescu, "The thermal runaway mechanism of second breakdown phenomenon", *Sol. St. Elec.*, **13**, 887–901, 1970.

[Luchies94] J. Luchies, C. de Kort and J. Verweij, "Fast Turn-on of an nMOS ESD Protection Transistor; Measurements and Simulations", in *Proceedings of the 16th EOS/ESD Symposium*, pp. 266–272, 1994.

[Orvis83] W.J. Orvis, C.F. McConaghy, J.H. Yee, G.H. Khanaka, L.C. Martin and D.H. Lair, "Modeling and Testing for Second Breakdown Phenomena", in *Proceedings of the 5th EOS/ESD Symposium*, pp. 108–117, 1983.

[Pierce88] D.G. Pierce, W. Shiley, B.D. Mulcahy, K.E. Wagner and M. Wunder, "Electrical Overstress Testing of a 256 K UVEPROM to Rectangular and Double Exponential Pulses", in *Proceedings of the 10th EOS/ESD Symposium*, pp. 137–146, 1988.

[Polgreen89] T. Polgreen and A. Chatterjee, "Improving the ESD Failure Threshold of Silicided nMOS Output Transistors by Ensuring Uniform Current Flow", in *Proceedings of the 11th EOS/ESD Symposium*, pp. 167–174, 1989.

[Ridley63] B.K. Ridley, "Specific negative resistance in solids", *Proc. Phys. Soc.*, **82**, 954–966, 1963.

[Roman70] G. Roman, "A model for computation of second breakdown in transistors", *Sol. St. Elec.* **13**, 961–980, 1970.

[Runyan65] W.B. Runyan, *Silicon Semiconductor Technology*, McGraw-Hill, New York, 1965.

[Scarlett63] R.M. Scarlett, W. Shockley and R.H. Haitz, "Thermal Instabilities and Hot Spots in Junction Transistors", in *Physics of Failure in Electronics*, ed: M.F. Goldberg and J. Vaccaro, Spartan Books, Baltimore, Md, pp. 194–203, 1963.

[Schafft62] H.A. Schafft and J.C. French, "Second breakdown in transistors", *IRE Trans. Elec. Dev.*, **ED-9**, 129–136, 1962.

[Schafft66] H.A. Schafft and J.C. French, "A survey of second breakdown", *IEEE T. Electron Dev.*, **ED-13**, 613–618, 1966.

[Scott86] D. Scott, G. Giles and J. Hall, “A Lumped Element Model for Simulation of ESD Failures in Silicided Devices”, in *Proceedings of the 8th EOS/ESD Symposium*, pp. 41–47, 1986.

[Selberherr84] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer-Verlag, New York, 1984.

[Shaw92] M.P. Shaw, V.V. Mitin, E. Schöll and H.L. Grubin, *The Physics of Instabilities in Solid State Electron Devices*, Plenum, New York, 1992.

[Smith73] W.B. Smith, D.H. Pontius and P.P. Budenstein, “Second breakdown and damage in junction devices”, *IEEE T. Electron Dev.*, **ED-20**, 731–744, 1973.

[Tauc57] J. Tauc and A. Abraham, “Thermal breakdown in silicon $p-n$ junctions”, *Phys. Rev.*, **108**, 936–937, 1957.

[Speakman74] T.S. Speakman, “A model for the failure of bipolar silicon integrated circuits subjected to electrostatic discharge”, in *Proc. 12th IRPS*, 60–69, 1974.

[Steele62] M.C. Steele, K. Ando and M.A. Lampert, “Avalanche breakdown double injection induced negative resistance of semiconductors”, *J. Phys. Soc. Jpn.*, **17**, 1729–1736, 1962.

[Stratton72] R. Stratton, “Semiconductor current-flow equations (diffusion and degeneracy)”, *IEEE T. Electron Dev.*, **ED-19**, 1288–1300, 1972.

[Sze81] S.M. Sze, *Physics of Semiconductor Devices*, 2nd edn., John Wiley & Sons, New York, 1981.

[Tasca70] D.M. Tasca, “Pulse power failure modes in semiconductors”, *IEEE. T. Nucl. Sci.*, **NS-17**, 364–372, Dec. 1970.

[Tasca72] D.M. Tasca, J.C. Peden and J. Miletta, “Non-destructive screening for thermal second breakdown”, *IEEE T. Nucl. Sci.*, **NS-19**, 57–67, 1972.

[Ward76] A.L. Ward, “An electrothermal model of second breakdown”, *IEEE T. Nucl. Sci.*, **NS-23**, 1679–1684, 1976.

[Ward77] A.L. Ward, “Studies of second breakdown in silicon diodes”, *IEEE T. Parts Hyb. Pac.*, **PHP-13**, 361–368, Dec. 1977.

[Watchutka90] G. Wachutka, “Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling”, *IEEE T. Comput. Aid. D.*, **CAD-9**, 1141–1149, 1990.

[Wunsch68] D.C. Wunsch and R.R. Bell, “Determination of threshold failure levels of semi-conductor diodes and transistors due to pulse voltages”, *IEEE T. Nucl. Sci.*, **NS-15**, 244, Dec. 1968.

[Yang93] P. Yang and J.-H. Chern, “Design for reliability: the major challenge for VLSI”, *Proc. IEEE*, **81**, 730–744, 1993.

[Yee82] J. Yee, W. Orvis, L. Martin and J. Peterson, “Modeling of Current and Thermal Mode Second Breakdown Phenomena”, in *Proceedings 4th EOS/ESD Symposium*, pp. 76–81, 1982.

11 Circuit Simulation Basics, Approaches, and Applications

Sridhar Ramaswamy

11.1 INTRODUCTION

Advances in processing technology and very-large-scale integration (VLSI) scaling are increasing demand for effective protection circuits. Circuit simulation enables designers to optimize protection circuits through simulation before final evaluation on silicon, thus reducing the number of learning cycles. Protection circuits are designed to meet certain specifications (ESD/EOS failure levels) and CAD tools can be effectively used in the design stage. Circuit-level simulations were used to develop and optimize novel protection circuits without going through a number of design cycles [Chatterjee91][Duvvury92]. Such simulation approaches use SPICE and depend on the existing or extracted SPICE parameters for a given protection circuit [Diaz93B][Diaz93C].

This chapter describes circuit-level device models that can be incorporated into SPICE-like circuit simulators [Nagel75]. The device models extend to the avalanche breakdown and high-current regime, which is the operating condition of devices under EOS/ESD. Previous attempts at modeling the devices used existing circuit models to describe the behavior under ESD conditions. In order that these models be used to design effective protection circuits, they should be scalable and accurately predict the behavior of the devices under high-injection conditions. The models should comprehend the importance of the layout-dependent substrate resistance. The model parameters must be easily extracted using simple measurements. The simulators need to be fast and accurate to be useful to circuit designers. This chapter describes models that fulfill these requirements.

The first section of this chapter is devoted to modeling MOS devices for ESD simulation. The subsequent sections discuss modeling other important circuit elements like bipolar devices, diffusion resistors and junction diodes for high-current operation. The application of the models in a SPICE-like circuit simulator with

some protection circuit examples is detailed, and a discussion on fully coupled electrothermal simulators is presented. While electrothermal simulators are not fully predictive they do offer useful insight into the high-temperature behavior of protection elements under ESD stress.

11.2 MODELING THE MOSFET

Most commercially available circuit simulators do not cover the high-current region of the MOSFET operation, but allow an approximate analysis of their behavior under ESD [Duvvury92][Krakauer94]. The use of MOS-based protection circuits [Duvvury92][Worley95] and the integration of the protection circuits with internal circuitry [Krakauer94][Dabral94] increase the need for simulators that are more capable of accurately reproducing the behavior of the circuit under ESD conditions. During an ESD event, the internal gates of the chip can be at different potentials depending on the coupling of the power supply bus to the ESD voltage [Chen88][Mistry90]. The turn-on of the ESD protection circuit and its effectiveness are dependent on the turn-on behavior of the I/O buffers and the state of the internal logic during the stress. Hence, an ESD circuit simulator must include the high-current behavior of the MOS transistors under gate bias conditions. It also allows circuit operation during an EOS or overvoltage stress under power-up conditions to be determined. This enables designers to identify potential weaknesses in the circuit under both ESD and EOS conditions. In this section, the behavior and equations of high-current conduction are presented, and the development of ESD simulators is reviewed. First, the operation of the parasitic bipolar transistor associated with the MOSFET, which is active under ESD, and then the operation of the forward-biased diode that is built into the MOSFET are described. The important role played by the layout-dependent substrate resistance is also discussed in detail.

11.2.1 Modeling the Parasitic Bipolar Transistor

Standard circuit simulators use bipolar model equations (e.g., Gummel-Poon) that cover the normal operating ranges for circuit performance. These models have a number of parameters that are generated using data measured under normal operating conditions of the device. The current densities during ESD are significantly higher than typical operating conditions for which MOS devices are designed. The regions of the $I-V$ curve are depicted schematically in Figure 11.1. Regions 1 and 2 are the linear and saturation regions governed by standard MOS equations. Region 3 is the avalanche breakdown region in which standard MOS equations are no longer valid. Region 4 is the bipolar or *snapback* region where the standard device model equations do not extend. It is in regions 3 and 4 that the ESD/EOS operation takes place.

Figure 11.2 shows a cross section of an nMOS transistor indicating the currents during a high-current stress. The avalanche generation of carriers in the high-field

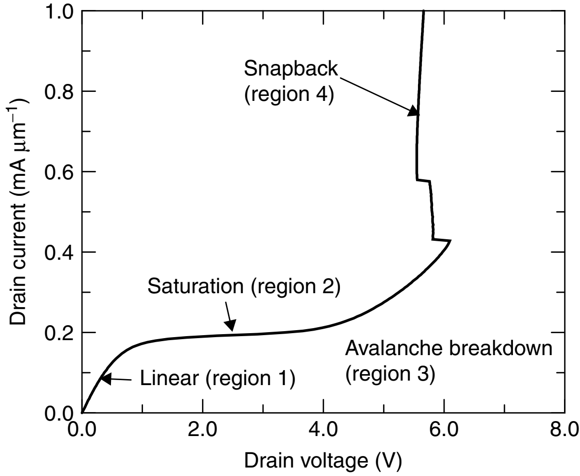


Figure 11.1 Generic $I-V$ curve for an nMOS transistor under gate bias showing the different regions of operation. The standard SPICE models cover regions 1 and 2

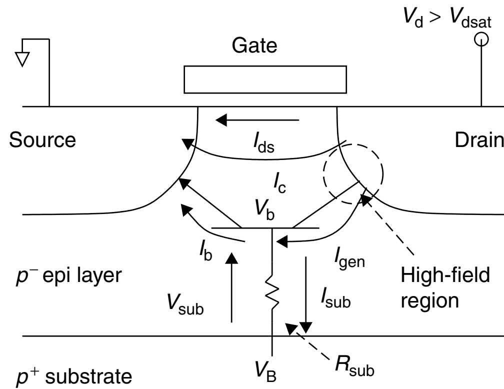


Figure 11.2 Cross section of an nMOS transistor showing the currents in the parasitic *npn* transistor

region near the drain results in hole current into the substrate, I_{sub}. The voltage dropped across the substrate resistance (R_{sub}) due to I_{sub} raises the local substrate potential (V_{sub}) and causes the source-substrate junction to become forward-biased. Electrons injected from the source into the substrate are collected at the drain and form a lateral *npn* bipolar transistor with the drain as the collector, the source as the emitter and substrate as the base. The effectiveness of this transistor is dependent on the emitter injection efficiency and the base transport factor, which is dependent on the effective channel length [Sun78].

Bipolar transistor models under avalanche conditions for circuit simulation were developed prior to ESD requirements [Dutton75][Reisch92][Guedes88][Scott86]. Table look-up techniques have also been used to simulate the complex $I-V$ curves during and beyond snapback [Kurimoto94]. Advances were made [Diaz94A] in developing the first ESD circuit simulators, which showed good correlation to measurement using standard Gummel-Poon (GP) bipolar models to describe the parasitic bipolar transistor. The problem with the GP models is the number of additional parameters required to define the bipolar transistor. This increases the number of measurements and complicates the device parameter set for the circuit simulator.

The readers should note that the parasitic bipolar transistor operates under a limited set of bias conditions and that the complex parameter sets associated with the GP models are not required to reproduce the MOS high-current behavior. The ideal extraction methodology should be targeted at the specific $I-V$ curves, which are to be reproduced; the high-current MOS $I-V$ curves including drain (I_d), source (I_s) and substrate (I_{sub}) terminal currents contain all the relevant information for parameter extraction. A simple extraction methodology is described to extract the intrinsic bipolar parameters for the MOS transistor.

A number of models were developed to simulate different aspects of the parasitic bipolar device in nMOS transistors. In [Luchies94] the transient turn-on of the parasitic bipolar within an nMOS transistor was studied. An Ebers-Moll model with extensions describing the mechanisms playing an important role in breakdown along with a coupled electrothermal model was used. A compact model to explain the behavior of a grounded-gate nMOS under CDM stress was developed in [Russ96]. The model had inherent symmetry to cope with oscillating CDM/MM pulses and accounted for relevant turn-on and turn-off mechanisms and addressed issues of clamping and power dissipation. An extension of a bipolar model for MOS transistors was proposed [Wolf98], which considered gate-coupling effects during HBM stresses. This model combines the high-current bipolar and MOS operations and includes the modulation of bipolar gain at high-current levels. Another model, which can easily be incorporated into commercial simulators is discussed in [Lim97]. This model uses existing elements in commercial simulators to model parasitic bipolar action with lumped resistors used for substrate resistance. The model is symmetric and can be used for CDM/MM simulations. The impact of gate RC elements on HBM and CDM behavior was investigated and modeled in [Mergens00]. The RC elements cause a phase shift between the gate and source transients leading to possible gate oxide breakdown under CDM stress.

In order for a MOS model to be useful for designing ESD protection circuits, it should model the following effects accurately. The parasitic bipolar associated with the MOS device must be modeled. The substrate resistance and the built-in diodes must also be modeled to account for high-injection effects. The number of parameters should be limited and easy to extract. The circuit equations describing the models should be simple, should have good convergence properties and not degrade the accuracy and speed of the simulator. The models should be predictable,

that is, they must be scalable and should comprehend the impact of layout on the device performance.

11.2.2 Parasitic Bipolar Action

In the nMOS transistor of Figure 11.2, the voltage drop in the substrate (with $V_B = 0$) is given by $V_{sub} = I_{sub}R_{sub}$. As V_{sub} approaches 0.7 V, the forward biasing of the source-substrate junction will cause the parasitic bipolar to begin to turn on. The emitter current is given by I_e and the collector current by I_c with I_b as the base-emitter current. I_{sub} is a function of the avalanche multiplication factor M in the high-field region of the drain. The avalanche generation current at the high-field region because of an incident current I_p is given by [Dutton75]

$$I_{gen} = (M - 1)I_p \tag{11.1}$$

With $V_g = 0$, the incident current at the drain junction is solely due to thermal generation, minority carrier diffusion, and band-to-band tunneling. Before the bipolar transistor turns on, $I_{gen} = I_{sub}$. A typical value of I_{sub} for bipolar turn-on is a few hundred $\mu A\,\mu m^{-1}$, while I_p at room temperature can be as low as 10^{-19} $\mu A\,\mu m^{-1}$. Therefore, M must tend to infinity as the drain voltage V_d approaches the avalanche breakdown voltage V_{av}.

A gate voltage V_g greater than the MOS threshold voltage V_{th} will result in an MOS current I_{ds} between the drain and the source. $I_p \approx (I_{ds})$ will now be much larger, and a lower M can sustain the same I_{sub}. Hence, the drain voltage at which the bipolar turn-on is initiated reduces as a function of V_g, as illustrated in Figure 11.3.

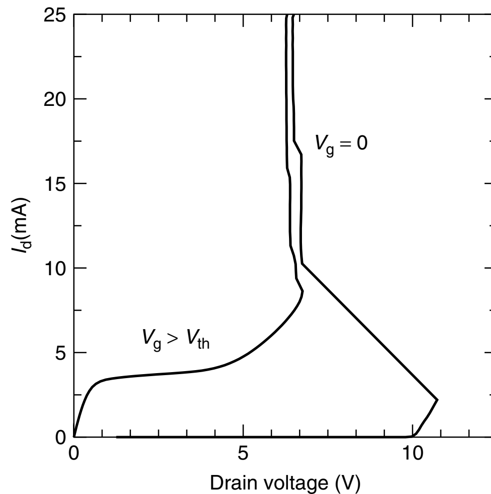


Figure 11.3 I_d–V_d curve of an nMOS device as a function of gate bias

As the bipolar transistor turns on, the bipolar current, I_c, provides an additional current source for multiplication, further reducing the value of M required to sustain the bipolar in the on-state. V_d can reduce even more and voltage 'snapback' is observed, where V_d drops to a sustaining level V_{sp}. The value of V_{sp} is dependent on the V_d required to maintain the V_{sub} at the level needed to sustain the bipolar current. The condition for sustaining snapback is given by (from Chapter 4)

$$\beta(M-1) \geq 1 \tag{11.2}$$

Hence V_{sp} is a function of R_{sub}, M, and the gain of the intrinsic parasitic bipolar, β [Gupta98].

It is important to note that the β of the self-biased parasitic bipolar is different from the β of the lateral bipolar with external bias on V_B. The parasitic bipolar is formed by the sidewall regions of the source and drain junctions acting as emitter and collector [Krieger89A]. The voltage across the base-emitter junction, V_{be}, is provided by the local substrate potential and the base current is usually provided by the internal current source due to avalanche generation at the drain junction. In contrast, the lateral bipolar is biased through the substrate (base) contact, and V_{be} is provided externally [Lindmeyer67][Vandebroek91]. Such a device has a larger emitter area, most of which does not contribute to the bipolar itself but will influence the measurement of parameters, such as current gain and resistance. The base-emitter current is given by [Muller86]

$$I_b = I_{oe}\left[\exp\left(\frac{V_{be}}{V_T}\right) - 1\right] \tag{11.3}$$

and the bipolar collector current is given by [Muller86]

$$I_c = I_{oc}\left[\exp\left(\frac{V_{be}}{V_T}\right) - \exp\left(\frac{V_{bc}}{V_T}\right)\right] \tag{11.4}$$

where I_{oe} is the reverse saturation current owing to diffusion of holes into the emitter of the *npn*, I_{oc} is the reverse saturation current because of the diffusion of electrons into the base of the *npn* and $V_T = kT/q$ is the thermal voltage. For an *npn* transistor, I_{oc} is given by

$$I_{oc} = \frac{qn_i^2 A_E D_n}{N_B L} \tag{11.5}$$

where n_i is the intrinsic concentration, A_E is the effective emitter area, D_n is the effective diffusion constant, N_B is the doping in the base and L is the base width. Typically, A_E is given by the sidewall junction area and L is the channel length. Similarly, I_{oe} for an *npn* is given by

$$I_{oe} = \frac{qn_i^2 A_E D_p}{N_E L_{pE}} \tag{11.6}$$

where D_p is the effective diffusion constant for holes in the emitter, N_E is the emitter doping concentration and L_{pE} is the hole diffusion length in the emitter. In shallow junction silicided processes, L_{pE} will be larger than the effective junction depth (X_{jeff}); therefore, X_{jeff} can be used instead. The current gain β of the bipolar is given by

$$\beta = \frac{I_c}{I_b} = \frac{I_{oc}}{I_{oe}} \tag{11.7}$$

Since β is related to the emitter injection efficiency (γ) and to the base transport factor (α_T) through

$$\beta = \frac{\gamma \cdot \alpha_T}{1 - \gamma \cdot \alpha_T} \tag{11.8}$$

the parameters I_{oc} and I_{oe} also define γ and α_T. Part of the generated holes will form the base-emitter current I_b, and the rest will provide the substrate current. Thus

$$I_{gen} = I_{sub} + I_b \tag{11.9}$$

where I_b is given by Equation (11.3). Therefore,

$$I_{sub} = (M - 1) \cdot (I_{ds} + I_c) - I_b \tag{11.10}$$

Once the bipolar transistor is fully turned on, V_g dependence becomes negligible if I_c is very much larger than I_{ds}, which is the case under high-injection conditions.

11.2.3 Avalanche Multiplication

The multiplication factor M can be described in terms of the impact ionization coefficient α [Chynoweth60][Sze91]

$$M = \frac{1}{1 - \int_0^{x_d} \alpha \, dx} \tag{11.11}$$

where x_d is the width of the depletion region, and α is given by [Chynoweth60],

$$\alpha = A \cdot \exp\left(\frac{-B}{E}\right) \tag{11.12}$$

where A and B are constants, and E is the electric field in the high- field region. E will vary across the depletion region for weak avalanche. However, for strong avalanche, as in the case of snapback, E could be assumed to be constant over the depletion width and M can be written as

$$M = \frac{1}{1 - \alpha x_d} \tag{11.13}$$

Empirically, M has been described in the form [Miller57]

$$M = \frac{1}{1 - (V_d/V_{av})^n} \tag{11.14}$$

where V_d is the applied drain voltage, V_{av} is the avalanche breakdown voltage, and n is a fitting parameter ranging from 2 to 6 depending on the type of junction being considered. This equation does not take into account the effect of V_g on the electric field at the drain junction under MOS conditions. It is, therefore, better to combine Equations (11.12) and (11.13) to give

$$M = \frac{1}{1 - A_i \exp\left(\frac{-B_i}{V_{ds} - V_{dch}}\right)} \tag{11.15}$$

Alternatively, Equation (11.9) could have been written as

$$M = \frac{1}{1 - \left(\frac{V_{ds} - V_{dch}}{V_{av}}\right)^n} \Rightarrow \left(\frac{V_{ds} - V_{dch}}{V_{av}}\right)^n = A_i \exp\left(\frac{-B_i}{V_{ds} - V_{dch}}\right) \tag{11.16}$$

In the above equations, A_i and B_i are fitting parameters, where $A_i \approx A \cdot x_d$ and $B_i \approx B \cdot x_d$. The voltage across the high- field region of width x_d is given by $(V_d - V_{dch})$. The effect of V_g on the drain electric field is included through the drain-source saturation voltage V_{dch} given by [Huang94][Ramaswamy97]

$$V_{dch} = \frac{\alpha(V_{gs} - V_{th})}{\alpha + \gamma(V_{gs} - V_{th})} \tag{11.17}$$

where α and γ are constants and V_{th} is the threshold voltage. As pointed out in Chapter 4, there is a direct impact of the gate voltage on the failure current of the device. At gate biases below V_{th}, V_{dch} is zero and a large drain bias is needed to trigger the bipolar device. At moderate V_{gs} greater than V_{th} the channel current adds to the incident current generating additional electron-hole pairs. This reduces the M required and leads to a lower V_{ds} needed to trigger the parasitic bipolar. From Equation (11.17) it can be seen that as V_{gs} increases, V_{dch} increases to a point where the effective voltage across the drain junction needed to sustain the bipolar action starts to increase. This leads to a higher snapback voltage and therefore, a lower failure current.

For $V_g = 0$ as V_d approaches V_{av}, M approaches infinity, and Equation (11.15) can be rewritten as

$$V_{av} = \frac{B_i}{\ln(A_i)}. \tag{11.18}$$

This equation can be used to relate A_i and B_i if V_{av} is already extracted and is used in the parameter extraction procedure described later. As M approaches infinity, the simulator could encounter convergence problems. Different techniques have been used to overcome these problems as discussed in [Ramaswamy96B][Lim97].

When the bipolar transistor turns on, Equation (11.1) can be rewritten as

$$I_{gen} = (M - 1)I_{ds} + (M_0 - 1)I_c \tag{11.19}$$

where I_{ds} is the drain-source current due to MOS action, while I_c is the collector-emitter current due to bipolar action. The multiplication factor M_0 is computed

using Equation 11.15 with $V_{gs} = 0$. The drain current, I_{ds}, for a given V_{gs} is determined by the MOS model equations.

11.2.4 Circuit-Level Model Equations

The equivalent circuit of the bipolar as implemented in the model along with the MOS transistor is shown in Figure 11.4. The following equations summarize the snapback model:

$$V_{dch} = \frac{(V_{gs} - V_{th})}{[\alpha + \gamma(V_{gs} - V_{th})]} \tag{11.20}$$

$$M(V_{gs}) = \frac{1}{1 - A_i \exp\left(\frac{-B_i}{V_{ds} - V_{dch}}\right)} \tag{11.21}$$

$$I_c = \frac{\beta I_s}{Q_b}\left[\exp\left(\frac{V_{bs}}{V_T}\right) - \exp\left(\frac{V_{bd}}{V_T}\right)\right] \tag{11.22}$$

$$Q_b = \frac{1}{2}\left(1 + \sqrt{1 + 4\frac{I_b}{I_{kf}}}\right) \tag{11.23}$$

$$I_b = I_s\left[\exp\left(\frac{V_{bs}}{V_T}\right) - 1\right] \tag{11.24}$$

$$I_{gen} = (M(V_{gs}) - 1)I_{ds} + (M(V_{gs} = 0) - 1)I_c \tag{11.25}$$

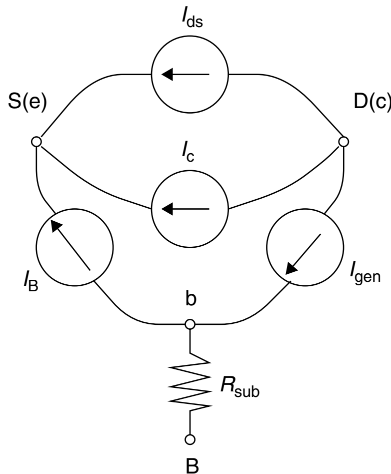


Figure 11.4 Current sources that represent the extension of the MOS model to include bipolar snapback

In addition to the forward characteristics, the reverse characteristics must also be modeled correctly. The equations governing the reverse mode of operation are

$$I_c = \frac{\beta I_s}{Q_{br}} \left[\exp\left(\frac{V_{bd}}{V_T}\right) - \exp\left(\frac{V_{bs}}{V_T}\right) \right] \tag{11.26}$$

$$Q_{br} = \frac{1}{2}\left(1 + \sqrt{1 + 4\frac{I_b}{I_{kfr}}}\right) \tag{11.27}$$

In the aforementioned equations, Q_b and Q_{br} represent the base charge factors, which model the high-current β degradation phenomena. This degradation is caused by an increase in minority carrier charges in the base under high-injection conditions [Sze91]. The knee currents I_{kf} and I_{kfr} determine the current levels at which the β degradation effects become prominent and they are inversely proportional to the lifetime of minority carriers in the base. The saturation current I_s scales with source area and it should be noted from the above equations that the diode model has been merged into the snapback model through the different current sources.

A comparison between experimental and simulation results is shown in Figure 11.5 for an nMOS transistor. Figure 11.6 shows a comparison between experimentally obtained pMOS $I-V$ data and the simulated results. An important observation from the experimental data is that the pMOS does not show snapback at high-current levels. The reason for the absence of snapback in the pMOS has been determined from device simulations to be because of the low hole mobility, which results in $I_{oc} < I_{oe}$ and a low β for the lateral *pnp*.

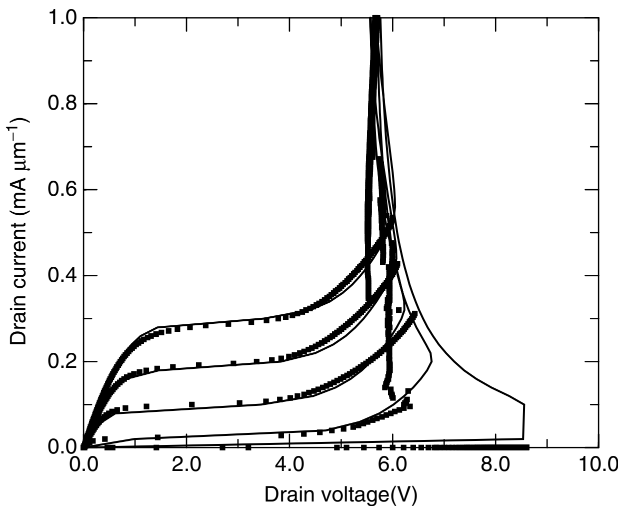


Figure 11.5 Comparison of experimental data (points) and simulated results (solid lines) for an nMOS transistor with $L = 0.525$ µm. Note the avalanche breakdown voltage is 8.6 V

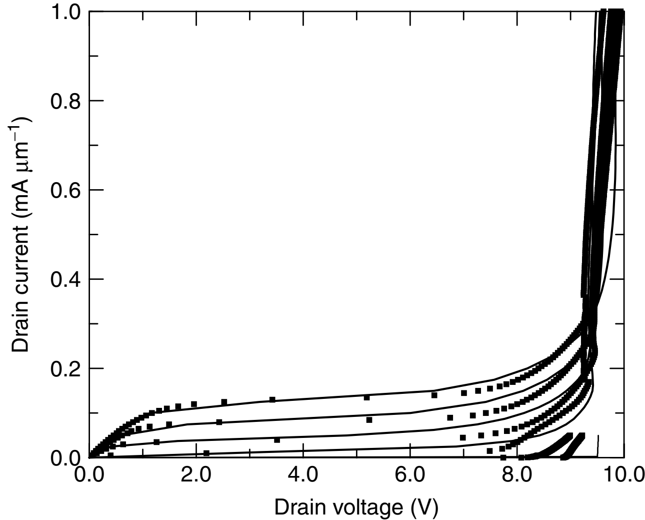


Figure 11.6 Comparison of experimental data (points) and simulated results (solid lines) for a pMOS transistor with $L = 0.525$ µm

11.2.5 Substrate Resistance

The on-resistance for diodes in forward-bias in deep submicron technologies is typically 5 Ω per micron width, while in reverse bias the resistance could be 10 to 100 times higher. Accurate modeling of this resistance is essential for circuit simulations of MOS and diode ESD protection circuits. A forward-biased diode or MOS junction modulates the substrate in its vicinity owing to minority carrier injection. On the other hand, reverse-biased junctions generate majority carriers in the substrate which at high fields experience velocity saturation. Hence, the direction of current flow determines the effective substrate resistance.

One can extract the low-field low-injection substrate resistance from a particular layout by solving the three-dimensional Poisson's equation,

$$\nabla^2 \psi = -\frac{\rho}{\epsilon} \tag{11.28}$$

where $\nabla^2 \psi$ is the Laplacian of the potential ψ, ρ is the charge density and ϵ is the permittivity. There are two different approaches to solving this equation. One approach uses the boundary element technique [Gharpurey96] while the other uses finite-difference methods [Verghese93][Li98]. Both these methods include spreading effects caused by three-dimensional current flow.

While these techniques for substrate resistance extraction are accurate for low-level injection, they do not account for conductivity modulation associated with

high injection of minority carriers or mobility degradation because of high fields. The carrier mobility depends on the electric field (E) as

$$\mu = \frac{\mu_o}{\left[1 + \left(\frac{E}{E_{sat}}\right)^n\right]^{\frac{1}{n}}} \tag{11.29}$$

where E_{sat} is the saturation electric field ($\approx 10^4\,\mathrm{V\,cm^{-1}}$) and n is a fixed constant (≈ 2 for electrons). The modified low-field resistance can, therefore, be written as

$$R = R_o\left[1 + \left(\frac{V_r}{V_c}\right)^n\right]^{\left(1+\frac{1}{n}\right)} \tag{11.30}$$

where R_o is the low-field resistance, V_r is the voltage across the resistor, and V_c is a fitting parameter. Experimental data and device simulations can be used to model $V_c = AR_o$, where A is a process dependent constant, which accounts for the three-dimensional effects.

When the substrate diode is forward-biased, there is a high injection of minority carriers into the substrate, which can significantly change the resistance. In MOS devices, the conductivity modulation manifests itself either through an increase in the substrate current after bipolar turn-on or through a reduction in resistance while forward biasing the source or drain junction diodes.

At low-injection levels the current density, J_f, is dominated by the drift current. However, at moderate and high-injection levels, J_f is dominated by diffusion currents in the vicinity of the junctions. It has been shown [Boselli01] that at high-injection levels $J_f \propto (\ln(V))^2$ where V is the voltage across the junction including the region being modulated. This results in the resistance varying as $R \propto \exp(\sqrt{J_f})/\sqrt{J_f}$. Therefore, the modulation phenomenon can be modeled as a current dependent resistor. The diode under high-injection conditions can be modeled as a series combination of an ideal diode and a variable resistor whose resistance is dependent on the direction of current flow. When the diode is forward-biased, the resistor is modeled as $R = K\exp(\sqrt{J_f})/\sqrt{J_f}$, where K is proportional to the separation of the contact from the junction. Figure 11.7 compares the experimental $I-V$ curve of a forward and reverse-biased well-diode indicating the low-injection, high-injection and velocity saturation regions.

As described earlier, the substrate resistor is connected to the body node of the MOS device and is modulated depending on the bias conditions of the device. Figure 11.8 shows the substrate resistance as a function of the current measured at the body node when (a) the parasitic bipolar is on and (b) when the drain junction is forward-biased. It can be seen for (a) when the drain junction is reverse-biased, the resistance is modulated after the parasitic bipolar turns on. When the drain-substrate junction (b) is forward-biased the substrate is conductivity modulated due to high-level electron injection. Figure 11.9 shows the good fit between simulation

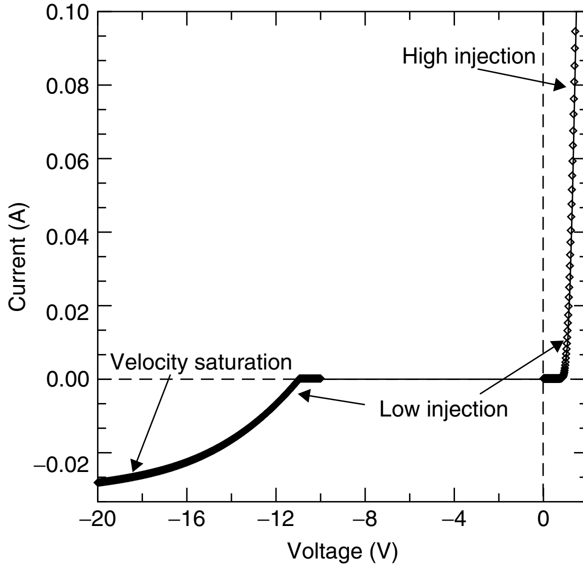


Figure 11.7 Experimental $I-V$ curve of a P^+/N^+ diode in an n-well compared with simulation results

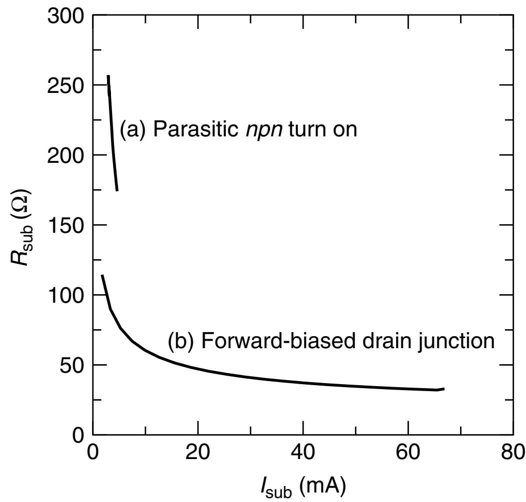


Figure 11.8 Substrate resistance as a function of substrate current when (a) the parasitic *npn* is turned on; and (b) when the drain junction diode is forward-biased

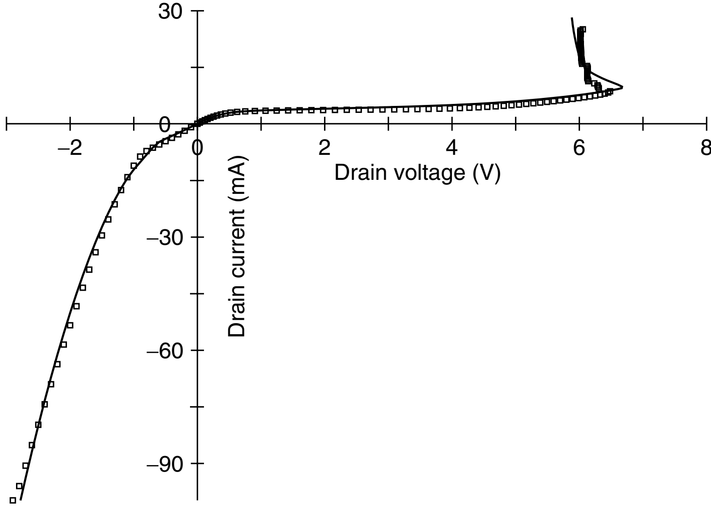


Figure 11.9 Drain current characteristics compared with simulation data for both forward and reverse modes of operation of an nMOS device

and experiment for the complete DC $I-V$ curve for an nMOS transistor for both positive and negative drain bias.

11.2.6 Base Transit-Time Effects

The time-dependent turn-on of the parasitic bipolar transistor is defined by the base transit time τ_{be} [Krieger89A][Muller86]. As the bipolar turn-on time is less than 250 ps for $L \leq 1$ µm, and the rise times for HBM and MM ESD stress are greater than 1 ns, τ_{be} is not an issue for these cases. However, very fast discharges such as with the Charged Device Model (CDM) test method have rise times, which can be less than 250 ps, and for simulating CDM events, τ_{be} must be included. This is accomplished through the addition of an effective diffusion capacitance (C_{be}) between the base and emitter of the bipolar, where C_{be} is a function of τ_{be} and the base charge [Luchies94]. The circuit elements C_{be} and C_{bc} are introduced to account for the finite diffusion time of the minority electrons across the base region. These capacitances control the turn-on time of the parasitic bipolar device, which is important when simulating fast ESD transients. The base-emitter diffusion capacitance is modeled by

$$C_{\mathrm{be}} = \tau_{\mathrm{be}} \frac{\partial I_{\mathrm{B}}}{\partial V_{\mathrm{be}}} = \frac{L^2}{\eta_{\mathrm{be}} \mu_{\mathrm{eff}}} \frac{I_{\mathrm{S}}}{V_{\mathrm{T}}{}^2} \exp\left(\frac{V_{\mathrm{be}}}{V_{\mathrm{T}}}\right) \tag{11.31}$$

where μ_{eff} is the effective electron mobility [Huang94], and η_{be} is a constant, which accounts for: (1) high-level injection effects; and (2) two-dimensional charge

injection into the base (substrate) [Krieger89A][Muller86]. A similar equation can be derived for the collector-base diffusion capacitance.

11.2.7 Parameter Extraction and Model Scalability

As with any circuit-level model, an extraction methodology is required to obtain the parameters for the model. This methodology needs to be simple, repeatable, and definite. It should have a short turnaround time for model parameter extraction and should be scalable. Keeping the above requirements in mind, for the model described in the preceding text, the following set of measurements is needed. A majority of the measurements are DC at probe level using simple test structures.

The avalanche breakdown voltage (V_{av}) of the drain junction is measured as shown in Figure 11.10. It is important that while taking this DC measurement a current limiting resistor be added in series to prevent damage to the drain junction. This allows us to relate the impact ionization parameters A_i and B_i using Equation (11.18). Then the substrate current as a function of gate bias (see Figure 11.11) is measured. This allows the extraction of α and γ and coupled with Equation (11.18) this also gives unique values for A_i and B_i. The forward and reverse $I_d - V_d$ curves shown in Figures 11.12 and 11.13 provide a measure for β and the knee currents. Figure 11.14 shows β as a function of the drain current obtained from the forward $I_d - V_d$ curve.

For the model to be useful for circuit designers it has to be scalable, that is, it must account for variations in channel length and width of the MOS devices. The impact ionization parameters A_i, B_i, α and γ are functions of channel length.

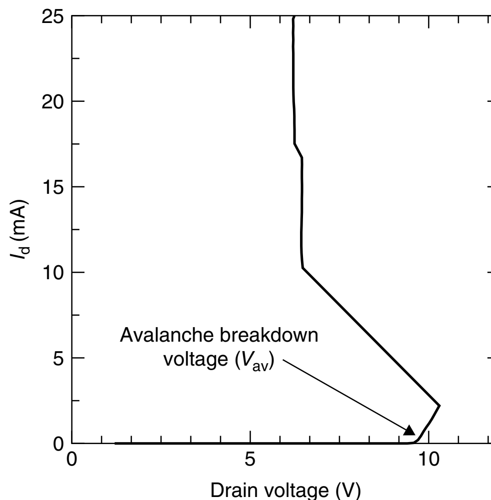


Figure 11.10 Avalanche breakdown voltage measurement curve

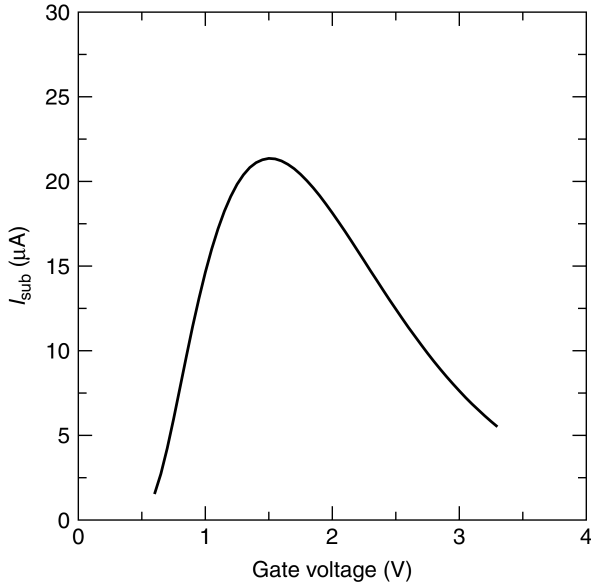


Figure 11.11 Substrate current as a function of gate bias. Initially I_{sub} increases with I_{ds}. However, at a certain point the depletion region at the drain increases reducing the E-field thereby reducing M and causing a drop in I_{sub}

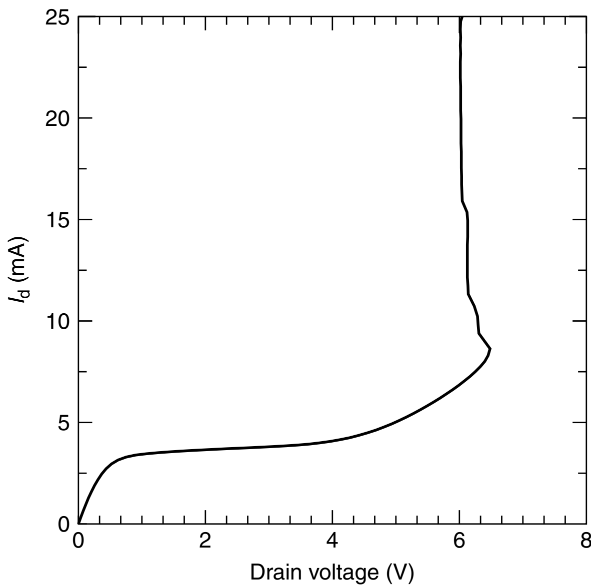


Figure 11.12 I_d–V_d curve with $V_g > V_{th}$ including snapback

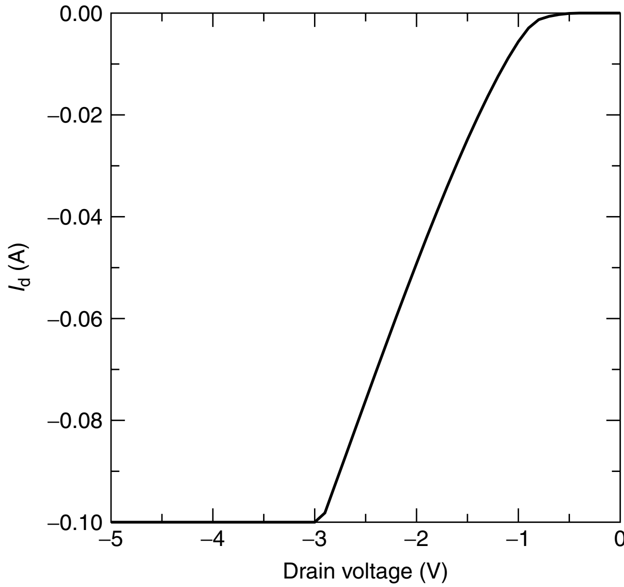


Figure 11.13 $I_d - V_d$ curve with the drain junction forward biased

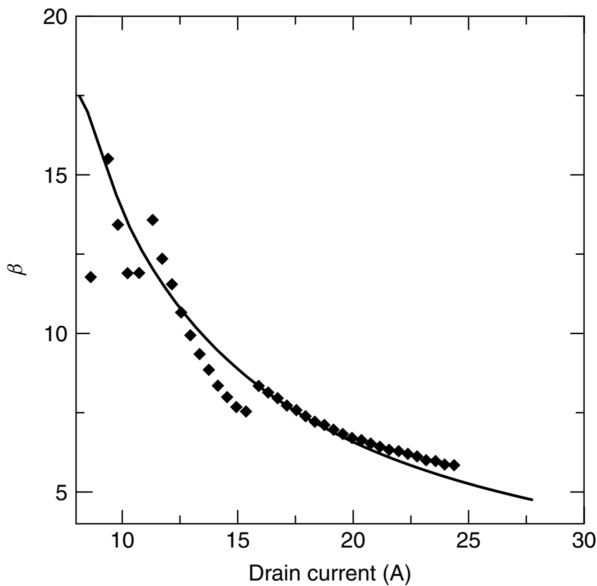


Figure 11.14 β as a function of drain current extracted from the $I_d - V_d$ curve beyond snapback. The data (points) compares well with the simulation (solid lines), which includes the high-current β degradation effect

The knee currents are scaled with width and length while β is scaled with length. Typically $\beta \propto L^{-m}$ from Equations (11.5) and (11.6) where $0 < m < 1$.

11.3 MODELING BIPOLAR JUNCTION TRANSISTORS

In order to accurately model the vertical parasitic bipolar devices, which are present in CMOS integrated circuits, one can use the GP model [Antognetti88][Gummel70] [Getreu78]. The GP model is based on an integral charge relation that relates electrical terminal characteristics to the base charge. As the collector and substrate nodes are the same in the vertical bipolar device, the additional substrate node that accounts for the collector-substrate capacitance can be omitted. The GP model accounts for base-width modulation and high-level injection effects. The model also includes the nonlinear charge elements, or equivalently, the voltage-dependent capacitances.

The primary dc parameters can be extracted using the Gummel plots (I_c, I_b vs. V_{be}) and the collector characteristics shown in Figures 11.15 and 11.16 below using techniques described in [ICCAP00]. The only parameter available for scalability over geometry is the emitter area, and, as shown in Figure 11.15 the model scales well with the emitter area.

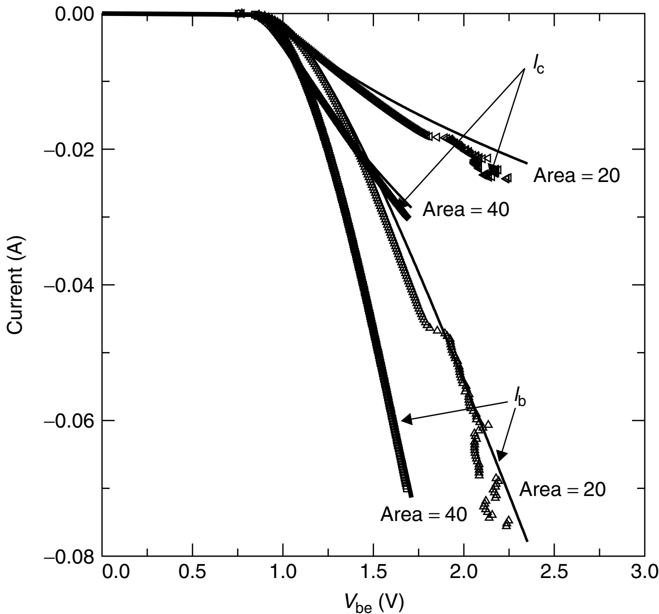


Figure 11.15 Gummel plots for vertical *pnp* devices with emitter widths of 20 and 40 μm. Simulation results show good agreement with measured data

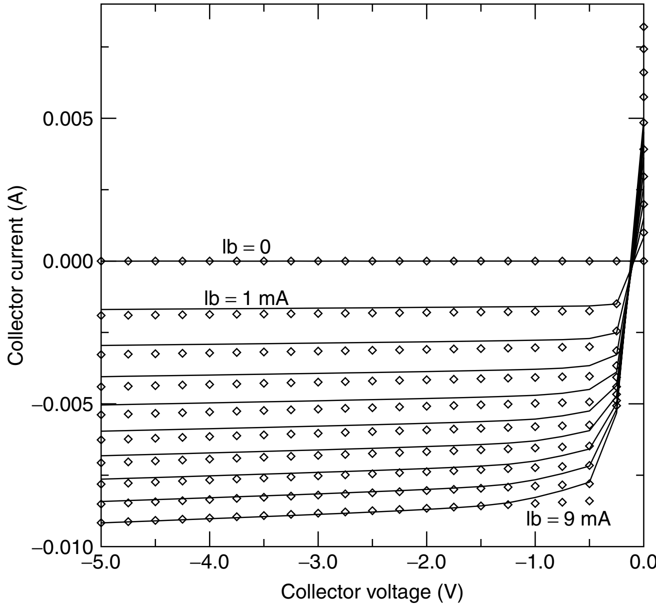


Figure 11.16 Collector characteristics for a vertical *pnp* device with $W = 40\ \mu\text{m}$. Simulation results match the measured data

Avalanche multiplication models in bipolar transistors are computationally intensive and can cause convergence problems as described in [Dutton75] [Severson86][Hebert87][Liou90A]. An improved approach to modeling bipolar transistor avalanche breakdown in circuit simulators was introduced in [Diaz92]. To include the avalanche breakdown of the collector-base junction, the static GP model is modified as shown in Figures 11.17(a) and (b) [Diaz92].

The additional nonlinear current source, i_{cm}, can be expressed in terms of the device terminal voltages and collector current without avalanche (i_{c}) as $i_{\text{cm}} = (M - 1)i_{\text{c}}$. The multiplication factor is once again given by

$$M = \frac{1}{1 - A_{\text{i}} \exp\left(\frac{-B_{\text{i}}}{V_{\text{cb}}}\right)} \tag{11.32}$$

where A_{i} and B_{i} are technology-dependent parameters and V_{cb} is the collector-base junction voltage. These parameters can be obtained by measuring the breakdown voltage under open base and open emitter conditions and using Equation (11.33) that follows.

In BiCMOS and bipolar processes, a vertical bipolar *npn* structure has better power dissipation capability than MOS device, which operates as a lateral bipolar under high-current conditions. The simulated output characteristics for a vertical *npn* transistor in a common-base and common-emitter configuration are

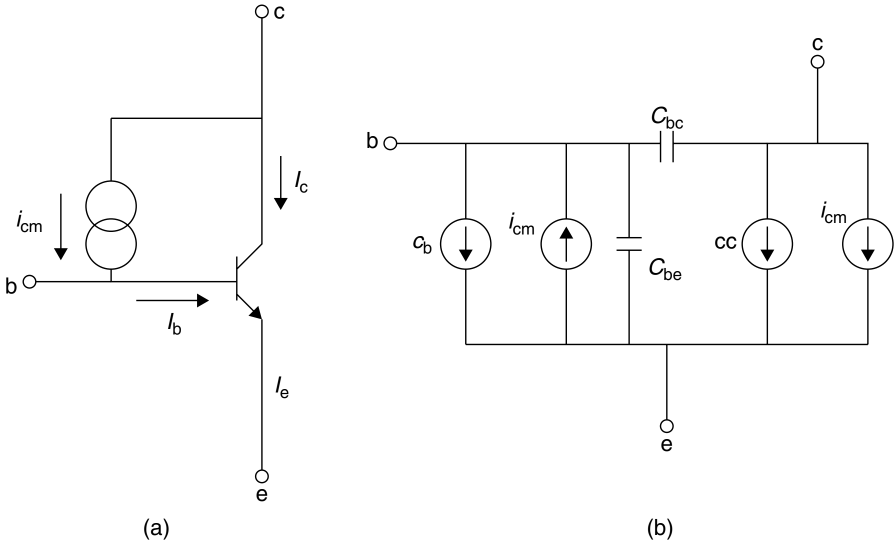


Figure 11.17 (a) The *npn* bipolar transistor with base-collector avalanche breakdown. (b) Corresponding transport model. The current sources *cc* and *cb* are calculated according to the standard GP model

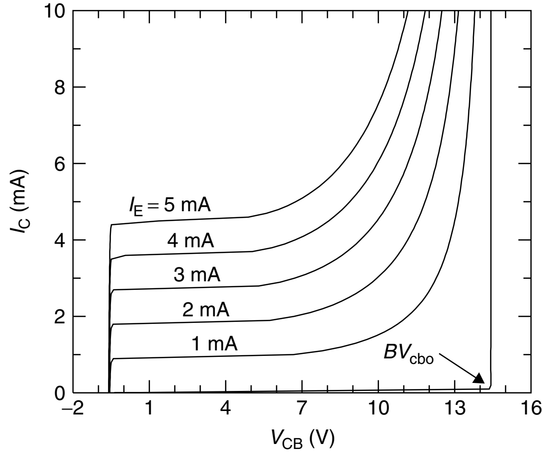
shown in Figures 11.18(a) and (b). The breakdown voltage under open base condition (BV_{cbo}) and the breakdown voltage under open emitter condition (BV_{ceo}) are given by

$$BV_{cbo} = \frac{B_i}{\ln(A_i)}, \quad BV_{ceo} = \frac{B_i}{\ln((\beta + 1)A_i)} \tag{11.33}$$

where β is the gain of the bipolar transistor. The two breakdown voltages are related by

$$BV_{ceo} = \frac{1}{\dfrac{\ln(\beta + 1)}{B_i} + \dfrac{1}{BV_{cbo}}} \tag{11.34}$$

For a grounded base device under a high-current stress on the collector, the collector voltage increases until the base-collector junction breaks down at BV_{cbo}. The avalanche generation current i_{cm} flows into the base terminal as a base current. This current would eventually turn on the bipolar transistor through the extrinsic base resistor and result in the *npn* entering the avalanche region to conduct the high stress current from collector to emitter with the collector voltage snapping back to BV_{ceo}. The *npn* can also be turned on by externally biasing the base. Initially, when the *npn* operates in the linear region, the base current flows from the external source into the base. However, it has been observed that after the *npn* enters the avalanche region, the base current changes direction and flows out of the base. This base current reversal effect has also been reported in [Lu89][Liou90B]. The effect indicates



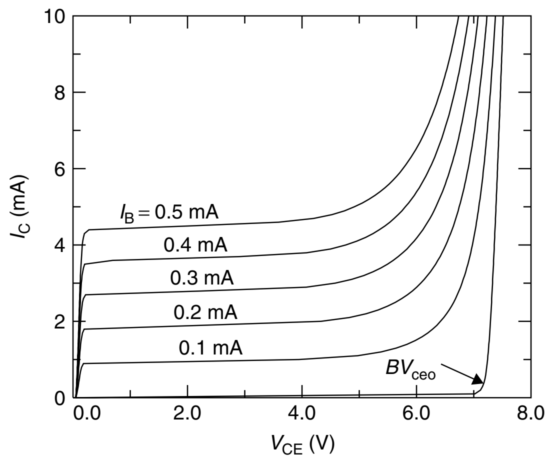


Figure 11.18 (a) Simulated output characteristics for an *npn* transistor in a common-base configuration. (b) Simulated output characteristics for the *npn* transistor in a common-emitter configuration. The main simulation parameters for the *npn* are $A_{\mathrm{i}} = 8$, $B_{\mathrm{i}} = 30\,\mathrm{V}$, $I_{\mathrm{S}} = 0.83\,\mathrm{pA}$ and $\beta = 10$

that whenever the *npn* is triggered into the avalanche region, a large avalanche generation current would supply the base current to bias the *npn* independent of the external bias condition. In turn, the self-biased *npn* would conduct the large stress current and clamp the collector at a lower voltage with the small on-resistance of the avalanche region. The high-current $I-V$ curves of an *npn* with different base biases (see Figure 11.19(a)) is shown in Figure 11.19(b). Hence, by adding

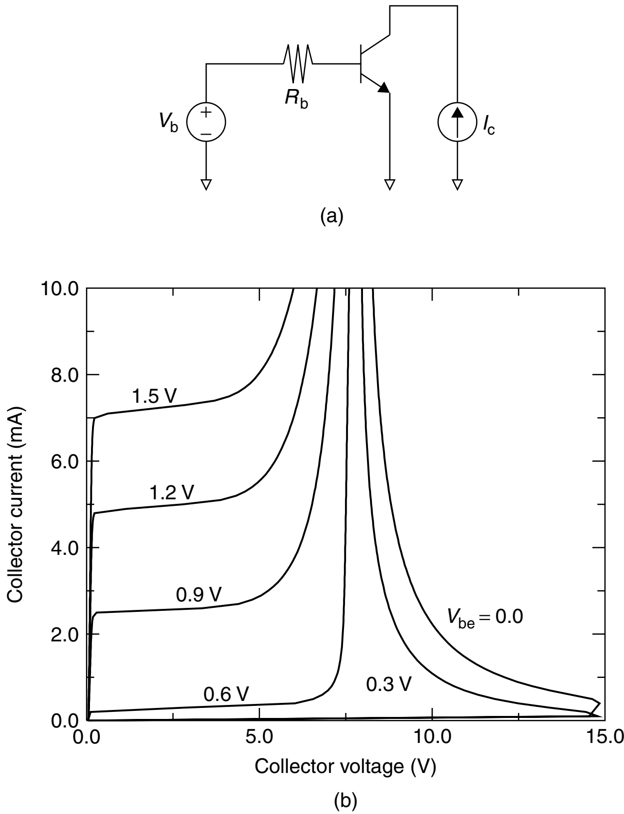


Figure 11.19 (a) Circuit schematic for simulating the high-current $I-V$ curves of an *npn*. (b) Simulated high-current $I-V$ curves of the *npn* with different base biases

an external base bias, the trigger voltage is reduced and makes the *npn* a useful ESD protection device for advanced BiCMOS and bipolar technologies [Chen96].

11.4 MODELING DIFFUSION RESISTORS

The diffusion resistor is a useful element in ESD protection circuit networks [Krieger89B]. Diffusion resistors are used to decouple primary and secondary protection devices in an input or output protection circuit. If they are not properly designed, however, they can become the bottleneck in the overall performance of the protection circuit. Taking advantage of velocity saturation effects at high-currents, such a resistor can be designed to have a very low resistance at normal operating currents, while effectively being much more resistive at current levels compatible with ESD.

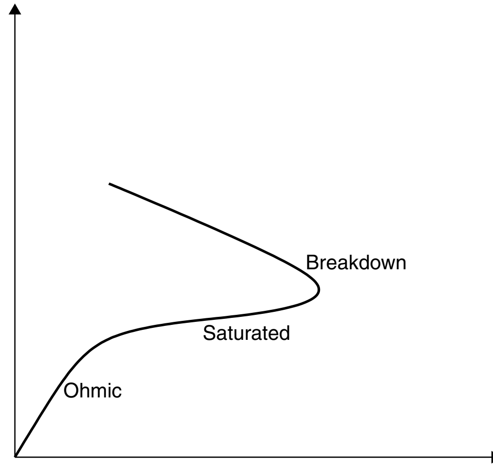


Figure 11.20 $I-V$ curve of a diffusion resistor, which indicates the different regions of operation

11.4.1 High-Field Transport

A typical $I-V$ curve for a diffusion resistor is shown in Figure 11.20. Low-field single carrier transport is usually described by the ideal ohmic relation

$$I_n = qn\mu_n W X_{\mathrm{j}} V_{\mathrm{ij}}/L \tag{11.35}$$

where I_n is the electronic current, V_{ij} is the voltage across the resistor, L is the effective resistor length, X_{j} is the effective junction depth, n is the electron concentration, and μ_n is the low-field electron mobility, which depends on the doping level (N_{d}) with quasi-neutrality ($n \approx N_{\mathrm{d}}$) implicitly assumed. With the continuing trend of scaling down device geometries, the related current densities are increased, and even heavily doped diffused layers may be subjected to mobility degradation due to the high electric field. At high fields the velocity saturation effect dominates and the current is limited to a saturation value given by

$$I_n = qnWX_{\mathrm{j}}v_{\mathrm{sat}} \tag{11.36}$$

where v_{sat} is the saturation velocity of electrons in silicon. A model that accounts for both the low and high field effects can, therefore, be written as

$$I_n = \frac{qnWX_{\mathrm{j}}V_{\mathrm{ij}}}{L\left(\dfrac{1}{\mu_n} + \dfrac{V_{\mathrm{ij}}}{v_{\mathrm{sat}}L}\right)} \tag{11.37}$$

11.4.2 Circuit-Level Snapback Model

To model the high-injection effect we have modified the simple diffusion resistor model by adding extra current sources that dominate the high-current snapback

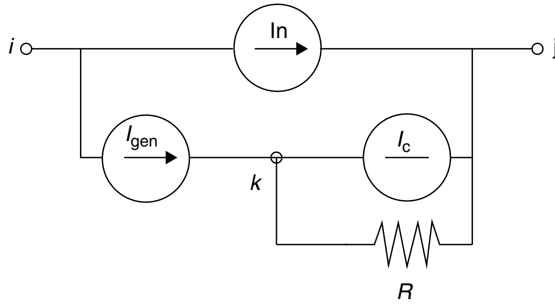


Figure 11.21 Circuit-level model for a diffusion resistor that includes avalanche breakdown and velocity saturation effects

regime. The circuit schematic in Figure 11.21 shows three current sources and a linear resistor, which model the high-current effects in the diffusion resistor. The extra node (k) in the model corresponds to the anode edge of the drift region. After snapback, device simulations show that the electric field drops at the cathode end of the resistor. The potential contours are concentrated around the anode junction while the potential remains constant throughout the drift region. As the potential is fixed in the drift region and, therefore, across the cathode junction, the injection of electrons at the forward-biased cathode junction is governed by [Sze91][Shur90]

$$np \approx n_i^2 \exp(qV_{kj}/kT) \tag{11.38}$$

where V_{kj} is the potential drop across the forward-biased cathode junction. Hence, the excess electron concentration, n, can be obtained from charge neutrality as

$$n = \frac{N_D}{2}\left(\sqrt{\left(1 + \frac{4n_i^2}{N_D^2}\exp(qV_{kj}/kT)\right)} - 1\right) \tag{11.39}$$

Under low-level injection conditions this is approximately

$$n = \frac{n_i^2}{N_D} \tag{11.40}$$

while under high-injection conditions this reduces to

$$n = n_i \exp(qV_{kj}/2kT) \tag{11.41}$$

These excess electrons drift towards the anode and are included in the current source, I_n, which can be written as

$$I_n = \frac{q(n + N_d)WX_jV_{ij}}{L\left(\frac{1}{\mu_n} + \frac{V_{ij}}{v_{sat}L}\right)} \tag{11.42}$$

The generation of carriers at the anode junction can be modeled similarly to the MOS snapback model as

$$I_{\text{gen}} = \theta(M - 1)(I_n) \tag{11.43}$$

where θ is a parameter that controls the extent to which the drift current I_n contributes to the generation of additional carriers. The multiplication factor is modeled similarly to the MOS snapback model as

$$M = \frac{1}{1 - A_{\text{i}} \exp\left(\frac{-B_{\text{i}}}{V_{\text{ik}}}\right)} \tag{11.44}$$

where A_{i} and B_{i} are technology-dependent parameters, which control the breakdown voltage of the anode junction. The final current source in the model is I_{c}, which is the excess impact ionization current, which flows from the anode to the cathode. This current is a fraction of the generation current and is given by

$$I_{\text{c}} = \delta I_{\text{gen}} \tag{11.45}$$

The parameters θ and δ, and the resistor R control the snapback characteristics and can be tuned to fit experimental data. Typically $0 < \theta, \delta < 1$ and the resistor R can be approximated by the low-field resistance, which is given by

$$R = \frac{L}{q(n\mu_n + p\mu_p)WX_{\text{j}}} \tag{11.46}$$

Figure 11.22 compares the simulated $I-V$ curve with those obtained from pulsed $I-V$ data and two-dimensional device simulations.

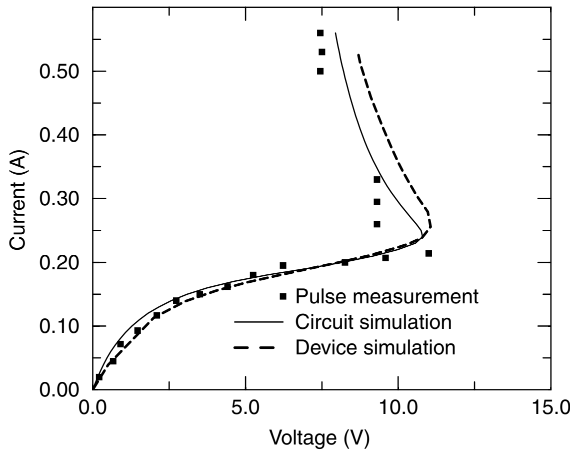


Figure 11.22 $I-V$ curve of a diffused n-well resistor with a saturation current of around 200 mA

11.4.3 Scalability Issues

As with any SPICE model it is important that the model be scalable. While the aforesaid model gives first-order scalability it is important that it be validated with silicon data. Scalability issues related to resistor modeling are discussed in detail in [Puvvada00]. The authors describe a scalable analytical model for the resistor valid up to the snapback point. They also describe a simple procedure for extracting the critical parameters that model the linear, saturation, and avalanche multiplication regions. The scalability of the extracted parameters with respect to variations in length, width, and doping was studied extensively using device simulations and then verified against experimental data. The low-field resistance is more accurately modeled as

$$R = \frac{L}{q(n\mu_n + p\mu_p)WX_{\mathrm{j}}} = \frac{m[L+a]}{[W+b]} \tag{11.47}$$

where m, a, and b are scaling parameters.

The equivalent circuit model of an n-well resistor consists of the actual resistor between the n^+ moats and the two parasitic diodes that are formed by the n-well and p^+ substrate junction. In some cases, these reverse biased diodes could further constrict the current flow leading to an increase in the low-field resistance. In addition, for very high reverse voltages and low series resistance to the p^+ contacts, the breakdown of these diodes could lead to a large increase in current. However, both these effects have a small impact in prevalent CMOS technologies since the n-well is much higher doped than the p substrate and is relatively deep.

11.5 MODELING PROTECTION DIODES

As described in the MOS model, the high-level injection operation of the diodes needs to be accurately modeled, for accurate ESD simulation. A forward-biased diode or MOS junction modulates the substrate in its vicinity due to minority carrier injection. On the other hand, reverse-biased junctions generate majority carriers in the substrate that at high fields experience velocity saturation. Hence, the direction of current flow determines the effective substrate resistance. Therefore, a simple circuit-level model for the diode is shown in Figure 11.23. The model for the variable resistor has been described in the substrate resistance modeling section. Any standard diode model that includes avalanche breakdown may be used. However, care must be taken not to include the built-in series resistance in the model. This is a fixed resistance, which could lead to an overestimation of the voltage dropped across the diode under high-current injection. Figure 11.7 shows the simulated $I-V$ curve of a diode under forward and reverse bias compared with measured data.

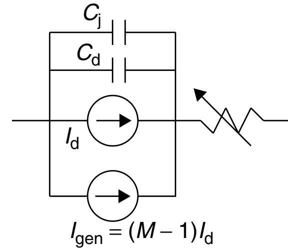


Figure 11.23 Circuit-level model for a protection diode including the diode current (Id), the impact-ionization current (Igen), the junction capacitance (cj), and diffusion capacitance (cd). The series resistor is modeled as a voltage or current dependent variable resistor depending on the direction of current flow

11.6 SIMULATION OF PROTECTION CIRCUITS

In Chapter 7, the STNMOS (gate-coupled nMOS with vertical *pnp* pump) and the NTNMOS (gate-coupled nMOS with nMOS pump) were described for circuit protection applications. In this section, we examine simulation examples using these devices.

In the first example, circuit simulations were used to analyze the protection circuit in Figure 11.24 subject to an ESD stress. The gain of the *pnp* is low at high current levels and most of the stress current appears as the base current to charge C_c. The simulation results for a 500 mA ESD stress current are shown in Figure 11.26 [Ramaswamy96B]. A sample pulse measurement is shown in Figure 11.25 which clearly indicates the linear charge-up phase ($t_c \approx 1.5$ μs), which is observed from simulations. The complex interactions between various protection elements

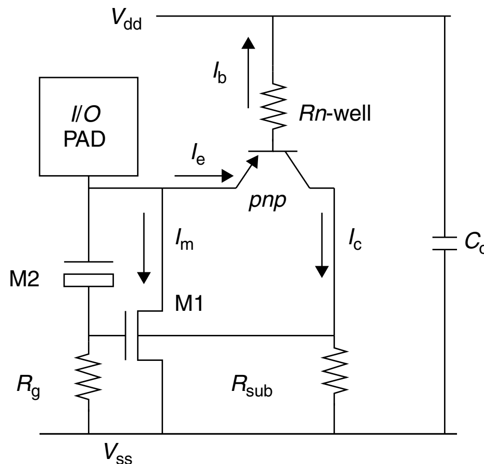


Figure 11.24 Circuit schematic of the I/O protection circuit showing the lateral *pn* diode to V_{dd} and chip capacitance between V_{dd} and V_{ss}. Also shown are the currents through the MOSFET (I_m) and *pnp* devices (I_e, I_b and I_c)

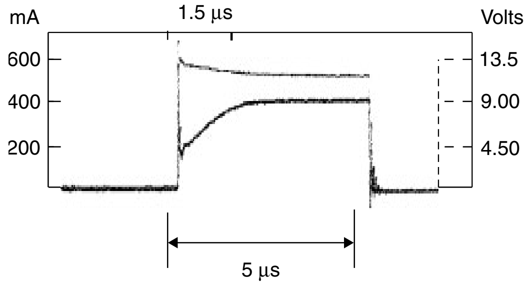


Figure 11.25 Measured current (upper) and voltage (lower) waveforms for the I/O protection circuit under a 5-μs current stress of 500 mA. Note that the charge-up time (t_C) is almost 1.5 μs

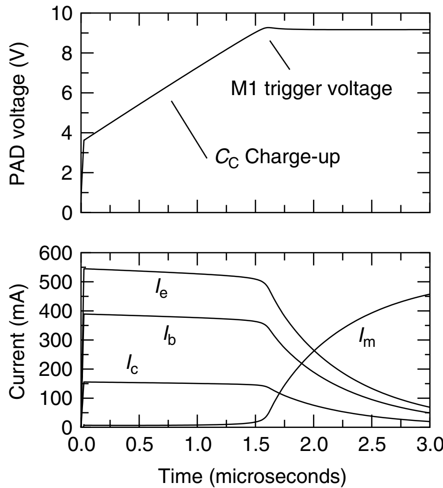


Figure 11.26 Circuit-level simulation for 500-mA stress current with a 100 nF chip capacitance. Note that the stress current is conducted by the *pnp* device for over 1.5 μs (I_e, I_b and I_c), while *M1* conducts the current after 3 μs (I_m)

can clearly be observed using a simulator that could allow the designer to further optimize the performance of the circuit.

In the second example, the importance of the layout parasitics including the substrate resistance on the ESD performance of an I/O cell in an ASIC library, is illustrated. The schematics for similar protection devices along with their layouts are shown in Figures 11.27 and 11.28. The main difference between them is in the layout of the gate-coupled MOS device (GCD). In the first case the GCD is placed in banks, with stripes of substrate taps connected to pump devices separating the individual banks. The second case has two pumps hooked to a floating substrate tap with the GCD device in a single bank in the center. As seen from

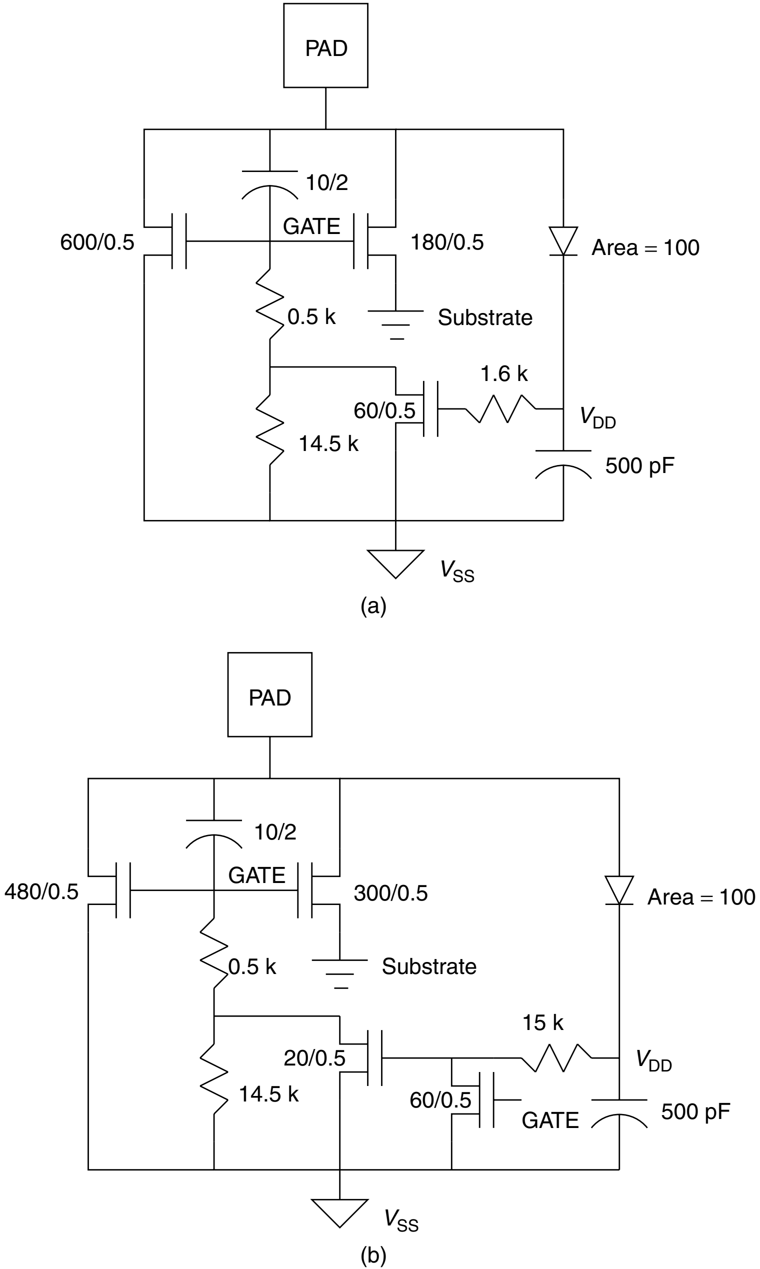


Figure 11.27 Schematics for two similar protection circuits

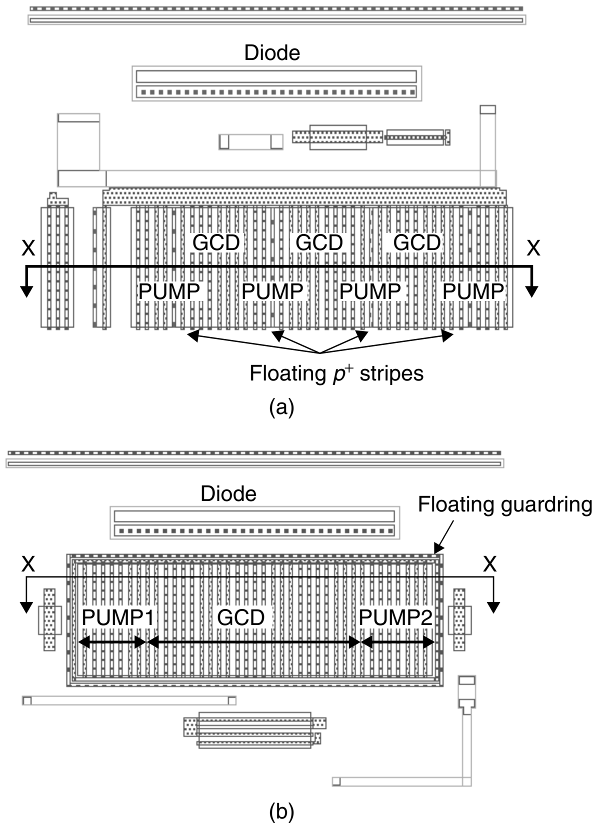
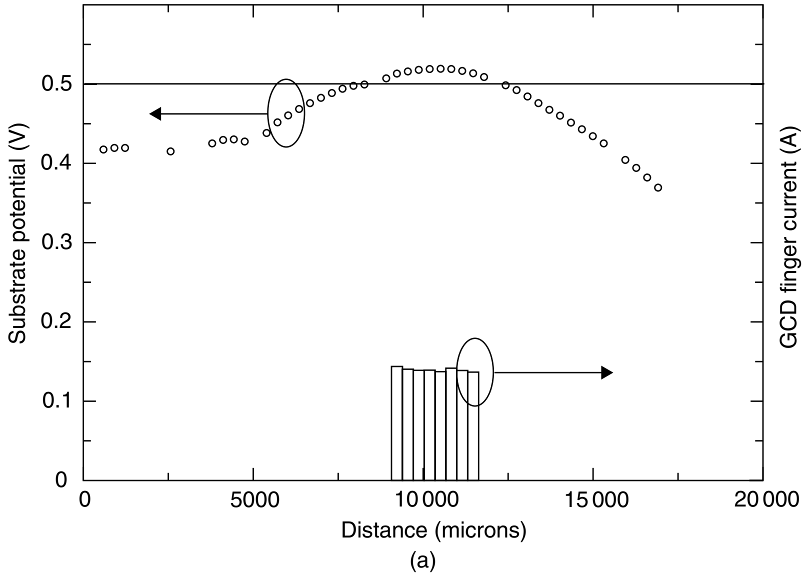


Figure 11.28 Two different layouts for the protection circuits in Figure 11.27. The substrate potential and current through the GCD fingers are plotted in Figure 11.29 along the line (X-X) shown in the two layouts

the simulation results in Figure 11.29 the second layout has a much better ESD performance. This is because the substrate potential is more uniformly raised in this case leading to conduction of ESD current through all the GCD fingers. In the stripe layout, however, owing to the large variation in substrate potential only one of the banks turns on leading to a much lower ESD performance. This example clearly demonstrates the need to comprehend the three-dimensional aspect of the substrate resistance and the multifinger turn-on of the ESD protection device for accurate simulations.

The impact of interconnect parasitics on the substrate-pumped GCD is comprehended in the next example. A simple schematic, which indicates the critical parasitic elements in addition to the primary protection elements is shown in Figure 11.30. The circuit is placed on two different substrates, which have different resistivity profiles shown in Figure 11.31. Figure 11.32 shows the impact of bus



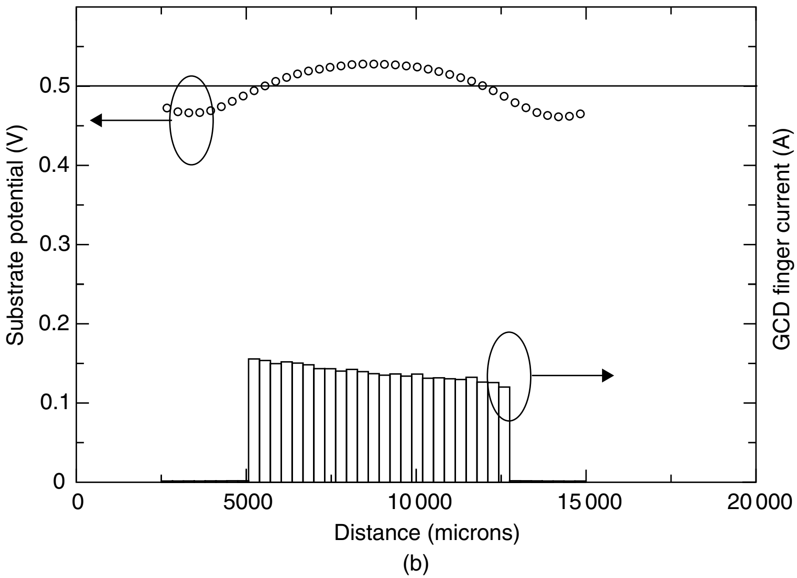


Figure 11.29 Simulation results showing the substrate potential variation and current through the GCD fingers for the two layouts. In case (a) the substrate potential prior to device turn-on shows a lot more variability than case (b). This causes a limited number of fingers to turn on in (a) while all GCD fingers turn-on in (b)

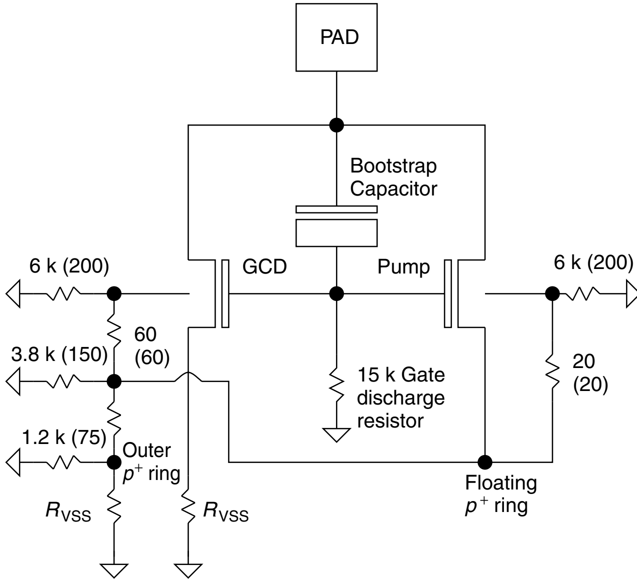


Figure 11.30 Equivalent schematic of the protection circuit used in an ASIC I/O cell. The various parasitic resistors for non-epi and epi(in brackets) substrates are indicated that need to be accounted for accurate simulation results. RVSS is the metal bus resistance connecting the source of the GCD and the outer p^+ guardring to the actual VSS pad

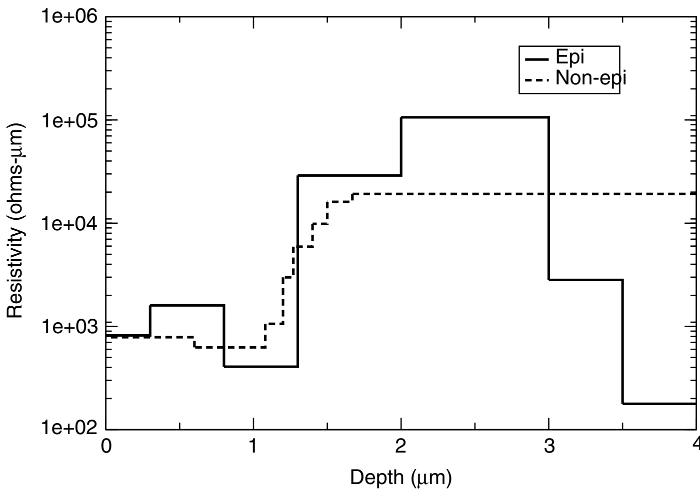


Figure 11.31 Discretized resistivity profiles for the two substrates (non-epi dashed line and epi solid line), which are used to calculate the parasitic substrate resistances in the ASIC I/O cell

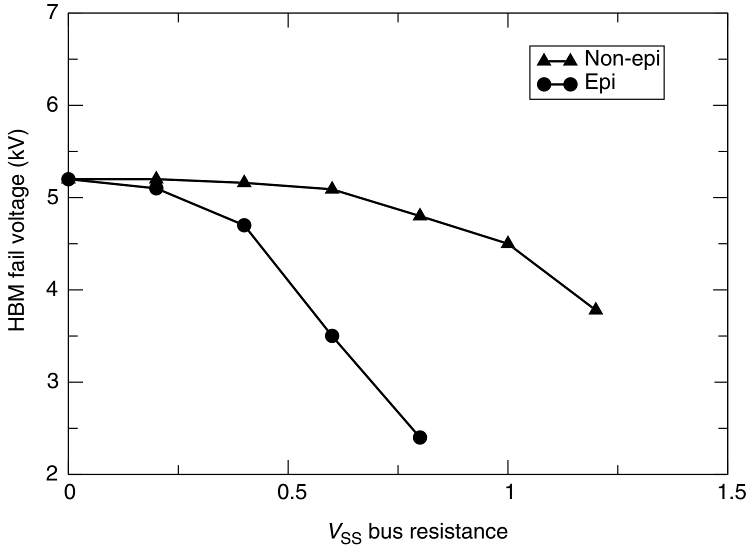


Figure 11.32 HBM-ESD performance results for the cell as a function of increasing V_{SS} bus resistance for the two substrates

resistance on the HBM performance of this cell. As can be seen the same cell has very different HBM performance levels depending on the type of substrate used. The HBM performance for the device in the epi substrate is much poorer especially for a higher bus resistance. This is because the effective substrate resistance for self-biasing the parasitic bipolar in the protection element is lower. Therefore, it is more susceptible to debiasing effects caused by increased bus resistance [Duvvury00]. Based on this example, a designer can use simulations to generate safe operating conditions for the ASIC cell in different environments.

11.7 ELECTROTHERMAL CIRCUIT SIMULATIONS

In the previous section, the usefulness of electrical simulations in analyzing and designing I/O protection circuits was demonstrated. In Chapter 10, it was shown that the onset of second breakdown was dependent on the rate of change of the avalanche and thermal generation currents with temperature. Hence, the coupling between the temperature and the current densities, impact ionization coefficients, mobilities, and electric fields are important in simulating ESD phenomena in devices. However, this increases the complexity of the simulations, resulting in longer computation times and increasing the difficulty in obtaining convergent solutions. Therefore, different approximations have been introduced to reduce the simulation complexity. Some of these approximations include: (1) calculation of

diffusion coefficients under isothermal conditions; (2) using fixed thermal parameters calculated at an elevated temperature typically around 900 K; and (3) assuming that the heat is dissipated in a semi-infinite medium. Electrothermal simulations of simple semiconductor structures have shown reasonable agreement with experimentally observed behavior. The main benefit from electrothermal simulators is obtained from studying the physical behavior of devices under high current stress. They can also be used to study in more detail the impact of process changes on ESD performance. However, for more complex circuits, it is impractical to use electrothermal simulations and one must resort to isothermal electrical simulations described earlier. Unlike electrothermal simulators, however, the failure criteria in electrical simulators is defined simply by the maximum current carrying capacity of various protection devices obtained from experimental data.

In this section, we outline the techniques used for circuit-level electrothermal simulation of EOS/ESD protection devices. We present the thermal models for describing second breakdown in the different protection devices discussed in Chapter 4. Previous work on electrothermal simulation using network analysis techniques has been limited in scope owing to the lack of avalanche breakdown modeling capability and models to efficiently describe the temperature dynamics. Bryant and Latif [Bryant82] used an electrical network analog to solve the discretized three-dimensional heat diffusion equation. This network was solved simultaneously with the temperature-dependent electrical characteristics of a power bipolar transistor using network analysis or relaxation methods. This approach, however, is not useful in analyzing devices subjected to ESD, which involve multiple devices and also steep temperature gradients ($>200\,\mathrm{K\,\mu m^{-1}}$). Scott *et al.* [Scott86] used a lumped element circuit model of MOS devices using cubic splines to simulate the avalanche breakdown phenomena under ESD. The steady state temperature distribution was calculated using a Green's function approach. While this model provides a good theoretical understanding of the effect of silicidation on the current distribution and steady state heating in the device, it is not applicable to transient electrothermal simulation. Beltman *et al.* [Beltman90] developed a circuit-level simulator to predict thermal runaway failures in BiCMOS circuits. The electrical and thermal characteristics were fully coupled to analyze the ESD susceptibility of protection diodes. A quasi-Gaussian distribution model of the heat source was implemented using a weighted resistor network. In addition, a temperature-dependent model for both the thermal conductivity and the specific heat of silicon was implemented. The scheme was successfully used to study the thermal response from a hot-spot during an ESD event.

A physically rigorous treatment of device modeling, which allows for self-heating due to coupled electrical and thermal effects under both steady state and transient conditions was presented by Wachutka [Wachutka90]. It has been shown that at second breakdown a device suffers permanent damage [Shafft67]. The problem of thermal breakdown in semiconductor devices was solved by Dwyer *et al.* [Dwyer90] using a Green's function formalism. The heat diffusion equation was solved for the specific case of a rectangular box (which is used to model the defect

site) subject to a constant input power. This model was successfully applied to the thermal breakdown in GaAs MESFETs caused by EOS [Franklin90]. This model was extended by Amerasekera *et al.* [Amerasekera91] to model second breakdown in silicon nMOS transistors subjected to ESD-type stresses.

To simulate the thermal breakdown of devices under EOS/ESD, an accurate description of temperature-dependent device behavior including breakdown is necessary. By modeling device behavior up to the onset of second breakdown, we can determine the safe operating limits for a given protection circuit. As second breakdown is thermally originated, its model should be based on the solution of the heat diffusion equation. The heat diffusion equation is given by

$$\frac{\partial T}{\partial t} - \frac{\kappa}{\rho C_p}\nabla^2 T = \frac{P(t)}{\Delta\rho C_p} = \frac{I(t)V_{\mathrm{H}}}{\Delta\rho C_p} \tag{11.48}$$

where T is the temperature, κ is the thermal conductivity, ρ is the density and C_p is the specific heat of the material (silicon). $P(t)$ is the power generated by a source in a volume Δ of the device, $I(t)$ is the current flowing through the device and V_{H} is the holding (breakdown) voltage of the device.

The temperature distribution in the vicinity of the heat source is obtained by solving the heat diffusion equation. For a rectangular box with volume $\Delta = abc$ the solution of the heat transfer equation yields [Dwyer90]

$$T(\vec{r}, t) = T_{\mathrm{o}} + \frac{\zeta}{\rho C_p \Delta}\int_0^t P(\tau)G(x, a, \tau)G(y, b, \tau)G(z, c, \tau)\, d\tau \tag{11.49}$$

where

$$G(x, a, \tau) \triangleq \frac{1}{2}\left\{\mathrm{erf}\left(\frac{a/2 + x}{2\sqrt{\tau\kappa/\rho C_p}}\right) + \mathrm{erf}\left(\frac{a/2 - x}{2\sqrt{\tau\kappa/\rho C_p}}\right)\right\} \tag{11.50}$$

and the factor $\zeta = 2$ for a semi-infinite medium (assuming the passivation layer is a perfect insulator). The integral over time in Equation (11.49) is evaluated in a circuit simulator using an electrical equivalent integrator circuit shown in Figure 11.33 [Diaz94A]. In this circuit, two special elements, a power monitor (P) and a time-dependent resistor (R), were introduced to convert the power dissipated in the device to a differential temperature (with respect to the ambient temperature T_{o}). The time-dependent resistor can be obtained from (11.49) and is given by

$$R(x, y, z, t) = \frac{\rho C_p \Delta}{\zeta C G(x, a, t)G(y, b, t)G(z, c, t)} \tag{11.51}$$

The capacitor C (typically around $1\,\mu\mathrm{F}$) is chosen so that the matrix entries in the linear system of equations are of the same order of magnitude. The thermal conductivity and the ρC_p product are temperature dependent and are given by [Selbeherr84]

$$\kappa = 1.5486\left(\frac{T}{300}\right)^{-4/3} [W\,\mathrm{cm}^{-1} - {}^\circ K]$$

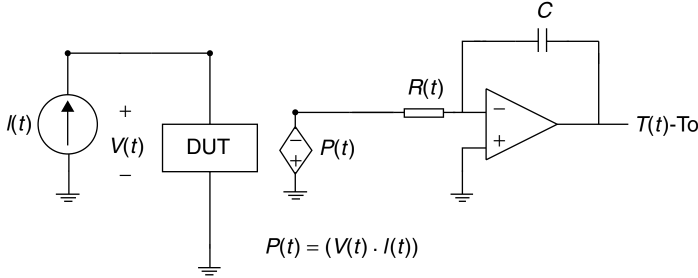


Figure 11.33 The integrator circuit including the special circuit elements that implements the closed-form solution of the three-dimensional heat diffusion equation

$$\rho C_p = 1.574 \left(\frac{T}{300}\right)^{0.1} [J \text{ cm}^{-3} - {}^\circ K]. \tag{11.52}$$

In silicon ICs for devices under EOS/ESD stress, a suitable choice of the thermal parameters can be obtained from Equation (11.52) by using a fixed temperature in the range $800 < T < 1000$ [Amerasekera91][Ash83][Diaz94B].

To simulate the complete coupling between various heat sources in a protection device (e.g., the drain junctions in a uniformly conducting nMOS output buffer), the current summation property at the integrator input is utilized.

11.7.1 MOSFET Electrothermal Model

Second breakdown occurs in an nMOS device when the thermal generation current increases significantly. As the temperature at the drain junction rises, the thermal generation current I_{th} increases. As I_{th} increases, the impact ionization generation current (I_{gen}) decreases because a relatively constant current is needed to sustain the bipolar action. It has been shown that M is a very weak function of temperature and decreases slightly with increasing temperature [Ramaswamy96B]. As I_{gen} is a rapidly varying function of the drain bias through M, a slight reduction in I_{gen} leads to a sharp drop in the drain voltage, which signals second breakdown in the nMOS device [Amerasekera94].

Figure 11.34 shows the three-dimensional and sectional views of a single finger MOS device. For an MOS device, the heat source dimensions are approximated by $a = W$, $b = b_o \ln \frac{V_H \mu_{eff}}{b_o v_{sat}}$ and $c = x_j$ where $b_o = \sqrt{\frac{\epsilon_{si} t_{ox} x_j}{\epsilon_{ox}}}$ [Ko89], v_{sat} is the electron saturation velocity, V_H is the holding voltage, μ_{eff} is the effective carrier mobility, x_j is the junction depth, t_{ox} is the gate oxide thickness and ϵ_{si}, ϵ_{ox} are the dielectric constants of silicon and silicon dioxide.

An additional current source, I_{th}, is added between the drain and internal base node for electrothermal simulations. The temperature at the drain junction (TDJ) controls I_{th}, while the source junction temperature (TSJ) affects the saturation

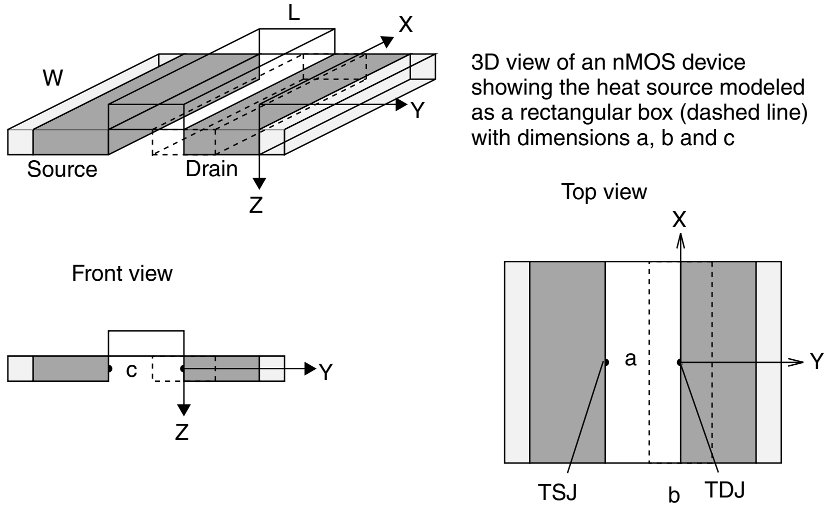


Figure 11.34 Three-dimensional and sectional views of a single finger MOSFET

current I_s (see Section 11.2). The temperatures TDJ and TSJ are obtained using the integrator circuit described in the previous section. These temperatures are used to evaluate the parameters of the parasitic bipolar device at each step of the transient simulation. The thermal generation current depends on the drain junction temperature as

$$I_{th}(T) = I_{th}(T_o)\left(\frac{T}{T_o}\right)^{3+\gamma/2} \exp\left(\frac{qE_g(T_o)}{kT_o} - \frac{qE_g(T)}{kT}\right)$$

where γ is a constant [Sze91] and T is the temperature (in K). The temperature dependence of the band gap energy is given by [Sze91][TMA93]

$$E_g = E_g(0) - \frac{E_{g\alpha}T^2}{T + E_{g\beta}} \tag{11.53}$$

where $E_{g\alpha}$ and $E_{g\beta}$ are material constants and T is the temperature (in °K) [Ramaswamy96]. The following equations describe the dependencies of the collector and base currents on the source temperature

$$I_c = \frac{\beta I_S(T_{SJ})}{Q_B}\left[\exp\left(\frac{qV_{BS}}{kT_{SJ}}\right) - \exp\left(\frac{qV_{BD}}{kT_{DJ}}\right)\right] \tag{11.54}$$

$$I_b = I_S(T_{SJ})\left[\exp\left(\frac{qV_{BS}}{kT_{SJ}}\right) - 1\right] \tag{11.55}$$

The saturation currents have the same temperature dependency as the thermal generation current but are controlled by the source junction temperature.

As an example of electrothermal simulations, we consider the simulation of a multifinger gate-coupled nMOS device using a fully coupled electrothermal simulation network. The device has eight fingers and the gate is coupled high during the ESD event to ensure uniform finger conduction (see Figure 11.35). As seen from the transient simulation results in Figure 11.36(a), all the fingers conduct the stress equally for the first 200 ns. However, as the temperature rises, full coupling across fingers results in the center two fingers becoming hotter than the outer six, as shown in Figure 11.36(b). This nonuniform temperature distribution across the fingers causes a redistribution of the current after around 200 ns. The thermal generation current in the central fingers increases and carries most of the stress current. This causes a further increase in the temperature of the central fingers and eventually leads to the thermal failure of the center fingers at around 400 ns.

This section described the electrothermal model for an nMOS transistor that can be used for circuit-level simulations. Similar models have been developed for silicon-on-insulator (SOI) nMOS devices and the reader is referred to [Raha97] for further details. In addition, the basic concepts described in this chapter can be used to develop electrothermal models for other circuit elements [Ramaswamy96B].

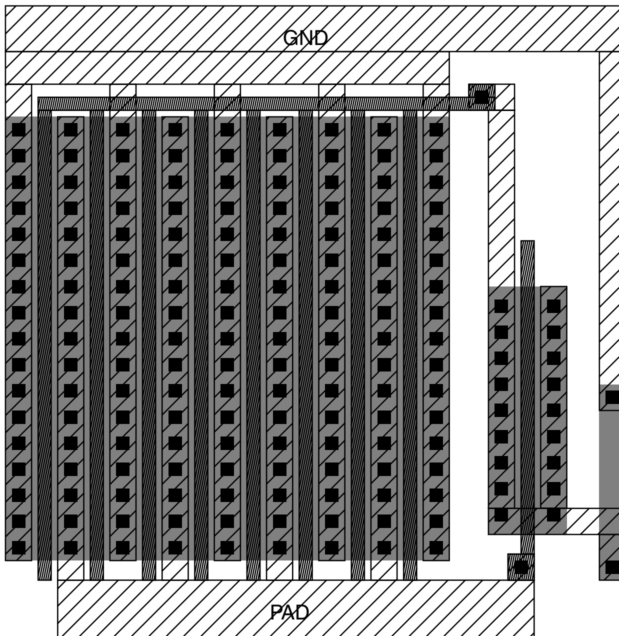
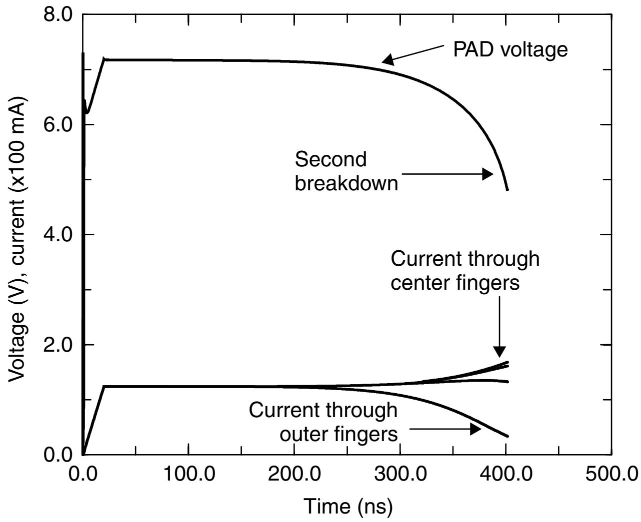


Figure 11.35 Layout of a multifinger gate-coupled nMOS protection device. The device has eight fingers each 20 μm wide. Also shown are the booting capacitor and gate discharge n-well resistor



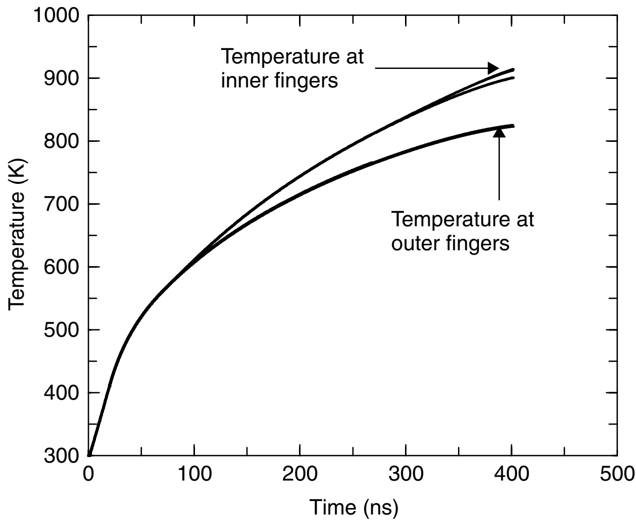


Figure 11.36 (a) Transient response of the multifinger nMOS device under a 1 A EOS stress. (b) Transient temperature waveforms at the center of each finger. The temperature coupling between adjacent fingers leads to current crowding in the center fingers after 200 ns

11.8 CONCLUSION

The basics of circuit simulation for ESD were described in this chapter. Circuit-level models for protection devices were discussed in detail. The high-current behavior of MOS devices and equations modeling the behavior of the parasitic bipolar device were presented. The importance of simple parameter extraction procedures and scalability of the model were emphasized. The Gummel-Poon model was shown to be adequate in describing the normal operation of the bipolar device. Extensions to the model to include avalanche breakdown of the junctions were reviewed. A new model for the diffusion resistor that incorporates the current saturation, avalanching, and snapback effects was presented. Scalability issues related to the model parameters were discussed. While most of the circuit models described are targeted towards advanced CMOS technologies, the basic concepts apply to other technologies as well.

Simulation examples were provided to show the usefulness of circuit simulations in designing protection circuits. Specific examples highlighting the importance of interconnect and substrate parasitics were presented. Finally, a discussion on electrothermal simulations with regard to their usefulness and their limitations, was presented. Circuit-level electrothermal simulation techniques were described and a simple example showing the thermal failure of a multifinger MOS device was illustrated.

It is well known that the circuit behavior is dictated by the environment in which it resides on a chip. It is important to include the parasitics associated with the interconnects while simulating protection circuits. While it is difficult, if not impossible, to simulate all the protection circuits on a chip simultaneously, one can use hierarchical modeling techniques as described in [Beebe98][Lee00]. The approach is certainly of value until more advanced and specialized tools are available for a full-chip simulation of ESD events.

BIBLIOGRAPHY

[Amerasekera91] A. Amerasekera, L.V. Roozendaal, J. Bruines and F. Kuper, "Characterization and modeling of second breakdown in nMOST's for the extraction of ESD-related process and design parameters", *IEEE T. Electron Dev.*, **38**, 2161–2168, 1991.

[Amerasekera94] A. Amerasekera and J. Seitchik, "Electrothermal behavior of deep submicron nMOS transistors under high current snapback conditions", *Tech. Dig. IEDM*, 455–458, 1994.

[Antognetti88] P. Antognetti and G. Massobrio, *Semiconductor Device Modeling with SPICE*, McGraw Hill, New York, 1988.

[Ash83] M. Ash, "Semiconductor junction non-linear failure power thresholds: Wunsch-Bell revisited", *EOS/ESD Symposium Proceedings*, pp. 122–127, 1983.

[Beebe98] S.G. Beebe, "Simulation of Complete CMOS I/O Circuit Response to CDM Stress", *EOS/ESD Symposium Proceedings*, pp. 259–270, 1998.

[Beltman90] R.A.M. Beltman, H. van der Vlist and A.J. Mouthaan, "Simulation of Thermal Runaway During ESD Events", *EOS/ESD Symposium Proceedings*, pp. 157–161, 1990.

[Boselli01] G. Boselli, S. Ramaswamy, A. Amerasekera, T. Mouthaan and F. Kuper, "Modeling Substrate Diodes under Ultra High ESD Injection Conditions", *EOS/ESD Symposium Proceedings*, pp. 71–81, 2001.

[Bryant82] P. Bryant and M. Latif, "Network analysis approach to multidimensional modeling of transistors including thermal effects", *IEEE Transactions on CAD*, **1**(4), 94–100, 1982.

[Chatterjee91] A. Chatterjee, T. Polgreen and A. Amerasekera, "Design and simulation of a 4 kV ESD protection circuit for 0.8 μm BiCMOS process", *Tech. Dig. IEDM*, 913–916, 1991.

[Chen88] K. Chen, "The effects of interconnect process and snapback voltage on the ESD failure threshold of nMOS transistors", *EOS/ESD Symposium Proceedings*, pp. 212–219, 1988.

[Chen96] J. Chen, X. Zhang, A. Amerasekera and T. Vrotsos, "Design and layout of a high ESD performance *npn* structure for submicron BiCMOS/bipolar circuits", *International Reliability Physics Symposium Proceedings*, pp. 227–232, 1996.

[Chynoweth60] A. Chynoweth, "Uniform silicon $p-n$ junctions II, ionization rates for electrons", *J. Appl. Phys.*, **31**, 1161–1164, 1960.

[Dabral94] S. Dabral, R. Aslett and T. Maloney, "Core clamps for low voltage technologies", *EOS/ESD Symposium Proceedings*, pp. 141–149, 1994.

[Diaz92] C. Diaz and S.-M. Kang, "New algorithms for circuit simulation of device breakdown", *IEEE Transactions on CAD*, **11**(11), 1344–1354, 1992.

[Diaz93B] C. Diaz, C. Duvvury and S.-M. Kang, "Electrothermal simulation of electrical overstress in advanced nMOS ESD I/O protection circuits", *Tech. Dig. IEDM*, 899–902, 1993.

[Diaz93C] C. Diaz, *Modeling and Simulation of Electrical Overstress Failures in Input/Output Protection Devices of Integrated Circuits*, PhD Thesis, University of Illinois at Urbana-Champaign, 1993.

[Diaz94A] C. Diaz, S.-M. Kang and C. Duvvury, "Circuit-level electrothermal simulation of electrical overstress failures in advanced MOS I/O protection devices", *IEEE Transactions on CAD*, **13**(4), 482–493, 1994.

[Diaz94B] C. Diaz, S.-M. Kang and C. Duvvury, "Thermal failure simulation of electrical overstress in semiconductor devices", *IEEE T. Electron Dev.*, **41**(3), 1994.

[Dutton75] R. Dutton, "Bipolar transistor modeling of avalanche generation for computer circuit simulation", *IEEE T. Electron Dev.*, **ED-22**, 334–338, 1975.

[Duvvury92] C. Duvvury and C. Diaz, "Dynamic gate coupling of nMOS for efficient output ESD protection", *International Reliability Physics Symposium Proceedings*, pp. 141–150, 1992.

[Duvvury00] C. Duvvury, S. Ramaswamy, A. Amerasekera, R. Cline, B. Andresen and V. Gupta, "Substrate Pump nMOS for ESD Protection Applications", *EOS/ESD Symposium Proceedings*, pp. 7–17, 2000.

[Dwyer90] V. Dwyer, A. Franklin and D. Campbell, "Thermal failure in semiconductor devices", *J. Solid State Electr.*, **33**(5), 553–560, 1990.

[Franklin90] A. Franklin, V. Dwyer and D. Campbell, "Thermal breakdown in GaAs MES diodes", *J. Solid State Electr.*, **33**, 1055–1064, 1990.

[Getreu78] I. Getreu, *Modeling the Bipolar Transistors*, Elsevier, Amsterdam, 1978.

[Gharpurey96] R. Gharpurey and R.G. Meyer, "Modeling and analysis of substrate coupling in integrated circuits", *IEEE J. Solid-State Circ.*, **31**(3), 344–353 1996.

[Guedes88] M. Pinto-Guedes and P.C. Chan, "A circuit simulation model for bipolar-induced breakdown in MOSFET", *IEEE Transactions on CAD*, **7**(2), 289–294, 1988.

[Gummel70] H. Gummel and H. Poon, "An integral charge control model for bipolar transistors", *Bell System Technical Journal*, **49**, 827, 1970.

[Gupta98] V. Gupta, A. Amerasekera, S. Ramaswamy and A. Tsao, "ESD-related Process Effects in Mixed-Voltage Sub-0.5 μm Technologies", *EOS/ESD Symposium Proceedings*, pp. 161–169, 1998.

[Hebert87] F. Hebert and D. Roulston, "Modeling of narrow-base bipolar transistors including variable-base-charge and avalanche effects", *IEEE T. Electron Dev.*, **ED-34**, 2323–2327, 1987.

[Huang94] J. Huang, Z. Liu, M. Jeng, K. Hui, M. Chan, P. Ko and C. Hu, *BSIM3 manual (version 2)*, tech. rep., University of California, Berkeley, 1994.

[ICCAP00] Agilent Technologies, *IC-CAP 5.3 Nonlinear Device Models*, 2000.

[Ko89] P.-K. Ko, *Advanced MOS Device Physics*, ch. 1. Approaches to Scaling. Academic Press, San Francisco, 1989.

[Krakauer94] D. Krakauer and K. Mistry, "Circuit interactions during ESD", *EOS/ESD Symposium Proceedings*, pp. 113–119, 1994.

[Krieger89A] G. Krieger, "The dynamics of electrostatic discharge prior to bipolar action related snapback", *EOS/ESD Symposium Proceedings*, pp. 136–144, 1989.

[Krieger89B] G. Krieger and P. Niles, "Diffused resistors characteristics at high current density levels–analysis and applications", *IEEE T. Electron Dev.*, **ED-36**, 416–423, 1989.

[Kurimoto94] K. Kurimoto, K. Yamashita, I. Miyanaga, A. Hori and S. Odanaka, "An Electrothermal Circuit Simulation using An Equivalent Thermal Network for Electrostatic Discharge (ESD)", *Symposium on VLSI Technology Digest of technical papers*, pp. 127–128, 1994.

[Lee00] J. Lee, Y. Huh, J.-W. Chen, P. Bendix and S.-M. Kang, "Chip-level simulation for CDM failures in multi-power IC's", *EOS/ESD Symposium Proceedings*, 456–464, 2000.

[Li98] T. Li, C.-H. Tsai, E. Rosenbaum and S.-M. Kang, "Substrate Resistance Modeling and Circuit-Level Simulation of Parasitic Device Coupling Effects for CMOS I/O Circuits under ESD stress", *EOS/ESD Symposium Proceedings*, pp. 281–289, 1998.

[Lim97] S. Lim *et al.*, "Computationally Stable Quasi-Empirical Compact Model for the Simulation of MOS Breakdown in ESD-Protection Circuit Design", *Proceedings of the SISPAD conference*, pp. 161–164, 1997.

[Lindmeyer67] J. Lindmayer and W. Schneider, "Theory of lateral transistors", *J. Solid State Electr.*, **10**, 225–234, 1967.

[Liou90A] J. Liou and J. Yuan, "An avalanche multiplication model for bipolar transistors", *J. Solid State Electr.*, **33**(1), 35–38, 1990.

[Liou90B] J. Liou and J. Yuan, "Modeling the reverse base current phenomena due to avalanche effect in advanced bipolar transistors", *IEEE T. Electron Dev.*, **37**(10), 2274–2276, 1990.

[Lu89] P. Lu and T. Chen, "Collector-base junction avalanche effects in advanced double-polysilicon self-aligned bipolar transistors", *IEEE T. Electron Dev.*, **36**(6), 1182–1188, 1989.

[Luchies94] J. Luchies, C. de Kort and J. Verweij, "Fast turn-on of an nMOS ESD protection transistor; measurement and simulations", *EOS/ESD Symposium Proceedings*, pp. 266–272, 1994.

[Mergens00] M. Mergens, W. Wilkening, G. Kiesewetter, S. Mettler, H. Wolf, J. Hieber and W. Fichtner, "ESD-level Circuit Simulation - Impact of Gate RC-Delay on HBM and CDM Behavior", *EOS/ESD Symposium Proceedings*, pp. 446–455, 2000.

[Miller57] S. Miller, "Ionization rates for holes and electrons in silicon", *Phys. Rev. Lett.*, **105**, 1246–1249, 1957.

[Mistry90] K. Mistry, D. Krakauer and R. Titus, "Dependence of input ESD failure thresholds on IC design style", *EOS/ESD Symposium Proceedings*, pp. 214–218, 1990.

[Muller86] R. Muller and T. Kamins, *Device Electronics for Integrated Circuits*. John Wiley & Sons, New York, 1986.

[Nagel75] L. Nagel, *SPICE2 - A computer program to simulate semiconductor circuits*, tech. rep., University of California, Berkeley, ERL-M520, 1975.

[Puvvada00] V. Puvvada, V. Srinivasan and V. Gupta, "A Scalable Analytical Model For The ESD *N*-Well Resistor", *EOS/ESD Symposium Proceedings*, pp. 437–445, 2000.

[Raha97] P. Raha, S. Ramaswamy and E. Rosenbaum, "Heat flow analysis for EOS/ESD protection device design in SOI technology", *IEEE Trans. on Electron Dev.*, 464–471, 1997.

[Ramaswamy96] S. Ramaswamy and S.-M. Kang, *iETSIM Version 2.0 Users Manual*, tech. rep., University of Illinois at Urbana-Champaign, 1996.

[Ramaswamy96B] S. Ramaswamy, *Modeling Simulation and Design Guidelines for EOS/ESD Protection Circuits in CMOS Technologies*, PhD Thesis, University of Illinois at Urbana-Champaign, 1996.

[Ramaswamy97] S. Ramaswamy, A. Amerasekera and M. Chang, "A unified substrate current model for weak and strong impact ionization in sub-0.25 μm nMOS devices", *Tech. Dig. IEDM*, 885–888, 1997.

[Reisch92] M. Reisch, "On bistable behavior and open-base breakdown of bipolar transistors in the avalanche regime - modeling and applications", *IEEE T. Electron Dev.*, **ED-39**, 1398–1409, 1992.

[Russ96] C. Russ, K. Verhaege, K. Bock, P.J. Roussel, G. Groeseneken and H. Maes, "A Compact Model for the Grounded-Gate nMOS Behaviour under CDM ESD Stress", *EOS/ESD Symposium Proceedings*, pp. 302–315, 1996.

[Scott86] D. Scott, G. Giles and J. Hall, "A lumped element model for simulation of ESD failures in silicided devices", *EOS/ESD Symposium Proceedings*, pp. 41–47, 1986.

[Selbeherr84] S. Selbeherr, *Analysis and Simulation of Semiconductor Devices*. Springer-Verlag, New York, 1984.

[Severson86] F. Severson, "A new BV_{CEO} SPICE model", *Proc. 1986 IEEE Bipolar Circuits and Tech. Meet. (BCTM)*, 77–78, 1986.

[Shafft67] H. Schafft, "Second breakdown - a comprehensive review", *Proceedings of the IEEE*, **55**(8), 1272–1285, 1967.

[Shur90] M. Shur, *Physics of Semiconductor Devices*. Prentice-Hall, Englewood Cliffs, NJ, 1990.

[Sun78] E. Sun, J. Moll, J. Berger and B. Alders, "Breakdown mechanism in short-channel MOS transistors", *Tech. Dig. IEDM*, 478–481, 1978.

[Sze91] S. Sze, *Physics of Semiconductor Devices*. 2nd ed. John Wiley, New York, 1991.

[TMA93] MEDICI, *Two Dimensional Simulation Program*, Technology Modeling Associates, Inc., Palo Alto, California, 1993.

[Vandebroek91] S. Verdonckt-Vandebroek, S. Wong, J. Woo and P. Ko, “High gain lateral bipolar action in a MOSFET structure”, *IEEE T. Electron Dev.*, **ED-38**, 2487–2496, 1991.

[Verghese93] N.K. Verghese, D.J. Allstot and S. Masur “Rapid simulations of substrate coupling in mixed-mode IC’s”, *Proc. IEEE Custom Integrated Circuits Conference*, 18.3.1–18.3.4, May 1993.

[Wachutka90] G. Wachutka, “Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling”, *IEEE Transactions on CAD*, **9**(11), 1141–1149, November 1990.

[Wolf98] H. Wolf, H. Geiser and W. Stadler, “Bipolar model extension for MOS transistors considering gate coupling effects in the HBM ESD domain”, *EOS/ESD Symposium Proceedings*, 271–280, 1998.

[Worley95] E. Worley, R. Gupta, B. Jones, R. Kjar, C. Nguyen and M. Tennyson, “Sub-micron chip ESD protection schemes which avoid avalanching junctions”, *EOS/ESD Symposium Proceedings*, 13–20, 1995.

12 Conclusion

12.1 LONG-TERM RELEVANCE OF ESD IN ICs

ESD damage is directly responsible for approximately 10% of the total failure returns [Green88] [Wagner93], and the reported number may be higher but for the difficulty in distinguishing between some EOS and ESD failures. And in the new millennium, these issues continue to make it an important failure mechanism throughout the semiconductor industry. As technology feature sizes move into the deep submicron regime, concerns regarding ESD are increasing. Shallower junctions, very thin gate oxides, and small channel lengths associated with technology scaling will affect ESD performance. Earlier indications have been that the impact may be positive rather than negative [Lin93][Amerasekera94B]. More recent indications are that some ESD limits may be approaching as the technologies move towards the sub-0.25-µm regime [Bock99] where it was noted that for short channel nMOS transistors, the ESD robustness begins to significantly decrease. Other studies point to the gate oxide being more susceptible to damage [Amerasekera99] [Salman02]. Together with the smaller feature sizes, future generations of ICs will also consist of very high-density circuits with pad counts in excess of 1000. This will cause a severe crunch on the available area for ESD protection circuitry, requiring that the protection circuits become more efficient. One way out of this crunch is a trend to place the ESD protection devices underneath the bond pad [Anderson99].

The impact of ESD damage due to handling and testing can have a negative influence on product yield [Wagner93]. Complex, expensive IC's can eventually sell for anything between $300 to $3000 depending on the application and the availability of alternative products. Any product loss due to ESD damage has a direct impact on profitability and even fallouts on the order of 1% are not acceptable. There is strong motivation, therefore, to ensure that the present and future ICs have reasonable ESD levels to avoid damage during handling and testing. Another issue, which gives increasing importance to ESD, is the move towards replaceable ICs in electronic systems. Instead of replacing the whole circuit board, as used to be standard practice, users are encouraged to purchase upgrades to their microprocessors and memory cards and do the installation themselves. Since the installation does not necessarily take place in an ESD-safe environment, the ICs need to be

ESD robust. Moreover, with the explosion of the cell phones in the late 1990s, the ESD reliability at the system level has also become critical. For example, accumulated charges from human handling can discharge through a metallic object such as a screwdriver causing concern for the exposed pins on the phone. The same thing can be said about the chips that are used in printers. For these reasons a new standard such as the System Level Test described in Chapter 2 will be an important issue for the future applications.

The demand for ESD robustness has led to more consideration for ESD robustness during technology development and circuit design. At the same time ESD test methods are being better defined and correlated to real ESD events, thus improving the confidence levels in the ability of protection circuits to function as required.

12.2 STATE-OF-THE-ART FOR ESD PROTECTION

Since the mid-80s the advances in the process technologies have been making the design of ESD protection circuits very challenging. As shown in Figure 12.1, the impact of process has to be constantly overcome with novel protection designs or reoptimization of the known protection devices. For the standard field-oxide protection device (FOD), the LDD has had a major impact and combined with the silicided diffusions this device became totally ineffective. Following the development of the SCR, it is easy to see that the thin epi substrate had to be overcome with redesigned structure or the STI effect had to be improved with a blocking mask. Similarly, the

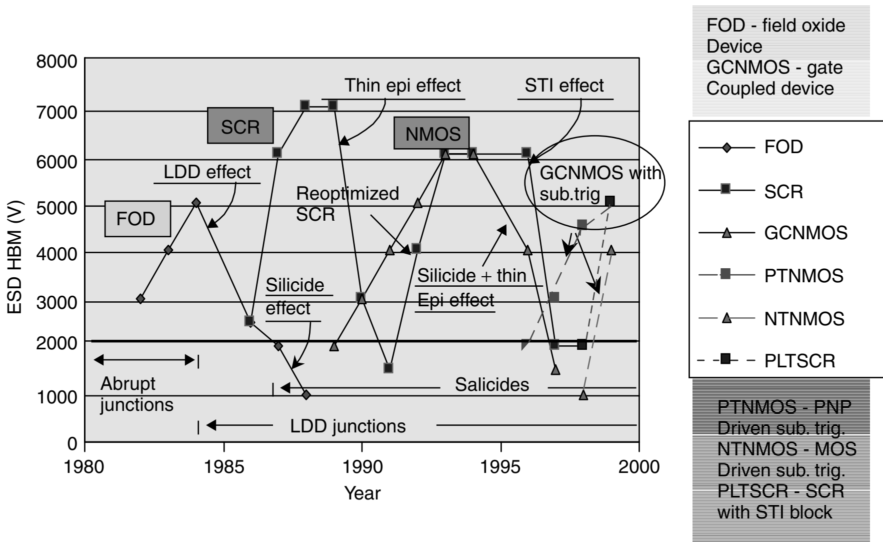


Figure 12.1 Evolution of ESD protection circuits

gate-coupled nMOS (GCNMOS) was initially effective but the combined silicide and thin epi substrate rendered to be a poor protection device. Substrate triggering with the PTNMOS or the NTNMOS improved the GCNMOS capability again. It is then reasonable to assume that newer phenomena might be in store for the next phase of the technologies. Some of these aspects are further discussed here.

The present ESD protection circuit elements for MOS processes are the nMOS devices designed as various types of efficient protection clamps, and PMOS devices used as high current MOS conduction devices. The bipolar *npn* devices are used for bipolar/BiCMOS processes.

The use of SCR structures has initially increased because of their lower power dissipation after they trigger, but the high trigger voltages have limited their application. Reducing the trigger voltages for both SCR as well as *npn* devices has been a major thrust in the development of ESD protection circuits with reasonable success. During the introduction of thin epi substrates and again with the introduction of Shallow Junction Isolation (STI), the SCRs have been absent for a while because of the difficulty in achieving their consistent trigger for protection designs. But more recently the SCRs are beginning to make a serious a comeback, the most common reason being the low associated capacitance with reasonable ESD levels.

The high chip capacitance in large ICs has increased the effectiveness of the diode protection circuit, which was popular in the early technologies. A diode between the pad and the positive power supply (V_{CC}) is usually supplemented by either an *npn* device or an SCR to the ground (V_{SS}) or negative supply. Diode circuits also require good protection between V_{CC} and ground, which can be an *npn* type circuit or an SCR depending on the application.

Large ICs have a large number of pin combinations, and consideration must be given to each of these when placing ESD protection circuits. In addition, in many cases multiple V_{CC} and V_{SS} buses may be used, which lead to added complications in terms of protection circuitry. In general, ESD protection circuits are placed between all the important combinations. This requirement makes the use of diode elements attractive because of their small areas. The diodes also offer minimum capacitance ESD protection, which makes it attractive for RF circuit applications.

As mentioned, the effect of technology on ESD protection circuit performance has been a major hindrance to achieving consistent ESD behavior. Advanced process development has begun to include ESD considerations in the technology roadmap. In some cases, process features have been implemented just to ensure consistent ESD performance.

12.3 CURRENT LIMITATIONS

The present approach to ESD protection circuit design and implementation is iterative. Circuits are designed and evaluated depending on the available area and the pin specifications. In many cases the ESD capability of the process has been previously characterized on test structures to enable basic design guidelines for the

protection circuits to be generated. However, in many cases the iterative approach involves too much time and can result in a delay in the release of a product to the market. Since a delay could be costly, the outcome is that ESD protection circuit performance may be sacrificed.

There is a need, therefore, to reduce the number of cycles required for the development of good ESD protection circuits. An important contribution in this direction would be the ability to use accurate simulation tools to evaluate the circuit and the technology before committing to silicon. Present generations of ICs are almost entirely designed using simulations, and the inability to include the ESD circuit into the simulation loop makes it more difficult to include ESD performance into the circuit design. The same is true for technology design. In recent years, there has been a significant thrust to use simulations for initial ESD designs [Fichtner01]. But even then, the lack of accurate simulation tools is one of the major reasons why ESD is considered to be almost a "black art" in the semiconductor industry.

Chapters 10 and 11 presented a summary of the state-of-the-art of modeling and simulating ESD events in semiconductors. It shows that there has been progress made towards the development of suitable ESD simulators. The two main obstacles to progress in the areas of circuit and technology simulation are the limited understanding of (a) the mechanisms governing the technology dependence of ESD sensitivity; and (b) the circuit interactions during an ESD event.

A further limitation has been the inability to automatically check a full-circuit design with regard to whether it follows all the ESD design rules. Many ESD problems in ICs are due to errors in the layout, which lead to a low ESD damage threshold although the protection circuit is well-designed and can pass the required ESD levels itself. ESD design rule checkers, which are invaluable to search for and identify transistors, which would be directly in the path of an ESD current, stress and do not have adequate ESD protection, have been introduced but their widespread use is still very limited [Sinha98] [Ngan01]. There is further need to improve both the checkers and the simulators to detect parasitic current paths. For example, they would need to ensure that unrelated diffusions would not trigger parasitic *npn* or SCR devices during an ESD event.

The present range of protection circuits all depend on the triggering of parasitic elements in order to provide suitable levels of ESD protection. Although these parasitic elements have successfully enabled very high ESD levels to be reached, it is difficult to ensure consistent triggering of these devices under all ESD conditions. At present this is not a major limitation, but with the advent of the CDM test method it is foreseeable that circuits, which consistently work under a range of stresses, would be needed. Such circuits may very well be active circuits whose operation is not based on the triggering of parasitic devices.

On a different note, almost all ICs, with the possible exception of a few high voltage devices used in automotive applications, use on-chip ESD protection techniques. As ICs become even larger and technologies become much smaller, there may be an advantage in moving towards off-chip ESD protection circuits [Cronin93] [Unger94] [Lin94]. One of the concerns in off-chip protection

techniques is that the dynamic impedance of the 'short' during a fast ESD event must be low [Lin94]. It is possible that even though the DC impedance is close to zero, the dynamic impedance during a CDM event can be large because of the high dI/dt. The governing parameter is the inductance of the wire or metalization used for the current shunt and this should be as small as possible.

12.4 FUTURE ISSUES

The current limitations listed in the previous section would form the basis of the main issues to be addressed as technologies move towards 0.10-μm feature sizes. It is essential to improve our understanding of the underlying mechanisms of the behavior of circuit elements during an ESD event. We would then be able to develop consistent predictive methods, which can be used to evaluate the impact of technology variations on ESD performance. In any case, ESD requirements should be part of the process development roadmap to ensure that good ESD performance is achieved in future generations of ICs.

A parallel action would be the development of circuit simulators, which incorporate the capability to extract the sensitivity to both HBM and CDM type ESD stress events. At present it is expected that such a simulator would require a large amount of memory and computing power because of the large number of parasitics involved. For example, in a large IC, each ESD event could select a different path between each of the power buses. These paths need to be accurately simulated to determine the chip layout factors influencing the ESD capability of these preferred paths or to channel the stress current through a dedicated ESD protection circuit.

The need for improved ESD design rule checkers is clear. Such a checker needs to be able to consistently identify ESD sensitive paths and violations of ESD design guidelines to be effective.

Attempts to develop new and improved ESD protection circuits are unlimited. Most of the present interest has been on better ways to triggering these protection devices at lower voltages. There was earlier a move towards using the path between the V_{CC} and V_{SS} buses as the primary protection in large ICs [Merrill93] [Voldman93] [Voldman94A] [Voldman94B] [Tandan94] [Croft94] [Dabral94]. This has now become more popular [Torres01]. As ICs become larger and more complex and on-chip area becomes more valuable, off-chip protection techniques may gain popularity. However, this would not eliminate the need to design robust circuits or technologies, but it will mean that large area on-chip protection circuits will not be used.

One area of protection circuit design, which will gain importance in the coming generations of Ics, is the protection of high speed telecom chips. Output drivers with bit rates exceeding 10 GBits/sec cannot tolerate much capacitive or resistive loading due to the ESD protection circuits. Some of the trade-offs involved between ESD protection and normal operation have been discussed in the earlier chapters.

However, there is a need for a protection circuit design methodology, which allows these high-speed circuits to be made self-protecting if possible. The alternative is to use off-chip protection techniques.

A second area of importance in protection circuit design for the present and future generations of ICs is the use of multivoltage supplies, that is, 3.3 V/5 V, 1.8V/3.3V/5 V, or 1.2 V/2.5 V/3.3 V, in these chips [Voldman94A] [Voldman94B]. The requirement that a circuit using elements designed for 3.3 V or 1.8 V operation is able to tolerate a higher voltage at the input or output pin means that in some cases it is not possible to use protection circuits where nMOS devices are the primary protection elements. New techniques are being developed for these ICs [Voldman94B] [Kunz01]. However, there are still many new challenges to be faced in the design of efficient and optimized ESD protection circuits for multivoltage applications.

Finally, the complexities and intricacies involved in ESD in silicon integrated circuits have generated interest in many academic and industrial research workers. A systematic approach towards understanding the fundamentals of the issues involved will eventually lead to solutions to some of the issues raised in this chapter. In the last 20 years, large strides have been made towards understanding and solving some of the most pressing ESD problems. There are still a number of issues, which, while making ESD in ICs an interesting and rewarding (if not sometimes frustrating!) field for research, need to be resolved if good, consistent ESD protection is to be designed for future generations of ICs.

BIBLIOGRAPHY

[Green88] T. Green, "A Review of EOS/ESD Field Failures in Military Equipment", in *Proceedings of the 10th EOS/ESD Symposium*, pp. 7–14, 1988.

[Wagner93] R.G. Wagner, J. Soden, C.F. Hawkins, "Extent and Cost of EOS/ESD Damage In An IC Manufacturing Process", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 49–55, 1993.

[Lin93] D.L. Lin, "ESD Sensitivity and VLSI Technology Trends: Thermal Breakdown and Dielectric Breakdown", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 73–82, 1993.

[Amerasekera94B] A. Amerasekera, C. Duvvury, "The Impact of technology Scaling on ESD Robustness and Protection Circuit Design", in *Proceedings of the 16th EOS/ESD Symposium*, pp. 237–245, 1994.

[Unger94] B. Unger, *private communication*, 1994.

[Cronin93] D. Cronin, "CRO-BAR: A New Technique For ESD Protection", *EMC Test And Design*, January 1993.

[Lin94] D. Lin, M.-C. Jon, "Off-Chip Protection: Shunting of ESD Current by Metal Traces on Circuit Packs", in *Proceedings of the 16th EOS/ESD Symposium*, pp. 279–285, 1994.

[Merrill93] R. Merrill, E. Issaq, "ESD Design Methodology", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 233–237, 1993.

[Voldman93] S. Voldman, V. Gross, "Scaling Optimization and Design Considerations of Electrostatic Discharge Protection Circuits in CMOS Technology", in *Proceedings of the 15th EOS/ESD Symposium*, pp. 251–260, 1993.

[Voldman94A] S. Voldman, "ESD Protection in a Multi-Rail Disconnected Power Grid And Mixed Voltage Interface Environments In 0.5 And 0.25 μm Channel Length CMOS Technologies", in *Proceedings of the 16th EOS/ESD Symposium*, pp. 125–134, 1994.

[Voldman94B] S. Voldman, G. Gerosa, "Mixed-voltage-interface ESD protection circuits for advanced microprocessors in shallow trench isolation and LOCOS CMOS technology", in *Tech. Dig. IEDM*, 1994.

[Tandan94] N. Tandan, G. Conner, "ESD Trigger Circuit", in *Proceedings of the 16th EOS/ESD Symposium*, pp. 120–124, 1994.

[Fichtner01] W. Fichtner, "TCAD-based ESD on-chip protection design", in *Tech. Dig. IEDM*, 2001.

[Bock99] K. Bock, B. Keppens, V. De Heyn, G. Groeseneken, L.Y. Chiang, and A. Naem, "Influence of Gate Length on ESD Performance for Deep Submicron CMOS Technology", in *Proceedings of the 21st EOS/ESD Symposium*, pp. 95–104, 1999.

[Anderson99] W. Anderson, W. Gonzalez, W. Fowler, and S. Knecht, "ESD Protection Under Wire Bonding Pads", in *Proceedings of the 21st EOS/ESD Symposium*, pp. 88–94, 1999.

[Kunz01] K. Kunz, C. Duvvury, and H. Shichijo, "5-V Tolerant Fail-Safe ESD Solutions for a 0.18 μm CMOS Process", in *Proceedings of the 23rd EOS/ESD Symposium*, pp. 12–21, 2001.

[Torres01] C.A. Torres, J. Miller, M. Stockinger, M.D. Akers, M.G. Khazhinsky, and J.C. Weldon, "Modular, Portable, and Easily Simulated ESD Protection Networks for Advanced CMOS Technologies", in *Proceedings of the 23rd EOS/ESD Symposium*, pp. 82–95, 2001.

[Ngan01] P. Ngan, D. Oliver, T. Smedes, R. Gramacy, and C-K. Wong, "A Automatic Layout Based Verification of Electrostatic Discharge Paths", in *Proceedings of the 23rd EOS/ESD Symposium*, pp. 96–101, 2001.

[Sinha98] S. Sinha, H. Swaminathan, G. Kadamati, and C. Duvvury, "An Automated Tool for Detecting ESD Design Errors", in *Proceedings of the 23rd EOS/ESD Symposium*, pp. 208–217, 2001.

[Amerasekera99] A. Amerasekera, V. Gupta, K. Vasanth and S. Ramaswamy, "Analysis of snapback behavior on the ESD capability of sub-0.20 μm NMOS", in *Proc. IRPS*, 159–166, 1999.

[Salman02] A. Salman, R. Gauthier, E. Wu, P. Riess, C. Putnam, M. Muhammad, M. Woo and D. Ioannou, "Electrostatic discharge induced oxide breakdown characterization in a 0.10 μm CMOS technology", in *Proc. IRPS*, 2002.

[Croft94] G. Croft, "ESD Protection Using a Variable Voltage Supply Clamp", in *Proceedings of the 16th EOS/ESD Symposium*, pp. 135–140, 1994.

[Dabral94] S. Dabral, R. Aslett, T. Maloney, "Core Clamps For Low Voltage Technologies", in *Proceedings of the 16th EOS/ESD Symposium*, pp. 141–148, 1994.

Index

Abrupt junction 76, 131, 133, 239, 249, 252, 253, 288, 289, 291, 292, 317
Adiabatic 336, 338
AEC 22, 24, 28, 29
Aluminium-silicon
  diffusions 304, 334
Aluminum-silicon 304
  eutectic temperature 334
Analytical CDM-testing 38
Analytical models 4, 335, 336, 339, 341, 344, 375
Anode 73, 76, 98–100, 150, 151, 157, 164, 218, 240, 297, 314, 315, 319, 373, 374
  to cathode spacing 99
Applications 6, 25, 35, 38, 39, 44, 45, 47, 57, 85, 98, 136, 138, 143, 144, 146, 153, 155–157, 162, 170, 171, 179, 183, 189, 192, 215, 216, 218–220, 223, 238, 247, 249, 326, 350, 376, 395–397, 399
  automotive 247
  industrial 247
  special 238
Arsenic implantation 290
ASIC 377, 381
  cell 381, 382
  library 377
Avalanche 11, 12, 70, 71, 73–76, 81, 83–86, 88–90, 95–97, 100, 101, 123, 131, 133, 138, 139, 141, 150, 152, 157–162, 164, 166, 170, 171, 180, 236, 241, 242, 250, 282, 284, 285, 287, 288, 290, 298, 299, 311, 313, 315, 327, 329, 333, 334, 339, 340, 343–345, 350, 351, 353–357, 359, 364, 367–370, 373, 375, 382, 383, 389
  conditions 83, 85
  generated current 11
  generation 11
  multiplication 12
Avalanche breakdown 70, 71, 73, 75, 76, 81, 89, 90, 100, 131, 133, 138, 139, 141, 150, 157, 158, 160, 162, 236, 241, 242, 250, 288, 290, 298, 299, 329, 333, 350, 351, 354, 357, 359, 364, 368, 369, 373, 375, 383, 389
  region 351
  voltage 71, 75, 76, 90, 100, 133, 139, 141, 157, 160, 236, 250, 290, 298, 354, 357, 359, 364

Background capacitance 23, 34
Bakeable leakage current 238
Ballast resistance 96, 283
Barrier metal 304
Base charge factor 359
Base current 77–79, 84, 86, 89, 95, 120, 175, 207, 212, 282, 284, 285, 292, 310, 327, 344, 355, 369, 370, 376, 386
Base transit time 21, 57, 80, 89, 314, 363
Base transport factor 79, 352, 356
Base-emitter resistance 219
Basic input protection scheme 156
BiCMOS 14, 95, 126, 157, 179–184, 188, 218, 219, 223, 327, 368, 371, 383, 396
BiCMOS protection circuits 184
Bipolar 4, 24, 48, 55, 68, 77, 78, 80, 83–86, 88–92, 95, 96, 98, 101, 126–129, 131, 133, 139, 144, 155, 179–182, 184, 202, 208, 209, 215, 216, 218, 219, 223, 250, 251, 257, 282–288, 292, 295, 299, 301–303, 307, 310–315, 317–319, 327–329, 344, 350–358, 361, 363, 367–369, 371, 382, 383, 385, 386, 389, 396
  collector current in the npn 77
  efficiency 129
  gain degradation 94, 95, 366
  lateral 88, 89, 127, 284, 285, 355

Bipolar (*Continued*)
npn transistor 68, 77, 78, 80, 83–86, 88, 89, 95, 98, 127, 180, 182, 219, 251, 283–288, 295, 312, 329, 351–357, 363, 368, 369, 383
pnp 68
protection circuits 181
SCR structure 180
self-biased 89, 101
technology 182
transistor 68, 77, 78, 80, 83–86, 88, 89, 95, 98, 127, 180, 182, 219, 251, 283–288, 295, 303, 310, 312–314, 318, 319, 329, 351–357, 363, 368, 369, 383
trigger voltage 128, 250, 284, 311, 344
Bipolar/BiCMOS 95, 180–182, 396
input Protection 181, 182
output protection scheme 181
technology 95

Calibration, tester 41
Calibration, testers 24
Capacitance at input/output 188
Capacitive loading 126, 179
Capacitive voltage suppression 9, 30
Cascode device 112, 201
Case Studies 267, 271
Case studies 6, 228, 238, 256
Cathode 73, 76, 98–100, 150, 151, 157, 240, 254, 315, 319, 373, 374
Caughey-Thomas empirical mobility 340
Channel length 82, 89, 92, 115, 128–130, 134, 136, 138, 152, 154, 170, 201, 204, 247, 249, 250, 252, 253, 311, 352, 355, 364, 394
ESD dependence 115, 128
Charged Board Model (CBM) 13
Charged Device Model (CDM) 5, 12–14, 17, 28–42, 44–47, 50, 57, 60, 61, 73, 105, 108, 117, 118, 121, 122, 127, 144, 145, 148, 198, 214, 215, 220, 221, 249, 300, 353, 363, 397, 398
direct charging method 32, 34
field-induced method 28, 30, 32
non-socketed test method 28
protection 214, 215, 220
socket 33
Charged Package Model (CPM) 33
Circuit models 20, 61, 90, 180, 327, 350, 375, 383, 389
Circuit simulations 6, 21, 39, 61, 327, 350, 353, 360, 376, 382, 389
Clamping voltage 21, 73, 88, 114, 150, 160, 172, 178, 206, 208, 210
Clock buffer 264
CMOS
input/output buffer 152, 180, 229, 230
CMOS processes 98, 100, 106, 119, 126, 127, 150–152, 157–159, 172, 173, 175, 178–183, 202, 203, 206, 208, 219, 231, 239, 282, 283, 285–287, 300, 327, 331, 352, 354, 355, 357, 364, 367–370, 386, 396
inverter 258
latchup 150, 151
Collector 77, 78, 80, 81, 83–89, 95, 96, 98, 99, 101
Collector-base breakdown 78, 85, 86, 203, 314
Collector-base junction 78, 84–88, 95, 96, 282, 283, 285, 287, 327, 331, 352, 354, 355, 357, 364, 367–370, 386
Collector-emitter breakdown 85, 203, 313
Common-base gain 83
Conductivity 9, 10, 69, 72, 73, 76, 87, 90, 93, 95, 96, 100, 133, 160, 306, 311, 334, 335, 340, 342, 360, 361, 383, 384
Conductivity modulation 73, 87, 90, 93, 95, 342, 360, 361
Constant current pulse 158, 339, 343
Contact 3, 5, 9–11, 23, 29–36, 39, 53, 56–58, 76, 86, 89, 90, 97, 98, 112, 116, 123, 127, 128, 130–132, 134–138, 144, 145, 147, 150, 154, 159, 161, 164, 170, 172, 182, 199, 219, 221, 236, 237, 243, 244, 246, 283, 296, 297, 300–304, 307, 310, 312, 314, 318, 330, 334, 342, 343, 355, 361, 375
burn-out 330
damage 170
drain spacing 145, 154
resistance 57, 135, 283, 300–303
spacing 123, 131, 137, 138, 182
spiking 131, 236, 237
transfer length 301
Contact damage 304

Contact spiking 304
Contamination 9, 36, 255
Core power supply 200
Correlation 17, 18, 21, 22, 24, 25, 27–30, 34, 35, 41, 47, 50, 56, 57, 60, 61, 231, 353
Current 1, 3–6, 8–14, 17–24, 26–33, 35–51, 53–61, 68–91, 93, 95–101, 105–107, 109–124, 126, 127, 135–139, 142, 144, 149–152, 156–162, 164, 165, 167, 168, 170–172, 175–181, 183, 188–193, 195–199, 201, 204, 205, 207–210, 212, 214, 218, 220–223, 228, 230–233, 235, 236, 238, 240, 242–244, 246, 249–251, 256–258, 282–292, 294–298, 300–307, 310, 311, 313–319, 326–331, 333, 334, 339, 340, 342–345, 350–377, 379, 380, 382–389, 396–398
  constriction 70, 331
  crowding at the emitter 80
  dependent resistor 361
  filamentation 96, 98, 101, 204, 233, 235, 304
  localization 283, 291, 292, 301, 330
  uniformity 135
Current discharge paths 17
Current flow equations 339
Current gain 79, 80, 84, 91, 99, 178, 208, 209, 285, 294, 301, 314, 317, 355, 356
  common-base 79, 84
  common-emitter 79
Current localization 303, 304
Current saturation 77, 159, 160, 389

Darlington effect 207
Dead Bug 29, 30, 39
Debugging 229, 230, 271
Decay time 18, 23
Design concepts 5, 105, 126
Design errors 221–223
Design methods, comparison of 112
Design of secondary element 155–157, 170, 183
Design synthesis 127, 155, 170, 181, 183
Device
  model 5, 6, 12–14, 17, 28–33, 40, 50, 57, 60, 61, 73, 105, 108, 117, 118, 121, 122, 127, 144, 145, 148, 198, 214, 215, 220–222, 249, 300, 326, 327, 350, 351, 353, 363, 383, 397, 398
  simulator 331
Differential amplifier 58
Diffusion
  silicided 6, 126, 155, 189, 192, 195, 196, 239, 242, 289, 300, 395
Diffusion capacitance 363, 364, 376
Diffusion Current 74, 76, 361
Diffusion current 74, 76, 361
Diffusion resistor 157, 159, 160, 162, 164–166, 168, 170, 171, 183, 240, 242, 250, 316, 319, 350, 371–373, 389
Diffusion sheet 135, 138
Diffusion times 338, 363
Diffusions 6, 70, 72–74, 76, 78, 79, 81, 90, 98, 100, 101, 112, 120, 126–128, 131, 132, 135, 137, 138, 143, 150, 155–157, 159–162, 164–166, 168, 170–173, 176, 180, 183, 189, 190, 192, 195, 196, 201, 202, 204, 206, 216, 218, 219, 221–223, 233–235, 237, 239, 240, 242, 249, 250, 258, 283, 289, 300, 301, 303, 304, 314, 316, 319, 334, 336, 338–340, 350, 354–356, 361, 363, 364, 371–373, 376, 383–385, 389, 395, 397
  salicided 150
Diode
  forward-biased 23, 70, 137, 153, 172, 257, 351, 360, 375
Diode-based protection 305, 312
Diodes 5, 11, 21–23, 68, 70, 72, 76, 77, 101, 120, 132, 171, 172, 178, 183, 198, 206–212, 214, 216, 220, 221, 223, 282, 297, 314, 319, 334, 340, 350, 353, 360, 361, 375, 383, 396
  reverse-biased 5
Direct charging 9, 10, 32, 34
Doping concentrations 69, 71–73, 82, 91, 96, 288, 291, 293–296, 301, 331, 332, 356
Double diffusion drain (DDD) 291
Double snapback 311, 312
Drain contact spacings 131, 137, 138
Drain Extended nMOS (DENMOS) 157, 216, 217
Drain Induced Barrier Lowering (DIBL) 95, 288
DRAM 5, 126, 171, 255, 260

Drift current 76, 361, 374
Drift velocity 69
Dual gate oxide transistor 201, 214
Dual SCR 153
Dual-diode 219, 294, 296, 297
Dummy nMOS 199
Dummy pMOS 199
dV/dt-triggering 21

Electrical failure modes 228
Electrical Overstress (EOS) 1–5, 223, 335, 339, 350, 351, 383–385, 388, 394
Electrical signature 228, 231
Electron concentration 372, 373
Electron ionization coefficients 74, 341
Electron mobilities 82, 363, 372
Electrostatic charging 12
Electrostatic voltage 8–11, 13, 17
Electrothermal effects 95, 327, 334
Electrothermal model 327, 341, 353, 385, 387
Electrothermal simulations 25, 334, 339, 343, 383, 385, 387, 389
Electrothermal simulators 342, 351, 383
Electrothermomigration 304
Element 4, 5, 11, 13, 17, 18, 20–23, 25, 27, 28, 39, 40, 42, 44, 46–48, 50, 52–58, 61, 68, 86, 105, 107, 109, 123, 126, 127, 138, 140, 146, 147, 149, 155, 156, 158, 160, 164, 165, 171, 176, 177, 179, 180, 183, 184, 189, 233, 256, 257, 282, 283, 294, 296, 317, 318, 326, 327, 334, 345, 350, 351, 353, 360, 363, 367, 371, 376, 379, 382–385, 387, 396–399
Emission Microscopy 230
Emission microscopy 250
Emitter area 80, 90, 95, 183, 208, 355, 367
Emitter crowding 183
Emitter current 77, 79, 80, 84, 91, 120, 180, 190, 207–209, 212, 354–357
Emitter efficiency 79, 80, 285, 314, 315
Emitter fingers 175, 182, 183, 219
Emitter length 181
Emitter width 175, 182, 367
Emitter-base diodes 207, 208, 210, 211, 214
Emitter-base diodes 206
Emitter-coupled-logic (ECL) 181
Energy gap 340
Epitaxial layer 99, 151, 288, 295, 296, 299, 316
Epitaxial thickness 151, 296–298
Error functions 338
ESD 1–15, 17–19, 21–35, 37, 38, 40–44, 46, 47, 49–52, 56, 58–61, 68, 69, 73, 76, 79, 81, 83, 85–88, 90, 95–98, 100, 101, 105–120, 122–124, 126–128, 130–140, 142–145, 149–153, 155–158, 160–162, 166, 168, 170–173, 175–184, 188, 193, 196–212, 214–216, 218–223, 228–233, 236, 238–241, 243–247, 249, 250, 252, 253, 255–258, 282–285, 287–306, 308, 310–320, 326, 327, 329, 330, 333–335, 339, 345, 350, 351, 353, 360, 363, 371, 375–377, 379, 382–385, 387, 389, 394–399
  design parameters 130, 135
  design rule checker 140, 144, 145, 231, 397, 398
  failure distribution 140, 144, 145
  implant 291
  modeling technique 326
ESD Association 10, 14, 18, 22, 24–26, 28–35, 37, 41, 42, 50–52, 56, 202, 203, 205–207
ESD Checker 221
Eutectic temperature 249, 301, 304, 334
  Al-Si 304
Extrinsic base current 327

Fail safe operation 109
Fail-safe 124, 189, 192, 216
Failure analysis 6, 60, 156, 169, 228, 229, 231, 239, 243, 247, 271
Failure analysis techniques 6, 229
Failure criteria 17, 47, 58, 60, 61, 383
Failure mechanism 17, 28, 47, 56, 58, 161, 228, 304, 394
Failure mode 1, 2, 4, 6, 131, 133, 161, 204, 228, 229, 231–233, 236, 304, 330
  distribution 1, 2
  electrical 228
  physical 228
Failure mode analysis 229, 233
Failure power 335
Failure voltage distribution 161

Failure window 157
Feature size 127, 132–134, 183, 287, 288, 394, 398
Field oxide 28, 127, 139, 140, 142, 151, 183, 188, 190, 201, 203, 233, 234, 254
Field oxide device (FOD) 127, 128, 132, 133, 139, 140, 142, 156, 183, 201, 203, 214, 234, 237, 239, 395
Field plated diode (FPD) 132, 133, 156–168, 170, 182, 237, 240–244, 246, 247, 249–253
Field-Induced Charged Device Model 13
Field-induced charging 5, 9, 10, 13, 28, 30, 32
Field-induced model FIM 11, 13
Filament 70, 96, 98, 101, 133, 170, 204, 233, 235–237, 241, 283, 304, 330
  molten 330
  polycrystalline 236, 237
Finger length 138
Finger Structure
  multifinger 142, 144, 147, 191, 195, 379, 387–389
  non-uniformity of turn-on 142, 147, 379
  number of fingers 136, 138–140, 380
Finger structure 136, 138, 312
Floating base transistor 85
Floating substrate 5, 126, 136, 137, 153, 192, 194, 215, 377
Floating substrate ICs 126
Forward biased diode 71, 72, 77
FPD
  snapback 162, 165
Freak failures 231

Gate drain overlap capacitance 139
Gate driven nMOS (GDNMOS) 116, 146
Gate oxide 9, 13, 28, 39, 48, 50, 57–60, 82, 88, 105, 117, 120, 121, 158, 166, 182, 200, 201, 214, 221, 223, 236, 237, 240, 297–300, 317, 318, 320, 353, 385, 394
  breakdown 158, 182, 298–300, 353
  breakdown voltage 298
  damage 300
  thickness 298, 299
Gate voltage 81–83, 90, 138, 142, 150, 153, 214, 284, 310, 354, 357
  coupling 138, 142
  effect of voltage on ESD performance 310
  grounded 138
  modulation 136
Gate-control 198
Gate-coupled nMOS (GCNMOS) 112, 116, 117, 123, 136, 138, 140–149, 153, 197, 201, 215, 223, 376, 387, 396
Gate-driven 116, 149, 183, 188, 205, 223
Gated diode 100, 132
Generation of carriers 86, 331, 351, 374
  avalanche 86, 351
  heat 334
  thermal 331
Generation rate 340
Graded junction 239, 289–291, 308, 317
Green's function 337, 383
Grounded gate nMOS 140, 143, 156, 158
Gummel plot 209, 367
Gummel-Poon model 353, 367–369, 389

HBT (Si-Ge bipolar transistor) 219
Heat capacity 342
Heat damage 171
Heat dissipation 137, 310, 316
Heat equation 327, 335, 336
Heat generation 339
Heat source 96, 131, 335–340, 383–385
High current I-V curve 70, 87, 89, 95, 284, 286, 311, 329, 333, 343
High pin count device testing 22
  ganging 26, 27
  rotation 27
  split-IO 27
High speed applications 170
  telecom 398
High-power applications 85
Holding voltage Vh 21, 49, 87, 99, 113, 117, 119, 120, 133, 150, 151, 164, 165, 180, 201, 204, 223, 240, 251, 284, 286, 287, 295, 297, 311, 327, 329, 343, 385
Hole concentrations 69, 78, 339

Hole impact ionization coefficient 74
Hole mobilities 69, 339, 359
Hostile environments 25
Hot carrier effects 289
Hot spot 96, 98, 101, 229, 301, 304
Hot swap 109
Human Body Model (HBM) 5, 12–14, 17–29, 36, 38, 44, 46, 52, 53, 56–59, 61, 105–107, 112, 117–120, 128, 131, 132, 137, 144, 156, 158, 161, 162, 166–170, 177, 210, 214, 219–221, 243, 247, 249, 296, 301, 302, 353, 363, 382, 398
  correlation to MM 21
  discharge waveform 24
  failure thresholds 21
  relationship to TLP 21
  reproducibility of the HBM test 17

Ideal diode equation 71
Impact ionization 70, 74–76, 82, 86, 205, 339, 341, 344, 356, 364, 374, 382, 385
Impact ionization coefficient 74, 339, 341, 344, 356, 382
Impact ionization generation current 385
Implants
  drain-source 287–289, 291, 292, 303
  ESD 291
  well 288
Impulse metrology 33, 43
Infra-red emission 230
Input buffer 145, 148, 181, 220, 221, 240
Input capacitance 188, 197, 198
Input leakage 57, 243, 254, 255
  (IIH) leakage failures 244
  (IIL) leakage failures 243
Input protection 117, 118, 120, 122, 127, 148, 154–157, 159–161, 164, 171, 181, 182, 230, 233, 238–244, 246, 247, 249, 253, 254, 257, 258
  circuits 160
  design 156, 238, 242, 247, 253
  optimization 242
  scheme 127, 154–156, 159, 182, 238, 242, 247, 253, 254
Input/output buffer 198, 199, 351
Interconnect parasitics 379
Interconnects 41, 305, 318, 389
Internal chip ESD damage 256, 257
Internal chip layout 256
Interval between pulses 25
Intrinsic base current 327
Intrinsic carrier concentration, ni 69, 78, 331
Intrinsic temperature, Ti 331
Ionic charging 9, 10
Isolation resistor 127, 145, 148, 152, 156, 159, 166, 182, 190, 223, 242, 246
Isothermal conditions 95, 383
It2
  degradation with finger width 196
  gate bias 146
  roll-off 146, 149, 308

JEDEC 22, 24, 26, 28–33
Joule heating 340
Junction breakdown voltage 86, 156, 298, 316
Junction capacitance 294, 376
Junction depth 72, 287, 288, 301–303, 356, 372, 385
Junction edge damage 330

Knee currents 359, 364, 367

Ladder layout 135
Latchup 5, 150, 151, 171, 176, 189–192, 194, 220, 288, 293, 294, 296, 316, 320
Latency effect 3
Latent effects 60
Lateral CMOS SCR 101, 180
Lateral diode 174, 189, 190, 197, 199, 242
Lateral diode to Vdd 174, 189, 197
Lateral DMOS (LDMOS) 157, 216, 218
Lateral npn 156, 240
Lateral pnp 99, 188, 197, 359
Lattice heating 340
Lattice mobility 340
Layout 3–6, 24, 42, 123, 126–131, 134–136, 153–155, 160, 164, 165, 167, 169–174, 176, 180, 182, 188, 190–194, 204, 206, 221, 223, 238, 240, 247–249, 253, 255, 256, 283, 293, 296, 297, 304–306, 315, 350, 351, 354, 360, 377, 379, 380, 387, 397, 398
  errors 221

guidelines 4, 6
parasitics 135
Leakage current
change in the Leakage current 231
non-backable leakage 231, 233, 236, 238
post-stress leakage current 231, 233, 235, 271
Leakage current 5, 47, 49, 58–61, 151, 210, 212, 231–233, 236, 317
Leakage problems 238, 247, 253
Light Emission Microscopy 230
Lightly doped drain (LDD) 2, 6, 116, 131–133, 156–158, 203, 233, 239, 242, 249, 252, 253, 287–291, 300, 308, 309, 316, 317, 395
Liquid crystal analysis 229, 258
Localized heating 21, 204
Low voltage trigger SCR (LVTSCR) 152–154, 168
cross section of 152
lateral 98, 150, 239
Low-power IC 310
Low-voltage IC 11
Low-voltage trigger SCR
cross section of 152
LSCR 150–155, 157, 158, 160, 164–171, 217, 239, 240, 242, 247, 250
Lumped circuit 327
Lumped element models 18
Lumped element thermal model 327

Machine Model (MM) 5, 12–14, 17, 21, 27, 28, 46, 52, 59, 61, 160, 169, 290, 353, 363
correlation to HBM 28
current 27, 28
reproducibility 28
Mean free path 75, 341
Melt filaments 133, 241, 283
Melting temperature of silicon 334
Microplasma 330
Microprocessors 177, 394
MIL-STD-883C method 24
Mixed mode simulators 327
Mixed Voltage 6, 199–201, 206, 212, 214, 221, 223
I/O 199, 201, 206, 212, 214
Mixed-mode protection 200
Mobile ions 253–255
Mobility degradation 361, 372
Modified LSCR (MLSCR) 151–154, 157, 158, 164–171, 217, 247
critical layout parameters 153
gate 154
trigger 165
MOS 3, 11, 13, 81, 83, 88, 90–92, 95, 101, 123, 127, 132, 136, 138, 146, 152, 154, 156, 180–182, 188, 189, 191, 201, 204, 205, 218, 257, 282, 285, 287, 288, 302–304, 306–312, 316–319, 327, 343, 350, 351, 353, 354, 357, 358, 360, 361, 364, 368, 374, 375, 377, 383, 385, 389, 396
gate coupling 144
protection circuits 179, 184
schematic cross section of MOS transistor 81
transistors 5, 48, 59, 68, 81–83, 88–90, 94, 95, 100, 101, 127, 132–136, 138–141, 144, 145, 149, 152, 156, 174, 180–182, 184, 188, 190, 233, 236–238, 282–288, 292, 295, 297, 299, 328–330, 343, 351–354, 358–360, 363, 384, 387, 394
Multi-voltage supplies 399
Multifinger nMOS 191, 195, 388
Multiplication 74, 90, 92, 123, 212, 285, 343, 355, 356, 368, 375
Multiplication factor, M 12, 75, 83, 84, 87, 88, 90, 92, 285, 292, 317, 354, 356, 357, 368, 374

n^+ diffusion 98, 100, 112, 120, 189, 218, 219, 223
n^+ guardring 258
n^+ resistor 255
n-substrate 24, 69
n-well 69, 70, 73, 76, 98–101, 106, 112, 127, 131, 143, 147, 150, 151, 157, 159–162, 166, 170–173, 180, 183, 188, 189, 216, 218, 239, 240, 242–244, 246, 249, 253–255, 283, 286, 287, 294–296, 314–319, 362, 374, 375, 387
doping 294–296, 313, 314
resistor 69, 70, 98, 143, 159, 160, 162, 166, 183, 315, 316, 319, 374, 375, 387
n-well resistor 315, 316, 319
n-well sheet resistance 294
Negative stress 137, 168, 215, 257, 260

Negative-resistance 70, 77, 89, 95, 96, 98, 99, 101, 159, 344
nMOS transistor 5, 48, 59, 81, 82, 88, 89, 95, 100, 127, 133–136, 139–141, 144, 145, 149, 156, 180, 181, 184, 237, 238, 282–284, 287, 288, 295, 297, 328–330, 351–354, 359, 363, 387
  failure of 133
nMOS transistors 302, 306–308, 310, 311
nMOS triggered nMOS (NTNMOS) 123, 191–197, 215, 220, 376, 396
  I/O application 197
  layout 192–194
Non-LDD process 233
Non-silicide 131, 133, 138, 140, 151
Nonuniform conduction 196
Normal circuit operation 140, 149, 150, 197, 201, 205, 207
npn 5, 68, 77, 78, 85–88, 97–100, 123, 128, 136–140, 150–153, 156–158, 160, 161, 165, 176, 180, 181, 184, 188, 201–203, 205, 216, 218, 219, 221–223, 238–240, 242, 247, 249, 251, 257, 282, 283, 286, 327, 328, 331, 340, 343–345, 352, 355, 362, 368–371, 396, 397
  breakdown 157
  conduction 139
  pull-down 181
  see also, laternal npn 123, 156, 240
  snapback 123, 128, 140, 158, 161
NTNMOS
  design 193

Off-chip protection 397–399
Off-state Current 329
Oscillating frequency 14
Output buffer 5, 126, 136, 152, 153, 180–182, 198, 199, 221, 223, 228–231, 233, 234, 236, 259, 260, 284, 286, 298, 316, 351, 385
Oxide breakdown 3, 39, 95, 118, 158, 182, 231, 298–300, 317, 318, 353
Oxide thickness 250, 298, 299, 385

p^+ diffusion resistor 242
p-i-n Diode 76
p-substrate 24, 69, 76, 106, 137, 150–152, 160, 171, 172, 239, 240, 285, 287, 295, 296
  resistance 151, 240
p-well 98–100, 150, 285, 290, 291, 293, 294, 296, 297, 310, 314, 317
Package 10, 24, 26,
  HBM 26, 28, 30, 35, 37–40, 42, 118
  CDM 37, 38,
  bonding 24
  high pincount 30
  plastic 40
Parameter extraction 353, 357, 364, 389
Parasitic current paths 221, 223, 397
Parasitic npn 14
Parasitic npn 5
Peak ESD current 3
Peak temperature 337
Phosphorus implantation 232, 289, 291
  channel 291
Photon Emission Microscopy (EMMI) 230
Physical failure mode 228
Pin combinations 24, 25, 28, 109, 110, 112, 171, 256, 396
Pin-to-pin 24, 28, 61
pMOS damage 238
pMOS effects 268
pMOS transistor 68, 81, 88, 94, 174, 188, 190, 237, 284–287, 360
pMOS transistor layout 174
  styles 175
PN diodes 5, 68, 71
PNP 5, 98, 188–191, 197, 199, 206–208, 210–212, 214, 219, 295
pnp 68, 98–100, 110, 119, 120, 123, 124, 150, 152, 172, 173, 175, 177, 178, 181, 183, 188, 197, 238–240, 246, 286, 287, 294, 314, 315, 318, 319, 359, 367, 368, 376, 377
  layout 173
  turn-on 150
PNP driven NMOS (PDNMOS) 123, 188, 189, 192
pnpn (SCR) 150
Poisson's equation 360
Polysilicon 127, 128, 133, 154, 156, 160–164, 166–171, 176, 183, 230, 231, 236, 250, 287, 288, 300, 312, 314–316, 319
Polysilicon
  resistor 156, 160, 162–164, 166–171, 183, 315
Polysilicon resistor 315, 316, 319
Power density 96, 282

Power dissipation 8, 96, 133, 137, 150, 160–162, 204, 216, 285, 292, 306, 307, 312, 327, 334, 336, 353, 368, 396
Power supply bus 351
  protection 351
Power vs. time curve 338, 339
Pre-charge voltage 8, 9
Primary element 127, 233, 317
Primary ESD protection 132, 171, 172
Process effects on ESD 6, 282
Process parameters 151, 196, 228, 282, 283, 308, 381
Protection circuits 1–6, 27, 49, 59, 68, 73, 76, 85, 88, 95, 98, 100, 101, 105, 106, 126–128, 140, 155–158, 160, 164, 165, 169, 170, 175, 176, 179–184, 208, 209, 218, 228–230, 233, 236, 238–244, 247, 249–253, 255–257, 282–284, 287, 289, 293–300, 304–306, 310, 312, 314–319, 326, 327, 329, 334, 336, 339, 345, 350, 351, 353, 360, 371, 375–379, 382, 384, 389, 394–399
  primary 239, 284
  secondary 155, 283, 296
  selecting a 170
Protection devices 2, 5, 6, 21, 52, 88, 107, 114, 116, 123, 124, 126, 127, 129, 132, 136–138, 142, 144, 145, 147, 150, 152, 153, 155–157, 171, 176, 177, 183, 184, 188–190, 192, 193, 195–199, 214–216, 218–221, 223, 228, 234, 257, 300, 305, 312, 371, 377, 379, 383, 385, 387, 389, 394–396, 398
  primary 127, 132, 155, 157, 183, 184, 219, 220
  secondary 127, 157, 183, 228, 371
Proximity effects 126, 171
Punchthrough 129, 136, 157

Qualification 9, 14, 15, 17, 23, 25, 28, 31, 32, 34, 36, 38, 40, 43, 46, 47, 56, 58, 60, 61, 238

RC delay 168, 183
Recombination 8, 9, 77–79, 285, 340
Reliability trade-offs 316, 317, 320
Resistance 8, 11, 13, 14, 18, 24, 26, 28, 30, 32, 34, 35, 39, 44, 46, 53, 57, 70–73, 76, 77, 86, 87, 89, 90, 95, 96, 98–101, 107, 112, 114–117, 120–123, 127, 131, 132, 135, 136, 138, 139, 142–144, 151, 157, 159, 161, 162, 167, 168, 170–172, 175, 177, 179, 180, 182, 183, 189, 190, 192–196, 204, 216, 219, 223, 240, 246, 249, 257, 283, 285, 288, 291, 293–297, 300–303, 305, 306, 311, 314–319, 327, 330, 332, 342–344, 350–353, 355, 359–362, 370, 371, 374, 375, 377, 379, 381, 382
  diffusion sheet 135, 138
  effective spreading resistance 303
  in the output transistor 179
  in the secondary clamp 106
  metal finger 135
  negative 70, 77, 89, 95, 96, 98, 99, 101, 154, 344
Resistance change 76
Resistive element 127
Resistive loading 127, 179, 398
Resistor 5, 14, 18, 20–22, 26, 28–32, 46, 51, 53–55, 57, 68–70, 76, 85, 86, 88, 98, 101, 105, 106, 111, 115–117, 120, 127, 140, 143, 145–149, 152–154, 156–171, 178, 180–183, 188–190, 197–199, 205, 219–221, 223, 240, 242–244, 246–250, 253, 255, 283, 315, 316, 319, 342, 343, 350, 353, 361, 364, 369, 371–376, 381, 383, 384, 387, 389
  n-type 69, 342, 343
  n-well 69, 70, 98, 143, 159, 160, 162, 166, 183, 315, 316, 319, 374, 375, 387
  p-type 69
  physics 5
  polysilicon 156, 160, 162–164, 166–171, 183, 315
  snapback 70
  time-dependent 384
Resistor 315
Reverse-biased diode 70, 181
RF applications 218, 219
Rise time 11, 18, 20–22, 32, 33, 41, 42, 44–46, 50, 52, 56, 58, 108, 140, 145, 177, 363
Robustness, design concept of 106, 107
Runt 33

Salicided processes *see* silicided processes

Saturation voltage 357
Scalability 364, 367, 374, 375, 389
Scanning Electron Microscopy (SEM) 169, 230, 231, 234–237
SCR 23, 68, 98–101, 113, 123, 127, 150–162, 164, 166–168, 170, 171, 180, 183, 215–219, 223, 239–247, 253–255, 295–297, 395–397
  layout 154, 240
  operation 98, 164
  primary protection 157
  protection 154, 157, 166
  trigger current 170
  trigger voltage 101, 151, 153
Second breakdown 48, 49, 96–98, 101, 107, 115, 116, 140, 158, 159, 165, 201, 204, 285, 291, 297, 299, 303, 304, 329–331, 333, 334, 339, 340, 342, 344, 345, 382–385
  condition for 333
  onset of 329, 331, 339, 342, 382, 384
Secondary element 127, 165
Secondary protection 106, 117, 127, 155, 156, 158, 160–162, 167, 169, 170, 181, 183, 240, 242, 246
  design 156, 161, 170, 183
  devices 127, 157, 183, 228, 371
  scheme 158, 169
Seebeck effect 339
Semi-infinite medium 336, 342, 383, 384
Shallow Trench Isolation (STI) 171, 172, 190, 201, 216, 217, 395, 396
Short-channel effects 291
Silicide 2, 114, 156, 236, 289, 300–303, 307, 311, 312, 314, 316, 318
Silicide-blocked processes 116, 204, 205
Silicon-on-Insulator (SOI) 37, 142, 143, 215, 216, 223, 310–312, 319, 387
Simulation 3, 5, 6, 21, 25, 39, 47, 57, 61, 70, 97, 139, 145, 146, 189, 190, 193–195, 199, 209, 221, 222, 326, 327, 331, 333, 334, 339, 340, 342–345, 350, 353, 359–363, 366–368, 370, 373–377, 379–383, 385–387, 389, 397
Simulation
  circuit 6, 21, 39, 61, 327, 350, 353, 360, 376, 382, 389
Simulators 327, 331, 342, 350, 351, 353, 357, 368, 377, 383, 384, 397, 398
Small Capacitor Model (SCM) 33
Snapback 20, 21, 23, 48–50, 70, 83, 87, 90–97, 101, 107, 108, 110, 112, 115–117, 119, 120, 123, 124, 128, 133, 139, 140, 149, 150, 158–162, 165, 166, 177, 178, 180, 191, 195, 201, 202, 204–206, 214, 250, 251, 253, 284, 285, 287, 298–300, 311, 312, 315, 316, 318, 319, 329, 334, 340, 342–344, 351, 353, 355–359, 365, 366, 372–375, 389
  FPD 162, 165
  holding voltage 87, 117, 119, 120, 133, 150, 165, 180, 201, 204, 251, 311, 343
  model 123, 206, 358, 359, 372, 374
  region 95, 101, 300
  resistance 139
  trigger current 87
  trigger voltage 87, 119, 180
  voltage 107, 139, 160, 165, 177, 250, 253, 329, 357
Sneak current paths 199, 221
Snubber diodes 212
Socket Discharge Model (SDM) 17
Soft failure 236
Source-contact-to-gate spacing (SCG) 134
Source/drain doping profiles 289
Source/drain junction effects 289
Source/drain sheet resistance 283
Specific heat 335, 383, 384
Speed of ESD clamp activation 108
SPICE 120, 122, 123, 139, 145, 175, 180, 183, 189, 193, 195, 199, 206, 207, 209, 222, 327, 350, 352, 375
SPICE parameters 350
Square pulse 14, 18, 39, 47–51, 54, 61
SRAMs 58
Stacked nMOS 204–207, 223
Static charge generation 4
Statistical time lag 12, 36
Step stress 25, 244
Step width 23
STI effect 395
STNMOS 123, 189–192, 196, 197, 215, 220, 221, 376
Stress hardening 23, 25

Stress polarities 39, 137
negative 137
positive 39
Sub-0.25 um 3
Substrate
potential 90, 95, 97, 190, 284, 352, 355, 379, 380
resistance 89, 90, 144, 151, 161, 171, 189, 190, 192, 193, 196, 216, 240, 288, 293, 295, 297, 314, 350–353, 359–362, 375, 377, 379, 381
Substrate bias effect 144
Substrate current 57, 82, 83, 89, 90, 142, 144, 192, 193, 221, 284, 306, 356, 361, 362, 364, 365
Substrate trigger nMOS, STNMOS 123, 189–192, 196, 197, 215, 220, 221, 376
Substrate trigger protection 189
System Level HBM 14, 26

Telecom 179, 398
Temperature dependence 76, 209, 332, 386
of alpha 74
of bandgap 386
of beta 341
of mobility 332
saturation velocity 341
Temperature-dependent thermal conductivity 340
Test board capacitance 20–22
Test procedures 23, 24, 26–28, 35, 36, 57, 58
Thermal boundary conditions 341, 342, 344
Thermal breakdown 95, 96, 282, 300, 329–332, 383, 384
Thermal capacitances 342
distributed 342
Thermal conductivity 133, 160, 306, 311, 335, 340, 383, 384
Thermal current 344
Thermal diffusion constants 339
Thermal diffusion length 336
Thermal diffusivity 335
Thermal equations 339
Thermal generation 74, 84, 90, 96, 97, 101, 334, 339, 345, 354, 382, 385–387
Thermal generation currents 84, 96, 97, 101, 334, 339, 345, 382, 385–387
Thermal model 196, 327, 335, 341, 344, 353, 383, 385, 387
Thermal resistance 311, 319, 327, 342, 343
distributed 342
lumped 327
Thermal snapback 342
Thermally generated carriers 73, 96
Thermistor devices 334
Thick field device (*see* Thick oxide device) 127, 128, 130, 131, 136, 141, 154, 156, 157, 160, 176, 236
Thick oxide 127, 129, 130, 134, 136, 141, 154, 156, 157, 160, 176, 236
Thick oxide Protection 129
Thick oxide transistor 156
Thin oxide 132–134, 136–139, 151, 299
Thin oxide device 132–134, 136–138
Thompson effect 340
Thompson heating 340
Thyristors 98, 327
Time domain metrology 43
dispersion 44
failure current It2 49, 58
failure voltage Vt2 49, 58
holding Voltage vh 49, 50
loss 44
multiple reflections 54, 55, 57
sampling 34
single shot 44
step response 44
trigger voltage vt1 49, 50, 56, 58
Time to failure, Tr 339
Time-dependent regions 336, 338
Titanium 243, 244, 247, 300
Total IC chip protection 255, 256
Total input protection 156, 164, 171, 247
Trace 9, 35, 38, 39, 41–44
Transient Latch-up (TLU) 14
Transient thermal simulations 342
Transistor Operation 77, 85, 95, 288
Transistor width 138
Transmission Line Pulsing (TLP) 14, 18, 21, 25, 46–59, 61, 106, 203
current source TLP 51, 53
pulsed characterization 47, 48
Time Domain Reflectometer (TDR) 54
Time Domain Transmission (TDT) 51–53, 55

Transmission Line Pulsing (TLP) (*Continued*)
  Time Domain Transmission Reflectometer (TDTR) 51, 53, 55
  trigger current It1 49, 56, 58
  trigger voltage 50
Transparency, design concept of 109
Triboelectric charging 9, 10
Triboelectric series 9
Trigger circuit 153, 180
Trigger current, Itrig 49, 87, 89, 99, 100, 158, 165, 170, 240, 296, 297
Trigger diffusion (in SCR) 157, 158, 164
Trigger element 86, 180, 184, 294
Trigger voltage 49, 50, 87, 99–101, 107, 119, 128, 130, 136, 151–153, 157, 165, 168, 177, 179–181, 183, 184, 217, 218, 240, 247, 250–253, 257, 284, 298, 312, 344, 371, 396

Uniform trigger 147, 191, 195, 201, 219

V_{dd} capacitance 189–192
V_{dd} charging 197, 198
V_{dd}-based protection 109–112
V_{dd}-referenced modulation 149
Velocity saturation 92, 360, 361, 371–373, 375
Verification Standard Test Modules 32
Vertical *npn* 180, 184, 327, 328, 368
Vertical *pnp* 189–191, 206, 207, 210, 295
Voltage tolerance 109, 172, 200, 201, 210
V_{sp} 87, 90, 119, 120, 165, 201, 284, 329, 343, 344, 355
V_{ss} bus 396, 398
V_{t1} 49, 50, 56, 58, 87–89, 99, 142, 145, 146, 180, 181, 193, 195, 201, 284, 311, 329, 343, 344
V_{t2} 49, 58, 96, 139, 142, 146, 284, 329, 344
V_{trig} 165

Wells 253, 254, 288, 293–295, 307, 317–319
Width dependence 140, 196
Window effect 23
Wunsch-Bell equation 336, 338

Zener diode 127, 149, 219
Zener triggered *npn* 219