INVESTIGATION OF ESD PERFORMANCE IN ADVANCED CMOS TECHNOLOGY

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF

ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES OF

STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Kwang-Hoon Oh

October 2002

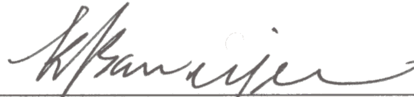
© Copyright by Kwang-Hoon Oh 2002

All Rights Reserved

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

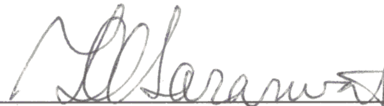
Robert W. Dutton
(Principal Advisor)

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



Kaustav Banerjee
(Associate Advisor)

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



Krishna Saraswat

Approved for the University Committee on Graduate Studies:

ABSTRACT

Electrostatic discharge (ESD) is one of the most important reliability issues in the integrated circuit (IC) industry and it is known that nearly 40% of all IC failures are associated with ESD/EOS (electrical overstress) related modes. Therefore, controlling ESD is indispensable for achieving higher quality and reliability standards of IC chips.

As VLSI technology continues to scale, new processes and materials can have profound effect on the ESD performance of sub-150 nm CMOS devices. Hence, characterization and modeling of ESD phenomenon for these technology nodes become necessary to comprehend the physical aspect of the ESD phenomenon and to consequently propose suitable protection schemes.

In this thesis, by investigating various aspects of ESD behavior involved in advanced 0.13 μm CMOS technology, it is identified that the non-uniform bipolar conduction phenomenon during ESD events results in a severe reduction in ESD protection strength. This non-uniform bipolar conduction becomes more serious with silicided protection devices, compared with non-silicided devices.

Using experiments and device simulations, it is inferred that the root cause of this non-uniform conduction is likely related to the intrinsic process defects that lock-in the bipolar conduction to a local area along the finger width. Furthermore, an extensive investigation into the impact of the substrate-bias and gate-bias on ESD performance has been carried out to provide new insight into the bias effect involved in the non-uniform conduction process.

The improvement of ESD failure threshold with the gate-to-contact spacing for fully silicided NMOS transistors has also been investigated. Results from this work suggest that even for silicided processes, the gate-to-contact spacing should be carefully engineered to achieve efficient and robust ESD protection designs.

Along with the non-uniform bipolar conduction and the relevant gate-to-contact spacing effects, it is also identified that for deep submicron devices an observed unusual reverse channel

length dependence of ESD performance is due to the reduction of thermal capacity arising from the reduced effective device width associated with severe non-uniform bipolar conduction.

Finally, to further enhance the understanding of the impact of silicided process on ESD hardness, an analytic silicide contact resistance model that can describe high-temperature and high-current behavior of the silicided contact system was formatted using physical models. The model shows how the current can be localized under ESD conditions and thus it extends the design capability of ESD protection providing information relevant to ESD failure limit.

Results from this work can be used to construct suitable design windows for efficient and robust ESD on-chip protection and have been demonstrated at Texas Instruments Inc., in practical ESD protection designs. These results can also be used to establish design guidelines for use in high performance 0.13 μm technology.

ACKNOWLEDGMENTS

I am indebted to many wonderful people who made my life at Stanford a very rewarding and invaluable experience. Without their support and concern, this dissertation could not have been completed.

First of all, I would like to express my sincerest gratitude to my principal advisor Professor Robert W. Dutton for his excellent guidance and ceaseless support throughout my Ph.D. study. He gave me such a great opportunity to work and study in TCAD group and the freedom to independently pursue my research interests. His extraordinary engineering perspective and insight have led me to many successful researches. Also, he made me realize my research potential and encouraged me to challenge comprehension of ESD reliability in deep submicron CMOS technologies. He has given me incessant motivation, inspiration and constructive suggestions through technical discussion and many paper reviews.

I would like to show my gratitude to oral examinations committee members Professor Kenneth E. Goodson, and Professor Shan Wang. Also I wish to give special thanks to Professor Krishna Saraswat for his careful review of my dissertation as well as for serving on my oral examination committee.

I would like to express special thanks to my associate advisor Professor Kaustav Banerjee at University California, Santa Barbara. I am greatly benefited from him. His advice was extremely valuable to my research and his extraordinary engineering insight enabled me to achieve many successful performances. Collaborations with him for a couple of years were very pleasant, memorable and productive.

I would like to thank Senior Research Scientist Dr. Zhiping Yu at Dutton's group. Dr. Yu helped and advised me a lot in various ways. In particular, he assisted me to define and concrete my research interests in my first two years at Stanford. Detailed technical discussions with him are always helpful.

I would like to express special gratitude to my former group mates Dr. Chang-Hoon Choi (currently Research Scientist at Dutton's group) and Dr. Jung-Suk Goo. We have shared almost every moment together in the office and technical discussions and chatting with them are always useful as well as pleasant.

Special thanks go to the group mates, Tae-Young, Li-Chang, Choshu, Eric, Jae-Wook, Reza, Jung-Hoon and Hai, and the visiting scholars, Prof. Hyungsoon Shin, Mr. Youngkwan Park and Dr. Olof Tornblad. I would also like to acknowledge the help of the group staff, Fely, Miho, and Dr. Yergeau in taking care of administrative issues and gloworm computer systems.

The time I have spent at Stanford has been unforgettable and enjoyable due to the all the friends that I have made. Since they have always been beside me, I could have been motivated and inspired. In particular, I wish to give special thanks to my friends, Dr. Hyun-Jin Cho, Wonil, Kyungsuk, Dr. Nahmsuk Oh, Eui-Young, Kiyoung, Jaewon, Donghyun, Moonjung, Soyoung, Jaeha, Ulrich, Sangmin and the KSCS (Korean Student Catholics at Stanford).

My most important experiences during my study were the two summers that I spent at Texas Instruments Inc. in Dallas, TX. Although extremely hot Texan weather bothered my family and me, in return I could have made significant progress through collaboration with excellent people there. I would especially like to thank Drs. Charvaka Duvvury, Young-Min Kim, Craig Salling, Gianluca Boselli, and Tim Rost of Silicon Technology Development for giving me the wonderful opportunity to work with them on a variety of interesting and stimulating topics related to design of ESD protection. In particular, I am greatly benefited from the interaction with Dr. Duvvury. His exceptional engineering insight and technical knowledge led me to many successful investigations. He coauthored conference (*IRPS*, *IEDM*) and journal (*IEEE T-ED, T-DMR*) papers with me and reviewed every work that I did for the last couple of years with constructive comments.

Also, I would like to many people in industry: Drs. Stephen Beebe of Advanced Micro Devices (AMD) and Jeremy Smith of Texas Instruments Inc. for mentoring the *IRPS* papers; Dr. Deok-Jung Kim of Fairchild Semiconductor for his firm support during my Ph.D. study.

Finally, special thanks go to my family - my parents, my sisters, and my wife, Hye-Jung and my sons, Tae-Seong and Jun-Seong for their endless support, love, patience, sacrifice and understanding throughout my study.

TABLE OF CONTENTS

Abstract v

Acknowledgments vii

List of Tables xiii

List of Figures xv

1 Introduction.. 1

1.1 Motivation 1

1.2 What is ESD ? 3

1.3 On-chip ESD Protection.. 4

1.4 ESD Protection Devices: Physics and Operation.. 6

1.4.1 Resistor.. 6

1.4.2 Diode.. 8

1.4.3 NMOS Transistor.. 8

1.4.4 Silicon Controlled Rectifier (SCR) 10

1.5 Thesis Outline.. 13

2 Characterization of ESD Phenomena.. 15

2.1 Types of ESD Events.. 15

2.1.1 Discharge to the Device.. 15

2.1.2 Discharge from the Device.. 16

2.1.3 Field Induced Discharge.. 17

2.2 ESD Related Failures.. 18

2.2.1 Catastrophic Failure.. 18

2.2.2 Latent Damage.. 20

2.3 Experimental Techniques.. 21

2.3.1 Transmission Line Pulsing (TLP)……………………………..……….……. 21
2.3.2 Emission Microscopy (EMMI)…………………………………....….…..... 25
2.4 Technology under Investigation……………………………………….….... 27

3 Non-uniform Bipolar Conduction under ESD Conditions………...………….……… 29
3.1 Introduction……………………………………..…………...….....…... 29
3.2 Experimental Evidence of Non-uniform Bipolar Conduction ...……….....…..….... 30
3.2.1 Transmission Line Pulsing (TLP) Tests…………………….……………. 31
3.2.2 Emission Microscopy (EMMI) Analysis……………………..…….………. 34
3.3 Physical Modeling of Non-uniform Bipolar Conduction……………………...… 37
3.4 Implications for the Design of ESD Protection ……………………………..… 41
3.5 Summary…………………………………...……………………………..… 42

4 Bias Dependencies of ESD Robustness……………………………………….……. 43
4.1 Substrate Bias Effect……………………………..…………………..…….…. 44
4.1.1 Intrinsic Second Breakdown Triggering Current, I_{t2i} ……………..…….…. 45
4.1.2 Device simulation Study: Principle of Operation with V_{sub}……………...… 48
4.1.3 Effective Finger Width under ESD Stress…………………..……..……… 50
4.2 Gate Bias Effect…………………………………………...…………......… 52
4.2.1 Output NMOS Failure………………………..…………………………...… 52
4.2.2 Experiments and Analysis…………………………………………..……… 56
4.2.3 Simulations and Discussion…………………………………………..……… 59
4.3 Design Window for Advanced ESD Protection…………….…………..……..… 66
4.4 Summary………………………………………..………………………..…. 67

5 Impact of gate-to-Contact Spacing……………………………..……………….……... 69
5.1 Introduction………………………………………………………………….…... 70
5.2 Experiments…………………..…………………………………………....….. 71
5.2.1 I_{t2} Dependence on Contact Spacing……..…………………...………....….… 71
5.2.2 Effect of Substrate Bias…………………..……………………………...… 75
5.3 Analysis and Discussion………………………...……….…………………….….. 78
5.3.1 Ballasting Current Distribution……………………...…………………....… 78
5.3.2 Characteristics of the Lateral n-p-n Transistor……………….………..….... 80

5.3.3 Thermal Effects……… 86
5.4 Summary……… 92

6 Reverse Gate Length Dependence……… 93
6.1 Introduction……… 93
6.2 ESD Performance with Gate Length……… 94
6.3 Analysis and Discussion……… 97
6.4 Summary……… 103

7 Modeling of Contact Resistance……… 105
7.1 Introduction……… 105
7.2 Temperature Dependent Specific Contact Resistance……… 106
7.3 Implications on ESD Reliability……… 114
7.4 Summary……… 118

8 Conclusions……… 119
8.1 Contributions……… 120
8.2 Suggested Future Work……… 122

Bibliography……… 124

LIST OF TABLES

2.1 Summary of ESD related process information of the technology under investigation…………..…………………………………………………………………….. 27

LIST OF FIGURES

1.1 Reported typical failure threshold current of ESD protection devices (NMOS transistors) for various process conditions and technology nodes (data from Philips, TI, IBM and IMEC)………………………………………………………….……..…... 2

1.2 The generic configuration of the ESD protection circuit in a bi-directional I/O circuit. The diode from V_{ss} to V_{dd} represents the p-substrate to n-well diode inherent in any CMOS process……………………………………….……………………….……..…. 5

1.3 Typical I-V characteristics of a diffused resistor…………….……………….…….….… 7

1.4 A typical diode ESD protection scheme…………………….……….………….……. 8

1.5 Typical operation of the gate-grounded NMOS (ggNMOS) (a) and I-V characteristics (b) under ESD conditions where I_{gen} is the avalanche generation current, I_{sub} is the substrate current, I_c is the collector current, I_b is the base current (i.e., I_b=I_{gen}-I_{sub}), V_{t1} (I_{t1}) is the triggering voltage (current), V_h (I_h) is the holding voltage (current), and V_{t2} (I_{t2}) is the second breakdown triggering voltage (current)……….……….………… 9

1.6 (a) Cross section of a lateral SCR in a CMOS process showing the parasitic p-n-p and n-p-n transistors. (b) High current I-V curve for an SCR (the inset shows the equivalent circuit schematic of the SCR)……………………...…………….….…... 11

2.1 Simplified typical ESD stress models (after [30]). (a) Human Body Model (HBM), (b) Machine Model (MM), and (c) Charged Device Model (CDM).…………….……. 16

2.2 RLC discharge current waveforms of the three basic ESD stress modes (after [31])…………………………………………………………………….…….….……... 17

2.3 Current induced damage mechanisms under ESD conditions (after [5]). (a) Thin film fuse, (b) Junction filamentation, and (c) Junction spiking……….……….…..…...… 19

2.4 Voltage induced damage mechanisms under ESD conditions (after [5]). (a) Charge injection, and (b) Oxide rupture……………………………….….………………… 20

2.5 Schematic of a TLP setup. (a) Circuit representation of the TLP and principle of pulsed characterization. (b) ESD test setup of the automated TLP …………………… 23
2.6 (a) High current snapback curves from the TLP measurement. The drain voltage and current waveforms (b) before triggering, (c) after triggering, and (d) at failure ……… 24
2.7 ESD design window for an ESD protection element…..…………………..……….. 25
2.8 EMMI setup used in this work for observing ESD behavior…………………...…..… 26
2.9 TEM image of the low voltage (1.5 V) NMOS transistor using the 0.13 μm CMOS technology……………………………………………………...……………..….….. 28
3.1 The layout of the single finger NMOS transistor. The contact opening (CNT) is 0.15 μm for both 1.5 V and 3.3 V devices. The gate to source/drain contact spacings (SS/DS) are 0.1 μm and 0.225 μm, and n+ overlaps of source/drain contact (S_{OL}/D_{OL}) are 0.4 μm and 0.125 μm for the 1.5 V and 3.3 V transistors, respectively. Also the body space (BS) from the source diffusion to the substrate diffusion is the same as the finger width (W)……….…………………………………………….……….….... 31
3.2 Sample high current I-V curves for 3.3 V NMOS transistors with different finger widths for (a) non-silicided, and (b) silicided processes where L_{poly} = 0.5 μm. The current per unit finger width at second breakdown I_{t2} [mA/μm] strongly depends on the finger width (W), which illustrates non-uniform bipolar currents flow under ESD conditions………………………...…………………………………….…….….…. 32
3.3 Silicide process dependent I_{t2} with the single finger width for (a) the1.5 V and (b) 3.3 V NMOS transistors. The I_{t2} roll-off with W indicates that the failure current essentially remains constant as W is further increased beyond this roll-off point…………………………………………………..…………..……..….…..... 33
3.4 EMMI images showing spatial extent of lateral current conduction at different current levels for 3.3 V (W/L_{poly}=20/0.5 μm) (a) silicided, and (b) non-silicided single finger NMOS devices. (T_p = 300 ns, f = 400 Hz, and T_{exp} = 6 min) …………... 34
3.5 Emission microscopy images showing spatial extent of lateral current conduction at different current levels for 3.3 V (W/L_{poly}= 80/0.5 μm) (a) silicided (b) non-silicided single finger NMOS devices. (T_p = 300 ns, f = 400 Hz, T_{exp} = 6 min). The small bright spot indicates a failure or a partial failure of the devices. The discontinuity of the conduction region in (b) was not observed in the EMMI images at the lower current level than the 60 mA and the two spots were included within the conduction region……………..……………………………………………………………… 36

3.6 Schematic of the segmented n-p-n transistors for a gate-grounded single finger NMOS transistor. Each n-p-n transistor has different intrinsic characteristics due to the statistical variations. R_s, R_d, R_{sub}, and R_b denote the parasitic resistance in the source, the drain, the substrate, and the intrinsic base, respectively …………………. 37

3.7 Non-uniform current conduction with a mixed mode transient simulation for M1 and M2 (W/L_{poly}=1/0.5 μm). (a) the schematic of simulation. (R = 500 Ω and r = 0.81 Ω), (b) drain current and voltage with elapsed time for M1 and M2, and (c) the current flowlines at the two different time conditions, A and B……………………………… 40

4.1 Schematic of the biasing scheme for investigating impact of substrate bias on ESD performance……………………………………………………………..…….... 44

4.2 Emission microscopy images that show the spatial extent of lateral current conduction with different V_{sub} for 3.3 V (W/L_{poly}= 80/0.5 μm) (a) silicided and (b) non-silicided single finger NMOS devices. (T_p= 300 ns, f= 400 Hz, T_{exp}= 6 min) ….. 45

4.3 Second breakdown triggering current I_{t2} with V_{sub} for the 3.3 V silicided devices with two different R_{sub} (6300 Ω-μm and 4800 Ω-μm) where W/L= 20/0.5 μm. I_{t2} approaches its intrinsic value I_{t2i} as the substrate bias is increased ……………….….. 46

4.4 I_{t2} as a function of transistor width for different substrate bias V_{sub} for (a) non-silicided, and (b) silicided 3.3 V NMOS transistors……………….………………..... 47

4.5 Static I-V characteristics for the ggNMOS with L_{poly}=0.5 μm. (a) High current I-V curve, and (b) current flowlines at the drain current (I_D) of 500 μA/μm with V_{sub}=0 V and V_{sub}=0.75 V. The current flows more deeply into the substrate with V_{sub} ………. 49

4.6 Effective finger widths (W_{eff}) vs. designed finger widths (W) for 3.3 V NMOS transistors with different substrate bias (a) non-silicided devices, and (b) silicided devices …………………...…………………………………………………………..... 51

4.7 Simplified output buffer protection scheme with various protection NMOS transistor options. Irrespective of protection transistor options, under ESD conditions, high ESD voltage can be fed into the gate of the output NMOS transistor, which could lead to early ESD failures. (The gray arrows indicate the ESD current paths.) (a) gate-grounded NMOS (ggNMOS), (b) gate-coupled NMOS (gcNMOS), (c) substrate pump NMOS (stNMOS), and (d) generalized high current I-V characteristics of ggNMOS……………..………………………………………………………….…. 53

4.8 The schematic of the standard 2 kV Human Body Model (HBM) tests for simple output buffer protection circuit………………………………...………..……….….. 54

4.9 Simulated voltage waveforms of each node of the output buffer protection under 2kV HBM stress with the four different pre-drive circuit conditions where (a) $p_1 = p_2 =$ low and $n_1 = n_2 =$ high, (b) $p_1 = p_2 = n_1 = n_2 =$ low, (c) $p_1 = p_2 =$ high and $n_1 = n_2 =$ low, and (d) $p_1 = p_2 = n_1 = n_2 =$ high………….……………………………...……. 55

4.10 Failure image of the output NMOS transistor in HBM test mode. Since increased gate voltage of the device due to high gate coupling lowers its ESD strength, the device fails earlier than the protection devices……….…………………………..……….. 56

4.11 Second breakdown triggering current (I_{t2}) vs. gate voltage for a salicided ($TiSi_2$) 0.35 μm technology node. The dotted line is the well-known substrate current (in arbitrary unit) for conventional NMOS transistor, which shows that the degradation of I_{t2} is not due to the substrate current behavior as a function of gate bias, and further suggests that the turn-on efficiency of the parasitic n-p-n transistor is not the primary cause of this phenomenon……………………………………..…………………... 56

4.12 Second breakdown triggering current (I_{t2}) with gate bias for the two different NMOS transistors for the 0.13 μm technology. The dependence of I_{t2} on gate bias is influenced by the finger widths. Gate bias improves I_{t2} for wide finger-width NMOS transistors, but degrades I_{t2} of narrow finger-width devices. (a) 1.5 V NMOS and (b) 3.3 V NMOS……………………………..……………………………………… 57

4.13 (a) I_{t2} for the output NMOS transistors (both of 1.5 V and 3.3 V transistors) with V_{gs} = 0 V for different finger widths and (b) EMMI image of the spatial distribution of ESD current (I_{ESD}) of 30 mA for the 3.3 V NMOS transistor. They both show that strong non-uniform conduction occurs in salicided NMOS transistors for the 0.13 μm technology. For the EMMI analysis, a pulsed bias with duration (T_p) of 300 ns was applied at a frequency of 400 Hz for the exposure time (T_{exp}) of 6 min ……………… 58

4.14 The current density at the edge of source (S_E) and drain (D_E) extension along y-axis with gate bias (see the *x* and *y* axis in the rectangle underneath the gate) and the schematic of the transient simulation circuit is shown in the inset …………………… 60

4.15 The local temperature rises with (a) V_{gs}=0 V and V_{gs}=3 V, and temperature contours with (b) V_{gs}=0 V and (c) V_{gs}=3 V. The simulations indicate that the channel and drain extension area undergo more heating and that the temperature distribution is more localized with gate bias…………..…………………...………………………………. 61

4.16 The location of peak temperature and the peak temperature value with gate bias. With increase in gate bias, Y_{peak} moves closer to the surface and the peak temperature, T_{peak} also increases moderately………………………………………………….………..…… 62

4.17 Impact of substrate bias on the I_{t2} with finger widths…………….……………..……….…. 63

4.18 Simulated current flowlines in the presence of gate bias with (a) V_{sub}=0 V and (b) V_{sub}=1 V. With substrate bias, the current flows more deeply into substrate and this leads to increase in effective volume for power dissipation……..………………...….. 64

4.19 Temperature distribution with substrate bias in the presence of gate bias. (a) temperature distribution along the x-axis and (b) temperature distribution along y-axis where T_{peak}=775 K for V_{sub}=0 V and T_{peak}=760 K for V_{sub}=1 V……………….…... 65

4.20 Effect of substrate bias on the I_{t2} degradation for high voltage (3.3 V) NMOS transistors with different finger width and gate bias, which shows that I_{t2} degradation is alleviated with substrate bias. The inset also shows the impact of substrate bias on I_{t2} degradation for narrow finger, high and low voltage transistors ……………… 66

4.21 I_{t2} of the high voltage (3.3 V) NMOS transistor with finger widths for the various gate voltages. Two competing trends are clearly shown ……..……….……………..... 68

4.22 Design window for optimizing the performance of deep submicron ESD protection…………………………………………………………………………..….…. 68

5.1 The schematic of a silicided NMOS transistor indicating the gate to source/drain contact spacing (GSCS/GDCS) and the n^+ overlap of the source/drain contact ($S_{_OL}/D_{_OL}$). For the 1.5 V NMOS, the S/D contact opening (S/D) = 0.15 μm and $S_{_OL} = D_{_OL}$ = 0.1 μm. For the 3.3 V NMOS, the S/D contact opening (S/D) = 0.15 μm and $S_{_OL} = D_{_OL}$ = 0.125μm. In the test structures, the S/D contact opening and n^+ overlap of S/D contact ($S_{_OL}/D_{_OL}$) remain unchanged despite the variations of GSCS/GDCS ……………………………………………………………………… 71

5.2 The measured drive currents (I_{drive}) for the 1.5 V NMOS transistors with different gate-to-contact spacings show no apparent differences, where V_{gs}=V_{ds}=1.5 V. The inset shows high current TLP curves for the 1.5 V NMOS transistor, which clearly show the impact of the gate-to-contact spacings on I_{t2} despite the silicided diffusion………………………………………………………….………...……..…… 72

5.3 The second breakdown triggering current (I_{t2}) with the gate-to-source/drain contact spacing for two different silicided NMOS transistors. The two dotted circles indicate I_{t2} for each device with minimum contact spacing: 0.1 μm and 0.225 μm for the 1.5 V NMOS (W/L_{poly} = 20/0.175 μm) and 3.3 V NMOS (W/L_{poly} = 20/0.5 μm), respectively …… 73

5.4 I_{t2} values for the 1.5 V NMOS transistors with various GDCS and GSCS. I_{t2} depends on both GDCS and GSCS within the scatter of data. W= 20 μm.…… 74

5.5 (a) I_{t2} of the 1.5 V NMOS transistors is dependent on GSCS and GDCS with an external substrate bias of 0.7 V, and is nearly independent of GSCS and GDCS with sufficient external substrate bias of (b) 1 V and (c) 1.25 V…… 76

5.6 Total failure threshold current (IT_2) with the finger width for (a) 1.5 V NMOS transistors and (b) 3.3 V NMOS transistors where GDCS/GSCS= 0.1 μm; the total failure current can be scaled with finger width within only a limited finger width (W ≤ 20 μm) and the inset shows that the 3.3 V device (W= 20 μm and GSCS= 0.225 μm) with nearly uniform current conduction shows negligible dependence of I_{t2} on GDCS…… 77

5.7 High current TLP curves for the 1.5 V NMOS transistor with two different GDCS of 0.1 μm and 1 μm where GSCS is 0.1 μm. The slope of high current regions is almost identical, but I_{t2} is nearly doubled with the increased GDCS …… 78

5.8 Equivalent circuit of the NMOS including the parasitic lateral n-p-n bipolar transistor when the lateral n-p-n is on. I_{DS} is the channel current, I_{gen} is the avalanche-generated current, and I_{sub} is the substrate current …… 80

5.9 The avalanche-generation current (I_{gen}) and multiplication factor (M) for the variation of (a) the GDCS and (b) GSCS …… 81

5.10 To track the current gain of a lateral n-p-n transistor, the triggering voltage (V_{t1}) and holding voltage (V_h) are measured for the 1.5 V and 3.3 V NMOS transistors. The data clearly show that the current gain is reduced by increase in GSCS (a), while remains nearly constant with GDCS (b)…… 83

5.11 The base current (hole current) density vector at $I_D = 1$mA/μm for the device with (a) minimum gate to contact spacing, GSCS/GDCS=0.1 μm and (b) increased gate to source contact spacing, GSCS=0.5 μm. As GSCS increases, wider emitter and base junction is utilized for the current conduction and this results in a drop in the current gain due to increase in base current for a given collector current. The dashed arrow is the schematic of base current …………………………..……………………….… 84

5.12 (a) The current gain (β) vs. the drain current (I_D), and (b) the base current (I_b) vs. the generation current (I_{gen}) for the device with different gate-to-contact spacings ……... 85

5.13 The current flowlines and temperature distribution contours at the drain current of 1 mA/μm for the three different structures are shown. (a) GSCS=0.1 μm and GDCS=0.1 μm, (b) GSCS=0.1 μm and GDCS=0.5 μm, and (c) GSCS=0.5 μm and GDCS=0.1 μm……………………………………………………………………… 87

5.14 The temperature distribution along the *x* and *y* direction [as indicated in Fig. 13 (a)] at the drain current of 1 mA/μm for the three different structures: (a) vertical temperature distribution and (b) lateral temperature distribution……………………... 88

5.15 The simulated maximum temperature for the two different test structures with injected drain current. The maximum temperature (T_{max}) increases more rapidly as the power dissipating volume decreases for the shorter gate-to-source contact spacing. GDCS = 0.1 μm…………………………………………………………...… 89

5.16 Total failure threshold current (IT_2) for the two different test structures having different n^+ overlap lengths of S/D contact of 0.1 μm and 0.4 μm ($D_{_OL}$ and $S_{_OL}$). The inset shows the schematic of the test structure. GDCS/GSCS = 0.1 μm ……..…. 90

6.1 I_{t2} dependence on the channel length (L_{poly}) for the silicided 1.5 V NMOS devices. I_{t2} drops rapidly as the channel lengths enter the submicron regime ($L_{poly} < 1$ μm) ……. 95

6.2 I_{t2} dependence on the channel length for the non-silicided 1.5 V NMOS devices. I_{t2} dependence on the channel length agrees with the conventional bipolar transistor triggering model [9] ………………………………………………………..……. 95

6.3 I_{t2} versus the drive current (I_{drive}) for the 1.5 V NMOS transistors where W= 20 μm. The reverse channel length dependence of I_{t2} on the effective channel length for the silcided devices is clearly shown …………..………………………………………. 96

6.4 The V_{t1}/V_h with L_{poly} for the 1.5 V NMOS devices. It shows that the bipolar current gain (β) of both silicided and non-silicided devices increases with decrease in L_{poly} … 96

6.5 Finger width dependence of I_{t2} for the 1.5 V silicided and non-silicided devices where L_{poly}=0.175 μm ………………………………….………………………. 97
6.6 I_{t2} with finger width for the silicided 1.5 V devices with different channel lengths L_{poly}. For the smaller channel length devices, uniform conduction is not reached even at W=5 μm. This implies that the non-uniform conduction becomes more serious for shorter channel devices ……………………………..…….…………………..…… 98
6.7 I_{t2} with L_{poly} for the fully silicided 1.2V devices (W=20 μm) with process splits. (a) baseline (b) *As* supplement source/drain implant, and (c) *P* supplement source/drain implant ………………………………….………………………………….…….……… 99
6.8 The temperature contours 3-D distribution (insets) just before thermal failure (T=1680 K) for the devices with (a) L_{poly}= 0.175 μm, (b) L_{poly}= 0. 25 μm, and (c) L_{poly}=0.35 μm ……………..……………………………………………….…………… 100
6.9 The temperature distribution along the channel (*x*-direction) for devices with different channel length at thermal failure (at *y*=0.05 μm) …………….…………..… 101
6.10 The maximum temperature (T_{max}) with the power dissipation for the devices with the different gate lengths. The T_{max} increases with decrease in channel length for a given power ……………………………………….……….…………………….……………...... 101
6.11 The failure power P_f with L_{poly} for 1.5 V devices …..……………………………... 103
6.12 (a) approximated heat source region for a reverse biased NMOS transistor and (b) input power as a function of stress time, for a constant ΔT, across a reverse biased junction from [16, 73]…………………….………………….…………………..………. 104
6.13 The predicted effective finger width (W_{eff}) with L_{poly} at second breakdown for 1.5 V devices……………………………………………………………………….……...… 104
7.1 The comparison of *Boltzmann* approximation and *Joyce-Dixon* approximation [86] for determining the position of Fermi energy E_F for various doping concentrations as a function of temperature. The difference of Fermi energy between the two models increases as the doping concentration increases ……………………………………… 108
7.2 E_{00}/E_0 and k_BT/E_{00} versus doping concentration. It shows that the relative importance of field emission (FT) with respect to thermionic field emission (TFE) decreases as temperature increases for a given doping concentration ……………………………..….. 113

7.3 Theoretical curves for the specific contact resistance ρ_c with temperature for $CoSi_2$-contact system varying background doping concentration. The thick arrows indicate that the carrier transport mechanism changes from the TE to the TFE as temperature increases …………………………..………………….…………………….….. 113

7.4 Schematic of the contact of a MOS transistor for analyzing impact of the temperature dependent ρ_c on current localization effects where L_a is the distance from the spacer edge to contact hole, L_b is the width of the contact hole and L_c is the distance from the edge of the contact hole to the STI boundary ……………..……………………… 114

7.5 Fraction of the drain current as a function of the position from the edge of the spacer *x* [Fig. 7.1] where L_a=0.2 µm, L_b=0.2 µm, L_c=0.5 µm and W=1 µm. (a) highly doped source/drain (ρ_d=50Ω/□) and (b) moderately doped source/drain (ρ_d=500 Ω/□) ……. 117

CHAPTER 1

INTRODUCTION

1. 1 Motivation

Electrostatic discharge (ESD) is a charge re-balancing process between two adjacent objects, which involves a rapid discharge of accumulated static electricity. It is one of the most important quality and reliability concerns in the integrated circuit (IC) industry since a substantial number of IC failures (either catastrophic or latent failures) result from ESD/EOS[1] related modes [1-3]. According to the failure analysis results by R. Merrill et al. [2], ESD/EOS related causes are responsible for nearly 40 % of the failures of IC customer returns.

Therefore, to achieve higher quality and reliability standards for IC products and to reduce the IC product loss due to such ESD/EOS failure mechanisms, ESD phenomenon should be well controlled and corrective actions should also be taken through all phases of IC device's life. In order to obtain higher ESD robustness of IC chips, significant progress has been made in understanding the implications of different types of ESD events on the design of on-chip protection circuits, and in implementing highly effective ESD protection circuits at each device technology node.

However, the advent of new generation of device technology always leads to new challenges to higher ESD reliability and more efficient on-chip ESD protection circuits [4, 5]. Under ESD stress, the typical failure threshold current[2] [mA/μm] of ESD protection devices for industry standard process conditions and technology nodes is shown in Fig. 1.1. As can be seen, ESD protection levels of the devices are strongly process and technology dependent. This shows how ESD performance of protection devices has varied as the CMOS technology evolves across the nodes. For a certain ESD protection device, as a new process technology is introduced, the ESD level of the devices improves since ESD protection circuit design matures. However, with the introduction of each new process technology, ESD performance of the protection devices degrades since the past protection scheme is no longer valid for the new process technology. The same trend in variability of the ESD strength of the protection devices has been repeated through device technology generations and as a result, the design scheme of ESD protection has not

[1] Electrical Overstress.

[2] In the context, the failure threshold current means the second breakdown triggering current I_{t2}, which is discussed later.

simply transferred across the technology nodes. In a word, there is no universal protection strategy against ESD that transcends technology generations.

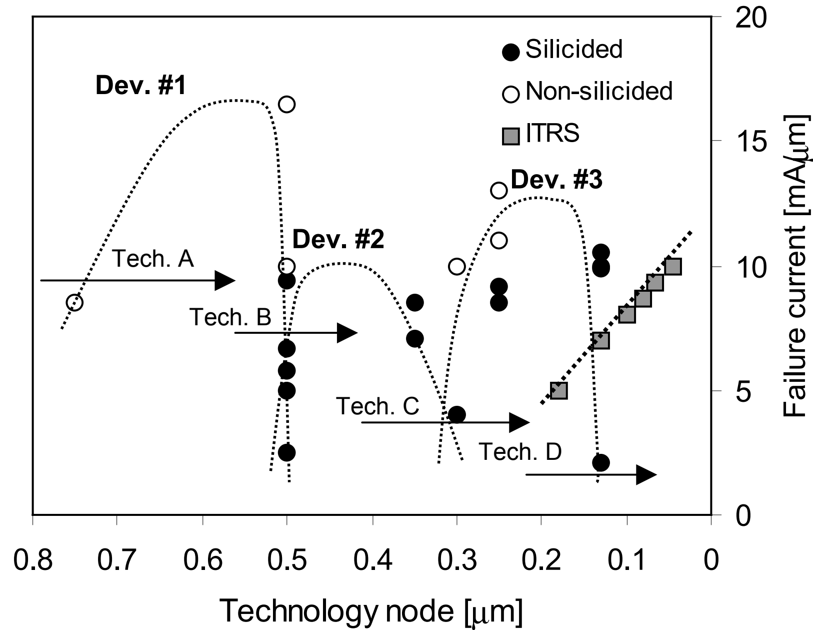


Figure 1.1: Reported typical failure threshold current of ESD protection devices (NMOS transistors) for various process conditions and technology nodes (data from Philips, TI, IBM and IMEC).

Therefore, as technology advances, new ESD protection designs are always necessary to restore the ESD levels to their previous performance level. Furthermore, the *International Technology Roadmap for Semiconductor* (ITRS) [6] predicts that the ESD performance of the protection devices for future IC chips should improve with the advancement of very-large-scale-integration (VLSI) technology. However, as can be seen, ESD performance data of the recent technologies in Fig. 1.1, in particular for the ESD level of a 0.13 μm advanced CMOS technology, show that there is a substantial gap between the required ESD levels from the ITRS[3] and the typical performance for an advanced CMOS technology (i.e., the 0.13 μm technology). This implies that as VLSI technology enters the nano-scale regime, associated technology features such as shallower junctions, thinner gate oxides, higher complexity of doping profiles,

[3] ESD levels were computed based on the human body model (HBM) ESD voltage values in the ITRS.

narrower width of metal lines and vias, and higher levels of interconnects [7] make the protection devices more vulnerable to ESD induced damage since the ESD stress does not scale independently from the device scaling. Furthermore, the ESD levels required for future VLSI applications look even more challenging, considering the ESD protection levels achieved at the current technology node.

Although significant progress has been made in understanding ESD and solving ESD related problems [4, 8-19], there are still knowledge voids in ESD phenomena observed in advanced technologies. Therefore, there is a growing need to comprehend ESD behavior of individual protection elements, in particular the NMOS transistor[4], and to develop the physical base modeling for ESD phenomenon in order to improve design guidelines and to develop effective protection strategies for higher ESD reliability in advanced CMOS technologies.

1. 2 What Is ESD?

In order to investigate implications of ESD events in IC technologies, it is also important to review the fundamental aspects of ESD phenomena. ESD is a subclass of the failure causes known as EOS[5]. The ESD events happen everywhere, in areas including: home environment, office, laboratories, wafer fabrication facilities, and assembly/test sites. ESD is the transfer of electrical charge between two bodies at different potentials, either through direct contact or through an induced electric field, so it is a charge driven physical mechanism resulting from a charge imbalance.

Electrostatic generation can occur due to friction between different materials, which is called a *triboelectric* charging and the potential induced by charges depends on the triboelectric property of materials. Therefore, people as well as equipment are able to easily generate ESD events during all stages of device manufacturing. For instance, the simple act of walking across a vinyl floor can generate up to 12000 V of static electricity depending on the relative humidity (RH) [20]. The accumulated charge due to ESD is discharged in typically 0.2 to 200 ns when the object contacts an effectively grounded object. ESD is an extremely fast event, and occurs from various sources, including human beings, machines, electromagnetic, nuclear and other harsh environments.

---

[4] An NMOS transistor is the commonly used protection element and can be used as a technology monitor of ESD performance in advanced CMOS technology.
[5] Electrical overstress typically lasts longer than 1µs.

Hence, depending on the source of the ESD event, various stress modes have been developed and routinely used for testing in IC industry. The discharge of the static electricity gives a harmless shock to human but is lethal to ESD sensitive microelectronic components since the associated current level with discharge event reaches up to 10 A. Such high ESD currents can easily cause a thermal failure of devices and have threatened reliability of operation for IC chips in numerous applications.

1. 3 On-Chip ESD Protection

It has been reported that ESD costs the microelectronics industry billions of dollars each year in damaged components, non-functional circuit boards and loss of information. As demands for higher performance and levels of integration on the IC chips increase, the minimum feature size of the devices in IC chips has to shrink. This device scaling also drives the development of new process technologies and a reduction of device size making the IC chips more susceptible to ESD induced failures.

Along with the issues of device miniaturization as well as new process technology, ESD concerns are critically important to meet reliability criteria. The ESD related reliability problems can encounter during the manufacturing, shipping, receiving, and field handling of integrated circuits or computer boards. A malfunction caused by ESD stress in these components may be immediately noted or undetected for even years with causing an unpredictable field failure. Therefore, ESD is one of the most serious threats to realize higher reliability in microelectronics.

To increase manufacturing yields, reduce overall cost and improve the quality and reliability of IC products, ESD events should be avoided or protected against. Although the cost of implementing an effective design of ESD protection is not trivial, the costs incurred by IC product failures and customer dissatisfaction can be far more difficult and costly to deal with after beginning manufacturing.

The solutions for avoiding or reducing ESD failures include: 1) identifying and rectifying possible ESD sources, 2) identifying and undertaking adequate prevention measures while handling the ESD sensitive devices, 3) incorporating built-in ESD protection networks in devices and 4) providing awareness and training to users at all levels.

In general, there are two ways to reduce IC failures under ESD conditions: reducing the likelihood of the ESD event occurring and improving the robustness of the device against ESD.

The first approach focuses on reducing the amount of ESD induced charges and redistributing them through proper handling of devices and controlling the handling environments. The second approach is to implement on-chip protection circuits in order to improve the circuit robustness against ESD events by improving ESD performance of the individual circuit components.

As shown in Fig. 1.2, the generic ESD protection configuration for a bi-directional I/O circuit is illustrated, which is composed of the primary protection, secondary protection and power ESD clamp. In the process of designing ESD protection, it is important that the functional requirements of the particular I/O should be satisfied while providing ESD immunity to the necessary level.

The function of the primary ESD clamp is to protect the driver by limiting the I/O pad voltage below the failure level of the output driver through bypassing most of the ESD stress current to the power rail. Devices such as p-n diodes, well resistors, transistors, and silicon controlled rectifiers (SCR's) can be used as efficient voltage clamps in ESD protection circuits. The secondary ESD clamp is for auxiliary protection and the series resistors, R_s and R_{in}, can lower the drain voltage of the output NMOS transistor and gate voltage of the input receivers. Finally, the power ESD clamp network can absorb the ESD energy, protecting the devices in the I/O circuits from ESD induced damage.

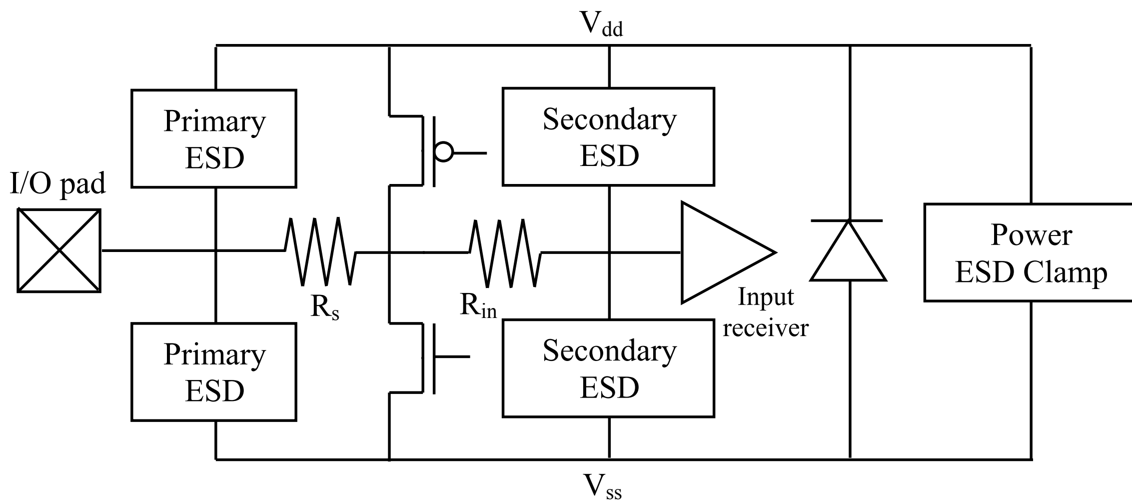


Figure 1.2: The generic configuration of the ESD protection circuit in a bi-directional I/O circuit. The diode from V_{ss} to V_{dd} represents the p-substrate to n-well diode inherent in any CMOS process.

Therefore, it is essential to understand the ESD phenomenon and its impact on ESD reliability of protection devices. Furthermore, it is also necessary to provide an insight into the design options in order to develop improved ESD levels of IC chips.

1. 4 ESD Protection Devices: Physics and Operation

For successful on-chip ESD protection design, it is crucial to choose a proper protection element. A good protection device should be capable of handling multiple ESD events without itself being damaged. In addition, it should not interfere with the operation of the protected circuits. Therefore, there are several requirements for an ideal protection device along with the above discussion [5, 21].

- Zero on-resistance; allows the device to shunt large amount of ESD current with no ohmic voltage drop.
- Finite clamping voltage; to avoid unintentional triggering of the protection device, the sustaining voltage of the protection device has to be higher than the supply voltage (V_{dd}) with a safety margin.
- Instantaneous turn-on time and infinite energy absorption; before the IC chip fails, protection device has to absorb full ESD induced energy as soon as ESD event occurs.
- Transparent to circuit operation (i.e., no parasitics); the protection device should activate only under ESD conditions and cause no parasitic effects on the functioning of IC chip.

As effective protection elements, resistors, diodes, NMOS transistors and silicon-controlled rectifiers (SCRs) are frequently used in standard CMOS technology [4]. The typical operating principles and characteristics of the protection devices in an ESD protection network are reviewed.

1. 4. 1 Resistor

A resistor is typically used to drop voltage or as an isolation mechanism in ESD protection networks as shown in Fig. 1.2. In addition, it can be sometimes used to add series resistance in order to ensure simultaneous triggering of multi-finger structures used for protection devices [8,

22]. Thin film resistors and diffused resistors are typical options available in standard CMOS processes. In general, the thin film resistors are made from polysilicon or alloys of NiCr or SiCr and fail primarily by fusing open. The ESD induced energy melts a region, causing a physical separation of the resistor terminating the current flow. The diffused resistors, which are made of a lightly doped diffusion region in the substrate, can handle large amounts of current. In this regard, the high current behavior of the diffused resistors has been investigated [23] and the typical I-V characteristics of the diffused resistors are shown in Fig. 1.3.

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

$$J = N_B q v_d \tag{1.1}$$

where N_B is the background doping concentration and v_d is the drift velocity. As shown, the diffused resistor has a region of the velocity saturation and the breakdown as the current level increases. With high electric field, the current density saturates and the saturation current density can be written by

$$J_{sat} = N_B q v_{sat} \tag{1.2}$$

where v_{sat} is the saturation velocity ($\approx 10^7$ cm/s).

The high current behavior of diffused resistors in view of ESD reliability has also been investigated [22]. With higher ESD currents, the resistor eventually enters permanent thermal failure.

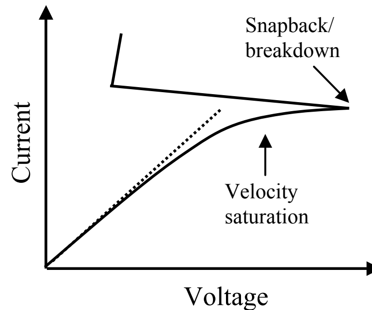


Figure 1.3: Typical I-V characteristics of a diffused resistor.

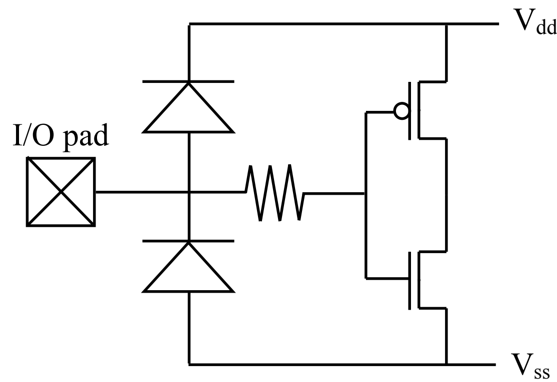


Figure 1.4: A typical diode ESD protection scheme.

1. 4. 2 Diode

The p-n diode or zener diode is the simplest voltage clamping device with minimum required area and is commonly implemented using reverse connections to I/O pad as shown in Fig. 1.4. The diode has a good power handling capability in the forward bias, but is a poor clamp with a high on-resistance in the reverse bias. One main disadvantage is the constant diode forward turn-on voltage that limits its applications to special circuit configurations and high V_{dd} cases. Multiple diode strings can also solve the problems of high voltage, but the current handling capability degrades due to increases in the on-resistance.

1. 4. 3 NMOS Transistor

In CMOS technology, NMOS transistor has a parasitic n-p-n lateral bipolar transistor buried within it as shown in Fig. 1.5 (a). Since the inherent parasitic bipolar transistor has a large current handling capability, NMOS transistor can bypass a significant amount of the ESD current in the snapback conduction mode. Using different bias conditions for better protection efficiency[6], NMOS protection devices can be implemented in the protection circuits with several configurations such as the gate-grounded NMOS (ggNMOS), the gate-coupled NMOS (gcNMOS) [24, 25], and the substrate-triggered NMOS (stNMOS) [26-28]. The ggNMOS is the most basic configuration among them and it has a self-protection capability against ESD.

[6] In general, ESD protection structure is composed of multi-fingers. Here, the protection efficiency means how the multi-fingers uniformly turn on under ESD stress.

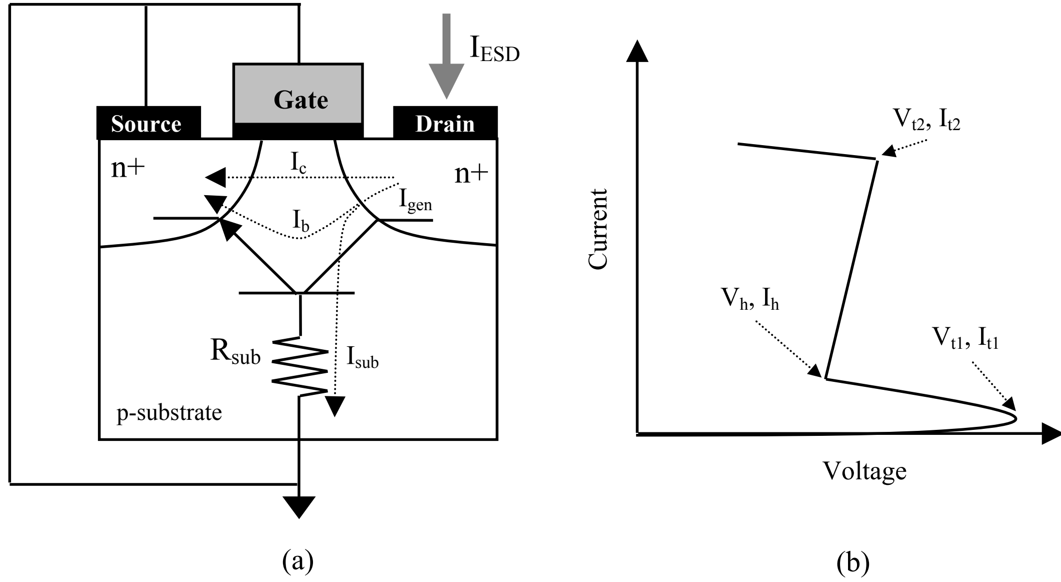


Figure 1.5: Typical operation of the gate-grounded NMOS (ggNMOS) (a) and I-V characteristics (b) under ESD conditions where I_{gen} is the avalanche generation current, I_{sub} is the substrate current, I_c is the collector current, I_b is the base current (i.e., $I_b=I_{gen}-I_{sub}$), V_{t1} (I_{t1}) is the triggering voltage (current), V_h (I_h) is the holding voltage (current), and V_{t2} (I_{t2}) is the second breakdown triggering voltage (current).

The operation of ggNMOS under ESD is shown in Fig. 1.5. Once the ESD event occurs, the ESD induced current is injected into the drain terminal of the ggNMOS. Then, a high electric field builds up across substrate to drain junction due to the associated reverse bias, resulting in an avalanche multiplication process. The avalanche generation current I_{gen}[7] in the high field region due to an incident collector current I_c is modeled by:

$$I_{gen} = (M-1) \cdot I_c \tag{1.3}$$

where M is an empirically determined multiplication factor. Avalanche generated electrons flow towards the drain contact while holes flow towards the substrate (i.e., I_{sub}) resulting in a voltage drop across the effective substrate resistance, R_{sub}.

As the substrate current, I_{sub} increases with increased drain current, the voltage drop across R_{sub} (i.e., V_{sub}) eventually reaches ~ 0.7 V, which is the turn-on voltage of the emitter-base (i.e., source-substrate) junction. The substrate voltage V_{sub} at this condition is given by

[7] $I_{gen}=I_{sub}$ before the transistor triggers and $I_{gen}=I_{sub}+I_b$ after the transistor triggers.

$$V_{sub} = I_{sub} \cdot R_{sub} \approx 0.7V \tag{1.4}$$

Finally, once the voltage drop V_{sub} reaches the turn-on voltage of the lateral n-p-n transistor, the transistor triggers and operates in a self-biasing mode. At this condition (i.e., V_{t1}), ESD current flows through the substrate bulk and the current driving capability of the ggNMOS is significantly increased compared to that for surface channel conduction. The regenerative condition for snapback to occur is given by [9]

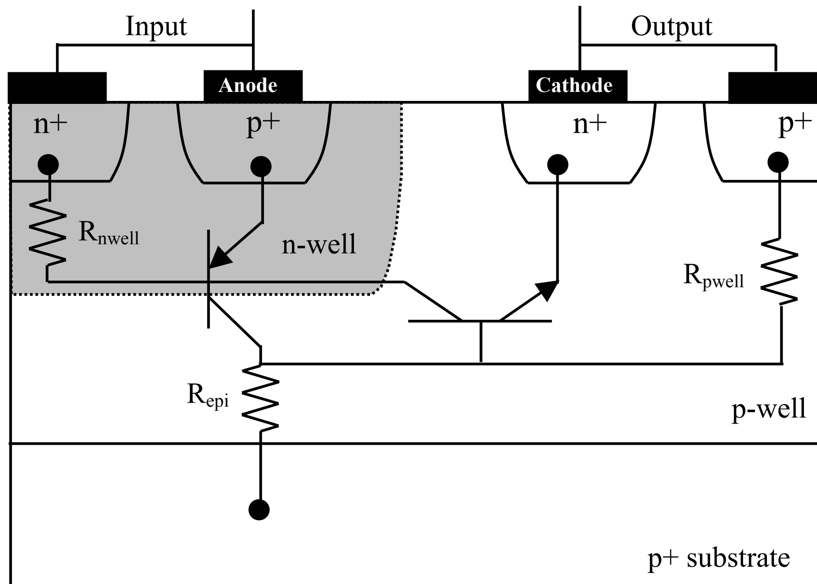
$$\beta \cdot (M - 1) \geq 1 \tag{1.5}$$

where β is the current gain of the parasitic lateral bipolar n-p-n transistor.

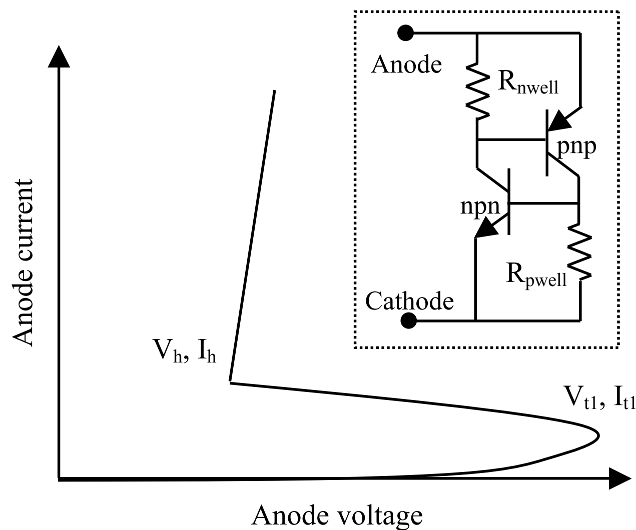
However, with further increases in the drain current, the ggNMOS enters the thermal breakdown regime at V_{t2} (I_{t2}), which is called *second breakdown*. Second breakdown is generally considered to be caused by current localization due to the negative resistance coefficient of silicon beyond a critical temperature, T_c at which the thermally generated carrier concentration is equal to the background doping concentration. It is also shown that second breakdown occurs when the heat produced by an increase in current generates the exact amount of minority carriers required to support the increase in current without raising the electric field [29]. The second breakdown results in the irreversible thermal damage to the device giving rise to a large junction leakage current.

1. 4. 4 Silicon Controlled Rectifier (SCR)

The silicon controlled rectifier (SCR) can serve as an excellent ESD protection device since it has a very high current handling capability arising from the combination of two bipolar transistors in a self-regenerative on condition. Fig. 1.6 shows the cross sectional view of SCR and its typical high current I-V characteristics. As can be seen, the SCR is a combination of the vertical p-n-p and lateral n-p-n bipolar transistors. When used as a protection device, it is configured as a two-terminal device; the cathode and p-well are tied together and the anode and n-well are also tied together as shown in the inset of Fig. 1.5 (b). The I-V curve of the two-terminal SCR shows a behavior similar to that of the ggNMOS. When ESD stress is applied to the anode, the n-well/p-well junction is forced into reverse bias until it reaches breakdown. As discussed for the ggNMOS, avalanche generated hole current flows toward the p-well contact which also forward biases the base of the lateral n-p-n transistor, eventually reaching its turn-on point.



(a)



(b)

Figure 1.6: (a) Cross section of a lateral SCR in a CMOS process showing the parasitic p-n-p and n-p-n transistors. (b) High current I-V curve for an SCR (the inset shows the equivalent circuit schematic of the SCR).

At this point, the base of the vertical p-n-p transistor is also biased by the current of the n-p-n from the n-well to the cathode and the anode voltage begins to decrease resulting in a negative differential resistance region (NDR). When the SCR is in the latched mode (regeneratively held in the "on" state), the required condition for sustaining this condition is given by [4]

$$\beta_{npn} \cdot \beta_{pnp} \geq 1 \tag{1.6}$$

The holding voltage, V_h is the minimum voltage required for this conduction state and the SCR is operating with extremely low power consumption leading to a very high ESD robustness. However, despite of the excellent ESD performance, one major drawback of the SCR protection is that a large triggering voltage is needed in advanced CMOS processes; the breakdown voltage of the well-substrate junction is about 20 V, depending on the well doping profile and substrate doping level.

Therefore, this large triggering voltage should be lowered to provide good protection capabilities. Inserting an additional device structure between the n-well and p-well, reduction of the breakdown voltage of n/p-well can be achieved [4], but it is generally difficult to obtain the triggering voltage (V_{t1}) below 10 V in typical advanced CMOS technology.

1. 5 Thesis Outline

This thesis is designed to provide detailed insight into ESD behavior involved in advanced technologies and to develop design guidelines for deep submicron ESD protection based on both device simulation and intensive experimental studies of the protection NMOS transistors used at Texas Instruments Inc. ESD behavior has been investigated using single finger NMOS test structures, which were manufactured by an advanced 0.13 μm CMOS technology[8], under various bias conditions as well as variations of design parameters. The organization of the thesis is as follows:

Chapter 2 presents a general overview of the characterization of the ESD phenomena as preliminary information for better understanding of the dissertation. The types of ESD events, ESD related failure modes, and analysis techniques for ESD behavior are discussed. Finally, ESD related process information is also provided.

Chapter 3 presents a detailed study of the non-uniform bipolar conduction phenomenon under electrostatic discharge (ESD) events in *single finger* NMOS transistors and analyses implications for the design of ESD protection for deep submicron CMOS technologies. It is shown that the uniformity of the bipolar current distribution under ESD conditions is severely degraded and is identified as a root cause of the severe reduction in ESD performance for advanced silicided technologies.

Chapter 4 investigates the substrate and gate bias dependencies of ESD robustness of advanced NMOS transistors. It is shown that the substrate bias can alleviate the non-uniform conduction so that it makes the design of ESD protection more practical by enlarging the effective turned-on width of the device. Also, it is presented that the gate bias induced heating is a root cause of the degradation of ESD performance for strongly gate-coupled protection devices.

Chapter 5 presents a detailed investigation of the influence of gate-to-source and gate-to-drain contact spacings for the salicided 0.13 μm CMOS technology. It is shown that the reduction in current localization and the increase in the power dissipating volume with increase in the gate-to-contact spacings are primary causes of the improvement of ESD immunity even for the devices with silicided diffusion.

---

[8] The information of process technology is given in Chapter 2.

Chapter 6 introduces the reverse channel length dependence of ESD performance for advanced silicided NMOS transistors. It is shown that the reverse channel length dependence of ESD performance is mainly due to severe non-uniform lateral bipolar conduction, which reduces the effective device width. Furthermore, it is demonstrated that substrate bias can be effective in controlling this reverse channel length effect.

Chapter 7 presents a model of the temperature dependent specific contact resistance that can be employed for analyzing the impact of silicide diffusion on ESD hardness. It is shown that the specific contact resistance for the silicide-silicon contact system strongly depends on temperature and thus the variation of the specific contact resistance value has significant implications for ESD reliability through localized heating by current crowding effects since the effective volume for power dissipation is reduced with the extent of the current localization for a given ESD stress.

Finally, Chapter 8 summarizes the contributions of the dissertation and discusses recommended researches for future generations of IC technologies.

CHAPTER 2

CHARACTERIZATION OF ESD PHENOMENA

2. 1 Types of ESD Events

According to the source of the ESD stress condition, the IC industry has standardized on three basic models related to ESD phenomenon. The ESD stress models are intended to reproduce typical discharge pulses to which the IC chips may be exposed during manufacturing or handling. In general, ESD induced damage is caused by one of three ESD events: direct electrostatic discharge to the device, direct electrostatic discharge from the device, or field induced discharges.

The equivalent circuit for each ESD event can be represented by modeling the discharge current waveforms using RLC elements shown in Fig. 2.1 [30], and idealized current waveforms for each ESD event are also shown in Fig. 2.2 [31]. In practice, using the models based on the real-world ESD events, ESD robustness of devices under different discharge processes can be systematically characterized.

2. 1. 1 Discharge to the Device

An ESD event can occur when any charged conductor (including the human body) discharges to an ESD sensitive device. The most common cause of electrostatic damage is the direct transfer of electrostatic charge from the human body or a charged material to the electrostatic discharge sensitive device due to improper handling of IC devices. When one walks across a floor, an electrostatic charge accumulates on the body. Simple contact of a finger to the leads of an ESD sensitive device or assembly allows the body to discharge, possibly causing device damage. The model used to simulate this event is the Human Body Model (HBM). This HBM is the most common industrial test method to measure ESD robustness and consists of a charging capacitor, and a discharging capacitor through a finger resistor (~ 1.5 kΩ) into an I/O or a supply pin of a packaged IC chip with another pin grounded and all other pins floating [Fig. 2.1 (a)]. The capacitor and resistor values are selected to create a pulse similar to that occurring due to an

electrostatically charged human digit. The HBM current waveform has a rise time of a few nanoseconds and a decay time of about 150 ns.

As a HBM event, a similar discharge can occur from a charged conductive object, such as a metallic tool or fixture. The model used to characterize this event is known as the Machine Model (MM). In the MM, a capacitor is charged up to a high voltage and then discharged through the pins of an IC chip, resulting in an oscillatory current waveform as shown in Fig. 2.2. Since in the case of MM, the parasitic series resistance is very small, the parasitic inductance and capacitors of test equipment as well as the dynamic impedance of the device under test have a significant impact on shaping the current pulse.

Therefore, unlike the HBM, the standardization of MM is more difficult. But the failure signature is generally the same for HBM and MM since discharge occurs through the devices. Due to the similarity of discharge process, MM ESD robustness could be guaranteed by HBM test in some cases. For instance, 100 V MM is generally equivalent to 2 kV HBM in IC chips [32].

2. 1. 2 Discharge from the Device

The transfer of charge from another device can also result in an ESD event. Static charge may accumulate on the device itself through handling or contact with packaging materials, work surfaces, or machine surfaces. This frequently occurs when a device moves across a surface or vibrates in a package. The model used to simulate the transfer of charge from a device is referred to as the Charged Device Model (CDM). The capacitance and energies involved in this phenomenon are different from those of a discharge to the device.

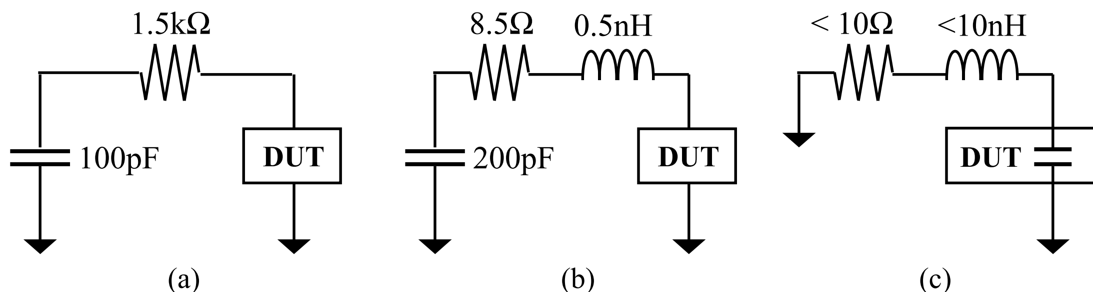


Figure 2.1: Simplified typical ESD stress models (after [30]). (a) Human Body Model (HBM), (b) Machine Model (MM), and (c) Charged Device Model (CDM)

While thermal destruction is the primary failure mechanism in the HBM and MM, dielectric failure is the typical failure mode in the CDM type ESD. Since as shown in Fig. 2.2, CDM event is so fast that the protection device may not be able to turn on or clamp the input voltage to a safe level during such a short time period (i.e., ≤ 1 ns) without dielectric damage having already occurred.

Although having an automated assembly system would seem to solve the problems of HBM ESD events, it has been shown that components may be more sensitive to ESD related damage when assembled by automated equipment. Therefore, significant work still needs to be done to understand the mechanisms of the CDM and to prevent the CDM related ESD failures.

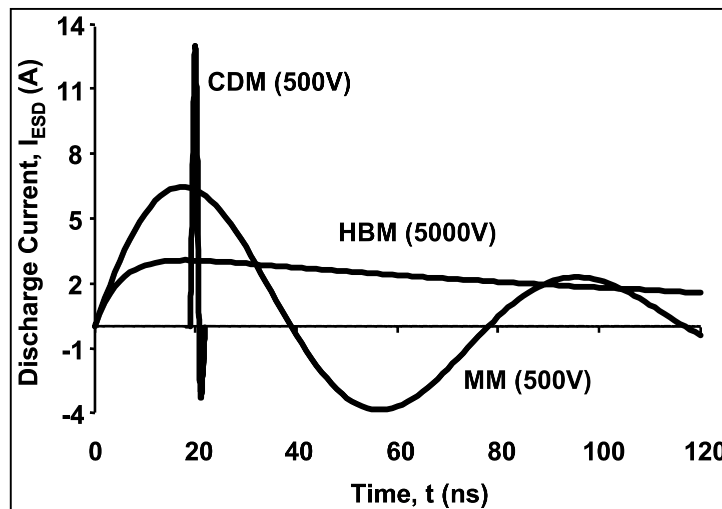


Figure 2.2: RLC discharge current waveforms of the three basic ESD stress modes (after [31]).

2. 1. 3 Field Induced Discharge

Any charged object can generate electrostatic fields. If a device is placed in that electrostatic field, charge may be induced on the device. As a CDM event, discharge occurs from the device when it is temporarily grounded in the field. This type of ESD event also can directly or indirectly cause damage to devices.

2. 2 ESD Related Failures

Electrostatic discharge can cause damage to microelectronic devices at any stage from manufacturing to field application. The damage to devices results from handling the devices in inappropriately ESD controlled environments. The damage to the devices by an ESD event is determined by the device's ability to dissipate the energy of the discharge or withstand the voltage levels involved. The currents induced by ESD events are extremely high. The HBM-generated current peaks, for example, are in excess of 2 A and the MM and CDM-generated current peaks are even higher than this. Since these current levels are in excess of the normal operating currents[9] for the circuits, the currents can directly or indirectly cause physical damage observed in ESD failures.

In order to design IC chips with robust ESD strength, it is important that the main design and process parameters, which influence the functions of IC chips under ESD stress, should be known. Therefore, the electrical and physical failure modes caused by ESD as well as underlying failure mechanisms should be fully analyzed [33-38]. In general, device failures under ESD conditions are classified as either causing a catastrophic failure or creating latent defects that impair functionality of the devices.

2. 2. 1 Catastrophic Failure

When an electronic device is exposed to an ESD event, the device may no longer operate properly. The ESD event may have caused a metal melt, junction breakdown, or oxide failure. In this case, the on-chip device is permanently damaged due to power dissipation during the ESD event. Such failures can be electrically sorted when the device is tested before shipment. However, ESD damage that occurs after test goes undetected until the device fails during field operation. The failure mechanisms are either current- or voltage-induced damage mechanisms. These mechanisms are illustrated schematically in Figs 2.3 and 2.4.

Current-induced damage result in thin film resistor burnout, junction filamentation and junction spiking. The temperature rise by Joule heating under ESD condition can melt and damage the thin film resistor; at this point the resistor exhibits a permanent failure. Filamented junction damage can be observed based on the degraded current-voltage characteristics, which show an abnormal increase in leakage current across the junction. In the extreme case of junction

[9] The normal operating currents refer to ones used for analog or digital applications.

filamentation, junction spiking occurs with a melted metal flow as a result of ESD-induced power dissipation. On the other hand, ESD induced voltage can cause either charge injection or rupturing of the oxide. The device failure due to charge injection cannot be observable by typical de-processing techniques since the charge state of dielectric materials change with the charge injection.

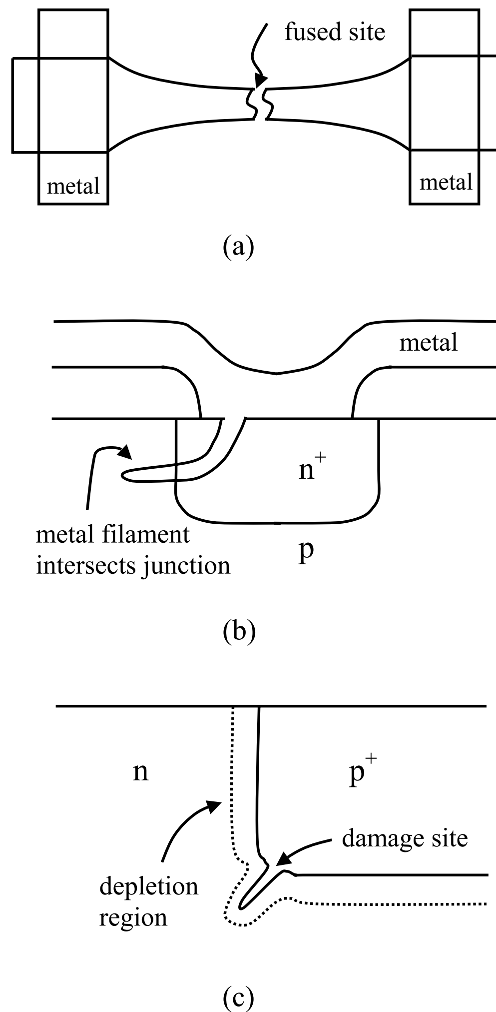


Figure 2.3: Current induced damage mechanisms under ESD conditions (after [5]). (a) Thin film fuse, (b) Junction filamentation, and (c) Junction spiking

Trapped charges are reversible, using either an unbiased bake or high energy ultraviolet light irradiation. The junction associated with trapped charge also shows leaky current-voltage characteristics, but it differs from junction filamentation since the junction damage cannot be recovered by baking. With increased amounts of charge, the dielectric material ruptures and forms a resistive current path.

2. 2. 2 Latent Damage

On the other hand, the latent damage due to ESD stress is more difficult to identify since a device can continue to perform its intended function independent of partial or subtle degradation. However, a product containing such a device with latent damage may experience premature failure or a malfunction.

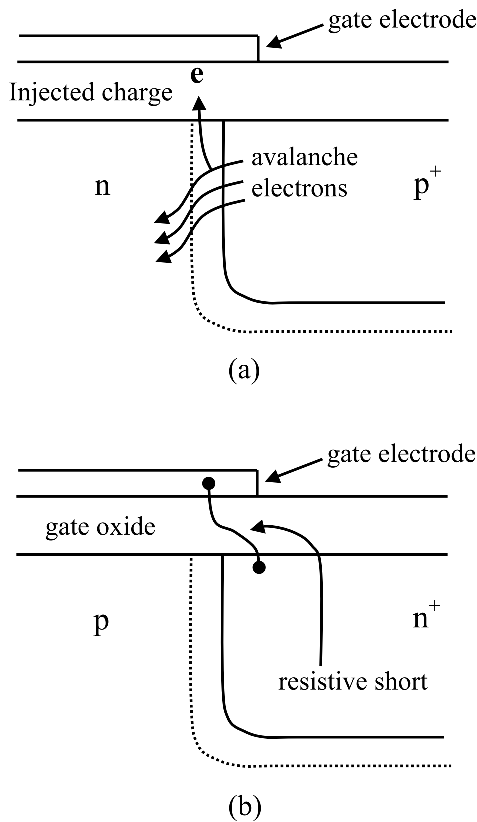


Figure 2.4: Voltage induced damage mechanisms under ESD conditions (after [5]). (a) Charge injection, and (b) Oxide rupture

Compared to latent damage, it is a relatively easy task with an appropriate test equipment to identify catastrophic failure of the protection device since the catastrophic damage can be substantiated using basic performance tests such as leakage current measurements.

However, latent defects are extremely hard to detect, especially after the device is assembled into a finished product. Although it is still debated whether any significant increase in the leakage current should be considered as a failure or potential failure, in most cases, leakage current has been attributed to more or less latent damage in p-n junctions or dielectric layers. Therefore, the failure criterion based on the leakage current level is currently widely used depending on the process technology. In this regard, it has been shown that the distribution of leakage current at a given stress level is linked to the distribution of the ESD failure threshold and indications of leakage current can be used to characterize the ESD behavior of an IC [4, 38, 39].

2. 3 Experimental Techniques

In order to investigate and characterize ESD behavior of protection devices, as well as the impact on ESD hardness, the transmission line pulsing (TLP) [40] and photon emission microscopy (EMMI) techniques have been used throughout this work.

2. 3. 1 Transmission Line Pulsing (TLP)

Real-world ESD events occur within the time period of a few hundred nanoseconds. Therefore, it is important to characterize behavior of devices in the ESD relevant times. Since DC measurements generally cause strong self-heating for higher currents and do not address transient characteristics of devices, pulsed measurement techniques are required for ESD characterization. In this sense, the TLP measurement [40] is commonly used to monitor the ESD hardness and provide dynamic and quasi-static high current behavior of protection elements.

The schematic of a TLP system and the principle of pulsed characterization are illustrated in Fig. 2.5 (a). A transmission line is charged to given input voltage and then discharged through the device under test when the switch closes. Duration of the discharged voltage pulse is proportional to the length of the transmission line[10], and thus the pulse width as well as amplitude of input voltage to the device can be simply controlled by the physical length of the transmission line and

---

[10] The duration of the discharging voltage pulse is $2L/\upsilon$ where L is the length of the transmission line and υ is the phase velocity of the line.

the initial charging voltage. As shown in Fig. 2.5 (b), the automated TLP system is more commonly used in the IC industry, which is more efficient in measurement as well as analysis of data.

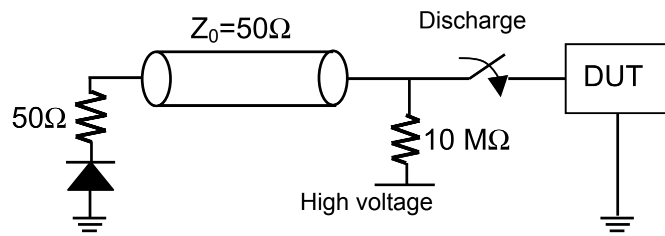
The I-V snapback curves and waveforms of the ggNMOS from TLP measurements are also shown in Fig. 2.6. Using a pulse generator, increased voltage pulse trains are applied to the drain of the ggNMOS transistor. At each voltage pulse (i.e., during t_d), using an oscilloscope, the drain current and drain voltage waveforms are measured. After each I-V measurement (i.e., during the off time), using the HP4156[11], leakage current is measured to see whether the device under test is failed. After complete the I-V measurements and recorded the I-V data, high current I-V snapback curves can be reconstructed as illustrated in Fig. 2.6 (a). As briefly discussed in Chapter 1, the TLP test [Fig. 2.6 (a)] clearly shows the typical I-V snapback curve of ggNMOS.

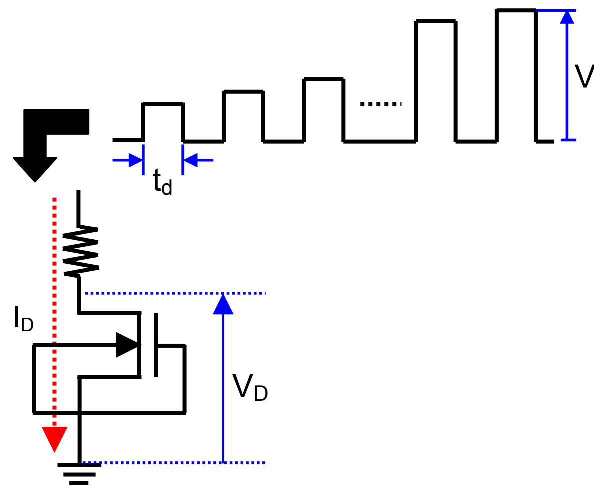
Before triggering of the ggNMOS, only the voltage waveform (with minor current leakage) can be observed [Fig.2.6 (b)]. However, after triggering, the voltage and current waveforms clearly show the triggering (V_{t1}, I_{t1}) and holding points (V_h, I_h) [Fig. 2.6 (c)] with observation of appreciable current values. With further increases in the voltage pulse to the drain, ggNMOS finally goes into the second (thermal) breakdown regime [Fig. 2.7 (d)]. This can be clearly seen from the abrupt changes in the voltage and current waveforms.

Second breakdown triggering current I_{t2} from TLP test can generally be used as a technology monitor of ESD robustness for a given technology although the correlation between data from TLP measurement and results for the HBM, MM, and CDM tests are still under discussion [38, 39]. However, recent work [37] also shows that there is a good correlation as long as the device is damaged in the hard failure mode.

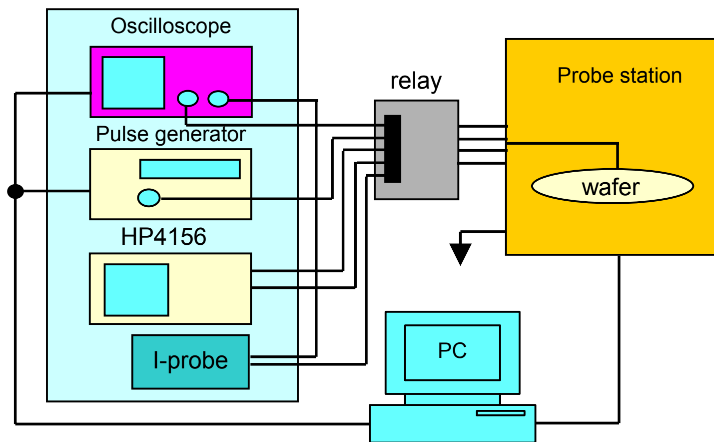
Based on the high current snapback curves of the protection device, the ESD design window for the protection device [21, 41, 42] can be obtained as shown in Fig. 2.7. The triggering voltages (V_{t1} and V_{t2}) of the protection device must be kept lower than the gate oxide breakdown voltage (BV_{ox}), which is dependent on the thickness of the gate oxide as well as the duration of the ESD stress, in order to avoid dielectric failures. On the other hand, to avoid unintentional triggering of the protection elements, the holding voltage (V_h) must be higher than the supply voltage (V_{dd}) plus a safety margin due to noise or voltage overshoot.

[11] The HP4156 is a typical semiconductor parameter analyzer.



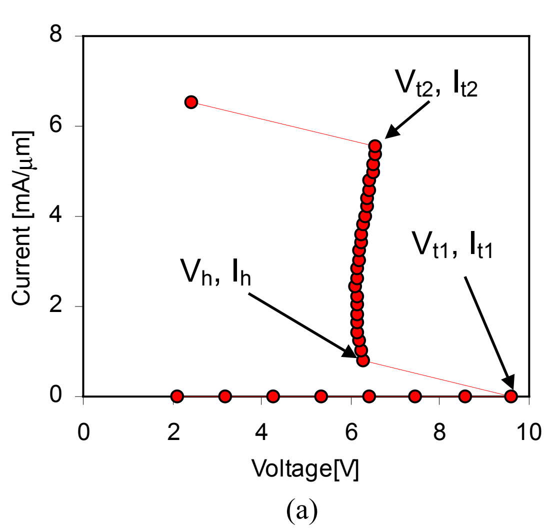


(a)

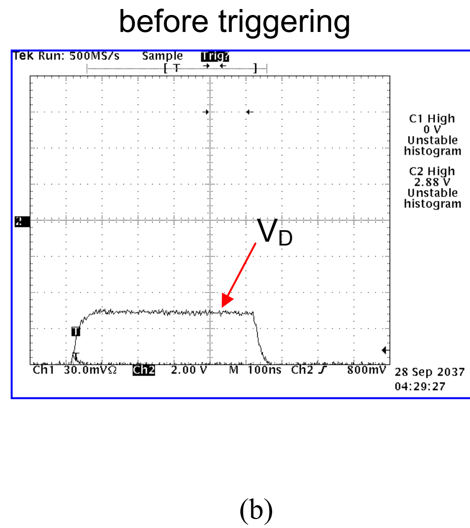


(b)

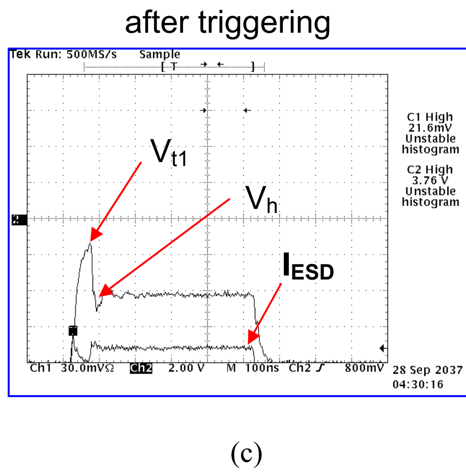
Figure 2.5: Schematic of a TLP setup. (a) Circuit representation of the TLP and principle of pulsed characterization. (b) ESD test setup of the automated TLP.



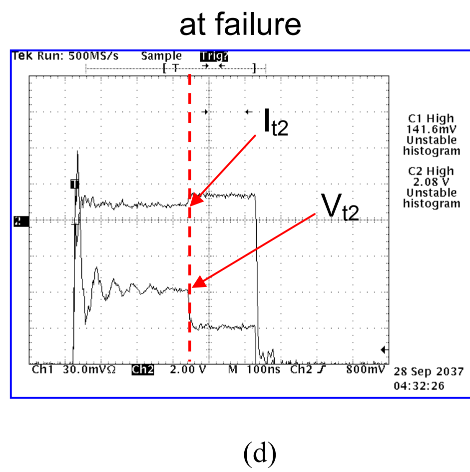
(a)



(b)



(c)



(d)

Figure 2.6: (a) High current snapback curves from the TLP measurement. The drain voltage and current waveforms (b) before triggering, (c) after triggering, and (d) at failure.

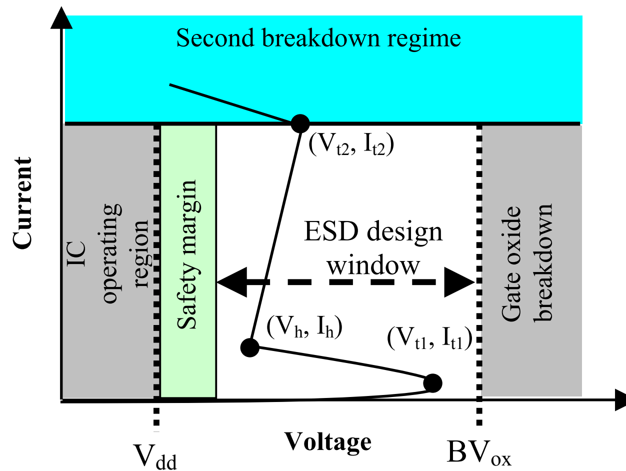


Figure 2.7: ESD design window for an ESD protection element.

2. 3. 2 Emission Microscopy (EMMI)

Photon emission microscopy is widely used as a failure and reliability analysis technique for semiconductor devices. Sensing photon emission during current conduction, the location of defects and failure sites in IC chips can be observed. These analyses have typically been performed by collecting visible and near infrared (390 ~ 1000 nm) photons emitted under device operation.

The general mechanisms of photon emission in devices are electron-hole recombination, intraband transition, thermal radiation and tunneling currents in the oxide. For MOS transistors, the radiative intraband transition is the predominant emission mechanism under high electric fields and currents. Thereby, generated photons, transmitted through relatively transparent dielectric layers, passing between or scattered around the patterned, and opaque metal interconnections, are detected. The analysis based on the photons emerging from these overlying layers is referred to as frontside light emission microscopy analysis. Correspondingly, imaging light passing through the silicon substrate and emerging from the bottom is referred to as backside light emission analysis.

Therefore, under ESD conditions the spatial distribution of current, which is commensurate with the intensity of emitted photons, can be also observed. In addition to high current behavior of devices, it is possible to use EMMI to pinpoint failure location of devices. In this sense,

investigations of ESD behavior using EMMI analysis are available [43-46] and it should be emphasized that EMMI is an invaluable tool in the study of ESD phenomenon in IC chips and in the design debugging process for IC chips [46].

Commercial EMMI systems are normally used for ESD reliability analysis in IC industry. In this work, the FA-1000 model EMMI tool, manufactured by Alpha Innotech Corporation, was used and the equipment setup is shown in Fig 2.8.

For analyzing ESD behavior with EMMI, it is important to use a pulsed bias since the real-world ESD phenomenon is a very short time event[12]. Therefore, the EMMI analysis with DC bias conditions cannot provide reasonable indications of ESD behavior of semiconductor devices and it should be only used for detecting the failure location of damaged IC chips.

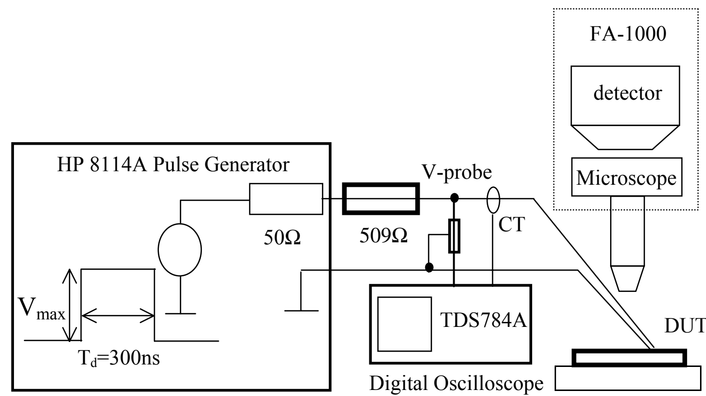


Figure 2.8: EMMI setup used in this work for observing ESD behavior.

[12] In general, an ESD event lasts less than a few hundred nanoseconds.

2. 4 Technology under Investigation

Analysis of ESD performance for NMOS transistors presented in this thesis are carried out using the 0.13 μm mixed voltage CMOS technology[13] for high performance logic circuits, developed by Texas Instruments Inc.

Summary of ESD related process information of the technology under investigation is documented in Table 2.1. In addition, the TEM[14] image of the 1.5 V NMOS transistor fabricated using the 0.13 μm technology is shown in Fig. 2.9.

Table 2.1: Summary of ESD related process information of the technology under investigation.

Technology node	0.13 μm mixed-voltage technology	
Supply voltage	1.5 V	3.3 V
Starting material	3.5 μm thick p-type epi layer on p-type substrate (100)	
Device isolation	Shallow trench isolation (STI)	
Well process	Retrograded n-well and p-well are implanted and followed by RTA[15] process	
Gate oxide thickness	27 Å	70 Å
Source/Drain engineering	Highly doped drain (HDD) Pocket implantation	Lightly doped drain (LDD) No pocket implantation
Junction breakdown voltage	~ 6 V	~ 9 V
Silicide	$CoSi_2$ salicide[16]	

[13] The 1.5 V and 3.3 V NMOS transistors are investigated.
[14] Transmission Electron Microscopy.
[15] Rapid Thermal Anneal.
[16] Self-aligned silicide.

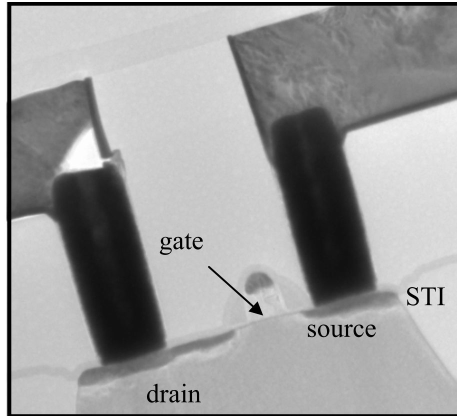


Figure 2.9: TEM image of the low voltage (1.5 V) NMOS transistor using the 0.13 μm CMOS technology.

The test structures used in this study are single finger NMOS transistors although the multi-finger NMOS studructures are generally used as protection elements in practice. Since the multi-finger structure experiences non-uniform triggering under ESD stress, for estimating ESD performance of a given process technology and design, analysis of the single finger structure is more informative. Varying layout parameters of the single finger structures as well as silicide process, ESD robustness of the devices are analyzed through this thesis and the layout dimensions of the test structure are given later.

CHAPTER 3

NON-UNIFORM BIPOLAR CONDUCTION UNDER ESD CONDITIONS

This chapter presents a detailed study of the non-uniform bipolar conduction phenomenon under electrostatic discharge (ESD) events in *single finger* NMOS transistors and analyzes its implications for the design of ESD protection for deep submicron CMOS technologies [47, 48]. It is shown that the uniformity of the bipolar current distribution under ESD conditions is severely degraded depending on device finger width (W) as well as silicide processing. This non-uniform current distribution is identified as a root cause of the severe reduction in ESD failure threshold current for the devices with advanced silicided processes.

3. 1 Introduction

For on-chip ESD protection circuits, the first requirement for achieving good protection structures is to provide a low-impedance discharging current path to shunt ESD currents and clamp the I/O pad voltage to a safe level without causing damage to internal circuits. The size of protection devices is mainly determined based on the current handling capability required in the ESD protection circuits. The multi-finger structures are the most common way of designing the protection devices in various sizes due to layout efficiency.

In multi-finger structures, uniform triggering across all the fingers is important to achieve maximum current handling capabilities. This requires that the single fingers in the multi-finger structures are fully turned on under ESD conditions. Typically, multi-finger gate-grounded NMOS (ggNMOS) devices are widely used as protection structures owing to the effectiveness of parasitic lateral n-p-n bipolar transistors in handling high ESD currents.

Non-uniform triggering behavior of lateral n-p-n transistors was first reported by Scott et al. for silicided single finger NMOS devices under ESD stress [49], and subsequently the phenomenon of non-uniform triggering in multi-finger NMOS transistors and the implications for the design of ESD protection was discussed by Polgreen et al. [8]. The simultaneous triggering of

the multi-finger NMOS protection structures has been considered as a critical aspect for the effectiveness of ESD protection designs [24-26].

However, in recent years the dependence of the second breakdown triggering current (I_{t2}), which is widely used as the figure-of-merit of ESD robustness of a protection structure, on the single finger device width (W) has become quite an important criterion for optimum design of ESD protection circuits, since under ESD events current localization occurs in advanced silicided single finger transistors [44, 45].

We have recently addressed this issue for the designs of protection devices involving advanced submicron technologies [47, 48]. In this study, detailed experimental investigations into the non-uniform current conduction phenomenon have been performed including transmission line pulsing (TLP) measurements and emission microscopy (EMMI) analysis for various silicided and non-silicided test structures. Moreover, the impact of both substrate and gate bias conditions[17] on this non-uniform current distribution and their implications for the design of ESD protection have been discussed in detail, which provides new physical insight into the ESD behavior of advanced protection devices and useful basis for constructing efficient design windows for robust ESD protection design to overcome early ESD failures in advanced deep submicron technologies.

3. 2 Experimental Evidence of Non-Uniform Bipolar Conduction

In order to investigate the spatial current distribution of a single finger protection device under ESD conditions, various test structures including 1.5V and 3.3V NMOS transistors with shallow trench isolation (STI) in a 0.13 μm CMOS technology were investigated. Both silicided[18] ($CoSi_2$) and non-silicided devices, which were formed on a 3.5 μm thick epi layer, were also explored for comparison. The layout of the single finger NMOS transistor is illustrated as shown in Fig. 3.1. The drawn poly gate length (L_{poly}) was 0.175 μm and 0.5 μm for the low voltage and high voltage transistor, respectively, and the substrate contact was located parallel to the source contact to keep the substrate resistance constant along the finger width direction.

[17] Impact of bias conditions on ESD is discussed in Chapter 4.

[18] The source/drain and gate are silicided.

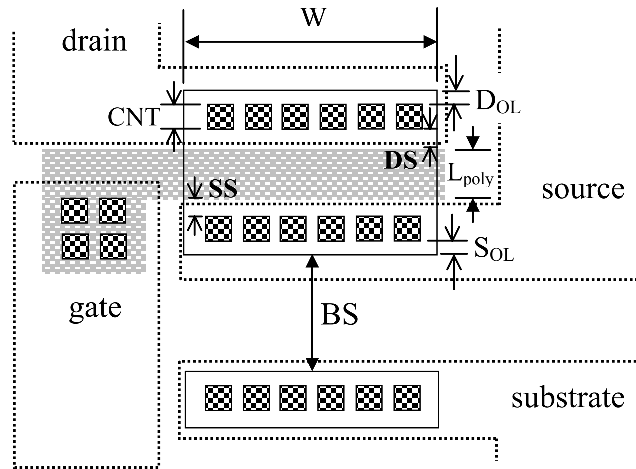
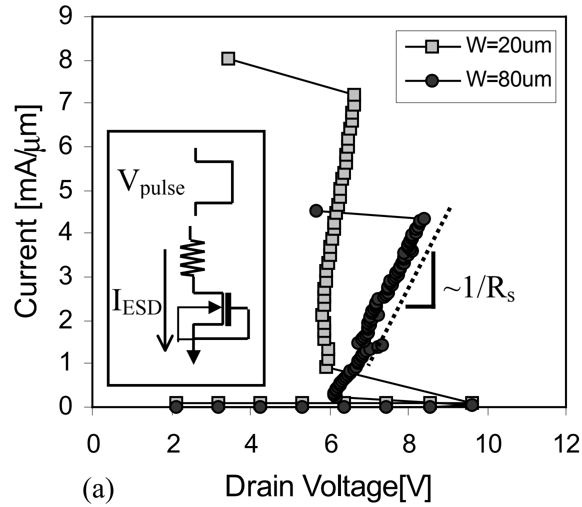


Figure 3.1: The layout of the single finger NMOS transistor. The contact opening (CNT) is 0.15 μm for both 1.5 V and 3.3 V devices. The gate to source/drain contact spacings (SS/DS) are 0.1 μm and 0.225 μm, and n+ overlaps of source/drain contact (S_{OL}/D_{OL}) are 0.4 μm and 0.125 μm for the 1.5 V and 3.3 V transistors, respectively. Also the body space (BS) from the source diffusion to the substrate diffusion is the same as the finger width (W).

3. 2. 1 Transmission Line Pulsing (TLP) Tests

As discussed in Chapter 2, the high current behavior of protection devices can be analyzed by applying a short time-scale constant-current pulse, generated using a transmission line, to protection structures with increased pulse magnitude at each step [40]. Using the TLP system (as also shown in Chapter 2), I_{t2} measurements were performed with a 200 ns long voltage pulse for various test structures, and sampled high current I-V snapback curves for the 3.3 V silicided and non-silicided devices are shown in Fig. 3.2.

As can be seen, the single finger device shows a significant width dependence of I_{t2} as measured in [mA/μm], and the effect becomes more apparent with silicide process. According to the I-V curves, the application of silicided technology shows no significant differences in the n-p-n transistor triggering voltage (V_{t1}), since the triggering voltage V_{t1} is dominantly determined by the avalanche multiplication process across the drain-substrate junction and both devices were drain engineered in the same manner.



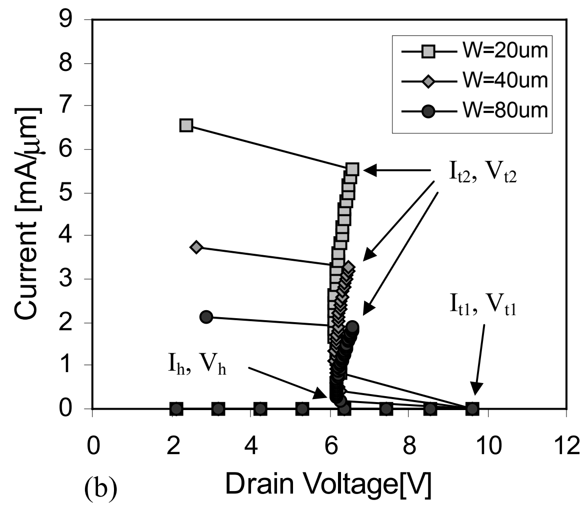
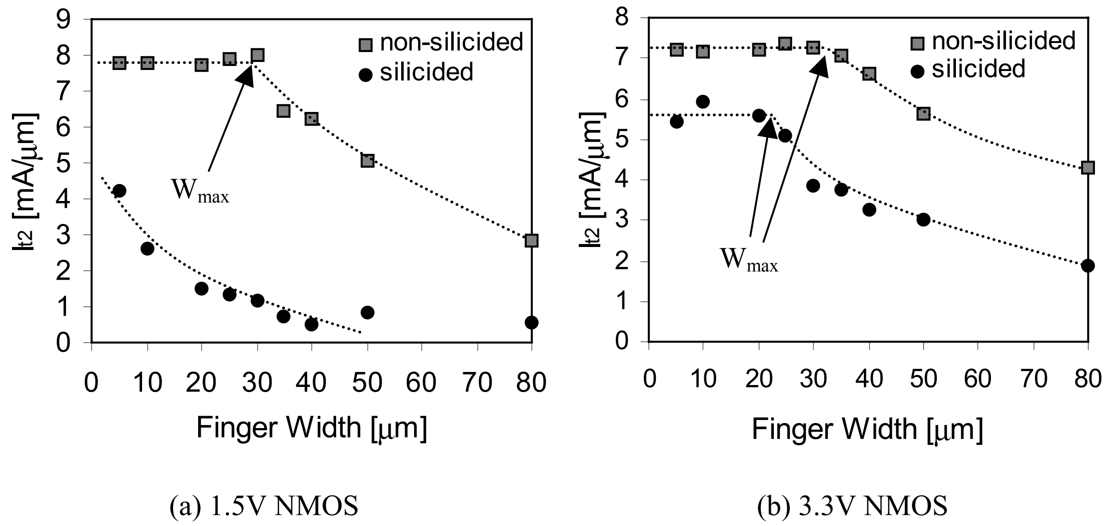


Figure 3.2: Sample high current I-V curves for 3.3 V NMOS transistors with different finger widths for (a) non-silicided, and (b) silicided processes where L_{poly} = 0.5 μm. The current per unit finger width at second breakdown I_{t2} [mA/μm] strongly depends on the finger width (W), which illustrates non-uniform bipolar currents flow under ESD conditions.



(a) 1.5V NMOS

(b) 3.3V NMOS

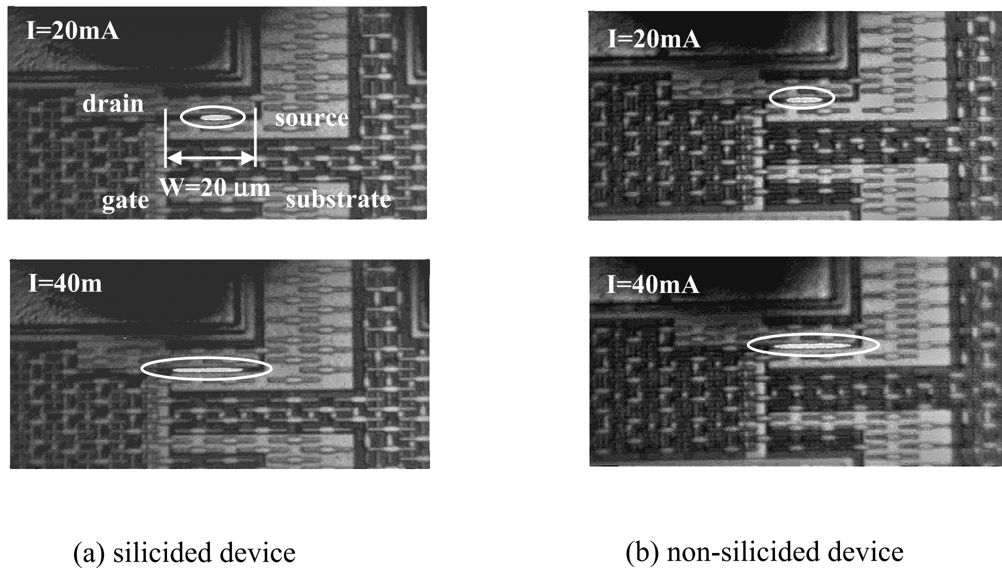
Figure 3.3: Silicide process dependent I_{t2} with the single finger width for (a) the1.5 V and (b) 3.3 V NMOS transistors. The I_{t2} roll-off with W indicates that the failure current essentially remains constant as W is further increased beyond this roll-off point.

In addition, the snapback holding voltage V_h of both the silicided and non-silicided transistors is nearly the same, which implies that the presence of silicide diffusion over the source/drain has no observable impact until the parasitic lateral n-p-n transistor snaps back.

However, the series resistance R_s [Ω-μm] in the high current regime for silicided and non-silicided devices shows considerable differences as expected. The high current I-V curves indicate that the current flows non-uniformly along the finger width (W) after the lateral n-p-n transistor snaps back. For both the low and high voltage (the 1.5 V and 3.3 V) devices, only a limited portion of the finger is effective for ESD current conduction beyond the I_{t2} roll-off points as is apparent from the channel width dependence shown in Fig. 3.3. The width dependence of I_{t2} gives clear evidence of the strong non-uniformity of bipolar conduction under ESD stress, even in single finger NMOS transistors. Moreover, for advanced silicided technologies, the degree of severity of this geometry dependence is even more significant in the effective design of on-chip ESD protection circuits.

3. 2. 2 Emission Microscopy (EMMI) Analysis

In order to visualize the strong non-uniformity of lateral n-p-n bipolar current conduction inferred from the I_{t2} data for the single finger NMOS transistors, the spatial distribution of ESD current was directly observed using EMMI for the silicided and non-silicided, 20 μm and 80 μm wide finger, high voltage (3.3 V) NMOS transistors. In this work, the commercial EMMI system was utilized, and a pulsed bias with duration (T_p) of 300 ns was applied at a frequency of 400 Hz to avoid any thermal failure due to self-heating during the exposure.



(a) silicided device (b) non-silicided device

Figure 3.4: EMMI images showing spatial extent of lateral current conduction at different current levels for 3.3 V (W/L_{poly}=20/0.5 μm) (a) silicided, and (b) non-silicided single finger NMOS devices. (T_p = 300 ns, f = 400 Hz, and T_{exp} = 6 min)

The emitted photons during each voltage pulse were integrated over the exposure time (T_{exp}) of 6 min. In Fig. 3.4, the observed current distributions are shown at current levels of 20 mA and 40 mA for the 20 μm wide transistors with silicided and non-silicided processes.

Despite the symmetrical layout of the substrate contacts [Fig. 3.1] with respect to the source contacts, at 20 mA of current stress, only a small section of the finger is turned on for both silicided and non-silicided devices. The turned-on location for each test structure is observed to be

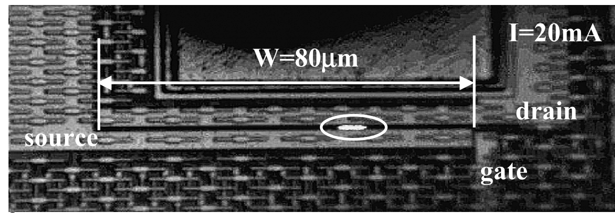
randomly positioned along the channel width, which is believed to result from the inhomogeneities in process conditions resulting in statistical random distribution of defects or dopant fluctuations, which causes different device behavior. The turned-on width expands with increased drain current and it is observed that most of the finger width is eventually turned on at a current of 40 mA. The same phenomenon has also been reported in the literature from observations using an infrared laser interferometric technique for the ggNMOS in a 0.35 μm technology [50].

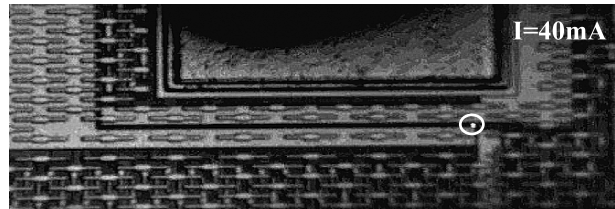
The EMMI results show strong qualitative correlation with the I_{t2} data for high voltage devices where the normalized failure currents (I_{t2}) are almost constant up to a finger width of 20 μm for both processes [Fig. 3.3 (b)]. For the 80 μm wide devices as shown in Fig. 3.5, the initial turned-on location is also randomly placed as observed for the 20 μm wide device.

However, both silicided and non-silicided devices failed with permanent or partial damage resulting in high drain junction leakage current, before the bipolar current conduction width extended to the entire 80 μm finger width, at 40 mA and 60 mA, respectively. Throughout the repetitive tests, full triggering of the 80 μm wide transistors was not observed for either the silicided or the non-silicided processes. This observation implies that severe degradation of I_{t2} with increase in (single) finger width results from inhomogeneous bipolar triggering phenomenon, and as shown in Fig. 3.3, the maximum turned-on width (W_{max}) of the lateral n-p-n transistor under ESD events can be regarded as the roll-off point in the data of I_{t2} vs. W. This qualitatively correlates with the results of the EMMI analysis to the TLP measurements.

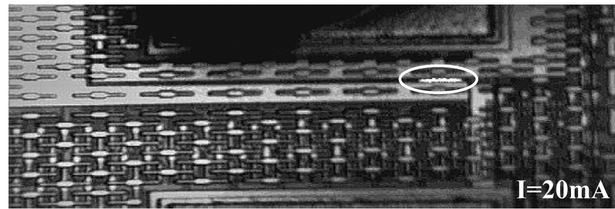
According to the data shown in Fig. 3.3 (a), for low voltage transistors, the estimated W_{max} value for non-silicided devices seems to be about 30 μm, and for silicided devices, W_{max} seems to be smaller than 5 μm. Moreover, for high voltage transistors, W_{max} for non-silicided devices and silicided devices are about 35 μm and 20 μm, respectively.

Hence, in practice, designed finger widths greater than W_{max} cause no improvement in I_{t2}. Thus, the obvious way of improving ESD strength is to expand the turned-on width, W_{max} for any given NMOS structures.





(a) silicided device



(b) non-silicided device

Figure 3.5: Emission microscopy images showing spatial extent of lateral current conduction at different current levels for 3.3 V (W/L_{poly}= 80/0.5 μm) (a) silicided (b) non-silicided single finger NMOS devices. (T_p = 300 ns, f = 400 Hz, T_{exp} = 6 min). The small bright spot indicates a failure or a partial failure of the devices. The discontinuity of the conduction region in (b) was not observed in the EMMI images at the lower current level than the 60 mA and the two spots were included within the conduction region.

3. 3. Physical Modeling of Non-Uniform Bipolar Conduction

For greater insight into the experimental results, a simple physical model for describing the non-uniform bipolar conduction is proposed. As the experimental results show, the spatial extent of bipolar conduction is determined by the triggered portion of the finger width. The single finger transistor can be considered as a parallel-connected network of narrow (segmented) n-p-n transistors as shown in Fig. 3.6. Since each segmented n-p-n transistor has slightly different intrinsic characteristics, which stems from the inherent statistical variations, the location of the triggered segmented n-p-n transistor (or transistors) is expected to be uncertain in this sense. According to the study by Russ et al. [45], avalanche multiplication starts at the corner of the drain structure where the field is highest due to the spherical junction curvature and the avalanche current is rather uniformly distributed along the channel width before the snapback of the lateral n-p-n transistor occurs.

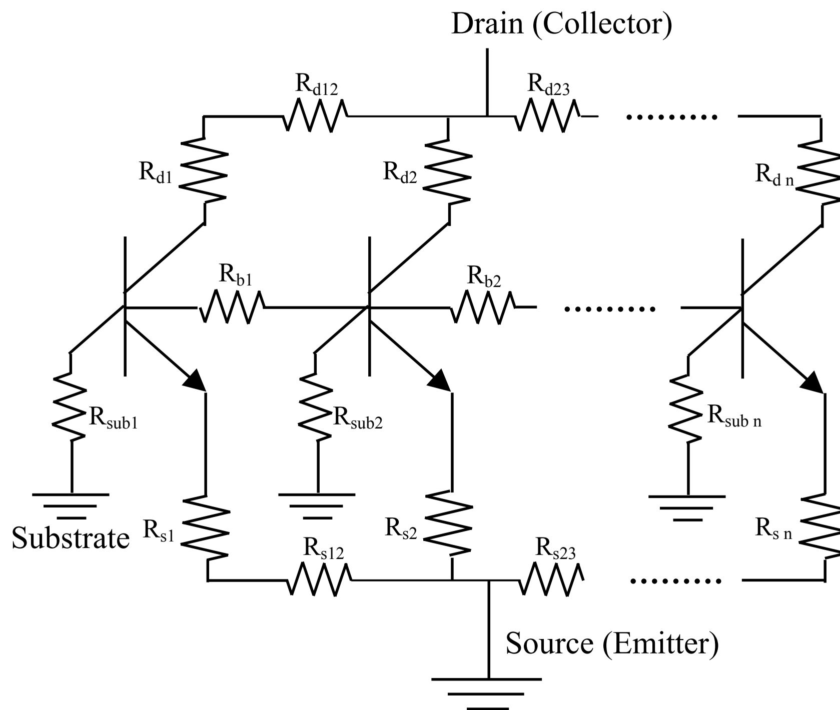


Figure 3.6: Schematic of the segmented n-p-n transistors for a gate-grounded single finger NMOS transistor. Each n-p-n transistor has different intrinsic characteristics due to the statistical variations. R_s, R_d, R_{sub}, and R_b denote the parasitic resistance in the source, the drain, the substrate, and the intrinsic base, respectively.

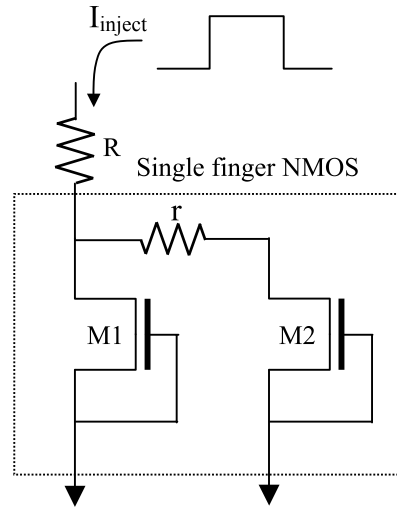
However, for the devices with STI used in this study, the initially turned-on location along the finger width has been observed to be random. Moreover, the avalanche region does not seem to be spread out enough to trigger the entire n-p-n transistor structure for 80 μm wide single finger devices. The parasitic bipolar triggering mechanism has been described in [51] in terms of three main device parameters - the current gain β , the substrate resistance R_{sub}, and the avalanche multiplication factor M. The substrate hole current (for NMOS) is strongly influenced by the electric field distribution of the drain junction, which depends on the doping profile and drain engineering of the structures. The resulting effective forward bias to the source-substrate junction, imposed by the local substrate potential for each segmented n-p-n transistor is unlikely to be equal due to local variations in substrate current (I_{sub}) and substrate resistance (R_{sub}).

Assuming that a small portion of the finger width (source-substrate junction) is sufficiently forward biased (i.e., $V_{sub} \approx 0.8$ V) because of the rather strong impact ionization process, the segmented transistors within this portion can be immediately triggered while the transistors along the rest of the finger width still remain off. This model agrees well with the EMMI observation that the location of initially turned-on segment is random over the finger width. Once the n-p-n transistors turn on, to maintain the on-state of the transistors, the snapback condition of $\beta \cdot (M-1) > 1$ [52] should be satisfied. However, both β and M are also functions of the injected drain current [51]. Therefore, the location of turned-on segmented n-p-n transistor should be strongly influenced by the drain current. With increased drain current, the number of turned-on segmented transistors should increase because the maximum current capability for each segmented transistor is limited. This is in agreement with EMMI analysis which showed that the turned-on width spreads out with increase in the injection current.

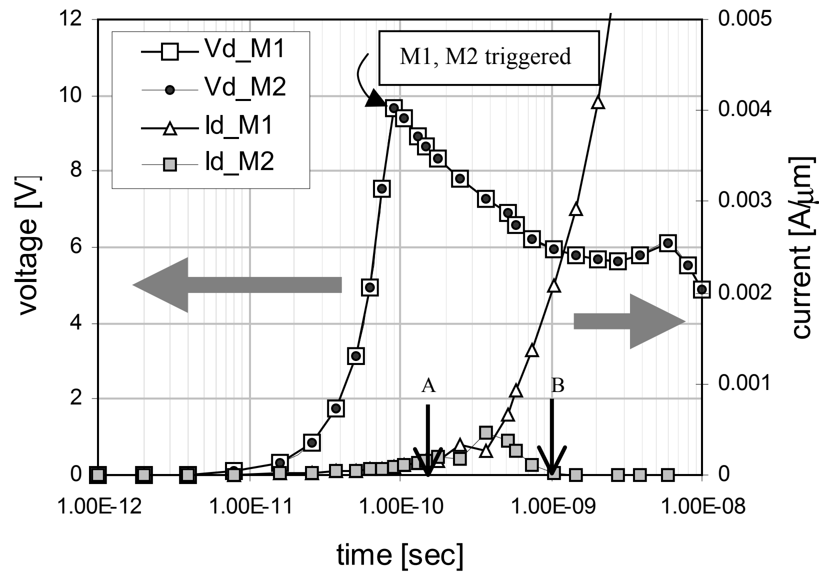
However, even with increased injection current, lateral bipolar conduction currents tend to flow through the portion of the finger width where impact ionization occurs most strongly. The rest of the finger width, where the impact ionization is relatively small, hardly turns on. This happens because the drain voltage drops to the holding voltage V_h, after a part of bipolar transistor structure triggers at V_{t1}. This non-isotropic device behavior along the width is technology dependent, and furthermore, even small asymmetric device properties can easily induce this non-uniform current conduction, which can be supported by 2D electro-thermal transient device simulations.

As shown in Fig. 3.7 (a), two NMOS structures, M1 and M2, are tied together to represent each half of a single finger transistor. For simplicity, non-isotropic properties of devices are represented by a parasitic resistance (r) between the two drains, which represents variations in

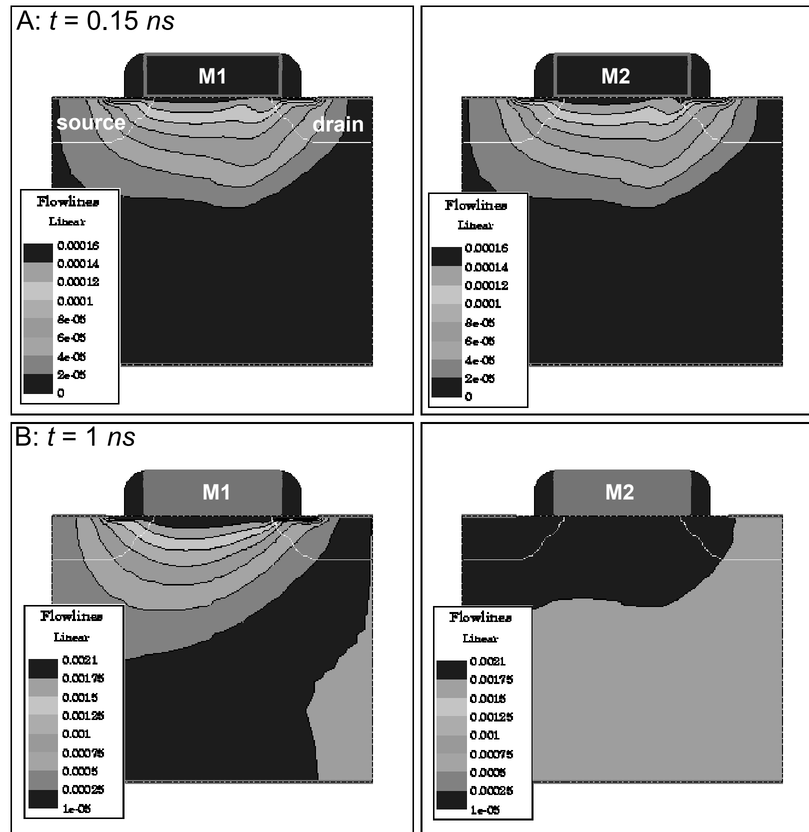
drain series resistance. This affects the strength of impact ionization through electric field reduction and also changes the effective series resistance for the two NMOS transistors.



(a)



(b)



(c)

Figure 3.7: Non-uniform current conduction with a mixed mode transient simulation for M1 and M2 (W/L_{poly}=1/0.5 μm). (a) the schematic of simulation. (R = 500 Ω and r = 0.81 Ω), (b) drain current and voltage with elapsed time for M1 and M2, and (c) the current flowlines at the two different time conditions, A and B.

The increase in the injection current (with elapsed time) triggers both the transistors, M1 and M2, at t = 0.09 ns. After the snapback, the current for the two transistors increase together up to 0.5 mA/μm as shown in Fig. 3.7 (b). However, as the current increases for the transistor (M2), the increase in voltage drop across the parasitic resistance can reduce the strength of avalanche multiplication of M2. In that case, M2 turns off. This can be observed from the current flowlines shown in Fig. 3.7 (c). The simulation results confirm that the inequality of intrinsic characteristics of each segmented transistor causes asymmetry in current conduction, and subsequently results in current localization which becomes dependent on the injected drain current level.

3. 4 Implications for the Design of ESD Protection

As is well known, the ESD strength of silicided technology is lower than that of non-silicided devices due to either the reduction in emitter efficiency [26] or early current localization associated with the reduced series resistance [14], which has also been verified in this work. EMMI analysis shows that different bipolar turned-on widths can be obtained depending on silicide process. Therefore, these facts indicate that wider current conduction is associated with a higher ESD failure threshold for a given protection structure.

For advanced silicided NMOS transistors, I_{t2} drops off rapidly beyond W_{max}. This obviously places a severe restriction on determining the useful width of a single finger for multi-finger structures used in ESD protection. From a practical design point of view, the minimum value for W_{max} should be at least 30 μm with a minimum I_{t2} (at T_p = 200 ns) of ~ 4 mA/μm. This will ensure that > 8 V/μm for HBM is available for the design of multi-finger protection devices.

The aforementioned I_{t2} degradation with finger width, which is very serious for devices with silicide diffusion, is related to the current crowding effect that allows only a portion of the finger to conduct the ESD current. This implies that total failure threshold current (i.e., $IT_2 = I_{t2} \cdot W$) under ESD stress is effectively reduced, and further implies that the ESD performance of the protection NMOS transistors cannot be increased by merely increasing the designed finger width in advanced silicided CMOS technology. Therefore, the width of a single finger in multi-finger protection devices should be determined within the maximum turned on width W_{max}. This also places some restrictions on how to efficiently layout multi-finger structures since the single finger width available is limited to W_{max}.

Therefore, it is proposed that during process and design characterization the ESD performance (i.e., I_{t2}) of the protection elements should be evaluated at the same conditions as the protection device itself.

3. 5 Summary

In summary, it is shown that non-uniform bipolar conduction phenomenon in advanced single finger NMOS transistors results in severe reduction in ESD protection strength depending on the device finger width. This non-uniform bipolar conduction becomes more serious with silicided protection devices and this can explain abnormal decrease in ESD performance of the silicided devices, compared with non-silicided devices.

The reason for this non-uniform conduction could be related to the intrinsic process defects that effectively lock in the bipolar conduction to a local area in the finger width. The results from experiments and 2D device simulations in this work provide improved understanding of ESD behavior and new physical insight into the current localization effects involved in advanced ESD protection devices.

Results from this work can be used to construct suitable design windows for efficient and robust ESD on-chip protection to overcome early ESD failures in advanced deep submicron CMOS technologies.

CHAPTER 4

BIAS DEPENDENCIES OF ESD ROBUSTNESS

In addition to the detailed study of the non-uniform bipolar conduction phenomenon, the substrate and gate bias dependencies of ESD robustness of advanced NMOS transistors are investigated. It is shown that the uniformity of the bipolar current distribution under ESD conditions, which is known to be a primary cause of the degradation of ESD performance, is significantly influenced by the substrate and gate bias conditions. In this regard, as a means of improving the uniformity of ESD current distribution, the substrate triggering and the gate coupling techniques have been proposed [25, 26, 28]. It is shown that the substrate bias can alleviate the non-uniform conduction so that it makes the design of ESD protection more practical by enlarging the effective turned on width of the device.

However, contrary to general understanding, it has been recently reported that the ESD strength with high gate bias is degraded depending on the finger width for deep submicron salicided ESD protection NMOS transistors [53, 54] and this observation also requires the identification of the underlying physical mechanism involved in the degradation.

Additionally, the concept of an *intrinsic second breakdown triggering current* (I_{t2i}) is introduced, which is substrate bias independent and represents the maximum achievable ESD failure strength for a given technology. With improved understanding of ESD behavior of protection devices under bias conditions, an efficient design window with extended design capabilities can be constructed and the on-chip protection design can be more practical for robust deep submicron ESD protection.

4. 1 Substrate Bias Effect

For comprehensive understanding of the influence of substrate bias on ESD robustness, EMMI has been performed with an external forward substrate bias (V_{sub}) to the emitter and base (the source and substrate) junction of the NMOS transistor [Fig. 4.1]. With increase in substrate bias, the local substrate potential can be sufficiently raised to trigger the parasitic lateral n-p-n transistor without relying on the self-biasing mode. The influence of the substrate bias on the spatial current distribution is apparent as shown in Fig. 4.2. It can be observed that the transistor turned-on width spreads out with increase in V_{sub} (from 0 V to 1 V) for both silicided and non-silicided devices at the constant drain current of 30 mA.

In fact, the total turned-on width is enlarged by 3 to 4 times with $V_{sub} = 1$ V regardless of the silicide process. However, even with the substrate bias, full n-p-n transistor triggering was not observed for the 80 μm wide transistors. Nevertheless, EMMI images clearly illustrate that the substrate bias can increase the effective finger width by extending the lateral bipolar conduction width, which can lead to the improvement of I_{t2} as shown in Fig. 4.3.

It is important to note that while this positive impact of substrate bias on I_{t2} has been reported before for a 0.35 μm process [26], the physical mechanism responsible for the I_{t2} improvement with V_{sub} has not been adequately explored. It can be observed from Fig. 4.3 that despite the difference in substrate resistance[19] (R_{sub}), the substrate bias is effective, and it also implies that the improvement of the ESD performance can be realized without changes in the process or structure of the devices.

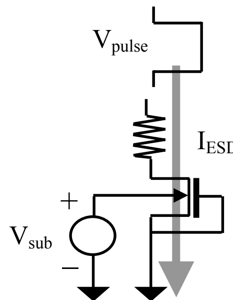
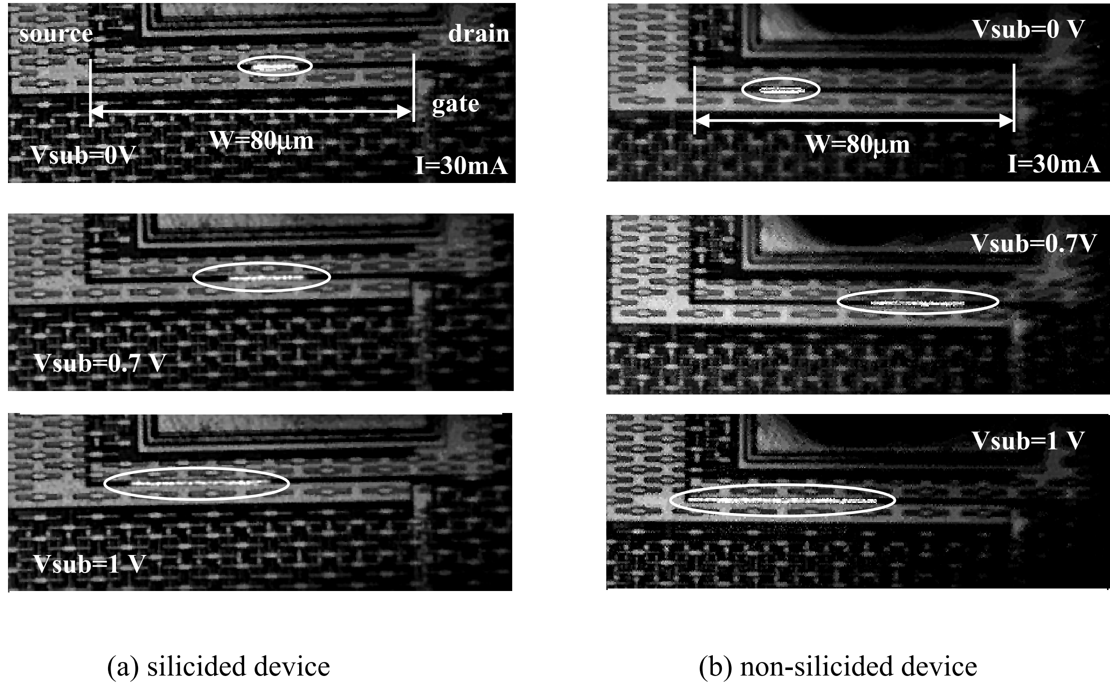


Figure 4.1: Schematic of the biasing scheme for investigating impact of substrate bias on ESD performance.

[19] Resistivity of epi layer is different for the two devices under test.



(a) silicided device (b) non-silicided device

Figure 4.2: Emission microscopy images that show the spatial extent of lateral current conduction with different V_{sub} for 3.3 V (W/L_{poly}= 80/0.5 μm) (a) silicided and (b) non-silicided single finger NMOS devices. (T_p= 300 ns, f= 400 Hz, T_{exp}= 6 min)

4. 1. 1 Intrinsic Second breakdown Triggering Current, I_{t2i}

As the forward substrate bias increases, the n-p-n bipolar triggering voltage (V_{t1}) reduces and eventually the bipolar device turns on without snapback when the effective emitter-base (source-substrate) junction bias reaches ~ 0.8 V. From Fig. 4.3, at V_{sub} = 1 V, it can be inferred that weak avalanche generation is adequate to supply the required substrate current for triggering the lateral n-p-n transistor. For higher V_{sub} (≥1 V), I_{t2} values tend to saturate. This means that the effective bipolar conduction width gets pinned and the local substrate potential near emitter-base junction cannot be altered by applying additional substrate bias. The associated I_{t2} at this substrate bias will be the maximum achievable value for both high and low substrate resistance devices. This current is substrate bias independent as shown in Fig. 4.3 and can be called as the *intrinsic second breakdown triggering current*, I_{t2i} [47] for a given technology.

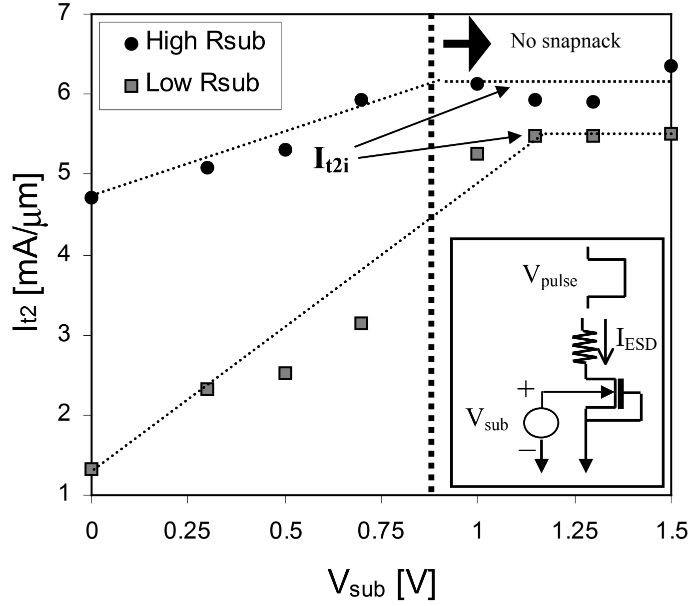
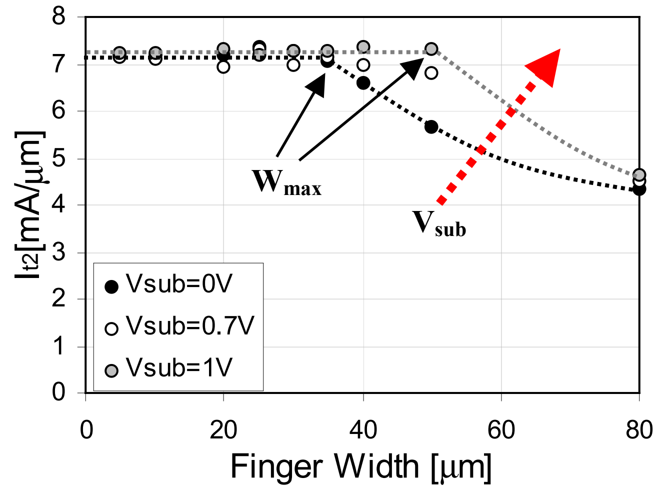


Figure 4.3: Second breakdown triggering current I_{t2} with V_{sub} for the 3.3 V silicided devices with two different R_{sub} (6300 Ω-µm and 4800 Ω-µm) where W/L= 20/0.5 µm. I_{t2} approaches its intrinsic value I_{t2i} as the substrate bias is increased.

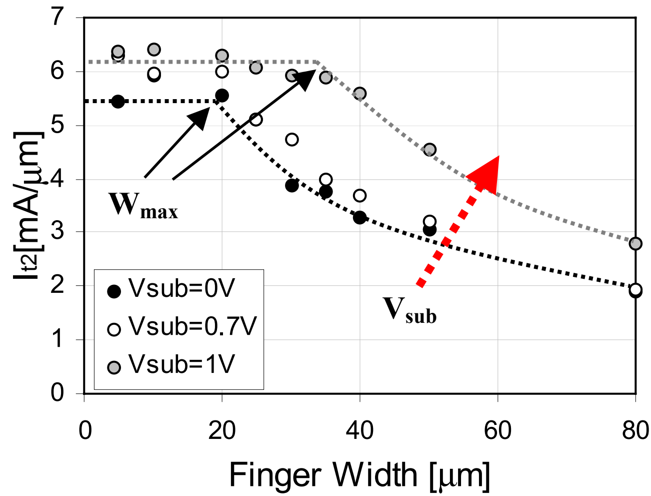
In addition, the non-uniform triggering that arises from the inhomogeneity of local substrate potential (along the width of NMOS) can be alleviated by applying a V_{sub} since the effective finger width can be increased by enhanced bipolar current uniformity. For the test structures with various finger widths, I_{t2} values are shown as a function of V_{sub} in Fig. 4.4.

It can be observed that $W_{max,}$ where I_{t2} rolls off, increases with substrate bias and the I_{t2} values for silicided devices approach that of non-silicided devices with substrate bias. However, it can also be observed that the values of I_{t2} for $W < W_{max}$ remain almost independent of the substrate bias within the scatter among the data.

Hence, this value of I_{t2} for finger widths less than W_{max} can be thought of as the maximum obtainable I_{t2} (defined as I_{t2i} earlier) under uniform bipolar conduction for a given process technology, and ranges from 6 to 7.2 mA/µm for both the processes, which is solely determined by process effects such as silicide/non-silicide process, gate-to-contact spacing, source/drain engineering and substrate resistance.



(a)



(b)

Figure 4.4: I_{t2} as a function of transistor width for different substrate bias V_{sub} for (a) non silicided, and (b) silicided 3.3 V NMOS transistors.

4. 1. 2 Device Simulation Study: Principle of Operation with V_{sub}

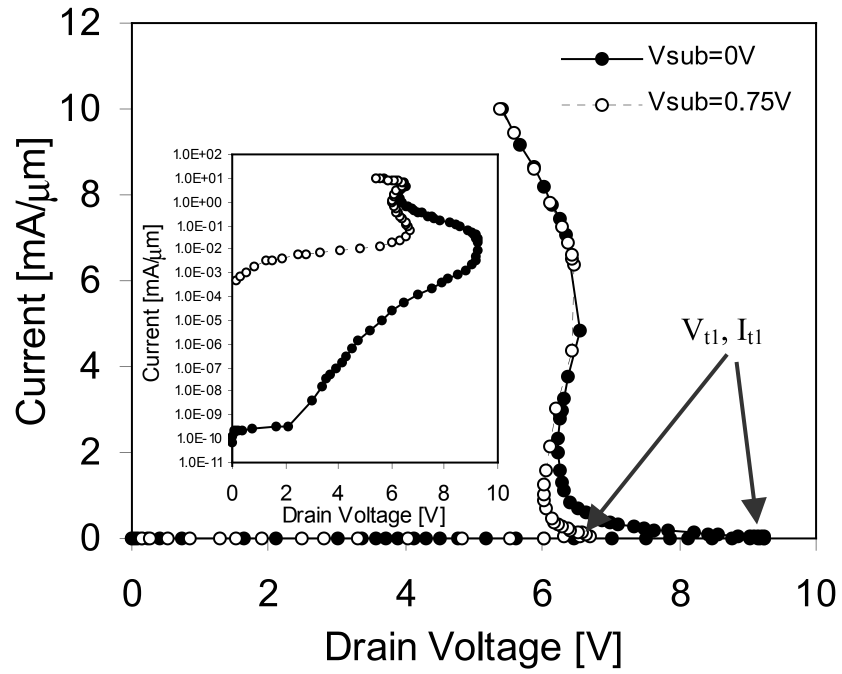
Aside from three-dimensional effects on ESD current distribution, insight concerning the substrate bias effect can be attained by 2D device simulations. In Fig. 4.5, static high current characteristics of a gate-grounded NMOS transistor with L_{poly} = 0.5 μm is simulated for two different substrate bias conditions, V_{sub} = 0 V and V_{sub} = 0.75 V. The high current I-V curves show the same physical trends as the measured data. At a drain current (I_D) of 500 μA/μm after n-p-n transistor triggering, the current flowlines are compared for the two substrate bias conditions, which shows that the current flows more deeply into the substrate with a substrate bias. The altered local substrate potential changes the snapback triggering voltage, V_{t1} [Fig. 4.5 (a)] and eventually turns on the n-p-n transistor without snapback. However, the I-V curves at high current levels show no significant differences between the two substrate bias conditions since any three-dimensional behavior cannot be taken into account.

The second snapback at the drain current of ~ 5 mA/μm in the I-V curves is attributed to a rapid increase in the base current resulting from a significant increase in the current component due to thermally generated carriers although the carriers generated by impact ionization decrease slightly with temperature rise [29]. In the case of the self-biasing mode, sufficient substrate current due to impact ionization is required to maintain the forward bias to the emitter and base junction (source and substrate) and the base current is also supplied by impact ionization. Hence, rather strong avalanche multiplication is required for triggering the lateral n-p-n transistor.

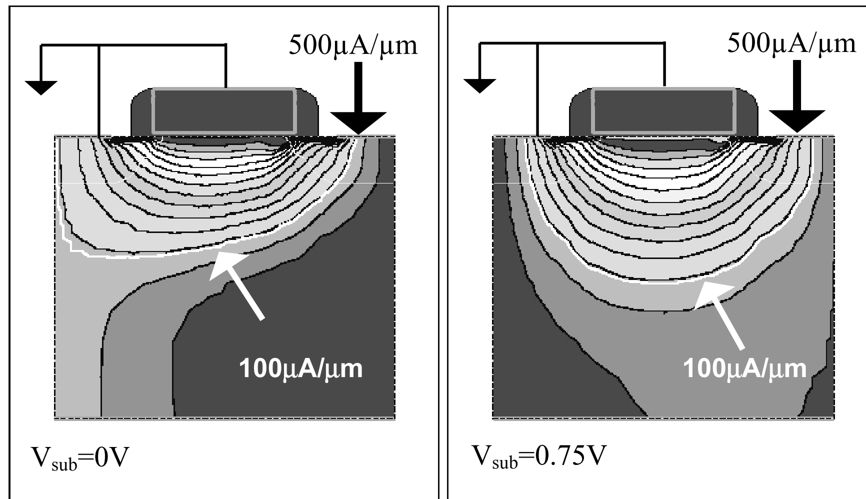
On the other hand, under adequate external substrate bias, the lateral n-p-n transistor operates in a normal biasing mode (common-emitter). Even in the absence of the drain current, the source-substrate and the drain-substrate junctions are fully turned on. However, since both the parasitic diodes[20] in an NMOS transistor have a relatively long base, most of the injected carriers from the source and drain to the substrate recombine resulting in small diffusion currents. As the drain current and the associated drain bias increase, the drain-substrate junction is eventually reversed biased and thus the lateral n-p-n transistor operates under normal bias conditions.

However, the bipolar conduction current increases very slowly with the increase in the drain current (the associated drain bias) until the impact ionization is initiated.

[20] The source-substrate and drain-substrate junction diodes.



(a)



(b)

Figure 4.5: Static I-V characteristics for the ggNMOS with L_{poly}=0.5 μm. (a) High current I-V curve, and (b) current flowlines at the drain current (I_D) of 500 μA/μm with V_{sub}=0 V and V_{sub}=0.75 V. The current flows more deeply into the substrate with V_{sub}.

Before the avalanche multiplication occurs, most of the drain current, I_D is supported by electrons injected from the source (emitter), but these injected carriers mainly contribute to the source-substrate diode current rather than to the drain (collector) current due to the low current gain β of the lateral n-p-n transistor arising from the significant electron-hole recombination in the base region. A further increase in the drain current induces impact ionization and eventually the avalanche multiplication process becomes regenerative as in the self-biasing mode. Since the emitter base junction is already fully turned on, small values of the avalanche-generation current, I_{gen}, are sufficient to initiate the regenerative process and accordingly the corresponding threshold value of M should be lower than the value needed in a self-biasing mode. As a result, the drain current flows through the low field area in the drain-substrate junction. Fig. 4.5 (b) clearly shows that relatively wider area of the drain junction is utilized for the same current conduction (i.e., I_D=500 μA/μm) with V_{sub} because less impact ionization current is required. Therefore 2D device simulation results suggest that the bipolar conduction could take place over a wider area of the drain and substrate junction under sufficient substrate bias.

4. 1. 3 Effective Finger Width under ESD Stress

As discussed previously, the turned-on width of the silicided wider finger devices is not the same as the designed finger width under ESD stress. Therefore, when determining the width of a single finger of a multi-finger structure, the effective finger width needs to be considered to design on-chip protection circuits more efficiently.

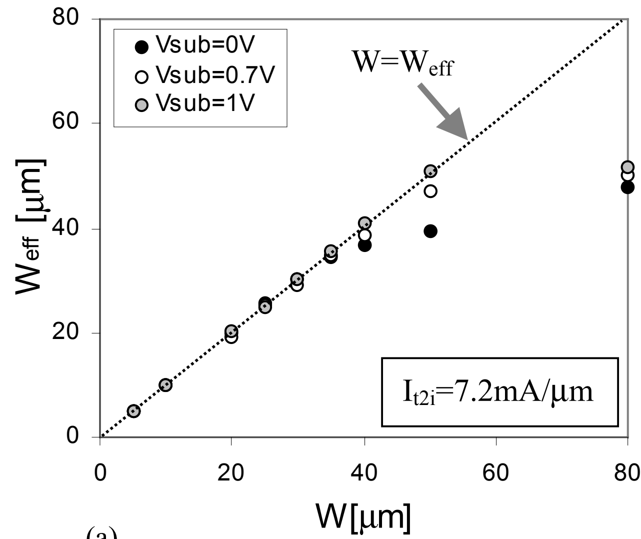
Based on the value of I_{t2i} for each technology, the effective finger width W_{eff}, for bipolar conduction can be determined from the following simple relation.

$$W_{eff} = (I_{t2} / I_{t2i}) \cdot W \tag{4.1}$$

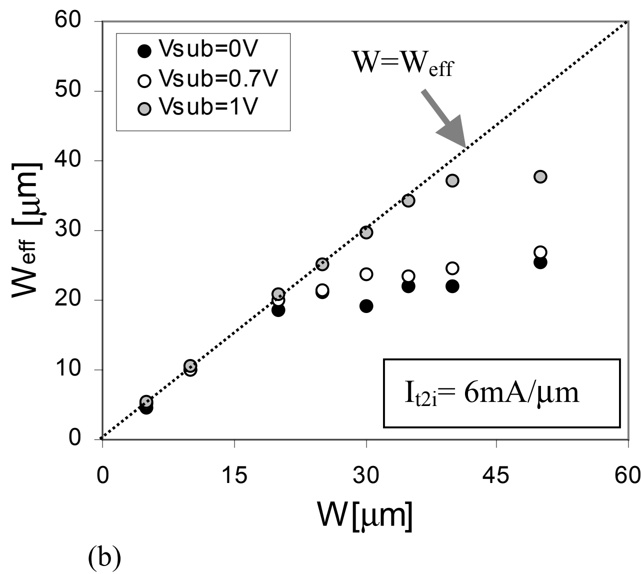
Since the total failure threshold current (i.e., $IT_2 = I_{t2} \cdot W$) is constant independent of the current uniformity for a given technology, the effective finger width W_{eff} can be directly obtained from the extracted I_{t2i} and the plots of W_{eff} vs. W for the high voltage (3.3 V) NMOS transistors are shown in Fig. 4.6.

As can be seen, at V_{sub}=1 V, W_{max} is 50 μm and 35 μm for the non-silicided and silicided technologies respectively, showing significant improvement over the W_{max} values for the zero substrate bias case. Therefore, the use of substrate bias can extend ESD design capabilities

beyond present design and technology limits, and one potential application has recently been implemented [28, 55].



(a)



(b)

Figure 4.6: Effective finger widths (W_{eff}) vs. designed finger widths (W) for 3.3 V NMOS transistors with different substrate bias (a) non-silicided devices, and (b) silicided devices.

4. 2 Gate Bias Effect

For multi-finger NMOS protection, it has been recognized that the gate coupling technique is efficient since it ensures uniform triggering of the lateral n-p-n bipolar transistors [24, 46]. However, its effectiveness is dubious in silicided processes [56]. Additionally, it is also well known that excess gate coupling degrades the second breakdown triggering current (I_{t2}) of NMOS devices [8, 25] and thus design techniques have been used to limit the gate coupling.

Even with controlled gate coupling on the protection device, under ESD stress, high gate coupling on the output NMOS transistor causes human body model (HBM) or charged device model (CDM) failures in output buffer protection, thereby placing some restrictions on the design of ESD protection. This phenomenon of degradation of I_{t2} is technology dependent and its impact on the advanced protection design has not been fully explored before. Moreover, for advanced salicided technologies, the severity of this effect has been shown to be dependent on the finger width and the extent of lateral uniformity of ESD current conduction [53, 54].

In this section, we identify the root cause of this degradation of I_{t2} with high gate bias for the 0.13 μm technology, and also investigate the impact of substrate bias on the gate coupling effect. Accounting for this ESD behavior involved in advanced salicided devices, design guidelines useful for designing protection structures are also presented.

4. 2. 1 Output NMOS Failure

A typical output buffer protection scheme with different protection device options is shown in Fig. 4.7. In ESD protection circuits, as briefly discussed in Chapter 1, there are several NMOS protection structures available such as the gate grounded NMOS (ggNMOS), the gate coupled NMOS (gcNMOS), and the substrate pump NMOS (stNMOS) structures. Especially, both the gcNMOS and stNMOS have been proposed to ensure uniform lateral bipolar current conduction [25, 28].

Under ESD stress from the I/O pad to ground, the ESD current is shared by the several different competing current paths, mostly through the NMOS protection structures, and partially through the lateral V_{dd} diode and the output NMOS transistor itself. During such conditions, the potential of the I/O pad reaches a voltage level sufficient to trigger the NMOS protection structure and finally the voltage of I/O pad remains near the holding voltage (V_h) depending on the NMOS protection options. The ESD current through the lateral diode to V_{dd} node charges up

the V_{dd} capacitance up to about V_h - 0.7 V. It should be noted that the maximum voltage of the V_{dd} node is also limited by the presence of V_{dd} clamp.

As a result, depending on the pre-drive circuit conditions, this voltage (at V_{dd} node) can be fed into the gate of the output NMOS transistor through internal circuit blocks, which can influence the effectiveness of the ESD protection design.

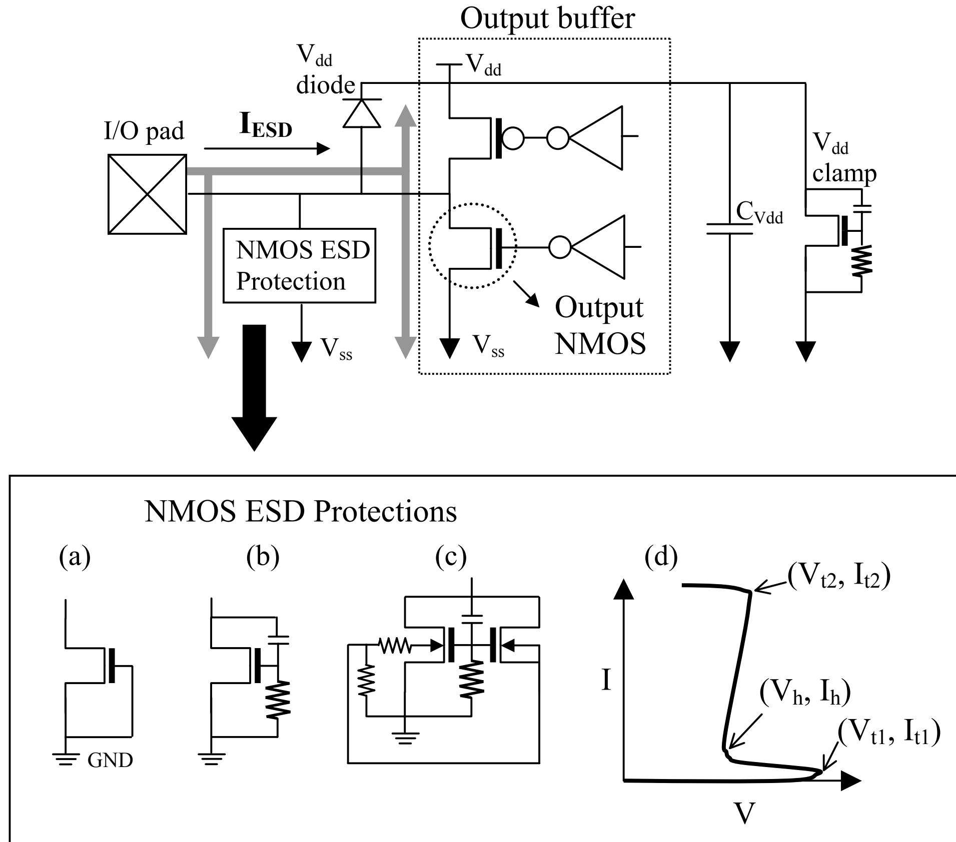


Figure 4.7: Simplified output buffer protection scheme with various protection NMOS transistor options. Irrespective of protection transistor options, under ESD conditions, high ESD voltage can be fed into the gate of the output NMOS transistor, which could lead to early ESD failures. (The gray arrows indicate the ESD current paths.) (a) gate-grounded NMOS (ggNMOS), (b) gate-coupled NMOS (gcNMOS), (c) substrate pump NMOS (stNMOS), and (d) generalized high current I-V characteristics of ggNMOS.

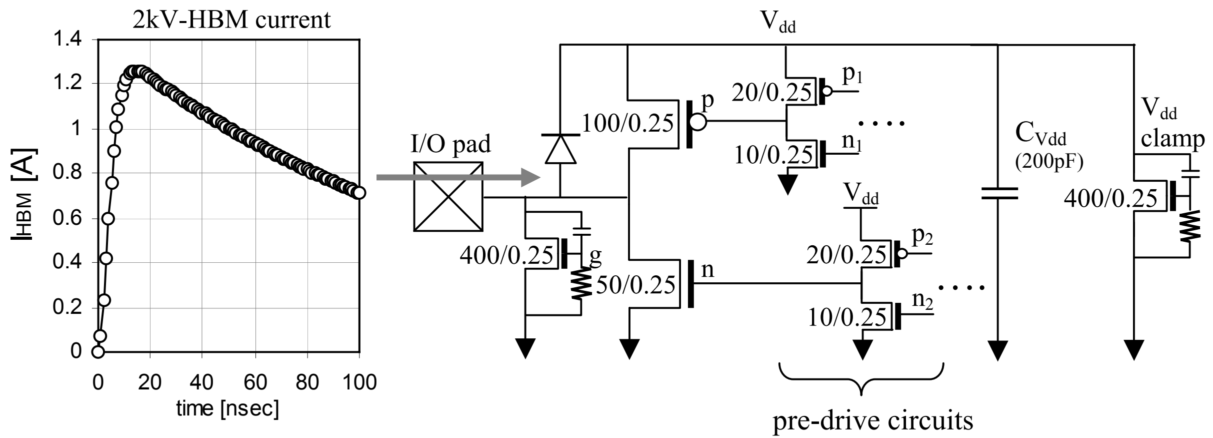


Figure 4.8: The schematic of the standard 2 kV Human Body Model (HBM) tests for simple output buffer protection circuit.

By using HSPICE, 2kV HBM test mode was reproduced for a simplified output buffer protection scheme in Fig. 4.8 [57]. A 3.3 V NMOS transistor model was used, which has a holding voltage (V_h) of ~ 5.3 V. Initially, all the nodes are floating, and then voltages at the I/O pad and V_{dd} line start to increase with the injection of the standard 2 kV HBM current. The gate voltage of the output NMOS transistor is determined depending on the condition of the pre-drive circuits of the output devices. Typically, the pre-drive circuits can have the logic value of either high (i.e., 1) or low (i.e., 0). Since the pre-drive circuit condition is unpredictable at the ESD event, by assuming different conditions of the pre-drive circuits, the gate potential of the output NMOS and protection NMOS transistors was investigated.

As shown in Fig. 4.9, with the conditions of (a) and (b), the gate voltage (V_n) of the output NMOS transistor increases with a slight delay, and finally goes higher than that (V_g) of the protection NMOS device, which can lead to a reduction in the ESD strength relative to the protection device. On the other hand, for conditions (c) and (d), the gate potential of the output NMOS stays below 0.5 V and thus, no severe I_{t2} degradation is expected. For the simulation, the gate voltage of the output NMOS transistor ranges from ground to the voltage at V_{dd} (~ 4.8 V) depending on the conditions of pre-drive circuits and similar effects can take place for any other potential values induced at V_{dd} node.

Therefore, the gate bias effect for the output NMOS transistor has significant implications for the overall design of ESD protection. In Fig. 4.10, the early ESD failure of the output NMOS transistor has been shown for the HBM test due to the high gate coupling induced I_{t2} degradation, regardless of the ESD strength of the protection device itself. The same failure mode has also been observed for the CDM tests.

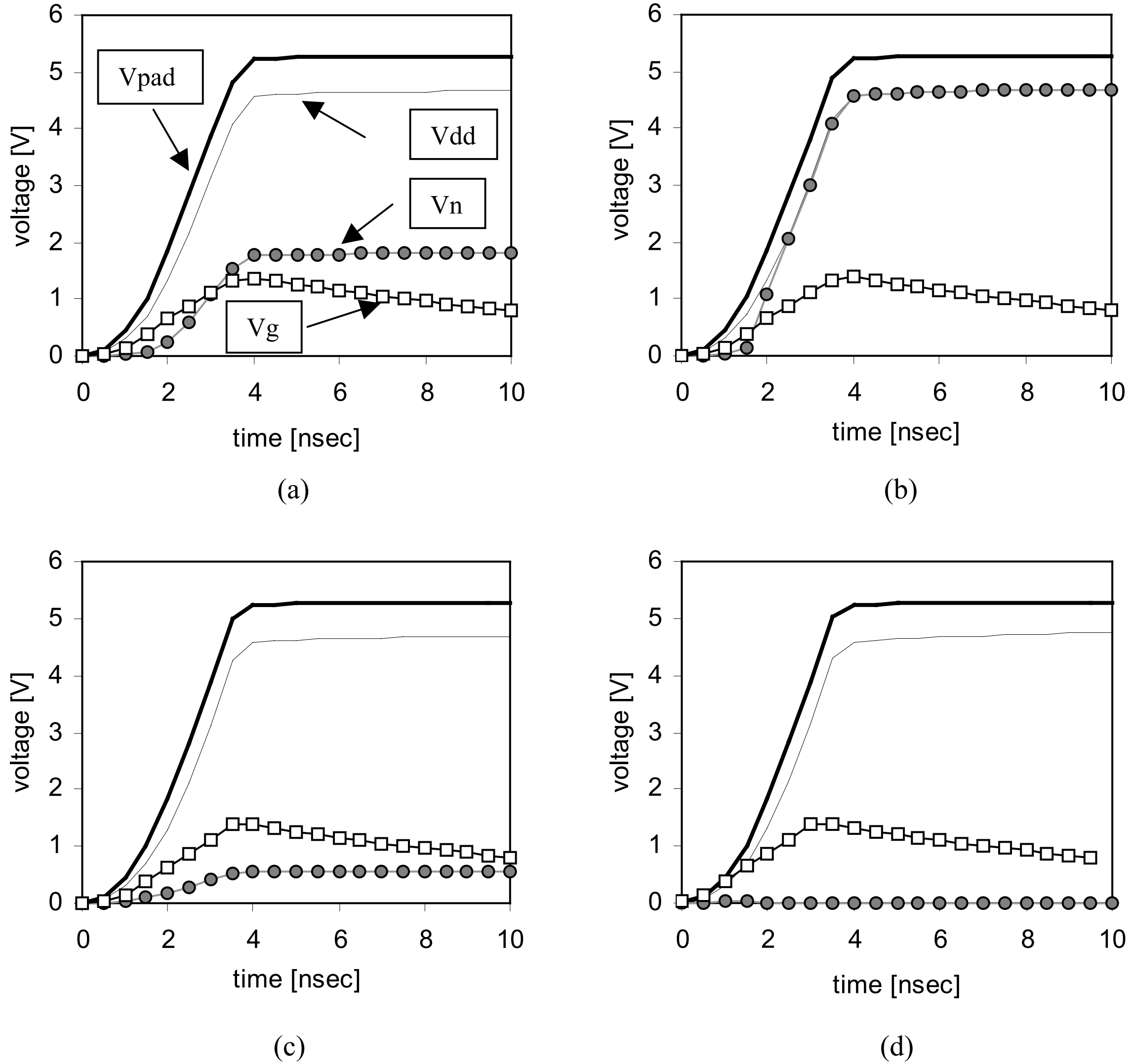


Figure 4.9: Simulated voltage waveforms of each node of the output buffer protection under 2kV HBM stress with the four different pre-drive circuit conditions where (a) $p_1 = p_2 =$ low and $n_1 = n_2 =$ high, (b) $p_1 = p_2 = n_1 = n_2 =$ low, (c) $p_1 = p_2 =$ high and $n_1 = n_2 =$ low, and (d) $p_1 = p_2 = n_1 = n_2 =$ high.

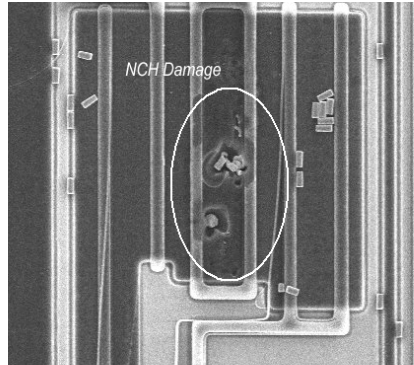


Figure 4.10: Failure image of the output NMOS transistor in HBM test mode. Since increased gate voltage of the device due to high gate coupling lowers its ESD strength, the device fails earlier than the protection devices.

4. 2. 2 Experiments and Analysis

In order to observe early ESD failure with excess gate coupling, involving various gate bias conditions, I_{t2} was measured with the transmission line pulsing (TLP) method with a 200 ns voltage pulse. In Fig. 4.11, the I_{t2} values with gate bias are shown for a 0.35 μm technology.

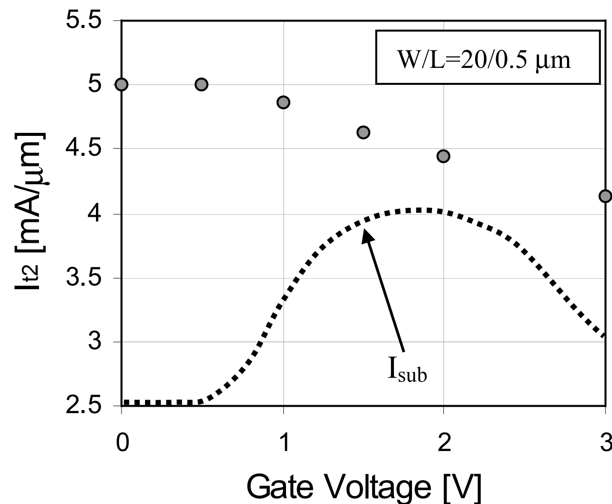
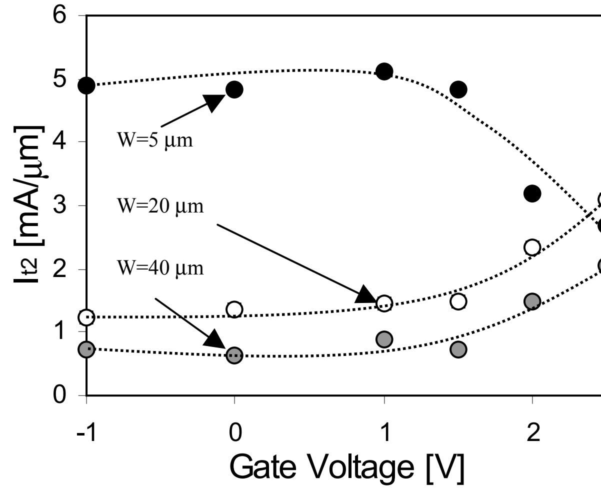
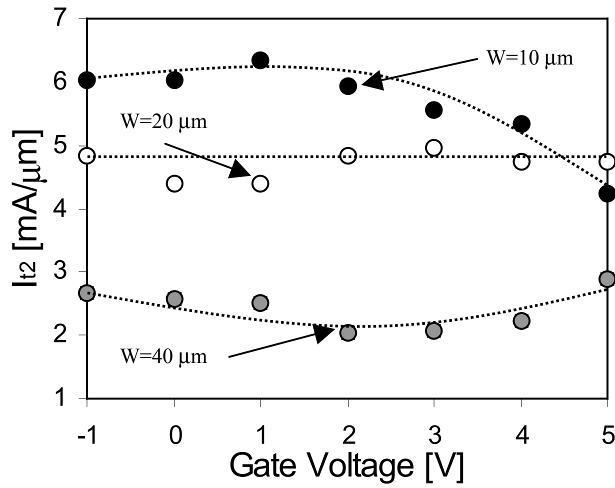


Figure 4.11: Second breakdown triggering current (I_{t2}) vs. gate voltage for a salicided ($TiSi_2$) 0.35 μm technology node. The dotted line is the well-known substrate current (in arbitrary unit) for conventional NMOS transistor, which shows that the degradation of I_{t2} is not due to the substrate current behavior as a function of gate bias, and further suggests that the turn-on efficiency of the parasitic n-p-n transistor is not the primary cause of this phenomenon.

As can be seen, contrary to the conventional belief, the phenomenon of degradation of I_{t2} does not follow the substrate current behavior with gate bias. This implies that the parasitic bipolar turn-on efficiency cannot be the primary cause of the I_{t2} degradation, and furthermore, that there is a new physical mechanism governing this phenomenon.



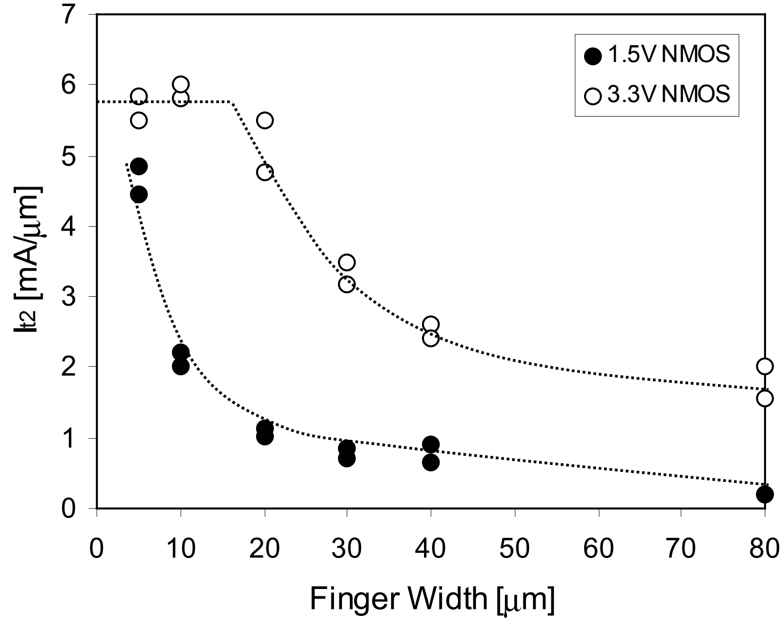
(a) 1.5V NMOS



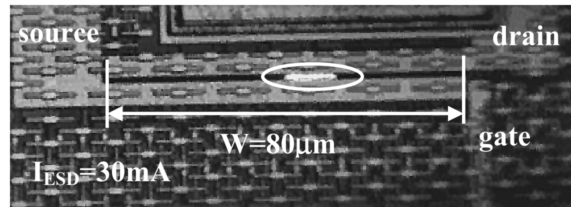
(b) 3.3V NMOS

Figure 4.12: Second breakdown triggering current (I_{t2}) with gate bias for the two different NMOS transistors for the 0.13 μm technology. The dependence of I_{t2} on gate bias is influenced by the finger widths. Gate bias improves I_{t2} for wide finger-width NMOS transistors, but degrades I_{t2} of narrow finger-width devices. (a) 1.5 V NMOS and (b) 3.3 V NMOS

In addition, it has been recently reported that ESD current distributes non-uniformly depending on the finger width, and under ESD conditions the effective width of devices is limited by the extent of the non-uniformity of the bipolar current distribution [47, 48], which throws new light on the gate bias induced phenomenon.



(a)



(b)

Figure 4.13: (a) I_{t2} for the output NMOS transistors (both of 1.5 V and 3.3 V transistors) with V_{gs} = 0 V for different finger widths and (b) EMMI image of the spatial distribution of ESD current (I_{ESD}) of 30 mA for the 3.3 V NMOS transistor. They both show that strong non-uniform conduction occurs in salicided NMOS transistors for the 0.13 μm technology. For the EMMI analysis, a pulsed bias with duration (T_p) of 300 ns was applied at a frequency of 400 Hz for the exposure time (T_{exp}) of 6 min.

For advanced salicided devices in the 0.13 μm technology under investigation, the impact of gate bias on I_{t2} can be clearly observed in Fig. 4.12, and which is also finger width dependent for both low and high voltage NMOS transistors. The I_{t2} dependence on the gate bias is no longer consistent with Fig. 4.11 for the W=20 μm and W=40 μm devices of the 0.13 μm technology node. Contrary to the dependence observed in the 0.35 μm technology node, opposite trends appear depending on the gate finger width of the NMOS transistor. This implies that the gate bias can result in two different competing physical mechanisms depending on the finger width for a given structure.

As shown in Fig. 4.13 (a), the I_{t2} values of the advanced salicided transistors are severely degraded with increasing finger widths. The ESD current distribution is uniform within a very narrow finger width such as $W < 5$ μm for the low voltage (1.5 V) transistors and $W \leq 10$ μm for the high voltage (3.3 V) transistors. In addition, the emission microscopy (EMMI) images with a pulsed bias condition show that only a small part of the finger width is effective for the ESD current conduction in the 80 μm wide salicided NMOS devices [Fig. 4.13 (b)].

Further increase in the drain current level caused thermal damage without full triggering of the NMOS transistor. This strong width dependence of I_{t2} for advanced technologies is attributed to the localized (non-uniform) bipolar conduction. As discussed in [47, 48], this non-uniform bipolar conduction becomes more serious for devices with low resistance substrates and silicided diffusions. Considering the results in Fig. 4.12 and Fig. 4.13, it can be inferred that gate bias can improve I_{t2} of the wide finger devices (W=20 μm and W=40 μm) where the ESD currents are non-uniform.

On the other hand, I_{t2} of the narrow finger device (W=5 μm for the 1.5 V NMOS and W=10 μm for 3.3 V NMOS) where ESD current is known to conduct almost uniformly, is degraded with gate bias.

4. 2. 3 Simulations and Discussion

As is well known, boosting of the substrate current with gate bias can minimize the current localization under ESD conditions. This mechanism seems to work for the wide finger 1.5 V NMOS devices [Fig. 4.12 (a)] with considerable improvement of I_{t2}, while the improvement of I_{t2} for the 3.3 V devices is less apparent [Fig. 4.12 (b)]. However, the severe reduction in I_{t2} with gate bias for the narrow transistors cannot be described by this mechanism involved in the efficiency of the lateral n-p-n structure. Hence, to comprehend the underlying physical

mechanism that leads to early ESD failure, electro-thermal transient simulations have been performed for a structure that was devised using *TSUPREM-4* [58]. Due to the relatively lower thermal conductivity of the passivation layers and shallow trench isolation (STI) structures, all the generated heat in the device is assumed to flow towards the substrate of NMOS structure, and the temperature of the substrate is set to the ambient temperature. This thermal boundary condition reasonably represents the actual wafer level test environment, and is verified by using the thermal simulator, *ANSYS* [59].

As shown in Fig. 4.14, the current density within the source/drain extension junction depth is strongly modulated by gate bias. This implies that the distribution of the local temperature rise in the drain extension and the channel area (indicated by the rectangle in NMOS) can also be influenced by the applied gate bias. At a given drain current of 10 mA/μm (at t=10 ns), the local temperature rise in the box, depending on the gate bias, is shown in Fig. 4.15.

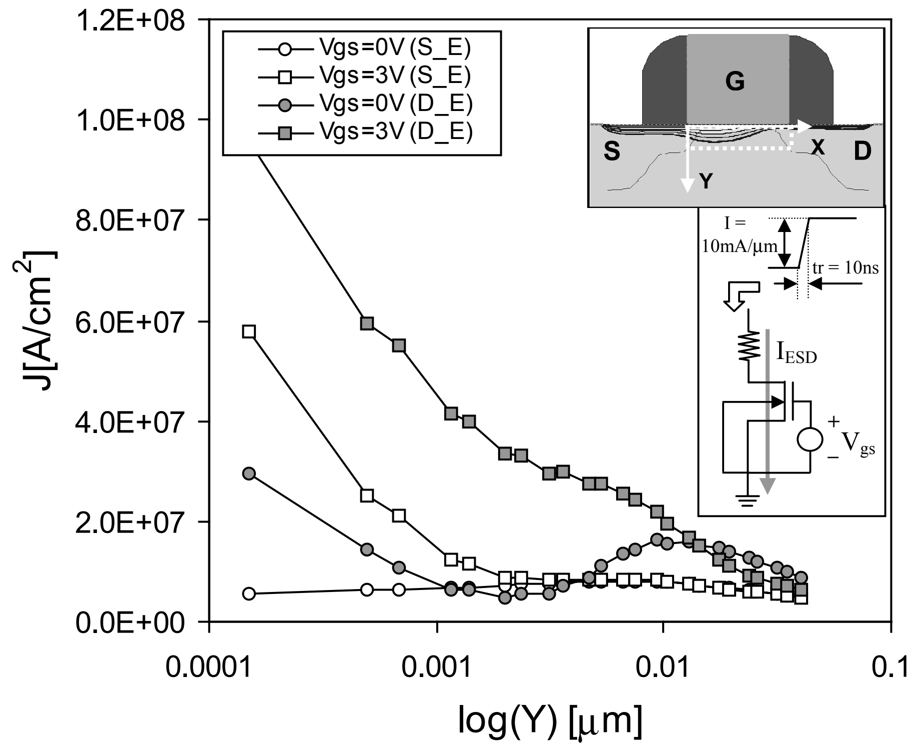
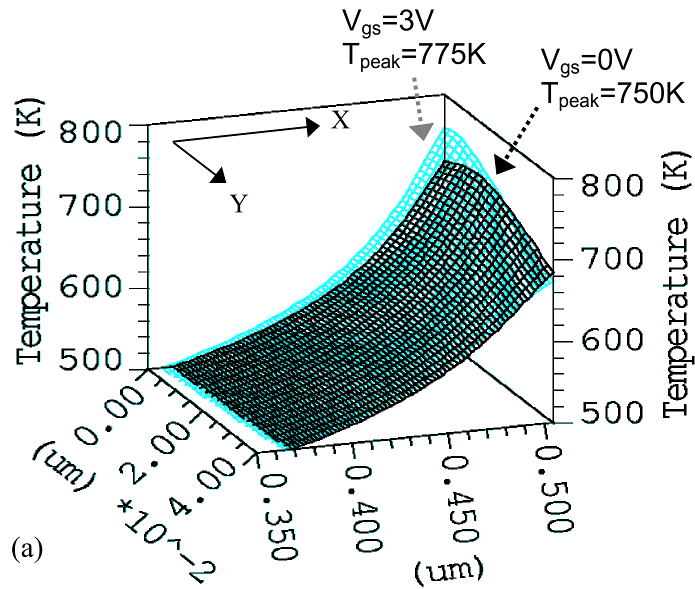


Figure 4.14: The current density at the edge of source (S_E) and drain (D_E) extension along y-axis with gate bias (see the *x* and *y* axis in the rectangle underneath the gate) and the schematic of the transient simulation circuit is shown in the inset.



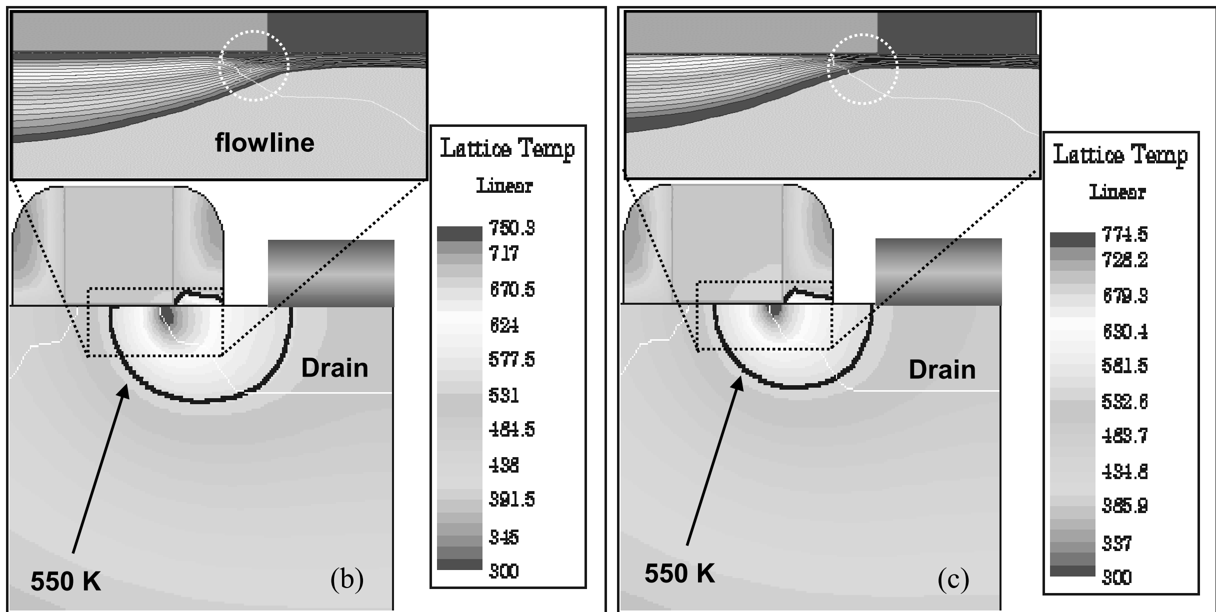


Figure 4.15: The local temperature rises with (a) V_{gs}=0 V and V_{gs}=3 V, and temperature contours with (b) V_{gs}=0 V and (c) V_{gs}=3 V. The simulations indicate that the channel and drain extension area undergo more heating and that the temperature distribution is more localized with gate bias.

The simulation results show that the distribution of the local temperature near the channel area as well as the peak temperature increases as gate bias increases. Also, the isothermal line of 550 K shows that the temperature distribution becomes more localized with the gate bias. In addition, as shown in Fig. 4.16, the location of the maximum temperature has been simulated with gate bias. For the negative gate bias, the location of the peak temperature does not change at all. This simulation result suggests that I_{t2} remains the same with the negative gate bias. This observation agrees well with the measured data in Fig. 4.12.

However, the surface heating becomes stronger with the gate bias since the location of the peak temperature approaches the Si/SiO_2 interface. The overall simulation results indicate that more heat can be accumulated within a smaller volume near the surface with gate bias, and the device tends to be more vulnerable to thermal failures at the surface. Hence this gate bias induced heating effect can lead to I_{t2} degradation in devices where the lateral ESD currents flow uniformly. At this point, it is instructive to note that a recent work has reported that for sub-180 nm devices heat diffusion equation underestimates the maximum device temperature compared to the one based on phonon Boltzmann transport equation [60].

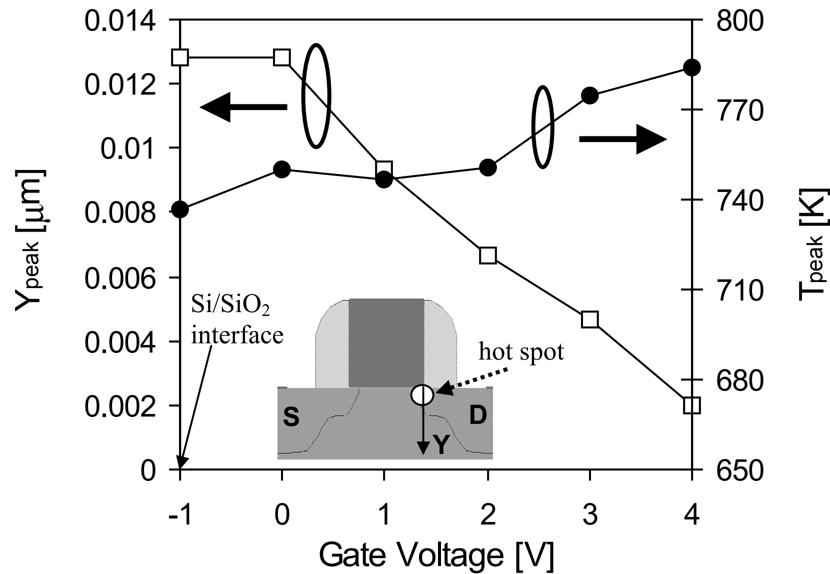


Figure 4.16: The location of peak temperature and the peak temperature value with gate bias. With increase in gate bias, Y_{peak} moves closer to the surface and the peak temperature, T_{peak} also increases moderately.

Therefore, the temperature value from the device simulation (*MEDICI* [61]) may not be accurate enough to predict the exact failure threshold of the device, but it provides a relative dependence of thermal effects on the bias conditions for a structure of interest.

To verify this heating effect, I_{t2} was also measured with both the gate and the substrate bias. As shown in Fig. 4.17, with substrate bias, I_{t2} increases since the maximum turned on width of the NMOS device is increased. Also, with substrate bias, the ESD currents spread out deeper into the silicon substrate despite the presence of gate bias as shown in Fig. 4.18.

As a result, the effective volume for power dissipation can be augmented. This suggests that the reduction in I_{t2} with gate bias could be alleviated by applying appropriate substrate bias, since the lateral ESD currents conduct more deeply into the silicon substrate with reduced current density near the surface, and in turn the heating is reduced as shown in Fig. 4.19. The experimental I_{t2} values confirm our argument as shown in Fig. 4.20. Despite high gate bias, the degradation of I_{t2} for both low and high voltage transistors is mitigated with substrate bias.

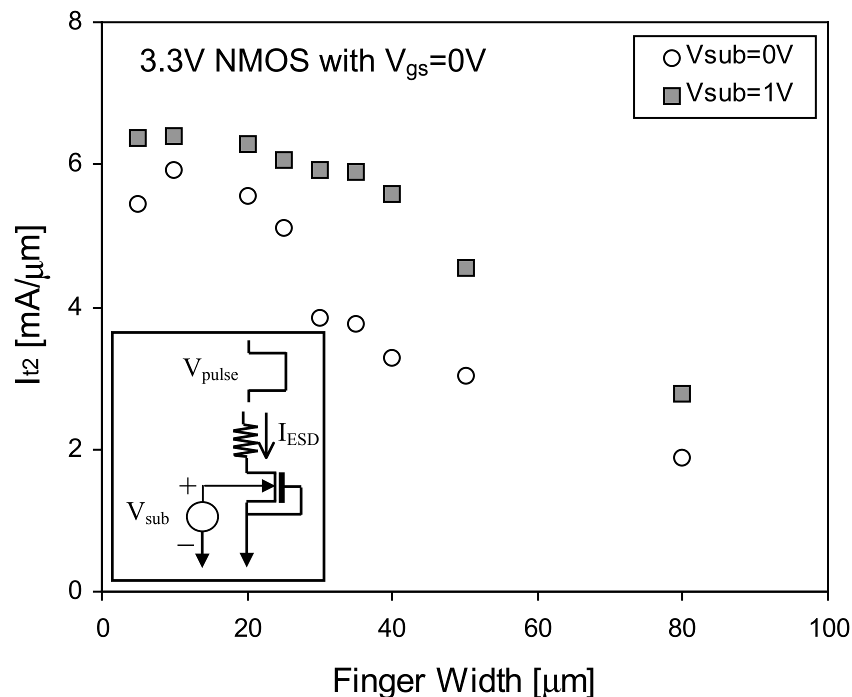


Figure 4.17: Impact of substrate bias on the I_{t2} with finger widths.

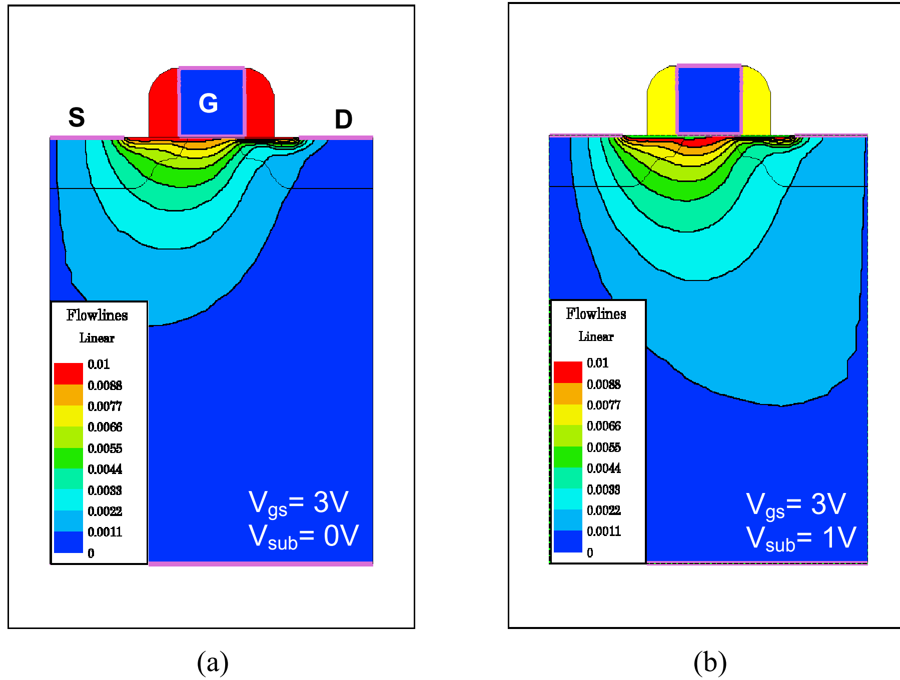
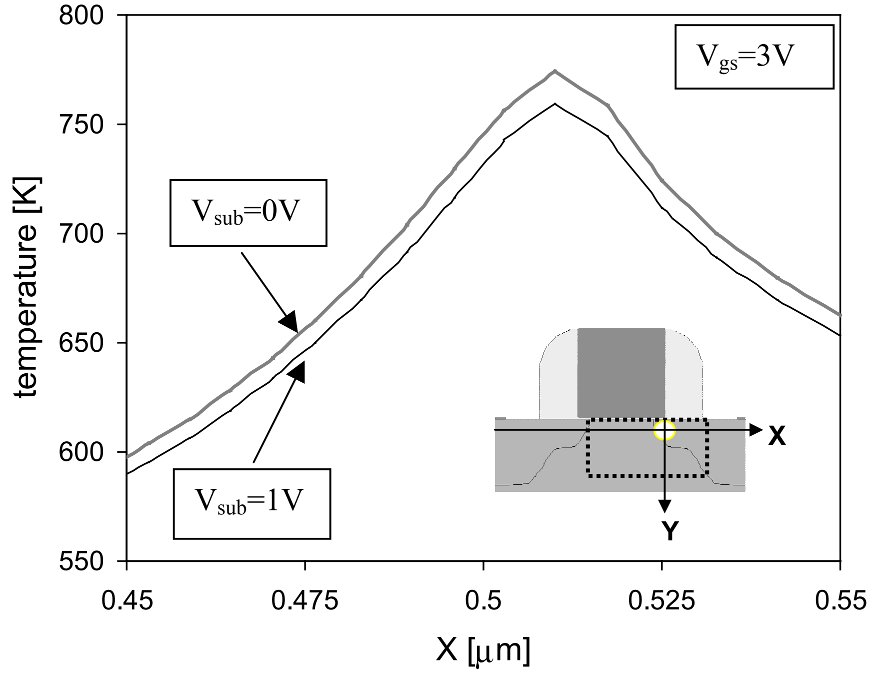


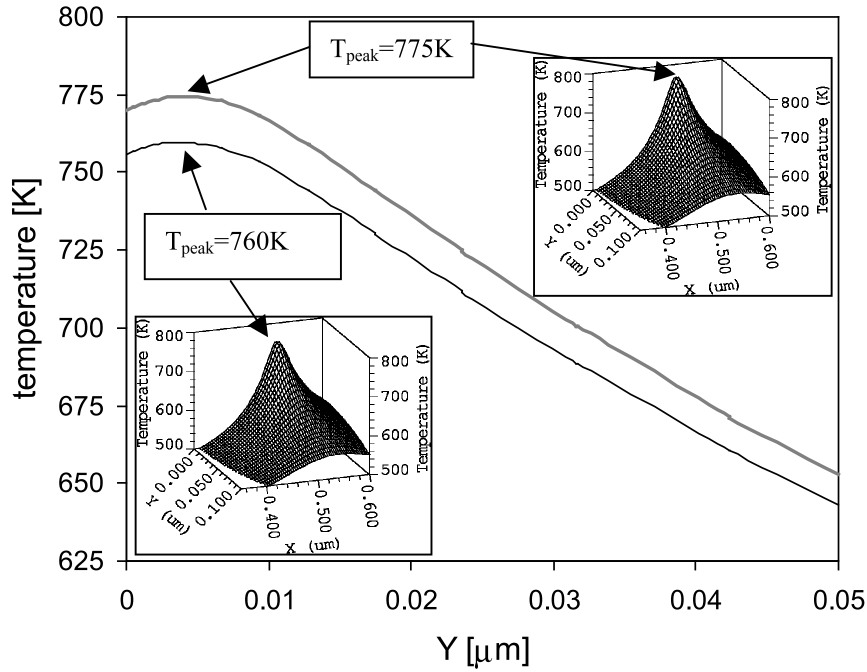
Figure 4.18: Simulated current flowlines in the presence of gate bias with (a) V_{sub}=0 V and (b) V_{sub}=1 V. With substrate bias, the current flows more deeply into substrate and this leads to increase in effective volume for power dissipation.

Hence, using the insightful results from device simulation and experimental data, it can be concluded that gate bias induced heating effect primarily accounts for the reduction in I_{t2} for devices with uniform lateral ESD current conduction. Based on these results, useful design consideration can be drawn for advanced salicided ESD protection. Hence for the substrate trigger protection [28, 47], I_{t2} roll-off with gate bias is not important. In fact, protection can be designed with the gate grounded NMOS structures as long as substrate bias is supplied for an efficient multi-finger n-p-n. On the other hand, for the design of the gate coupled ESD protection devices without substrate bias, the gate of the protection device should be designed with appropriate values of R and C [Fig. 4.7 (b)] to maintain the gate bias below the level above which I_{t2} begins to roll-off with the gate bias. Finally, for the output transistor, since the substrate bias is not available and the gate coupling is unpredictable, the buffer size should be designed based on the failure current component that it can handle, which depends on its gate coupling level.

For the above described ESD protection design, the gate coupling level and the optimum size of devices can be determined with high current ESD circuit simulations, which show the primary ESD current paths as well as the weak points for ESD [9].



(a)



(b)

Figure 4.19: Temperature distribution with substrate bias in the presence of gate bias. (a) temperature distribution along the x-axis and (b) temperature distribution along y-axis where T_{peak}=775 K for V_{sub}=0 V and T_{peak}=760 K for V_{sub}=1 V.

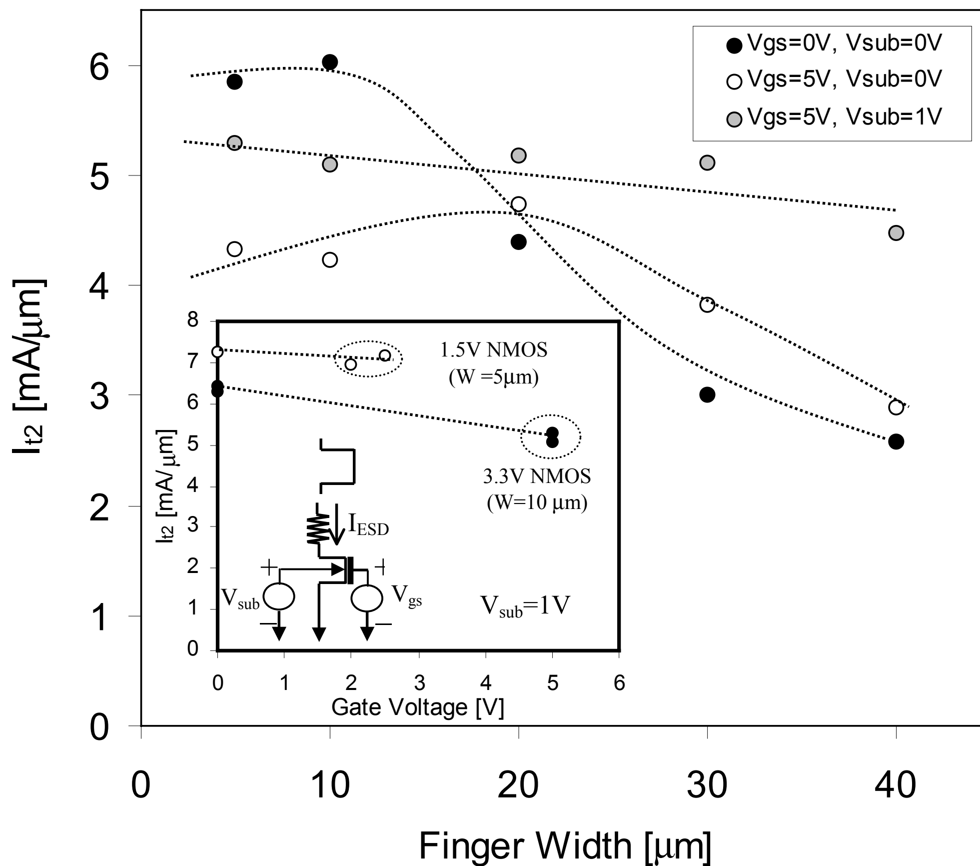


Figure 4.20: Effect of substrate bias on the I_{t2} degradation for high voltage (3.3 V) NMOS transistors with different finger width and gate bias, which shows that I_{t2} degradation is alleviated with substrate bias. The inset also shows the impact of substrate bias on I_{t2} degradation for narrow finger, high and low voltage transistors.

4. 3. Design Window for Advanced ESD Protection

EMMI analysis and TLP measurements show that different bipolar turned-on widths can be obtained depending on the substrate and gate bias conditions at a given ESD current level. Therefore, these facts indicate that a higher ESD failure threshold for a given protection structure can be realized by wider current conduction under ESD stress using appropriate biasing scheme.

As investigated previously, the substrate triggered protection is highly effective in overcoming non-uniform bipolar conduction, which is turned out to be a root cause of severe degradation of ESD performance in advanced silicided technology. In addition to the impact of substrate bias,

the gate bias dependence has been more clearly quantified with the experimental results for 3.3 V NMOS transistor in Fig. 4.21. Depending on the finger width, two competing trends of the gate bias effect are clearly observed. As can be seen, the gate bias is only effective in raising ESD hardness where ESD current is non-uniformly distributed (i.e., for the device with wider finger width). Otherwise, the gate bias even degrades the ESD strength by increasing the local temperature rise.

Therefore, combining the overall results and considering both the impact of the gate bias and the substrate bias, a design window for advanced protection devices can be presented as shown in Fig. 4.22. It should be noted that for substrate trigger protection [28, 47], the I_{t2} roll-off with gate bias is less important. Since the substrate bias can compensate for the adverse effects of gate bias, protection devices can be designed with either the gate grounded or gate coupled configurations, as long as the substrate bias is efficiently supplied for multi-finger n-p-n structures.

On the other hand, for the design of gate coupled ESD protection devices without substrate bias, the gate potential of the protection device should be controlled with reasonable values of coupling resistance and capacitance to maintain its value below the level above which I_{t2} begins to roll-off with the gate bias.

4. 4. Summary

An extensive investigation into impact of the substrate bias and gate bias has been carried out to provide improved understanding of ESD behavior and new insight into the bias effect involved in advanced salicided NMOS devices.

It is shown that ESD performance can be improved with the substrate bias by enlarging the effective turned-on device width. Additionally, the concept of an intrinsic second breakdown triggering current (I_{t2i}) is introduced, which is substrate bias independent and represents the maximum achievable ESD failure strength for a given technology.

It is also shown that gate bias induced heating near the drain extension region close to the Si/SiO_2 surface is the primary cause of the degradation of ESD performance for the devices with uniform bipolar conduction. Moreover, it is established that substrate biasing can help eliminate the negative impact of the gate bias effect. Results from this work can be used to construct suitable design windows for efficient and robust ESD protection design to overcome ESD failures in advanced deep submicron CMOS technologies.

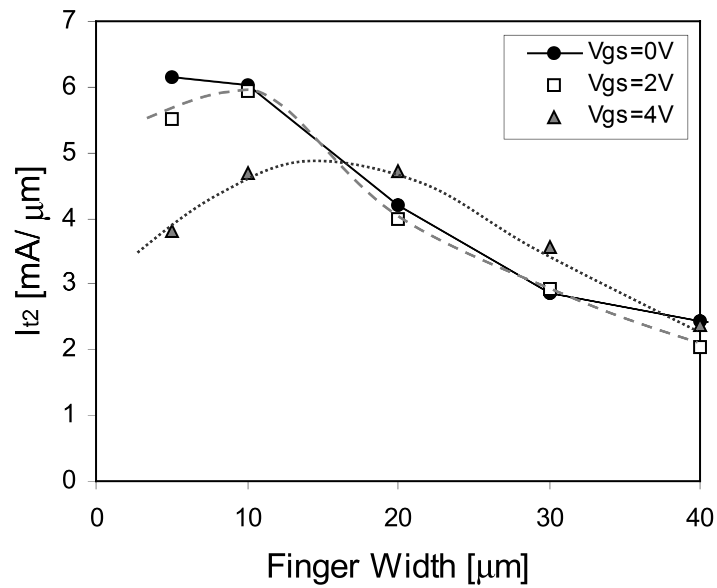


Figure 4.21: I_{t2} of the high voltage (3.3 V) NMOS transistor with finger widths for the various gate voltages. Two competing trends are clearly shown.

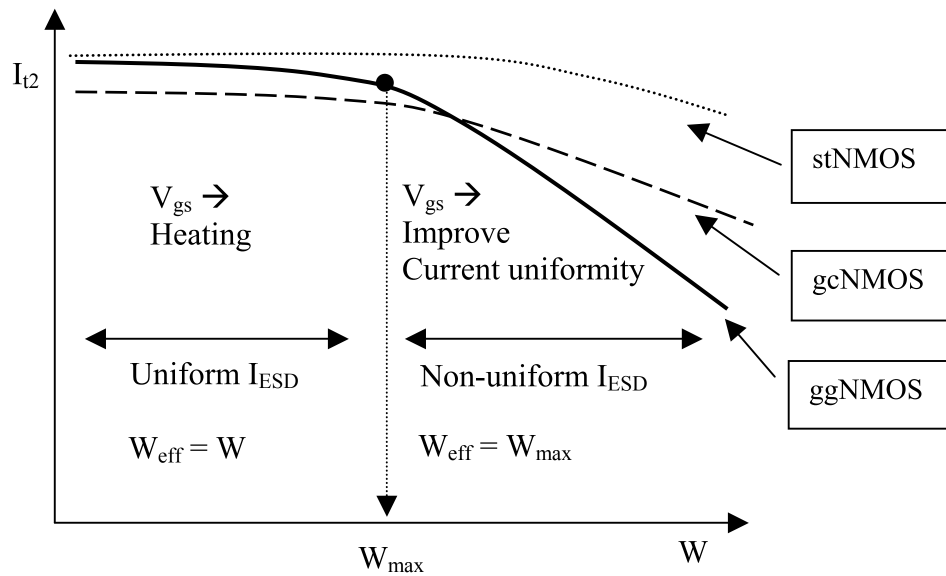


Figure 4.22: Design window for optimizing the performance of deep submicron ESD protection.

CHAPTER 5

IMPACT OF GATE-TO-CONTACT SPACING

ESD failure threshold of NMOS transistors is known to degrade with the use of silicided diffusions owing to insufficient ballast resistance, making them susceptible to current localization, which leads to early ESD failure. In general, the gate-to-contact spacing of salicided devices is known to have little impact on their ESD strength. However, experimental results presented in this chapter show that the ESD strength depends on the gate-to-contact spacing independent of the silicided process. Subsequently, a detailed investigation of the influence of gate-to-source and gate-to-drain contact spacings is carried out for the salicided 0.13 μm CMOS technology that provides new insight into the behavior of deep submicron ESD protection devices. It is shown that the reduction in current localization and increase in the power dissipating volume with increase in the gate-to-contact spacings are the primary causes of this improvement, which implies that even for silicided processes, the gate-to-contact spacing should be carefully engineered for efficient and reliable ESD protection designs.

5. 1 Introduction

NMOS transistors are widely used as protection devices against electrostatic discharge (ESD), which is a major reliability concern for all categories of integrated circuits [57]. It is well known that for non-silicided or silicide-blocked NMOS protection transistors, the second breakdown triggering current (I_{t2}), which is widely used to monitor the ESD strength, can be increased with larger drain contact spacing because of more uniform triggering of the lateral n-p-n structure obtained with ballast resistance[21] effects [57]. In addition, it is also well established that effectiveness against ESD is reduced in the case of devices with silicided diffusions [14, 47, 57] since the ballast resistance is negligible and even zero ESD performance is reported for standard silicide devices [62].

In silicided CMOS processes, as discussed so far, the primary cause of the degradation of ESD failure threshold is known to be non-uniform lateral bipolar conduction, which is attributed to insufficient ballasting resistance in the fully silicided source/drain structures [47, 48]. This decrease in ESD strength imposes severe restrictions on the efficient design of ESD protection. Therefore, to avoid localized current conduction and improve I_{t2}, device structures with sufficient ballasting resistance are realized by introducing the silicide blocking option, or by implementing well resistors on the drain side, or by inserting local interconnect layers [14, 33, 63].

However, these options either require an extra mask or more process complexity and result in increased process cost and chip area. Hence, use of salicided devices is often preferred for cost effectiveness in providing advanced ESD protections. The I_{t2} of the silicided devices is generally believed to be independent of the gate-to-contact spacing parameter. The MM and HBM tests of a salicided 0.18 μm CMOS technology have also shown that the gate-to-drain contact spacing has negligible impact on ESD hardness [64]. Therefore, the impact of the gate-to-contact spacing of a salicided NMOS transistor on the ESD failure strength has been overlooked.

However, contrary to conventional understanding, for advanced deep submicron salicided technologies with shallow trench isolation (STI) structures, we have recently reported that the gate-to-drain contact spacing (GDCS) has an impact on I_{t2} [65, 66]. Moreover, the gate-to-source contact spacing (GSCS) has also been observed to affect I_{t2} [65, 66].

This chapter investigates the above new phenomenon in advanced salicided transistors and describes the different mechanisms that are observed at the source and drain sides, respectively. The physical mechanism causing unexpected I_{t2} improvement in silicided devices with increased

[21] Ballast resistance effect makes the ESD currents flow more uniformly in the drain diffusion region.

gate-to-contact spacing is identified. Furthermore, it is shown that the ESD strength of the protection device becomes independent of the gate-to-contact spacing when adequate substrate bias is applied. These observations have significant implications for ESD performance improvement simply through optimization of the device layout, even without introducing expensive process options.

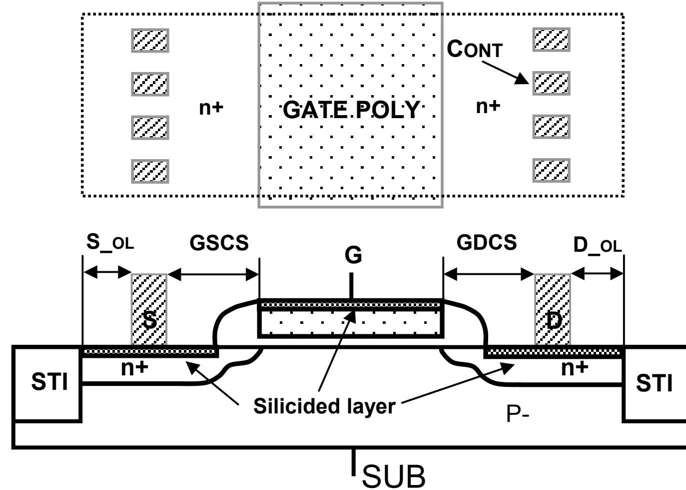


Figure 5.1: The schematic of a silicided NMOS transistor indicating the gate to source/drain contact spacing (GSCS/GDCS) and the n^+ overlap of the source/drain contact ($S_{_OL}/D_{_OL}$). For the 1.5 V NMOS, the S/D contact opening (S/D) = 0.15 µm and $S_{_OL} = D_{_OL}$ = 0.1 µm. For the 3.3 V NMOS, the S/D contact opening (S/D) = 0.15 µm and $S_{_OL} = D_{_OL}$ = 0.125µm. In the test structures, the S/D contact opening and n^+ overlap of S/D contact ($S_{_OL}/D_{_OL}$) remain unchanged despite the variations of GSCS/GDCS.

5. 2 Experiments

5. 2. 1 I_{t2} Dependence on Contact Spacing

In this work, the dependence of the second breakdown triggering current, I_{t2} [mA/µm], is investigated as a function of the gate-to-source/drain salicided ($CoSi_2$) contact spacing (GSCS/GDCS) for various test structures with grounded gates.

Test structures were fabricated using the 0.13 μm mixed-voltage CMOS technology and the low (1.5 V) and high (3.3 V) voltage NMOS transistors are investigated. The 1.5 V NMOS transistor has a 27 Å thick gate oxide and 0.175 μm long gate poly, while the 3.3 V NMOS transistor has a 70 Å gate oxide and 0.5 μm long gate poly. However, the finger width for both transistors is 20 μm. Fig. 5.1 shows the schematic cross section of the ESD NMOS transistor used in this study.

The contact spacing is measured from the gate poly edge to the near edge of contact opening and the contact opening width is fixed at 0.15 μm for all the test structures. Since the test structure uses a shallow trench isolation (STI) with constant n^+ overlap with the source and drain contact ($S_{_OL}/D_{_OL}$) for a given device rating, the total size of the source and drain structures are changed proportional to the variation of the GSCS and GDCS.

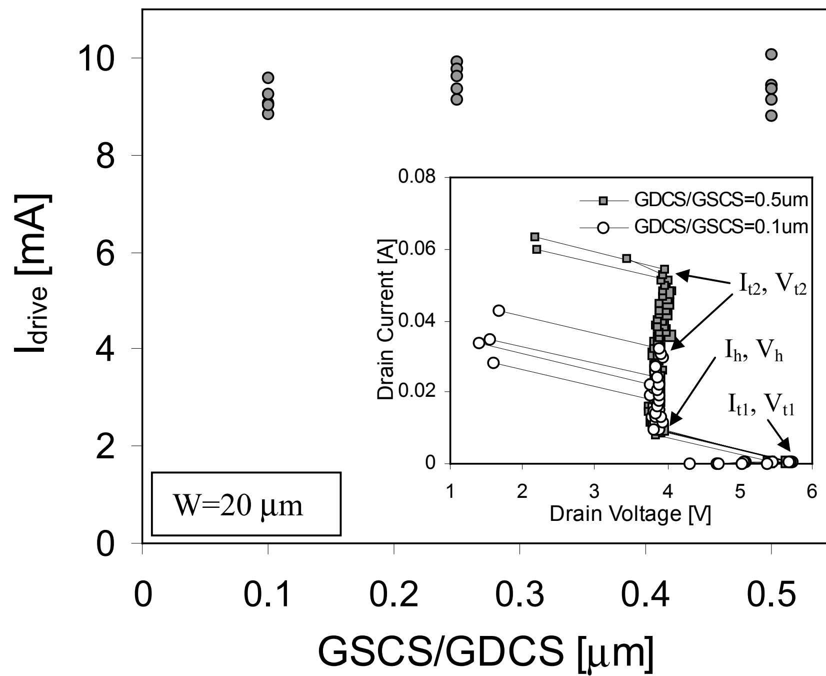


Figure 5.2: The measured drive currents (I_{drive}) for the 1.5 V NMOS transistors with different gate-to-contact spacings show no apparent differences, where V_{gs}=V_{ds}=1.5 V. The inset shows high current TLP curves for the 1.5 V NMOS transistor, which clearly show the impact of the gate-to-contact spacings on I_{t2} despite the silicided diffusion.

The second breakdown triggering current (I_{t2}) was measured using the transmission line pulsing (TLP) method for a voltage pulse width of 200 ns. As expected, since the resistance of silicided region is relatively small compared to other parasitic resistances in the source/drain structure, a change in resistance proportional to the GDCS/GSCS is not apparent from the DC I-V measurements. The drive current (I_{drive}) of the 1.5 V NMOS transistor was tested with GDCS/GSCS variations.

As shown in Fig. 5.2, the measured drive currents show that the difference in the resistance due to increased gate-to-contact spacing with the salicided diffusion is hardly apparent. In addition, the inset in Fig. 5.2 shows a sample of TLP curves for the 1.5 V transistor with two different GDCS and GSCS. It can be clearly seen that the I_{t2} value is about doubled with an increase in GDCS/GSCS from 0.1 μm to 0.5 μm. However, the slope of the high current regions is almost identical for the two different test structures, which implies that the on-resistance of the test structures is nearly the same despite the different gate-to-contact spacings.

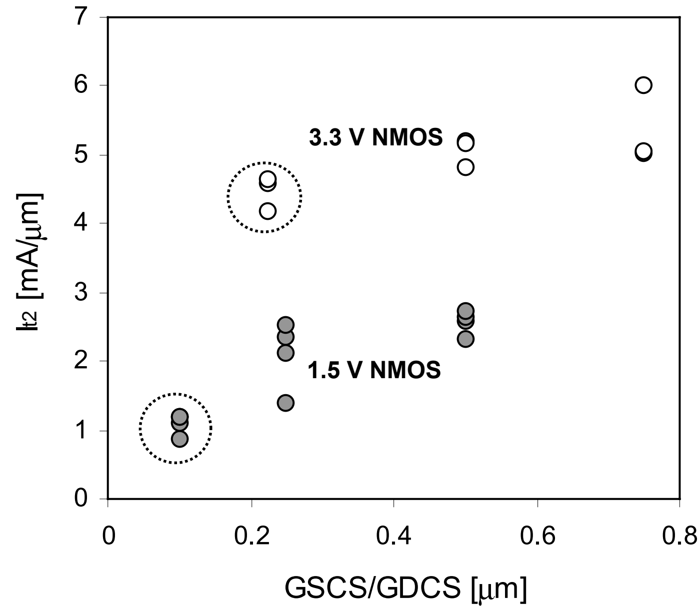


Figure 5.3: The second breakdown triggering current (I_{t2}) with the gate-to-source/drain contact spacing for two different silicided NMOS transistors. The two dotted circles indicate I_{t2} for each device with minimum contact spacing: 0.1 μm and 0.225 μm for the 1.5 V NMOS (W/L_{poly} = 20/0.175 μm) and 3.3 V NMOS (W/L_{poly} = 20/0.5 μm), respectively.

For the TLP measurements of the 1.5 V and 3.3 V transistors as shown in Fig. 5.3, I_{t2} of both the devices surprisingly improves with the GSCS/GDCS. Compared with I_{t2} values of the transistor with the minimum gate-to-contact spacing (0.1 μm for 1.5 V NMOS and 0.225 μm for 3.3 V NMOS), the improvement of I_{t2} is approximately 100 % and 40 % for the low voltage and high voltage transistors, respectively. This implies that the gate-to-contact spacing is an important design parameter determining ESD strength for the gate grounded ESD protection devices. It also suggests the possibility of achieving increased ESD robustness through optimizing the layout of the silicided protection devices without any extra processing steps or structure options.

However, the primary cause of this improvement of I_{t2} has not been explored and needs comprehensive modeling and analysis in order to improve understanding of the device physics involved in this effect. This will also enable the establishment of robust ESD protection design approaches through proper design of the devices.

As shown in Fig. 5.3, for the 1.5 V NMOS transistors, the I_{t2} dependence on the gate-to-contact spacing is more apparent than that for the 3.3 V devices. In general, for advanced devices with salicided diffusion, the improvement of I_{t2} is not easily achieved due to early failure caused by current localization. In this regard, any amount of improvement in I_{t2} for salicided technology with no process changes has significant implications.

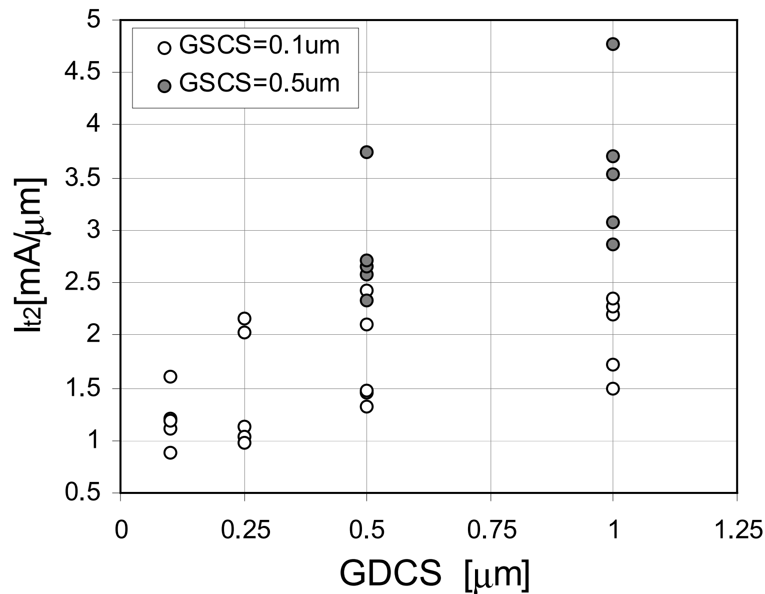


Figure 5.4: I_{t2} values for the 1.5 V NMOS transistors with various GDCS and GSCS. I_{t2} depends on both GDCS and GSCS within the scatter of data. W= 20 μm.

As can be seen in Fig. 5.4, I_{t2} values are influenced by both GDCS and GSCS within the scatter of data. Despite the salicided process, the increasing trend of I_{t2} is obvious as GDCS and GSCS increase. For the minimum gate-to-contact spacing of 0.1 µm, I_{t2} ranges from 1 to 1.5 mA/µm. However, with the increased GDCS=1 µm and GSCS=0.5 µm, I_{t2} values are clustered around 3.5 mA/µm. The data clearly show that for a given spacing (GDCS), increases in GSCS strongly affect the ESD hardness. This experimental result implies that the impact of gate-to-source contact spacing is as significant as that of the gate-to-drain contact spacing for a salicided technology.

5. 2. 2 Effect of Substrate Bias

For silicided devices, it has been reported that I_{t2} improves with a forward substrate bias by enlarging the turn-on finger width of the devices [47]. With sufficient external substrate bias (V_{sub} > 0.7 V), the It_2 dependence on the gate-to-contact spacing disappears as the emitter-base junction of the lateral bipolar transistor fully turns on [Fig. 5.5].

As shown in [47], the phenomenon of I_{t2} improvement is associated with the extent of uniformity of the lateral current distribution, since the ESD current becomes more uniform along the channel width with increased external substrate bias. In addition, it should be noted that the I_{t2} dependence on the gate-to-contact spacing for the 1.5 V transistor is stronger, since the extent of uniformity of the lateral current distribution of the two devices varies [Fig. 5.6]. It is believed that the devices with shorter channel length and shallower junction depth experience non-uniform bipolar conduction more strongly since the relative sensitivity to the statistical random variation is higher for given process conditions.

In order to identify the underlying physical mechanisms for the drain side and source side contact spacing effects respectively, the following effects were investigated:

1) the influence of the ballasting ESD current distribution.
2) the improvement of the current driving capability of the lateral n-p-n transistor.
3) the increase in thermal capacity due to the enlargement of power dissipating volume along with the increase in GSCS/GDCS.

Each of these effects has been investigated in detail in the next section.

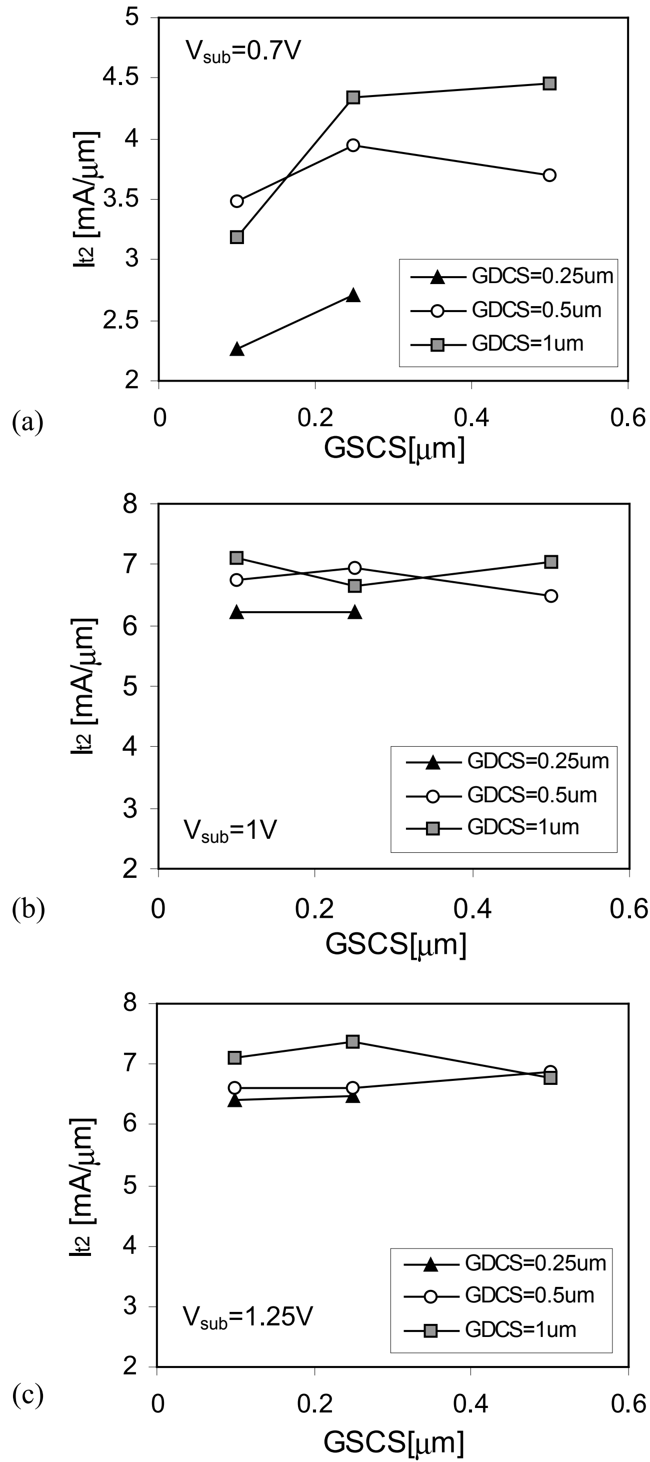
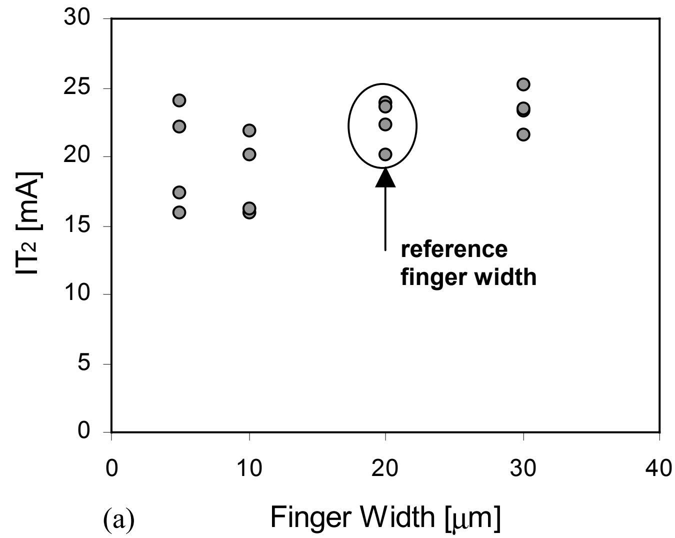


Figure 5.5: (a) I_{t2} of the 1.5 V NMOS transistors is dependent on GSCS and GDCS with an external substrate bias of 0.7 V, and is nearly independent of GSCS and GDCS with sufficient external substrate bias of (b) 1 V and (c) 1.25 V.



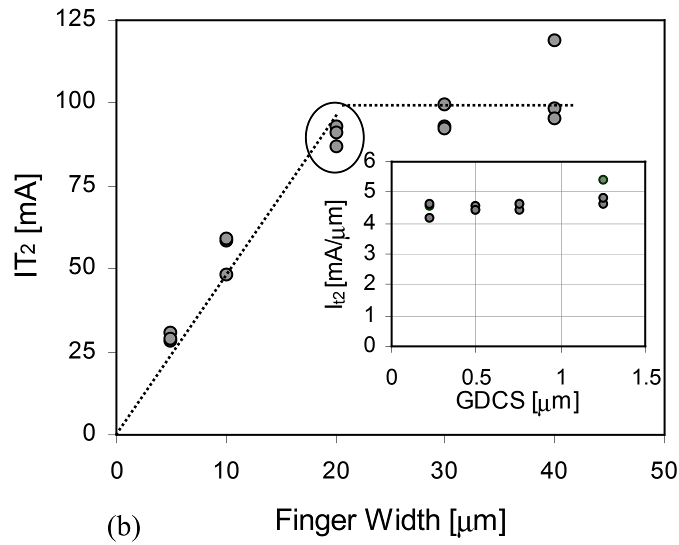


Figure 5.6: Total failure threshold current (IT_2) with the finger width for (a) 1.5 V NMOS transistors and (b) 3.3 V NMOS transistors where GDCS/GSCS= 0.1 μm; the total failure current can be scaled with finger width within only a limited finger width (W ≤ 20 μm) and the inset shows that the 3.3 V device (W= 20 μm and GSCS= 0.225 μm) with nearly uniform current conduction shows negligible dependence of I_{t2} on GDCS.

5. 3 Analysis and Discussion

5. 3. 1 Ballasting Current Distribution

As shown in Fig 5.2, from the DC and TLP measurements for 1.5 V NMOS transistors, the drive current and the on-resistance show no differences between the structures with the minimum gate-to-contact spacing and the ones with increased gate-to-contact spacing as expected. First, to analyze the impact of the GDCS, for a fixed GSCS of 0.1 μm, I_{t2} was measured for the structures with GDCS of 0.1 μm and 1 μm, respectively. As can be observed from Fig. 5.7, consistent with Fig. 5.2 (inset), even with further increase in GDCS, the on-resistance remains unaffected.

However, I_{t2} is nearly doubled with GDCS=1 μm. It is instructive to note that, contrary to the experimental results presented in this work, it has been reported in the past that the on-resistance is changed depending on the silicide thickness and junction depth [14].

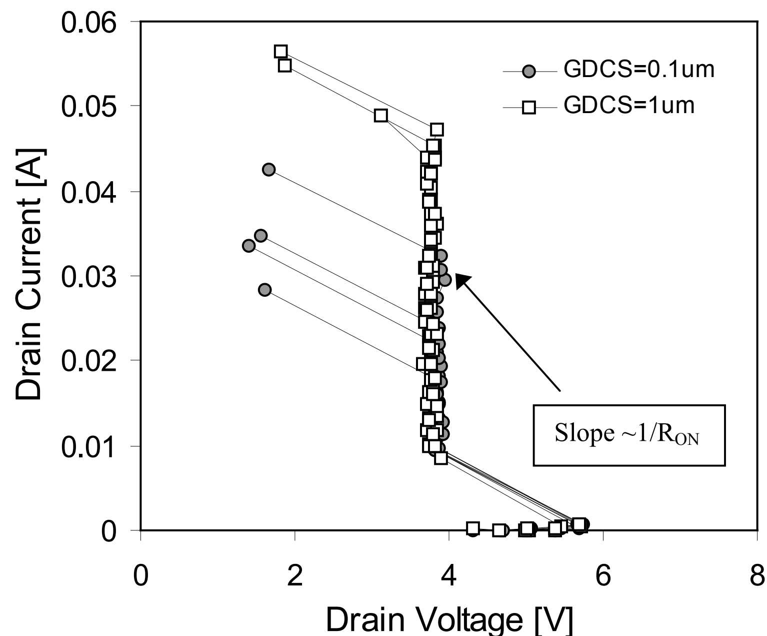


Figure 5.7: High current TLP curves for the 1.5 V NMOS transistor with two different GDCS of 0.1 μm and 1 μm where GSCS is 0.1 μm. The slope of high current regions is almost identical, but I_{t2} is nearly doubled with the increased GDCS.

Since it is well known that the silicide contact resistance is a strong function of interfacial doping concentration (Silicide/Si), the silicide contact resistance is influenced by the variation of the silicide thickness [67]. Moreover, the junction depth can also affect the drain diffusion resistance. Therefore, it is likely that the impact of these factors was observed as slight changes in the on-resistance in [14], even with silicided structures.

Nevertheless, based on the empirical results presented in this work, it can be conjectured that the increase of GDCS alleviates the current localization problem by further expanding the turned on portion of the finger width, which appears only for devices with strongly non-uniform current distributions under ESD stress. For devices with various finger widths, total failure threshold currents, IT_2 [mA], were measured for the low and high voltage transistors. As can be seen in Fig. 5.6, the total failure threshold current does not scale with the drawn finger width for the 1.5 V transistor. For the 3.3 V transistor, IT_2 scales only for a limited range of widths. This implies that the ESD current distribution for 1.5 V NMOS devices is highly non-uniform while it is nearly uniform for the narrow 3.3 V NMOS transistors ($W \leq 20$ μm).

The I_{t2} data shown in Fig. 5.6 (b) support the observation that only devices with uniform ESD current distribution do not show dependence on GDCS. Thus, it can be concluded that the increase in GDCS helps spread out the ESD current more uniformly along the finger width, which leads to improvement in the effective value of I_{t2}, though the increase in the ballasting resistance with the GDCS is hardly noticeable.

This is apparent from Fig. 5.3 and from the total failure threshold current (IT_2) data in Fig. 5.6 (a) for the 1.5 V NMOS devices where the ESD current distribution is strongly non-uniform. In addition, for low voltage transistors, with adequate external substrate bias, the dependence of I_{t2} on GDCS disappears as shown in Fig. 5.5. This result also supports the argument that the increase in GDCS alleviates current localization for low voltage transistors. In other words, it seems that for the salicided devices, the shorter the gate-to-drain contact spacing is, the stronger the current localization. Therefore for the devices requiring higher ESD strength, minimum gate-to-contact spacing should be avoided unless substrate bias can be used in the protection circuit design [28].

5. 3. 2 Characteristics of the Lateral n-p-n Transistor

In general, the ESD hardness of NMOS devices can be described in terms of the primary device parameters of the parasitic lateral n-p-n transistors, such as the current gain (β), avalanche multiplication factor (M), and effective substrate resistance (R_{sub}) [9, 10, 11]. When the lateral n-p-n turns on [Fig. 5.8], the M, β, and R_{sub} are given by [68]

$$M = \frac{(\beta + 1) I_D}{\beta (I_D - I_{sub})} \quad (5.1)$$

$$\beta = \frac{1}{(M - 1) - M (\frac{I_{sub}}{I_D})} \quad (5.2)$$

$$R_{sub} \approx \frac{0.8}{I_{sub}} \quad (5.3)$$

Despite variation of the GSCS/GDCS, R_{sub} can be assumed to remain constant, since the substrate doping concentration remains unchanged. In addition, the substrate contact is designed 20 μm away from the STI boundary at the source side and the maximum variation in the distance of substrate contact from the gate poly edge can be 0.65 μm due to increase in GSCS. This 0.65 μm variation in the distance of substrate contact will have a negligible impact on the effective substrate resistance value.

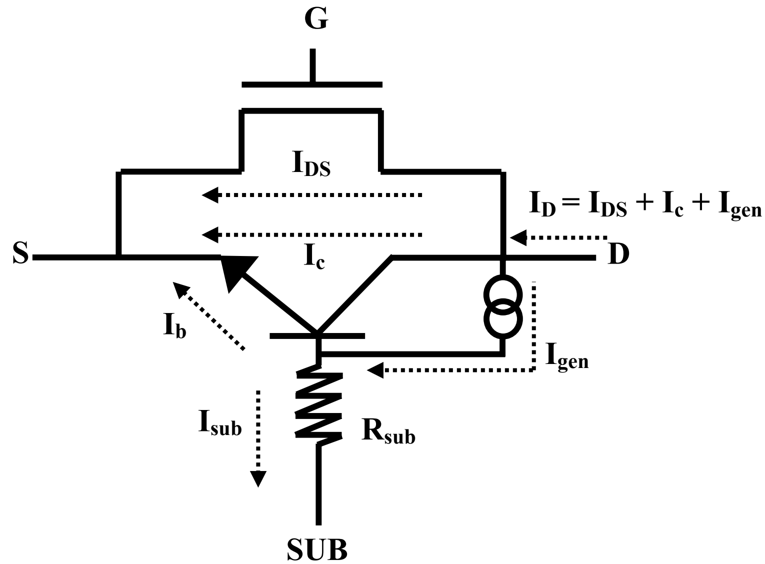


Figure 5.8: Equivalent circuit of the NMOS including the parasitic lateral n-p-n bipolar transistor when the lateral n-p-n is on. I_{DS} is the channel current, I_{gen} is the avalanche-generated current, and I_{sub} is the substrate current.

Therefore, the current gain and avalanche multiplication are of primary interest for studying the impact of variations of gate-to-contact spacing on I_{t2}. The lateral n-p-n operation depends on a combination of M, β, and R_{sub} for a given power dissipation. For a given ESD current, less avalanche multiplication and higher current gain is preferable for higher ESD strength, since strong avalanche multiplication results from high fields, which in turn, leads to locally higher temperatures.

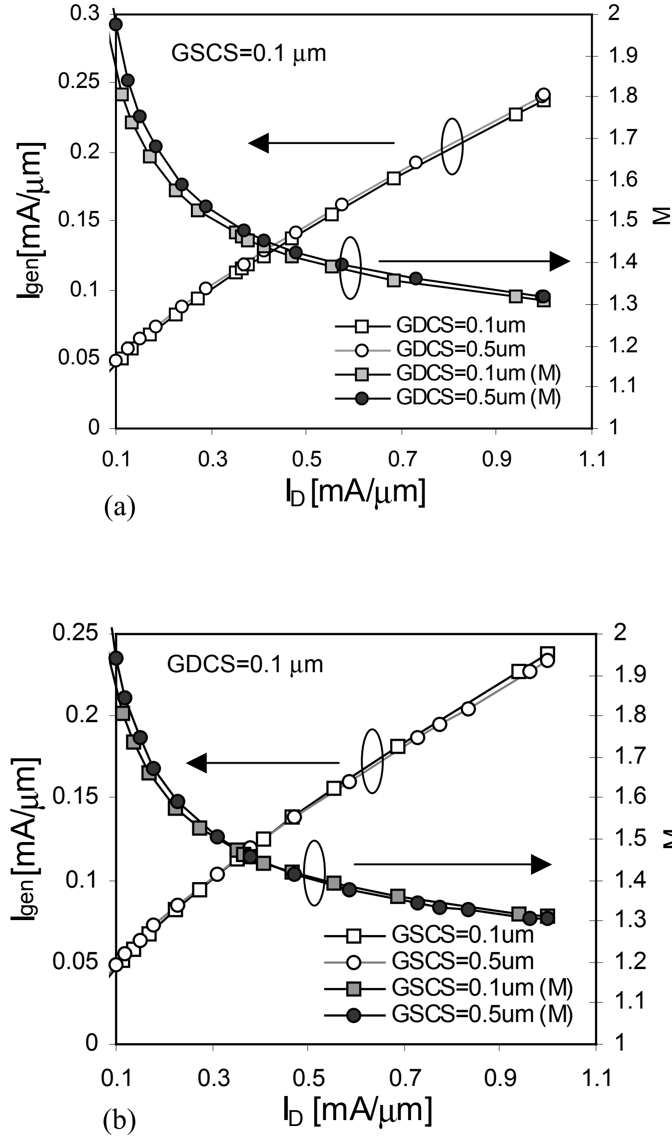


Figure 5.9: The avalanche-generation current (I_{gen}) and multiplication factor (M) for the variation of (a) the GDCS and (b) GSCS.

The variation of the avalanche multiplication for the test structures can be observed by employing 2D device (*MEDICI*) simulation. It should be noted that in the test structures used in this study, as the contact spacing is increased, the source or drain area did not remain constant since the distance from the edge of the gate poly to the STI boundary on the source and drain side also simultaneously increased [Fig. 5.1]. In addition, in the simulations, the silicide layers are treated as virtual electrodes and the effective substrate resistance of 5 kΩ/μm is attached to the bottom substrate contact.

For the thermal boundary conditions, a thermal electrode is defined at the bottom of the substrate and the temperature of this thermal electrode is assumed to be the same as the ambient temperature. It is important to note that we assume the simple thermal boundary conditions for the simulations since the results of the thermal simulations are intended to provide only a relative comparison between different structures for a given thermal boundary condition. Furthermore, from the thermal point of view, the size of the simulation structure should be fixed, otherwise the thermal boundary conditions change.

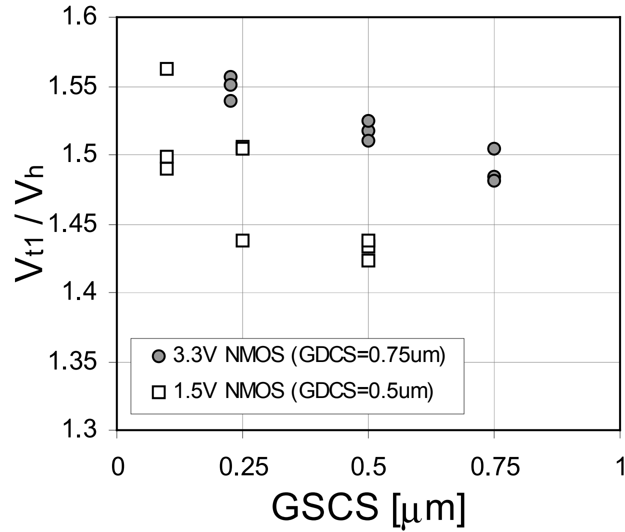
Using the DC current sweep simulation mode, high current characteristics were reproduced for the structures with minimum gate-to-contact spacings (GSCS/GDCS= 0.1 μm) and also for longer GDCS (= 0.5 μm) or GSCS (= 0.5 μm). As shown in Fig. 5.9, the avalanche-generated current (I_{gen}) and multiplication factor (M) are compared for the three cases. As the drain current increases the generation current increases while the multiplication factor decreases rapidly, regardless of the gate-to-contact spacing. Based on the simulation results, variations of the gate-to-contact spacing appear to have no impact on the avalanche process.

The increase in the current gain of the lateral n-p-n transistor can result in an improvement of I_{t2} by conducting more current for a given ESD stress. In order to track the current gain of a lateral n-p-n transistor with the variation of GSCS and GDCS, the ratio of the triggering voltage (V_{t1}) and holding voltage (V_h) is monitored for various test structures as shown in Fig. 5.10, since the current gain of a lateral n-p-n transistor (for a gate-grounded NMOS) in a self-biasing mode is proportional to the V_{t1}/V_h ratio [69],

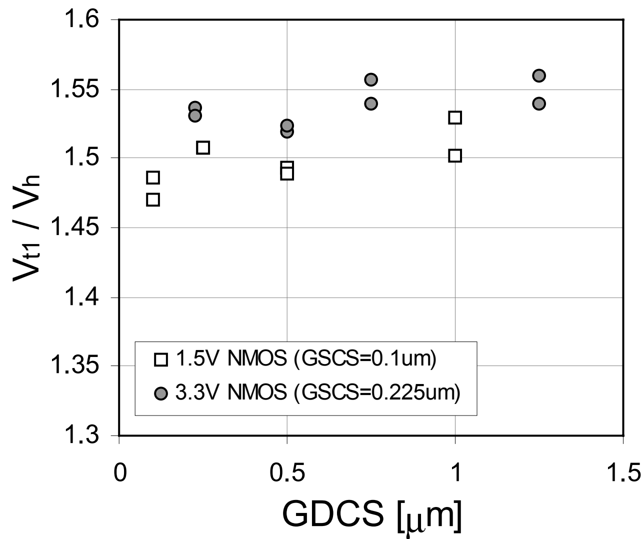
$$(1+\beta)^{\frac{1}{n}} = \frac{BV_{CBO}}{BV_{CEO}} \approx \frac{V_{t1}}{V_h} \tag{5.4}$$

where BV_{CBO} is the common-base breakdown voltage (with emitter open-circuited), BV_{CEO} is the breakdown voltage for the common-emitter configuration (with base open-circuited), and n is a constant. For a constant GDCS of both the low and high voltage devices, the ratio of V_{t1} and V_h

decreases with increase in GSCS. However, compared with the V_{t1}/V_h values for GSCS variations, the V_{t1}/V_h values are nearly independent of the GDCS variations.



(a)



(b)

Figure 5.10: To track the current gain of a lateral n-p-n transistor, the triggering voltage (V_{t1}) and holding voltage (V_h) are measured for the 1.5 V and 3.3 V NMOS transistors. The data clearly show that the current gain is reduced by increase in GSCS (a), while remains nearly constant with GDCS (b).

As described earlier, for the test structures, size of the source/drain is increased as the GSCS/GDCS increases. Therefore, the effective area of the emitter (source) of the lateral n-p-n is enlarged with increase in the GSCS, and in turn, the effective current path is also increased. As shown in Figs. 5.11 and 5.12, for a given collector current, with the increased effective size of the emitter, more base current (hole current component) flows into the emitter for a given generation current [Fig. 5.11], which results in a slight decrease in the current gain [Fig. 5.12].

However, despite the decreased current gain with GSCS, improvement in I_{t2} is observed with GSCS. Consequently, the increase of I_{t2} with the GSCS is not attributed to the lateral n-p-n current gain. These results suggest that the mechanism for improvement of I_{t2} with the gate-to-source contact spacing variations is not dependent on the efficiency of lateral parasitic n-p-n transistor.

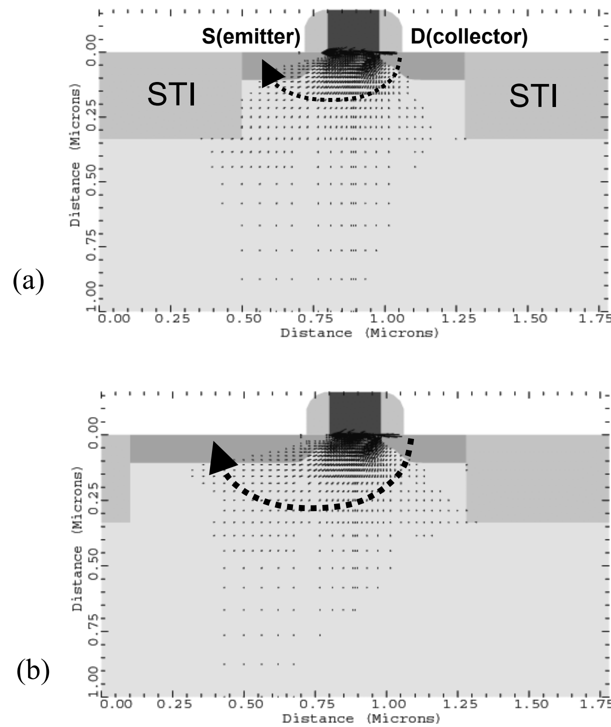
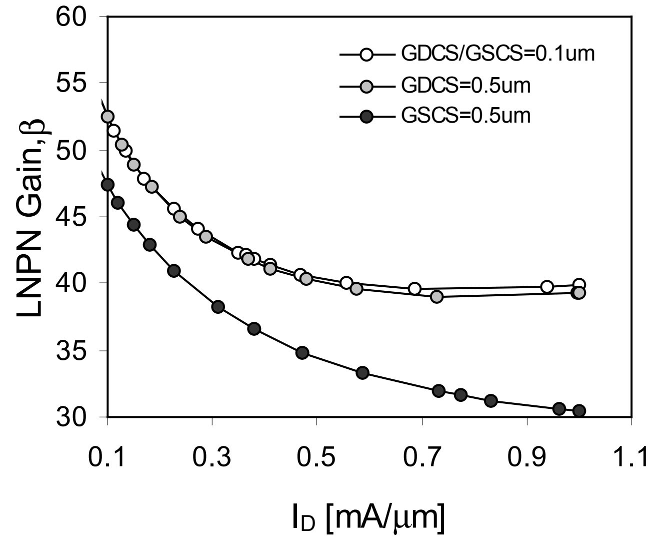
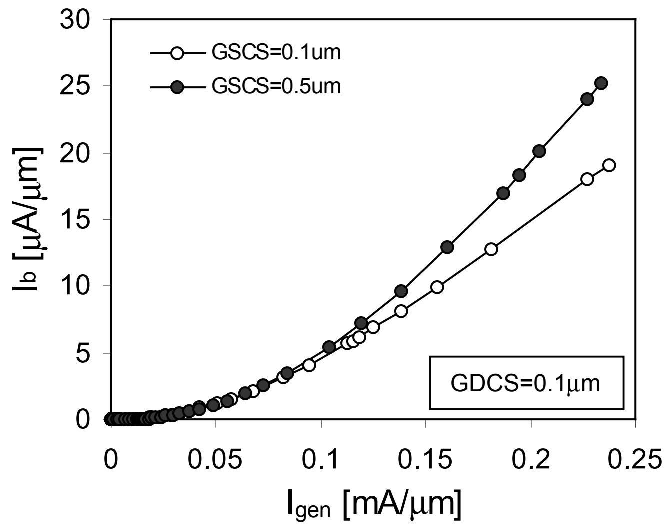


Figure 5.11: The base current (hole current) density vector at $I_D = 1mA/\mu m$ for the device with (a) minimum gate to contact spacing, GSCS/GDCS=0.1 μm and (b) increased gate to source contact spacing, GSCS=0.5 μm. As GSCS increases, wider emitter and base junction is utilized for the current conduction and this results in a drop in the current gain due to increase in base current for a given collector current. The dashed arrow is the schematic of base current.

In fact, the efficiency of the lateral n-p-n for conducting ESD current seems to be degraded or unchanged with increases in gate-to-source contact spacing. Therefore the main cause of the improvement in I_{t2} due to increase in GSCS is still unclear. Hence, to gain better insight into the physical behavior we next explored possible thermal effects involved in the gate-to-source contact spacing by performing electro-thermal simulations.



(a)



(b)

Figure 5.12: (a) The current gain (β) vs. the drain current (I_D), and (b) the base current (I_b) vs. the generation current (I_{gen}) for the device with different gate-to-contact spacings.

5. 3. 3 Thermal Effects

Before investigating any thermal effects involved in the mechanism of I_{t2} improvement with gate-to-source contact spacings, it is important to discern any possible role of the uniformity of current distribution (arising solely due to increase in GSCS or the effective emitter area of the lateral n-p-n transistor) as the primary physical mechanism responsible for the improvement. As shown in Fig. 5.6 (b), the current distribution is nearly uniform for the 20 μm wide high voltage (3.3 V) transistors and increase in GDCS has no impact on I_{t2} since the current is already uniformly distributed.

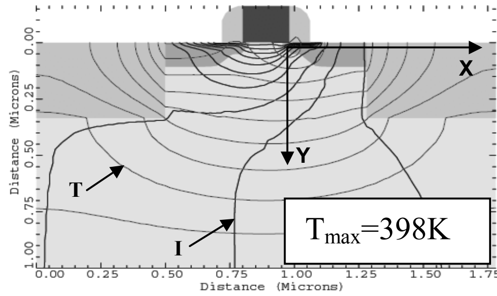
Using the same argument, increase in GSCS should also have little impact on the uniformity and on I_{t2}. Nevertheless, the I_{t2} data for high voltage devices in Fig. 5.3 clearly show the impact of GSCS even though the current is already uniform (since W=20 μm). Hence, it can be concluded that the observed improvement of I_{t2} with GSCS is not primarily due to any improved uniformity of current distribution. The above arguments are particularly valid for the high-voltage transistors.

In order to investigate any increase in thermal capacity due to enlargement of power dissipating volume along with the increase in GSCS/GDCS, the temperature distributions in the device have been compared using electro-thermal simulations. Current flowlines and temperature distribution contours for a drain current (I_D) of 1 mA/μm in the three different structures are shown in Fig. 5.13. According to the simulation results, the location of the maximum temperature in the device remains the same, despite differences in the maximum temperature value itself. Due to the higher thermal resistance of STI structures and reduced thermal conductivity of upper passivation layers, the heat[22] generated in the device under the ESD stress is confined and mostly dissipated through the substrate. Therefore, changes in the STI boundary associated with variations of GSCS/GDCS influence the overall temperature distribution and the peak temperature in the device as well.

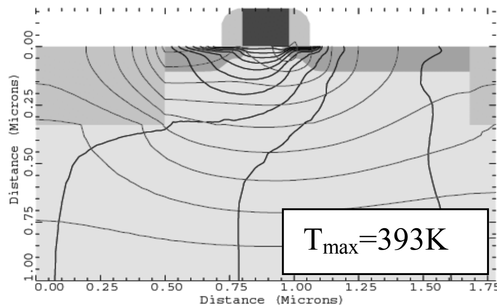
In Fig. 5.14, we show the effect of the gate-to-contact spacings on the temperature distribution. The temperature for the device with longer GSCS (= 0.5 μm) is significantly lower than that of the two other structures with GDCS of 0.1 μm and 0.5 μm. The augmented power dissipating volume for larger GSCS results in a lower peak temperature for a given drain current. Thus for the device with larger GSCS, a higher ESD failure threshold can be obtained due to a reduction in the peak temperature. Therefore, under ESD conditions, the maximum temperature

[22] Joule heating is the main source of heat generation.

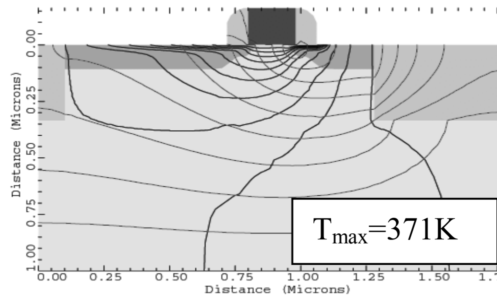
of the device for a given drain current is higher due to reduction in the effective source size resulting from smaller GSCS.



(a)

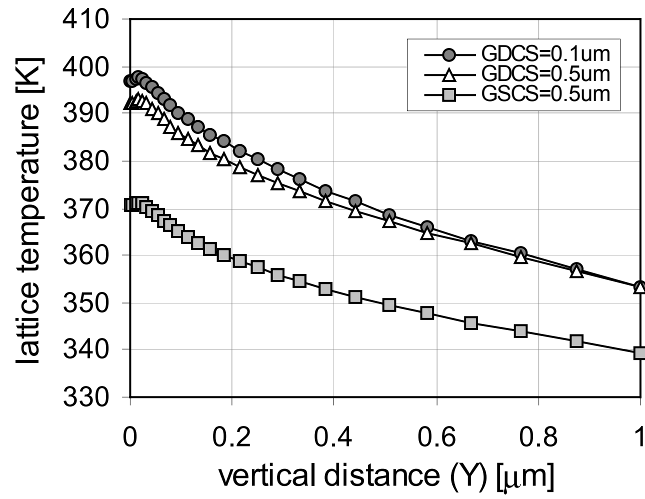


(b)

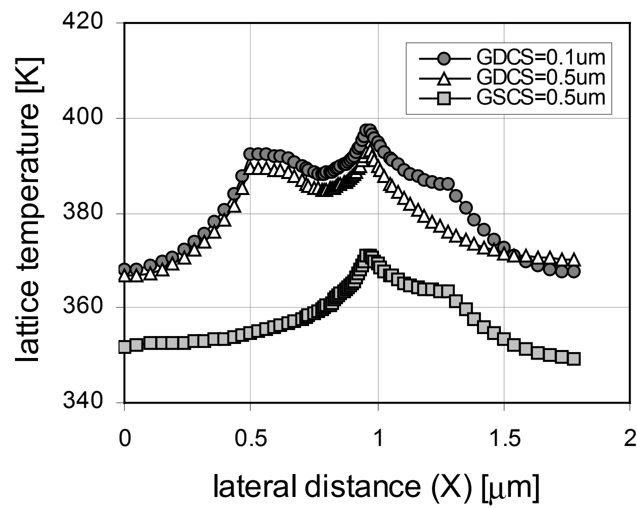


(c)

Figure 5.13: The current flowlines and temperature distribution contours at the drain current of 1 mA/µm for the three different structures are shown. (a) GSCS=0.1 µm and GDCS=0.1 µm, (b) GSCS=0.1 µm and GDCS=0.5 µm, and (c) GSCS=0.5 µm and GDCS=0.1 µm



(a)



(b)

Figure 5.14: The temperature distribution along the *x* and *y* direction [as indicated in Fig. 13 (a)] at the drain current of 1 mA/μm for the three different structures: (a) vertical temperature distribution and (b) lateral temperature distribution

The simulated maximum temperature with the drain current is also plotted in Fig. 5.15. The maximum temperature increases more rapidly with the drain current as the power dissipating volume decreases for the shorter gate-to-source contact spacing. Note that a significant difference in the maximum temperature can be expected at higher drain currents. To provide further support for the thermal effect involved in the gate-to-contact spacing, the total failure threshold currents (IT_2) of the 1.5 V test devices having two different power dissipating volumes (arising due to increased n^+ overlap length of S/D contacts) are shown in Fig. 5.16. It clearly indicates the dependence of IT_2 on the size of the power dissipating volume. For the device with $S_{_OL}/D_{_OL}$ of 0.1 μm, the total failure current IT_2 is less than 25 mA.

However, with increase in the power dissipating volume due to the extension of $S_{_OL}/D_{_OL}$, significant improvement in the failure current can be obtained. It should be noted that this improvement of I_{t2} for low voltage transistors is not attributed to any improved uniformity of current distribution since the gate-to-source/drain contact spacing is unchanged for both the test structures in Fig. 5.16. The experimental result in Fig. 5.16 agrees with the predictions based on electro-thermal simulations [Figs 5.13 to 5.15]; both sets of data suggest that the thermal effect is the root cause of I_{t2} improvement with increase in GSCS. Therefore, it can be concluded that the observed improvement of I_{t2} with GSCS for the low voltage transistors is also primarily due to thermal effects.

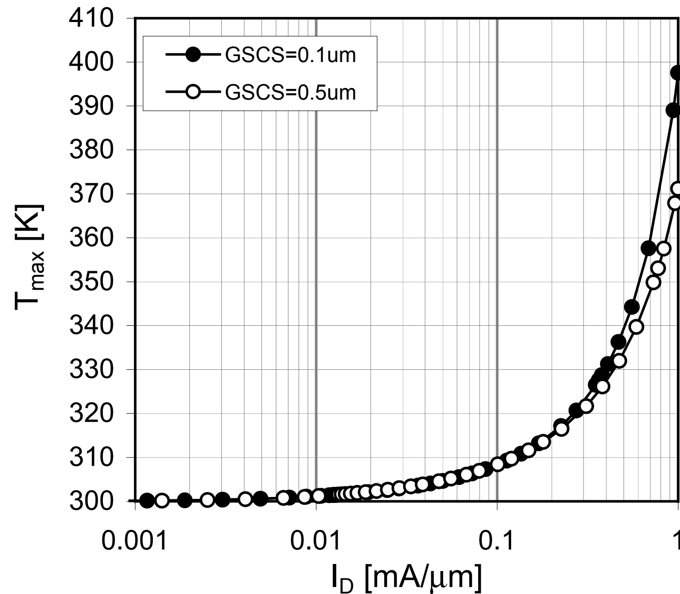


Figure 5.15: The simulated maximum temperature for the two different test structures with injected drain current. The maximum temperature (T_{max}) increases more rapidly as the power dissipating volume decreases for the shorter gate-to-source contact spacing. GDCS = 0.1 μm.

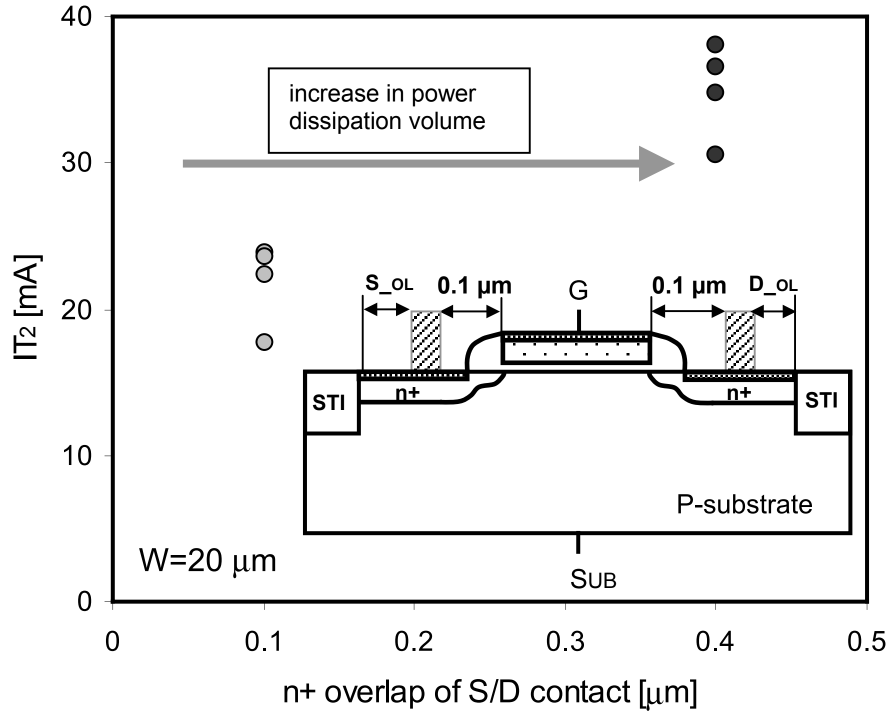


Figure 5.16: Total failure threshold current (IT_2) for the two different test structures having different n^+ overlap lengths of S/D contact of 0.1 µm and 0.4 µm (D_{OL} and S_{OL}). The inset shows the schematic of the test structure. GDCS/GSCS = 0.1 µm.

Also, it should be noted from Fig. 5.5 (b) and (c) that both GSCS and GDCS have little impact on I_{t2} under substrate bias condition, which makes the current distribution more uniform for the low voltage transistors. As shown in [47, 48], the current flowlines, with substrate bias, spread out more uniformly and deeper into the substrate compared with the flowlines without substrate bias. Hence, Fig. 5.5 implies that the volume associated with current (or temperature) distribution is significantly increased by substrate bias and that this volume is substantially larger than the one arising due to increase in GSCS. This is reflected in the dramatic improvement of I_{t2} in Fig. 5.5 (b) and (c) while showing almost no sensitivity to GSCS and GDCS.

Finally, the overall physical mechanisms involved in the gate-to-contact spacing can be summarized. Although the changes in ballast resistance cannot be observed directly, increases in the gate-to-drain contact spacing are effective in mitigating severe non-uniform ESD current conduction. The increase in the GDCS improves I_{t2} for devices with non-uniform ESD current distributions, primarily the low voltage (1.5 V) transistors used in this work. Despite reduction in

current gain of the lateral n-p-n transistor, increases in GSCS lead to higher I_{t2} primarily due to thermal effects arising from increase in the power dissipation volume. This implies that I_{t2} for salicided deep submicron devices with STI is sensitive to the thermal capacity of the structures; and that the lateral n-p-n model is not sufficient for describing the device behavior of these devices under ESD conditions.

Therefore, analysis of ESD behavior of advanced devices should consider both thermal effects and the non-uniform bipolar conduction. Based on this study, it is recommended that the minimum gate-to-contact spacing should be avoided for the design of protection devices. However, the minimum contact spacing can be used if sufficient substrate bias can be supplied to the NMOS device under ESD conditions, because the substrate-triggered lateral n-p-n transistor is independent of the gate-to-contact spacings as confirmed experimentally in this work. Furthermore, on-going experimental work also show that the results obtained from the single finger structure correlate very well with the results from multi-finger structures provided the multi-fingers are uniformly triggered.

5. 4 Summary

The improvement of ESD failure threshold with the gate-to-contact spacing for fully silicided NMOS transistors have been investigated; the results provide new insight into ESD design rules for deep submicron technology, based on detailed experimental and simulation results. It has been shown that the reduction in current localization and increase in the power dissipation volume with increases in the gate-to-contact spacing are the primary factors influencing improvement of ESD performance. It has also been established that substrate biasing can help eliminate the impact of the gate-to-contact spacing on the ESD robustness. Results from this work suggest that even for silicided processes, the gate-to-contact spacing should be carefully engineered to achieve efficient and robust ESD protection designs.

CHAPTER 6

REVERSE GATE LENGTH DEPENDENCE

Contrary to the general understanding, ESD performance of NMOS devices can degrade for shorter channel length transistors in advanced silicided CMOS technologies. In this chapter, using test structures in the 0.13 μm CMOS process, detailed characterization has been carried out for the first time to comprehend and model the physical mechanism causing this effect.

It is shown that the reverse channel length dependence of ESD performance is mainly due to severe non-uniform lateral bipolar conduction, which reduces the effective device width. Furthermore, it is demonstrated that substrate bias can be effective in controlling this reverse channel length effect. The findings from this investigation will have significant implications for designing ESD protection circuits in deep submicron silicided CMOS technologies with minimum gate lengths.

6. 1 Introduction

In advanced CMOS technologies, the gate-grounded NMOS transistor is widely used as an ESD protection device due to its effectiveness during ESD events. Based on the traditional bipolar transistor model under ESD conditions [9, 68], it is generally believed that the shorter channel length (L_{poly}) devices with higher current gain (β)[23] will show better ESD performance since the power dissipation of such devices is smaller for a given ESD stress.

However, recent experimental observations indicate that it is not always the case, and that the dependence of ESD performance on the channel length of the device can become reversed. In this regard, K. Bock et al. [70] qualitatively proposed that the trade-off between the power dissipation and melt volume of the entire parasitic bipolar transistor determine the ESD performance, and not just the parasitic bipolar triggering model.

However, along with the reverse channel length dependence, the new experimental results in our work also show a strong non-uniformity in the lateral bipolar current conduction, which was

[23] $\beta \propto \coth(L_{poly})$ [68, 69].

not observed in the previous study [70]. This additional experimental observation has thrown new light on the problem. Therefore, to gain a better understanding of the physical mechanism responsible for this unusual ESD behavior involved in the advanced silicided technologies, detailed characterization has been carried out using test structures in the 0.13 μm CMOS process.

6. 2 ESD Performance with Gate Length

As a monitor of ESD hardness, the transmission line pulsing (TLP) tests have been performed for the various test structures with silcided ($CoSi_2$) and non-silicided diffusions. Fig. 6.1 shows the second breakdown triggering current (I_{t2}) with L_{poly} for the silicided devices, and the strong degradation of I_{t2} can be observed for L_{poly} < 1μm. This is contradictory to the expectation from the parasitic bipolar transistor model. However, as shown in Fig. 6.2, the same measurements for the non-silicided devices show that I_{t2} values increase with decrease in L_{poly}. This implies that the channel length dependence of I_{t2} is not always consistent for different technologies, and furthermore the application of silicide diffusion strongly influences the ESD behavior of the advanced NMOS transistors. The data in Fig. 6.2 also demonstrates that the model proposed in [70] to account for the short channel length effect is not appropriate.

Also, in advanced CMOS technology, an effective channel length (L_{eff}) of MOS transistors can fluctuate due to statistical process variation, such as the line edge roughness effect[24] [71]. Therefore, the I_{t2} dependence on the effective channel length (not the drawn gate length, L_{poly}) can also be observed through the data of I_{t2} versus drive current (I_{drive}) in Fig. 6.3, since the drive current of a standard NMOS transistor is inversely proportional to the effective channel.

In addition, as shown in Fig. 6.4, despite the decrease in I_{t2} with lower L_{poly}, the β of the silicided devices increases with decreased L_{poly} as expected, since the ratio of triggering voltage (V_{t1}) to holding voltage (V_h) (i.e. V_{t1}/V_h) is proportional to β [69]. Therefore, the results imply that the degradation mechanism of I_{t2} with lower L_{poly} cannot be attributed to the operation of the parasitic lateral n-p-n transistors. It further implies that the bipolar triggering model for ESD performance analysis is not very reliable in advanced CMOS devices.

[24] Gate line edge roughness effects come from lithographic error during resist patterning and combined etch process errors.

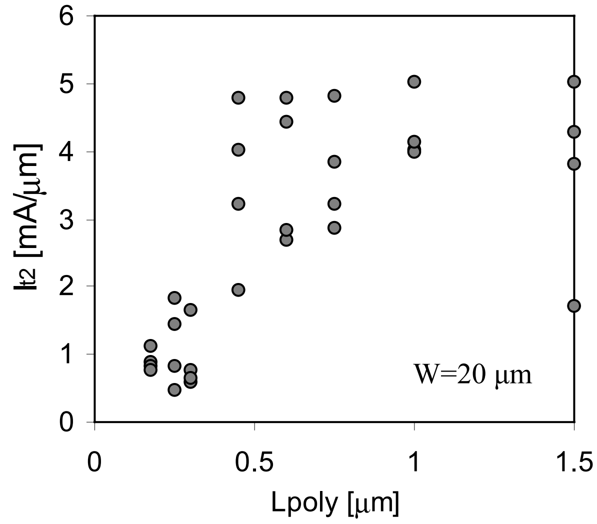


Figure 6.1: I_{t2} dependence on the channel length (L_{poly}) for the silicided 1.5 V NMOS devices. I_{t2} drops rapidly as the channel lengths enter the submicron regime ($L_{poly} < 1$ µm).

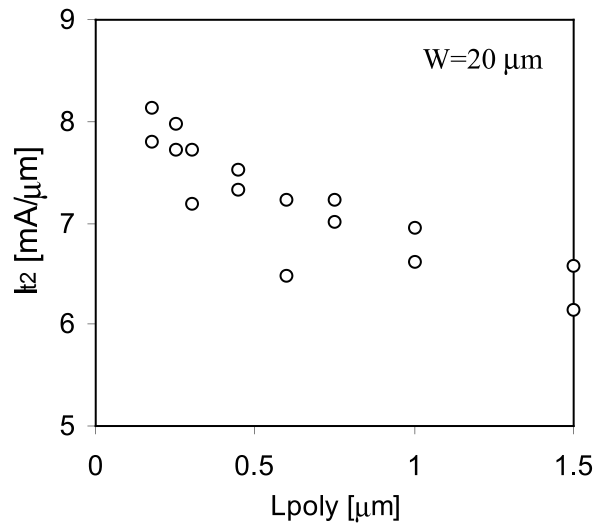


Figure 6.2: I_{t2} dependence on the channel length for the non-silicided 1.5 V NMOS devices. I_{t2} dependence on the channel length agrees with the conventional bipolar transistor triggering model [9].

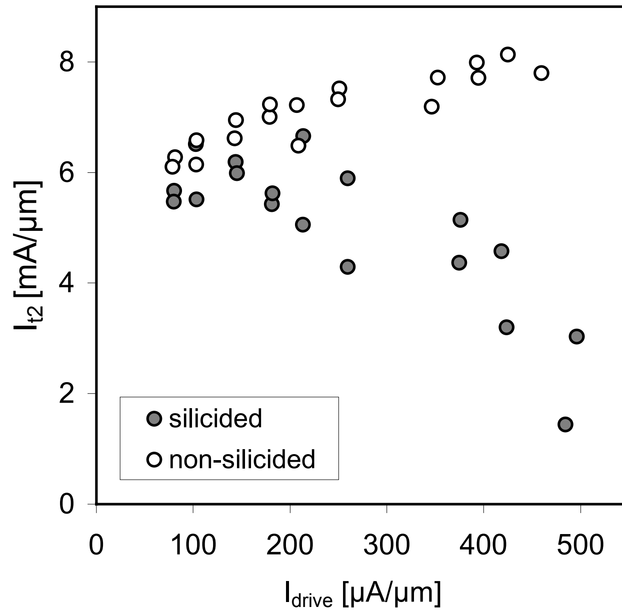


Figure 6.3: I_{t2} versus the drive current (I_{drive}) for the 1.5 V NMOS transistors where W= 20 µm. The reverse channel length dependence of I_{t2} on the effective channel length for the silcided devices is clearly shown.

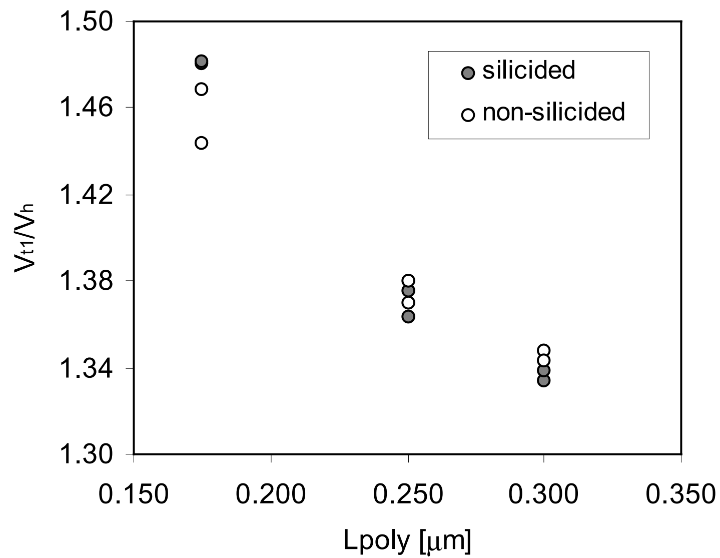


Figure 6.4: The V_{t1}/V_h with L_{poly} for the 1.5 V NMOS devices. It shows that the bipolar current gain (β) of both silicided and non-silicided devices increases with decrease in L_{poly}.

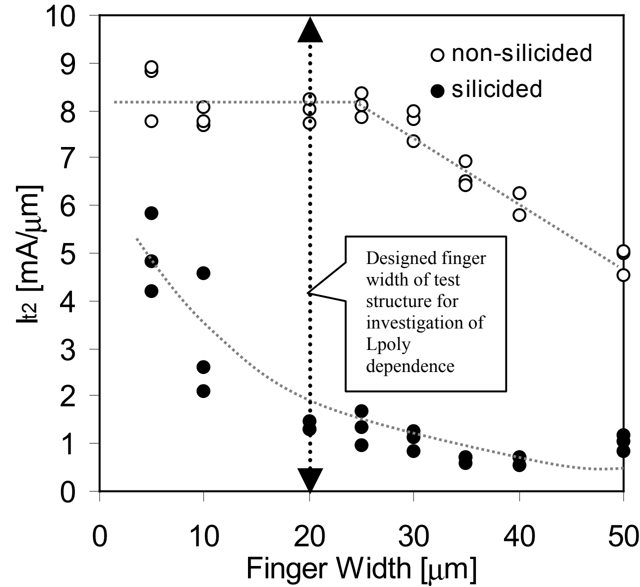


Figure 6.5: Finger width dependence of I_{t2} for the 1.5 V silicided and non-silicided devices where L_{poly}=0.175 µm.

6. 3 Analysis and Discussion

In order to identify the root cause of the reverse channel length dependence, the I_{t2} has been characterized for the devices with various finger widths (W) as shown in Fig. 6.5. Without silicided diffusion, the test devices show the uniform ESD current distribution up to a finger width of about 25 µm. On the other hand, the silicided devices show a strong non-uniformity in the ESD current, and this implies that the ESD behavior is strongly influenced by the 3-D current conduction phenomenon. Although this behavior was not reported in the earlier work [70], the non-uniform bipolar conduction effect has been predominantly observed for advanced silicided devices [22, 28, 46, 47, 48].

Therefore, it is important to investigate whether this non-uniform current conduction has any possible impact on I_{t2} degradation with decreasing L_{poly}. It can be observed from Figs 6.1 and 6.5 that for the silicided devices, I_{t2} of the device with the narrow finger (W=5 µm) is nearly the same as the It_2 values of the devices for wider W (=20 µm) with L_{poly} > 1 µm. This also suggests that the I_{t2} degradation with L_{poly} can be shown to be different according to the reference finger width.

However, it has not been understood as to how the extent of the non-uniformity depends on the L_{poly}. In order to investigate the dependency of the strength of this non-uniform conduction effect on the L_{poly}, I_{t2} was tested for the devices with varying W and L_{poly}.

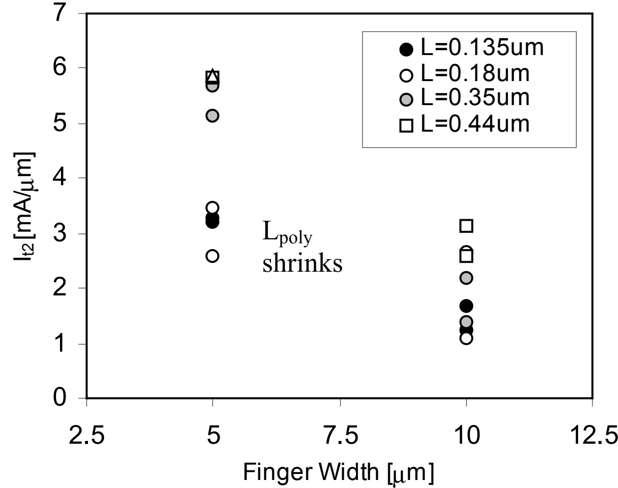


Figure 6.6: I_{t2} with finger width for the silicided 1.5 V devices with different channel lengths L_{poly}. For the smaller channel length devices, uniform conduction is not reached even at W=5 μm. This implies that the non-uniform conduction becomes more serious for shorter channel devices.

As shown in Fig. 6.6, for the smaller channel length devices, uniform conduction is not reached even at W=5 μm. This implies that the non-uniform conduction seems to be stronger as the gate length L_{poly} shrinks. Furthermore, since it is known that the substrate bias improves I_{t2} by enlarging the turned-on finger width under ESD conditions [47]. The impact of substrate bias on the reverse channel length effect was next explored.

Applying an external substrate bias, I_{t2} dependence on the L_{poly} for the 1.2 V devices[25] with various process splits is shown in Fig. 6.7. As can be seen, with substrate bias, the I_{t2} degradation with L_{poly} disappears independent of the process variations considered. The above result also suggests that the dominant cause of the reverse channel length effect is non-uniformity in ESD current distribution, which appears strongly in the devices with silicided processes.

In addition, we have also carried out the detailed investigations into thermal effects that might be involved in this reverse channel length phenomenon. Electrothermal simulations (*MEDICI*) shown in Figs 6.8 and 6.9 indicate that the temperature distribution at thermal failure changes significantly with the L_{poly}.

[25] The 1.2 V NMOS transistors are additional test structures for investigating ESD behavior along with the 1.5V and 3.3 V devices.

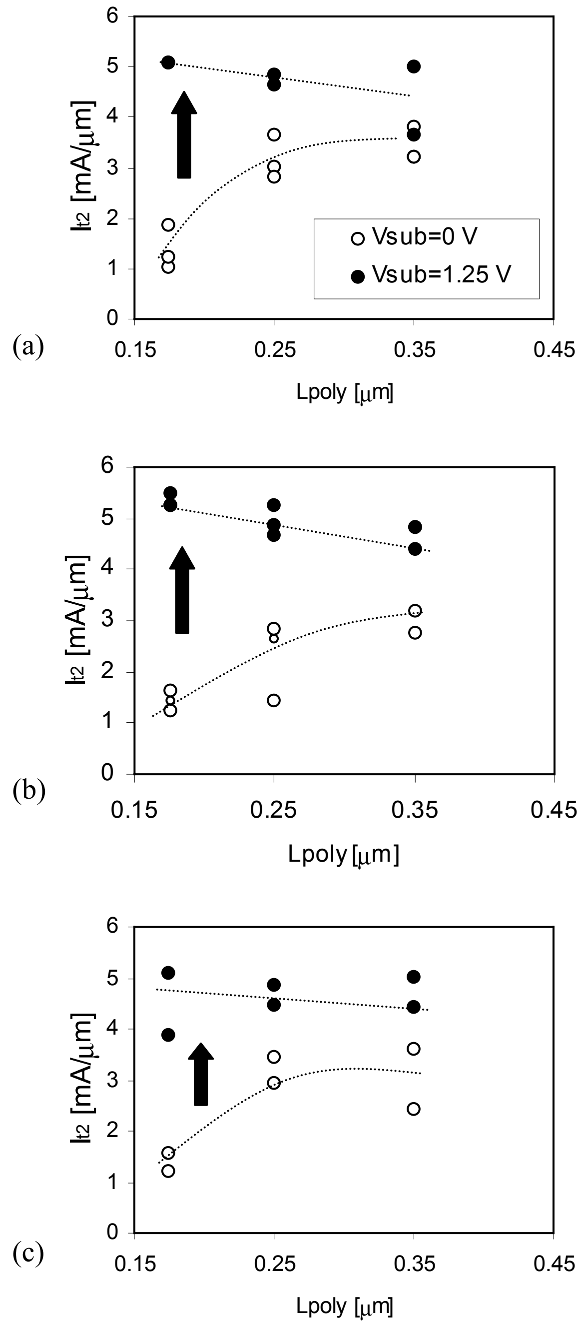


Figure 6.7: I_{t2} with L_{poly} for the fully silicided 1.2V devices (W=20 μm) with process splits. (a) baseline (b) *As* supplement source/drain implant, and (c) *P* supplement source/drain implant.

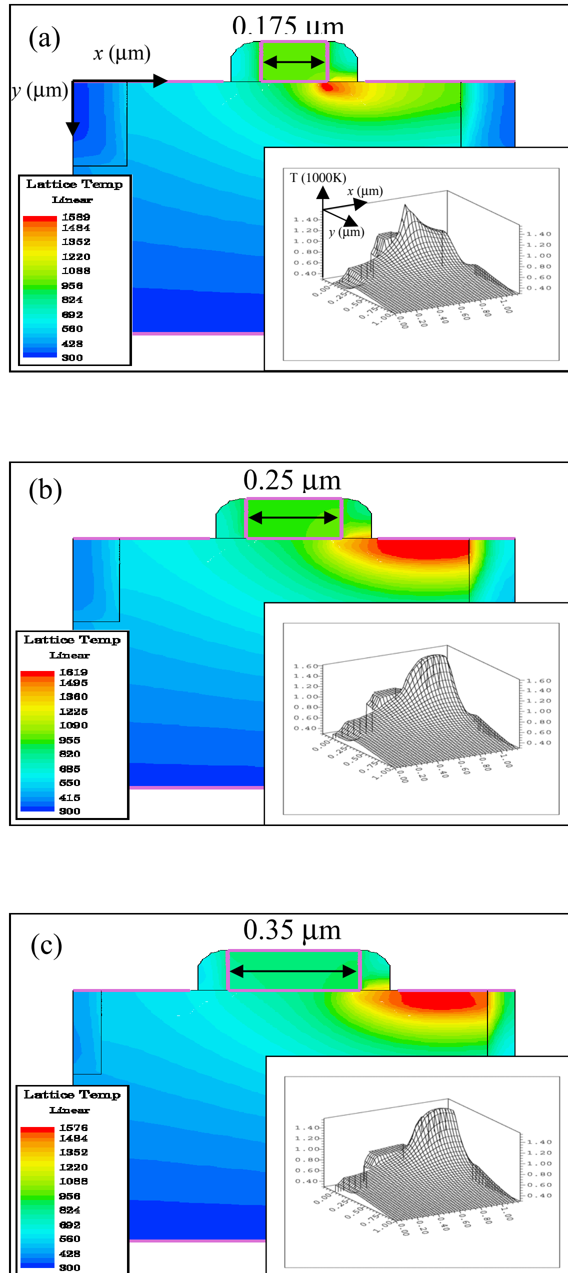


Figure 6.8: The temperature contours 3-D distribution (insets) just before thermal failure (T=1680 K) for the devices with (a) L_{poly}= 0.175 µm, (b) L_{poly}= 0. 25 µm, and (c) L_{poly}=0.35 µm

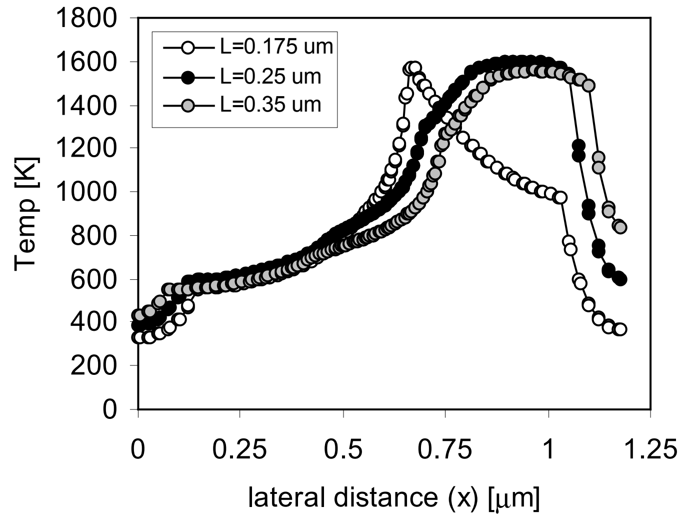


Figure 6.9: The temperature distribution along the channel (*x*-direction) for devices with different channel length at thermal failure (at y=0.05 µm).

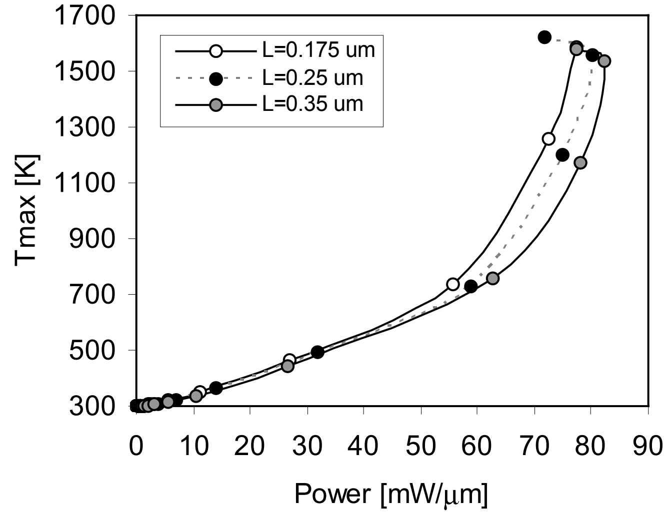


Figure 6.10: The maximum temperature (T_{max}) with the power dissipation for the devices with the different gate lengths. The T_{max} increases with decrease in channel length for a given power.

As the L_{poly} shortens, the temperature distribution seems to be more localized in accordance with the model in [70]. Furthermore, for a given power dissipation, the maximum temperature (T_{max}) increases with decrease in L_{poly} [see Fig. 6.10] since the power density in the shorter device becomes higher. However, as shown in Fig. 6.11, the measured failure power[26] (P_f) for the non-silicided devices still increases despite the decrease in L_{poly} or the power dissipating volume, which contradicts the observation from Figs 6.8 and 6.10.

In fact, Fig. 6.11 implies that the thermal effects resulting from the decreased L_{poly} are negligible for the non-silicided devices, but apparent for the silicided devices since the $CoSi_2$ layer is known to be more vulnerable to thermal damage (than non-silicided device) under high current stress [72].

However, the trend in P_f shown for the two technologies in Fig. 6.11 cannot both be explained at the same time if thermal effect with L_{poly} reduction was the main cause of the reverse channel length effect since the failure power P_f increases for the shorter channel non-silicided devices.

To further justify the role of the non-uniformity mechanism in the reverse channel length effect, consider a first-order model of the rectangular box heat source with dimensions a, b, and c (a> b> c) at the second breakdown [Fig. 6.12 (a)]. The input power as a function of time is shown in Fig. 6. 12 (b), which shows that the failure power decreases with increase in the stress time, depending on the thermal diffusion time, t_a, t_b, and t_c, related to the heat source.

In general, the time scale of ESD (i.e., a few nanoseconds to a few hundred nanoseconds) event ranges from t_b to t_a. Therefore, the P_f, before the drain-substrate junction reaches thermal equilibrium, is given by [16],

$$P_f = \frac{4\pi ka\,\Delta T}{\log(t/t_b) + 2 - c/b} \tag{6.1}$$

where ΔT is temperature rise, k is the thermal conductivity, and t_b ($=b^2/4\pi D$) is the thermal diffusion time, particularly determined by b.

Using (1) and the measured P_f data for silicided devices, the W_{eff} (*a* in Eq. (6.1)) along with L_{poly} were calculated as shown in Fig. 6.13. It can be observed that the effective finger width W_{eff} decreases with decreasing L_{poly}. These results confirm that the reduction in effective finger width arising from non-uniform ESD current distribution is the primary cause of reverse channel length dependence of ESD performance in silicided devices.

---

[26] $P_f = I_{t2} \cdot V_{t2}$

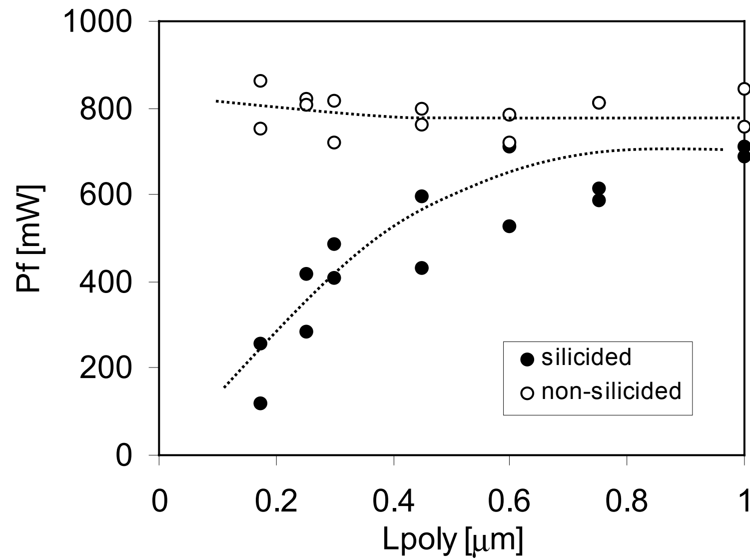


Figure 6.11: The failure power P_f with L_{poly} for 1.5 V devices.

6. 4 Summary

In conclusion, through device simulations and experimental data, this work has clearly shown that for deep submicron devices the unusual reverse channel degradation of It_2 is due to reduced effective device width associated with severe non-uniform bipolar conduction. For ESD applications, on the other hand, the protection design with minimum L_{poly} is critical for achieving efficient input gate oxide and internal core circuit reliability. This work has further shown that with substrate bias the reverse channel effect can be overcome to make this possible. The concept has been successfully tested for a multi-finger NMOS device with minimum channel length using substrate bias.

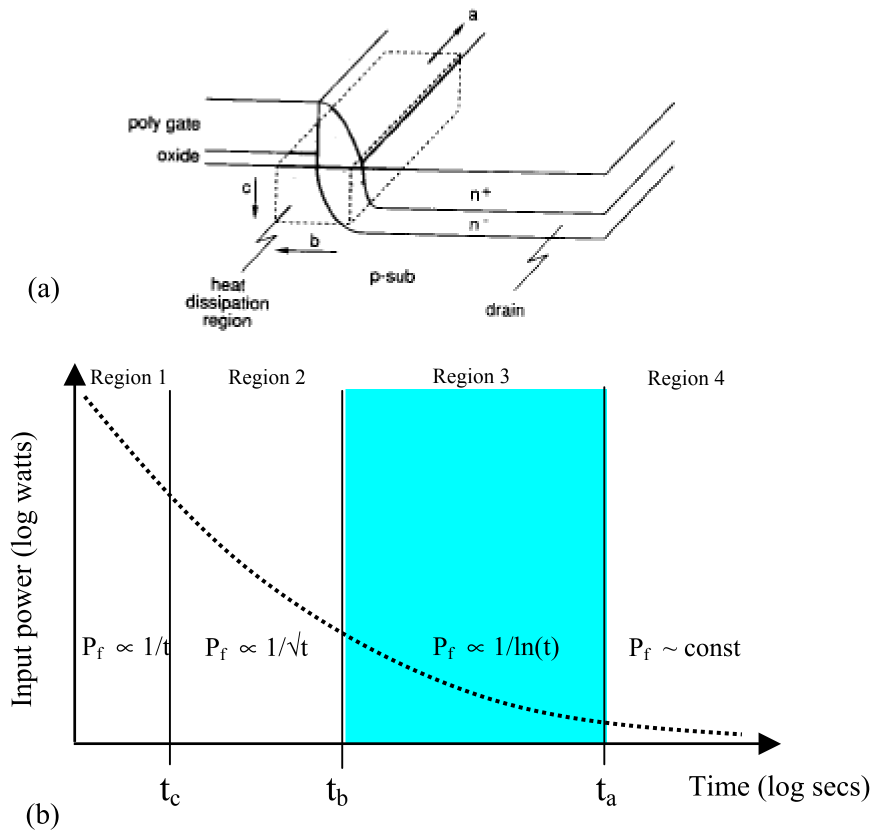


Figure 6.12: (a) approximated heat source region for a reverse biased NMOS transistor and (b) input power as a function of stress time, for a constant ΔT, across a reverse biased junction from [16, 73]

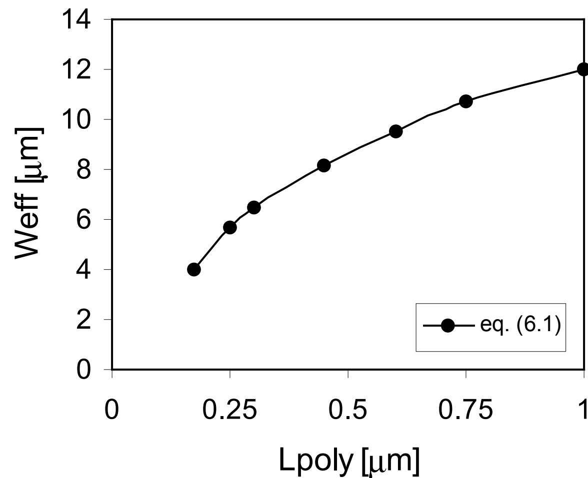


Figure 6.13: The predicted effective finger width (W_{eff}) with L_{poly} at second breakdown for 1.5 V devices.

CHAPTER 7

MODELING OF CONTACT RESISTANCE

7. 1 Introduction

As CMOS technology advances, the junction depth of the source/drain and size of other associated device features are aggressively decreased. Therefore, the impact of source/drain parasitic resistances on device performance becomes more significant since the parasitic resistance component is not scaled with device scaling and contributes an appreciable fraction of channel resistance resulting in significant current degradation [74-78]. Hence, in advanced technologies, accurate understanding of the physics involving parasitic resistance is required for optimization of high performance devices used in various applications.

In particular, the contact resistance component of parasitic resistances is not scalable since it simply increases with decrease in contact size commensurate with a device scaling. In this sense, it is important to identify physical parameters determining the contact resistance value for improved understanding its implications and for optimizing device performances. Furthermore, the current flow or current distribution associated with the contact system is strongly influenced by the contact resistance value itself [79, 80]. As a result, the impact of the contact resistance on the current distribution also has important implications for ESD reliability through localized heating by current crowding effects.

Nevertheless, the temperature dependence of contact resistance has not been fully investigated. Thermal effects involved in the contact resistance become more important in the deep submicron device regime, especially under high current stress conditions such as ESD. Therefore, in this chapter we theoretically propose a temperature dependent specific contact resistance model for generating high temperature contact resistance values and for predicting high current behavior of silicided deep submicron devices, which is useful for analysis of ESD reliability.

7. 2 Temperature Dependent Specific Contact Resistance

For a metal semiconductor junction, the specific contact resistance ρ_c is defined as

$$\rho_c = (\frac{dV}{dJ})_{V=0} \tag{7.1}$$

where J is the current density of a metal semiconductor contact and V is the applied voltage. Using the theoretical relation of current-voltage characteristics for a metal semiconductor junction, ρ_c can be expressed. In general, for a metal semiconductor contact, the carrier transport mechanisms can be separated into three regions: field emission (FE) [69, 81, 82], thermionic-field emission (TFE) [82, 83], and thermionic emission (TE) [69, 81, 82], depending on the doping concentration of the semiconductor.

In advanced CMOS devices, the source and drain are formed with a high or moderate doping concentration (i.e., typically, $N \geq \sim 10^{18}$ cm^{-3}). Therefore, the carrier transport mechanisms in the source/drain region are limited to the thermionic field emission (TFE) or the field emission (FE). For a metal semiconductor contact system with a higher doping concentration, ρ_c can be written by [82, 84]

$$\rho_c = [\frac{A^* T \pi q}{k_B \sin(\pi c k_B T)} \exp(\frac{-q\phi_b}{E_{00}}) - \frac{A^* q}{c k_B{}^2} \exp(\frac{-q\phi_b}{E_{00}} - cE_F)]^{-1} \tag{7.2}$$

where

$$c = \frac{1}{2E_{00}} \ln(\frac{4q\phi_b}{E_F}) \tag{7.3}$$

$$E_{00} = \frac{qh}{4\pi} \sqrt{\frac{N}{\varepsilon_s m^*_{tun}}} \tag{7.4}$$

A^* ($=4\pi q k^2 m^*/h^3$) is the effective *Richardson* constant, T is the temperature, ϕ_b is the barrier height, E_F is the Fermi energy with respect to the energy band edge in the semiconductor bulk, and the quantity E_{00} is a characteristic energy related to the tunneling probability, depending on the doping concentration N, the tunneling effective mass m_{tun}^*, which is the effective mass component of in the direction of current flow [85].

As E_{00} increases with N, the barrier for the tunneling process becomes thinner, and in turn results in increased tunneling carriers, since the carrier transmission rate *P*, is exponentially increasing with E_{00} as given by [69]

$$P \sim \exp(-q\phi_b / E_{00}) \tag{7.5}$$

Therefore, the ratio of k_BT/E_{00} can be used as a measure indicating the relative significance of the thermionic emission process with respect to the field emission (tunneling) process [82, 83]: For a lightly doped semiconductor where $k_BT/E_{00} >> 1$, the TE dominates, and for a heavily doped semiconductor where $k_BT/E_{00} <<1$, the FE dominates. Otherwise, for an intermediately doped semiconductor where $k_BT/E_{00} \approx 1$, the TFE dominates.

The above criterion also implies that the carrier transport mechanism can be determined by the temperature rise for a given doping concentration. For a given high doping concentration, Eq. (7.2) for where the FE dominates is valid if [82]

$$T < \frac{1}{k_B[c + (2E_{00}E_F)^{-1/2}]} \tag{7.6}$$

The above condition also gives the lower limit of doping concentration for the field emission process for a given temperature. At room temperature, the lower limit of doping concentration is approximately $N \approx 6.6\times10^{19}\text{cm}^{-3}$ with $q\phi_b = 0.6\text{eV}$ and $m_{tun}^* = 0.19m_o$ [84].

For a highly doped region where the field emission dominates, the approximate carrier distribution by *Boltzmann* statistics is erroneous in determining the location of E_F. The position of Fermi energy E_F with respect to energy band edge should be estimated by more accurate distribution model such as *Joyce-Dixon* approximation, which is given by [86],

$$E_F = k_BT[\ln(\frac{N}{N_c}) + \frac{1}{\sqrt{8}}\frac{N}{N_c} - (\frac{3}{16} - \frac{\sqrt{3}}{9})(\frac{N}{N_c})^2] \tag{7.7}$$

where N_c is the effective density of states in the conduction band.

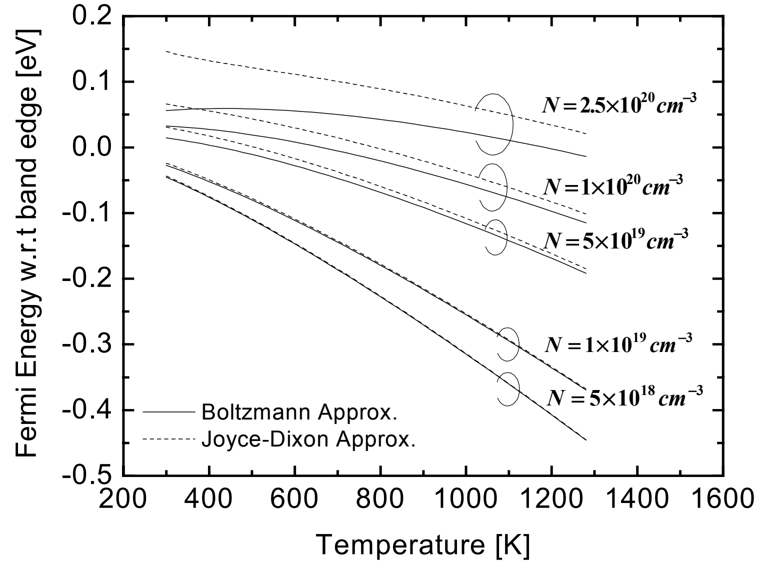


Figure 7.1: The comparison of *Boltzmann* approximation and *Joyce-Dixon* approximation [86] for determining the position of Fermi energy E_F for various doping concentrations as a function of temperature. The difference of Fermi energy between the two models increases as the doping concentration increases.

For silicon, the density of states in the conduction and valence bands, N_c and N_v, respectively, can be expressed as a function of temperature by [87]

$$N_c = 2.86\times10^{19}(T/300)^{1.58}\,cm^{-3} \tag{7.8}$$

$$N_v = 3.10\times10^{19}(T/300)^{1.85}\,cm^{-3} \tag{7.9}$$

The intrinsic concentration n_i is also given by

$$n_i = \sqrt{N_c N_v}\exp(-E_g(T)/2k_B T) \tag{7.10}$$

where E_g (T) is the temperature dependent energy bandgap.

In Fig. 7.1, using the two different approximations, the theoretically estimated Fermi energy is shown with the temperature for the various doping concentration. As the doping concentration increases (i.e., $N > N_c$), error arising from using the *Boltzmann* distribution becomes more significant.

For advanced CMOS technologies, the sheet resistivity of the source/drain junction typically ranges from ~50 to ~500 Ω/□ [7], which corresponds to an effective doping concentration[27] in the source/drain junction of $5.1\times10^{20} \sim 3.8\times10^{19}$ cm^{-3}. Therefore, for analysis of contact resistance in advanced technologies, the position of E_F based on *Joyce-Dixon* approximation should be used for accuracy.

For an intermediate doping regime, the thermionic-field emission (TFE) is the dominant mechanism for carrier transport. In this condition, the specific contact resistance, ρ_c is also given by [82]

$$\rho_c = [\frac{k_B{}^2}{qA^*}] \frac{\cosh(\frac{E_{00}}{k_B T})}{\sqrt{\pi(q\phi_b + E_F)E_{00}}} \sqrt{\coth(\frac{E_{00}}{k_B T})} \exp(\frac{E_F}{E_0} - \frac{E_F}{k_B T} + \frac{q\phi_b}{E_0}) \tag{7.11}$$

where

$$E_0 = E_{00} \coth(\frac{E_{00}}{k_B T}) \tag{7.12}$$

The above Eq. (7.11) is valid if

$$\frac{\cosh^2(\frac{E_{00}}{k_B T})}{\sinh^3(\frac{E_{00}}{k_B T})} < \frac{2(q\phi_b + E_F)}{3E_{00}} \tag{7.13}$$

and which also gives a lower bound on the doping concentration for needed thermionic-field emission. Typically, the transition from the thermionic-field emission (TFE) to the thermionic emission (TE) occurs at the doping concentration of $10^{17} \sim 10^{18}$ cm^{-3} at room temperature. Therefore, for advanced CMOS technology, the carrier transport in metal-silicon junction in source/drain structure is governed by a combination of FE and TFE. In order to predict the dependence of ρ_c on the temperature, primary physical parameters should also be expressed as a function of the temperature. In general, ρ_c is exponentially dependent on the barrier height, doping concentration, and tunneling effective mass (m_{tun}^*).

[27] The effective doping concentration is calculated using the relation $N=(\rho x_j q\mu)^{-1}$ where x_j=0.07 μm is assumed for advanced devices.

For a metal-semiconductor junction, the barrier height is strongly dependent on the density of surface states and properties of interfacial layer, which is strongly process dependent; The Fermi energy level E_F is known to be pinned roughly at about one third of band gap, E_g [69, 88, 89]. Assuming that the temperature dependence of density of surface states is negligible, we can express the temperature dependence of the barrier height to be approximately equal to that of energy band gap E_g [88]:

$$\frac{\partial q\phi_b}{\partial T} \cong \frac{\partial E_g}{\partial T} \tag{7.14}$$

Using the empirical model of temperature dependent energy band gap $E_g(T)$, which given by [90, 91],

$$E_g(T) = 1.17 - \frac{7.02\times10^{-4}T^2}{T+1108} \tag{7.15}$$

the temperature dependence of the barrier height can be simply expressed by

$$q\phi_b(T) \cong q\phi_b(T_0) + \frac{\partial E_g}{\partial T}(T - T_0) \tag{7.16}$$

where T_0 is the ambient temperature of interest.

As mentioned previously, the tunneling effective mass m_{tun}* is also a critical parameter in the field emission process [92]. Since a large value of tunneling effective mass gives rise to a small transmission coefficient *P* [in Eq. (7.5)], the specific contact resistance is exponential with the tunneling effective mass as shown in the equation given by

$$\rho_c \propto \exp(\frac{q\phi_b\sqrt{m_{tun}^*}}{\sqrt{N}}) \tag{7.17}$$

For any particular ellipsoid on a constant energy surface in k-space, the m_{tun}* is given by [85]

$$m_{tun}^* = (\frac{l^2}{m_x} + \frac{m^2}{m_y} + \frac{n^2}{m_z})^{-1} \tag{7.18}$$

where *l*, *n*, and *m* are the direction cosines of the ellipsoid relative to the principle axis, and m_x, m_y, and m_z are the tensor masses. Silicon has a total of six ellipsoids in the conduction band. For N-type silicon on <100> surface, there are two tunneling effective masses associated with the

transverse (m_t*) for four ellipsoids and the longitudinal effective mass (m_l*) for two ellipsoids in the ellipsoidal constant energy surfaces. Based on the low temperature cyclotron resonance studies [87], the m_{tun}* is given by 0.916m_0 (from m_l*) and 0.19m_0 (from m_t*) where m_0 is the electron mass in free space [92]. Since the tunneling process is exponential with m_{tun}*, the m_{tun}* from transverse m_t* will dominate.

The temperature dependence of m_t* is experimentally known to be relatively weak and that for m_l* is theoretically considered to be even weaker. In this work, as used in [87], m_l* is assumed to be independent of temperature and m_t* is weakly temperature dependent as much as on temperature dependent energy scaling factor, which is given by

$$m_t^*(T) = 0.19 m_0 \frac{E_{g0}}{E_g(T)} \tag{7.19}$$

where E_{g0} is the value of energy band gap E_g(T) at T=0 K. The effective mass m* in *Richardson* constant A* is different from m_{tun}* and is given by [92, 93]

$$m^* = (l^2 m_y m_z + m^2 m_x m_z + n^2 m_x m_y)^{1/2} \tag{7.20}$$

For computing the *Richardson* constant, the number of equivalent ellipsoids corresponding to the m_{tun}*, *M* should be included (*M* = 4 from m_t* and $A^* = M 4\pi m^* q k_B^2/h^3$).
Using the m_{tun}* values, m* for <100> Silicon is given by [92]

$$m^* = \sqrt{m_l^* m_t^*} \tag{7.21}$$

If the temperature dependence of m_t* is neglected, m* is calculated to be 0.417m_0. K. Varahramyan et al. [84] proposed unified model of specific contact resistance for the TFE and FE based on the similarity of the functional forms of ρ_c in the two different emission mechanisms, and which is given by

$$\rho_c = \frac{k_B}{qA^*T} \underline{C} \exp(\frac{q\phi_b}{E_0}) \tag{7.22}$$

where $\underline{C}$ = 0.425 for n-type silicon and $\underline{C}$ = 0.335 for p-type silicon, for a wide range of doping concentration (10^{16} cm^{-3} $\leq$ N $\leq$ 10^{21} cm^{-3}).

Although the proposed model is valid at room temperature, it starts to deviate from the FE or the TFE models as temperature increases. Since several primary physical parameters in the specific contact resistance model are strong functions of temperature, the average value of $\underline{C}$ is no longer constant with variations of temperature.

Fig. 7.2 clearly shows that the dominant carrier transport mechanism changes depending on the temperature for a given doping condition. Furthermore, it also implies that for analysis of device behavior at high temperatures the thermionic field emission process should be accounted even for highly doped regions.

In order to demonstrate the proposed temperature dependent contact resistance model, $CoSi_2$ silicide process is considered since $CoSi_2$ is currently very popular as a replacement of $TiSi_2$ for advanced CMOS technologies. Also it does not suffer significantly from linewidth effects[28]. At room temperature, the barrier height ($q\phi_b$) of $CoSi_2$/N-type Silicon is reported to be 0.64 eV [69]. For doping concentrations ranging from 5×10^{19} to 5×10^{20} cm^{-3}, which are practical values of interest for advance technologies, theoretical ρ_c curves are created as shown in Fig. 7.3. As can be seen, the specific contact resistance dramatically decreases with increase in temperature. For instance, the ρ_c of 5×10^{19} cm^{-3} at about 1000 K is almost the same as that of 2.5×10^{20} cm^{-3} at about 300 K. For the temperature range from 300 K to 1000 K, the specific contact resistance is at least a factor of 10 reduced. In addition, it is shown that the carrier transport mechanism changes from the TE to the TFE with increase in temperature as indicated by the thick arrows.

Since the theoretical $\rho_c(T)$ shows a considerable dependence on temperature, it is also important to understand its impact on ESD performance. In the next section, implication of high temperature behavior on ESD reliability is discussed in terms of temperature dependent current localization effects.

[28] For $TiSi_2$, thin Ti lines have higher resistance than other silicides.

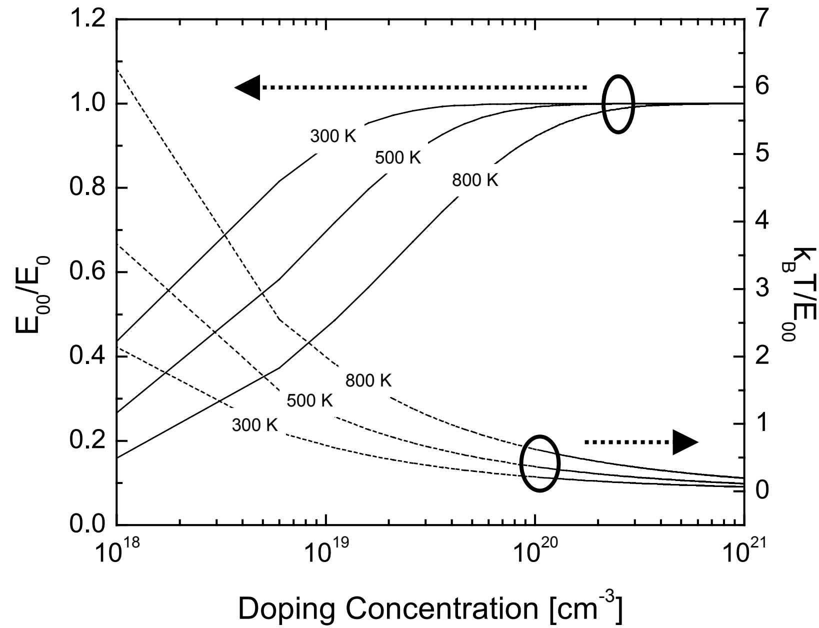


Figure 7.2: E_{00}/E_0 and k_BT/E_{00} versus doping concentration. It shows that the relative importance of field emission (FT) with respect to thermionic field emission (TFE) decreases as temperature increases for a given doping concentration.

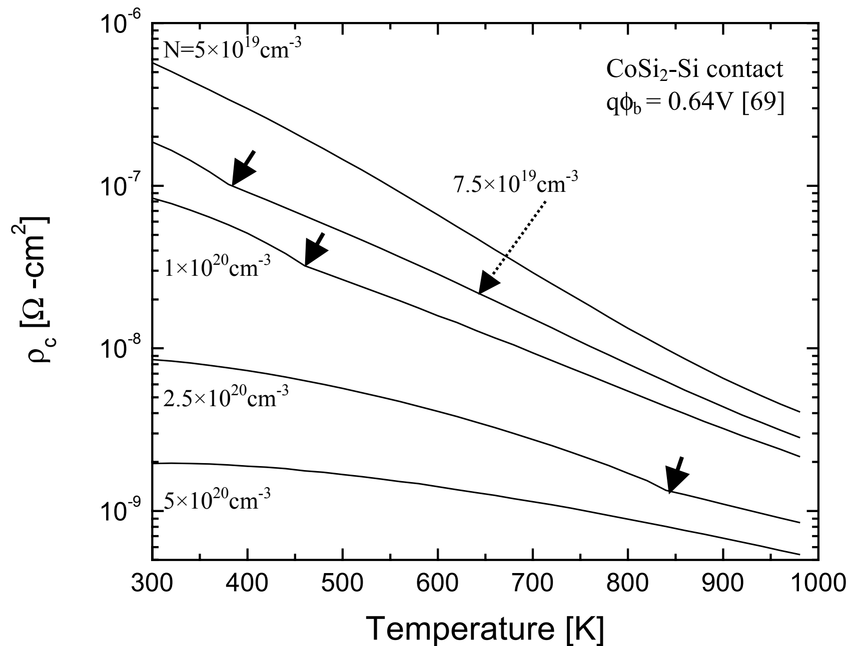


Figure 7.3: Theoretical curves for the specific contact resistance ρ_c with temperature for $CoSi_2$-contact system varying background doping concentration. The thick arrows indicate that the carrier transport mechanism changes from the TE to the TFE as temperature increases.

7. 3 Implications on ESD Reliability

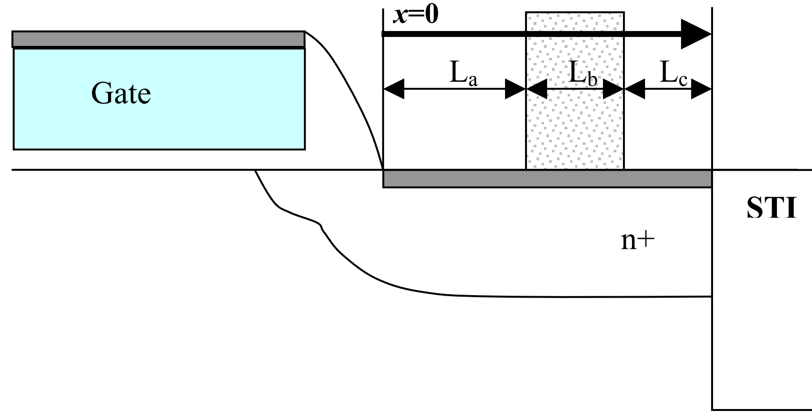


Figure 7.4: Schematic of the contact of a MOS transistor for analyzing impact of the temperature dependent ρ_c on current localization effects where L_a is the distance from the spacer edge to contact hole, L_b is the width of the contact hole and L_c is the distance from the edge of the contact hole to the STI boundary.

In order to investigate silicide effects on ESD reliability, the derived temperature dependent specific contact resistance model is applied to the structure as shown in Fig. 7.4. The basic idea of this work is to examine how severe the current localization effect is with temperature rise arising from ESD stress.

For advanced devices, the parameter L_a, L_b and L_c are generally comparable in the submicron regime. Therefore, using a 1-D transmission line model, the total contact resistance R_T including L_a, L_b and L_c can be written by [80]

$$R_T = \frac{\rho_s \rho_d}{\rho_s + \rho_d} \frac{L_a}{W} + R_C \tag{7.23}$$

where

$$R_C = \frac{\dfrac{\rho_d^{\,2}}{\rho_s \beta W} \sinh \beta L_a + \dfrac{R'_C \rho_d}{\rho_d + \rho_s} \left(2 + \dfrac{\rho_s^{\,2} + \rho_d^{\,2}}{\rho_s \rho_d} \cosh \beta L_a \right)}{\dfrac{\rho_s + \rho_d}{\rho_s} \cosh \beta L_a + \dfrac{R'_C}{\rho_s} \beta W \sinh \beta L_a} \tag{7.24}$$

$$R'_C = \frac{\rho_d}{\alpha W} \frac{\cosh \alpha L_b + \dfrac{\rho_d}{\rho_s + \rho_d} \dfrac{\beta}{\alpha} \tanh \beta L_c \sinh \alpha L_b}{\sinh \alpha L_b + \dfrac{\rho_d}{\rho_s + \rho_d} \dfrac{\beta}{\alpha} \tanh \beta L_c \cosh \alpha L_b} \tag{7.25}$$

ρ_d is the sheet resistivity of the diffusion, ρ_s is the sheet resistivity of the silicide diffusion, W is the diffusion width. Also, the parameter α and β are given by $(\rho_d/\rho_s)^{1/2}$ and $[(\rho_d+\rho_s)/\rho_c]^{1/2}$, respectively. In Eq. (7.23), the first term is the geometrical line resistance and the second term is the contact resistance. Since the sheet resistivity of the silicide diffusion ρ_s is negligible compared with the sheet resistivity of the diffusion ρ_d, it is assumed that $\rho_s \approx 0\ \Omega/\Box$. Therefore, the total resistance is approximately equal to the contact resistance R_c itself [in Eq. (7.24)]. Also, the temperature dependence of ρ_d is much less significant than that of ρ_c, for simplicity in this work ρ_d is considered to be constant with temperature.

However, for increased accuracy, it is recommended to account for the temperature dependence of ρ_d, which can be realized in terms of the temperature dependence of Silicon bulk mobility [91, 94]. Another important parameter of the silicided contact system is the transfer contact length L_T, which is given by

$$L_T = \sqrt{\rho_c / \rho_d} \qquad (7.26)$$

The contact transfer length L_T characterizes the distance over which the current moves from the silicide into the diffusion. It simply means that the contact resistance abruptly increases if $L_a \ll L_T$. Therefore, for previous technologies, L_a was designed to be much longer than L_T. For advanced technologies, using the ρ_c at room temperature (~ 10^{-6} to ~10^{-9} Ω–cm^2) in Fig. 7.3, however, L_T is estimated to be a few tenth nanometers to ~ 1 μm. These values are roughly comparable to the values of L_a, L_b, and L_c and thus it can be expected that the current distribution in the silicide contact system is strongly influenced by such design parameters.

Using the derived temperature dependent ρ_c, the current crowding effect has been investigated. According to the 1-D transmission line model for contact resistance [80], the position (x) dependent fraction of current in the diffusion $I_D(x)/I$ for each segmented region can be expressed based on Eqs. (7.27) to (7.36).

$$I_D(x)/I = -\frac{W}{\rho_d}\left[\frac{\beta\rho_d}{\rho_d+\rho_s}(A_a e^{\beta x} - B_a e^{-\beta x}) + D_a\right] \qquad \text{for } x < L_a \qquad (7.27)$$

where

$$A_a = -\frac{\rho_d[R_c'/\rho_d + (R_c'/\rho_s)e^{-\beta L_a} - (\rho_s+\rho_d)/\rho_s \cdot (1/\beta W)e^{-\beta L_a}]}{(R_c'/\rho_s)\beta W(e^{\beta L_a} - e^{-\beta L_a}) + (\rho_s+\rho_d)/\rho_s \cdot (e^{\beta L_a} + e^{-\beta L_a})} \qquad (7.28)$$

$$B_a = -\frac{\rho_d[R_c'/\rho_d + (R_c'/\rho_s)e^{-\beta L_a} + (\rho_s+\rho_d)/\rho_s \cdot (1/\beta W)e^{-\beta L_a}]}{(R_c'/\rho_s)\beta W(e^{\beta L_a} - e^{-\beta L_a}) + (\rho_s+\rho_d)/\rho_s \cdot (e^{\beta L_a} + e^{-\beta L_a})} \qquad (7.29)$$

$$D_a = \frac{\rho_d\rho_s}{W(\rho_s+\rho_d)} \qquad (7.30)$$

$$I_D(x)/I = -\frac{\alpha W}{\rho_d}(A_b e^{\beta x} - B_b e^{-\beta x}) \qquad \text{for } L_a \le x < L_a + L_b \tag{7.31}$$

where

$$A_b = \frac{\rho_d (1/\alpha W - R_c'/\rho_d)[\cosh(\beta L_a) - \rho_d/\rho_s]}{(R_c'/\rho_s)\beta W (e^{\beta L_a} - e^{-\beta L_a}) + (\rho_s + \rho_d)/\rho_s \cdot (e^{\beta L_a} + e^{-\beta L_a})} \tag{7.32}$$

$$B_b = -\frac{\rho_d (1/\alpha W + R_c'/\rho_d)[\cosh(\beta L_a) + \rho_d/\rho_s]}{(R_c'/\rho_s)\beta W (e^{\beta L_a} - e^{-\beta L_a}) + (\rho_s + \rho_d)/\rho_s \cdot (e^{\beta L_a} + e^{-\beta L_a})} \tag{7.33}$$

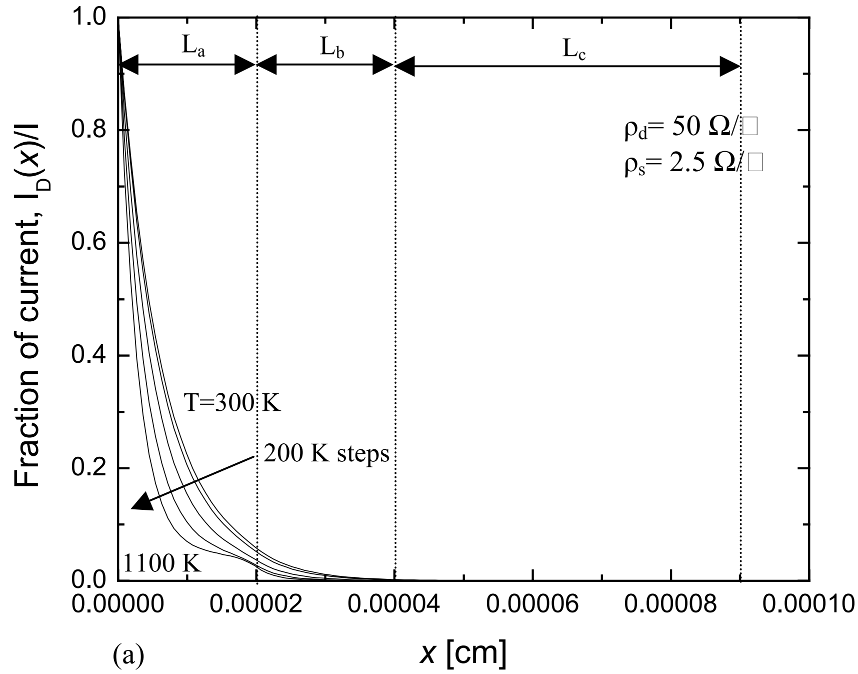
$$I_D(x)/I = -\frac{W}{\rho_d}[\frac{\beta \rho_d}{\rho_d + \rho_s}(A_c e^{\beta x} - B_c e^{-\beta x}) + D_c] \qquad \text{for } L_a + L_b \le x \le L_a + L_b + L_c \tag{7.34}$$

where

$$A_c = (\rho_s + \rho_d)/\rho_d \cdot (\alpha/\beta) \cdot [A_b e^{\alpha L_b} - B_b e^{-\alpha L_b} /(1 - e^{2\beta L_c})] \tag{7.35}$$

$$B_c = A_c \cdot e^{2\beta L_c} \tag{7.36}$$

Since temperature dependencies of other parameters except ρ_c are ignored, thermal effect on the current distribution can be seen directly from thermal behavior of ρ_c. Fig. 7.5 clearly shows how current flow under silicide contact is affected by temperature, depending on the sheet resistivity of the diffusion. Compared to the current distribution at room temperature, the current is severely localized with temperature since the most of current flows for the region of L_a with increase in temperature. Also, the impact of temperature is more prominent for the case of the higher sheet resistivity [Fig. 7.5 (a)]. This implies that effective volume for power dissipation for a given current stress (i.e., ESD current stress) is dramatically reduced as temperature increases. Furthermore, it implies that the device with silicide diffusion could fail thermally at much lower current levels than predicted when ignoring thermal effects because the temperature dependence of the current localization is severe which in turn causes enhanced heating near the edge of silicided contact (i.e., near the spacer edge).



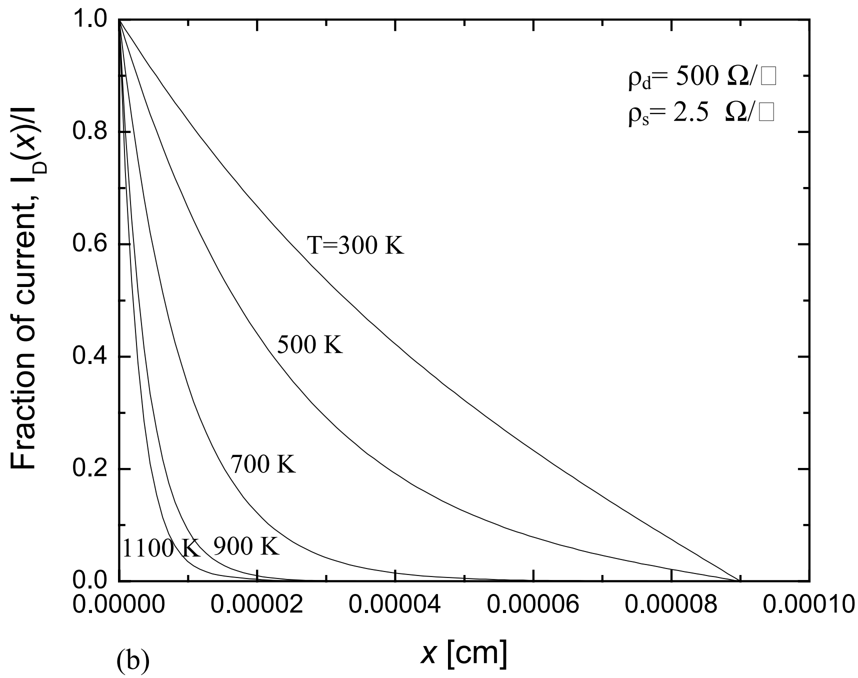


Figure 7.5: Fraction of the drain current as a function of the position from the edge of the spacer *x* [Fig. 7.1] where L_a=0.2 µm, L_b=0.2 µm, L_c=0.5 µm and W=1 µm. (a) highly doped source/drain (ρ_d=50Ω/□) and (b) moderately doped source/drain (ρ_d=500 Ω/□).

7. 4 Summary

Based on theoretical approaches, a temperature dependent specific contact resistance model has been derived. The specific contact resistance for $CoSi_2$-Silicon dramatically decreases with increase in temperature depending on the doping concentration in the diffusion. Also, it is shown that variation of the specific contact resistance value has significant implications for ESD reliability through localized heating by current crowding effects since the effective volume for power dissipation is reduced with the extent of the current localization for a given ESD stress.

The theoretically proposed temperature dependent contact resistance model can generate silicide contact resistance values at high temperature and it is capable of predicting high current behavior of silicided deep submicron devices, which is useful for analysis of ESD reliability.

CHAPTER 8

CONCLUSIONS

Electrostatic Discharge (ESD) is a common phenomenon arising from charge restoring process in nature and the current and voltage levels associated with ESD events are up to several amperes and k-volts. In the IC industry, it is known that ESD/EOS related damage is responsible for nearly 40 % of IC failures from customer returns. Since microelectronic components are very susceptible to ESD stress, the protection of IC chips against ESD has become a pacing factor in developing new technologies. Given the major impact of ESD on reliability of IC chips, there is a strong need of developing protection strategies.

Therefore, there have been ongoing efforts to improve design of on-chip ESD protection while growing demands for new applications drive CMOS technology into nano-scale regime. For technology scaling, earlier indications have been that the impact of technology scaling on design of ESD protection may be positive rather than negative [13], but more recent work [70] as well as this work have shown that some ESD performance limits may be approaching as CMOS technology moves far below the sub-0.25 μm regime where it was observed that ESD hardness begins to severely degrade for short channel NMOS transistors. In addition, other studies have reported that dielectric damage is the primary failure mode with technology scaling [36, 68].

While circuit designers have successfully created robust ESD protection for past CMOS technologies, a lack of understanding of the failure mechanisms underlying ESD damage observed in advanced technologies limited the amount of transferable knowledge from one technology node to the next.

Therefore, in order to improve understanding of ESD behavior involved in advanced technology the focus of this thesis has been directly towards non-uniform conduction and relevant issues during ESD events in single finger NMOS transistors. The investigation has considered both experimental results and numerical simulations. The knowledge acquired on ESD phenomena is fundamental to understand and predict the behavior of more complex and effective protection networks. Moreover, it is essential to generate ESD device models that can be implemented in compact model simulators to provide accurate and reliable simulations prior to committing to final silicon implementation. This chapter reviews the contributions of this thesis and proposes future work that is needed for developing high performance ESD protection.

8. 1 Contributions

Investigating various aspects of ESD behavior involved in the advanced 0.13 μm CMOS technology of Texas Instruments Inc., it is shown that the non-uniform bipolar conduction phenomenon during ESD events result in a severe reduction in ESD protection strength depending on the device finger width as well as silicide process. This non-uniform bipolar conduction becomes more serious with silicided protection devices and thus is identified as a root cause of the abnormal degradation of ESD hardness of the silicided device, compared with non-silicided devices [47, 48].

Using experiments and device simulations, it is inferred that the root cause of this non-uniform conduction could be related to the intrinsic process defects that lock in the bipolar conduction to a local area in the finger width and it has also been provided improved understanding of ESD behavior and new physical insight into the current localization effects. Furthermore, an extensive investigation into impact of the substrate bias and gate bias on ESD performance has been carried out to provide new insight into the bias effect involved [48, 53, 54].

It is shown that ESD performance can be improved with the substrate bias by enlarging the effective turned-on device width. Additionally, the concept of the intrinsic second breakdown triggering current (I_{t2i}) is introduced, which is substrate bias independent and represents the maximum achievable ESD failure strength for a given technology.

For devices with significant gate coupling, it is also shown that gate bias induced heating near the drain extension region close to the Si/SiO_2 surface is the primary cause of the degradation of ESD performance. Moreover, it is established that substrate biasing can help eliminate the negative impact of the gate bias effect.

The improvement of ESD failure threshold with the gate-to-contact spacing for fully silicided NMOS transistors have also been investigated; the results provide new ESD design rules for deep submicron technology, based on detailed experimental and simulation results. It has been shown that the reduction in current localization and increase in the power dissipation volume with increases in the gate-to-contact spacing are the primary factors influencing improvement of ESD robustness. It has also been established that substrate biasing can help reduce the impact of gate-to-contact spacing on the ESD strength of the silicided devices. Results from this work suggest that even for silicided processes, the gate-to-contact spacing should be carefully engineered to achieve efficient and robust ESD protection designs [65, 66].

Along with the non-uniform bipolar conduction and the relevant gate-to-contact spacing effects, it is also identified that for deep submicron devices an observed unusual reverse channel

length dependence of I_{t2} is due to the reduction of thermal capacity arising from the reduced effective device width associated with severe non-uniform bipolar conduction. Also it is demonstrated that this degradation phenomenon can be mitigated with the application of substrate bias, which can be effective in protection circuits for various applications, since for ESD applications, the protection design with minimum L_{poly} is critical for achieving efficient input gate oxide and internal core circuit reliability.

Finally, to further enhance understanding impact of silicided process on ESD hardness depending on source/drain engineering, an analytic silicide contact resistance model that can describe high temperature and high current behavior of the silicided contact system was formatted using physics based approaches. The model shows how the current can be localized under ESD conditions depending on device parameters and thus it extends the design capability of ESD protection providing information relevant to ESD failure limit.

Results from this work can be used to construct suitable design windows for efficient and robust ESD on-chip protection to overcome early ESD failures in advanced deep submicron CMOS technologies and have been demonstrated at Texas Instruments Inc., in practice ESD protection designs as well as helping to establish design guidelines for use in high performance 0.13 μm technology.

8. 2 Suggested Future Work

As CMOS technology moves toward the nanometer regime, it is more critical to further improve understanding of mechanisms underlying ESD behaviors in protection circuit networks for developing consistent predictive models, which can be used to evaluate the impact of process technology and design variations on ESD performance of protection elements. Therefore, the development of robust ESD reliability should be a part of the process and technology roadmap to ensure that good ESD level is realized in future generations of IC chips.

As a part of work that extends our knowledge on ESD in advanced CMOS technology, this thesis has been presented on specific aspects of ESD behavior observed, but there is still a lack of in-depth understanding of many aspects of ESD behavior. Thus, several recommendations for future research are discussed.

With aggressive technology scaling, minimum feature size of a transistor is being dramatically decreased. This implies that the device behavior becomes more susceptible to thermal effects due to an ongoing reduction in effective volume for power dissipation. Moreover, the importance of thermal effects with the advancement of technology has significant implications on ESD behavior since ESD is a phenomenon involving high temperature operation. Therefore, as discussed through this thesis the non-uniform ESD current distribution effects should also be understand including impact of temperature rise during ESD event. Especially, for analysis of the devices with silicided process, temperature dependent current localization is more important and thus it should be investigated in detail for future technologies.

In addition, as IC chips become more complex and larger, area-efficient on-chip protection is more important. Thus, optimization of the protection circuits for a given process are indispensable and of course, this can be realized based on circuit simulation that includes ESD compact models. Significant work has already been done to create compact models for MOS snapback and thermal failure [73-75]. However, as presented in this thesis the 3-D non-uniform behavior of devices is critical in modeling the physics underlying real-world ESD behavior of advanced transistors. Although recent work [95] has demonstrated the capability of 3-D device simulation for ESD analysis, there are still issues under discussion such as model parameter calibration and statistical process variations.

Furthermore, taking into account the combined 3-D and thermal effects in advanced technologies, ESD compact model with an appropriate thermal network, which is able to

reproduce the ESD behavior of protection elements in both steady state and transient state, should be implemented in a suitable circuit simulator environment.

Finally, it should also be noted that the estimated thermal effects during ESD events from electorthermal device simulation may be underestimated or produce erroneous results. Since the device simulators commercially available are based on the classical heat diffusion equation, results obtained from those simulators seem to be inaccurate for the deep submicron technology. As an intensive approach for understanding thermal effect in deep submicron technology, new electrothermal models are being developed based on *Phonon Boltzmann Transport* equation, yet they are not commercially available [60]. In this regard, improved electorthermal models based on advanced thermal physics are required for understanding ESD behavior of nano-scale transistors.

There are virtually limitless productive avenues toward increased understanding of fundamentals of ESD phenomena. Moreover, the impact of ESD on future scaling and reliability is a major concern for IC product reliability: As such the topic of ESD continues to be a very productive area for further research efforts.

BIBLIOGRAPHY

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

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

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

[4] A. Amerasekera and C. Duvvury, *ESD in Silicon Integrated Circuits*, 2nd ed., Wiley, 2002.

[5] J. E. Vinson and J. J. Liou, "Electrostatic discharge in semiconductor devices: protection techniques," *Proceedings of the IEEE*, Vol. 88, pp. 1878-1900, 2000.

[6] http://public.itrs.net/Files/2001ITRS/Home.htm, *The International Technology Roadmap for Semiconductors*, 2001 edition.

[7] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, 1998.

[8] T. L. Polgreen and A. Chatterjee, "Improving the ESD failure threshold of silicided n-MOS output transistor by ensuring uniform current flow," *IEEE Trans. Electron Device*, Vol. 39, No. 2, pp.379-388, 1992.

[9] A. Amerasekera, S. Ramaswamy, M-C Chang, and C. Duvvury, "Modeling MOS snapback and parasitic bipolar action for circuit-level ESD and high current simulations," *Proceedings of the IRPS*, pp. 318-326, 1996.

[10] R. N. Rountree, "ESD protection for submicron CMOS circuits issues and solutions," *IEDM Tech. Dig.*, pp. 580-583, 1988.

[11] S. Beebe, "Characterization, Modeling and Design of ESD Protection Circuits," *PhD thesis*, Stanford University, 1998.

[12] C. Duvvury and A. Amerasekera, "ESD: A pervasive reliability concern for IC technologies," *Proceedings of the IEEE*, Vol. 81, No. 5, pp. 690-702, 1993.

[13] A. Amerasekera and C. Duvvury, "The impact of technology scaling on ESD robustness and protection circuit design," *IEEE Trans*. *Comp*., *Packag*., *Manufact.*, *Tech*., Part A, Vol. 18, No. 2, pp. 314-320, 1995.

[14] G. Notermans, A. Heringa, M. van Dort, S. Jansen and F. Kuper, "The effect of silicide on ESD performance," *Proceedings of the IRPS*, pp. 154-158, 1999.

[15] A. Amerasekera, V. McNeil, and M. Rodder, "Correlating drain junction scaling, salicide thickness, and lateral NPN behavior with the ESD/EOS performance of a 0.25 mm CMOS process," *IEDM Tech*. *Dig*., pp. 893-896, 1996.

[16] 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 parameter," *IEEE Trans*. *Electron Device*, Vol. 38, No. 9, pp. 2161-2168, 1991.

[17] C. H. Diaz, S.-M. Kang, and C. Duvvury, *Modeling of Electrical Overstress in Integrated Circuits*, Kluwer Academic Publishers, 1995.

[18] S. H. Voldman, "The state of the art of electrostatic discharge protection: physics, technology, circuits, design, simulation, and scaling," *IEEE J. Solid-State Circuits*, Vol. 34, No. 9, pp. 1272-1282, 1999.

[19] P. Yang and J.-H. Chern, "Design for reliability: the major challenge for VLSI," *Proceedings of the IEEE*, Vol. 81, No. 5, pp. 730-744, 1993.

[20] R. Y. Moss, "Caution-Electrostatic discharge at work," *IEEE Trans*. *Comp*. *Hybr*. *and Man*., CHMT-5, pp. 512-515, 1982.

[21] Kai Esmark, *Device Simulation of ESD Protection Elements*, Hartung-Gorre Verlag Konstanz, 2002.

[22] G. Notermans, "On the use of N-well resistors for uniform triggering of ESD protection elements," *Proceedings of the EOS/ESD Symposium*, pp. 221-229, 1997.

[23] G. Krieger and P. Niles, "Diffused resistors characteristics at high current density levels-analysis and applications," *IEEE Trans. Electron Device*, Vol. 36, No. 2, pp. 416-423, 1989.

[24] C. Duvvury, and C. Diaz, "Dynamic gate coupled NMOS for efficient output ESD protection," *Proceedings of the IRPS*, pp. 141-150, 1992.

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

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

[27] J. C. Smith, "A substrate triggered lateral bipolar circuit for high voltage tolerant ESD protection applications," *Proceedings of the EOS/ESD Symposium*, pp. 63-71, 1998.

[28] C. Duvvury, S. Ramaswamy, A. Amerasekera and R. A. Cline, "Substrate pump NMOS for ESD protection application," *Proceedings of the EOS/ESD Symposium,* pp. 7 -17, 2000.

[29] A. Amerasekera, M.-C. Chang, J. A. Seitchik, A. Chatterjee, K. Mayaram, and J.-H. Chern, "Self-heating effects in basic semiconductor structures," *IEEE Trans. Electron Device*, Vol. 40, No. 10, pp. 1836-1844, 1993.

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

[31] H. Gieser, “Methods for the characterization of integrated circuits employing very fast high current impulses,” *PhD thesis*, Technical University of Munich, 1999.

[32] C. Duvvury, “Issues in deep submicron state-of-the-art ESD design,” *Tutorials of the ISQED*, session. B1, 2002.

[33] A. Amerasekera, W. van den. Abeelen, L. van Roozendaal, M. Hannemann, P. Schofield, “ESD failure modes: characteristics, mechanisms, and process influences,” *IEEE Trans. Electron Device*, Vol. 39, No. 2, pp. 430-436, 1992.

[34] K.-L. Chen, “Effects of interconnect process and snapback voltage on the ESD failure threshold of NMOS transistors,” *Proceedings of the EOS/ESD Symposium*, pp. 211-219, 1988.

[35] S. Aur, A. Chatterjee, and T. Polgreen, “Hot-electron reliability and ESD latent damage,” *IEEE Trans. Electron Device*, Vol. 35, No. 12, pp. 2189-2193, 1988.

[36] A. Salman, R. Gauthier, W. Stadler, K. Esmark, M. Muhammad, C. Putnam, and D. Ioannou, “NMOSFET ESD self-protection strategy and underlying failure mechanism in advanced 0.13 μm CMOS technology,” *IEEE Trans. Device and Materials Reliability*, Vol. 2, No. 1, pp. 2-8, 2002.

[37] G. Meneghesso, S. Santirosi, E. Novarini, C. Contiero, and E. Zanoni, “ESD robustness of smart-power protection structures evaluated by means of HBM and TLP tests,” *Proceedings of the IRPS*, pp. 270-275, 2000.

[38] G. Notermans, P. De Jong, and F. Kuper, “Pitfalls when correlating TLP, HBM and MM testing,” *Proceedings of the EOS/ESD Symposium*, pp. 170-176, 1998.

[39] W. Stadler, X. Guggenmos, P. Egger, H. Gieser, and C. Musshoff, “Does the ESD failure current obtained by transmission line pulsing always correlate to human body model tests?,” *Proceedings of the EOS/ESD Symposium*, pp. 366-372, 1997.

[40] T. J. Maloney and N. Khurana, "Transmission line pulsing techniques for circuit modeling of ESD phenomena," *Proceedings of the EOS/ESD Symposium*, pp. 49-54, 1985.

[41] C. Russ, "ESD protection design for CMOS technologies: processing impact, modeling and testing issues," *PhD thesis*, Technical University of Munich, 1999.

[42] G. Boselli, "On high injection mechanisms in semiconductor devices under ESD conditions," *PhD thesis*, University of Twente, 2001.

[43] H. Ishizuka, K. Okuyama, and K. Kubota, "Photon emission study of ESD protection devices under second breakdown conditions," *Proceedings of the IRPS*, pp. 286-291, 1994.

[44] P. Salome, C. Leroux, J. P. Chante, P. Crevel, and G. Reimbold, "Study of a 3D phenomenon during ESD stress in deep submicron CMOS technologies using photon emission tool," *Proceedings of the IRPS*, pp. 325-332, 1997.

[45] C. Russ, K. Bock, M. Rasras, and I. D. Wolf, "Non-uniform triggering of gg-nMOST investigated by combined emission microscopy and transmission line pulsing," *Proceedings of the EOS/ESD Symposium*, pp. 177-186, 1998.

[46] C. Duvvury, C. Diaz, and T. Haddock, "Achieving uniform nMOS power distribution improving ESD/EOS reliability," *IEDM Tech. Dig.*, pp. 131-134, 1992.

[47] K-H. Oh, C. Duvvury, C. Salling, K. Banerjee, and R. W. Dutton, "Non-uniform bipolar conduction in single finger NMOS transistors and implications for deep submicron ESD design," *Proceedings of the IRPS*, pp. 226-234, 2001.

[48] K-H. Oh, C. Duvvury, K. Banerjee, and R. W. Dutton, "Analysis of non-uniform ESD current distribution in deep Submicron NMOS Transistors," *IEEE Trans. Electron Device*, Vol. 49, No. 10, to appear, 2002.

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

[50] M. Litzenberger, K. Esmark, D. Pogany, C. Furbock, H. Gossner, E. Gornik, and W. Fichtner, "Study of triggering inhomogeneities in gg-nMOS protection devices via thermal mapping using backside laser interferometry," *Microelectronics Reliability*, Vol. 40, pp. 1359-1364, 2000.

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

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

[53] K-H. Oh, C. Duvvury, K. Banerjee, and R. W. Dutton, "Gate bias induced heating effects and implications for the design of deep submicron ESD protection," *IEDM Tech*. *Dig*., pp. 315-318, 2001.

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

[55] C. Salling, J. Hu, J. Wu, C. Duvvury, R. Cline, and R. Pok, "Development of substrate pumped nMOS protection for a 0.13 mm technology," *Proceedings of the EOS/ESD Symposium*, pp. 192–204, 2001.

[56] S. Ramaswamy, C. Duvvury, and S-M. Kang, "EOS/ESD reliability of deep-submicron nMOS protection devices," *Proceedings of the IRPS*, pp. 284-291, 1995.

[57] A. Amerasekera and C. Duvvury, *ESD in Silicon Integrated Circuits*, Wiley, 1995.

[58] *TSUPREM -4 Manual*, TMA Inc., Sunnyvale, CA, 1996.

[59] *ANSYS Manual*, Ansys Inc., Canonsburg, PA, 1998.

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

[61] *MEDICI Manual*, TMA Inc., Sunnyvale, CA, 1997.

[62] J. Smith, "An anti-snapback circuit technique for inhibiting parasitic bipolar conduction during EOS/ESD events," *Proceedings of the EOS/ESD Symposium*, pp. 62-69, 1999.

[63] K. Verhaege and C. Russ, "Wafer cost reduction through design of high performance fully silicided ESD devices," *Proceedings of the EOS/ESD Symposium*, pp. 18-28, 2000.

[64] T. Suzuki, S. Mitarai, S. Ito, H. Monma, and N. Higashi, "A study of fully 0.18 μm CMOS ESD protection devices," *Proceedings of the EOS/ESD Symposium*, pp. 78-87, 1999.

[65] K-H. Oh, C. Duvvury, K. Banerjee, and R. W. Dutton, "Investigation of gate to contact spacing effect on ESD robustness of salicided deep submicron single finger NMOS transistors," *Proceedings of the IRPS*, pp. 148-155, 2002.

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

[67] K. Banerjee, A. Amerasekera, G. Dixit, and C. Hu, "Temperature and Current Effects on Small-Geometry-Contact Resistance," *IEDM Tech. Dig.*, pp.115-118, 1997.

[68] A. Amerasekera, V. Gupta, K. Vasanth, and S. Ramaswamy, "Analysis of snapback behavior on the ESD capability of sub-0.20 μm NMOS," *Proceedings of the IRPS*, pp. 159-166, 1999.

[69] S. M. Sze, *Physics of Semiconductor Devices*, New York: Wiley, 1981.

[70] 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," *Proceedings of the EOS/ESD Symposium*, pp. 95-104, 1999.

[71] P. Oldiges, Q. Lin, K. Petrillo, M. Sanchez, M. Ieong, M. Hargrove, "Modeling line edge roughness effects in sub 100 nanometer gate length devices," *Proceedings of the SISPAD*, pp. 131-134, 2000.

[72] K. Banerjee, C. Hu, A. Amerasekera, and J. A. Kittl, "High current effects in silicide films for sub-0.25 μm VLSI technologies," *Proceedings of the IRPS*, pp. 284-292, 2002.

[73] V. M. Dwyer, A. J. Franklin, and D. S. Campbell, "Thermal failure in semiconductor devices," *Solid-State Electron*., Vol. 33, pp. 553-560, 1990.

[74] C-H. Choi, "Modeling of Nanoscale MOSFETs," *PhD thesis*, Stanford University, 2002.

[75] K. K. Ng and W. L. Lynch, "The impact of intrinsic series resistance on MOSFET scaling, "*IEEE Trans*. *Electron Device*, Vol. ED-34, No. 3, pp. 503-511, 1987.

[76] Y-S. Chieh, A. H. Perera, and J. P. Krusius, "Series resistance of silicided ohmic contacts for nanoelectronics," *IEEE Trans*. *Electron Device*, Vol. 39, No. 8, pp. 1882-1888, 1992.

[77] K-S. Kim, C-M. Park, and J. C. S. Woo, "Advanced model and analysis of series resistance for CMOS scaling into nanometer regime-part I: theoretical derivation," *IEEE Trans*. *Electron Device*, Vol. 49, No. 3, pp. 457-466, 2002.

[78] K-S. Kim, C-M. Park, and J. C. S. Woo, "Advanced model and analysis of series resistance for CMOS scaling into nanometer regime-part II: quantative analysis," *IEEE Trans*. *Electron Device*, Vol. 49, No. 3, pp. 467-472, 2002.

[79] D. B. Scott, W. R. Hunter, and H. Shichijo, "A transmission line model for silicided diffusions: impact on the performance of VLSI circuits," *IEEE Trans*. *Electron Device*, Vol. ED-29, No. 4, pp. 651-661, 1982.

[80] D. B. Scott, R. A. Chapman, C-C. Wei, S. S. Mahant-shetti, R. A. Haken, and T. C. Holloway, "Titanium disilicided contact resistivity and its impact on 1-μm CMOS circuit performance," *IEEE Trans. Electron Device*, Vol. ED-34, No. 3, pp. 562-573, 1987.

[81] F. A. Padovani and R. Stratton, "Field and thermionic-field emission in Schottky barriers," *Solid-State Electronics*, Vol. 9, pp. 695-707, 1966.

[82] A. Y. C. Yu, "Electron tunneling and contact resistance of metal-silicon contact barriers," *Solid-State Electronics*, Vol. 13, pp. 239-247, 1970.

[83] C. R. Crowell and V. L. Rideout, "Normalized thermionic-field (T-F) emission in metal-semiconductor (Schottky) barriers," *Solid-State Electronics*, Vol. 12, pp. 89-105, 1969.

[84] K. Varahramyan and E. J. Verret, "A model for specific contact resistance applicable for titanium silicide-silicon contacts," *Solid-State Electronics*, Vol. 39, No. 11, pp. 1601-1607, 1996.

[85] C. R. Crowell, "Richardson constant and tunneling effective mass for thermionic and thermionic-field emission in Schottky barrier diodes," *Sold-State Electronics*, Vol. 12, pp. 55-59, 1969.

[86] W. B. Joyce and R. W. Dixon, "Analytic approximations for the fermi energy of an ideal fermi gas," *Appl. Phys. Lett.*, Vol. 31, No. 5, pp. 354-356, 1977.

[87] M. A. Green, "Intrinsic concentration, effective densities of states, and effective mass in Silicon," *J. Appl. Phys.*, vol. 67, no. 6, pp. 2944-2954, 1990.

[88] A. M. Cowley and S. M. Sze, "Surface states and barrier height of metal-semiconductor systems," *J. Appl. Phys.*, Vol. 36, No. 10, pp. 3212-3220, 1965.

[89] Y-S. Lou and C-Y. Wu, "A self-consistent characterization methodology for Schottky-barrier diodes and ohmic contacts," *IEEE Trans. Electron Devices*, Vol. 41, No. 4, pp. 558-566, 1994.

[90] F. H. Gaensslen, R. C. Jaeger, "Temperature dependent threshold behavior of depletion mode MOSFET," *Solid-State Electronics*, Vol. 22, pp. 423-430, 1979.

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

[92] K. K. Ng and R. Liu, "On the calculation of specific contact resistivity on <100> Si," *IEEE Trans*. *Electron Device*, Vol. 37, No. 6, pp. 1535-1537, 1990.

[93] C. R. Crowell, "The Richardson constant for thermionic emission in Schottky barrier diodes," *Solid-State Electronics*, Vol. 8, pp. 395-399, 1965.

[94] N. Arora, *MOSFET Models for VLSI Circuit Simulation*, Springer-Verlag, Wien New York, 1993.

[95] W. Fichtner, K. Esmark, and W. Stadler, "TCAD software for ESD on-chip protection design," *IEDM Tech*. *Dig*., pp. 311-314, 2001.