ANALYTICAL STUDY OF CMOS-SUBTHRESHOLD OPERATION FOR LOW POWER VLSI APPLICATIONS

A THESIS

Submitted in fulfilment of the
requirements for the award of the degree
of
DOCTOR OF PHILOSOPHY
in
ELECTRONICS AND COMPUTER ENGINEERING

by

BHAVANA JHARIA

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)

OCTOBER, 2007

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **'ANALYTICAL STUDY OF CMOS-SUBTHRESHOLD OPERATION FOR LOW POWER VLSI APPLICATIONS'** in fulfilment of the requirements for the award of the Degree of Doctor of Philosophy submitted in the Department of Electronics and Computer Engineering of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from July 2001 to April 2006 under the supervision of Prof. R. P. Agarwal, Prof. S. Sarkar and Prof. S. C. Handa.

The suggestions have been incorporated in compliance with the examiners' remarks.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Bhavana

(BHAVANA JHARIA)

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

Date:29/10/07	**Dr. R. P. Agarwal** Professor Dept. of E & CE	**Dr. S. Sarkar** Professor (Rtd) Dept. of E & CE	**Dr. S. C. Handa** Prof.& Coordinator (Rtd) Q. I. P. Centre

---

The Ph.D. Viva-Voce examination of **Ms. BHAVANA JHARIA**, Research Scholar, has been held on 24/12/2008.

Signature of Supervisor (s)

Signature of External Examiner

Signature of HOD

ABSTRACT

In recent years, there has been increasing prominence of portable battery operated low power systems such as hearing aids, pacemakers, cell phones, pagers, and portable computers become more complex, prevalent, and demand increased battery life. Therefore, the demand for increased battery life will require designers to seek out new technologies and circuit techniques to maintain high performance with longer battery lifetime. Portable devices, however, are not the sole motive behind the low power and low energy design efforts. The increasing power dissipation for fixed supply devices is almost equally challenging as for portable devices.

As technology feature size is reduced, the number of transistors on the chip is increased and more power is dissipated. The main components of power dissipation are switching power and leakage power. Switching power, being the dominant power component, has caught special attention in recent years. Many techniques have been introduced to control this ever-increasing power component on all levels of design abstraction. Increased leakage current due to technology feature downsizing is another challenge that faces circuit designers in the VLSI and ULSI era. In the VLSI circuits a large number of MOSFETs are often simultaneously biased in the subthreshold region of operation due to ultra-low power consumption with moderate performance degradation. In such a situation, the subthreshold currents due to the individual devices add up to a source of substantial power dissipation. This power dissipation is highly undesirable, as it causes many reliability problems. In submicron devices the problem is further aggravated by the high electric field existing in the channel region. Under the influence of high electric field the carriers multiply by impact-ionization. The result is an increased drain current.

In this thesis, an analytical approach is developed to analyze effect of scaling and impact ionization in submicron and deep submicron MOSFETs.

To include the effect of impact ionization on drain current, multiplication factor is first determined using well-established expressions. The analysis shows that for subthreshold operation it is important that the effect of impact ionization should be included in the analytical expression for drain current. Calculations show that multiplication factor increases with drain-source voltage, V_{DS}, and gate voltage and decreases with increase in gate oxide thickness. Calculations also show that multiplication factor increases with increase in drain/source pn-junction depth. However, multiplication factor levels off at larger junction depths. Shallower junctions are therefore desirable if carrier multiplication is to be restricted.

Present study considered Fjeldly and Shur's physical model for subthreshold current. While this model yields accurate results for submicron devices, it does not take into account the effect of impact ionization, which is more prominent in devices of nanometer channel length. The subthreshold current model of Fjeldly and Shur is therefore, first modified to include the effect of impact ionization, for all analytical purposes. The modified current model agrees well with results obtained from ATLAS simulation. It shows that impact ionization increases subthreshold current by few percents. Such an increase in subthreshold current model, the DIBL –coefficient was reconsidered. It has been found that this parameter is not independent of the terminal voltages of a MOSFET. The DIBL- coefficient has a strong dependence on V_{DS} and relatively lesser dependence on gate-source voltages V_{GS}. This analysis shows that, because of the strong V_{DS} dependence of the DIBL-coefficient, the threshold voltage shift is a non-linear function of V_{DS}, in contradiction to a constant DIBL-coefficient as assumed previously. It is also found that DIBL- coefficient increases with increase in channel length and decreases with increase in temperature.

Effect of scaling methods on the subthreshold conduction of a small geometry device is also analyzed. The results show that as the channel length is scaled down, multiplication factor increases slowly in the higher regime of channel length. In the lower regime of channel length, multiplication factor rises rapidly. This result also justifies the inclusion of impact –

ionization effect on subthreshold current. However, the analysis shows that there is insignificant dependence of multiplication factor on the method of scaling. Constant electric field scaling produces the largest threshold voltage shift due to DIBL. Selective scaling has the least DIBL, with generalized scaling produced DIBL falling in between the two extremes.

Finally in this thesis, an analytical approach is developed to obtain optimal V_{DD} for minimizing total power dissipation. It is well known that in above threshold CMOS operation, a tradeoff exists between propagation delay and power dissipation, when voltage is scaled. When the supply voltage is scaled down to reduce power dissipation, propagation delay is increased. SPICE simulation, however, shows that this does not hold good for subthreshold logic. For a threshold CMOS– logic operation, propagation delay in most cases first increases and then decreases with supply voltage scaling. The subthreshold current model mentioned above is used to calculate voltage dependence of propagation delay of a CMOS inverter. The results thus obtained show that, the delay increases with increase in supply voltage. An important requirement of a logic circuit is symmetric input-output characteristics. The ratio of pMOS and nMOS widths that meets this requirement is determined using the modified subthreshold current model. The ratio thus obtained is primarily mobility dependent, as in the case of above threshold logic operation. The analytical model presented shows very well match with SPICE simulations.

The results of the present investigations can be very useful in designing low power VLSI circuits, which is an immediate necessity in the modern portable electronic gadgets.

ACKNOWLEDGMENTS

No work can be accomplished successfully and proficiently without the motivation, guidance and support of valuable persons. I am blessed to have **Prof. R. P. Agarwal**, **Prof. S. Sarkar** and **Prof. S. C. Handa** as my supervisors, who were the pillars of strength for me throughout the course of this work and supported me in hours of distress. I feel obliged to **Prof. R. P. Agarwal**, **Ex Vice-Chancellor, Bundelkhand University Jhansi (U.P.)**, who persuaded me to take up research in the area of Low Power VLSI. I am indeed grateful to **Prof. R. P. Agarwal** and **Prof. S. Sarkar** for their valuable assistance, dedicated support and helpful technical suggestions during the entire course of research. It was memorable experience for me to enrich myself from their vast experience, great wisdom and deep knowledge. I am beholden to **Prof. S. C. Handa** for his priceless support, advices, and encouragement throughout this research.

I also feel indebted to the Research Committee comprising of **Prof. A. K. Sarje, Prof. D. K. Mehra, Prof. A. K. Saxena and Prof. I. S. Tyagi** for their critical and constructive suggestions and comments during the initial and final scrutiny of this work.

I also want to extend my sincere thanks to **Prof. G. S. Visweswaran**, **I.I.T Delhi** for allowing me to use the laboratory facility and specially ATLAS software at **VLSI Lab.** at **IIT Delhi** to carry out my simulation work. I am highly grateful to him for providing me ample inspiration, holding useful discussion and giving me suggestions from time to time. Thanks are due to Shiva and Babu, M. Tech. students for helping me in the VLSI Lab. at IIT Delhi.

I sincerely thank all faculty members and the staff of Department of Electronics & Computer Engineering, IIT Roorkee for rendering desired help as and when needed. My special thanks are due to Co-ordinator QIP and staff who paid due attention to our problems and also showed a way to come out of it.

I acknowledge my sincere gratitude to the **AICTE, MHRD Govt. of India**, for financial assistance under QIP and **M.I.T. Govt. of India** for providing sophisticated VLSI Lab. facility under the project 'Special Manpower Development for VLSI Design and Related Softwares' Dept. of E & CE, IIT Roorkee. I owe my sincere gratitude to the **Governing Body** of my parent institute **Govt. Engineering College, Jabalpur** for allowing me to undertake this research under the QIP scheme. I have a special word of thanks for **Dr. S. B. L. Beohar,** Principal of my parent institute, **Prof. S. K. Grover,** former H.O.D. and **Prof M. P. Parsai**, H.O.D., Dept. of Electronics and Telecommunications for giving useful tips and suggestions.

My special thanks and heartfelt gratitude goes to **Dr. (Mrs.) Laxmi Agarwal** for ever making me feel owned, loved and a part of her family. I am thankful to **Mrs. Yashodhara Sharma**, Lecturer & Dy. Dir (HRD), Institute of Economics & Finance, Bundelkhand University, Jhansi for extending her help and support.

Thanks are due to my fellow research scholars and friends for sharing some very odd and even moments during the period.

I express my deep gratitude to my mother **Smt. Shashi Devi Jharia** for always posing a great confidence and instilling faith in me. I owe my respect and love to **Chandarji**, my brothers, sisters and brothers-in-laws for their support and good wishes.

I owe a deep sense of gratitude to All-pervading Spirit whose Divine light provided me the perseverance, guidance, inspiration, faith and strength to carry on even when the going got tough.

Bhavana Jharia

ABBREVIATIONS

ICs	Integrated Circuits
VLSI	Very Large Scale Integration
ULSI	Ultra Large Scale Integration
DSM	Deep Submicron
VTC	Voltage Transfer Characteristic
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
nMOS	n-channel MOSFET
pMOS	p-channel MOSFET
CMOS	Complimentary Metal Oxide Semiconductor
I/O	Input-Output
SOI	Silicon-on-Insulator
SOIAS	Silicon-on-Insulator with Active Substrate
DGSOI	Double Gate Silicon-on-Insulator
RAM	Random Access Memory
DRAM	Dynamic Random Access Memory
DTMOS	Dynamic-Threshold Metal Oxide Semiconductor
VTCMOS	Variable-Threshold Complimentary Metal Oxide Semiconductor
MTCMOS	Multi-Threshold Complimentary Metal Oxide Semiconductor
CPL	Complimentary Pass Logic
DIBL	Drain-Induced Barrier Lowering
GIDL	Gate Induced Drain Leakage
SPICE	Simulated Program for Integrated Circuit Emphasis

LP	Linear Programming
3G	3rd Generation
nm	Nano-Meter
µm	Micro-Meter
MC	Monte-Carlo
EED	Electron Energy Distribution
MBD	Maxwell-Boltzmann Distribution
BTE	Boltzmann Transport Equation
CEFS	Constant Electric Field Scaling
GS	Generalized Scaling
SS	Selective Scaling
LOCOS	Local Oxidation of Silicon
HCE	Hot Carrier Effects
GTC	Gate Tunneling Current
MLV	Minimum Leakage Vector
EDP	Energy-Delay Product

LIST OF SYMBOLS

W	Width of gate
L	Length of Gate
V_{DS}	Drain-Source Voltage
V_{GS}	Gate-Source Voltage
V_{GB}	Gate-Bulk Voltage
V_{DD}	Supply Voltage
V_{TH}	Threshold Voltage
V_{THn}	Threshold Voltage for nMOS
V_{THp}	Threshold Voltage for pMOS
V_{TH0}	Zero-Drain-Source bias threshold Voltage
V_T	Thermal Voltage
V_{sat}	Saturation Voltage
P_{SW}	Average Switching Power
α	Switching Activity Factor
C	Switching Capacitance
ΔV	Voltage Swing
f	Switching/Signal Frequency
$P_{parallel}$	Parallel Power
P_{org}	Original Power
C_{org}	Original effective Capacitance

f_{org}	Original Signal Frequency
E	Electric Field
E_m	Maximum Electric Field
E_{th}	Threshold Energy
λ	Mean Free path
α_i	Impact-Ionization Coefficient
$A_i(A_n),\ B_i(B_n)$	Impact-Ionization Rate Parameters
M	Multiplication Factor
M_{sat}	Multiplication Factor in Saturation region
M_{lin}	Multiplication Factor in Linear region
M_{sub}	Multiplication Factor in Subthreshold region
l_d	Characteristic Channel Length
t_{ox}	Oxide Thickness
x_j	Junction Depth
L_{eff}	Effective Gate Length
x_s	Source-Channel Depletion Width
x_d	Drain-Channel Depletion Width
I_{DS}	Drain Current With Impact-Ionization
I_D	Drain Current (Without Impact-Ionization)
I_{dd}	Drift-Diffusion Current
I_{te}	Thermionic Emission Current
A^*	Richardson's Constant
δ	Effective Channel Thickness

T	Temperature ^{0}K
η	Subthreshold Ideality Factor
χ	Constant
d_{dep}^0	Depletion width at Zero Drain Source Bias
C_{ox}	Gate-Oxide Capacitance
C_{dep}	Depletion Layer Capacitance
φ_b	Bulk Potential
n_i	Intrinsic Carrier Density
N_a	Acceptor Doping Density
D_n	Diffusion Coefficient
μ_n	Electron Mobility
σ	Drain-Induced Barrier Lowering Coefficient
ψ_s^0	Potential at Si-SiO_2 Interface at Zero Drain Bias
V_{bi}	Built-in Potential
$E(x)$	Lateral Electric Field at Distance 'x' from Source
V_{FB}	Flat-band Voltage
N_{sub}	Substrate Doping Density
ϕ_F	Fermi Potential
S	Scaling Factor
S'	Voltage Scaling Factor
S_d	Device Vertical Dimension Scale Factor
S_w	Device Width and Wiring Dimension Scale Factor
W_p	pMOS Channel Width

W_n	nMOS Channel Width
I_{DSn}	nMOS Subthreshold Current
I_{DSp}	pMOS Subthreshold Current
C_{load}	Total Load Capacitance
C_{eff}	Effective Switched Capacitance
$I_{DS,on}$	Drain Current with $V_{GS} = V_{THO}$
t_{delay}	Propagation Delay
E_{total}	Total Energy
$E_{dynamic}$	Dynamic Energy
E_{leak}	Leakage Energy
n	Number of Inverter Stages Used

LIST OF FIGURES

Figure No.	Title	Page
2.1	Static (Leakage) and Dynamic (Switching) Power for Different Technology Generations	14
3.1	Comparison of Multiplication factor in Subthreshold and Saturation Regions of Operations. The Threshold Voltage of The Device is 0.33V	41
3.2	Current Multiplication Model	42
3.3	Lateral Electric Field as a Function of Normalized distance from source. L is the Channel Length	46
3.4	Maximum Electric Field as a Function of Effective Length. V_{DS} is the Parameter	47
3.5	Multiplication factor as a Function of Maximum Electric Field	48
3.6	Multiplication factor as a Function of Normalized V_{DS}	48
3.7	Multiplication factor versus Gate Bias for Different Drain Voltages	48
3.8	Multiplication factor as a Function of V_{DS} for V_{GS} of 0.3V and oxide Thickness as parameter	49
3.9	Multiplication factor as a Function of Doping Concentration and Junction x_j is a Parameter	50
3.10	Plot Showing The Comparison Between Subthreshold Current With and Without Impact ionization	51

3.11	Comparison of Calculated Subthreshold Current With ATLAS Simulation Result as a Function of V_{DS} for $V_{GS} = 0.1V$ (Data given in Appendix-1)	51
4.1	Drain-Induced Barrier Lowering (DIBL)	58
4.2	Variation of DIBL-Coefficient with Drain Source Voltage for Different Channel Lengths	65
4.3	Threshold Voltage Shift as a Function of Drain Bias for Different Channel Lengths	65
4.4	Shift-in-threshold Voltage as a Function of effective Channel Length with V_{DS} (1, 1.5, 2, 2.5 and 3V) as the Parameter	66
4.5	Variation of Threshold Voltage Shift with Oxide Thickness for Different doping Concentrations	67
4.6	DIBL-coefficient versus Temperature for Different Channel Lengths for V_{DS} =3V	67
4.7	Lateral Electric Field as a Function of Normalized Channel Position for Selective Scaling Method. L is Channel Length	71
4.8	Maximum Electric Field as a Function of Normalized V_{DS} Different Scaling Method (L =90nm)	71
4.9	Multiplication factor as a Function of Normalized V_{DS} for Three CEFS, GS and SS Scaling scheme	72
4.10	Plot of Subthreshold Current with Impact ionization Versus Normalized V_{DS} for CEFS Scaling Method	73
4.11	Subthreshold Current with Impact ionization Versus Normalized V_{DS} Plot for CEFS, GS and SS Method for S =0.1	73

4.12	Subthreshold Current with Impact ionization Versus Normalized V_{DS} Plot for CEFS, GS and SS Method for S=0.9	74
4.13	Subthreshold Current with Impact ionization Versus Normalized V_{DS} Plot for CEFS, GS and SS Method for S=0.7	74
4.14	Effect of Scaling on Shift-in-threshold Voltage in 70nm Device on CEFS, Gs and SS Techniques	75
5.1	Different power Dissipation Types in CMOS Circuits	79
5.2	CMOS Inverter for Switching Power Calculation	81
5.3	Sizing an Inverter to have a Minimum Delay Gives a Good Estimate of The Optimum Width to Minimize V_{DD} Because the nMOS and pMOS are Balanced	91
5.4	Comparison of Analytical Results with SPICE Results for 100nm Device. Minimum energy occurs at $V_{DD(min)}$=0.16V	93
5.5	Optimum V_{DD}/V_{TH} Biasing Point Trading-off Dynamic Power with Leakage Power for Constant Performance	95
5.6	Energy Versus Supply Voltage Plot with Oxide thickness as Parameter	96
5.7	Effect of Switched Capacitance variation as Optimal V_{DD}/V_{TH} Biasing Point	97
5.8	Energy-Delay product for Various Power Supply Voltage	98
5.9	(a) VTC (b) Noise Margin (High and Low) at the Minimum V_{DD} (Simulation). Optimum pMOS/nMOs width ratio of 3 is Shown with others for Comparison	101

LIST OF TABLES

Table No.	Title	Page
2.1	VTCMOS vs MTCMOS Techniques	27
3.1	MOSFET Parameters	40
4.1	Scaling Rules for Various Scaling Techniques	55
4.2	MOSFET Parameters for DIBL Analysis	64
4.3	Parameters for the Scaled Devices for CEF Scaling Method	70
4.4	Subthreshold Current trends in CEFS, GS, SS at Normalized of V_{DS}=1V After Scaling 100nm MOSFET	74
4.5	Various device and Technology Parameters Obtained After scaling 100nm Device	75
5.1	Parameter values used in Comparing the Subthreshold Current of 100nm MOSFET [KPW02] with SPICE	99

CONTENTS

CANDIDATE'S DECLARATION i

ABSTRACT ii

ACKNOWLEDGEMENTS v

ABBRIVIATIONS vii

LIST OF SYMBOLS ix

LIST OF FIGURES xiii

LIST OF TABLES xvi

CHAPTER 1 INTRODUCTION **1**

1.1 Overview 1

1.2 Designing for Low Power 2

1.3 Motivations and Objectives 6

1.4 Organization of Thesis 7

CHAPTER 2 POWER DISSIPATION in CMOS: AN OVERVIEW **11**

2.1 Introduction 11

2.2 Power Dissipation Components in Digital CMOS Circuits 13

2.2.1 Switching Power 14

2.2.2 Leakage Power 15

2.3 Power and Energy Reduction Techniques 17

2.3.1 Voltage Scaling 18

2.4.2 Circuit Level Techniques 20

2.3.3 Device Level Optimizations 23

2.4 Leakage Reduction Techniques 24

2.4.1 Multi-Threshold COMS (MTCMOS) 25

2.4.2 Variable-Threshold CMOS (VTCMOS) 26

2.4.3 Transistor Stacking 28

2.4.4 Gate Level Leakage Reduction 29

2.5 Ultra-Low Voltage Circuit Techniques 29

2.6 Conclusions 30

CHAPTER 3 EFFECT OF IMPACT-IONIZATION ON SUBMICRON MOSFET SUBTHRESHOLD CURRENT 31

3.1 Introduction 31

3.2 Impact –Ionization in MOS-Channel 32

3.3 Carrier Multiplication in Subthreshold Operation 38

3.3.1 Carrier Multiplication in The Three regions of MOSFET Operation 38

3.3.2 Subthreshold Impact Ionization Current Model 41

3.4 Analysis and Results 45

3.5 Conclusions 52

CHAPTER 4 SUBTHRESHOLD CURRENT OF SUBMICRON AND DEEP SUBMICRON MOSFET 53

4.1 Introduction 53

4.2 Scaling Techniques 54

4.3 Challenges of MOSFET Scaling 57

4.3.1 Short Channel effects 57

4.3.1.1 Length Dependent Threshold Voltage 58

4.3.1.2 Width Dependents threshold voltage 59

4.3.1.3 Reverse Short Channel Effects 59

4.3.2 Subthreshold Leakage Currents 59

4.3.2.1 Gate-Induced Drain Leakage 60

4.3.2.2 Gate Tunneling Current 60

4.3.4 Hot Carrier Effects 60

4.3.5 Dielectric Breakdown 61

4.3.6 Scaling the Interconnects 61

4.4 Controlling Factors of Drain Induced Barrier Lowering Coefficient in Short Channel MOSFET 61

4.4.1 Numerical analysis 62

4.5 Effect of Technology Scaling on MOSFET 68

4.6 Conclusions 76

CHAPTER 5 OPTIMUM SUPPLY VOLTAGE FOR MINIMUM POWER AND ENERGY 77

5.1 Introduction 77

5.2 Power and Energy Management in CMOS Circuits 79

5.2.1 Existing Dynamic Power Estimation Techniques 80

5.2.2 Existing Leakage Power Estimation Techniques 84

5.2.3 Existing Total Power Estimation Techniques 86

5.3 A Simplified Analysis for Minimum Power and Energy 87

5.3.1 Optimum Sizing for Minimum Energy at a Minimum V_{DD} 88

5.3.2 Model Verification 92

5.3.3 V_{DD}-V_{TH} Optimization Overview 93

5.3.4 Role of Switched capacitance on Optimal V_{DD}-V_{TH} Scaling 97

5.3.5 Energy-Delay Product 98

5.3.6 Voltage Transfer Characteristic and Noise Margin 99

5.4 Conclusions 102

CHAPTER 6 CONCLUSIONS **103**

6.1 Summary of Important Findings and Contributions 104

6.2 Suggestions for Future Work 108

APPENDIX-1 110

BIBLIOGRAPHY 122

LIST OF PUBLICATIONS 142

CHAPTER 1

Introduction

1.1 Overview

In recent years, low power circuit operation is becoming an increasingly important metric in portable battery powered devices such as hearing aids [MBSH+99] [KR01], pacemakers, cell phones, pagers, and portable computers become more complex, prevalent, and demand increased battery life. Circuit designers are seeking out new technologies and circuit techniques to maintain high performance and long operational lifetimes. In non-portable applications, reducing power dissipation is also becoming an increasingly important issue. Thus, obtaining the required performance within a limited power budget is the most challenging goal in custom designs and in modern VLSI circuits.

Literature survey shows that sub-threshold logic circuits are preferred due to ultra-low power consumption with moderate performance degradation. Hearing aid devices are clearly one of the most suitable application areas where performance is of secondary importance [SRP00]. In subthreshold logic, circuits are powered by using supply voltage below the threshold voltage. In such circuits the delay increases but ultra-low power can be achieved without major alteration of the circuit. In the subthreshold region, the current through a transistor has an exponential dependency upon gate voltage, threshold voltage and temperature. However the width to length (W/L) ratio has a linear dependency upon the transistor current, so sizing has much less effect on the transistor current than it had in the normal strong inversion region. Therefore the disadvantages of ratioed logic in the strong inversion region such as degradation of noise margin and VTC (voltage transfer characteristic) are diminished in the sub-threshold region.

Keeping all these in mind low power design have become one of the most important design parameters for VLSI (Very Large Scale Integration) systems, which largely rely on complementary metal-oxide-semiconductor (CMOS) technology.

1.2 Designing for Low Power

The power dissipation attributable to the three sources viz.: short-circuit, switching, and leakage, and influences the design at different levels of the overall design process. Since the dominant component of power dissipation in CMOS circuits (due to logic transitions) varies as the square of the supply voltage, significant savings in power dissipation can be obtained by operating the devices at a reduced supply voltage. If the supply voltage is reduced while the threshold voltages stay the same, reduced noise margins result. To improve noise margins,

the threshold voltages need to be made smaller too. However, the subthreshold leakage current increases exponentially when the threshold voltage is reduced. The higher static dissipation may offset the reduction in transitions component of the dissipation. Hence the devices need to be designed to have threshold voltages that maximize the net reduction in the dissipation.

The transitions component of the dissipation also depends on the frequency or the probability of occurrence of the transitions. If a high probability of transitions is assumed and correspondingly low supply and threshold voltages chosen, to reduced the transitions components of the power dissipations and provide acceptable noise margins, respectively, the increase in the static dissipation may be large. As the supply voltage is reduced, the power-delay product of CMOS circuits decreases and the delays increase monotonically. Hence, while it is desirable to use the lowest possible supply voltage, in practice only as low a delay that can be compensated by other means, and steps can be taken to retain the system level throughput at the desired level.

One way of influencing the delay of a CMOS circuit is by changing the channel-width-to-channel-length ratio of the devices in the circuit. The power-delay product for an inverter driving another inverter through interconnects of certain length varies with the width-to-length ratio of the devices. If the interconnect capacitance is not insignificant, the power-delay product initially decreases and then increases when the width-to-length ratio is increased and the supply voltage is reduced to keep the delay constant. Hence, there exists a combination of the supply voltage and the width-to-length ratio that is optimal from the power-delay product consideration.

The way to assure that the system level throughput does not degrade as supply voltage is reduced is by exploiting parallelism and pipelining. Hence as the supply voltage is reduced, the degree of parallelism or the number of stages of pipelining is increased to compensate for the increased delay. But then the latency increases. Overhead control circuitry also has to be added. As such circuitry itself consumes power; there exist a point beyond which power, rather than decreasing, increases. Even so, great reductions in power dissipation by factors as large as 10, have been shown to be obtainable theoretically as well as in practice [CB95].

As more circuit can be accommodated per unit area, off-chip input-output (I/O) power may become the dominant power-consuming function [LS93] unless a significant amount of memory [usually static random-access memory (RAM) and in some cases dynamic RAM] and analog functions are integrated on-chip.

Circuit level choices also impact the power dissipation of CMOS circuits. Usually a number of approaches and topologies are available for implementing various logic and arithmetic functions. Choices have to be made between static versus dynamic style and pass-gate versus normal CMOS for realization of asynchronous circuits.

At the logic level, automatic tools can be used to locally transform the circuit and select realizations for its pieces from a pre-characterized library so as to reduce transitions and parasitic capacitance at circuit nodes and therefore circuit power dissipation [IP95]. At a higher level various structural choices exist for realizing any given logic functions for example for an adder one can select one of ripple-carry, carry-look-ahead, or carry-select realizations.

In synchronous circuits, even when the outputs computed by a block of combinational logic are not required, the block keeps computing its outputs from observed inputs every clock cycle. In order to save power, entire execution units comprising of combinational logic and their state registers can be put in stand-by mode by disabling the block and/or powering down the unit. Special circuitry is required to detect and power-down unused units and then power those up again when they need to be used [Kap94]. The rate of increase in the total amount of memory per chip as well as rate of increase in the memory requirement of new applications has more than kept pace with the rate of reduction in power dissipation per bit of memory. As a result, in spite of the tremendous reductions in power dissipation obtained from each generation of memory to the next, in many applications, the major portion of the instantaneous peak-power dissipation occurs in the memory.

In case of dynamic RAM (DRAM) memory, the most effective way to reduce power of any memory size is to lower the voltage and increase the effective capacitance to maintain sufficient charge in the cell. The new array organizations introduced recently [KR01] present many possibilities of lowering the power.

The literature survey reveals that many advances have been made in various research aspects, including scaling of MOS devices, low-voltage low-power CMOS design style, threshold voltage scaling for high performance designs, leakage current in CMOS circuits, leakage control techniques such as multiple VTCMOS, dynamic VTCMOS, subthreshold logic etc. [CB95] [IF97] [FS93] [LHHC+93] [DDS95a&b] [CSB92] [Mei95] [SRP01] [Sta01]. However, there are still many open research

issues, which need to be solved before low-power CMOS VLSI technology can be viable.

1.3 Motivations and Objectives

As the device is scaled down to submicron level the electric field in the channel region of a MOSFET becomes considerably large. The high electric field in the submicron channel has been found to be responsible for many undesirable effects. One of these effects is impact ionization. The effect of impact ionization on MOSFET operation has been the subject of study of many authors in connection with substrate current. However, the effect of impact ionization on subthreshold current has been overlooked until recently. Park et al. [PDW01] carried out a computer simulation to study the effect of impact ionization on the subthreshold current of a double-gate MOSFET. The results of the simulation show that impact ionization has substantial effect on subthreshold current of such a device. While the double-gate MOSFETs are still under study, the conventional (single gate) MOSFETs find application over a wide range of gate-length. Thus it is of interest to analyze the influence of impact ionization on the subthreshold operation of the conventional MOSFETs.

It is well known that drain induced barrier lowering (DIBL) adversely affect the operation of the submicron MOSFETs. Several workers have studied this effect in the past and currently a parameter, commonly termed as DIBL coefficient is used as a measure of the effect. However, the dependence of this parameter on gate and drain voltages, need further study.

For VLSI applications, MOS technology applies various scaling technologies viz: Ideal or Constant Electric Field Scaling, Generalized Scaling, and Selective Scaling. The different scaling schemes have their own merits and demerits. For subthreshold operation, the effects of these scaling methods need further studies. As the activities in the development of subthreshold logic grow optimization of transistor size and supply voltage for such applications has become important. Such effects should take into account both delay minimization and control of power dissipation.

More succinctly, the following course of action was adopted to achieve the planned objectives as emerging from the motivations discussed above.

- Study of the effect of impact ionization on the subthreshold current in sub-micron MOSFET
- An improved analysis of Drain-Induced-Barrier-Lowering (DIBL) and its effect on subthreshold current. Study of the controlling factors of DIBL coefficient
- Examination of the effects of various scaling techniques on the subthreshold current
- Optimization of transistor size and supply voltage for low-power subthreshold CMOS operation

1.4 Organization of the Thesis

The thesis has been organized as follows: Second chapter discusses the basic models and provides the necessary background to set the pace for going through this work. In addition, it overviews the evolution of minimum power strategies spanned over the last four decades. The survey has, primarily, been confined to minimum power based approaches, which are separately the focus of three

subsequent chapters. Chapter 3 deals with the effect of impact ionization on drain current. Multiplication factor has been determined using well-established expressions [IF97]. The drain-source voltage (V_{DS}) dependence of multiplication factors for subthreshold, linear and saturation regions of operation of a 100nm device are compared. It is found that, subthreshold multiplication factor is highest and linear region has the lowest multiplication factor. The analysis shows that for subthreshold operation it is important that the effect of impact ionization is included in the analytical expression for drain current.

Chapter 4 focuses on subthreshold current of submicron and deep submicron MOSFET. In subthreshold CMOS logic circuit, the input voltage swing is restricted below the threshold voltage of the MOSFETs. Fjeldly et al. [FS93] developed a very useful physical model for subthreshold current. This model yields accurate results for submicron devices, but does not take into account the effect of impact ionization, which is more prominent in devices of nanometer channel length. The subthreshold current model of Fjeldly et al. [FS93] is suitably modified to include the effect of impact ionization for analytical study. The modified current model agrees well with results obtained from ATLAS simulation. It shows that impact ionization increases subthreshold current by few percents. This increase in subthreshold current modifies DIBL–coefficient. It is observed that the DIBL-coefficient has a strong dependence on V_{DS} and relatively lesser dependence on gate-source voltages V_{GS}. The analysis shows that, because of the strong V_{DS} dependence of the DIBL-coefficient, the threshold voltage shift is a non-linear function of V_{DS}, in contradiction to a constant DIBL-coefficient as assumed previously [LHHC+93]. It is also found that DIBL- coefficient increases with increase in channel length and

decreases with increase in temperature. Methods of device scaling in VLSI technology have also been discussed in this chapter. The dependence of subthreshold current on applied scaling technique has been analyzed in this chapter. The results show that as the channel length is scaled down, multiplication factor increases slowly in the higher regime of channel length. In the lower regime of channel length, multiplication factor rises rapidly. This result also justifies the inclusion of impact–ionization effect on subthreshold current. Similar variations in subthreshold current with channel length scaling are shown by the analytical results. However, the analysis shows that there is insignificant dependence of multiplication factor on the method of scaling.

It has also been observed during analysis that subthreshold current also affects DIBL, and is dependant on the scaling technique. Dependence of DIBL is more significant in the higher regime of V_{DS}. Constant electric field scaling produces the largest threshold voltage shift due to DIBL. Selective scaling has the least DIBL, with generalized scaling produced DIBL falling in between the two extremes. As a result, the scaling dependence of subthreshold current follows the same pattern.

Chapter 5 addresses calculation of optimum supply voltage for minimum power and energy requirement. This chapter also deals with the proposed Modified Subthreshold Current model, which is used to study power dissipation in subthreshold CMOS operation and design of low-power CMOS logic. A tradeoff exists between propagation delay and power dissipation, when voltage is scaled. When the supply voltage is scaled down to reduce power dissipation, propagation delay is increased. SPICE simulation, however, shows that this does not hold good for subthreshold logic [Sta01] [PB03]. For a threshold CMOS–logic operation,

propagation delay in most cases first increases and then decreases with supply voltage scaling. The subthreshold current model mentioned above is used to calculate voltage dependence of propagation delay of a CMOS inverter. The results thus obtained show that, the delay increases with increase in supply voltage. An important requirement of a logic circuit is symmetric input-output characteristics. The ratio of pMOS and nMOS widths that meets this requirement is determined using the modified subthreshold current model. The ratio thus obtained is primarily mobility dependent, as in the case of above threshold logic operation.

Chapter 6 summarizes contributions and conclusions. Some future directions have also been suggested here for further exploration.

CHAPTER 2

Power Dissipation in CMOS: An Overview

2.1 Introduction

Low power and low energy issues have captivated VLSI circuit designers for the past few years in the quest for enhancing performance and extending battery lifetime [Pow95]. The increasing demand for integration of more functions with faster speeds is met by an increase in battery. For example, the third generation (3G) wireless protocol provides real-time streaming video at a high data rate on a 3G-enabled cellular phone. Such a computation intensive application can impact the battery life of the portable device. Therefore, the demand for increased battery life will require designers to seek out new technologies and circuit techniques to maintain high performance with longer battery lifetime.

Portable devices, however, are not the sole motive behind the low power and low energy design efforts. The increasing power dissipation for fixed supply devices is almost equally challenging as for portable devices. As technology feature size is reduced, the number of transistors on the chip is increased and more power is dissipated. According to Moore's law, the number of transistors quadruples every two to three years. One hundred billion transistors on a single chip are projected before 2020 [Mei95]. Expensive packing techniques are essential for dissipating such extensive power generated from that large number of transistors. Also, increased power dissipation has a negative impact on device's reliability.

Several methods for power and energy reduction have been proposed in [LS93], [LM93], [KS94], [MDMA+95], [BDAM96], [KFMN+96], [KNC98], [KNBH+99], [BDAM00], [IHS01], [KHK02], [KKK03] and [WCK02]. Voltage supply, V_{DD}, scaling is considered one of the most effective elements for the process of reducing power dissipation in CMOS circuits [FSRS97]. Threshold voltage, V_{TH}, has also to be reduced to maintain the delay but reduction in V_{TH} results in an exponential increase in leakage power. With leakage power under control, the ratio (V_{DD}/V_{TH}) tends to decrease with technology scaling [KC00] [WCK02].

Lowering supply voltage has been proved to save V_{TH} energy. Recently, a fast Fourier transform (FFT) unit was shown to work at 350 mV to provide optimal energy efficiency [WC05]. The FFT unit was also shown to function correctly at a supply voltage of 180 mV. Peak supply voltage is selected based on peak performance requirements. Occasionally, the processing unit does not require peak performance. Therefore, supply voltage can be scaled if some liberty can be taken on the performance front [BP91] [ASPB+97] [NS00]. When supply voltage is scaled,

the average energy of the system is reduced while maintaining the required throughput. As a result, battery lifetime can be extended. A typical example of such an application is the burst-mode operation of the active devices. Such an application has only two modes of operation: active and idle. The device operates at full throughput for a small period of time (active mode) before entering the idle mode (standby). Given a deadline to be met for a given task, reducing supply voltage alone would lead to timing violations. Meeting the deadline requires satisfying the circuit delay constraint. In order to meet the specified deadline at the scaled supply voltage, the threshold voltage of the device, V_{TH} is reduced. However, threshold voltage reduction leads to an exponential increase in subthreshold leakage. Therefore, leakage reduction methods are utilized to reduce leakage power during the idle mode.

2.2 Power Dissipation Components in Digital CMOS Circuits

Power consumption in CMOS circuits can be divided into three main components: short-circuit power, switching power, and leakage power. Short-circuit power arises when a conducting path between supply and ground is formed. Switching power is a result of the power consumed in charging and discharging internal capacitances in the circuit. Leakage power is the power dissipated while the device is turned off. Leakage power has started to form a significant portion of the total power consumption as a result of increase in device density. Figure 2.1 shows the increase in static (leakage) power for different technology generations [ITRS03]. It is apparent that static power is dramatically increasing with technology scaling.

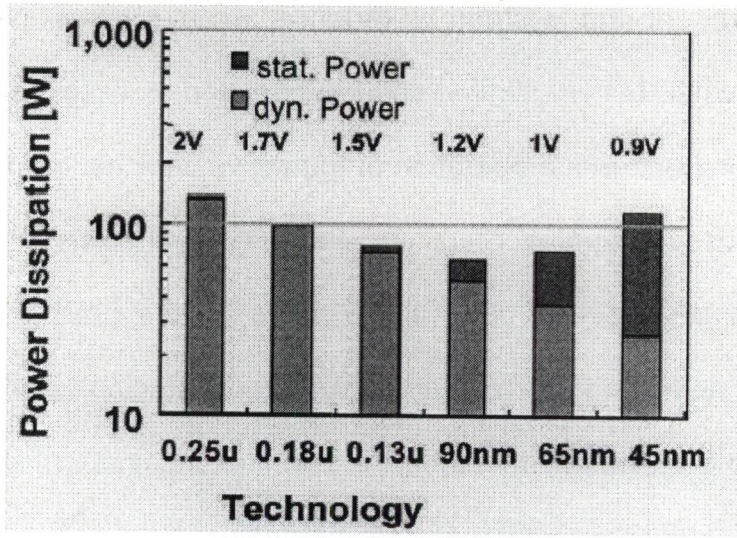


Figure 2.1: Static (Leakage) and Dynamic (Switching) Power for different technology generations

The ratio of leakage to total power is expected to exceed 50% in 45nm designs from about 10% in 90nm designs. Since switching and leakage power are the dominant components of power consumption, they are discussed in detail below.

2.2.1 Switching Power

Switching power is the largest contributor to the total power dissipation in conventional CMOS technologies. It is a result of switching the junction, diffusion, and interconnects capacitances. In any logic gate, if inputs to the gate are assumed to switch at a rate of f times per second, then the average switching power for that gate is given by

$$P_{SW} = \alpha C \Delta V^2 f \tag{2.1}$$

where α is the switching activity factor which represents the probability of the output switching from 0 to 1, C is the switching capacitance, ΔV is the voltage swing, and f is the switching frequency.

Generally, α is less than 1. In case of a logic network of several levels of gates, the activity factor of the gate becomes a function of its inputs probabilities. For certain logic styles, however, glitching can form a non-trivial part of the overall consumption. Glitching often arises when paths with unbalanced proportional delays converge at the same node in the circuit. If glitching due to signal races is to be accounted for, α might be greater than one [CSB92]. Calculations of this activity in a circuit are very difficult and require careful logic and/or circuit level characterization of the gates. Obviously, reducing any term in (2.1) will result in a reduction in switching power. However, low power techniques need to address power reduction without affecting performance or device functionality. For example, frequency reduction is beneficial in terms of power consumption but it affects the overall system speed. Therefore, it is often a challenge to reduce power dissipation while maintaining the system performance.

2.2.2 Leakage Power

The different leakage current components are: (1) the reverse bias p-n junction leakage caused by barrier emission, minority carrier diffusion and band-to-band tunneling; (2)subthreshold conduction current; (3) a current component which results from the drain-induced barrier lowering (DIBL) ; (4) gate induced drain leakage (GIDL); (5) a component due to channel punch through; (6) current owing to hot carrier injection current; (7) oxide leakage current; (8) gate current due to hot carrier injection; The first six are OFF currents are while the last two are ON and switching currents.

- Junction Reverse Bias Current: Junction Reverse Bias Current has two components: One is minority carrier diffusion/drift near the edge of the

depletion region, and the other is due to electron hole pair generation in the depletion region of the reverse biased junction [KNBH+99]. Heavily doped junctions are also prone to Zener and band-to-band tunneling. The p-n reverse bias leakage is a function of junction area and doping concentration. It is normally a minimal contributor to total OFF current.

- Subthreshold Conduction Current: Subthreshold conduction or weak inversion current flows between source and drain when the gate voltage is below the threshold voltage V_{TH}. The subthreshold current occurs due to carrier diffusion when the gate source voltage, V_{GS}, has exceeded the week inversion point, but still below the threshold voltage, where carrier drifts is dominant. Subthreshold conduction typically dominates modern device off-state leakage as they are mostly of low threshold voltages.
- Drain-induced Barrier Lowering, DIBL: DIBL is the effect of lowering the source potential barrier near the channel surface as a result of the applied drain voltage. Ideally, DIBL does not change the subthreshold slope but does lower V_{TH}. Higher surface and channel doping, and shallow source/drain junction depths work to reduce the DIBL mechanism.
- Gate-Induced Drain Leakage, GIDL: GIDL current arises in the high electric field under gate/drain overlap region causing a thinner depletion region of drain to well junction. GIDL results in an increase in leakage current when applying a negative voltage to the gate (NMOS case). GIDL is small for normal supply voltage but its effect rises at higher supply voltages (near burn-in).
- Punchthrough: Punchthrough occurs when source and drain depletion regions approach each other and the gate voltage loses control over the

channel current in the sub gate region. Punchthrough current varies quadratically with drain voltage. Punchthrough is often regarded as a subsurface version of DIBL.

- Hot-carrier injection current: Threshold voltage tends to decrease in trench-isolated small effective channel width devices. The narrow width effect causes the threshold voltage to decrease in trench-isolated technologies for channel widths on the order of W <= 0.5μm. It can be ignored for device sizes >> 0.5μm.

Subthreshold leakage current is the largest leakage current component. It increases exponentially as a result of threshold voltage reduction. Various techniques have been developed to keep both active and leakage power under control. In the next section, some of the effective power and energy reduction methodologies are described. The intent is to focus on these particular methodologies since the work presented in this thesis builds on these methodologies.

2.3 Power and Energy Reduction Techniques

Since switching power is the major source of power dissipation in CMOS technologies, various techniques have been proposed on a variety of design levels to achieve switching power reduction. Keeping power-speed trade off in view some effort can be made at the architecture and circuit level to maintain speed at a scaled voltage. Furthermore, considering a top-down design paradigm, power and energy reduction can be achieved on the architecture, circuit, and device levels. Starting at the top level, the architecture is modified to cater to lower power dissipation and high performance requirements by introducing or adding parallelism or pipelining. When

such modifications are implemented, power can be further reduced via supply or frequency scaling. Moving down the design paradigm, both circuit and device level optimizations are required to enable energy efficient operation.

2.3.1 Voltage Scaling

Many designers have focused on power supply scaling as a means for low power operation. Reducing frequency or switching capacitance provides a linear reduction in switching power. On the other hand supply voltage reduction leads to quadruple savings. Furthermore, subthreshold leakage current can be reduced exponentially with supply voltage reduction. Thus, both dynamic and leakage power can be effectively reduced through supply scaling [Sak03]. Voltage reduction enables architectural level power optimizations. Parallelism or pipelining can be employed to reduce power dissipation [CSB92] [CB95]. For example if a signal of frequency f drives a logic gate at maximum applicable supply voltage, then the gate delay can be reduced by putting a duplicate unit in parallel with the original one in that case the clock frequency can reduced to $\frac{f}{2}$ (doing the computations in parallel). Slashing the operating frequency by half allows for a 40% reduction in the supply voltage, say this reduction may vary from design to design and from one technology to another. Due to the parallelism used, the capacitance increases by a factor of 2 as a result of using a duplicate circuit. In addition, capacitance increases by another 20% (say) due to the extra routing required. Thus, the resulting reduction in power consumption of the parallel architecture compared to the original one is given by

$$P_{parallel} = CfV^2 = (2.2C_{org})(0.6V_{org})2(0.5f_{org})$$

$$= 0.4P_{org} \tag{2.2}$$

where C_{org} is the original effective capacitance being switched per clock cycle. Apparently, the primary limitation of using parallelism to reduce overall power is the area. A considerable part of the extra area required for parallelism is the extra routing area. Wiring capacitance represents a significant part of the total capacitance of a chip. In addition, wiring capacitance does not scale as much as the feature size. Therefore, careful optimization and sophisticated routing techniques have to be utilized to fully exploit the advantage of parallelism and minimize its side effects.

For area constrained designs, pipelining is a viable option with much less area overhead compared to parallelism but yet a comparable throughput. It is a technique that is used to increase the throughput of a sequential set of distinct data input through a synchronous logic cascade. Pipeline architecture divides the cascade into smaller sections, the input to each section being provided by a register. That is the registers are placed between two successive sections. A faster clock is used to initiate activities in all sections at a given time. This way the overall circuit performance is improved. The power reduction achieved by this method can be more than what can be obtained from a parallel structure. Balancing the delay of all the pipelined stages is extremely important to achieve further reduction in power. That would allow for more supply voltage reduction and hence more power savings. In addition, increasing the level of pipelining also reduces the logic depth and hence the power contributed by hazards and critical races. Furthermore, exploiting both pipelining and parallelism is more attractive. This architectural choice results in further speedup and more room for supply voltage reduction.

2.3.2 Circuit Level Techniques

Different static and dynamic logic styles have been introduced for the sole aim of reducing power. It is also a common design practice to combine both static and dynamic logic styles to optimize for delay and power at the same time. The merits of each logic style are explained below.

- Conventional CMOS logic style: Static CMOS logic refers to conventional CMOS circuits, which are constructed using an nMOS pull-down network and a complementary pMOS pull-up network. Due to the complementary nature of the circuit, conventional CMOS logic style is inherently able to reject noise. The layout of CMOS gates is simple and regular due to the similar, yet complementary, pull-up and pull- down network structure. On the other hand, conventional CMOS suffers from inherent disadvantages due to the pull-up pMOS network. One of the main disadvantages is the increased gate capacitance resulting from the large size pMOS transistors. Furthermore, the pMOS transistor is usually made larger to compensate for the speed difference with nMOS due to the lower hole mobility compared to electron mobility. However, this disadvantage is diminishing as technology feature size is shrunk. The carrier drift velocities of both pMOS and nMOS approach the saturation velocity and therefore the size ratio between pMOS and nMOS devices is quickly approaching one. Another drawback of static CMOS logic is the relatively weak driving current. By adding output buffers, the driving current can be enhanced.
- Dynamic Logic Style: Dynamic logic operates in two phases: precharge and evaluation. During the precharge phase, a clock signal charges up the output node. During the evaluation phase, the clock signal switches high. Depending

upon input values, the output node is discharged or remains charged. Dynamic logic is usually faster than static CMOS due to less capacitance (pMOS network is eliminated). However, dynamic logic consumes more power. Many dynamic logic styles with improved delay and power compared to the conventional dynamic style have been reported [SRP00].

- Pass-Transistor Logic Style: Unlike static and dynamic logic, pass-transistor logic provides complementary output. Moreover, inputs are connected to both the gates and the sources of transistors. Pass-transistor gates have two input categories: pass inputs and control inputs. Pass inputs are connected to the sources of the devices while control inputs are connected to the gates. The strongest advantage of pass-transistor implementation is that it can use just one network, usually nMOS, to build the logic. Also, the dual rail nature of the logic style can be used efficiently to implement multiplexing functions. However, connecting some inputs to the source causes a V_{TH} drop. As a result, the voltage swing is reduced and it requires restoration at the output stage to increase noise margin and to minimize short circuit currents. As a consequence, two nMOS networks would be used in addition to the output buffering circuitry. This overhead annihilates the advantage of the low transistor count and small input capacitance. Moreover, pass-transistor logic is sensitive to voltage scaling and transistor sizing. Finally, the layout of pass-transistor logic is complicated due to the extra wiring normally required. One example of pass-transistor logic is the complementary pass-transistor logic (CPL). A CPL has two nMOS networks, one for each rail, and two inverters for level restoration [YY90]. CPL has small input capacitance, a fast differential output stage, and a high driving current. However, CPL, as a

member of the pass-transistor logic family, suffers from short circuit currents at the output and wiring complexity due to the dual rail. Other pass-transistor logic styles have been proposed. A good comparison between the different styles can be found in [ZF97]. In [ZF97], static CMOS has been shown to have superior performance over pass-transistor logic. Therefore, static and dynamic logic usually occupy a larger share of the circuit design space.

From a low power perspective, static logic dissipates less power compared to dynamic logic due to the following reasons:

1. Spurious Transitions: Static designs are prone to spurious transitions more than dynamic circuits due to critical races and dynamic hazards in static logic. The magnitude and the number of those undesirable transitions in a logic structure is a function of the logic design, delay skew, and logic depth. For example, an 8-bit ripple carry adder consumes an extra 30% of power due to spurious transitions [CSB92]. Dynamic logic, however, intrinsically does not suffer from spurious transitions, since any node can undergo at most one power consuming transition per clock cycle.
2. Switching Capacitance: Dynamic logic has fewer devices, typically N+2 for N-input gate compared to 2N in case of CMOS. This is reflected directly on the switching capacitance and thus has a direct impact on delay and power dissipation.
3. Switching Activity: Dynamic logic is notorious for its high switching activity. The dynamic node has to be precharged every clock cycle even if it going to be discharged immediately after evaluation starts. For example, for a 2-input dynamic NOR gate, the switching activity is (3/4) compared to just (3/16) in case of static logic implementation. If spurious transitions are

neglected, then using dynamic logic would result in a 4 times increase in power. But if reduction in capacitance and spurious transitions were taken into account, the resulting power increase would not be that dramatic.

With fewer transistors required to implement a certain dynamic logic function compared to static logic, standby leakage current of dynamic logic can be less than its static logic counterpart. In some applications where fast evaluation time is followed by a long idle period, dynamic logic can be more attractive than static logic for its low standby leakage.

2.3.3 Device Level Optimizations

As mentioned previously, CMOS is regarded as the technology of choice for low power and low energy applications. It offers a good performance and a considerable stability. However, as supply voltage is reduced, threshold voltage has to be reduced to maintain the required performance. A reduced threshold voltage directly results in an exponential increase in subthreshold current.

Some technologies have been offering a solution for the increase in subthreshold current resulting from the reduced threshold voltage. Silicon on insulator (SOI) technology has emerged with a good potential in low power and low voltage applications. In SOI technology, the thin film is totally isolated from the body by a thick film oxide. The thick oxide serves to suppress the radiation-induced current. Also, due to the thick oxide layer, the gate to source/drain capacitance is greatly reduced. As a consequence, SOI devices are faster and consume less dynamic power compared to CMOS. In terms of integration and technology down scaling, the depletion regions in bulk CMOS, which are, used for isolation put a lower

limit on feature size in bulk CMOS. The buried thick oxide in SOI makes it easier to down scale device dimensions.

Two other forms of MOS devices are of importance. These are: DTMOS SOI and DGSOI. The DTMOS is a Dynamic Threshold MOS structure; the gate is tied to the body of the SOI device. This type of connection allows for low threshold during the ON state and high threshold during the OFF state [ASPB+97]. The DGSOI is a Double-Gate SOI device in which there is a back gate separated from the body of the device by the back oxide [WZRC+02]. The DGSOI has a higher current drive for high output load in addition to an excellent ability of leakage control.

However, the history effect of SOI devices, and low thermal conductivity of the buried oxide that results in an increase in temperature are among the drawbacks of using SOI technology. Further development and innovations are required to enable a cost-effective and efficient SOI solution.

In addition to switching power, leakage power is forming an increasing portion of the total power dissipated in modern technologies; several techniques have been developed to reduce its impact. Some of these techniques are summarized in the next section.

2.4 Leakage Reduction Techniques

Modern DSM technologies are suffering from a dramatic increase in leakage current. Constant field scaling dictates that the supply voltage has to be reduced when downsizing the technology feature size. Low threshold voltage devices are used to maintain the required current drive and to satisfy performance specifications. Low threshold devices have caused a dramatic increase in leakage current. A direct and effective solution for that is to utilize low threshold devices in the critical path and

high threshold devices elsewhere. The threshold voltage can be controlled utilizing the well bias of the device in the so-called Variable Threshold CMOS (VTCMOS) [KFMN+96].

Dual threshold technology is another way to address the increasing active and leakage power problem. The technology is a CMOS process with two types of devices, low threshold and high threshold device. Placing the low threshold devices on the critical path to increase performance and place the high threshold devices on the noncritical paths to decrease leakage enhances performance. Several mechanisms have been developed to optimize the process of placing the low/high threshold devices on the gate level or on the transistor level [KC00] [KNC98]. This method was presented in [MDMA+95] and referred to as Multi-Threshold CMOS (MTCMOS). These two methodologies are discussed in detail below. Some recent enhancements and design considerations are also summarized.

2.4.1 Multi-Threshold CMOS (MTCMOS)

The leakage current can be dynamically controlled using multi-threshold devices as was proposed in [MDMA+95]. In this scheme, a low V_{TH} logic is used for faster evaluation while a high V_{TH} nMOS device; the sleep device placed in series with the logic circuit is used to disconnect the logic from the supply during standby. A Sleep control signal is used to turn the high V_{TH} nMOS device ON and OFF depending upon the mode of operation. A clear drawback of this technique is the impact of the Sleep device sizing on performance. Increasing the Sleep transistor size more than necessary would add to the circuit capacitance and power dissipation while sizing it too small would result in a supply current limitation and speed degradation. Another potential problem in the MTCMOS scheme is the bounce of

virtual ground line bouncing. In fact, the capacitance of the virtual ground line is much larger than that of the real ground resulting in a ground bounce. This bounce adversely affects both noise margin and delay. A methodology for properly sizing the Sleep device to minimize the delay based on mutual exclusive discharge patterns was proposed in [KNC98]. The advantage of low leakage during standby mode is stressed by back biasing the sleep transistors to more than V_{DD} [KNS98] [KNS00]. By reverse biasing the body of the sleep transistor, threshold voltage is increased and leakage current is decreased. Therefore, a low threshold voltage device can be used without an increase in leakage current during standby. The low threshold sleep device limits voltage drop during the active mode and provides more current drive. Improving the current drive during the active mode is highly desirable in order to achieve more speed. By increasing the voltage swing of the gate of the sleep transistor, the gate-to-source voltage becomes greater than zero and boosts the current drive [IHS00].

2.4.2 Variable Threshold CMOS (VTCMOS)

VTCMOS technique uses all low threshold devices [KFMN+96]. However, the threshold voltage is controlled using the well bias of the devices in a triple-well CMOS process. During the ON state, the p-well and the n-well are biased for zero body bias for n and p transistors so that low threshold voltages are realized for both. During standby, the source-body junction is strongly reverse biased to increase the threshold voltage and to reduce leakage current. One potential problem with this approach is that the threshold voltage varies as the square root of the body-source voltages. Therefore, the body-source voltage has to significantly increase to change the threshold voltage to a relatively higher value. VTCMOS is even more efficient in

leakage current suppression for series connected transistors due to the increased body-effect [IHS01].

With technology scaling, the body-effect is reduced from one technology generation to the next. The body effect is primarily reduced due to the short channel effects. Techniques such as well doping can be applied to enhance the short channel effects. However, well doping causes the doping levels in the vicinity of source-body and drain-body junctions to increase significantly. As the doping limit approaches the tunneling limit, the junction current increases exponentially, and becomes the dominant leakage component. As a consequence, body-effect is reduced which limits the effectiveness of the VTCMOS scheme [KMNB+01].

SOI technology can also be used in the implementation of VTCMOS. In [Y 97], a silicon-on-insulator-with-active-substrate (SOIAS) was used to dynamically control the threshold voltage. The dynamic threshold MOS (DTMOS) scheme is another means to provide low threshold during the ON state and high threshold during the OFF state [ASPB+97].

A summary of the different features of the MTCMOS and VTCMOS techniques is presented in Table 2.1 [Sak00].

Table 2.1: VTCMOS vs. MTCMOS Techniques

	MTCMOS	**VTCMOS**
Principle	Sleep-mode switch	Well-bias threshold control
Low leakage in standby	Yes	Yes
Products are already rolled out	Yes	Yes
Scalability	Yes	No
V_{TH} fluctuations compensation	No	Yes
I DDQ testing	No	Yes
Serial Sleep Device	Yes slower, lower yield..	No
Sleep-mode storage	Dual V_{TH} FF's	Conventional FF's
Process	Dual threshold	Triple Well or SOI

Moving towards smaller feature size, the MTCMOS technique seems to be a better choice. However, VTCMOS is more effective in reducing process variations, which are increasing with technology scaling. Therefore, the choice between MTCMOS and VTCMOS is application dependent.

2.4.3 Transistor Stacking

Narendra et. al. [NBDA+01] examined the effect of transistor stacking on subthreshold leakage current reduction. It has been shown that the stacking effect increases as the technology scales. Therefore, by forcing transistor stacking speed can be traded for leakage reduction. The term *stack effect* refers to the reduction of subthreshold leakage by stacking multiple FETs in series. The stack effect provides leakage reduction in the same way as switched-source-impedance approach without the resistor. This approach just uses the resistance of the non-ideal switching FET in an off state. The last device in the stack essentially appears as impedance to any leakage current, so the voltage becomes non-zero. The stack effect can reduce leakage current by 2 orders of magnitude for low V_{TH} devices and 3 orders of magnitude for high V_{TH} devices [NBDA+01].

Stacking of transistors during the standby mode can be accomplished by controlling the sleep mode input vector to maximize the number of transistors in the stack during sleep. Such a technique has a negligible speed penalty. However, the minimum leakage state is difficult to achieve by using a specific vector that maximizes the use of stacking since it is not a default feature in all logic gates (e.g. Inverter, NOR, etc.). By combining the use of sleep vector control and forcing stacks in stackless structures, a 30-90% reduction is leakage can be achieved [JSYR02]. In addition, transistor stacking has been shown to effectively reduce gate leakage.

2.4.4 Gate Level Leakage Reduction

Wei et. al. [WH99] proposed a gate level optimization method for leakage reduction. In their work, gates are divided into groups, one is low threshold and the other is high threshold. Gates in the critical path are low V_{TH} for faster evaluation while non-critical path gates are high V_{TH} to reduce leakage. The optimization method is run iteratively to find the optimum gate assignment for the given V_{TH} value. Leakage reduction through the use of multiple supply voltages can be also achieved. The normal supply voltage is assigned to the gates on the critical path while reduced voltages are applied to gates not on the critical path. An earlier work was proposed in [LS93] where optimizations of supply and threshold voltages are performed to achieve low power implementations.

2.5 Ultra - Low Voltage Circuit Techniques

Voltage supply, V_{DD}, reduction has been utilized to achieve low power, energy efficient operation due to the quadratic relationship between power and V_{DD}. The transistor's threshold voltage, V_{TH}, is often reduced to maintain a decent performance. In a limiting case, fully static CMOS logic works when V_{DD} is slightly greater than max $\{|V_{THp}|, V_{THn}\}$ where V_{THp} and V_{THn} are the threshold voltages of the pMOS and nMOS transistors respectively. However, when V_{DD} is reduced below that value, the switching delay increases appreciably.

Reducing supply voltage is often accompanied by threshold voltage reduction in order to prevent current drive degradation and the resulting delay increase. Assaderaghi et. al. [ASPB+97] introduced the concept of dynamically controlling the threshold voltage (DTMOS) by connecting the gate of the MOS

transistor to its substrate in silicon on insulator (SOI) technology. Threshold voltage is reduced during the active mode and is restored back to the normal value during the standby mode. This results in a significant speedup during the active mode and normal standby leakage current. However, an exponential increase in active mode leakage current is observed due to threshold voltage reduction. A potential remedy to this increase in leakage current is to use one of the leakage reduction techniques described earlier.

2.6 Conclusions

Designing for power and energy efficient designs has become a necessity for modern VLSI technologies. With doubling integration capacity every two to three years, power dissipation presents a real threat for reliability and even functionality of the devices. As a result, tremendous effort has been devoted to achieve lower power dissipation without affecting performance. The main components of power dissipation are switching power and leakage power. Switching power, being the dominant power component, has caught special attention in recent years. Many techniques have been introduced to control this ever-increasing power component on all levels of design abstraction. Increased leakage current due to technology feature downsizing is another challenge that faces circuit designers in the deep sub-micron era. System, circuit, and device levels are all examined for potential solutions for overall power reduction.

CHAPTER 3

Effect of Impact-Ionization on Submicron MOSFET Subthreshold Current

3.1 Introduction

Metal-oxide-semiconductor field–effect transistor operates in sub-threshold region when applied gate voltage is below the threshold voltage *i.e. conducting channel is weakly inverted.* Normally this is considered to correspond to the OFF state of the MOS device. Actually in subthreshold operation, there exists a leakage current due to diffusion of carriers between source and drain. Naturally this current leads to power dissipation, however small it may be. In the VLSI circuits a large number of MOSFETs are often simultaneously biased in the subthreshold region of operation. In such a situation, the subthreshold currents due to the individual devices add up to a source of substantial power dissipation. This power dissipation is highly undesirable, as it causes many reliability problems. In submicron devices the

problem is further aggravated by the high electric field existing in the channel region. Under the influence of high electric field the carriers multiply by impact-ionization. The result is an increased drain current. This chapter deals with impact–ionization in submicron MOSFET channel and it's effect on subthreshold current.

3.2 Impact-Ionization in MOS-Channel

As the feature size of integrated MOS devices decreases continuously, the high electric field near the drain region is becoming more crucial and is imposing limits on device operations. The limiting factors are large gate current, substrate current, and substantial threshold voltage shift and drain breakdown. Hot electron generation caused by the high field near the drain is also responsible for effects like impact ionization and gate leakage current. Another crucial parameter for these effects is the length of the velocity saturation region.

Impact ionization is the process of electron-hole pair creation through the breaking of a lattice bond by a charged carrier whose kinetic energy exceeds the threshold for bond breaking. This threshold is called the ionization threshold and is comparable to the band gap energy of the semiconductor. In the case of impact ionization by electrons, the impact-ionizing electron loses most of its energy by interacting with a valence band electron. The ionizing carriers usually gain their energy from the electric field. The high field region near the drain of an n-channel MOSFET is one such cause, leading to the generation of electron-hole pairs by electron impact ionization. The biasing of a typical n-MOSFET in normal modes of operation causes the generated electrons to be drawn into the drain terminal, while the holes are collected at the substrate contact, in the form of the substrate current.

The earlier impact-ionization investigations approached the phenomenon analytically, describing the probability of impact ionization as a function of the local electric field. In Shockley's lucky electron model [Sho61] the impact ionization rate is proportional to the probability of an electron gaining the threshold energy E_{th} from the electric field E after traveling a distance $l = E_{th} = qE$ without collision, i.e.

$$P_{ii} \sim \exp\left(-\frac{E_{th}}{qE\lambda}\right) \tag{3.1}$$

where λ is the mean free path. An even earlier result belonged to Wolff [Wol54], who investigated the high field regime above 200 kV/cm, assuming a constant electric field distribution, in space (not constant in time, as is customary for device simulation). Wolff Investigated impact ionization from the point of view of energy diffusion, and his approach yielded an ionization coefficient (rate of electron-hole pair generation per unit distance) with an $\exp(C/E^2)$ dependence on the electric field E, where C is a constant. Although physically sound, Wolff's model could not account for a lot of the data taken in small semiconductor devices. This situation was clarified by Chynoweth [Chy58] and later by Baraff [Bar62] who showed that for lower electric fields, when $qE\lambda \leq \hbar\omega$ (the optical phonon energy), the ionization rate is better represented by a dependence like Shockley's, proportional to $\exp(C/E)$.

Keldysh [Kel65] extended the above treatments to finite temperatures. Although originally developed only for direct gap and parabolic band materials, his model has been extensively used in applications ever since. A more rigorous treatment of impact ionization was adopted by Kane [Kan67], which took into account the effects of band structure. In his work he numerically determined the ionization rate using Fermi Golden Rule calculations, including a realistic band

structure and a momentum-dependent dielectric function. Bude et al. [BHI92] [BH92] refined Kane's work by including collision broadening and intra-collisional field effects. Instead of Fermi's Golden Rule, they applied a quantum transport approach (e.g. using density matrix formalism) combined with Monte Carlo methods to treat the impact ionization process. His work showed that the already "soft" threshold of the ionization process- as indicated in Kane's work and also suggested by other workers (e.g. Woods [Woo87]) - is considerably broadened by electron-phonon collisions, so that a well-defined threshold does not really exist.

In short channel nMOSFETs, the lateral electric field is very high. The high electric field in the submicron channel has been found to be responsible for impact ionization. The effects of impact ionization on MOSFET operation have been the subjects of study of many authors. The experimental and theoretical studies in linear and saturation region shows that ionization rate in the high-field region near the drain are gate-length and oxide thickness dependent [MB95] [MTIT+92] [BMTN+95] [DRJ98] [MNOY+00] [KTK76].

It is also observed that in submicron devices, for low drain voltages, impact ionization rates are more controlled by lattice temperature than electric field [HNYJ+99] [SGSH02] [AME03]. This is attributed to the decrease in critical energy for impact ionization with increase in temperature. Explanation of all these observations needs suitable impact – ionization model. In the submicron devices, electrons have an effective temperature that is much higher than the lattice temperature. Carrier temperature gradient leads to thermal diffusion currents. This can be accommodated by consistently solving the energy balance equations along with the drift-diffusion equations and the Poisson equation. Energy balance

equations include energy lost through phonon emission. The impact ionization generated leakage currents, can be modeled as functions of carrier temperature [Mei88]. In these models energy conservation equations accommodate non-local nature of carrier heating. However, the details of the band structure are not taken into consideration. Further, this model assumes that the electron energy distribution (EED) can be represented by a Maxwell-Boltzmann distribution (MBD) with a single effective temperature. Since the EED is represented by a MBD with a single effective temperature, it will extend much beyond the applied V_{DS}. This is considered an inherent shortcoming of the energy balance formulations.

State-of-the-art approach to carrier transport in semiconductor devices is to solve Boltzman Transport Equation (BTE) using Monte-Carlo (MC) techniques [SH81]. In MC, trajectories of one or more carriers is traced at the microscopic level in a given device structure. This is subject to the action of electric fields, device geometry and various scattering mechanisms. Carrier flight and scattering are stochastically treated in accordance with probabilities describing the physical processes. Consistent MC solutions also solve the Poisson equation. The outputs of MC are carrier energy distributions. Since the carriers are treated as particles, and BTE is a classical formulation, MC is essentially a classical approach to simulation. However, the band structure models used by various MC groups are different. For example, Fischetti et al. [FL95] and Bude et al. [BPS00] employed full-band structure using pseudo potential method. In some of the analyses, the probabilities describing scattering mechanisms are determined empirically to fit experimental observations. For example, Cartier et al. determined impact ionization rate as a function of carrier energy by fitting empirical models to soft X-ray photoemission spectroscopy data [CFEM93]. Bude et al. [BM95] used theoretically evaluated impact ionization rates.

Ghetti et al. [GSB99] used the impact ionization rates from Bude et al. However, the band structures used by both the groups are different and Ghetti et al. have fine-tuned deformation potentials to match experimental quantum yield data. Since MC is a stochastic approach, the number of carriers to be simulated must be sufficiently high to obtain reliable statistics. This problem is prohibitively serious in MOSFETs operating in the subthreshold regime, due to the low number of inversion layer carriers, and no reliable results can be obtained for this regime. The discussions above show that the exact calculation of the ionization rate is extremely difficult. The local field approximations are relatively simpler though. Such approximations exist for two separate cases, one for very high electric fields (Wolf's model) and one for lower values of the electric field (Shockley's model). However, the channel electric field near the drain region in MOS transistor is neither high enough to validate Wolf's model nor low enough to validate Shockley's model. Okuto and Crowell [OC72] proposed a model that is accurate in both high and low electric field regions. Unfortunately, since the evaluation of substrate current or drain avalanche break-down generally involves the integration of the ionization rate over the velocity saturation region [CKH85], the mathematical form of Okuto and Crowell approximation [OC72] is too complicated to be used in analytical calculation. Hence the Shockley's model is still widely used in modeling the hot electron generation and impact ionization in MOSFET's. According to Shockley's model, the impact ionization coefficient, α_i, is determined by the local electric field, E, and is given by

$$\alpha_i = A_i \exp\left(\frac{-B_i}{E}\right) \tag{3.2}$$

where A_i is a proportionality constant with a dimension of inverse distance and B_i is a characteristic electric field. As mentioned before, expression (3.2) is an approximation for the low electric case; it does not consider the energy distribution function of carriers. Hence, it requires the fitting parameter A_i and B_i to be modified for different processes and different biasing conditions.

As can be seen from (3.2), for assessment of carrier multiplication by impact-ionization in a MOS-channel, the two most important parameters are the length of the ionization region and the electric field in it. The electric field and length are both dependent on the device operating conditions. It is well known that a MOSFET has three distinct operating conditions, namely, subthreshold, linear and saturation. The magnitude profiles of the channel electric fields are different for these operating conditions. An effective means of estimating carrier multiplication by impact ionization is careful observation of substrate current. This is because the substrate current is primarily an impact – ionization generated current. For example, in an n-channel MOSFET, the electrons constituting the drain current may acquire sufficient energy to cause electron-hole pair generation by impact-ionization. The secondary electrons thus generated are collected by the drain along with the primary channel electrons. The substrate terminal, on the other hand collects the holes. As the result the current, which flows into the substrate, is the substrate current. Since the origin of this current is impact-ionization by electrons, it can be well described by an electron multiplication model.

The impact-ionization substrate current has been extensively studied by several authors [KTK76], [SS80], [MTIT+92], [BMTN+95], [PJYP+96], [IF97], [DRJ98], [HNYJ+99], [LYCZ+00], [MNOY+00], [SGSH02], and [AME03]. The

operating region of the MOS characteristics limits most of the substrate current models. In order to eliminate the transition region discontinuities arising out of the use of different models, Iniguez and Fjeldly [IF97] developed a unified subthreshold current model applicable to all the three operating regions of a MOSFET. In this model the effect of impact – ionization is accounted for by a multiplication factor, which by itself represents a unified impact – ionization model, applicable to all three MOSFET operating regions.

3.3 Carrier Multiplication in Subthreshold Operation

The scaled channel length and increased channel doping of the submicron MOSFETs give rise to high channel electric field even at scaled voltages. The high fields of these devices give rise to considerable impact ionization and increase subthreshold current. The controlling effects of various parameters on subthreshold impact-ionization and current will now be considered.

3.3.1 Carrier Multiplication in the Three Regions of MOSFET Operation

As MOS-channel is scaled, the state of arts device density in VLSI chip is increased. The result is an increase in standby power dissipation due to the increased number of transistors, which lie idle from time to time, during the operation of a chip. As standby power dissipation results from subthreshold conduction, the subthreshold current is of extreme importance for low-power design. However, the effect of impact ionization on subthreshold current has been overlooked until recently. Park et al. [PDW01] carried out a computer simulation to study the effect of impact ionization on the subthreshold current of a double-gate

MOSFET. The results of the simulation show that impact ionization has substantial effect on subthreshold current of such a device. While the double-gate MOSFETs are proposed for deep submicron technology and the technology is not developed yet, the conventional (single gate) MOSFETs find application over a wide range of gate-length. Thus it is of interest to analyze the influence of impact ionization on the subthreshold operation of the conventional MOSFETs.

The generalized expression for electron multiplication factor of an n-channel MOSFET as proposed by Iniguez and Fjeldly [IF97] is

$$M = 1 + \frac{A_n}{B_n}(V_{DS} - V_{DSC})\exp\left(\frac{-B_n l_d}{V_{DS} - V_{DSC}}\right) \tag{3.3}$$

where A_n and B_n is the impact ionization rate parameters, $V_{DSC} = V_{sat}$, the saturation voltage. In linear region $V_{DSC} = V_{DS}$, the drain-source voltage, and $V_{DSC} = 0$ in weak inversion. Characteristic channel length in velocity saturation is given by [CGMK+88]

$$l_d = 1.7 * 10^{-2} t_{ox}^{1/8} x_j^{1/3} L_{eff}^{1/5} \text{ m} \tag{3.4}$$

where x_j is the junction depth, t_{ox} is the oxide thickness and L_{eff} is the effective gate length and can be calculated by $L_{eff} = L - x_s - x_d$. Thus, by substituting the appropriate value for the critical voltage parameter V_{DSC} one can find the expression for the impact-ionization multiplication factor for the concerned region of operation of the MOSFET. The three expressions for the linear, saturation and subthreshold operation are:

$$M_{sat} = 1 + \frac{A_n}{B_n}(V_{DS} - V_{sat})\exp\left(\frac{-B_n l_d}{V_{DS} - V_{sat}}\right) \tag{3.5}$$

$$M_{lin} = 1 \tag{3.6}$$

$$M_{sub} = 1 + \frac{A_n}{B_n}(V_{DS})\exp\left(\frac{-B_n l_d}{V_{DS}}\right) \tag{3.7}$$

It is clear from (3.6) that in the linear operation there is no significant effect of impact–ionization on the drain current of a MOSFET. To compare the influence of impact-ionization on drain current, when the operations are in the saturation and subthreshold regions, the case of a MOSFET is taken, the device parameters of which are given in Table 3.1[KPW02].

Table 3.1 MOSFET Parameters

Parameters	Values
ε_s	1.04×10^{-10}
ε_i	3.36×10^{-11}
N_a	1.5×10^{24} m^{-3}
N_C	3.0×10^{26} m^{-3}
t_{ox}	2.0 nm
V_{TH0}	0.3305 V
μ_0	0.0134 m^2/V-s
η	1.21
x_j	11.1×10^{-8} m

Analytical results in Figure 3.1 shows that carrier multiplication in subthreshold region of operation of MOSFET is higher then the saturation region. It is approximately 1.6% more for subthreshold region in comparison with saturation region and unity in linear region of operation.

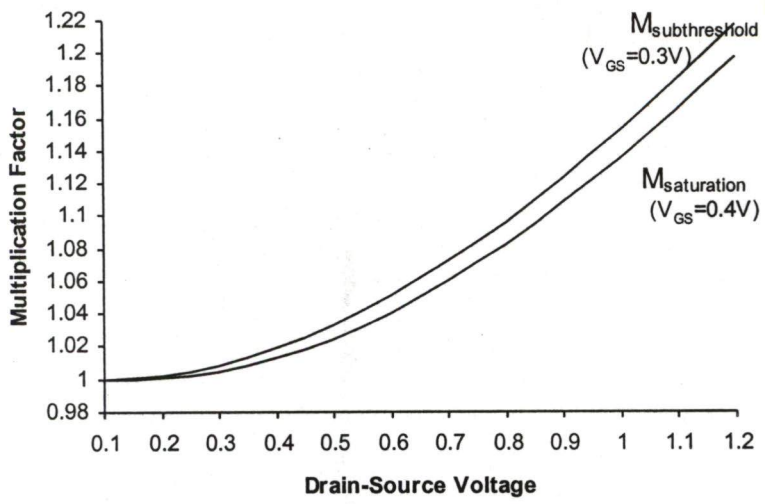


Figure 3.1: Comparison of Multiplication Factor in subthreshold and saturation regions of operation.
The threshold voltage of the device is 0.33V

3.3.2 Sub-threshold Impact-Ionization Current Model

Fjeldly et al. [FS93] proposed a reliable analytical model for subthreshold conduction. This model is applicable to submicron devices, as it takes into consideration the short channel effect on subthreshold current. However, the effect of impact ionization was assumed to be negligible. This is a good approximation when the electric field in the channel is not substantially large. As the channel length is reduced, electric field in the channel increases and for large drain voltages, impact ionization in the channel region cannot be ignored. This is true for all operating regions of MOSFET. To incorporate the effect of impact-ionization in the model of Fjeldly et al., the active region of the device is assumed to consist of two parts, of which one is the impact-ionization region. As shown in Figure 3.2, multiplication occurs in the region of length l_d, where the field is at it's maximum value of E_m.

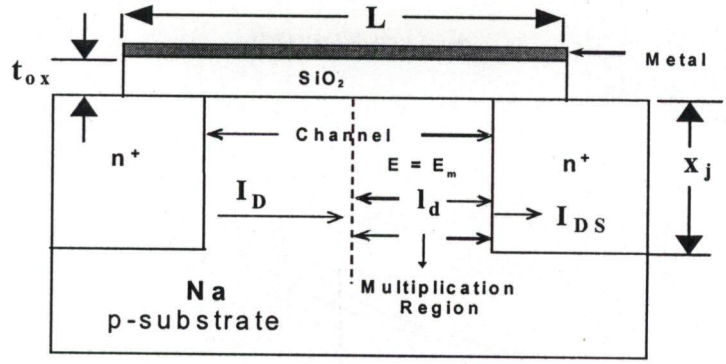


Figure 3.2 Current Multiplication Model

If M is the current multiplication factor due to impact ionization in channel of a MOSFET, then the drain current I_{DS} is

$$I_{DS} = MI_D \tag{3.8}$$

where I_D is the channel current without the effect of impact ionization. When the device operates in subthreshold region, I_D have two components: drift-diffusion current I_{dd} and thermionic emissions current I_{te}. These are respectively given by [FS93],

$$I_{dd} = \frac{qN_C W\delta D_n V_0}{\lambda V_T} \exp\left(\frac{\psi_s^0 + V(x_s)}{V_T}\right) \frac{1-\exp\left(\frac{-V_{DS}}{V_T}\right)}{1-\exp\left(\frac{-V_0(L-x_s-x_d)}{V_T}\right)} \tag{3.9}$$

$$\text{and, } I_{te} = I_0 \exp\left(\frac{\psi_s^0 + V(x_s)}{V_T}\right)\left(1-\exp\left(\frac{-V_{DS}}{V_T}\right)\right) \tag{3.10}$$

where $I_0 = W\delta A^* T^2$, A^* is the Richardson's constant, W is the width of the gate, δ is the effective channel thickness.

Various parameters in (3.9) and (3.10) are given by

$$\left.\begin{aligned}
&\delta = V_T\sqrt{\frac{\varepsilon_s}{2qN_a(2\varphi_b + V_{GT}/\eta)}} \\
&\eta = 1 + c_{dep}/c_{ox} \\
&x_s = \sqrt{\frac{2\varepsilon_s}{qN_a}\left(-\psi_s^o\right)} \\
&x_d = \sqrt{\frac{2\varepsilon_s}{qN_a}\left(V_{DS} - \psi_s^0\right)} \\
&d_{dep}^0 = \sqrt{\frac{2\varepsilon_s}{qN_a}\left(\psi_s^0 + V_{bi}\right)} \\
&\lambda = d_{dep}^0\left(1 + \frac{\varepsilon_i\, d_{dep}^0}{\varepsilon_s\, t_{ox}}\right)^{-1/2} \\
&V(x_s) = V_0 \sinh(\frac{x_s}{\lambda})
\end{aligned}\right\} \qquad (3.11)$$

where $V_0 = 2\chi V_{DS}\,\frac{d_{dep}^0}{\lambda}[\cosh(\frac{L - x_d}{\lambda}) - \cosh(\frac{x_s}{\lambda})]^{-1}$

and, $\psi_s^0 = -V_{bi} + 2\varphi_b + V_{GT}/\eta$

where η is the subthreshold ideality factor, χ is a constant (on the order of 0.5) which accounts for the charge sharing in depletion region, L is the gate length, x_s is the source-channel depletion width, x_d is the drain-channel depletion width, d_{dep}^0 is the depletion width at zero drain-source bias and λ is a parameter that accounts for the effect of gate oxide thickness on depletion layer width, c_{ox} is the gate oxide capacitance, c_{dep} is the depletion layer capacitance, bulk potential $\varphi_b = V_T \ln(N_a/n_i)$, where n_i is the intrinsic carrier density, diffusion coefficient $D_n = \mu_n V_T$, where

$V_T = kT/q$ is the thermal voltage, ψ_s^0 is the constant potential at the semiconductor-insulator interface at zero drain bias, N_a is the acceptor doping density in the substrate, ε_s is the dielectric permittivity of the semiconductor, ε_i is the electrical permittivity of the insulator, V_{bi} is built-in-potential, t_{ox} is the thickness of the gate oxide and $V(x_s)$ is the amount of bending from equilibrium potential across the channel region due to applied drain bias. For gate-source voltage of V_{GS} and zero drain-source bias threshold voltage V_{TH0}, the parameter V_{GT} is

$$V_{GT} = V_{GS} - (V_{TH0} - \sigma V_{DS}) \tag{3.12}$$

where σ is the drain-induced barrier lowering (DIBL) coefficient. The two components of currents given by (3.9) and (3.10) combine to give the total drain current [FS93] as

$$I_D \approx \left[\frac{1}{I_{dd}} + \frac{1}{I_{te}}\right]^{-1} \tag{3.13}$$

if impact-ionization is of negligible consequence. Substituting (3.9) and (3.10) in (3.13)

$$I_D \approx \left[\frac{1}{\frac{qN_C W\delta D_n V_0}{\lambda V_T}\exp\left(\frac{\psi_s^0 + V(x_s)}{V_T}\right)\frac{1-\exp\left(\frac{-V_{DS}}{V_T}\right)}{1-\exp\left(\frac{-V_0(L-x_s-x_d)}{V_T}\right)}} + \frac{1}{I_0\exp\left(\frac{\psi_s^0 + V(x_s)}{V_T}\right)\left(1-\exp\left(\frac{-V_{DS}}{V_T}\right)\right)}\right]^{-1} \tag{3.14}$$

For the case shown in Figure 3.2, for subthreshold operation the multiplication factor is of the form of (3.7). A consideration of (3.4) and (3.7) together shows that the multiplication factor is depends on the maximum electric field. The maximum electric

field E_m depends on drain-source voltage V_{DS} and si/sio$_2$ interface potential ψ_s^0. The expression for E_m is [IF96]

$$E_m = \sqrt{\frac{V_{DS}^2}{l_d^2} + \frac{2E_a V_{DS}}{l_d}} \tag{3.15}$$

where $E_a = \frac{qN_a l_d}{\varepsilon_s} - \frac{\left(V_{GB} - V_{FB} - \psi_s^0\right)}{l_d}$ (3.16)

Thus, if multiplication is significant, then combining (3.8), (3.9), (3.10), (3.13) and (3.14) the subthreshold current is of the form

$$I_{DS} = \left(1 + \frac{A_n}{B_n}(V_{DS})\exp\left(\frac{-B_n l_d}{V_{DS}}\right)\right)\left(qN_C W \delta A^* T^2 \exp\left(\frac{\psi_s^0 + V(x_s)}{V_T}\right)1 - \exp\left(\frac{-V_{DS}}{V_T}\right)\right) * \left(1 + \frac{A^* T^2 \lambda V_T \left(1 - \exp\left(\frac{-V_0(L - x_s - x_d)}{V_T}\right)\right)}{qN_C D_n V_0}\right)^{-1} \tag{3.17}$$

3.4 Analysis and Results

It has been shown [HK 92] that in a weakly inverted channel the potential distribution $V(x)$ and lateral electric field $E(x)$ are

$$V(x) = E_a l_d \left(\cosh\left(\frac{x}{l_d}\right) - 1\right) \tag{3.18}$$

$$E(x) = E_a \sinh\left(\frac{x}{l_d}\right) = \left[\frac{qN_{sub} l_d}{\varepsilon_s} - \frac{\left(V_{gb} - V_{FB} - \psi_0\right)}{l_d}\right]\sinh\left(\frac{x}{l_d}\right) \tag{3.19}$$

where $\psi_0 = 2\phi_F + \left(\frac{V_{GS} - V_{TH0}}{\eta}\right) + V_{FB}$ is the minimum of the surface potential relative to the bulk [FS93], N_{sub} is the substrate doping density, V_{FB} is the flat-band voltage, V_{GS} is the gate to-source voltage, V_{GB} is the gate-to-bulk voltage, V_{TH0} is the threshold voltage, η is the sub-threshold ideality factor, $\phi_F = V_{TH} \log\left(\frac{N_{sub}}{n_i}\right)$ is the Fermi potential, V_{TH} is the thermal voltage and x is the distance from source.

Figure 3.3 shows the lateral field distribution along the channel for two channel lengths, $L = 90nm$ and $L = 100nm$, using above equations. It also illustrates the effect of scaling on channel electric field. It can be seen that as channel is shorten field becomes high. Interestingly, the maximum electric field E_m, at the drain end is larger for longer channel. The crossover very close to drain shows that the longer device field is lower in most of the channel. That is the rate of variation of field with distance from source is larger in a longer device.

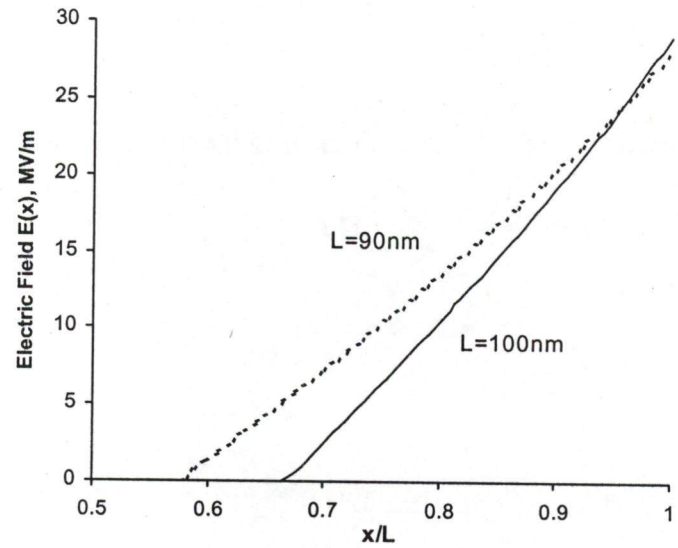


Figure 3.3: Lateral Electric Field as a function of normalized distance from source. L is channel Length, $V_D = 1V$

It can be easily seen from (3.14) that the Maximum electric field E_m greatly influences the carrier multiplication process. This field in turn depends on the important parameters of effective length L_{eff}, and drain-source voltage V_{DS}, as may be seen in (3.15) and (3.4). L_{eff} can be calculated by equation $L_{eff} = L - x_s - x_d$ [IF93]. Figure 3.4 illustrates the dependence of E_m on L_{eff} and V_{DS} for the device in consideration. The fall in E_m with increase in L_{eff} and decrease in V_{DS} indicates the influence of channel width modulation on the field. Increased channel length modulation increases the maximum electric field. The crossover in Figure 3.3 actually refers to drawn gate lengths.

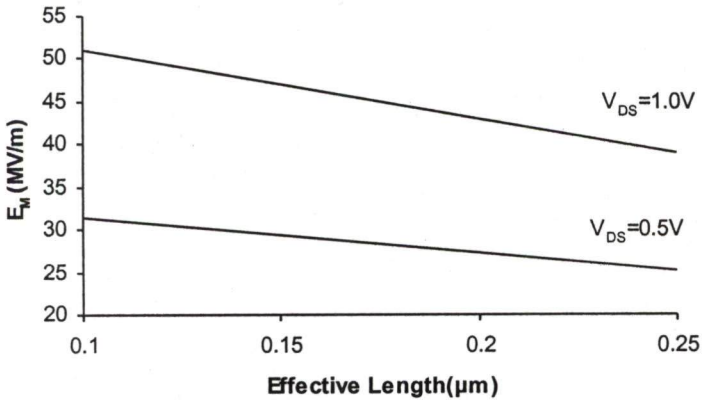


Figure 3.4: Maximum Electric Field as function of effective length. V_{DS} is the parameter

For the same device parameters multiplication factor M, is calculated using (3.14) through (3.16) for different V_{DS}. Figure 3.5, shows the dependence of M on the maximum electric field E_m. It shows that in the lower regime of E_m, the multiplication factor does not vary significantly with change in field. However, in the higher regime there is an exponential rise in multiplication factor with increase in electric field. As discussed previously, E_m, depends on V_{DS} and as V_{DS} increases E_m increases.

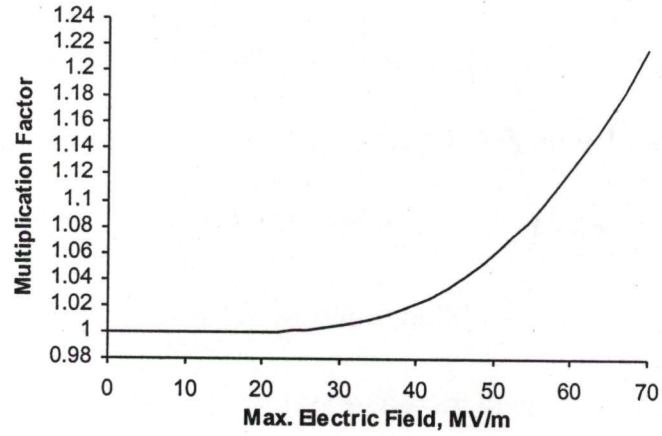


Figure 3.5: Multiplication Factor as a function of Maximum Electric Field

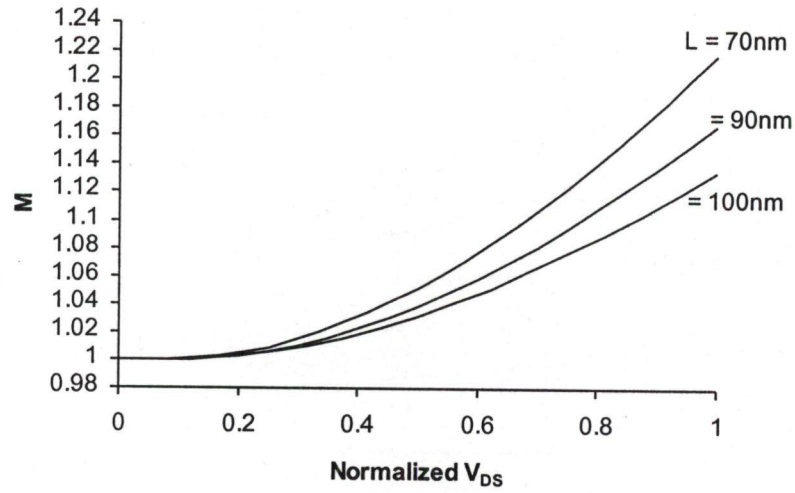


Figure 3.6: Multiplication factor as a function of normalized V_{DS}

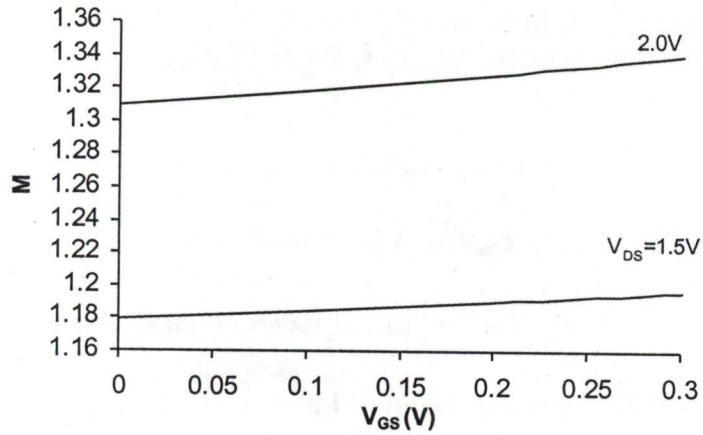


Figure 3.7: Multiplication Factor versus gate bias for different drain voltages

This results in corresponding increase in multiplication factor with increase in V_{DS} (Figures 3.6 and 3.7).

Careful examination of (3.14), (3.15) and (3.16) shows that multiplication factor has some dependence on gate-source voltage. This is illustrated in Figure 3.7, where multiplication factor M, is plotted as function of V_{GS}. The increase in M with V_{GS} shows that contribution to current from impact ionization is increased if gate voltage is increased. As may be seen from (3.4) and (3.15)-(3.16), the multiplication factor given by (3.14) is dependent on gate-oxide thickness through the parameters of l_d and E_m. The dependence of M on the gate oxide thickness can be seen from Figure 3.8. It shows that for lower V_{DS} the dependence on oxide thickness is insignificant. However, it becomes prominent as V_{DS} is increased. An important conclusion that can be drawn from the results presented in Figure 3.8 is that, undesirable increase in drain current owing to impact-ionization can be checked by the use of thicker gate oxide.

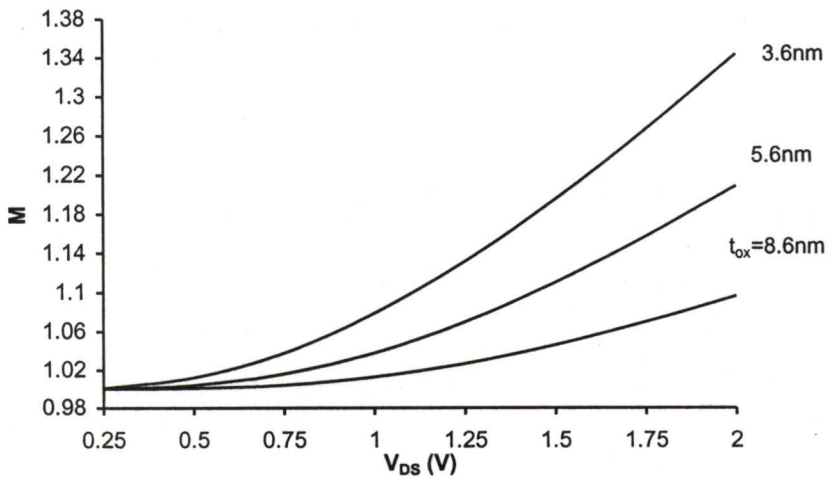


Figure 3.8: Multiplication factor as function of V_{DS} for V_{GS} of 0.3 V and oxide thickness as parameter

The junction depth x_j is another parameter that influences the saturation region length l_d (equation (3.4). Thus, x_j affects the multiplication factor M as may be seen in (3.14)-(3.16). Figure 3.9, shows the effect of junction depth on M. It can be seen that M increases with increase in x_j. This is because of the increase in the size of the region where impact-ionization takes place. The strong dependence of multiplication factor on maximum electric field makes it dependent on the channel doping. This may be verified from (3.4)-(3.7) and (3.14)-(3.17). The plots in Figure 3.9 are obtained from these equations. The increase in Multiplication factor with increase in doping concentration is the result of corresponding increase in maximum field.

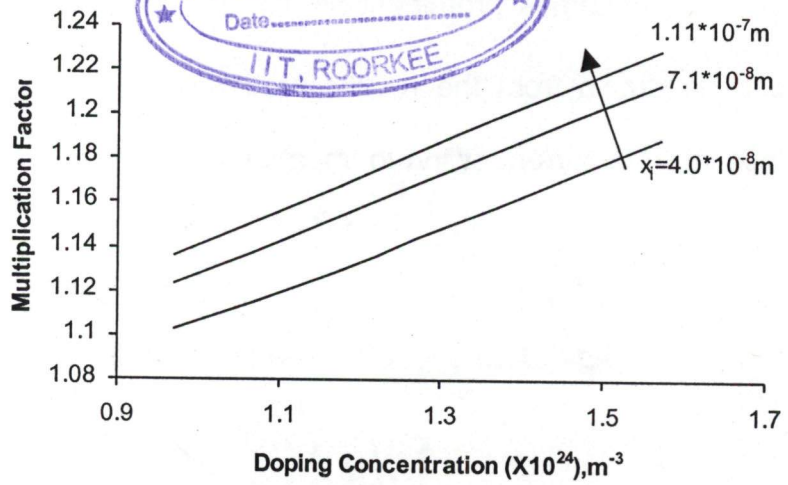


Figure 3.9: Multiplication Factor as a function of doping concentration and junction depth x_j is a parameter

Subthreshold current, excluding the effect of impact ionization is calculated from (3.14). The dependence of this current I_D on drain voltage is shown in Figure 3.10. The effect of impact ionization can be estimated, when the subthreshold current I_{DS} is calculated by (3.17). A plot of I_{DS} is also shown in Figure 3.10. A comparison of the two plots shows that, the effect of impact ionization becomes substantial at larger

drain voltages. For example, at V_{DS}=1.0V, there is an increase of 16% in current, when impact ionization is considered. Impact ionization, increases the current by 23%, when V_{DS} is raised to 1.2V.

Figure 3.11 shows a comparison of subthreshold current obtained from (3.17) with results of ATLAS-simulation. Data for analytical calculations and ATLAS simulation are taken from 90nm parameters [Apendix-1]. Good agreement between analytical and simulated result can be seen in Figure 3.11.

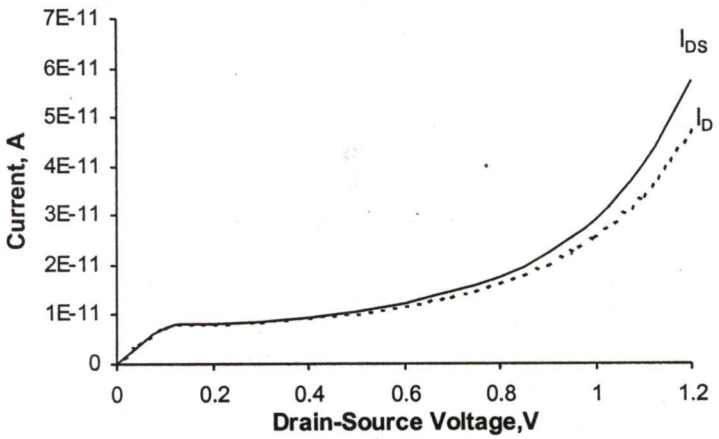


Figure 3.10: Plot showing the comparison between sub-threshold current with and without impact ionization

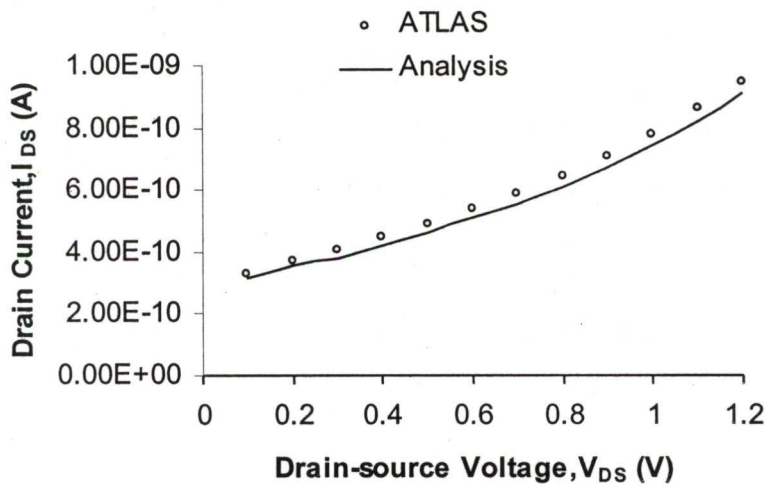


Figure 3.11: Comparison of calculated subthreshold current with ATLAS simulation result as a function of V_{DS} for V_{GS}=0.1V (Data given in Appendix-1)

3.5 Conclusions

A review of past work in the area of impact-ionization and consequent carrier multiplication has been presented in this chapter. The focus is on the analytical methods developed for the assessment of ionization rate and multiplication factor for MOSFETs operated with high channel fields. Taking help of the relevant analytical methods as obtained from the literature survey an analysis of the influence of impact-ionization on subthreshold current of MOSFET is presented. In the process an analytical model that takes into account carrier multiplication by impact-ionization has been developed.

The analysis show that, carrier multiplication is most prominent when a MOSFET operates in the subthreshold region of it's operating characteristics. The carrier multiplication is being least significant in the linear region. When a MOSFET operates in the subthreshold region, the carrier multiplication factor is strongly dependent on the maximum electric field, which occurs at the drain end of the channel.

As this maximum field depends on the effective length of the channel, it is dependent on scaling and channel length modulation parameter.

Multiplication factor increases exponentially with increase in drain-source voltage and is a slowly varying function of gate voltage. Further more, multiplication can be controlled by proper selection of gate-oxide thickness and drain/source junction depth. As the subthreshold current is proportional to multiplication factor, the behaviors of the later reflect on the behavior of the former. The analytical results obtained from the subthreshold current model presented in this chapter are in good agreement with ATLAS simulation results.

CHAPTER 4

Subthreshold Current of Submicron and Deep-Submicron MOSFET

4.1 Introduction

The secure downscaling of metal-oxide-semiconductor field-effect-transistor (MOSFET) dimensions over the past four decades has been the main motivation to the growth of silicon integrated circuits (ICs) and communication technology. The scaling of MOSFETs has been following the famous Moore's law, which is often stated as the doubling of transistor performance and quadrupling of the number of devices on a chip every three years. The downscaling reduces the capacitance of the component or MOSFET in the chip and increases the number of components. The reduction in the capacitance increases the operating speed of the circuits and decreases their power consumption.

As MOSFET dimensions are reduced, circuit requirements demand maintaining long channel behavior and minimizing short channel effects. In short channel devices, substrate doping must increase to prevent punch-through. The increase in substrate doping increases the junction electric field, thereby increasing drain-junction tunneling current into the substrate. To limit the tunneling current to a reasonable value, we reduce the supply voltage, thereby reducing the ratio of channel on-current to channel off-current. With proper scaling of all parameters of the process, device miniaturization may be achieved.

4.2 Scaling Techniques

The aim of scaling is to reduce device dimensions without disturbing its performance. In the literature, different circuit and technological solutions have been proposed, to study the scaling effects in sub-micron MOSFETs [CHHL80], [OTWC+81] [TKTA82] [THHM+83] [CSBP94] [Mea94] [PJYP+96] [IF97] [HNYJ+99] [KT02] [AME03], [Rei83]. Duncan et al. [DRJ98] studied the effects of various scaling techniques on hot carrier behavior in different kinds of n-channel MOSFET structures. Results provide the physical basis of understanding the overall behavior of impact ionization, gate oxide injection, and their relation with scaling. Somenov et al. [SVS02] investigated the impact of technology scaling and effect of temperature on the leakage mechanisms of sub-quarter micron MOSFET.

For VLSI applications, various MOS scaling techniques have been proposed in the literature such as: Constant Electric Field Scaling, Generalized Scaling and Selective Scaling etc. [BWD84] [DDS95a] [FIWS92] [FIWS94] [FTW98] [FDNS+01][MW95]. Table 4.1 sums-up the characteristic features of various techniques.

Table 4.1: Scaling rules for various scaling techniques

Quantity	Constant Electric-Field Scaling (S <1)	Generalized Scaling ($S < S'$)	Selective Scaling ($S_d < S_w < S'$)
W	S	S	S_w
L	S	S	S_d
t_{ox}	S	S	S_d
x_j	S	S	S_w
E	1	S'/S	S'/S
$N_a(N_C)$	$1/S$	S'/S^2	$S'/(S_d * S_d)$
V_{DS}, V_{TH}	S	S'	S'

S is the dimensional scaling parameter, S' is the voltage scaling parameter, and S_w and S_d are separate dimensional scaling parameters for the selective scaling case, S_d is applied to the device vertical dimensions and gate length, while S_w applies to the device width and the wiring.

In the Constant Electric Field scaling (**CEF**) [DDS95a&b], voltages and all the physical dimensions are scaled down by the same factor S (S <1) thereby maintaining electric field equal in the scaled and original device, without any undesirable side effect. In this technique, the circuit delay is affected by factor S ; whereas, the power dissipation per circuit gets reduced by S^2. However, scalability limits of power supply voltage, and non-scalability of weak inversion slope impose limitations on CEF scaling. Furthermore, threshold voltage scaling is limited by increase in sub-threshold current, which also affects the circuit speed. However, there are two problems in CEF scaling. The built in potentials do not scale because they are tied to the silicon energy band-gap, which does not change (except by changing to a different semiconductor).

Furthermore, the subthreshold slope cannot be scaled (except by lowering the temperature), since it is primarily determined by the thermodynamics of the

Boltzmann distribution of carriers. Consequently, the excessive leakage currents impose a limit on threshold voltage scaling. But both of these limitations are not valid below 1V supply voltages. These CEF scaling relations are summarized in column two of Table 4.1.

In practice, voltages are not usually scaled as fast as the linear dimensions due to these difficulties. Better results can be obtained by introducing an additional scaling factor S' for the electric field as summarized under "Generalized Scaling (**GS**)" [BWD84] (column three of Table 4.1). In GS, geometric dimensions and the substrate doping are scaled as in the CEF scaling, but voltages are not scaled as fast as linear dimensions. This scaling method also allows the field pattern to be conserved within the scaled device. As a result, punch-through and drain-induced barrier lowering remains unmodified, inspite of an increase in electric field strength. The disadvantage of this scaling is the increase in electric field, which poses a threat to the device reliability.

Selective-generalized Scaling (**SS**) [DDS95a] [FTW98] is a more flexible scaling method in which device width and wiring dimensions are not scaled as fast as the channel length. This improves the wiring yield without degrading the gate delay. In this separate scaling parameters are used to scale width and wiring dimensions.

The main problem with miniaturization is direct, and more importantly, indirect dependence of electrical characteristics on controllable physical parameters. This causes many non-ideal effects that restrict the performance of the devices.

4.3 Challenges of MOSFET Scaling

This section describes the various factors that limit the scaling of the MOSFET devices. These factors are: short channel effects, subthreshold leakage currents, gate induced drain leakage, gate tunneling, dielectric breakdown, hot carrier effects, and interconnect scaling.

4.3.1 Short Channel Effects

Short channel effects in sub-micron technologies are drain induced barrier lowering (DIBL), punch through and mobility degradation etc. These effects pose serious challenges for future MOSFET scaling. One of the most obvious consequences of scaling is the decrease in the threshold voltage as the channel length is reduced (i.e., DIBL). Short channel effects are controlled in state-of-the-art MOSFET devices by employing source/drain and channel engineered structures. Features such as source/drain extensions, halo implants and retrograde well profiles have allowed MOSFETs to be scaled to such dimensions that would have been unattainable with conventional device engineering [Rei83], [PDW01], [AME03].

At very high V_{DS}, the depletion regions of the source and drain may overlap causing large amounts of current to flow which can not be controlled by the gate. This phenomenon is known as punch through. In the ON state, reduction of the gate length decreases the channel resistance of MOSFETs. But, in the short channel MOSFET design, the source and drain resistances are increased to suppress the short channel effects. Thus, it is important to consider ways of reducing the overall resistance of MOSFETs while suppressing the short channel effects.

4.3.1.1 Length Dependent Threshold Voltage

The decrease in V_{TH} with effective length, L_{eff}, is also known as V_{TH} roll-off, which is an indication of short channel effect. The threshold voltage of MOSFETs cannot continue to be scaled down as the gate length is decreased. The subthreshold current increases as the V_{TH} decreased. An increase in subthreshold current is a serious threat to the continued performance enhancements of the MOSFET transistor. The V_{TH} decrease as the drain-source voltage is increased. To predict V_{TH} of a short channel device the shift in the threshold voltage, which is measure of DIBL coefficient, must be approximated.

Figure 4.1 shows the DIBL effect for an n-channel MOSFET. As the voltage drop increases, the depletion region under the drain may lower the potential barrier between the

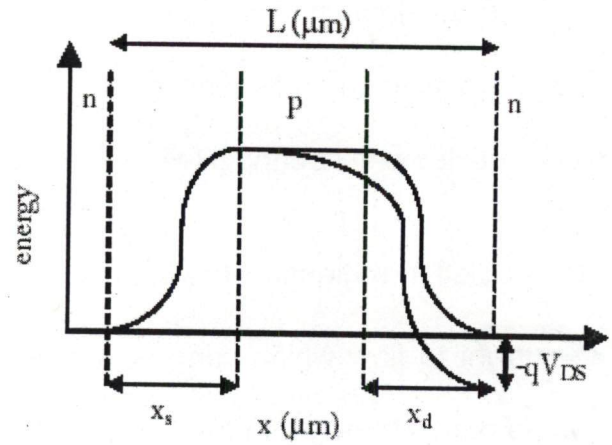


Figure 4.1: Drain Induced Barrier Lowering (DIBL)

source and channel junction and thus electrons are more freely injected into the channel region, which increases the current in the channel and varies exponentially with the barrier height. A slight decrease in the barrier increases the channel current significantly.

4.3.1.2 Width Dependent Threshold Voltage

Channel width scaling affects the threshold voltage in three ways but effects are not as severe as the length scaling. Two effects are due to the fabrication of isolation structures, either due to raised field-oxide or semi-recessed LOCOS (Local Oxidation of Silicon), due to this effect V_{TH} increases (opposite to length scaling)). The third effect is due to fully recessed LOCOS and decreases V_{TH}. This effect is also known as the inverse narrow-width effect.

4.3.1.3 Reverse Short Channel Effects

When threshold voltage increases with effective gate length, L_{eff}, scaling this is referred as V_{TH} - rollup. The two effects, V_{TH} roll-up and V_{TH} roll-off, compete with one another until V_{TH} roll-off becomes the dominant effect due to scaling. This is known as reverse short channel effect.

4.3.2 Subthreshold Leakage Currents

Short channel effects reduce the threshold voltage of MOSFET transistors, which in turn increases the leakage current. Therefore, subthreshold leakage currents restrict the scaling of MOSFET devices. Simple MOSFET models assume that the drain to source current is zero, when the gate to source voltage is less than the device threshold voltage ($V_{GS} < V_{TH}$). In reality the current does not drop immediately to zero but decreases exponentially as the gate voltage drops below the threshold voltage [TBCF+97]. This is due to in flow of thermally generated electrons, which overcome the potential barrier. The minimum values for the ratio between On and Off currents gives the lower limits of threshold voltage and to determine the

minimum value of V_{TH}, a minimum ratio of V_{DD}/V_{TH} must be required. As the ratio V_{DD}/V_{TH} decreases, circuit's performance deteriorates.

4.3.2.1 Gate Induced Drain Leakage

In addition to subthreshold leakage from the drain to the source, MOSFETs with high oxide fields may also show significant leakage from the drain to the bulk. This phenomenon is known as gate induced drain leakage (GIDL) [Cri87]. High fields in the oxide produce high fields in the surface of the silicon. GIDL limits the gate oxide thickness. In heavily doped gate to drain overlap region, GIDL produces band bending, greater than the silicon band gap, over a very short vertical distance. Within this depleted region at the surface of the drain, electrons may tunnel from the valence band into the conduction band producing a drain to bulk current.

4.3.2.2 Gate Tunneling Current

This phenomenon is known as tunneling. As gate oxide is scaled down, significant tunneling current may flow from drain to the gate in OFF devices or from the gate to the source in ON devices. Gate tunneling current (GTC) imposes the fundamental limits on the scaling of gate oxide thickness.

4.3.4 Hot Carrier Effects

With the increasing fields in the gate oxide, lateral fields within the channel also increase gradually and at sufficiently high fields, electrons may gain enough energy to cause impact ionization in the channel. The energetic carriers thus produced lead to gate and substrate current. Interface traps may also be formed in the gate oxide, and hot electrons may become trapped within the oxide. This buildup

of traps and negative charge in the oxide causes the threshold voltage to increase and which in turn decreases the transconductance. Eventually, increase in current due to hot carrier effects (HCE) degrades the performance. Both nMOS and pMOS devices experience hot carrier effects, but due to the lower mobility of holes hot carrier effects are less significant in pMOS devices.

4.3.4 Dielectric Breakdown

Dielectric breakdown limits the gate oxide thickness. A vertical field within the gate oxide increases steadily as power supply voltage is scaled more gradually than oxide thickness. Eventually, very high fields damage the oxide layer and cause breakdown.

4.3.5 Scaling the Interconnects

Interconnect scaling improves interconnect density at the cost of interconnect delay. Interconnect RC delay is a significant fraction of clock cycle time and has a significant impact on the overall chip performance. Due to the scaling of the interconnect pitch the line resistance increases thus interconnect RC delay increases. Therefore to minimize this delay, more interconnect metal layers are needed to meet both density and performance requirements.

4.4 Controlling Factors of Drain Induced Barrier Lowering Coefficient in Short Channel MOSFET

The threshold voltage of a MOSFET is known to vary with drain-source voltage, V_{DS} [FS93] [ST95] [TN97] [Won00] [CJMK+91] [LHHC+93] [FGD94] [Ghi99] [MR00], because of drain-induced barrier lowering (DIBL). Most of the analytical

works considered this shift in threshold voltage (with change in V_{DS}) for constant effective channel length. Theoretical as well as experimental observation shows that for constant effective length, the threshold voltage shift is linearly related to V_{DS}. The rate of change of threshold voltage-shift with V_{DS} is known as the DIBL-coefficient, σ, [FS93].

Both theoretical and experimental studies show that for constant effective length, the DIBL-coefficient is independent of V_{DS}. When a MOSFET, operates with variable V_{DS}, channel length modulation makes the effective length to vary. Thus, it is of interest to know the influence of V_{DS} on DIBL-coefficient of an individual MOSFET. An accurate estimate of the shift in the potential minimum is especially important since the channel current is exponentially dependent on the barrier height. The purpose of this analysis is to analyze the dependence of DIBL-coefficient and threshold voltage-shift of an individual MOSFET as its drain-source voltage.

The influences of some technological parameters are also examined.

4.4.1 Numerical Analysis

Fjeldly and Shur [FS93] proposed a new analytical model for the sub-threshold regime of operation of short-channel MOSFET's and developed an expression for the threshold voltage shift associated with the drain-induced-barrier lowering (DIBL) due to the application of a drain bias. This model is based on the concept of charge sharing, where the depletion charge under the gate is identified with counter charges on the gate electrode, the source and drain contacts. In particular, they estimated the amount of drain-bias-induced depletion charge in the channel and developed an expression for distribution of this charge along the

channel, based on a simplified two-dimensional solution of Poisson's equation. From this distribution, they found the lowering of the potential barrier between the source and the channel, and the corresponding threshold voltage shift ΔV_{TH}.

The threshold voltage shift in terms of DIBL-coefficient σ and V_{DS} is

$$\Delta V_{TH} = -\sigma V_{DS}. \tag{4.1}$$

The DIBL-coefficient is given by [FS97]

$$\sigma = \frac{2\eta\chi d_{dep}^{0} \sinh\left(\frac{x_s}{\lambda}\right)}{\lambda \cosh\left(\frac{L - x_d}{\lambda}\right) - \cosh\left(\frac{x_s}{\lambda}\right)} \tag{4.2}$$

where χ is a constant (on the order of 0.5) which accounts for the charge sharing in depletion region other parameters are as defined in (3.11) and (3.12).

Defining

$$\left.\begin{aligned} a_1 &= \frac{2\varepsilon_s}{qNa}\left(2\varphi_b + \frac{(V_{GS} - V_{TH0})}{\eta}\right); \\ a_2 &= \frac{2\varepsilon_s}{\eta qNa} V_{DS}; a_3 = \frac{2\varepsilon_s}{\eta qNa} V_{bi}; \\ \text{and,} \quad a_4 &= \frac{\varepsilon_i}{\varepsilon_s t_{ox}} \end{aligned}\right\} \tag{4.3}$$

Combining (4.3), (3.11) and (3.12) we get

$$x_s = \sqrt{a_3 - a_1 + a_2\sigma}; \tag{4.4}$$

$$x_d = \sqrt{a_3 - a_1 + (\eta + \sigma)a_2}; \tag{4.5}$$

$$d_{dep}^{0} = \sqrt{a_1 - a_2\sigma}\,; \tag{4.6}$$

$$\text{and, } \lambda = \frac{d_{dep}^{0}}{1 + a_4 d_{dep}^{0}}. \tag{4.7}$$

Substituting the parameters defined by (4.4)-(4.7) in (4.2) leads to the following expression for the threshold voltage shift:

$$\sigma - \frac{2\eta\chi\left(1 + a_4\sqrt{a_1 - a_2\sigma}\right)\sinh\left(\frac{\left(\sqrt{a_3 - a_1 + a_2\sigma}\right)\left(1 + a_4\sqrt{a_1 - a_2\sigma}\right)}{\sqrt{a_1 - a_2\sigma}}\right)}{\cosh\left(\frac{\left(L - \sqrt{a_3 - a_1 + a_2(1+\sigma)}\right)\left(1 + a_4\sqrt{a_1 - a_2\sigma}\right)}{\sqrt{a_1 - a_2\sigma}}\right) - \cosh\left(\frac{\left(\sqrt{a_3 - a_1 + a_2\sigma}\right)\left(1 + a_4\sqrt{a_1 - a_2\sigma}\right)}{\sqrt{a_1 - a_2\sigma}}\right)} = 0 \tag{4.8}$$

σ as a function of V_{DS} can be obtained by numerically solving (4.8).

To analyze the V_{DS} dependence of σ, the case of an n-channel MOSFET (L=0.25μm), characterized by the parameters given in the Table 4.2 are considered.

Table 4.2: MOSFET Parameters for DIBL Analysis

Parameters	Values
ε_s	1.04×10^{-10} Fm^{-1}
ε_i	3.36×10^{-11} Fm^{-1}
N_a	7.0×10^{22} m^{-3}
n_i	6.5×10^{14} m^{-3}
t_{ox}	8.6 nm
V_{TH0}	0.402 V
V_{GS}	$(0.9V_{TH0})$ V
η	1.196

Figure 4.2 illustrates the variation of DIBL-coefficient with drain- source voltage. σ is a non-linear function of V_{DS}, and it increases as V_{DS} is increased. Furthermore the increase in σ is more pronounced in the higher regime of V_{DS}. The concept of V_{DS} independent σ is valid in the lower regime of drain-source voltage. It also shows the expected channel length dependence of σ, which is higher for devices of shorter channels. The slope of the σ-V_{DS} characteristics is relatively higher in a device of short channel.

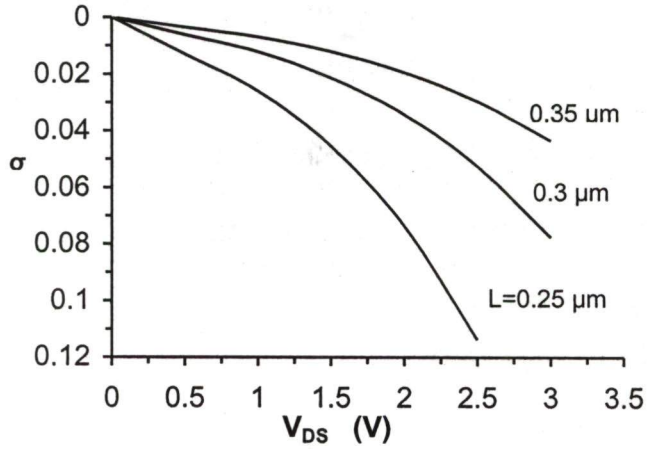


Figure 4.2: Variation of DIBL-coefficient with drain -source voltage for different channel lengths

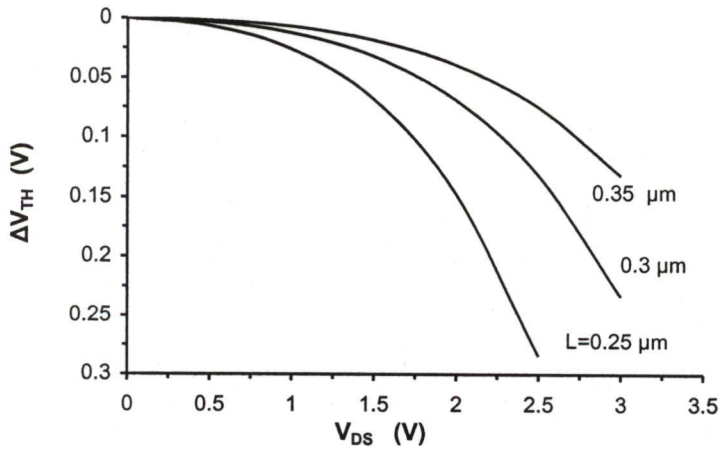


Figure 4.3 Threshold voltage shift as a function of drain bias for different channel lengths

The effect of non-linear variation of σ with V_{DS} results in a non-linear relationship between the threshold voltage-shift, ΔV_{TH} and V_{DS}. Figure 4.3 shows that in the higher regime of V_{DS}, a large swing in ΔV_{TH} may result for a small change in V_{DS}. The ΔV_{TH} swing is larger in cases of a short channel. The non-linearities in the σ-V_{DS} and ΔV_{TH}-V_{DS} curves shows that σ determined from ΔV_{TH}-V_{DS} plots with effective channel length as the parameter does not yield a correct picture of DIBL-effect on threshold voltage. Figure 4.4 shows the logarithmic plot of ΔV_{TH} as obtained from (4.8). This is in agreement with experimental results of Chung et al. [CJMK+91]. The dependence of σ on oxide thickness and substrate doping concentration are shown in Figure 4.5. Significant reduction in threshold voltage shift can be achieved by decreasing oxide thickness, t_{ox}, or increasing doping concentration, N_a.

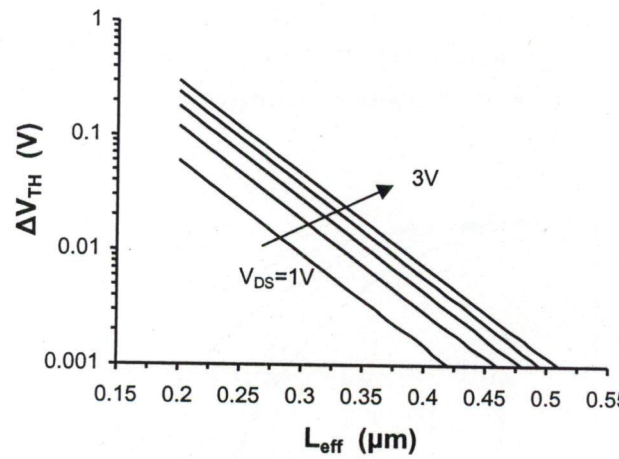


Figure 4.4: Shift in threshold voltage as a function of effective channel length with V_{DS} (1,1.5, 2, 2.5and 3 V) as the parameter

Fikry et al. [FGD94] showed that the DIBL parameter is almost independent of temperature between 50 and 300K but it is not so in our case. It is observed that

the σ decreases with increase in temperature and this decrease is more pronounced for short channel devices. Figure 4.6 shows the variation of σ with temperature. This is because of the variations of the η, ψ_s^0 and V_{bi} with temperature. This is in agreement with experimental results of Ghitani [Ghi99].

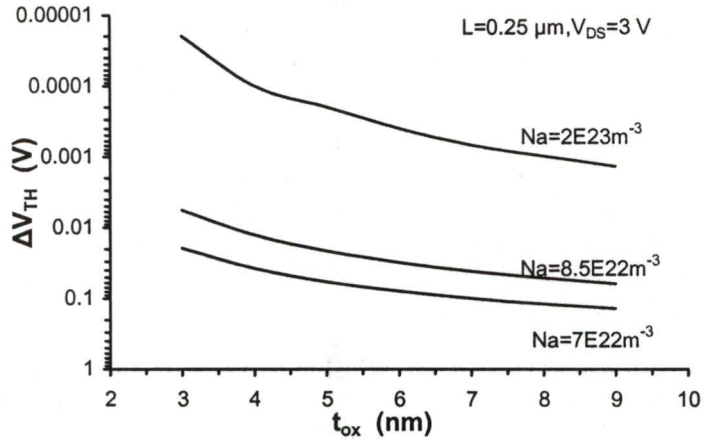


Figure 4.5: Variation of threshold voltage shift with oxide thickness for different doping concentrations

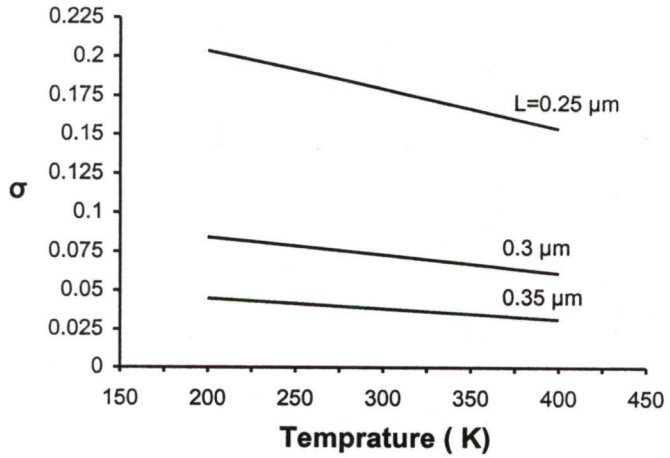


Figure 4.6 DIBL-coefficient versus temperature for different channel lengths for V_{DS}=3V

4.5 Effect of Technology Scaling on MOSFET

To study the effect of technology scaling on device performances in the sub-threshold condition, the various equations used in Chapter 3 has been considered.

A schematic diagram of the simulated n-channel MOSFET structures is used as shown in Figure 3.2 of Chapter 3. The parametric values for this device are tabulated in the Table-3.1. As discussed in section 4.2 scaling affects various parameters such as voltage drops, fields, doping, dimensions of the device etc. (refer table 4.1) hence affects other dependent parameters like current, power etc. Now, after scaling various parameters given in (3.9) and (3.10), obtained scaled parameters are as

$$
\begin{aligned}
&\delta' = V_T \sqrt{\frac{\varepsilon_s}{2qN_a'\left(2\varphi_b + \frac{(V_{GT} * S)}{\eta}\right)}} \quad \text{where } N_a' = \frac{N_a}{S};\\
&\eta' = 1 + c_{dep} * S/(c_{ox} * S);\\
&\left(\psi_s^0\right)' = -V_{bi} + 2\varphi_b + \frac{V_{GT} * S}{\eta};\\
&x_s' = \sqrt{\frac{2\varepsilon_s}{qN_a'}\left(-\psi_s^0\right)'};\\
&x_d' = \sqrt{\frac{2\varepsilon_s}{qN_a'}\left(V_{DS} * S - \left(\psi_s^0\right)'\right)};\\
&\left(d_{dep}^0\right)' = \sqrt{\frac{2\varepsilon_s}{qN_a}\left(\left(\psi_s^0\right)' + V_{bi}\right)};\\
&\lambda' = \left(d_{dep}^0\right)'\left(1 + \frac{\varepsilon_i\left(d_{dep}^0\right)'}{\varepsilon_s(t_{ox} * S)}\right)^{-1/2};\\
&\text{and} \quad V(x_s)' = V_0' \sinh\left(\frac{x_s'}{\lambda'}\right)\\
&\text{where}\\
&V_0' = 2\chi V_{DS} * S \frac{\left(d_{dep}^0\right)'}{\lambda'}\left[\cosh\left(\frac{L * S - x_d'}{\lambda'}\right) - \cosh\left(\frac{x_s'}{\lambda'}\right)\right]^{-1}.
\end{aligned}
\qquad (4.9)
$$

The expression for E_m is now become

$$E_m^{'} = \sqrt{\frac{(V_{DS} * S)^2}{(l_d^{'})^2} + \frac{2E_a^{'}V_{DS} * S}{(l_d^{'})^2}} \tag{4.10}$$

where $l_d^{'} = 1.7 * 10^{-2}(t_{ox} * S)^{1/8}(x_j * S)^{1/3}(L_{eff}^{'})^{1/5}$ m and $L_{eff}^{'} = L * S - x_s^{'} - x_d^{'}$.

since multiplication is significant, then combining (3.7), (3.8), (3.9), (3.10), (3.13), (3.14) and (4.9) the scaled subthreshold current is given as

$$I_{DS} = \left(1 + \frac{A_n}{B_n}(V_{DS} * S)\exp\left(\frac{-B_n l_d^{'}}{V_{DS} * S}\right)\right)\left(q\frac{N_C}{S}(W * S)\delta^{'}A^{*}T^2\exp\left(\frac{(\psi_s^0)^{'} + (V(x_s))^{'}}{V_T}\right)1 - \exp\left(\frac{-V_{DS} * S}{V_T}\right)\right) *$$
$$\left(1 + \frac{A^{*}T^2\lambda^{'}V_T\left(1 - \exp\left(\frac{-V_0^{'}(L * S - x_s^{'} - x_d^{'})}{V_T}\right)\right)}{q\left(\frac{N_C}{S}\right)D_n V_0^{'}}\right)^{-1} \tag{4.11}$$

The physical and geometrical parameters of the smaller devices have been obtained from the 100nm transistor following the constant electric field scaling, generalized scaling and selective scaling rules. An ideal structure with gate lengths of 100, 90 and 70nm; with scaling factors of 1.0, 0.9 and 0.7 are considered for performance comparison. Scaling factors considered in three scaling techniques are: $S^{'} = S + 0.01$ in generalized scaling, $S^{'} = S_d + 0.02$ and $S_w = S_d + 0.01$ in selective scaling, since GS required an additional scaling factor for field and wiring is not scaled to the same extent as the gate length in selective scaling.

First considering the **CEF** scaling technique in 100nm device and applying the rules we will have parameters as shown in Table 4.3.

Table 4.3: Parameters for the scaled device following the CEF scaling method

L (nm)	t_{ox} (nm)	N_a (10^{24}m^{-3})	N_C (10^{26}m^{-3})	x_j (10^{-8}m)	S
$L*S=100$	$t_{ox}*S=2.0$	$N_a/S=1.5$	$N_C/S=3.0$	$X_j*S=11.1$	1
$L*S=90$	$t_{ox}*S=1.8$	$N_a/S=1.66$	$N_C/S=3.33$	$X_j*S=9.99$	0.9
$L*S=70$	$t_{ox}*S=1.4$	$N_a/S=2.14$	$N_C/S=4.285$	$X_j*S=7.77$	0.7

Similarly by applying Generalized and Selective scaling methods we may obtain the corresponding parametric values for analysis.

Considering the equations (4.9), (4.10) and (4.11) it is clear that after scaling the parameters, values of dimension parameters and voltages are also get reduced. Ultimately field is getting increased except in CEF scaling method. Now, calculating the maximum field using equations (4.10). It can be seen that as channel is shorten field becomes high. Figure 4.7 shows the lateral field distribution along the channel for channel lengths (L) equal to 70nm, 90nm and 100nm using selective scaling techniques. It also illustrates the effect of scaling on channel electric field. Interestingly, the maximum electric field E_m', at the drain end is larger for longer channel. The crossover is close to the drain indicates that the longer device field is lower in most of the channel. This shows that the rate of variation of field with distance from source is larger in a longer device. It can easily be seen from (3.7) and (4.10) that the Maximum electric field E_m' greatly influences the carrier multiplication process. This field in turn depends on the important parameters of effective length L_{eff}', and drain-source voltage V_{DS}, as may be seen in (4.10) and (3.4).

Figure 4.8 illustrates the dependence of E_m' on normalized V_{DS} for the devices in consideration in CEF, GS and GS techniques. As discussed previously, E_m', depends on V_{DS} and E_m' increases with increase in V_{DS}. The increase in E_m' with

increase in normalized V_{DS} and decrease in L'_{eff} indicates the influence of channel width modulation on the field. Increased channel length modulation increases the maximum electric field, as in case of selective scaling (SS will have smallest L'_{eff}, since $L'_{eff} = L' - x'_s - x'_d$ and x'_d, x'_s are the function of voltages (3.11) which are scaled less drastically). The crossover in Figure 4.7 actually refers to drawn gate lengths.

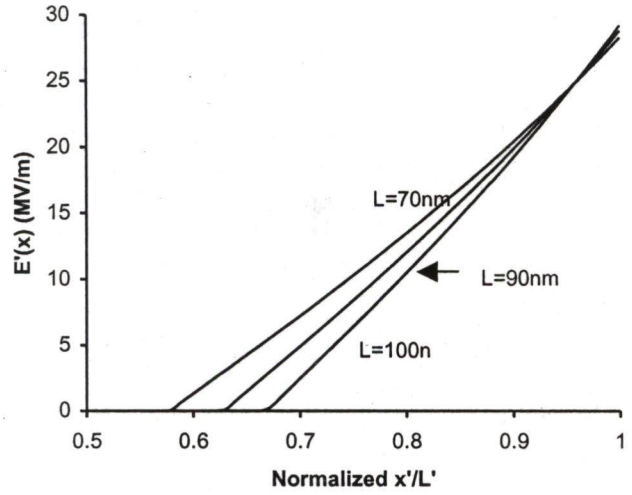


Figure 4.7: Lateral Electric Field as a function of normalized channel position for selective scaling method. Where L is channel length

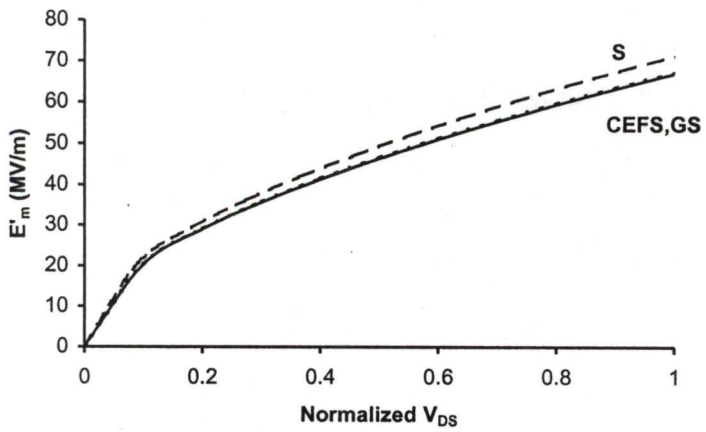


Figure 4.8: Maximum Electric Field as a function of normalized V_{DS} for different scaling methods (L= 90nm)

Figure 4.9, shows the dependence of M on the on normalized V_{DS} for three scaling methods. Results show that M increases with increases in V_{DS} and scaling. It also depicts that smaller devices have larger multiplication due to increase in high electric field in the channel and different scaling methods do have much affect on it.

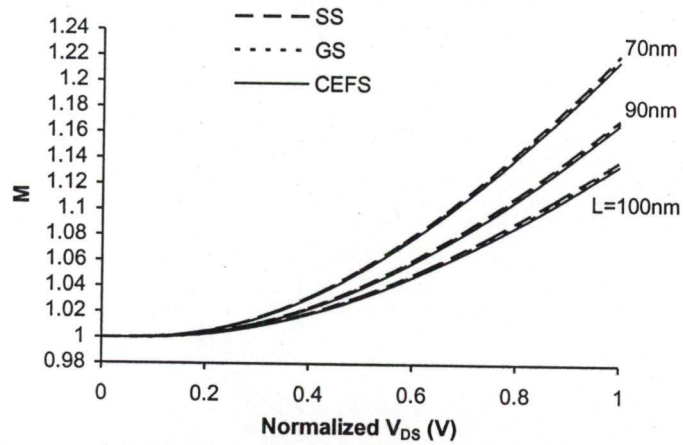


Figure 4.9: Multiplication factor as a function of normalized V_{DS} for three CEFS, GS and SS scaling scheme

Figure 4.10 shows the I_{DS} -V_{DS} characteristics of an NMOS transistor of 100nm length using (4.11) for CEFS with scaling factor S=0.7, 0.9 and 1.0. It is seen that for small drain voltages (V_{DS} <3V_{TH}), the I_{DS} current increases almost linearly. This is because the contribution from the term $\exp(-V_{DS}/V_{TH})$ is large. As V_{DS} increases, $(\psi_s^0)'$ increases and $(V(x_s))'$ decreases therefore the current now rises steeply due to increased injection from the source. There comes a point when for a very small increase in V_{DS} the current rises very much. For example, in case of 70nm CEF scaling, at V_{DS} =0.6V current I_{DS} is 9.44E-11 A and at V_{DS} =0.8V, I_{DS} =3.67E-10 A, I_{DS} increases approximately 4 times with increase in 0.2V.

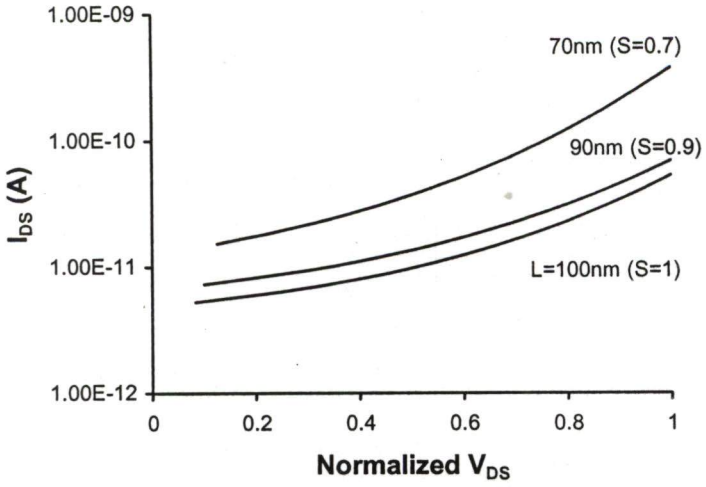


Figure 4.10: Plot of Subthreshold current with impact ionization versus normalized V_{DS} for CEF scaling methods

On comparing these three scaling it is observed that selective scaling gives the lowest current. For example, at normalized V_{DS} =1V, I_{DS} is 5.2E-11 A for CEF scaling, I_{DS} =4.52E-11 A for generalized scaling and I_{DS} =3.99E-11 A for selective scaling (Figure 4.11). Also, subthreshold current increases with scaling down the channel.

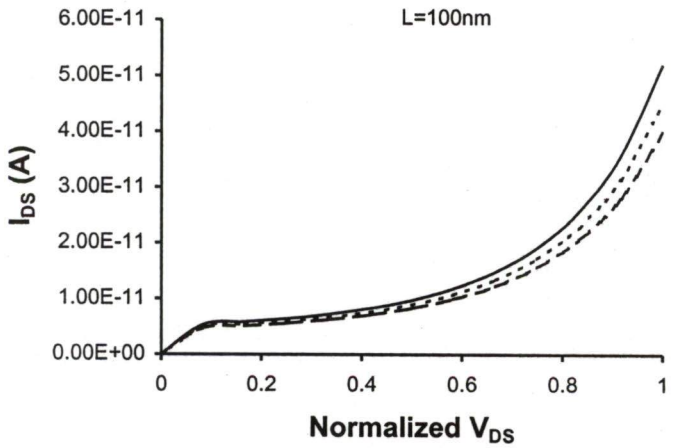


Figure 4.11: Subthreshold current with impact ionization versus normalized V_{DS} plot for CEFS, SS and GS methods for S=1.0

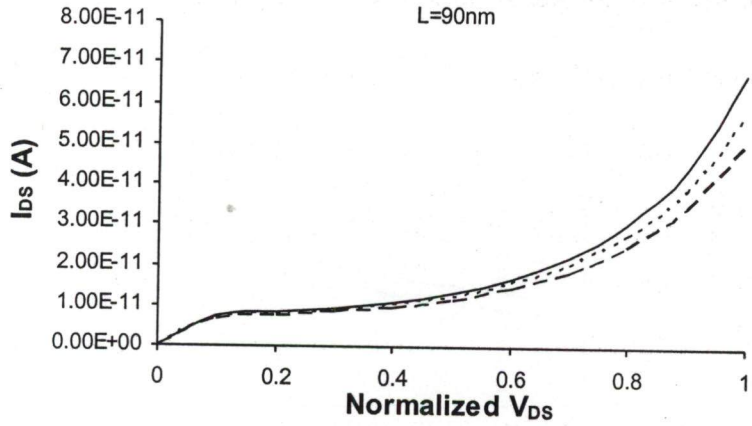


Figure 4.12: Subthreshold current with impact ionization versus normalized V_{DS} plot for CEFS, SS and GS methods for S=0.9

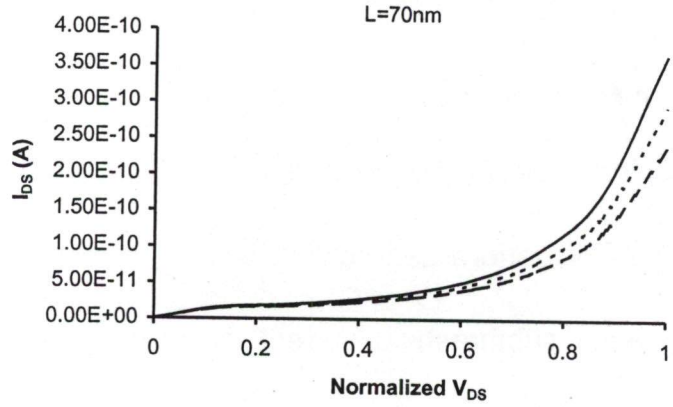


Figure 4.13: Subthreshold current with impact ionization versus normalized V_{DS} plot for CEFS, SS and GS methods for S=0.7

Figure 4.12 and Figure 4.13 shows the effect of different scaling techniques on device subthreshold current with impact ionization for scaling factor S=0.9 and S=0.7, respectively. Table 4.4 summarizes the current trends in deferent scaling methods.

Table 4.4: Subthreshold Current trends in CEFS, GS and SS at normalized V_{DS} =1 V after scaling 100nm nMOSFET

Length	CEF Scaling	Generalized Scaling	Selective Scaling
100nm	5.2E-11	4.53E-11	3.99E-11
90nm	6.77E-11	5.85E-11	5.11E-11
70nm	3.67E-10	2.93E-10	2.39E-10

An overview of effect of different scaling (as shown in Table 4.1) on various device and technology parameters are presented in Table 4.5 for scaling factor S =0.7.

Table 4.5: Various device and technology parameters obtained after scaling 100nm device

Physical Parameters	CEF Scaling (S =0.7)	Generalized Scaling (S =0.7, $S'=S+0.01$)	Selective Scaling (S_d =0.7, $S'=S_d+0.02$, $S_w=S_d+0.01$)
L (nm)	70	70	70
t_{ox} (nm)	1.4	1.4	1.4
N_a ($10^{24}m^{-3}$)	2.14	2.17	2.2
N_C ($10^{26}m^{-3}$)	4.28	4.35	4.41
x_j (10^{-8}m)	7.77	7.77	7.88
V_{DS} (V)	0.7	0.71	0.72

The effect of non-linear variation of σ with V_{DS} results in a non-linear relationship between the threshold voltage-shift, ΔV_{TH} and V_{DS}. Figure 4.14 shows that in the higher regime of V_{DS}, a large swing in ΔV_{TH} may result for a small change in V_{DS} for three scaling methods.

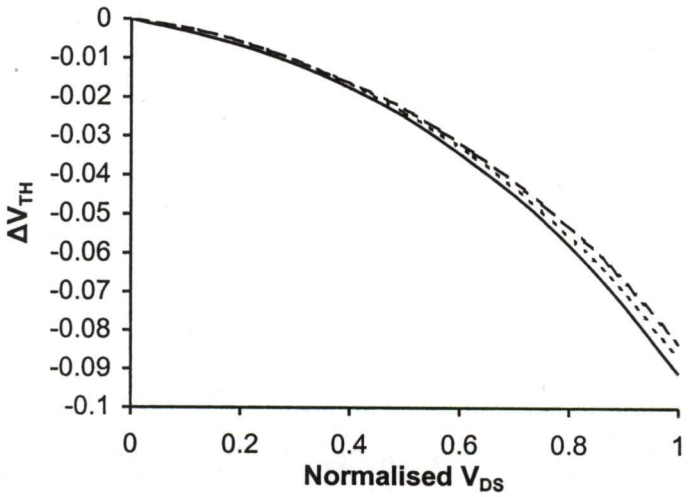


Figure 4.14: Effect of scaling on shift-in-threshold voltage in 70nm device on CEFS, GS and SS techniques

Constant electric field scaling produces the largest threshold voltage shift due to DIBL. Selective scaling has the least DIBL, with generalized scaling produced DIBL falling in between the two extremes.

4.6 Conclusions

The dependence of DIBL-coefficient and threshold voltage-shift due to drain-source voltage has been analyzed. It has been observed that the DIBL-coefficient being a non-linear function of drawn gate length, the threshold voltage-shift varies non-linearly with drain-source voltage. Furthermore, the threshold voltage shift can be controlled by proper selection of gate oxide thickness and substrate doping. It has also been observed that the DIBL-coefficient decreases with increase in temperature and this decrease is more pronounced for short channel devices.

Effects of the various scaling techniques on a 100nm device performances and the dependence of subthreshold current parameters on applied scaling technique has been analyzed. The results show that as the channel length is scaled down, multiplication factor increases slowly in the higher regime of channel length. In the lower regime of channel length, multiplication factor rises rapidly. This result also justifies the inclusion of impact – ionization effect on subthreshold current. However, the analysis shows that there is insignificant dependence of multiplication factor on the method of scaling. Similar variations in subthreshold current with channel length scaling have been observed in the analytical results for different scaling techniques. Results show that subthreshold current increases with scaling channel length down.

CHAPTER 5

Optimum Supply Voltage for Minimum Power and Energy

5.1 Introduction

The continuous development of recent mobile and portable devices [CASB94] and applications, such as hearing aids [MBSH+99], pacemaker [KR01] wearable computing [Sta96] and self-powered devices [AC98], has created a tremendous thrust for low power circuits' design. Various methods and techniques viz. voltage scaling, clock gating, leakage power reduction techniques etc. [CB95] [HIG95] have been applied successfully in the medium power, medium performance region of the design spectrum for low power consumption.

This chapter focuses on design techniques for low power applications where performance is of secondary importance. One way to achieve this goal is by

operating the digital circuits in subthreshold region. Subthreshold circuits can also be used in applications where the circuits remain idle for long period of time. For example, when the microprocessor goes into deep sleep, specific modules will survey the input devices (e.g. keyboard or mouse) at a regular period of time. This continuous periodic monitoring can be done in subthreshold mode where the circuits consume much less power, while running at much slower speed.

The proposed subthreshold current model is a well known analytical model of Fjeldly et al. [FS93], including the effect of impact ionization, has been used to study the power dissipation in subthreshold CMOS operation and to design low-power CMOS logic. It is well known that in above threshold CMOS operation, a tradeoff exists between propagation delay and power dissipation, when voltage is scaled. It has been observed that when the supply voltage is scaled down to reduce power dissipation, propagation delay is increased. SPICE simulation, however, shows that this does not hold good for subthreshold logic [Sta01]. For a sub-threshold CMOS– logic operation, propagation delay in most cases first increases and then decreases with supply voltage scaling. The subthreshold current model mentioned above is used to calculate voltage dependence of propagation delay of a CMOS inverter. Results obtained show that the delay increases with increase in supply voltage. An important requirement of a logic circuit is symmetric input-output characteristics. The ratio of pMOS and nMOS widths, $\left(\frac{W_p}{W_n}\right)$, that meets this requirement is determined by using the modified subthreshold current model. The ratio thus obtained is primarily mobility dependent, as in the case of above threshold logic operation.

5.2 Power and Energy Management in CMOS Circuits

Power dissipation in CMOS digital circuits is categorized into two types: peak power and time-averaged power consumption. Peak power is a reliability issue that determines both the chip lifetime and performance. The time-averaged power consumption in conventional CMOS digital circuits occurs in two forms: dynamic and static. Dynamic power dissipation occurs in the logic gates that are in the process of switching from one state to another. During this process, any internal and external capacitance associated with the gate's transistors has to be charged, thereby consuming power. Static power dissipation is associated with inactive logic gates (i.e. not currently switching from one state to another). Dynamic power is important during normal operation, especially at high operating frequencies, whereas static power is more important during standby, especially for battery-powered devices. An overview of the different power dissipation types is given in Figure 5.1.

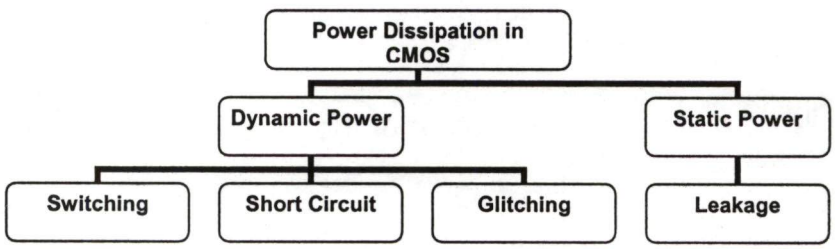


Figure 5.1: Different power dissipation types in CMOS circuits

Several methods for power and energy reduction have been proposed in [BNYT93], [CBSP95], [LS93], [LM93], [KS94], [PR94], [SL94], [SL96], [MDMA+95], [MWTL+94], [BDAM96], [KFMN+96], [KNC98], [PDC98], [KNBH+99], [BDAM00], [IHS01], [KHK02], [WCK02] and [KKK03]. Most of the circuit techniques proposed in

these papers have already been discussed in chapter 2. Some more techniques, which are proposing minimum power methodology, are presented below:

5.2.1 Existing Dynamic Power Estimation Techniques

Dynamic power dissipation primarily caused by the current flow from the charging and discharging of parasitic capacitances, consists of three components: switching power, short-circuit power, and glitching power. In digital CMOS circuits, the switching power is dissipated when current is drawn from the power supply to charge up the output node capacitance. During this switching event, the output node voltage typically makes a full transition from 0 to V_{DD}, and one-half of the energy drawn from the power supply is dissipated as heat in the conducting pMOS transistors. The energy stored in the output capacitance during charge-up is dissipated as heat in the conducting nMOS transistors, when the output voltage switches from V_{DD} to 0. A CMOS inverter circuit, shown in Figure 5.2, is used to illustrate the dynamic power dissipation during switching.

The total capacitive load C_{load} at the output of the inverter consists of the diffusion capacitance of the drains of the inverter transistors, the total interconnect capacitance, and the input gate oxide capacitance of the driven gates that are connected to the inverter's output. In most CMOS digital circuits, the switching power is the dominant component in power dissipation. The generalized expression for the switching power dissipation of a CMOS logic gate can be written as (2.1).

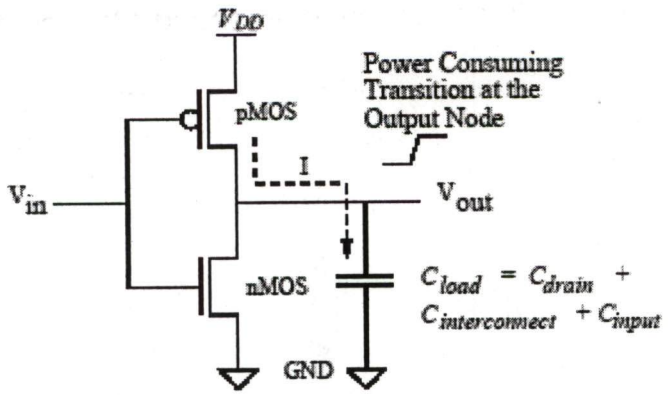


Figure 5.2: CMOS inverter for switching power calculation

Equation (2.1) indicates that the supply voltage is the dominant factor in the switching power dissipation. Thus, reducing the supply voltage is the most effective technique to reduce the power dissipation. Other methods such as reducing the switching activity and the load capacitance [KL03], for reducing the power consumption are also suggested by the equation.

The main components of logic level power estimation are switching activity and capacitance estimation. Switching activity depends upon input data and correlations between inputs and the internal nodes. There can be four types of correlation:

1. *Temporal Correlation*: This arises when the previous value of a signal can be correlated with the present value of the signal.

2. *Spatial Correlation*: Spatial correlation arises when the two spatially connected signals are dependent on each other. It is caused in general by re-convergent fan-out, by feedback and by already correlated primary inputs.

3. *Spatio-temporal*: Spatio-temporal correlation is dependence of a signal to the previous value of a spatially connected signal. Hence, it is a combination of spatial and temporal correlation.

4. *Sequential*: It is part of spatial correlation when the dependence is amongst the feedback state lines.

Switching activity can be estimated at various levels of abstraction namely at architectural, behavioral, RTL, logic, or transistor levels. As discussed earlier, estimation at each level has its own advantages and disadvantages. More power savings can be achieved as power is estimated and optimized at the higher levels, whereas the estimates are significantly more accurate at the transistor level. Switching activity estimation strategies can be divided into three broad categories: estimation by simulation, estimation by statistical simulation and estimation by probabilistic techniques.

Estimation by pure simulation is extremely accurate but time consuming [Kan86]. To decrease the time complexity, several improved simulation techniques have been proposed in [BFOR93] and [RHA92]. Many of them use vector compaction and modeling of input sample space and sequence generation to reduce the samples needed for simulation. The simulation-based techniques are strongly input pattern dependent. In general, all simulation tools have higher accuracy compared to existing methods but with higher time consuming.

In statistical simulation, statistical methods are used in conjunction with simulation in order to determine the stopping criterion for the simulation. The earliest works in this area can be found in [BNYT93] and [NGH95]. These methods are

efficient in terms of the time required and yield accurate estimates if the statistical distribution of the input data is modeled properly.

Probabilistic techniques are fast and better, but involve assumptions about joint correlations. The primary conceptual difference is that, the input statistics are first gathered in terms of probabilities and then these probabilities are propagated. Hence, the abstracted knowledge about inputs is used to estimate the switching activity of internal nodes. Therefore, these techniques can more easily model changes in input pattern efficiently than other methods. Unlike simulation and statistical simulation, one needs to know the dependencies in the circuit structure to propagate probabilities efficiently. Probabilistic techniques can be further classified into probabilistic simulation [DTP98] [NBYH90] [SH93], and purely probabilistic methods [GDKW92] [Bry92].

Najm et al. [NBYH90] estimated the mean and variance of current using probability waveforms. Probability waveform consists of probability of a signal to be 1 for a certain time interval and probabilities of transition from low to high and high to low at a particular time instant. Further, Najm et al. [Naj93] introduced the concept of transition density, which is also a propagation based strategy using Boolean Difference Algorithm (BDA).

Kapoor [Kap94] has modeled structural dependencies approximately by partitioning the circuit into local Binary Decision Diagrams (BDDs) for signal probability.

Schneider et al. [SS94] used one-lag Markov model to capture temporal dependence. The first order temporal model is valid only under zero delay models,

where, the present value of a node is independent of all the past values. Tsui et al. [TPD93] modeled first order spatial correlation efficiently using correlation coefficients and utilizing them in probabilistic propagation. Marculescu et al. [MMP94] studied temporal and spatial dependencies jointly.

5.2.2 Existing Leakage Power Estimation Techniques

Static power dissipation is due to leakage currents while the gates are idle i.e. no output transitions. Theoretically, CMOS gates do not consume any power in this condition. This is due to the fact that either pull-down or pull-up networks are turned off, thus preventing static power dissipation. In reality, however, there is always some leakage current passing through the transistors, which indicates that, the CMOS gates consumes certain amount of power. Though the static power consumption, associated with an individual logic gate is extremely small, the total effect becomes significant when millions of gates are utilized in VLSI circuits. Furthermore, as transistors shrink in size, due to high packaging density, the level of doping increases thereby causing increase in leakage currents in subthreshold region.

Previous researchers reported that both dynamic and static power dissipations are data dependent [SEOP+02] [WV98] [KS02] [ADRL+03]. The accuracy of leakage power estimation model is dependent on the stand-by leakage current model [CJWR98]. Since leakage power depends on the primary input combinations, [CJWR98], suggests that the leakage power could be minimized if we apply the input combinations corresponding to the minimum leakage power. To obtain the minimum and maximum values for leakage power dissipation, in [CJWR98], the authors developed an accurate leakage model considering the

effects of transistor stacks and implemented it in the genetic algorithm framework. [SEOP+02], introduces a new approach for accurate and efficient calculation of the average leakage current in circuits by determining the dominant leakage states and use of state probabilities. They also use graph reduction techniques and nonlinear simulation to speedup the simulation time with desired accuracy.

In most of the leakage estimation techniques only stand-by mode leakage currents have been considered. But the gates dissipate leakage power even during the active modes also [NDOC+3] [SSB04]. Hence, for accurate estimation of leakage power one have to consider the leakage occurring in the active or run-time mode as well. Run time leakage mitigation schemes [WCJR+98] [ASPB+97] have been proposed to dynamically change circuit conditions in response to low and high leakage situations. [ASPB+97] introduces a method to identify the Minimum Leakage Vector (MLV) for leakage power reduction. Since the leakage power depends on the input pattern, the central idea in [ASPB+97] is to apply minimum leakage producing input combination to the circuit when it is in the inactive mode.

Circuits which are in active mode most of the time, or switches frequently between active and stand-by modes will also have nodes that would leak significantly during the active mode. And, this leakage would be data dependent. This component is clearly captured by Nguyen et al. [NDOC+03], where the static component of power during active region is dependent on (1-α) where α is a measure of activity. They proposed a linear programming (LP) based optimization framework for simultaneous assignment of threshold voltage and sizing. They also proposed a dual V_{DD} extension of the problem by LP formulation. A heuristic is used

in [SSB04] for dual V_{DD}, dual V_{TH} and sizing. They presented optimization for three different switching scenarios.

5.2.3 Existing Total Power Estimation Techniques

The total power dissipation during the active mode is reduced with scaling. Further power savings can be achieved by controlling the subthreshold leakage currents. Due to scaling the subthreshold leakage currents will increase and thus the total power dissipation increases with micro-miniaturizations. Similarly, for low-power scaling, the optimum energy point for V_{DD} and V_{TH} will correspond to a larger subthreshold component. Moreover, during the standby mode, the standby power dissipation tends to increase since the leakage currents are large [KC00]. Therefore, to develop a power efficient system, the power dissipation in the active and standby modes must be minimized. To reduce dynamic power dissipation in the active period, traditional low-power circuit techniques such as pipelining and parallelism can be used in order to lower the supply voltage [CB95].

Standby subthreshold leakage currents are especially detrimental in burst mode type circuits, where computation occurs only during short bursty intervals, and the system is inactive for the majority of the time, while waiting for the next instruction. This problem is especially severe for portable electronics, where battery power is drained unnecessarily during long idle periods. Therefore, subthreshold leakage reduction techniques during the standby mode can significantly reduce the overall energy consumption.

Bhavnagarwala et al. [BDAM00] presented a "Transregional" MOSFET model for minimizing total power. Their methodology permits a complete evaluation

of tradeoffs between saturation drive current and subthreshold leakage current for a prescribed cycle time performance and operating temperature range. Optimal supply voltages, device threshold voltage and device channel widths, corresponding to minimum total power, are calculated out to the year 2012 for local and global critical paths. Calhoun and Chandrakasan modified transregional current equation to provide a numerical model for large circuits. They showed that the minimum energy point depends on the technology, the characteristic of the design and on operating conditions such as temperature, duty cycle and workload [CC04].

Next section presents analysis of the proposed subthreshold current model including impact ionization to study the power dissipation in subthreshold CMOS operation and design of low-power CMOS logic.

5.3 A Simplified Analysis for Minimum Power and Energy

In chapter 2 we noted that the power dissipation in CMOS circuits is mainly due to dynamic (switching and short-circuit) current and the subthreshold leakage current. With the scaling down of device sizes and transistor threshold, the subthreshold current will increase and can become a sizable component of total power dissipation. However, in the current-day technology, the dominant component power is still the switching component (about 80% of the total power dissipation). In this analysis it is assumed that the input rise and fall times are negligibly small and hence short-circuit power dissipation is negligible. Also, we have assumed simultaneous variations of signal and if all signals of a gate change simultaneously, no glitching occurs. Therefore, in this chapter we devoted a considerable effort in estimation of switching and subthreshold leakage powers in CMOS.

5.3.1 Optimal Sizing for Minimum Energy at a Minimum V_{DD}

Due to technology scaling, performance of digital and battery-operated systems has been increased tremendously even though at the cost of higher power consumption. Energy efficient operation has therefore become very pressing issue. Most applications do not always require the peak performance from the processor. Hence, in a system with fixed performance level, certain tasks complete ahead of their deadline and the processor enters a low-leakage sleep mode for the remainder of time. As the processor frequency is reduced, the supply voltage can be reduced. As shown by the equations below, the reduction in frequency combined with a quadratic reduction from supply voltage results in an approximately cubic reduction of power consumption. However, with reduced frequency the time to a task increases, leading to an overall quadratic reduction in energy to complete a task.

$$Delay = \frac{1}{f} = \frac{C_{load} V_{DD}}{2 I_{DS,on}} \qquad Power \propto f V_{DD}^2 \qquad Energy \propto V_{DD}^2$$

Therefore study of energy variation is preferred instead of power variation. It has been proposed that theoretically optimal minimum energy circuits should use minimum sized devices [Vit94]. Sizing influences the energy consumption of a circuit in two ways. First, sizing directly affects energy consumption by changing switched capacitance and leakage current. Secondly, the minimum voltage at which the circuit functions and which can change the absolute minimum energy point. Minimum V_{DD} operation occurs when the pMOS and nMOS devices have the same current [SS96].

As proposed in [FS93] expression for subthreshold drain current is

$$I_D \approx \frac{qW\delta D_n N_C}{L} \exp\left(\frac{\psi_s^0}{\eta V_T}\right)\left(1 - \exp\left(\frac{-V_{DS}}{V_T}\right)\right) \qquad (5.1)$$

Now assuming $V_{bi} \approx 2\varphi_b$ and neglecting the DIBL effect (since for lower supply voltage DIBL is very small), the basic equation after substituting values of D_n, δ and ψ_s^0 for modeling subthreshold current is

$$I_D \approx \frac{qW\mu_{eff}V_T N_C}{L} V_T \sqrt{\frac{\varepsilon_s}{2qN_a(2\varphi_b + V_{GT}/\eta)}} \exp\left(\frac{-V_{bi} + 2\varphi_b + V_{GS} - V_{TH0}}{\eta V_T}\right)\left(1 - \exp\left(\frac{-V_{DS}}{V_T}\right)\right) \tag{5.2}$$

$$\text{or, } I_D \approx \frac{\mu_{eff}WC_{GB}}{L} V_T^2 \exp\left(\frac{V_{GS} - V_{TH0}}{\eta V_T}\right)\left(1 - \exp\left(\frac{-V_{DS}}{V_T}\right)\right) \tag{5.3}$$

Substituting (5.3) in (3.8), sub-threshold drain current with impact ionization is

$$I_{DS} = MI_{DS,on} \exp\left(\frac{V_{GS} - V_{TH0}}{\eta V_T}\right) \tag{5.4}$$

where $I_{DS,on}$ is the drain current when $V_{GS} = V_{TH0}$ [RP00], given as

$$I_{DS,on} = \mu_{eff} C_{GB} \frac{W}{L} V_T^2 \left(1 - \exp\left(\frac{-V_{DS}}{V_T}\right)\right). \tag{5.5}$$

In (5.5), μ_{eff} and C_{GB} are the effective mobility and effective gate – body capacitance respectively. In the subthreshold condition C_{GB} is determined by depletion charge and oxide – substrate interface charge density.

For a step input signal the delay can be modeled as

$$\tau_{PHL} = \frac{C_{load}V_{DD}}{2I_{DS,onn}} \tag{5.6}$$

$$\tau_{PLH} = \frac{C_{load}V_{DD}}{2I_{DS,onp}} \tag{5.7}$$

Where $I_{DS,onn}$ and $I_{DS,onp}$ are $I_{DS,on}$ currents for n – MOS and p – MOS respectively. From (5.5) it is clear that V_{DS} has very little effect on $I_{DS,on}$. Assuming, $L_n = L_p$ and $C_{GBn} = C_{GBp}$, from (5.5) – (5.7)

$$\frac{\tau_{PHL}}{\tau_{PLH}} = \frac{\mu_{effn}W_n}{\mu_{effp}W_p} \tag{5.8}$$

For symmetrical output:

$\tau_{PHL} = \tau_{PLH}$

i.e. $\frac{W_p}{W_n} = \frac{\mu_{effn}}{\mu_{effp}}$ (5.9)

As μ_{effn} is about $3\mu_{effp}$, from (5.9) it follows that $W_p = 3W_n$ is a requirement in case of subthreshold inverters. Same was shown to be true in case of above threshold inverters [KL96]. Data given in table 5.1 when used to calculate $\frac{W_p}{W_n}$ through (5.1) – (5.9) also results in the value of 3.

Transistor sizing also affects the functionality of CMOS circuits at low supply voltages. Equation (5.10) shows the propagation delay of a characteristic inverter with a certain switched capacitance $C_{eff} = \alpha C_{load}$ in subthreshold, where α is activity factor and C_{load} is the load capacitance assuming average fan-out:

$$t_{delay,step} = \frac{0.5 C_{eff} V_{DD}}{I_{DS,on}} \quad (5.10)$$

The expression for current in the denominator of (5.10) models the ON current of the characteristic inverter, so it accounts for transitions through both nMOS and pMOS devices i.e. denotes the ideal inverter delay with step input and $t_{delay,actual}$ can be given as [ZBSF04]

$$t_{delay,actual} = \eta t_{delay,step} \quad (5.11)$$

To calculate the minimum size of transistor we have chosen a minimum delay to width ratio, i.e. $\left(\frac{\partial t_{delay}}{\partial W_n}\right)_{\min} = \left(\frac{t_{delay1} - t_{delay2}}{W_{n1} - W_{n2}}\right)_{\min}$, analytically and corresponding minimum size of device has been found with fixed V_{DD} and V_{TH} (which is at the approximate saturation point of the delay versus width of transistor plot). Finally, obtained minimum size and delay is plotted in Figure 5.3, for each value of V_{DD}.

Bhavnagarwala et al. [BDAM00], also proposed a methodology for optimizing device size. They have obtained the normalized propagation delay versus the supply to threshold voltage ratio γ, with fixed V_{TH0}, W_n and W_p to illustrate T_{cycle} saturation

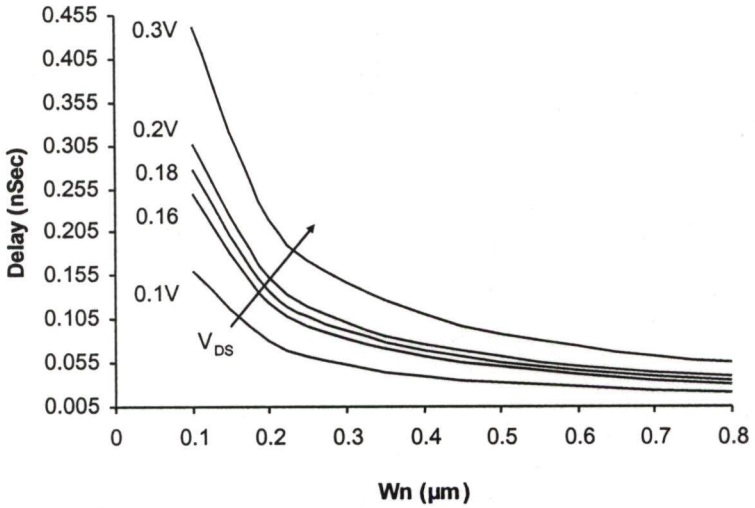


Figure 5.3: Sizing an inverter to have a minimum delay gives a good estimate of the optimum width to minimize V_{DD} because the nMOS and pMOS are balanced

at a $\frac{V_{DD}}{V_{TH0}} = \gamma \approx 6$ .The optimum $\frac{V_{DD}}{V_{TH0}}$ ratio was chosen at the approximate saturation point of the T_{cycle} curve, which occurs somewhat beyond a slope of negative one. It has limitations that increasing V_{DD} beyond five times V_{TH0} yields diminishing improvements in performance. But, in our case no such limitation as we have chosen the values for minimum size.

Minimum sized devices generally minimize energy consumption in subthreshold for a given V_{DD}. However, sizing also impacts the minimum operating voltage, which can affect the total energy per operation, E_{total}. For a single inverter, dynamic ($E_{dynamic}$), leakage (E_{leak}), and total energy (E_{total}) per cycle are expressed in (5.12)-(5.14), assuming rail-to-rail swing ($V_{GS} = V_{DD}$ for "on" current).

$$E_{dynamic} = n\frac{1}{2}C_{eff}V_{DD}^2 \tag{5.12}$$

$$E_{leak} = (nI_{DS}V_{DD})(nt_{delay,actual}) \tag{5.13}$$

$$E_{total} = n\frac{1}{2}C_{eff}V_{DD}^2 + (nI_{DS}V_{DD})(nt_{delay,actual}) \tag{5.14}$$

where n is the number of stages. Similarly, total power can be given as

$$P_{total} = n(C_{eff}V_{DD}^2 f + nI_{DS}V_{DD}) \tag{5.15}$$

Total energy per cycle is proportional to C_{eff}, so minimum sized devices give a minimum C_{eff} and minimize E_{total} for a single inverter at a given V_{DD}.

The energy dependence on supply voltage using a simple inverter chain consisting of 20 inverters has been analyzed. A single transition is used as a stimulus and energy is calculated over the time period necessary to propagate the transition through the chain.

5.3.2 Model Verification

In order to verify the accuracy of the proposed model, we compared the results obtained from (5.15) with SPICE simulations for inverter chain. In Figure 5.4, we compare the energy-V_{DD} relationship predicted by the proposed analytical model in the subthreshold region with SPICE simulation results for an industrial 100nm process. As shown in Figure 5.4, the analytical model results match with SPICE results well. The energy-V_{DD} relation plotted in Figure 5.4 also indicates that the dynamic energy, $E_{dynamic}$, reduces quadratically while the leakage energy, E_{leak}, increases with voltage scaling.

The increase in leakage energy in subthreshold operating regime is due to the voltage scaling below the threshold voltage. The on-current (and hence the

circuit delay) increases exponentially with voltage scaling while the off-current reduces less strongly.

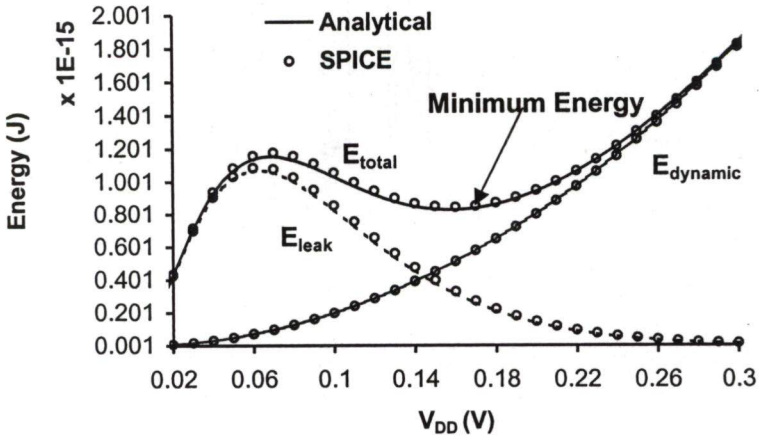


Figure 5.4: Comparison of analytical results with SPICE results for 100nm device. Minimum energy occurs at V_{DD}(min)=0.16V

Hence, the leakage energy, E_{leak}, will increase and supersede the dynamic energy, $E_{dynamic}$, at about 150 mV. This creates a minimum energy point in the inverter circuit that lies at about 160mV (Figure 5.4).

5.3.3 V_{DD} - V_{TH} Optimization Overview

For a given chip or circuit, it is possible to choose an appropriate V_{DD} and V_{TH} combination, that optimizes the combination of dynamic power and leakage power for a given performance. Typically, for a fixed performance requirement, there is a locus of V_{DD} and V_{TH} combinations that could be used. As the supply voltage drops, the dynamic power is reduced quadratically, but as the threshold voltage drops the leakage power is increased exponentially. The optimum V_{DD} and V_{TH} combination corresponds to the point where the incremental decrease in dynamic power with a change in V_{DD} is offset by the incremental increase in leakage power due to change

in V_{TH} [ST95]. J. T. Kao and A. Chandrakasan [KC00] showed that for a given circuit operating condition there is an optimal supply voltage and threshold voltage operating point that minimizes the overall energy. For the active leakage reduction fine grain adaptive body biasing technique is proposed in which a system can be divided into individual local regions (locally compensated). Unfortunately, this fine grain control cannot be directly applied to supply voltage scaling because of interfacing between local logic block and different voltage generated for each local logic block. As a result, V_{DD}/V_{TH} optimization is geared towards chip level control where the die as a whole is tuned to have a global optimum supply and global optimum threshold voltage. Although theoretically different sub-blocks in a system could operate with their own fine grain V_{DD}/V_{TH} operating point and results in lower power levels. By combining the adaptive body biasing scheme, for local block control, with a global V_{DD}/V_{TH} optimization scheme, for the whole chip one can achieve fine grain control over local block. This may be done by first finding an optimal V_{DD}/V_{TH} operating point, for the entire chip, and then selectively applying reverse body-biasing as necessary.

After global V_{DD}/V_{TH} scaling further reduction in subthreshold leakage currents during the active mode may be done by selectively slow down gates (within a block, or within the whole chip) which are not in the critical paths. Technically, it is not possible because only the critical paths need to scale aggressively and non-critical gates within these circuit blocks can still be operated with higher threshold voltages. As described earlier, the optimal bias setting corresponds with the point on the V_{DD}/V_{TH} locus (for a fixed performance), where the incremental change in

dynamic power (due to V_{DD} scaling) is compensated by the incremental change in leakage power (due to V_{TH} scaling).

On comparing the results shown by [KC00] with our plot it is seen the nature of curves shown in fig 5.5 are same. Figure 5.5 below shows performance curves for power versus supply voltage, where the threshold voltage is implicitly a one-to-one function of V_{DD}. The power curves are divided into two parts. The right curve $P_{dynamic}$ shows dynamic power versus V_{DD} while the left curve P_{leak} shows leakage power versus V_{DD}. The sum of these two curves thus provides the total power dissipation for the circuit.

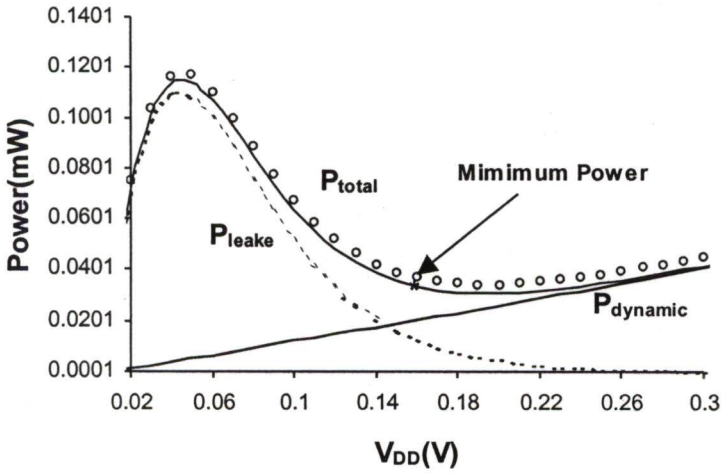


Figure 5.5: Optimal V_{DD}/V_{TH} biasing point trading off dynamic power with leakage power for constant performance

The minimum of the total power curve corresponds to the point where the slope of the two curves are opposite in sign but equal in magnitude. It is important to note that the minimum operating point isn't necessarily located where the two curves intersect, but rather where the two curves have equal and opposite slopes. This is true because the minimum power point occurs in a region where one curve is

monotonically increasing while the other curve is monotonically decreasing, and the slopes change monotonically as well. The plot shows that for the chain of 20 inverters minimum power dissipation is 33μW at V_{DD} of 160mV.

Figure 5.6 shows the energy versus supply voltage plot for different t_{ox}. As we already discussed that the multiplication factor given by (3.14) is dependent on gate-oxide thickness through the parameters of l_d and E_m; and undesirable increase in drain current owing to impact-ionization can be checked by the use of thicker gate oxide. Figure 5.6 also depicts that energy minima shifts towards higher V_{DD} as t_{ox} increases.

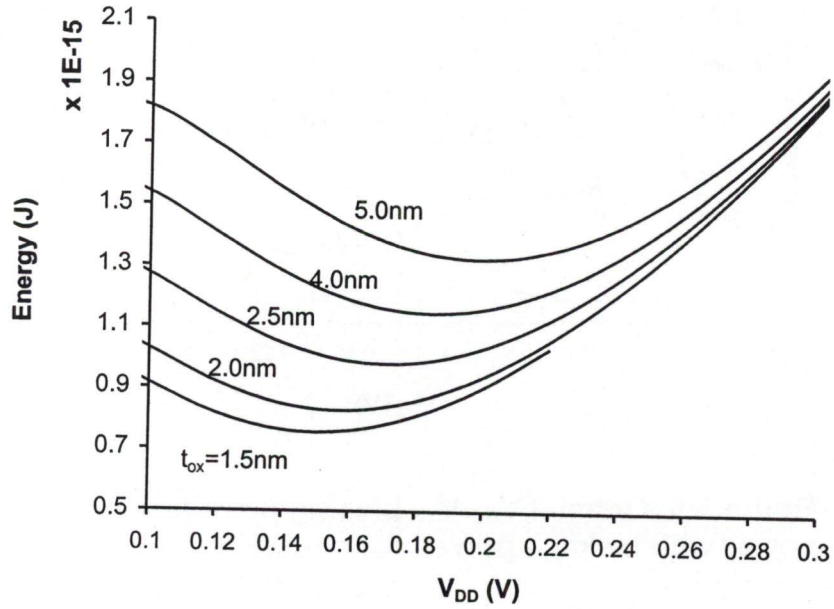


Figure 5.6: Energy versus supply voltage plot with oxide thickness as a parameter

5.3.4 Role of Switched Capacitance on Optimal V_{DD}-V_{TH} Scaling

The switched capacitance C_{eff} corresponds to the average capacitance that is switched per cycle. When this value increases, the dynamic power consumption increases without directly impacting leakage currents. As can be seen in the equation for total power consumption (5.15), an increase in C_{eff} will correspond to an increase in the dynamic power equation, and the equilibrium point will shift as seen in the Figure 5.7 below.

As a circuit becomes more dynamic power dominated (by increasing the switching capacitance), the optimal V_{DD}-V_{TH} point moves towards lower supply voltages and lower threshold voltages. This effectively weights the circuit more strongly with leakage currents because lowering supply voltages can more effectively control the larger dynamic power component.

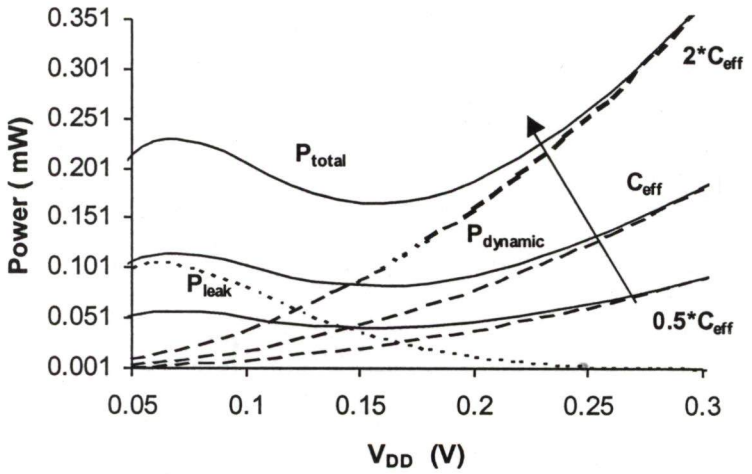


Figure 5.7: Effect of switched capacitance variation on optimal V_{DD}-V_{TH} biasing point

5.3.5 Energy Delay Product

Several different metrics have been used in the literature to characterize low power circuit performance [GGH97]. Power delay products, energy delay products, and average power consumption have all been useful measures for comparing different circuits because energy consumption as well as circuit speed both is taken into account in the metric.

Figure 5.8 shows another important advantage of subthreshold circuit: lower energy delay product (EDP). The EDP for V_{DD} =0.1V is about 4.85 times smaller than the EDP for V_{DD} =0.3V. The lower EDP is because the reduction in power consumption outweighs the increase in delay. Having a lower EDP means the subthreshold circuit consumes less energy than the normal strong inversion circuit when both circuits operate with the same amount of switching activities. This reduction is expected, as the energy per switching activity (Power delay product) is proportional to $C_{eff}V_{DD}^2$.

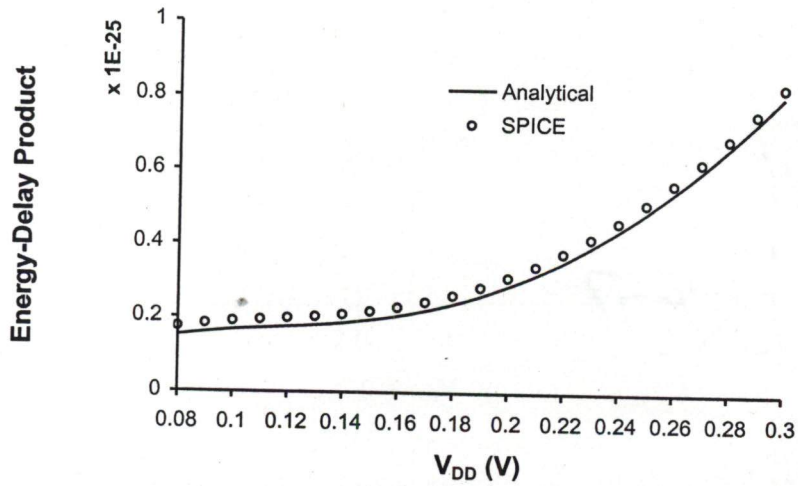


Figure 5.8: Energy-Delay Product for various power supply voltage

Table 5.1: Parameter Values used in comparing the subthreshold current 100nm MOSFET [KPW02] with SPICE. Some other parameters are: $A_n = 7.03*10^7\ \text{cm}^{-1}$, $B_n = 1.231*10^8\ \text{V/cm}$, $\alpha = 0.2$ $n = 20$ and $C_{load} = 10fF$

Parameters	nMOS	pMOS
L	100nm	100nm
ε_s	1.04×10^{-10} Fm^{-1}	1.04×10^{-10} Fm^{-1}
ε_i	3.36×10^{-11} Fm^{-1}	3.36×10^{-11} Fm^{-1}
$N_a(N_d)$	1.5×10^{24} m^{-3}	1.8×10^{24} m^{-3}
N_C	3.0×10^{26} m^{-3}	3.0×10^{26} m^{-3}
t_{ox}	2.0 nm	2.0 nm
V_{TH0}	0.3305 V	-0.34 V
μ_0	0.018 m^2/V-sec	0.0058 m^2/V-sec
η	1.21	1.22
x_j	11.1×10^{-8} m	11.1×10^{-8} m

5.3.6 Voltage Transfer Characteristic and Noise Margin

Considering the fact that in a CMOS inverter operating in steady state, the subthreshold drain current of nMOS transistor is always equal to subthreshold drain current of the pMOS transistor,

$I_{DS,n} = I_{DS,p}$ or,

$$\kappa_n V_T^2 e^{\left(\frac{V_{GS,n}-V_{TH0,n}}{\eta_n V_T}\right)}\left(1-e^{-\left(\frac{V_{DS,n}}{V_T}\right)}\right) = \kappa_p V_T^2 e^{\left(\frac{V_{GS,p}-V_{TH0,p}}{\eta_p V_T}\right)}\left(1-e^{-\left(\frac{V_{DS,p}}{V_T}\right)}\right) \tag{5.16}$$

where $\kappa_n = \mu_n C_{ox}\frac{W_n}{L_n}$ and $\kappa_p = \mu_p C_{ox}\frac{W_p}{L_p}$.

Now referring figure 5.2, we may define

$V_{GS,n} = V_{in}$;

$V_{DS,n} = V_{out}$;

$V_{GS,p} = V_{in} - V_{DD}$; and

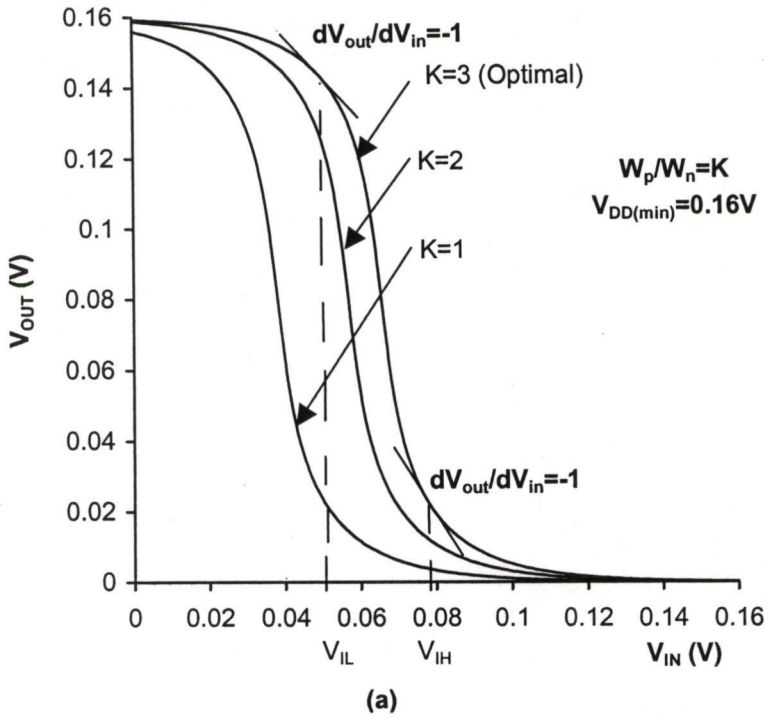
$V_{DS,p} = V_{out} - V_{DD}$.

Substituting above values in (5.16), we have

$$\kappa_n V_T^2 e^{\left(\frac{V_{in}-V_{TH0,n}}{\eta_n V_T}\right)}\left(1-e^{-\left(\frac{V_{out}}{V_T}\right)}\right) = \kappa_p V_T^2 e^{\left(\frac{V_{in}-V_{DD}-V_{TH0,p}}{\eta_p V_T}\right)}\left(1-e^{-\left(\frac{V_{out}}{V_T}\right)}\right) \quad (5.17)$$

Now, taking the different values of V_{in} corresponding V_{out} has been found out by iterations for different values of $K = \frac{W_p}{W_n}$, and voltage transfer characteristics (VTC) are plotted as shown in figure 5.9(a). The VTC of the inverter gate running in subthreshold mode is closer to ideal compared to the VTC in normal strong inversion region.

The exponential relationship between I_{DS} and V_{GS} in subthreshold region gives rise to an extremely high transconductance g_m, i.e. also exponentially related to V_{GS}. The much-improved VTC yields better noise margin. Figure 5.9 (a) shows the VTCs at the minimum V_{DD} of 160mV for several $K = \frac{W_p}{W_n}$ ratios. By definition, the slope of the VTC is equal to (-1), i.e. $\frac{dV_{out}}{dV_{in}} = -1$. To calculate the values of V_{IL} (lower value of V_{in}) and V_{IH} (higher value of V_{in}), ratio $\frac{dV_{out}}{dV_{in}} = \frac{V_{out2} - V_{out1}}{V_{in2} - V_{in1}}$ have been found out, using above calculated values, numerically for different values of V_{DD} and K. The sum of V_{IL} and V_{IH} is always equal to V_{DD} in a symmetric inverter, which are as shown on VTC plot.



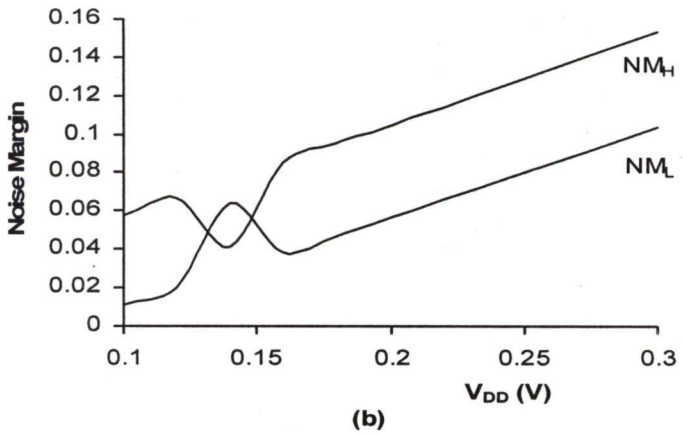


Figure 5.9: (a) VTC and (b) noise margins (high and low) at the minimum V_{DD} (simulation). Optimum pMOS/nMOS width ratio of 3 is shown with others for comparison

$$V_{IL} + V_{IH} = V_{DD} \tag{5.18}$$

Finally, we found the noise margins NM_L and NM_H for low voltage levels and high voltage levels using,

$$NM_L = V_{IL} - V_{OL} = V_{IL} \text{ and,} \tag{5.19}$$

$$NM_H = V_{OH} - V_{IH} = V_{DD} - V_{IH} \tag{5.20}$$

Results so obtained using (5.19) and (5.20) are shown in figure 5.9(b). The gain is somewhat degraded, but the optimum sized curve is symmetrical and shows poor noise margins.

5.4 Conclusions

We proposed an analytical model for most energy efficient supply voltage $V_{DD,\min}$ for CMOS circuits. A number of interesting conclusions can be drawn: (1) Energy shows clear minimum in the subthreshold region since the time over which a circuit is leaking (delay) grows exponentially in this region while leakage current itself does not drop rapidly with reduced V_{DD}, (2) $V_{DD,\min}$ does not fully depends on V_{TH0}, (3) minimum sized devices are theoretically optimal for minimizing power and dependency of delay on input transition time, and (4) as the optimum V_{DD} is a highly scaled voltage, the effect of optimizing V_{DD} as noise margin is also studied. It shows that the power minimization is achieved as the cost of noise margin.

The result shows that for a simple inverter chain consisting of 20 inverters minimum power dissipation as low as 33μW can be achieved at V_{DD} of 160mV at V_{TH0}=0.33 V.

The analytical model presented is shown to match very well with SPICE simulations.

CHAPTER 6

Conclusions

Power and energy efficient designs have become a necessity for modern VLSI technologies. With doubling integration capacity every two to three years, power dissipation presents a real threat for reliability and even functionality of the devices. As a result, tremendous effort has been devoted to achieve lower power dissipation without affecting performance. The main components of power dissipation are switching power and leakage power. Switching power, being the dominant power component, has caught special attention in recent years. Many techniques have been introduced to control this ever-increasing power component on all levels of design abstraction. Increased leakage current due to technology feature downsizing is another challenge that faces circuit designers in the VLSI and ULSI era.

Literature survey shows that subthreshold leakage currents can consume upwards 30% of the total power budget in submicron and deep submicron technology. New design considerations, which have been described in the thesis, must be developed to control subthreshold currents in both the active and standby modes in order to provide low power solutions. Fortunately, industry has begun to respond to this problem and subthreshold leakage and impact ionization has become an active research area.

6.1 Summary of Important Findings and Contributions

A large contribution of this thesis work was to add to the state-of-the-art of scaling techniques and impact ionization in submicron and deep submicron MOSFETs.

To assess the magnitude of electric field in the channel of a submicron MOSFET, the example of 90nm and 100nm devices were taken. The calculated fields turn out to be well within impact ionization range near the drain. To include the effect of impact ionization on drain current, multiplication factor was first determined using well-established expressions. The drain-source voltage (V_{DS}) dependence of multiplication factors for subthreshold, linear and saturation regions of operation of a 100nm device were compared. It has been observed that, subthreshold multiplication factor is highest and linear region has the lowest multiplication factor. The analysis shows that for subthreshold operation it is important that the effect of impact ionization should be included in the analytical expression for drain current.

The multiplication factor was expressed as a function of maximum of the channel field, which occurs at the drain end. The maximum field increases with V_{DS}.

As a result multiplication factor increases with V_{DS}. Calculations show that multiplication factor also increases with increase of gate voltage and decreases with increase in gate oxide thickness. This shows increase in gate field increases multiplication factor. At lower values of V_{DS}, variation of multiplication with oxide thickness was nominal. Calculations also show that multiplication factor increases with increase in drain/source pn-junction depth. However, multiplication factor levels off at larger junction depths. Shallower junctions were therefore desirable if carrier multiplication is to be restricted.

In case of a subthreshold CMOS logic circuit, the input voltage swing is restricted to be always below the threshold voltage of the MOSFETs used. Thus, in such a circuit, it is the threshold current that flows between the rails. Fjeldly and Shur developed a very useful physical model for subthreshold current. While this model yields accurate results for submicron devices, it does not take into account the effect of impact ionization, which is more prominent in devices of nanometer channel length. As technology scaling has made nanometer devices to be the more commonly used ones, the present study considered such devices. The subthreshold current model of Fjeldly and Shur was therefore, first suitably modified to include the effect of impact ionization, for all analytical purposes to be discussed subsequently. The modified current model agrees well with results obtained from ATLAS simulation. It shows that impact ionization increases subthreshold current by few percents. Such an increase in subthreshold current model, the DIBL –coefficient was reconsidered. It has been found that this parameter is not independent of the terminal voltages of a MOSFET. The DIBL- coefficient has a strong dependence on V_{DS} and relatively lesser dependence on gate-source voltages V_{GS}. This analysis shows that, because of the strong V_{DS} dependence of the DIBL-coefficient, the

threshold voltage shift is a non-linear function of V_{DS}, in contradiction to a constant DIBL-coefficient as assumed previously. It has also been found that DIBL- coefficient increases with increase in channel length and decreases with increase in temperature.

As mentioned already, there are three basic methods of device scaling in VLSI technology. One or the other of these scaling methods is many often adopted to obtain a small geometry device from a large geometry one of known behavior. To have an understanding of how the application of the scaling methods affects the subthreshold conduction of a small geometry device, the dependence of subthreshold current on applied scaling technique has been analyzed. The results show that as the channel length is scaled down, multiplication factor increases slowly in the higher regime of channel length. In the lower regime of channel length, multiplication factor rises rapidly. This result also justifies the inclusion of impact – ionization effect on subthreshold current. Similar variations in subthreshold current with channel length scaling were shown by the analytical results. However, the analysis shows that there is insignificant dependence of multiplication factor on the method of scaling.

Another factor, that affects subthreshold current is DIBL, it has been found that, this effect is dependant on the scaling technique adopted. Such dependence of DIBL is more significant in the higher regime of V_{DS}. Constant electric field scaling produces the largest threshold voltage shift due to DIBL. Selective scaling has the least DIBL, with generalized scaling produced DIBL falling in between the two extremes. As a result, the scaling dependence of subthreshold current follows the same pattern.

The subthreshold current model has been finally used to a study of power dissipation in subthreshold CMOS operation and design of low-power CMOS logic. It is well known that in above threshold CMOS operation, a tradeoff exists between propagation delay and power dissipation, when voltage is scaled. When the supply voltage is scaled down to reduce power dissipation, propagation delay is increased. SPICE simulation, however, shows that this does not hold good for subthreshold logic. For a threshold CMOS– logic operation, propagation delay in most cases first increases and then decreases with supply voltage scaling. The subthreshold current model mentioned above has been used to calculate voltage dependence of propagation delay of a CMOS inverter. The results thus obtained show that, the delay increases with increase in supply voltage. An important requirement of a logic circuit is symmetric input-output characteristics. The ratio of pMOS and nMOS widths that meets this requirement is determined using the modified subthreshold current model. The ratio thus obtained is primarily mobility dependent, as in the case of above threshold logic operation.

A final contribution of this research was to provide a framework for exploring optimal V_{DD} for minimizing total power dissipation.

The subthreshold operation is applicable to only those cases where delays in the logic circuits are not of primary importance. It is the low-power operation that is of major concern. Even then, it is desirable to minimize delay as far as it is possible. Keeping this in view, a low-power CMOS-design method has been proposed. In this method, for a given delay constraint, the nMOS width on V_{DD} was first determined. Using the information thus obtained optimum V_{DD} is obtained by minimizing total power dissipation. The result shows that for a 100nm technology, power dissipation as low as 33μW can be achieved at V_{DD} of 160mV. While, determining the optimum

V_{DD}, the requirement for the symmetry of input-output characteristic is also taken into account. As the optimum V_{DD} is a highly scaled voltage, the effect of optimizing V_{DD} as noise margin has also been studied. It depicts that the power minimization can be achieved at the cost of noise margin. The analytical model presented was shown to match very well with SPICE and ATLAS simulations.

6.2 Suggestions for Future Work

This thesis touched upon the effects of impact ionization on subthreshold current. Since it has been a very crucial issue for power minimization thus more efforts should be under taken for further research work in this area.

The challenge of power dissipation goes hand-in-hand with that of power delivery. Subsequent to the computation, power delivered to CMOS VLSI circuits gets dissipated as heat. Increase in power delivered with scaling results in increased power dissipation and higher power density.

In order to maintain junction temperature constant with increased power dissipation it maybe necessary to use more exotic cooling and enhanced heat spreading solutions such as carbon nano-tubes and electro-kinetic micro-channel cooling. Since sub-threshold leakage power will become more dominant with scaling, total power dissipated will have strong junction temperature dependence. Therefore, instead of keeping junction temperature constant with scaling, it might be beneficial to decrease the temperature. This temperature scaling will, not only, reduce leakage power but also improve drive current, interconnect resistance and reliability. Further optimization of device and circuits for low temperature operation can provide additional scaling benefits. Main challenges to achieve temperature scaling for future

CMOS generations include: (i) understanding the relationship between junction temperature, total power, reliability, and speed and (ii) invention of low-cost integrated cooling solutions.

APPEDIX-1

90nm Device Parameters used in ATLAS Simulation

```
//  DEVICE 1
simulator lang=spectre insensitive=yes model n_10_sp bsim4 type=n

+ version=4.3000e+00          binunit=1.000e+00          paramchk=1.000e+00
+ mobmod=0.0000e+00           capmod=2.0000e+00           igcmod=1.0000e+00
+ igbmod=1.0000e+00           geomod=0.0000e+00           diomod=2.0000e+00
+ rdsmod=0.0000e+00           rbodymod=0.000e+00         rgatemod=0.000e+00
+ permod=1.0000e+00           acnqsmod=0.000e+00         trnqsmod=0.000e+00
+ rgeomod=1.0000e+00          fnoimod=1.000e+00          tnoimod=0.0000e+00
+ toxe=2.2500e-09 + dtoxe_n_10_sp
+ toxp=1.8220e-09 + dtoxp_n_10_sp                         toxm=2.2500e-09
+ epsrox=3.9000e+00           xj=1.2000e-07              ngate=1.3000e+20
+ ndep=1.0000e+17             nsd=1.0000e+20             rsh=8.0000e+00
+ wint=2.0210e-08 + dwint_n_10_sp                        lint=-4.0910e-09
+ vth0= - 1.0000e-03 + dvth0_n_10_sp                     k1=1.5690e-01
+ k2=4.0000e-03               k3= - 1.2880e+00 + dk3_n_10_sp
+ k3b=2.9280e+00              w0=9.0000e-08              dvt0=3.9630e+00
+ dvt1=5.6320e-01             dvt2=-3.3200e-02           dvt0w=5.2510e-01
+ dvt1w=1.1170e+07            dvt2w=-7.7000e-01          dsub=3.9000e-02
+ minv=7.7040e-01             voffl= - 4.9270e-09 + dvoffl_n_10_sp
+ dvtp0=8.9100e-06            dvtp1=-8.0630e-01
+ lpe0=1.0000e-10 + dlpe0_n_10_sp                        lpeb=-1.6990e-07
+ phin=8.7670e-02             cdsc=4.6490e-04            cdscb=1.5000e-04
+ cdscd=0.0000e+00            cit=1.5520e-03             voff=-6.3870e-02
+ nfactor=1.0000e-01          eta0=5.0000e-05            etab=-1.8500e-04
+ vfb=-1.0000e+00             u0=2.3200e-02              ua=-1.5500e-09
+ ub=3.4800e-18               uc=1.7330e-10              vsat=1.6250e+05
+ a0=8.8340e+00               ags=1.0020e+00             a1=0.0000e+00
+ a2=1.0000e+00               b0=0.0000e+00              b1=1.0000e-08
+ keta=-4.4080e-02            dwg=-5.4000e-09            dwb=4.8000e-09
+ pclm=1.0000e-01             pdiblc1=1.0000e-07         pdiblc2=3.954e-02
+ pdiblcb=1.0000e-01          drout=5.5990e-01           pvag=8.6180e+00
+ delta=1.0000e-02            pscbe1=6.5350e+09          pscbe2=3.3000e-01
+ fprout=1.0000e-02           pdits=6.1100e-01           pditsd=8.8000e-01
+ pditsl=1.0000e+05           rdsw=5.0000e+01 + drdsw_n_10_sp
+ rdswmin=5.0000e+01          prwg=2.8000e-01            prwb=4.4700e-01
+ wr=1.0000e+00               alpha0=2.0000e-07          alpha1=4.0000e+00
+ beta0=1.5000e+01            agidl=1.1080e-08           bgidl=1.3900e+09
+ cgidl=2.9630e-01            egidl=9.4400e-01           toxref=2.2500e-09
+ dlcig=1.8000e-08            aigbacc=1.1980e-02         bigbacc=8.013e-03
+ cigbacc=6.2560e-01          nigbacc=4.3970e+00         aigbinv=1.530e-02
+ bigbinv=4.8520e-03          cigbinv=1.0000e-03       eigbinv=1.1000e+00
```

```
+ nigbinv=1.6000e+00        aigc=1.1380e-02          bigc=1.8790e-03
+ cigc=1.0000e-04           aigsd=9.8830e-03         bigsd=1.2690e-03
+ cigsd=1.5540e-01          nigc=1.000e+00          poxedge=1.0000e+00
+ pigcd=2.5000e+00          ntox=1.0000e+00          dlc=1.6400e-08
+ dwc=-3.0000e-08           xpart=1.0000e+00
+ cgso=5.0000e-11 + dcgso_n_10_sp
+ cgdo=5.0000e-11 + dcgdo_n_10_sp                    cgbo=0.0000e+00
+ cgdl=2.2000e-10 + dcgdl_n_10_sp
+ cgsl=2.2000e-10 + dcgsl_n_10_sp                    clc=1.0000e-07
+ cle=6.0000e-01            cf=9.2600e-11 + dcf_n_10_sp
+ ckappas=3.0000e+00        vfbcv=-1.0000e+00        acde=2.8080e-01
+ moin=1.1830e+01           noff=2.4860e+00         voffcv=-1.3720e-02
+ ef=0.9448
+ noia=3.8700000e+41        noib=1.8600000e+25      noic=6.7000000e+08
+ em=6.3600000e+06          ntnoi=1.0
+ xl= - 1.0000e-08 + dxl_n_10_sp
+ xw=0.0000e+00 + dxw_n_10_sp                        dmcg=1.6000e-07
+ dmci=1.0000e-07           dwj=0.0000e+00           jss=2.3350e-07
+ jsws=7.0330e-14           jswgs=3.2986e-14        ijthsfwd=3.445e-03
+ ijthsrev=1.6910e-03       bvs=1.1470e+01           xjbvs=1.0000e+00
+ pbs=6.1000e-01            cjs=1.0700e-03 + dcjs_n_10_sp
+ mjs=2.9000e-01            pbsws=9.9000e-01
+ cjsws=1.2600e-10 + dcjsws_n_10_sp                  mjsws=1.0000e-01
+ pbswgs=6.0000e-01         cjswgs=2.3100e-10 + dcjswgs_n_10_sp
+ mjswgs=9.8900e-01         tnom=2.5000e+01          kt1=-3.8000e-01
+ kt11=1.0000e-09           kt2=-4.0740e-02          ute=-1.0220e+00
+ ua1=4.3500e-09            ub1=-4.1040e-18          uc1=2.6360e-10
+ prt=0.0000e+00            at=1.0000e+05            njs=1.0560e+00
+ tpb=1.3000e-03            tcj=9.0000e-04           tpbsw=3.5150e-03
+ tcjsw=4.0000e-04          tpbswg=1.2470e-03        tcjswg=8.2290e-03
+ xtis=3.0000e+00           ll=4.3480e-16            wl=-4.0050e-15
+ lln=9.0000e-01            wln=1.0000e+00           lw=3.2080e-15
+ ww=-1.5010e-15            lwn=1.0000e+00           wwn=1.0000e+00
+ lwl=-1.6220e-21           wwl=1.7820e-22 + dwwl_n_10_sp
+ llc=-9.0100e-16           wlc=0.0000e+00           lwc=0.0000e+00
+ wwc=1.0000e-15            lwlc=0.0000e+00          wwlc=0.0000e+00
+ lmin=8.0000e-08           lmax=5.0000e-05          wmin=1.2000e-07
+ wmax=1.0000e-04           pvth0= - 1.2500e-03 + dpvth0_n_10_sp
+ lk3=7.2000e-01            wk3=-1.3000e-01          lk3b=-2.0000e-01
+ pk3b=2.0000e-02           ldsub=-1.2720e-03        wdsub=5.0000e-04
+ llpe0=3.8910e-08 + dllpe0_n_10_sp                  lcit=7.0000e-05
+ wvoff=-1.3400e-03         leta0=1.3000e-05         weta0=3.7760e-05
+ letab=8.2510e-06          wu0=2.4000e-04           pu0=-6.5000e-05
+ lub=-2.5220e-20           wub=-3.0000e-20          pub=-6.5270e-21
+ wuc=-5.5000e-12           pvsat=-7.3390e+02        lags=8.0000e-01
+ lketa=4.3920e-03          pketa=-5.0000e-04        ldelta=5.5800e-04
+ lvoffcv=-5.3220e-03       pkt1=1.0000e-03          lute=7.5240e-02
+ wute=2.5000e-02           pute=7.4000e-03          lub1=6.5000e-20
+ wuc1=-7.2000e-12          saref=1.7600e-06         sbref=1.7600e-06
+ wlod=0.0000e+00           kvth0=5.0000e-08         lkvth0=3.9000e-06
+ wkvth0=9.0000e-08         pkvth0=0.0000e+00       llodvth=1.0000e+00
+ wlodvth=1.0000e+00        stk2=0.0000e+00          lodk2=1.0000e+00
+ lodeta0=1.0000e+00        ku0=-1.5200e-08          lku0=-6.2900e-09
+ wku0=-1.0000e-08          pku0=1.2800e-15         llodku0=1.0500e+00
+ wlodku0=1.0000e+00        kvsat=9.9000e-01         steta0=-2.800e-11
+ tku0=0.0000e+00
```

```
//  ***User Drop Parameters*****************************************


// DEVICE 2
model p_10_sp bsim4 type=p
+ version=4.3000e+00          binunit=1.0000e+00       paramchk=1.0000e+00
+ mobmod=0.0000e+00           capmod=2.0000e+00          igcmod=1.0000e+00
+ igbmod=1.0000e+00           geomod=0.0000e+00          diomod=2.0000e+00
+ rdsmod=0.0000e+00           rbodymod=0.0000e+00        permod=1.0000e+00
+ acnqsmod=0.0000e+00         rgeomod=1.0000e+00       fnoimod=1.0000e+00
+ tnoimod=0.0000e+00          toxe=2.4500e-09 + dtoxe_p_10_sp
+ toxp=1.9110e-09 + dtoxp_p_10_sp                        toxm=2.4500e-09
+ epsrox=3.9000e+00           xj=1.2000e-07              ngate=1.0000e+20
+ ndep=3.6000e+16             nsd=1.0000e+20             rsh=8.0000e+00
+ wint=8.0090e-09 + dwint_p_10_sp                        lint=-2.1220e-08
+ vth0= - 5.8100e-02 + dvth0_p_10_sp                     k1=2.2500e-01
+ k2=-2.4750e-02              k3= - 8.8950e+00 + dk3_p_10_sp
+ k3b=3.9000e+00              w0=2.1220e-06              dvt0=4.6860e+00
+ dvt1=8.7290e-01             dvt2=1.2770e-02            dvt0w=3.0000e-01
+ dvt1w=3.9660e+06            dvt2w=2.4940e-01           dsub=1.0160e+00
+ minv=2.8230e-01             voffl= - 2.5000e-09 + dvoffl_p_10_sp
+ dvtp0=6.0620e-06            dvtp1=4.4890e-01
+ lpe0= - 1.2670e-07 + dlpe0_p_10_sp                     lpeb=6.2500e-08
+ phin=0.0000e+00             cdsc=0.0000e+00            cdscb=-8.0000e-03
+ cdscd=0.0000e+00            cit=2.7750e-04             voff=-1.2000e-01
+ nfactor=2.0000e+00          eta0=3.0000e-02            etab=-5.0310e-01
+ vfb=-1.0000e+00             u0=9.2600e-03 + du0_p_10_sp
+ ua=4.2790e-10               ub=1.1290e-18              uc=8.5910e-11
+ eu=1.0000e+00               vsat=1.3670e+05            a0=1.8600e+00
+ ags=1.4670e+00              a1=0.0000e+00              a2=1.0000e+00
+ b0=7.0000e-07               b1=6.0000e-07              keta=-5.1120e-02
+ dwg=-1.7240e-08             dwb=0.0000e+00             pclm=2.9400e-01
+ pdiblc1=5.1850e-08          pdiblc2=4.0800e-03         pdiblcb=-5.00e-01
+ drout=4.6980e-04            pvag=1.2960e+00            delta=2.3890e-03
+ pscbe1=6.3370e+09           pscbe2=3.0000e-03          fprout=3.0000e+02
+ pdits=2.9810e-01            pditsd=7.1760e-01          peditsl=5.0000e+05
+ rdsw=2.2500e+02 + drdsw_p_10_sp                      rdswmin=8.0000e+01
+ prwg=0.0000e+00             prwb=2.0000e-01            wr=1.0000e+00
+ alpha0=2.1400e-08           alpha1=7.0000e-02          beta0=1.2000e+01
+ agidl=4.4320e-09            bgidl=4.8080e+09           cgidl=9.1730e-03
+ egidl=-2.1800e+00           toxref=2.4500e-09          dlcig=3.2000e-08
+ aigbacc=1.1030e-02          bigbacc=6.7610e-03         cigbacc=5.770e-01
+ nigbacc=4.3960e+00          aigbinv=9.4660e-03         bigbinv=2.340e-03
+ cigbinv=1.8320e-03          eigbinv=1.6330e+00        nigbinv=3.1240e+00
+ aigc=6.7900e-03             bigc=8.8750e-04            cigc=6.3430e-04
+ aigsd=5.6520e-03            bigsd=7.8050e-05           cigsd=1.8030e-02
+ nigc=7.9250e-01             poxedge=1.0000e+00         pigcd=2.0000e+00
+ ntox=1.0000e+00             dlc=3.4200e-08             dwc=-3.0000e-08
+ xpart=1.0000e+00            cgso=4.2000e-11 + dcgso_p_10_sp
+ cgdo=4.2000e-11 + dcgdo_p_10_sp                        cgbo=0.0000e+00
+ cgdl=2.0000e-10 + dcgdl_p_10_sp
+ cgsl=2.0000e-10 + dcgsl_p_10_sp                        clc=1.0000e-07
+ cle=6.0000e-01              cf=9.0800e-11 + dcf_p_10_sp
+ ckappas=7.3000e-01          ckappad=7.3000e-01         acde=3.5090e-01
+ moin=6.7000e+00             noff=2.9360e+00            voffcv=-5.257e-02
+ ef=1.103336
```

```
+ noia=1.0635922e+41          noib=6.9613951e+26          noic=5.2897264e+09
+ em=4.1000000e+07           ntnoi=1.0
+ xl= - 1.0000e-08 + dxl_p_10_sp
+ xw=0.0000e+00 + dxw_p_10_sp                              dmcg=1.6000e-07
+ dmci=1.0000e-07            dwj=0.0000e+00               jss=1.9950e-07
+ jsws=1.0920e-13            jswgs=1.0000e-13             ijthsfwd=3.50e-03
+ ijthsrev=2.1750e-03        bvs=8.9640e+00               xjbvs=1.0000e+00
+ pbs=7.3000e-01             cjs=1.2600e-03 + dcjs_p_10_sp
+ mjs=3.1000e-01             pbsws=9.9000e-01
+ cjsws=1.2900e-10 + dcjsws_p_10_sp                       mjsws=1.0000e-01
+ pbswgs=6.0000e-01          cjswgs=2.4500e-10 + dcjswgs_p_10_sp
+ mjswgs=9.8900e-01          tnom=2.5000e+01              kt1=-3.4000e-01
+ kt11=-9.5660e-09           kt2=-1.0000e-02              ute=-1.9620e+00
+ ua1=-8.3500e-10            ub1=-1.3400e-18              uc1=0.0000e+00
+ prt=-1.6750e+02            at=1.0340e+05                njs=1.0540e+00
+ tpb=1.4000e-03             tcj=8.0000e-04               tpbsw=1.0000e-04
+ tcjsw=4.0000e-04           tpbswg=1.5050e-03            tcjswg=7.6180e-03
+ xtis=3.0000e+00            ll=5.5440e-16                wl=7.1650e-16
+ lln=1.0500e+00             wln=9.7350e-01               lw=-2.1170e-15
+ ww=-4.3920e-15             lwn=1.0000e+00               wwn=9.9400e-01
+ lwl=2.3760e-23             wwl= - 1.4950e-22 + dwwl_p_10_sp
+ llc=-6.7780e-16            wlc=0.0000e+00               lwc=0.0000e+00
+ wwc=1.0000e-15             lwlc=0.0000e+00              wwlc=0.0000e+00
+ lmin=8.0000e-08            lmax=5.0000e-05              wmin=1.2000e-07
+ wmax=1.0000e-04            pvth0=0.0000e+00 + dpvth0_p_10_sp
+ lk3=1.0000e+00             pk3=-1.4700e-01              lk3b=-7.6900e-01
+ wk3b=2.1290e+00            wdsub=1.1010e-02             pdvtp1=-2.000e-02
+ llpe0=2.9370e-08 + dllpe0_p_10_sp                       llpeb=1.3590e-08
+ lnfactor=2.6600e-01        letab=3.9610e-02             lags=1.8360e+00
+ pags=-8.4000e-02           lb0=-5.5000e-08              lb1=-5.8900e-08
+ lketa=-1.9200e-03          ldelta=2.4950e-03            wrdsw=1.0000e+01
+ lvoffcv=4.1250e-04         wkt1=5.0000e-03              wua1=3.9440e-11
+ saref=1.7600e-06           sbref=1.7600e-06             wlod=0.0000e+00
+ kvth0=-8.0000e-10          lkvth0=-1.5000e-06           wkvth0=6.0000e-07
+ pkvth0=0.0000e+00          llodvth=8.0000e-01         wlodvth=1.0000e+00
+ stk2=0.0000e+00            lodk2=1.0000e+00           lodeta0=1.0000e+00
+ ku0=5.3000e-07             lku0=5.8000e-04              wku0=-1.1000e-09
+ pku0=-2.5000e-10           llodku0=6.8000e-01           wlodku0=8.500e-01
+ kvsat=1.0000e+00           steta0=3.8000e-10            tku0=0.0000e+00

//  ***User Drop Parameters*****************************************

 simulator lang=spectre insensitive=yes
library MyLib

section tt
// DEVICE 1
parameters dtoxp_n_10_sp=0.000000e+000
parameters dwint_n_10_sp=0.000000e+000
parameters dllpe0_n_10_sp=0.000000e+000
parameters dxl_n_10_sp=0.000000e+000
parameters drdsw_n_10_sp=0.000000e+000
parameters dcgdo_n_10_sp=0.000000e+000
parameters dcgsl_n_10_sp=0.000000e+000
parameters dcf_n_10_sp=0.000000e+000
```

```
parameters dcjs_n_10_sp=0.000000e+000
parameters dcjsws_n_10_sp=0.000000e+000
parameters dcjswgs_n_10_sp=0.000000e+000
parameters dtoxe_n_10_sp=0.000000e+000
parameters dvth0_n_10_sp=0.000000e+000
parameters dcgso_n_10_sp=0.000000e+000
parameters dvoffl_n_10_sp=0.000000e+000
parameters dcgdl_n_10_sp=0.000000e+000
parameters dpvth0_n_10_sp=0.000000e+000
parameters dlpe0_n_10_sp=0.000000e+000
parameters dxw_n_10_sp=0.000000e+000
parameters dk3_n_10_sp=0.000000e+000
parameters dwwl_n_10_sp=0.000000e+000

// DEVICE 2
parameters dwwl_p_10_sp=0.000000e+000
parameters dtoxp_p_10_sp=0.000000e+000
parameters dwint_p_10_sp=0.000000e+000
parameters dllpe0_p_10_sp=0.000000e+000
parameters dxl_p_10_sp=0.000000e+000
parameters drdsw_p_10_sp=0.000000e+000
parameters dcgdo_p_10_sp=0.000000e+000
parameters dcgsl_p_10_sp=0.000000e+000
parameters dcf_p_10_sp=0.000000e+000
parameters dcjsws_p_10_sp=0.000000e+000
parameters dtoxe_p_10_sp=0.000000e+000
parameters dcjs_p_10_sp=0.000000e+000
parameters dcjswgs_p_10_sp=0.000000e+000
parameters dvth0_p_10_sp=0.000000e+000
parameters du0_p_10_sp=0.000000e+000
parameters dcgso_p_10_sp=0.000000e+000
parameters dvoffl_p_10_sp=0.000000e+000
parameters dcgdl_p_10_sp=0.000000e+000
parameters dpvth0_p_10_sp=0.000000e+000
parameters dxw_p_10_sp=0.000000e+000
parameters dk3_p_10_sp=0.000000e+000
parameters dlpe0_p_10_sp=0.000000e+000

include "L90_SP10_V051.mdl.scs"
endsection tt

section ss
// DEVICE 1
parameters dtoxp_n_10_sp=1.000000e-010
parameters dwint_n_10_sp=2.000000e-009
parameters dllpe0_n_10_sp=1.400000e-009
parameters dxl_n_10_sp=1.302000e-009
parameters drdsw_n_10_sp=1.700000e+001
parameters dcgdo_n_10_sp=-5.000000e-012
parameters dcgsl_n_10_sp=-2.200000e-011
parameters dcf_n_10_sp=-9.260000e-012
parameters dcjs_n_10_sp=1.070000e-004
parameters dcjsws_n_10_sp=1.260000e-011
parameters dcjswgs_n_10_sp=2.310000e-011
parameters dtoxe_n_10_sp=1.000000e-010
parameters dvth0_n_10_sp=3.300000e-002
parameters dcgso_n_10_sp=-5.000000e-012
```

```
parameters dvoffl_n_10_sp=4.500000e-010
parameters dcgdl_n_10_sp=-2.200000e-011
parameters dpvth0_n_10_sp=1.700000e-004
parameters dlpe0_n_10_sp=0.000000e+000
parameters dxw_n_10_sp=-5.000000e-009
parameters dk3_n_10_sp=2.200000e+000
parameters dwwl_n_10_sp=4.000000e-023

// DEVICE 2
parameters dwwl_p_10_sp=5.720000e-023
parameters dtoxp_p_10_sp=1.000000e-010
parameters dwint_p_10_sp=2.500000e-009
parameters dllpe0_p_10_sp=2.600000e-009
parameters dxl_p_10_sp=1.000000e-009
parameters drdsw_p_10_sp=1.000000e+001
parameters dcgdo_p_10_sp=-4.200000e-012
parameters dcgsl_p_10_sp=-2.000000e-011
parameters dcf_p_10_sp=-9.080000e-012
parameters dcjsws_p_10_sp=1.290000e-011
parameters dtoxe_p_10_sp=1.000000e-010
parameters dcjs_p_10_sp=1.260000e-004
parameters dcjswgs_p_10_sp=2.450000e-011
parameters dvth0_p_10_sp=-3.000000e-002
parameters du0_p_10_sp=0.000000e+000
parameters dcgso_p_10_sp=-4.200000e-012
parameters dvoffl_p_10_sp=-1.170000e-009
parameters dcgdl_p_10_sp=-2.000000e-011
parameters dpvth0_p_10_sp=0.000000e+000
parameters dxw_p_10_sp=-3.000000e-009
parameters dk3_p_10_sp=3.051000e+001
parameters dlpe0_p_10_sp=0.000000e+000

include "L90_SP10_V051.mdl.scs"
endsection ss

section ff
// DEVICE 1
parameters dtoxp_n_10_sp=-1.000000e-010
parameters dwint_n_10_sp=0.000000e+000
parameters dllpe0_n_10_sp=-1.550000e-009
parameters dxl_n_10_sp=-1.000000e-010
parameters drdsw_n_10_sp=-1.000000e+001
parameters dcgdo_n_10_sp=5.000000e-012
parameters dcgsl_n_10_sp=2.200000e-011
parameters dcf_n_10_sp=9.260000e-012
parameters dcjs_n_10_sp=-1.070000e-004
parameters dcjsws_n_10_sp=-1.260000e-011
parameters dcjswgs_n_10_sp=-2.310000e-011
parameters dtoxe_n_10_sp=-1.000000e-010
parameters dvth0_n_10_sp=-3.300000e-002
parameters dcgso_n_10_sp=5.000000e-012
parameters dvoffl_n_10_sp=-1.053000e-009
parameters dcgdl_n_10_sp=2.200000e-011
parameters dpvth0_n_10_sp=-1.400000e-004
parameters dlpe0_n_10_sp=0.000000e+000
parameters dxw_n_10_sp=1.800000e-009
parameters dk3_n_10_sp=-2.510000e+000
```

```
parameters dwwl_n_10_sp=-1.200000e-023

// DEVICE 2
parameters dwwl_p_10_sp=-4.200000e-023
parameters dtoxp_p_10_sp=-1.000000e-010
parameters dwint_p_10_sp=0.000000e+000
parameters dllpe0_p_10_sp=0.000000e+000
parameters dxl_p_10_sp=-3.337000e-009
parameters drdsw_p_10_sp=0.000000e+000
parameters dcgdo_p_10_sp=4.200000e-012
parameters dcgsl_p_10_sp=2.000000e-011
parameters dcf_p_10_sp=9.080000e-012
parameters dcjsws_p_10_sp=-1.290000e-011
parameters dtoxe_p_10_sp=-1.000000e-010
parameters dcjs_p_10_sp=-1.260000e-004
parameters dcjswgs_p_10_sp=-2.450000e-011
parameters dvth0_p_10_sp=2.360000e-002
parameters du0_p_10_sp=0.000000e+000
parameters dcgso_p_10_sp=4.200000e-012
parameters dvoffl_p_10_sp=-2.565000e-009
parameters dcgdl_p_10_sp=2.000000e-011
parameters dpvth0_p_10_sp=1.215000e-004
parameters dxw_p_10_sp=3.000000e-009
parameters dk3_p_10_sp=-3.100000e+001
parameters dlpe0_p_10_sp=1.600000e-009

include "L90_SP10_V051.mdl.scs"
endsection ff


section snfp
// DEVICE 1
parameters dtoxp_n_10_sp=0.000000e+000
parameters dwint_n_10_sp=0.000000e+000
parameters dllpe0_n_10_sp=1.700000e-009
parameters dxl_n_10_sp=3.040000e-010
parameters drdsw_n_10_sp=1.300000e+001
parameters dcgdo_n_10_sp=0.000000e+000
parameters dcgsl_n_10_sp=0.000000e+000
parameters dcf_n_10_sp=0.000000e+000
parameters dcjs_n_10_sp=1.070000e-004
parameters dcjsws_n_10_sp=1.260000e-011
parameters dcjswgs_n_10_sp=2.310000e-011
parameters dtoxe_n_10_sp=0.000000e+000
parameters dvth0_n_10_sp=2.200000e-002
parameters dcgso_n_10_sp=0.000000e+000
parameters dvoffl_n_10_sp=2.250000e-010
parameters dcgdl_n_10_sp=0.000000e+000
parameters dpvth0_n_10_sp=1.160000e-004
parameters dlpe0_n_10_sp=0.000000e+000
parameters dxw_n_10_sp=-3.200000e-009
parameters dk3_n_10_sp=1.100000e+000
parameters dwwl_n_10_sp=1.800000e-023

// DEVICE 2
parameters dwwl_p_10_sp=-2.200000e-023
parameters dtoxp_p_10_sp=0.000000e+000
```

```
parameters dwint_p_10_sp=0.000000e+000
parameters dllpe0_p_10_sp=0.000000e+000
parameters dxl_p_10_sp=-2.259000e-009
parameters drdsw_p_10_sp=0.000000e+000
parameters dcgdo_p_10_sp=0.000000e+000
parameters dcgsl_p_10_sp=0.000000e+000
parameters dcf_p_10_sp=0.000000e+000
parameters dcjsws_p_10_sp=-1.290000e-011
parameters dtoxe_p_10_sp=0.000000e+000
parameters dcjs_p_10_sp=-1.260000e-004
parameters dcjswgs_p_10_sp=-2.450000e-011
parameters dvth0_p_10_sp=1.600000e-002
parameters du0_p_10_sp=1.100000e-004
parameters dcgso_p_10_sp=0.000000e+000
parameters dvoffl_p_10_sp=-1.365000e-009
parameters dcgdl_p_10_sp=0.000000e+000
parameters dpvth0_p_10_sp=6.150000e-005
parameters dxw_p_10_sp=1.500000e-009
parameters dk3_p_10_sp=-1.600000e+001
parameters dlpe0_p_10_sp=-3.078000e-009

include "L90_SP10_V051.mdl.scs"
endsection snfp

section fnsp
// DEVICE 1
parameters dtoxp_n_10_sp=0.000000e+000
parameters dwint_n_10_sp=0.000000e+000
parameters dllpe0_n_10_sp=-1.000000e-009
parameters dxl_n_10_sp=-1.000000e-010
parameters drdsw_n_10_sp=-1.300000e+001
parameters dcgdo_n_10_sp=0.000000e+000
parameters dcgsl_n_10_sp=0.000000e+000
parameters dcf_n_10_sp=0.000000e+000
parameters dcjs_n_10_sp=-1.070000e-004
parameters dcjsws_n_10_sp=-1.260000e-011
parameters dcjswgs_n_10_sp=-2.310000e-011
parameters dtoxe_n_10_sp=0.000000e+000
parameters dvth0_n_10_sp=-2.100000e-002
parameters dcgso_n_10_sp=0.000000e+000
parameters dvoffl_n_10_sp=-5.000000e-010
parameters dcgdl_n_10_sp=0.000000e+000
parameters dpvth0_n_10_sp=-1.300000e-004
parameters dlpe0_n_10_sp=-1.640000e-008
parameters dxw_n_10_sp=1.000000e-009
parameters dk3_n_10_sp=-7.502000e-001
parameters dwwl_n_10_sp=-1.200000e-023

// DEVICE 2
parameters dwwl_p_10_sp=3.585000e-023
parameters dtoxp_p_10_sp=0.000000e+000
parameters dwint_p_10_sp=1.250000e-009
parameters dllpe0_p_10_sp=1.800000e-009
parameters dxl_p_10_sp=5.000000e-010
parameters drdsw_p_10_sp=1.000000e+001
parameters dcgdo_p_10_sp=0.000000e+000
parameters dcgsl_p_10_sp=0.000000e+000
```

```
parameters dcf_p_10_sp=0.000000e+000
parameters dcjsws_p_10_sp=1.290000e-011
parameters dtoxe_p_10_sp=0.000000e+000
parameters dcjs_p_10_sp=1.260000e-004
parameters dcjswgs_p_10_sp=2.450000e-011
parameters dvth0_p_10_sp=-2.000000e-002
parameters du0_p_10_sp=0.000000e+000
parameters dcgso_p_10_sp=0.000000e+000
parameters dvoffl_p_10_sp=-6.000000e-010
parameters dcgdl_p_10_sp=0.000000e+000
parameters dpvth0_p_10_sp=0.000000e+000
parameters dxw_p_10_sp=-1.500000e-009
parameters dk3_p_10_sp=1.440000e+001
parameters dlpe0_p_10_sp=0.000000e+000

include "L90_SP10_V051.mdl.scs"
endsection fnsp

endlibrary MyLib
```

ATLAS Simulation Results

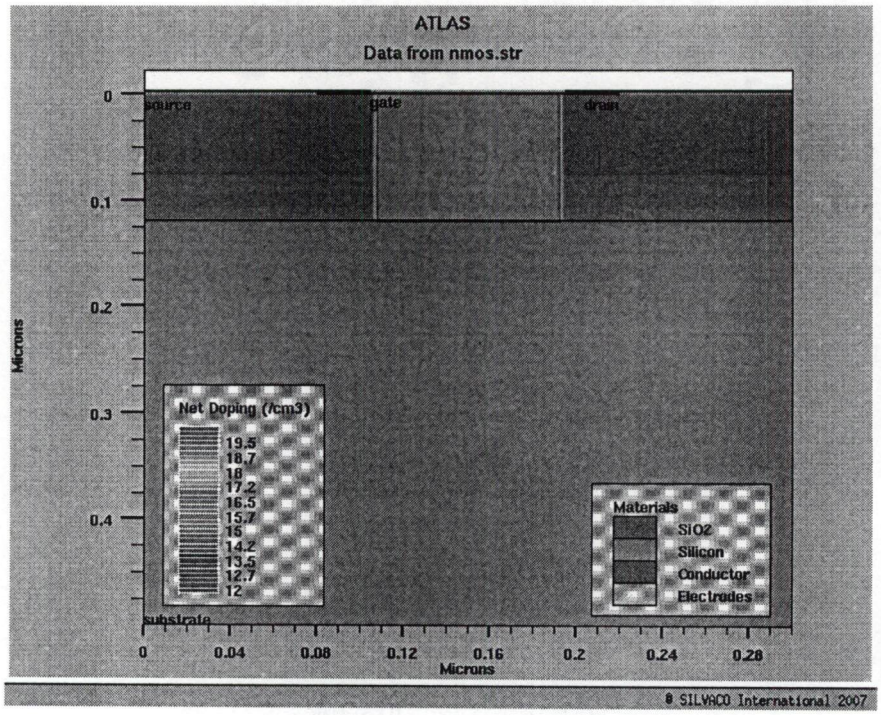


Figure1 Simulated Device Structure

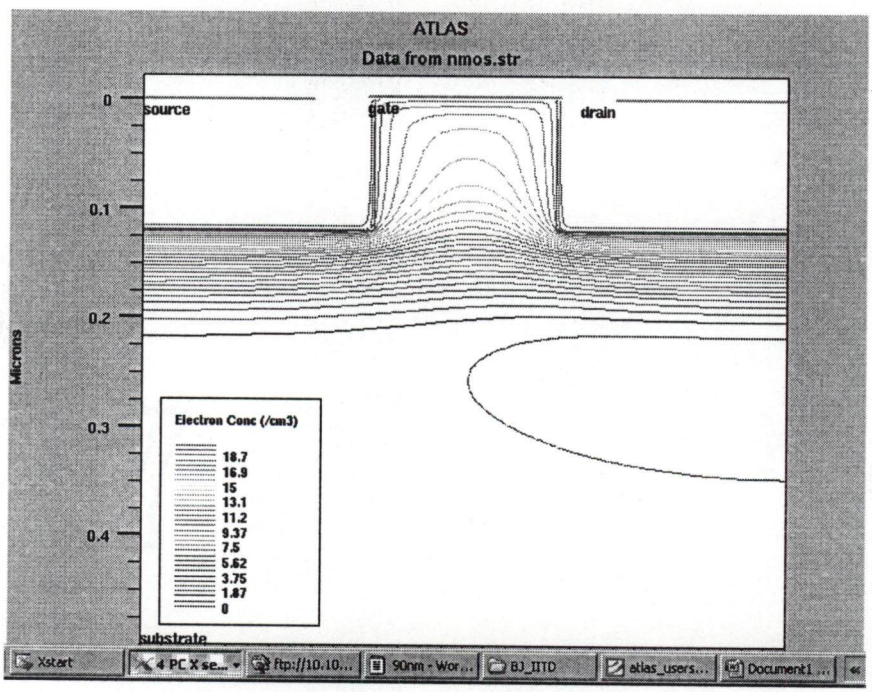


Figure2 Device Structure showing electron concentration

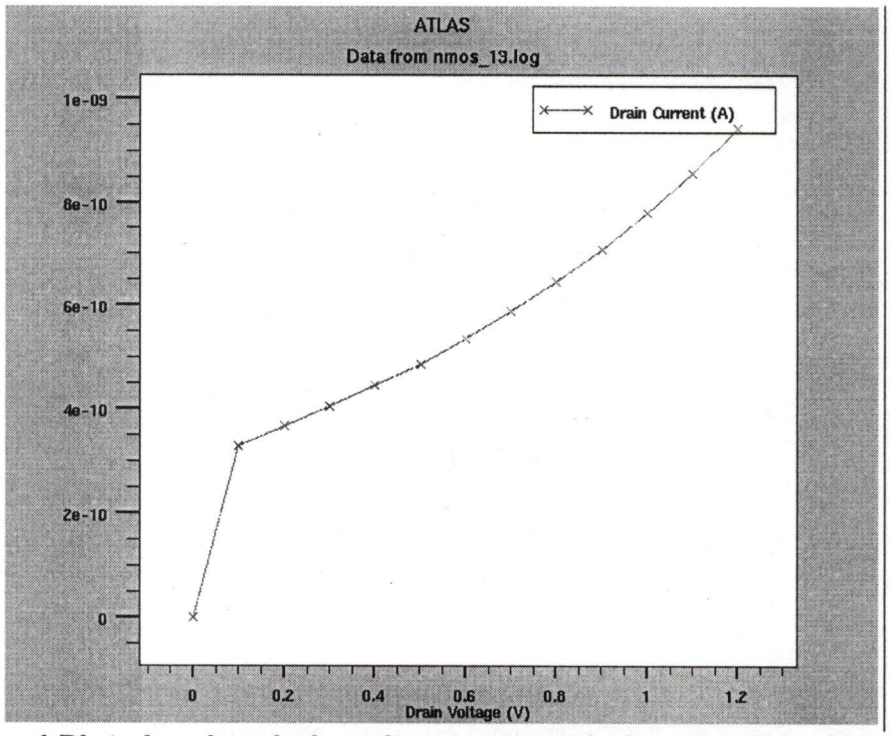


Figure 4 Plot showing drain voltage versus drain current for V_{GS}=0.1V

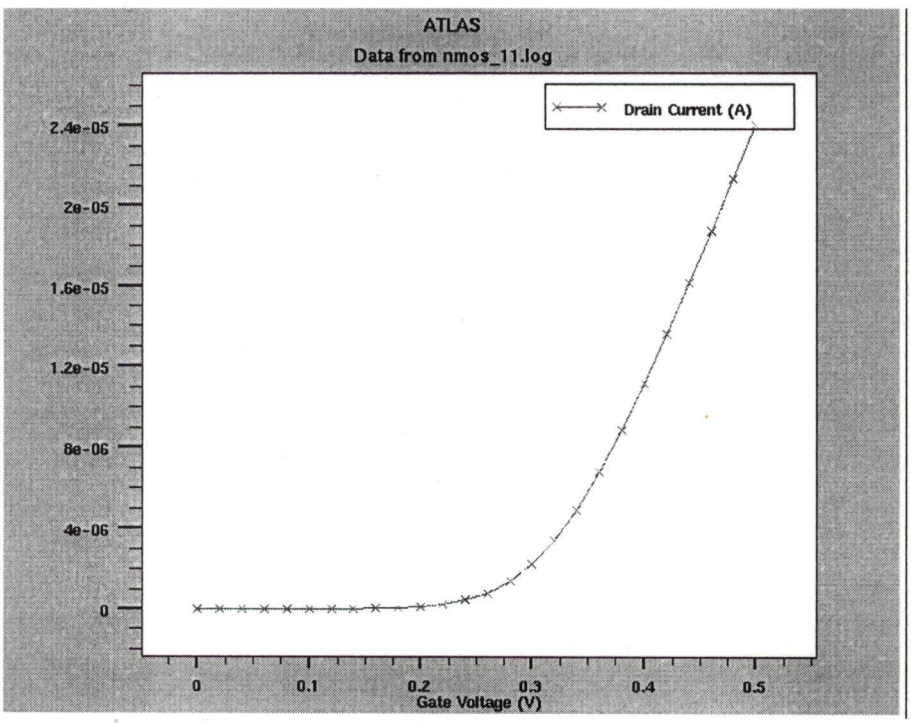


Figure 5 Gate voltage versus drain current for V_{DS}=1V

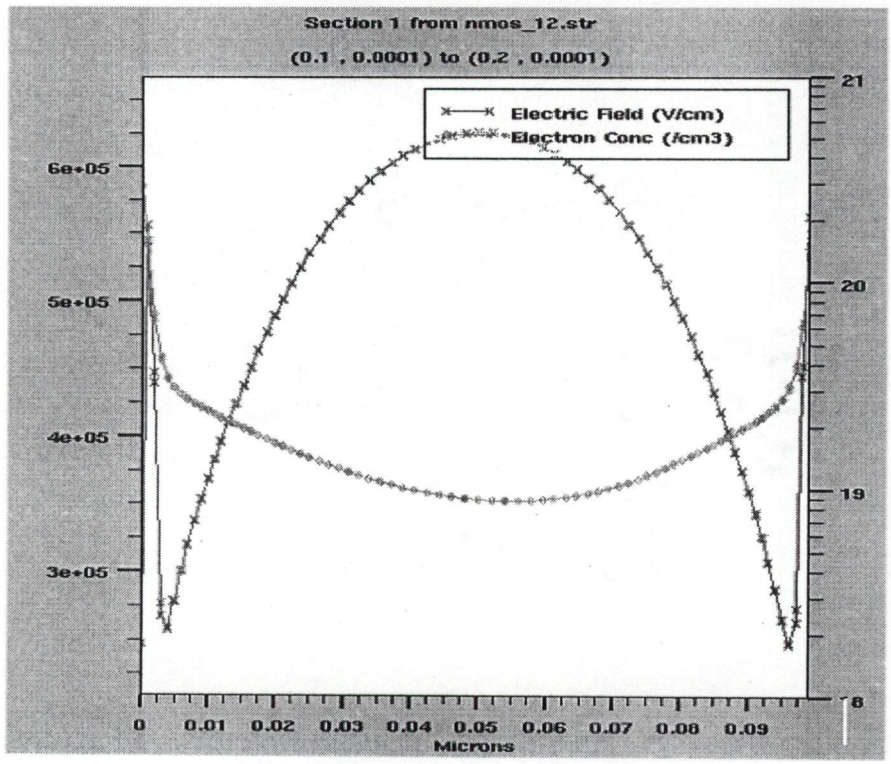


Figure 5 Plot showing electric field and electron concentration along channel

BILLIOGRAPHY

[AME03] K. G. Anil, S. Mahapatra and I. Eisele, "A Detailed Experimental Investigation of Impact Ionization in n-Channel Metal-Oxide-Semiconductor Field-Effect-Transistors at Very Low Drain Voltages," *Solid State Electronics*, Vol. 47, pp. 995-1001 (2003).

[APSB+94] F. Assaderaghi, S. Park, D. Sinitsky, J. Bokor, P. K. Ko, and C. Hu, "A Dynamic Threshold Voltage MOSFET (DTMOS) for Low Voltage Operation," *IEEE Electron Device Letter*, Vol. 15, pp. 510 (1994).

[Aro93] N. Arora, MOSFET Models for VLSI Circuit Simulation: Theory and Practice, Springer-Verlag Wein New York, (1993).

[ASPB+97] F. Assaderaghi, D. Sinitsky, S. Parke, J. Bokor, P. Ko, and C. Hu, "Dynamic Threshold Voltage MOSFET (DTMOS) for Ultra Low Voltage VLSI," *IEEE Transactions on Electron Devices*, vol. 44, no. 3, pp. 414-421 (1997).

[AW97] S. Asai and Y. Wada, "Technology Challenges for Integration Near and Below 0.1μm," *Proceedings IEEE*, Vol. 85, pp. 505, (1997).

[AYPM+94] J. G Ahn, C. S. Yao, Y. J. Park, H. S. Min, and R. Dutton, "Impact Ionization Modeling Using Simulation of High Energy Tail Distributions," *IEEE Electron Device Letters*, Vol. 15, No. 9, pp. 348–350 (1994).

[Bar62] G.A. Baraff, "Distribution Functions and Ionization Rates for Hot Electrons in Semiconductors," *Physical Review*, Vol. 128 (6), pp. 2507–2517 (1962).

[BBCK+01] A. Bryant, J. Brown, P. Cottrell, M. Ketchen, J. E. Monaghan and E, J, Nowak, "Low-power CMOs at Vdd=4kT/q," *Device Research Conference*, pp. 22-23 (2001).

[BBS99] V. K. Bhat, K. N. Bhat and A. Subrahmanyam, "Frequency Dependence of Accumulation Capacitance of MOS Structure with Ultra thin Oxide Layer," *Proceedings of the 10th International Workshop on Physics of Semiconductor Devices (IWPSD), New Delhi*, pp. 341-344 (1999).

[BCS91] R. W. Brodersen, A. Chandrakasan and S. Sheng, "Technologies for Personal Communications," *Proceedings of the VLSI Symposium*, pp. 5–9 (1991).

[BDAM00] A. J. Bhavnagarwala, V. K. De, B. Austin and J. D. Meindl, "A Minimum Total Power Methodology for Projecting Limits on CMOS GSI," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 8, No. 3, pp. 235-251 (2000).

[BDAM96] A. J. Bhavnagarwala, V. K. De, B. Austin and J. D. Meindl, "Circuit Techniques for Low Power CMOS GSI", *ISLPED Digest Technical Papers,* pp. 193-196 (1996).

[BFOR93] L. Benini, M. Favalli, P. Olivo, and B. Ricco, "A novel approach to cost-effective estimate of power dissipation in CMOS ICs," *European Design Automation Conference*, pp. 354–360 (1993).

[BFR94] L. Benini, M. Favalli and B. Ricco, "Analysis of Hazard Contribution to Power Dissipation in CMOS IC's," *Proceedings of the 1994 ACM/IEEE International Workshop on Low Power Design*, pp. 27–32 (1994).

[BH92] J. F. Bude and K. Hess, "Impact Ionization in Semiconductors: Effects of High Electric Fields and High Scattering Rates," *Physical Review B,* Vol. 45, No. 19 (1992).

[BHI92] J. Bude, K. Hess, and G.J. Iafrate, "Impact Ionization: Beyond the Golden Rule," *Semiconductor Science and Technology,* vol. 7, no. 3B (1992).

[BHMN+02] R. Brodersen, M. A. Horowitz, D. Markovic, B. Nikolic and V. Stojanovic, "Methods for True Power Minimization," *ICCAD*, pp. 35-42 (2002).

[BM95] J. D. Bude and M. Mastrapasqua, "Impact ionization and distribution functions in sub-micron nMOSFET technologies," *IEEE Electron Device Letters*, vol. 16, no. 10, pp. 439-441(1995).

[BMTN+95] F. Balestra, T. Matsumoto, M. Tsuno, H. Nakabayashi and M. Koyanagi, "New Experimental findings on Hot Carrier Effects in Sub-0.1 μm MOSFET's," *IEEE Electron Devices Letters*, Vol.16, No.10, pp.433-435 (1995).

[BNYT93] R. Burch, F. N. Najm, P. Yang, and T. Trick, "A Monte Carlo Approach for Power Estimation," *IEEE Transactions on VLSI Systems*, Vol. 1, No. 1, pp. 63-71 (1993).

[BOI95] M. Borah, R. M Owens and M. J Irwin, "Transistor Sizing for Minimizing Power Consumption of CMOS Circuits Under Delay Constraint," *Proceedings of the 1995 International Symposium on Low Power Design,* pp. 167-172 (1995).

[Bon98] P. Bondy, "Moore's Law Governs the Silicon Revolution," *Proceedings IEEE*, Vol. 86, pp. 78 (1998).

[BP91] J. Burr and A. Peterson, "Ultra-low Power CMOS Technology," *NASA VLSI Design Symposium*, pp. 4.2.1-13 (1991).

[BPS00] J. D. Bude, M. R. Pinto, and R. K. Smith, "Monte Carlo simulation of the CHISEL flash memory cell," *IEEE Transactions on Electron Devices*, vol. 47, no. 10, pp. 1873-1881(2000).

[BPVV+04] B. Bhaumik, P. Pradhan, G. S. Visweswaran, R. Varambally and A. Hardi, "A Low Power 256 KB SRAM Design," *DAC* (2004).

[BRCV+02] D. G. Borse, M. Rani K. N., A. N. Chandorkar, J. Vasi, V. Ramgopal Rao, B. Cheng, and J. C. S. Woo, "Optimization and Realization of sub 100nm channel Length lateral Asymmetric p-MOSFETs," *IEEE Transactions on Electron Devices*, Vol. 49, No. 6 (2002).

[Bry92] R. E. Bryant, "Symbolic Boolean Manipulation with Ordered Binary-Decision Diagrams," *ACM Computing Surveys*, vol. 24, No. 3, pp. 293-318 (1992).

[BS95] D. Burnett and S. W. Sun, "Statistical Threshold-Voltage Variation and its Impact on Supply-Voltage Scaling," *Proceedings SPIE*, Vol. 2636, pp. 83 (1995).

[BWD84] G. Baccarani, M.R. Wordeman and R. H. Dennard, "Generalized Scaling Theory and its Application to a ¼μm MOSFET Design," *IEEE Transactions on Electron Devices*, Vol. ED-31, No.4, pp. 452-462 (1984).

[CASB94] A. Chandrakasan, R. Allmon, A. Stratakos, and R. W. Brodersen, "Design of Portable Systems," *Proceedings of the IEEE Custom Integrated Circuit Conference, San Diego, California* (1994).

[CB95] A. P. Chandrakasan and R. W. Brodersen, "Minimizing Power Consumption in Digital CMOS Circuits," *Proceedings IEEE*, Vol. 83, pp. 498–523 (1995).

[CBSP95] Z. Chen, J. Burr, J. Shott, and J. D. Plummer, "Optimization of Quarter Micron MOSFETs for Low Voltage/Low Power Applications," *IEEE IEDM Technical Digest,* pp. 63– 65 (1995).

[CC04] B. H. Calhoun and A. Chandrakasan, "Characterizing and Modeling Minimum Energy Operation for Sub threshold Circuits," *ISLPED04*, pp. 90-95 (2004).

[CDWT+92] W. H. Chang, B. Davari, M. R. Wordeman, Y. Taur, C. C. H. Hsu, and M. D. Rodriguez, "A High-performance 0.25-m CMOS Technology-I:

Design and characterization," *IEEE Transactions on Electron Devices*, Vol. 39, pp. 959 (1992).

[CFEM93] E. Cartier, M. V. Fischetti, E. A. Eklund, and F. R. McFeely, "Impact ionization in Silicon," *Applied Physics Letters*, vol. 62, no. 25, pp. 3339-3341 (1993).

[Cha96] A. Chatterjee, "An Investigation of the Impact of Technology Scaling on Power Wasted as Short-circuit Current in Low Voltage Static CMOS Circuits," in *Proceedings of the International Symposium on Low Power Electronics and Design*, ISLPED, pp. 145-150 (1996).

[CHHL80] P. K. Chatterjee, W. R. Hunter, T.C. Holloway, and Y.T. Lin, "The Impact of Scaling Laws on the Choice of n-Channel or p-Channel for MOS VLSI," *IEEE Electron Device Letters*, Vol. EDL-1, No.10, pp. 220–223 (1980).

[Chy58] A. G. Chynoweth, "Ionization Rates for Electrons and Holes in Silicon," *Physical Review*, Vol. 109, No. 5, pp. 1537–1540 (1958).

[CJMK+91] J. E. Chung, M. C. Jeng, J. E. Moon, P. K. Ko, C. Hu, "Performance and Reliability Design Issues for Deep-sub-micrometer MOSFET's," *IEEE Transactions on Electron Devices*, Vol. 38, No. 3, pp. 545-554 (1991).

[CJWR98] Z. Chen, M. Johnson, L. Wei, and K. Roy, "Estimation of Stand by Leakage Power in CMOS Circuits Considering Accurate Modeling of Transistor Stacks", *International Symposium on Low Power Electronics and Design*, pp. 239-244, August, 1998.

[CKH85] T.Y. Chan, P.K. Ko and C. Hu, "Dependence of Channel Electric Field in Device Scaling," *IEEE Electron Device Letters*, Vol. EDL-6, pp. 551-553 (1985).

[Cri87] M. A. Cirit, "Estimating Dynamic Power Consumption of CMOS Circuits", *IEEE International Conference on Computer -Aided Design*, pp. 534–537 (1987).

[CSB92] A. Chandrakasan, S. Sheng, and R. Brodersen, "Low Power CMOS Digital Design," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 4, pp. 473--484, (1992).

[CSBP94] Z. Chen, J. Shott, J. Burr and J. D. Plummer, "CMOS Technology Scaling for Low Voltage Low Power Applications," *SLPE Digest Technical Papers*, pp. 6–57 (1994).

[DCPW+92] B. Davari, W. H. Chang, K. E. Petrillo, C. Y. Wong, D. Moy, Y. Taur, M. R. Wordeman, J. Y. C. Sun and C. C. H. Hsu, "A High-performance 0.25-μm CMOS technology—II: Technology," *IEEE Transactions on Electron Devices*, Vol. 39, pp. 967 (1992).

[DDS95a] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS Scaling, The Next Ten Years," *Proeedings IEEE*, Vol. 83, pp. 595-606 (1995a).

[DDS95b] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS Scaling for High Performance and Low Power," Proceedings IEEE, Vol. 83, No. 4, pp. 408–425 (1995b).

[DRJ98] A. Duncan, U. Ravaioli and J. Jakumeit, "Full-band Monte Carlo Investigation of Hot Carrier Trends in the Scaling of Metal-Oxide-Semiconductor Field-Effect Transistors," *IEEE Transactions on Electron Devices*, Vol. 45, No. 4, pp. 867–876 (1998).

[DTP98] C.-S. Ding, C.-Y. Tsui, and M. Pedram, "Gate-Level Power Estimation Using Tagged Probabilistic Simulation", *IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems*, vol. 17, no. 11, pp. 1099–1107, November 1998.

[FDNS+01] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur and H. S. P. Wong, "Device Scaling Limits of Si MOSFETs and Their

Application Dependencies," *Proceedings IEEE*, Vol. 89, No. 3, pp. 259-287 (2001).

[FGD94] W. Fikry, G. Ghibaudo and M. Dutoit, "Temperature Dependence of Drain-Induced Barrier Lowering in Deep-Submicrometer MOSFETs," *IEEE Electron Devices Letters*, Vol. 30, No. 11, pp. 911-912 (1994).

[FIWS+92] C. Fiegna, H. Iwai, T. Wada, T. Saito, E. Sangiorgi, and B. Ricco, "A New Scaling Methodology for the 0.1–0.025μm MOSFET," *Proceedings of the Symposium on VLSI Technology,* pp. 33 (1992).

[FIWS+94] C. Fiegna, H. Iwai, T. Wada, M. Saito, E. Sangiorgi, and B. Ricc´o, "Scaling the MOS transistor below 0.1 μm: Methodology, device structures, and technology requirements," *IEEE Transactions on Electron Devices*, Vol. 41, pp. 941 (1994).

[FL95] M. V. Fischetti and S .E. Laux, "Monte Carlo Study of Sub Band Gap Impact Ionization in Small Silicon Field-Effect Transistors," *IEDM Technical Digest,* pp. 305-308 (1995).

[FS93] T. A. Fjeldly and M. Shur, "Threshold Voltage Modeling and the Sub-threshold Regime of Operation of Short-Channel MOSFET's," *IEEE Transactions on Electron Devices*, Vol. 4, No. 1, pp. 137-145 (1993).

[FSRS97] D. J. Frank, P. Solomon, S. Reynolds and J. Shin, "Supply and Threshold Voltage Optimization for Low Power Design," *Proceedings of the International Symposium on Low Power Electronics and Design*, pp. 317-322 (1997).

[FTW98] D. Frank, Y. Taur, and H.-S. P. Wong, "Generalized Scale Length for Two-Dimensional Effects in MOSFET's," *IEEE Electron Device Letters,* Vol. 19, pp. 385 (1998).

[GDKW92] A. Ghosh, S. Devadas, K. Keutzer, and J. White, "Estimation of Average Switching Activity in Combinational and Sequential Circuits",

Proceedings of the 29th Design Automation Conference, pp. 253–259 (1992).

[GE96] R. X. Gu and M. I. Elmasry, "Power Dissipation Analysis and Optimization of Deep Sub-micron CMOS Digital Circuits," *IEEE Journal of Solid-State Circuits*, Vol. 31, No. 5, pp. 707-713 (1996).

[GGH97] R. Gonzalez, B. M. Gordon and M. A. Horowitz, "Supply and Threshold Voltage Scaling for Low power CMOS," *IEEE Journal of Solid-State Circuits*, Vol. 32, No. 8, pp. 1210-1216 (1997).

[Ghi99] H. El Ghitani, "DIBL-Coefficient in Short-Channel NMOS Transistors," *16th National Radio Science Conference (NRSC)*, pp. D4-1-5 (1999).

[GSB99] A. Ghetti, L. Selmi, and R. Bez, "Low-voltage hot electrons and soft programming lifetime prediction in nonvolatile memory cells," *IEEE Transactions on Electron Devices,* vol. 46, no. 4, pp. 696-702 (1999).

[HIG95] M. Horowitz, T. Indermaur, and R. Gonzalez, "Low-Power Digital Design," *Proceedings of the 1995 IEEE Symposium on Low Power Electronics*, pp. 8–11 (1995).

[HNYJ+99] S. H. Hong, S. M. Nam, C.G. Yu, S. H. Jeon and J. T. Pak, "New Experimental Findings on the Enhanced MOSFET Degradation at Elevated Temperature," *Proceedings IEEE TENCON*, Vol. 2, pp.1113-1115 (1999).

[IF97] B. Iniguez and T. A. Fjeldly, "Unified Substrate Current Model for MOSFET's," *Solid State Electronics*, Vol.41, pp. 87-94 (1997).

[IHS00] T. Inukai, T. Hiramot, and T. Sakurai., "Boosted Gate MOS (BGMOS): Device/Circuit Cooperation Scheme to achieve Leakage free Giga scale Integration," in *Proceedings of IEEE Custom Integrated Circuits Conference*, pp. 409 – 412 (2000).

[IHS01] T. Inukai, T. Hiramot, and T. Sakurai, "Variable Threshold Voltage CMOS (VTCMOS) is Series Connected Circuits," in *Proceedings of the International Symposium on Low Power Electronics and Design*, ISLPED, pp. 201-206 (2001).

[IP94] S. Iman, and M. Pedram, "Multi-level Network Optimization for Low Power," *Proceedings of the IEEE International Conference on Computer Aided Design*, pp. 372–377 (1994).

[IP95] S. Iman, and M. Pedram, "Two Level Logic Minimization for Low Power," *Proceedings of the IEEE International Conference on Computer Aided Design*, pp. 433–438 (1995).

[JP96] J. Rabaey and M. Pedram: Editors, Low Power Design Methodologies, Kluwer Academic Publishers (1996).

[JSYR02] M. Johnson, D. Somasekhar, C. Yih, and K. Roy, "Leakage Control with Effcient use of Transistor Stacks in Single Threshold CMOS," *IEEE Transactions on Very Large Scale Integration*, vol. 10, no. 1, pp. 1-5 (2002).

[Kan67] E.O. Kane, "Electron Scattering by Pair Production in Silicon," *Physical Review*, Vol.159 (3), pp. 624–631 (1967).

[Kan86] S. M. Kang, "Accurate Simulation of Power Dissipation in VLSI Circuits," *IEEE J. Solid State Circuits*, Vol. 21, No. 5, pp. 889–891(1986).

[Kap94] B. Kapoor, "Improving the Accuracy of Circuit Activity Measurement," *Proceedings of the 1994 International Workshop on Low Power Design*, pp. 111–116 (1994).

[KC00] J. T. Kao and A. Chandrakasan, "Dual-Threshold Voltage Techniques for Low Power Digital circuits," IEEE Journal of Solid-State Circuits, vol. 35, no. 7, pp.1009-1018 (2000).

[Kel65] L. V. Keldysh, "Concerning the Theory of Impact Ionization in Semiconductors," *Soviet Physics JETP*, Vol. 21, No. 6, pp. 1135-1144 (1965).

[KFMN+96] T.Kuroda, T.Fujita, S.Mita, T.Nagamatsu, S.Yoshioka, F.Sano, M.Norishima, M.Murota, M.Kato, M.Kinugawa, M.Kakumu and T.Sakurai, "A 0.9-V, 150-MHz, 10-mW, 4 mm^2 , 2-D Discrete Cosine Transform Core Processor with Variable Threshold-Voltage (VT) Scheme," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 11, pp. 1770-1779 (1996).

[KHK02] C. Kim, In-Chul Hwang and Steve Kang, "A Low-Power Small-Area ±7.28 ps Jitter 1-GHz DLL-based Clock Generator", IEEE Journal of Solid-State Circuits, vol. 37, no. 11, pp. 1414 - 1420, 2002.

[KKK03] C. Kim, K. Kim, and S. Kang, "Energy-efficient Skewed Static Logic(S^2L)with Dual Vt: Design and Synthesis", IEEE Tran on VLSI Systems, vol. 11, no. 1, pp. 64 - 70, 2003.

[KL96] S. M. Kang and Y. Leblebici, "CMOS Digital Integrated Circuits: Analysis and Design," *McGraw-Hill, New York* (1996).

[KMNB+01] A. Keshavarzi, S. Ma, S. Narendra, B. Bloechel, K. Mistry, S. Borkar, and V. De, "Effectiveness of Reverse Body Bias for Leakage Control in Scaled Dual Vt CMOS ICs," in *Proceedings of the International Symposium on Low Power Electronics and Design*, ISLPED, pp. 207-212 (2001).

[KNBH+99] A. Keshavarzi, S. Narenda, S. Borkar, C. Hawkins, K. Roy, and V. De, "Technology Scaling Behavior of Optimum Reverse Body Bias for Standby Leakage Power Reduction on CMOS ICs," in *Proceedings of the International Symposium on Low Power Electronics and Design*, ISLPED, pp. 252-254 (1999).

[KNC98] J. T. Kao, S. Narenda, and A. Chandrakasan, "MTCMOS Hierarchal Sizing based on Mutual Exclusive Discharge Patterns," in

Proceedings of the Design Automation Conference, pp. 495-500 (1998).

[KNS00] H. Kawaguchi, K. Nose, and T. Sakurai, "A Super Cut-Off CMOS (SCCMOS) Scheme for 0.5-V Supply Voltage with Picoampere Standby Current," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 10, pp. 1498-1501(2000).

[KNS98] H. Kawaguchi, I. Nose, and T. Sakurai, "A CMOS scheme for 0.5 V Supply Voltage with Pico-ampere Standby Current," in *Proceedings of the International Solid-State Conference*, pp. 192 – 193 (1998).

[KPW02] S. D. Kim, C. M. Park and J. C. S. Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling Into Nanometer Regime—Part II: Quantitative Analysis," *IEEE Transactions on Electron Devices*, Vol. 49, No. 3, pp. 467-472 (2002).

[KR01] H. Kim and K. Roy, "Ultra-low Power DLMS Adaptive Filters for Hearing Aid Applications," *ISPLED* (2001).

[KRH97] A. Keshavarzi, K. Roy, and C. Hawkins, "Intrinsic leakage in low power deep submicron CMOS ICs," *Proceedings of Intl. Test Conference*, pp. 146-155 (1997).

[KS02] M. Ketkar, and S. Sapatnekar, "Standby Power Optimization Via Transistor Sizing and Dual Threshold Voltage Assignment", *Proceedings of the International Conference on Computer Aided Design*, pp. 375-378 (2002).

[KS94] T. Kobayashi, and T. Sakurai, "Self-adjusting Threshold-voltage Scheme for Low Voltage High Speed Operation," *Proceedings of CICC*, pp. 271–274 (1994).

[KT02] N. Koike and K. Tatsuuma, "A Drain Avalanche Hot Carrier Lifetime Model For n- And p-Channel MOSFET's," *IEEE 40th Annual*

International Reliability Physics Symposium, Dallas, Texas, pp. 86-92 (2002).

[KTK76] T. Kamata, K. Tanabashi, and K. Kobayashi, "Substrate Current due to Impact Ionization in MOSFET," *Japanese Journal of Applied Physics*, Vol. 15, No. 6, pp. 1127–1133 (1976).

[LHHC+93] Z. H. Liu, C. Hu, J. H. Huang, T. Y. Chan, M. C. Jeng, P. Ko, and Y. Cheng, "Threshold Voltage Model for Deep Sub-Micrometer MOSFET's," *IEEE Transactions on Electron Devices*, Vol. 40, pp. 86 (1993).

[LM93] B. Lin, and H. D. Man, "Low-power Driven Technology Mapping under Timing Constraints," *Proceedings of the International Conference on Computer Design*, pp. 421–427 (1993).

[LS93] D. Liu and C. Svensson, "Trading Speed for Low Power by Choice of Supply and Threshold Voltages," *IEEE Journal of Solid-State Circuits*, Vol. 28, pp. 10 (1993).

[LYCZ+00] W. Li, J. S. Yuan, S. Chetlur, J. Zhou, A. S. Oates, "An improved substrate current model for deep submicron MOSFETs," *Solid-State Electronics*, Vol. 44, pp.1985-1988 (2000).

[MB95] M. Mastrapasqua and J. D. Bude, " Electron and hole impact ionization in deep sub-micron MOSFET", *Microelectronic Engineering*, Vol. 28, pp. 293-300 (1995).

[MBSH+99] J. S. Martinez, S. S. Bustos, J. S. Suner, R. R. Hernandez and M. Schellenberg, "CMOS Hearing Aid Device", *Journal on Analog Integrated Circuits and Signal Processing*, Vol. 21, No. 2, pp. 163-172 (1999).

[MDMA+95] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu and J. Yamada, "1-V Power Supply High-Speed Digital Circuit Technology

with Multi-threshold Voltage CMOS," *IEEE Journal of Solid State Circuits*, Vol. 30, No.8, pp. 847-853 (1995).

[Mea94] C. Mead, "Scaling of MOS Technology to Sub-Micrometer Feature Sizes," *Journal of VLSI Signal Processing,* pp. 9–25 (1994).

[MEDICI96] Technology Modeling Associates. MEDICI Manual, version 2.3.1 (1996).

[Mei88] B. Meinerzhagen, "Consistent gate and substrate current modeling based on energy transport and the lucky electron concept," *IEDM Technical Digest*, pp. 504-507(1988).

[Mei95] J. Meindl, "Low Power Microelectronics: Retrospect and Prospect," *Proceedings of the IEEE*, vol. 83, no. 4, pp. 619--635, (1995).

[MMP94] R. Marculescu, D. Marculescu, and M. Pedram, "Switching Activity Analysis Considering Spatiotemporal Correlations", *Proc. 1994 Int. Conference on Computer Aided Design*, pp. 294–299, November 1994.

[MNOY+00] H. S. Momose, S. Nakamura, T. Ohguro, T. Yoshitomi, E. Morifuji, T. Morimoto, Y. Katsumata and H. Iwai, "Hot-carrier Reliability of Ultra-thin Gate Oxide CMOS," *Solid State Electronics*, Vol. 44, pp. 2035-2044 (2000).

[MR00] A. A. Mutlu, M. Rahman, "Two-dimensional Analytical Model for Drain Induced Barrier Lowering (DIBL) in Short Channel MOSFET's," *Proceedings IEEE*, pp. 340-344 (2000).

[MTIT+92] T. Mizuno, A. Torriumi, M. Iwase, M Takahashi, H. Niiyama, M. Fukumoto, and M. Yoshimi, "Hot-Carrier Effect in 0.1 μm Gate Length CMOS Devices," *IEEE IEDM*, pp. 695-698 (1992).

[MW95] J. J. Maa and C. Y. Wu, "A new constant-field scaling theory for MOSFETs," *IEEE Transactions on Electron Devices*, Vol. 42, No. 7, pp.1262-1268 (1995).

[Naj93] F. N. Najm, "Transition Density: A New Measure of Activity in Digital Circuits", *IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, no. 2, pp. 310–323 (1993).

[Naj95] F. N. Najm, "Towards a High - level Power Estimation Capability," *Proceedings of the 1995 International Symposium on Low Power Design*, pp. 87–92 (1995).

[NBDA+01] S. Narendra, S. Borkar, V. De, D. Antoniads, and A. Chandrakasan, "Scaling of Stack Effect and its Application for Leakage Reduction," in *Proceedings of the International Symposium on Low Power Electronics and Design*, ISLPED, pp. 195-200 (2001).

[NBYH90] F. N. Najm, R. Burch, P. Yang, and I. N. Hajj, " Probabilistic Simulation for Reliability Analysis of CMOS Circuits", *IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems*, vol. 9, no. 4, pp. 439–450 (1990).

[NDOC+03] D. Nguyen, A. Davare, M. Orshansky, D. Chinnery, B. Thompson, and K. Keutzer, "Minimization of Dynamic and Static Power Through Joint Assignment of Threshold Voltages and Sizing Optimization", *International Symposium on Low Power Electronics and Design*, pp.158-163, Aug. 2003.

[NGH95] F. N. Najm, S. Goel, and I. Hajj, "Power Estimation in Sequential Circuits," *Proceedings of the 32nd Design Automation Conference*, pp. 635–640 (1995).

[NS00] K. Nose and T. Sakurai, "Optimization of V_{DD} and V_{TH} for low power and high speed application," *DAC*, pp. 469-474 (2000).

[OC72] Y. Okuto, C. R. Crowell, "Energy-conservation considerations in the characterization of impact ionization in semiconductors," *Physical Review* B, No. 6, pp. 3076-811 (1972).

[OTWC+81] S. Ogura, P. Tsang, W. Walker, D. Critchlow, and J. Shepard, "Elimination of Hot Electron Gate Current by Lightly Doped Drain-Source Structure," *Proceedings of the International Electron Devices Meeting,* pp. 651 (1981).

[PB03] M. Pattanaik and S. Banerjee, "A New Approach to Analyze a Sub-micron CMOS Inverter," *Proceedings of the 16th International conference on VLSI Design* (2003).

[PDC98] P. Pant, V. K. De and A. Chatterjee, "Simultaneous Power Supply, Threshold Voltage, and Transistor Size optimization for Low-Power Operation of CMOS Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 6, No. 4, pp. 538-545 (1998).

[PDW01] J. K. Park, H. V. Deshpande and J. C. S. Woo, "Enhanced Subthreshold Leakage Current due to Impact Ionization in Deep Sub-100nm N-Channel Double-Gate MOSFETs," *IEEE International SOI Conference*, pp. 147-148 (2001).

[PJYP+96] J. T. Park, S.J. Jang, C. G. Yu, B. G. Park and J. D. Lee, "New Experimental Findings on Hot Carrier Effects in Deep Sub-micrometer Surface Channel PMOS," *Microelectronics Reliability*, Vol. 36, No. 11/12, pp. 1659-1662 (1996).

[Pow95] R. A. Powers, "Batteries for Low Power Electronics," *Proceedings IEEE*, Vol. 38, No. 4, pp. 687–693 (1995).

[PR94] S. C. Prasad, and K. Roy, "Circuit Optimization for Minimization of Power Consumption under Delay Constraint," *Proceedings of the 1994 International Workshop on Low Power Design*, pp.15–20 (1994).

[Rei83] A. Reismann, "Device, Circuit, and Technology Scaling to Micron and Submicron Dimensions," *Proceedings IEEE*, Vol. 71, No. 5, pp. 550-565 (1983).

[RHA92] F. Rouatbi, B. Haroun, and A. J. Al-Khalili, "Power estimation tool for sub-micron CMOS VLSI circuits", *IEEE/ACM International Conference on Computer-Aided Design*, pp. 204–209 (1992).

[RP00] K. Roy and S. C. Prasad, Low-power CMOS VLSI circuit design, A wiley Interscience Publication, New York, 2000.

[Sak00] T. Sakurai, "Reducing power consumption of CMOS VLSI's through V_{DD} and V_{TH} Control," in *First International Symposium on Quality Electronic Design*, IEEE, pp. 417-423 (2000).

[Sak03] T. Sakurai, "Perspectives on Power-Aware Electronics," in *Proceedings of the International Solid-State Conference*, pp. 26—29 (2003).

[SEOP+02] S. Sirichotiyakul, T. Edwards, C. Oh, R. Panda, and D. Blaauw, Duet: An Accurate Leakage Estimation and Optimization Tool for Dual Vt Circuits", *IEEE Transactions on VLSI*, pp. 79- 90, 2002.

[SGSH02] P. Su, K. I. Goto, T. Sugii and C. Hu, "A Thermal Activation View of Low Voltage Impact Ionization in MOSFETs," *IEEE Electron Devices Letters*, Vol.23, No.9, pp.550-552 (2002).

[SH81] H. Shichijo and K. Hess, "Band structure dependent transport and impact - ionization in GaAs," *Physical Review B*, vol. 23, no. 8, pp. 4197-4207 (1981).

[Sho61] W. Shockley, "Problems Related to p-n Junctions in Silicon," *Solid-State Electronics*, Vol. 2, No. 1, pp. 35–67 (1961).

[SIA] Semiconductor Industry Association (SIA), The National Technology Roadmap for Semiconductors. [Online] Available WWW: http://www.sematech.org/public/roadmap/index.htm.

[SIA03] Semiconductor Industry Association, ITRS, 2003 Ed. http://www.public.itrs.net

[SK01] D. Sylvester and H. Kaul, "Power-driven challenges in nanometer deign," IEEE Design and Test of Computers, Vol. 18, No. 6, pp.12-21 (2001).

[SL94] C. Svensson and D. Liu, "A Power Estimation Tool and Prospects of Power Savings in CMOS VLSI Chips," *Proceedings of the 1994 International Workshop on Low-Power Design*, pp.171–176 (1994).

[SL96] C. Svensson and D. Liu, "Low Power Circuit Techniques," *Low Power Design Methodologies*, P. M. Rabaey and M. Pedram, *Eds. Dordrecht, The Netherlands*: Kluwer (1996).

[SPS02] O. Semenov, A. Pradzynski, and M. Sachdev, "Impact of Gate Induced Drain Leakage on Overall Leakage of sub-micron CMOS VLSI Circuits," *IEEE Transactions on Semiconductor Manufacturing*, pp. 9-18 (2002).

[SRP00] H. Soeleman, K. Roy and B. C. Paul, "Sub-Domino Logic: Ultra-Low Power Dynamic Sub-Threshold Digital Logic," *Proceedings IEEE*, pp.211-214 (2000).

[SRP01] H. Soleman, K. Roy and B. C. Paul, "Robust Sub threshold Logic for Ultra-Low Power Operation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 9, No. 1, pp. 90-98 (2001).

[SS94] P. Schneider, and U. Schlichtmann, "Decomposition of Boolean Functions for Low Power Based on a New Power Estimation Technique," *Proceedings of the 1994 International Workshop on Low Power Design,* pp. 123–128 (1994).

[SSB04] A. Shrivastava, D. Sylvester and D. Blaauw, "Power minimization using simultaneous gate sizing, dual-Vdd and dual-Vth assignment," *Design Automation Conference*, pp. 783-787, 2004.

[ST95] S. W. Sun and P. G. Y. Tsui, "Limitation of CMOS Supply Voltage Scaling by MOSFET Threshold-Voltage Variation," *IEEE Journal of Solid-State Circuits*, Vol. 30, pp. 947–949 (1995).

[Sta01] M. R. Stan, "Low–Power CMOS with Sub-Volt Supply Voltages", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 9, No. 2, pp. 394-399 (2001).

[SVS02] O. Semenov, A. Vassighi and M. Sachdev, "Impact of Technology Scaling on Thermal Behavior of Leakage Current in Sub-Quarter Micron MOSFETs: perspective of low temperature current testing," *Microelectronics Journal*, Vol. 33, pp. 985-994 (2002).

[Sze81] S. M. Sze, "Physics of Semiconductor Devices," Wiley *Eastern Limited, New Delhi* (1981).

[TAM93] V. Tiwari, P. Ashar, and S. Malik, "Technology Mapping for Low Power," *Proceedings of the 30th Design Automation Conference*, pp. 74–79 (1993).

[TBCF+97] Y. Taur, D. Buchanan, W. Chen, D. Frank, K. Ismail, S. H. Lo, G. Sai-Halasz, R. Viswanathan, H. J. C. Wann, S. Wind, and H. S. Wong, "CMOS Scaling into the Nanometer Regime," *Proceedings IEEE*, Vol. 85, pp. 486 (1997).

[TD00] S. B. Thakare and A. K. Dutta, "A New Improved Model for Sub Threshold Slope for Submicron MOSFETs," *Microelectronics Journal*, pp. 105-111 (2000).

[THHM+83] S. Tam, F. C. Hsu, C. Hu, R. S. Muller, and P. K. Ko, "Hot-Electron Currents in Very Short-Channel MOSFETS," *IEEE Electron Device Letters*, Vol. EDL-4, pp. 249 (1983).

[TKTA82] E. Takeda, H.i Kume, T.Toyabe, and S. Asai, "Sub-micrometer MOSFET Structure for Minimizing Hot-carrier Generation," *IEEE*

Transactions on Electron Devices, Vol. ED-29, No. 4, pp. 611-618 (1982).

[TN97] Y. Taur and E. Nowak, "CMOS Devices below 0.1M: How High Will Performance Go?", *Proceedings of the International Electron Devices Meeting*, pp. 215 (1997).

[TN98] Y. Taur and T. H. Ning, "Fundamentals of Modern VLSI Devices," Cambridge University Press, (1998).

[TPD93] C. Y. Tsui, M. Pedram, and A. M. Despain, "Efficient Estimation of Dynamic Power Dissipation under a Real Delay Model," *Proceedings of the IEEE International Conference on Computer Aided Design*, pp. 224–228 (1993).

[WC05] A. Wang and A. Chandrakasan, "A 180-mV Subthreshold FFT Processor Using a Minimum Energy Design Methodology," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 310-319 (2005).

[WCJR+98] L. Wei, Z. Chen, M. Johnson, K. Roy, K and V. De, "Design and Optimization of Low Voltage High Performance Dual Threshold CMOS Circuits", *Design Automation Conference*, pp. 489-494 (1998).

[WCK02] A. Wang, A. P. Chandrakasan and S. V. Kosonocky, "Optimal Supply and Threshold Scaling for Sub threshold CMOS Circuits," *International Symposium on VLSI*, pp. 7-11 (2002).

[WH99] G. Wei and M. Horowitz, "A Fully Digital, Energy-Effcient Adaptive Power-Supply Regulator," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 4, pp. 520-528 (1999).

[Wol54] P. A. Wol, "Theory of Electron Multiplication in Silicon and Germanium," *Physical Review,* Vol. 95, No. 6, pp.1415–1420 (1954).

[Won00] H. Wong, "Drain Breakdown in Submicron MOSFETs: A Review," *Journal of Microelectronics Reliability*, Vol. 40, pp. 3-15 (2000).

[Woo87] R. C. Woods, "'Soft' Energy Thresholds in Impact Ionization: A Classical Model," *IEEE Transactions on Electron Devices,* Vol. ED-34 (1987).

[WV98] Q. Wang and S. Vrudhula, "Static power optimization of deep submicron CMOS circuit for dual Vt technology", *International Conference on Computer-Aided Design*, pp. 490-496, 1998.

[WZRC+02] L. Wei, R. Zhang, K. Roy, Z. Chen, and D. Janes, "Vertically Integrated SOI Circuits for Low power and High-performance Applications," *IEEE Transactions on Very Large Scale Integration*, vol. 10, no. 3, pp. 351-362 (2002).

[Yan97] R. Yang, "Experimental Exploration of Ultra-low Power CMOS Design Space using SOIAS Dynamic VT Control Technology," in *The IEEE International SOI Conference*, pp. 76-77 (1997).

[YY90] K. Yano and T. Yamannaka, "A 3.8-ns CMOS 16×16 multiplier using complementary pass-transistor logic," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 4, pp. 388-393 (1990).

[ZBSF04] B. Zhai, D. Blaauw, D. Sylvester, K. Flautner, "Theoretical and Practical Limits of Dynamic Voltage Scaling," *DAC'04*, San Diego, California, USA, June 7-11 (2004).

[ZF97] R. Zimmermann and W. Fichner, "Low-Power Logic Styles: CMOS versus Pass-Transistor Logic," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 7, pp. 1079-1090 (1997).

LIST OF PUBLICATIONS

[1] Bhavana Jharia, S. Sarkar and R. P. Agarwal, "Analytical Study of Impact Ionization and Subthreshold Current in Submicron n-MOSFET," ***Proc. of IEEE 6th International Symposium on Quality Electronic Design***, San Jose, CA, U.S.A., pp.72-76, March21-23, 2005.

[2] Bhavana Jharia, S. Sarkar, and R. P. Agarwal, "Drain-Source Voltage Dependence of MOSFET Threshold Voltage," ***Proc. of the National Conference on Electronics Circuits and Communication Systems (ECCS-2004)***, Thapar Institute of Engineering & Technology, Patiala, pp. 343-345, Sept. 23-24, 2004.

[3] B. Jharia, S. Sarkar and R. P. Agarwal, "Effect of Impact Ionization on Subthreshold Current in Submicron n-MOSFET," ***WSEAS*** *(World Scientific and Engineering Academy and Society)* ***Transaction on Circuits and Systems***, Vol.-3, Issue -3, pp. 426-430, May 2004.

[4] B. Jharia, S. Sarkar and R. P. Agarwal, "Controlling Factors of Drain Induced Barrier Lowering Coefficient in Short Channel MOSFET," ***WSEAS Transaction on Electronics***, Vol.-1, Issue -2, pp. 252-256, April 04.

[5] Bhavana Jharia, S. Sarkar, R. P. Agarwal and S. C. Handa, " Low Power Techniques for VLSI Hearing Assistive Devices," ***Proc. of 10th National Seminar on Physics and Technology of Sensor (NSPTS-10)***, Pune, pp. 32-37, March 4-6, 2004.

[6] Bhavana Jharia, S. Sarkar and R. P. Agarwal, "Towards the Nanoscale: Influence of Scaling on the Impact Ionization and Subthreshold Behavior of MOSFETs," communicated to ***Holon Academic Institute of Technology (HAIT) Journal of Science and Engineering (Series B: Applied Science and Engineering)***.

[7] Bhavana Jharia, S. Sarkar and R. P. Agarwal, "Optimum Supply Voltage for Minimum Power and Energy," communicated to ***IEEE 8th International Symposium on Quality Electronic Design***, San Jose, CA, U.S.A., March 17-19, 2008.