Experimental Study of Electron Velocity Overshoot in Silicon Inversion Layers

by

Hang Hu

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1994

© Massachusetts Institute of Technology 1994. All rights reserved.

Author ....
Department of Physics
August 30, 1994

Certified by ....
Dimitri A. Antoniadis
Professor
Thesis Supervisor

Certified by ....
Henry I. Smith
Professor
Thesis Supervisor

Certified by ....
Mildred S. Dresselhaus
Institute Professor
Thesis Supervisor

Accepted by ....
George F. Koster
Chairman, Graduate Committee

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

OCT 14 1994

LIBRARIES
Science

Experimental Study of Electron Velocity Overshoot in Silicon Inversion Layers

by

Hang Hu

Submitted to the Department of Physics
on August 30, 1994, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

This thesis work is an investigation of the physical effects associated with the non-equilibrium electron transport dynamics in the inversion layers of deep-submicron silicon MOSFETs. These effects are the electron velocity overshoot, the hot-electron injection barrier lowering, and the hot-carrier "cooling" effect. These macroscopic effects reveal the microscopic mechanisms behind the unique transport dynamics of high-energy electrons and holes under non-equilibrium conditions such as high electric fields and electric-field gradients in deep-submicron MOSFETs. They also have profound impact on one of the most important issues regarding the evaluation of the silicon MOSFET devices and thus the ULSI industry, the deep-submicron MOSFET scaling. By pursuing the understanding of these physical effects from the perspective of MOSFET scaling, one can take a different view of the MOSFET scaling issues from the traditional one based on the classical scaling theory, and thus follow a new methodology that is theoretically plausible and practically efficient to unify all fundamental quantities in deep-submicron MOSFET scaling and providing the insight of deep-submicron MOSFET design.

High performance sub-0.1 μm MOSFET devices (SSR MOSFETs) using X-ray lithography, self-aligned $CoSi_2$ silicide formed by Ti/Co laminates, super-steep retrograde channel doping, and ultra-shallow source/drain extension structure with "halo" doping are demonstrated. These SSR MOSFETs exhibit the best-to-date performance with a given amount of short-channel effect. X-ray lithography is proven to be a highly promising lithography technology for deep-submicron MOSFET fabrication. The ultra-shallow source/drain extension structure coupled with Ti/Co bimetallic $CoSi_2$ silicide is demonstrated to be highly effective in controlling short-channel effects and minimizing parasitic resistance. Super-steep retrograde channel doping is shown to be highly effective in preventing device punchthrough, while maintaining the device electrostatic integrity. The excellent overall behavior of these SSR MOSFETs guarantees unambiguous device measurements.

The electron velocity overshoot in silicon inversion layers is investigated using sub-0.1 μm SSR MOSFETs. It is found that the average electron velocity is not yet in the overshoot regime. From the perspective of deep-submicron MOSFET scaling, there exists a trade-off between the electron velocity and the device short-channel effects, such as the drain-induced barrier lowering effect and the punchthrough.

The correlation between gate and substrate currents in n-channel MOSFETs with L_{eff} down to 0.1 μm is investigated within the general framework of the lucky-electron model. It is found for the first time that the correlation coefficient, Φ_b/Φ_i, decreases with decreasing L_{eff} in the 0.1 μm regime. This hot-electron injection barrier lowering effect is confirmed by numerical simulations which incorporate non-equilibrium dynamical effects. This effect indicates the increasing decoupling between channel hot-electron injection and impact ionization with decreasing L_{eff} in the deep-submicron regime.

The hot-carrier "cooling" effect is investigated at both 300 K and 77 K. The reduction in the normalized substrate current with decreasing L_{eff} is not observed at both temperatures. The same observation is made for p-channel MOSFETs. The normalized gate current is characterized with decreasing L_{eff}. It is found that there is no indication of gate current reduction with L_{eff} either.

The scaling relationships among all the fundamental quantities of deep-submicron MOSFETs, device speed, drain-induced barrier lowering, effective channel length, and hot-carrier-induced currents are investigated with both device measurements and numerical simulations following a new methodology using nonlinear regressions. The dependence of these relationships on channel and source/drain parameters is studied. With this new scaling methodology, all fundamental quantities of MOSFET scaling are unified in the deep-submicron regime. It is found that the scaling relationships among them can be expressed in appropriate power-law forms with excellent statistical significance for both experimental and simulation data samples, there exist universal trade-off relationships among the performance, the short-channel effect, and the hot carrier currents, with respect to the threshold voltage and the channel doping profile, and the trade-off between the performance and the short-channel effect is dominated by the source/drain parameters.

Thesis Supervisor: Dimitri A. Antoniadis
Title: Professor

Thesis Supervisor: Henry I. Smith
Title: Professor

Thesis Supervisor: Mildred S. Dresselhaus
Title: Institute Professor

Acknowledgments

I would like to thank my thesis advisor, Professor Dimitri Antoniadis, for the enormous support and courage during the cause of this thesis work. I would like to thank Professor Henry Smith for providing me the unique opportunity to work with him and produce what we all wanted. I would like to thank my thesis co-advisor, Professor Mildred Dresselhaus for supervising the final stage of the thesis work.

It would be simply impossible to carry out such a complex project without the tremendous support from the fellow graduate students and the staff members of the MTL community. I would like to thank Dr. Lisa Su and Mr. Jarvis Jacobs for all the help and the countless technical discussions, Ms. Isabel Yang for the excellent work on X-ray lithography, Dr. Hao Fang for the device processing, and Ms. Melanie Sherony for the device fabrication. The happy ending of this project would never come without the expert and much-needed help from Mr. Vincent Wong, Dr. William Chu, Mr. Euclid Moon, and Mr. Martin Burkhardt on X-ray lithography. And of course, nothing would happen without the brute-force management of Mr. James Carter, together with his army, Mr. Mark Mondol, Ms. Jeanne Porter, and Mr. Bob Sisson, who made our SSL (it is now called NSL) so much productive. The MIT ICL and TRL staff are gratefully acknowledged for their generous support on the device fabrication. Special thanks go to Mr. David Breslau of MIT CSR who was always there to provide expert help on machining.

This thesis project is a natural extension of some of the excellent early work done by Dr. Stephen Chou, Dr. Ghavam Shahidi, Mr. Shiang The, and Mr. Greg Carlin. I would like to thank Dr. Ghavam Shahidi for the technical discussions and for setting up an example that was so hard to reach that almost kept me in this place indefinitely. I would also like to thank Drs. Ran Yan and Kwen Lee of AT&T Bell Labs for allowing me to measure their devices.

My graduate life would be too boring to bear without the fun times with my lunch and dinner pals, Vince, Iz, Lisa, Joe, and Ken. I would like to thank Douglas for being such a special friend and such a wonderful guitar player. I had an increditable

experience with our southern blues. The endless fun of CD critic with Heng Li really got me going when I was too exhausted in the lab.

I would not have the honor of completing this thesis work without the help and guidance from my mentors and teachers during my early years at MIT. I owe Drs. Tom Chang, John Foster, and Stephen Jasperson forever for their recognizing me and providing me a chance when the times were rough and not as easy for a youngster with twenty-five dollars in his pocket when he first came here. I am indebted to Henry and Mary Ann for their generosity and kindness. This thesis is dedicated to my parents, who always love me, believe in me, and support me through the good times as well as the bad. I am indebted to my little sister for leaving her so early and not seeing her for so long, and I hope that she understands why. And to Yuan Tao, I thank you for all the times that we were together.

This thesis work is partly supported by SRC(Contract 93-SC-309), ARPA, JSEP, IBM, and Motorola.

To My Parents.

Contents

1 Introduction **17**

2 Sub-0.1 μm MOSFET: Design, Fabrication, and Characterization **20**

2.1 Sub-0.1 μm n-Channel MOSFET Design 20

2.1.1 Design objectives 20

2.1.2 Channel design 21

2.1.3 Source/drain design 23

2.2 Sub-0.1 μm n-Channel MOSFET Device
Fabrication 26

2.2.1 X-ray lithography 28

2.2.2 Channel doping 30

2.2.3 Source/drain junction formation 32

2.3 Sub-0.1 μm n-Channel MOSFET
Characterization 33

2.4 Conclusion 34

3 Physics of Velocity Overshoot in Perspective of MOSFET Scaling **37**

3.1 Electron Velocity Overshoot in Si MOSFETs 37

3.1.1 Hydrodynamic transport model 37

3.1.2 Dynamics of electron velocity overshoot in MOSFETs 41

3.2 Experimental Observations 46

3.2.1 Experimental techniques 46

3.2.2 Experimental results 49

3.3 Conclusion 60

4 **Deep-Submicron MOSFET Scaling: Methodology and Analysis** **62**

4.1 Background 62

4.2 Methodology 64

4.3 Analysis 66

4.3.1 Channel parameters 66

4.3.2 Source/drain parameters 75

4.4 Discussion 77

4.5 Conclusion 83

5 **Physics of Non-Equilibrium Hot-Carrier Effects** **85**

5.1 Theory of Hot-Carrier Current Generation in Si MOSFETs 85

5.1.1 Channel electric field 86

5.1.2 Substrate current generation 88

5.1.3 Gate current generation: the lucky-electron model 89

5.2 Hot-Electron Injection Barrier Lowering 90

5.2.1 Experimental observations 92

5.2.2 Analysis 97

5.3 Hot-Carrier "Cooling" Effect 100

5.3.1 Hot-electron "cooling" effect and non-equilibrium transport dynamics 102

5.3.2 Experimental observations at room temperature 105

5.3.3 Experimental observations at low temperature 111

5.4 Hot-Carrier Effects in Scope of MOSFET Scaling 113

5.4.1 Hot-carrier effect: a fourth dimension to MOSFET scaling . . 113

5.4.2 Methodology and analysis 114

5.5 Conclusion 120

6 **Conclusion** **122**

List of Figures

2-1 Super-Steep-Retrograde (SSR) and Step (STEP) channel doping profiles with ion-implant conditions summarized in Table 2.1. 22

2-2 Vertical source/drain extension junction profiles obtained by SIMS (after G. G. Shahidi *et al.* [55]). 24

2-3 The cross-sectional schematic of a sub-0.1 μm SSR n-channel MOSFET device (not to scale). The "halo" doping is an approximately 10 nm wide p-type region (indium implant) around the n^+-type arsenic source/drain extension. 25

2-4 The simulated subthreshold characteristics of a $L_{eff} = 0.1\ \mu m$ n-channel MOSFET with $t_{ox} = 5.0\ nm$ and a SSR-III channel doping profile described in Table 2.1 and Fig. 2-1. 26

2-5 The simulated saturation transconductance, g_m, of n-channel MOSFETs with $t_{ox} = 5.0\ nm$, $L_{eff} = 0.1\ \mu m \sim 0.5\ \mu m$, and SSR-III channel doping profiles described in Table 2.1 and Fig. 2-1. The saturation drain voltage $V_{ds} = 2.0\ V$. 27

2-6 The schematic of the proximity X-ray lithography. 28

2-7 The custom-built X-ray lithography alignment and exposure system [37]. 29

2-8 SEM micrograph of a resist(SAL-601)/LTO-mask gate structure defined by X-ray lithography and CHF_3 etching. The SAL-601 resist is about 400 nm thick, and the LTO hard-mask is about 50 nm thick. . 30

2-9 SEM micrograph of a polysilicon gate structure defined by LTO hardmask and Cl_2 etching. The polysilicon gate is about 300 nm thick. . 31

2-10 The drain current, I_d, *vs.* drain voltage, V_{ds}, characteristics of a $L_{eff} =$ 0.085 μm n-channel MOSFET with a SSR-III channel doping profile. V_{gs} steps of 0.2 V. 35

2-11 The drain current, I_d, *vs.* gate voltage, V_{gs}, characteristics of a $L_{eff} =$ 0.085 μm n-channel MOSFET with a SSR-III channel doping profile. V_{ds} steps of 0.2 V. 35

3-1 A schematic of the coordinate system used for demonstrating electron transport processes in a MOSFET. 42

3-2 Electron velocity calculated by Eqs. 3.24 and 3.25 for various low-field mobilities, μ_0. The right-hand Y-axis shows the applied lateral electric field. 45

3-3 Electron temperature calculated by Eqs. 3.24 and 3.25 for various low-field mobilities, μ_0. The right-hand Y-axis shows the applied lateral electric field. 45

3-4 Calculated (Monte Carlo simulation) electron velocity for three silicon MOSFETs plotted against the distance from the source along the Si-SiO_2 interface. For the channel length $L_c = 43\ nm$ device, $V_{gs} = 0.7\ V$ and $V_{ds} = 0.6\ V$; for the channel length $L_c = 103\ nm$ device, $V_{gs} =$ 1.0 V and $V_{ds} = 1.0\ V$; for the channel length $L_c = 233\ nm$ device, $V_{gs} = 2.5\ V$ and $V_{ds} = 2.5\ V$. All sources and substrates are grounded and $T_0 = 300\ K$. The velocity is in the source-to-drain direction, averaged over a depth of 10 nm below the interface. 47

3-5 Calculated (Monte Carlo simulation) electric field for three silicon MOSFETs in Fig. 3-4 plotted against the distance from the source along the Si-SiO_2 interface. $T_0 = 300\ K$. The electric field is in the source-to-drain direction, averaged over a depth of 10 nm below the interface. 47

3-6 Measured saturated transconductance per unit width, g_m/W, *vs.* effective channel length, L_{eff}, for NMOSFETs with the four channel doping profiles described in Table. 2.1. The right-hand Y-axis shows the corresponding electron velocity, g_m/WC_{ox}, and $T = 300\ K$. 50

3-7 Comparison between the uncorrected (extrinsic) and corrected (intrinsic) saturated transconductance per unit width, g_m/W, *vs.* effective channel length, L_{eff}, for NMOSFETs with the SSR-I channel doping profile described in Table. 2.1. The right-hand Y-axis shows the corresponding electron velocity, g_m/WC_{ox}, and $T = 300\ K$. 51

3-8 The $\log I_d$ *vs.* V_{gs} characteristics of a $L_{eff} = 0.21\ \mu m$ device with a SSR-I channel doping profile described in Table 2.1, and $T = 300\ K$. 54

3-9 The $\log I_d$ *vs.* V_{gs} characteristics of a $L_{eff} = 0.15\ \mu m$ device with a SSR-I channel doping profile described in Table 2.1, and $T = 300\ K$. 54

3-10 The $\log I_d$ *vs.* V_{gs} characteristics of a $L_{eff} = 0.10\ \mu m$ device with a SSR-I channel doping profile described in Table 2.1, and $T = 300\ K$. 55

3-11 The $\log I_d$ *vs.* V_{gs} characteristics of a $L_{eff} = 0.055\ \mu m$ device with a SSR-I channel doping profile described in Table 2.1, and $T = 300\ K$. 55

3-12 Electron velocity, g_m/WC_{ox}, as a function of drain-induced barrier lowering, $\delta V_t/\delta V_{ds}$, with L_{eff} as an implicit variable. All NMOSFETs have a SSR-I channel doping profile described in Table 2.1, $V_{ds} = 2.0\ V$, and $T = 300\ K$. 56

3-13 Electron velocity, g_m/WC_{ox}, as a function of off-current, I_{off}, with L_{eff} as an implicit variable. All NMOSFETs have SSR-I channel doping profiles described in Table 2.1, $V_{ds} = 1.4\ V$, and $T = 300\ K$. 56

3-14 A schematic diagram of the electron transport under the vertical electric field control due to the gate and the lateral electric field control due to the drain in a MOSFET. 57

4-1 Measured drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.085 μm to 0.4 μm, $t_{ox} = 5.3$ nm and the four channel doping profiles (SSR-I, II, III and STEP) as shown in Fig. 2-1 and Table. 2.1. A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7. 66

4-2 Simulated drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.07 μm to 0.4 μm, $t_{ox} = 3.3, 5.3, 7.3, 10.3$ nm, $x_j = 50$ nm and uniform channel doping profile ($N_a = 1 \times 10^{17}$ cm^{-3}). A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7. . . 67

4-3 Simulated drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.07 μm to 0.4 μm, $t_{ox} = 3.3, 5.3, 7.3, 10.3$ nm, $x_j = 30$ nm and uniform channel doping profile ($N_a = 1 \times 10^{17}$ cm^{-3}). A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7. . . 68

4-4 Measured device speed, g_m/WC_{ox}, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.085 μm to 0.4 μm, $t_{ox} = 5.3$ nm and the four channel doping profiles (SSR-I, II, III and STEP) as shown in Fig. 2-1 and Table. 2.1. A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7. . . 70

4-5 Simulated device speed, g_m/WC_{ox}, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.07 μm to 0.4 μm, $t_{ox} = 3.3, 5.3, 7.3, 10.3$ nm, $x_j = 50$ nm and uniform channel doping profile ($N_a = 1 \times 10^{17}$ cm^{-3}). A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7. 71

4-6 Measured device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for NMOSFETs with L_{eff} ranging from 0.085 μm to 0.4 μm, $t_{ox} = 5.3\ nm$ and the four channel doping profiles (SSR-I, II, III and STEP) as shown in Fig. 2-1 and Table. 2.1. A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7. 73

4-7 Simulated device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for NMOSFETs with L_{eff} ranging from 0.07 μm to 0.4 μm, $t_{ox} = 3.3, 5.3, 7.3, 10.3\ nm$, $x_j = 50\ nm$ and uniform channel doping profile ($N_a = 1 \times 10^{17}\ cm^{-3}$). A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7. . . 74

4-8 Measured device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for NMOSFETs with identical device structures except $t_{ox} = 6.5\ nm$ and $t_{ox} = 9.0\ nm$, respectively. 75

4-9 Experimental device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for two sets of NMOSFETs with and without an indium "halo" doping structure and with the same V_t. Empty symbols: $V_t = 0.15\ V$. Filled symbols: $V_t = 0.57\ V$. 76

4-10 Simulated device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for NMOSFETs with $x_j = 30, 100, 150\ nm$ and uniform source/drain junctions. 77

4-11 Comparison of measured device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for various recent 0.1 μm technologies given in the references. 78

4-12 A hypothetical trade-off curve in the form of a power-law as expressed in Eq. 4.5 in g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ space. The arrows show the movement of the "device design point", $P(L_{eff}, t_{ox}, V_t)$, according to the changes in t_{ox} and V_t. 79

4-13 Threshold voltage roll-off behavior, V_t *vs.* L_{eff}, for NMOSFETs with SSR and STEP channel dopings, as shown in Fig. 2-1 and Table. 2.1. 80

4-14 Drain-induced barrier lowering behavior, $\frac{\delta V_t}{\delta V_{ds}}$ *vs.* L_{eff}, for NMOSFETs with SSR and STEP channel dopings, as shown in Fig. 2-1 and Table. 2.1. 80

4-15 Punchthrough characteristics for NMOSFET with SSR-I,II,III and STEP channel doping profiles, as shown in Fig. 2-1 and Table. 2.1 The percentage change in the subthreshold slope, or the S-factor, $\frac{\Delta S(V_{ds})}{S(V_{ds})}$ is defined in Eq. 4.8. 83

5-1 A schematic of hot-carrier-induced current generation in a n-channel MOSFET. 86

5-2 Gate current, I_g, and substrate current, I_b, characteristics as a function of drain voltage, V_{ds}, for the $L_{eff} = 0.1\ \mu m$ SSR-III MOSFET with V_{gs} steps of 0.1 V from 2.01 V to 3.01 V. The arrows indicate the axes which the curve sets are associated to. 92

5-3 The correlation between normalized gate current, I_g/I_d, and normalized substrate current, I_b/I_d, for the $L_{eff} = 0.1\ \mu m$ SSR NMOSFET with constant $V_{gs} - V_{ds}$ steps of 0.1 V from $-1.5\ V$ to 1.0 V. 93

5-4 Measured correlation coefficient, Φ_b/Φ_i, *vs.* gate voltage, V_{gs}, with constant $V_{gs} - V_{ds} = 0\ V$ for the SSR NMOSFETs with $L_{eff} = 0.10\ \mu m$, 0.13 μm, 0.18 μm, and 0.20 μm. The straight lines are obtained from linear regressions on the data points. 94

5-5 Measured Φ_b/Φ_i ratio as a function of gate voltage V_{gs} with a set of fixed $V_{gs} - V_{ds}$ steps 0.1 V from -0.3 V to 0.3 V for the $L_{eff} = 0.1\ \mu m$ SSR NMOSFET shown in Fig. 5-2. The straight lines are obtained from linear regressions on the data points. 96

5-6 Simulated correlation coefficient, Φ_b/Φ_i, *vs.* gate voltage, V_{gs}, with constant $V_{gs} - V_{ds} = 0\ V$ for two NMOSFETs with $L_{eff} = 0.14\ \mu m$ and 0.20 μm. The straight lines are obtained from linear regressions on the data points. 100

5-7 Three displaced Gaussian distributions representing the energy distribution of channel hot electrons in three MOSFETs with different L_{eff}s. $L_{eff1} > L_{eff2} > L_{eff3}$, $v_{d3} > v_{d2} > v_{d1}$, $T_{e1} > T_{e2} > T_{e3}$, and $A_2 > A_1 > A_3$. 103

5-8 Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}s, and SSR-I channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2. 107

5-9 Measured normalized gate current, I_g/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-I channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2. 107

5-10 Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-II channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2. 108

5-11 Measured normalized gate current, I_g/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-II channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2. 108

5-12 Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-III channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2. 109

5-13 Measured normalized gate current, I_g/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-III channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2. 109

5-14 Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and STEP channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2. 110

5-15 Measured normalized gate current, I_g/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and STEP channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2. 110

5-16 Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 1\ V$ for NMOSFETs with various L_{eff} at $77\ K$. Courtesy of AT&T Bell Labs for device fabrication [33]. 112

5-17 Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 1\ V$ for NMOSFETs with various L_{eff} at $300\ K$. Courtesy of AT&T Bell Labs for device fabrication [33]. 112

5-18 Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = -1\ V$ for PMOSFETs with various L_{eff}at $77\ K$ and $300\ K$. The PMOSFETs used here are described in [24]. 113

5-19 Measured normalized substrate current, I_b/I_d, as a function of effective channel length, L_{eff}, at a given drain voltage, $V_{ds} = 1.5\ V$ and $3\ V$, for NMOSFETs with the three SSR and STEP channel dopings described in Fig. 2-1 and Table. 2.1 in Chapter 2. $V_{gs} = 2\ V$ and $3\ V$. 116

5-20 Measured normalized gate current, I_g/I_d, as a function of effective channel length, L_{eff}, at a given drain voltage, $V_{ds} = 3\ V$, for NMOSFETs with the three SSR and STEP channel dopings described in Fig. 2-1 and Table. 2.1 in Chapter 2. $V_{gs} = 2\ V$ and $3\ V$. 117

5-21 Measured normalized gate current, I_g/I_d, as a function of device speed, g_m/WC_{ox}, for NMOSFETs with the three SSR and STEP channel dopings described in Fig. 2-1 and Table. 2.1 in Chapter 2. 119

List of Tables

2.1 Ion-implant parameters for Super-Steep Retrograde (SSR) and uniform (STEP) channel doping profiles shown in Fig. 2-1. The threshold voltage, V_t, is the measured value. 22

5.1 $\gamma_b(V_{ds}, V_{gs})$ coefficient matrix according to Fig. 5-19 with index (V_{ds}, V_{gs}), as defined in Eq. 5.34, and channel doping profile, as shown in Fig. 2-1. 115

5.2 $\gamma_g(V_{ds}, V_{gs})$ coefficient matrix according to Fig. 5-20 with index (V_{ds}, V_{gs}), as defined in Eq. 5.35, and channel doping profile, as shown in Fig. 2-1. 118

Chapter 1

Introduction

The dynamics of electron transport in the inversion layers of extremely short-channel MOSFETs is of crucial importance to the understanding of deep-submicron MOSFET operations and their practical implications in today's ULSI (Ultra Large Scale Integrated-circuit) systems. When the energy relaxation time of high-energy electrons is comparable to their transport duration, as may be the case in the inversion layer of an extremely short-channel MOSFET with its effective channel length on the order of the electron energy relaxation length scale, the electrons undergo an essentially non-equilibrium transport process. That is, the energy exchange process between the electrons and the lattice never reaches steady-state equilibrium and the transport process is quasi-ballistic in nature. One of the most prominent phenomena associated with this non-equilibrium nature of dynamic electron transport is the so-called electron velocity overshoot. Namely, the electrons in the inversion layer of an extremely short-channel MOSFET may exhibit an average drift velocity higher than their presumed saturation velocity in bulk silicon. This is of great interest to the ULSI industry, as it gears towards making shorter and shorter channel length MOSFETs. Velocity overshoot was first predicted by Monte Carlo simulations [45], and then was observed experimentally at both room and low temperatures, $T = 77\ K$ and $T = 4.2\ K$ [11, 55]. The understanding of this phenomenon is of both theoretical and practical importance. It is expected that electron velocity overshoot in extremely short-channel MOSFETs provides the extra amount of current drive beyond what

is predicted by the conventional "drift-diffusion" model that does not take velocity overshoot into account.

In Chapter 3, the theoretical background of electron velocity overshoot in the framework of a hydrodynamic model is introduced. Then the experimental results from the past and present are presented and compared. The impact of electron velocity overshoot in extremely short-channel MOSFETs based on the new experimental observations from this thesis work is discussed from the perspective of MOSFET scaling theory which is discussed in more detail in Chapter 4.

As recent as a decade ago, there were arguments that silicon-based MOSFET devices would reach their performance or fabrication limit of 0.25 μm minimum channel length and 500 MHz switching speed imposed by the laws of semiconductor physics. Yet today there is no lack of successful demonstrations of practical and robust 0.1 μm channel length silicon MOSFETs which switch on and off at astonishing times in the 10 ps range [25, 55, 35, 64, 33]. How short silicon MOSFETs can go, how well they behave, how fast they can be switched on and off, and how long they last have been been the questions of ultimate importance and great interest. The answers to those questions are important because they determine the future of silicon-based ULSI (Ultra-Large-Scale-integrated circuit) systems and thus the future direction of the whole semiconductor and computer industry, and the answers are interesting because they reveal the fundamental principles of semiconductor device physics. The fundamental scaling limits of silicon MOSFETs and the improvement of the scaling theories [16, 5, 3] to provide a proper guide for deep-submicron MOSFET design are the two key issues that need to be addressed in order to answer those questions.

In Chapter 4, the scaling relationships among the three fundamental quantities of deep-submicron MOSFETs, device speed, g_m/WC_{ox}, drain-induced barrier lowering (DIBL), $\frac{\delta V_t}{\delta V_{ds}}$, and effective channel length, L_{eff}, are investigated with both device measurements and numerical simulations. The dependence of these relationships on the particular set of channel and source/drain parameters is also investigated experimentally and by numerical simulations in the deep-submicron L_{eff} regime from 0.5 μm down to sub-0.1 μm.

One of the most prominent manifestations of non-equilibrium carrier transport phenomena in the inversion layer of deep-submicron MOSFETs is the hot-carrier effects associated with the non-equilibrium dynamic transport of conduction electrons and holes under ultra-high electric field and electric field gradient. In a silicon MOSFET, hot-carrier phenomena result in the generation of substrate current and gate current through impact ionization and hot-carrier injection into the gate insulator (SiO_2). These two macroscopic quantities (currents) carry the information which reveals the physics of the microscopic hot-carrier scattering processes in the silicon inversion layer. By investigating the characteristic dependence of those currents with respect to an appropriate set of device parameters, specific information about the non-equilibrium transport dynamics of high-energy carriers can be extracted. On the practical side, understanding the physics of hot-carrier effects in silicon MOSFETs plays an important role in determining device degradation mechanisms and thus the improvement of MOSFET design, especially in the deep-submicron regime where hot-carrier-induced device degradation is expected to be significant.

In Chapter 5, the theory of hot-carrier current generation is introduced. A new effect, hot-electron injection barrier-lowering effect, discovered in this thesis work and believed to be associated with the non-equilibrium nature of hot-electron transport in extremely short-channel MOSFETs, is discussed. Another non-equilibrium hot-carrier effect, the hot-electron "cooling" effect, is investigated at both room and liquid nitrogen temperatures. Finally, the role of hot-carrier-induced currents in deep-submicron MOSFET scaling is introduced in the scope of Chapter 4, and a new MOSFET scaling methodology is introduced to unify all four fundamental aspects of MOSFET devices: device performance, device characteristic dimensions, device short-channel effect, and device hot-carrier effect.

In Chapter 6, the main conclusions from the previous chapters are summarized, and the possible future directions in the area of deep-submicron MOSFET physics and technology are discussed.

Chapter 2

Sub-0.1 μm MOSFET: Design, Fabrication, and Characterization

2.1 Sub-0.1 μm n-Channel MOSFET Design

2.1.1 Design objectives

The deep-submicron MOSFET transistors in a modern ULSI system should be designed according to the following three criteria: (1) maximization of device current and/or device speed, (2) minimization of device short-channel effect, and (3) maximization of device punchthrough resistance. The first design criterion ensures the current drive capability and switching speed, and the last two ensure unambiguous on-and-off states at low power supply voltages. To achieve high current drive or high device speed, the MOSFET effective channel length, L_{eff}, must be short enough, as the drain current, I_d, and the device speed, g_m/WC_{ox}, are proportional to ${L_{eff}}^{-\alpha}$, where α is a positive constant (see Chapter 4 for more detail). Also, the source-to-drain parasitic resistance, R_{sd}, must be low enough so that the extrinsic or "usable" transconductance, g_m, can be as close to the intrinsic transconductance, g_{mi}, as possible (see Section 2.3 of this chapter for more detail). To minimize the short-channel effect, which is represented by the drain-induced-barrier-lowering effect (DIBL), $\frac{\delta V_t}{\delta V_{ds}}$, the MOSFET source/drain junction must be shallow and abrupt enough to reduce the

drain influence on the device turn-on/off characteristics at operational drain voltages (see Section 2.3 of this chapter and Chapter 3 and 4 for more detail). To minimize the device punchthrough, the charge concentration in the device sub-surface bulk must be high enough to raise the electrostatic potential, suppress the depletion width expansion around the drain, and confine the conduction carriers in the inverted channel. The following sections address these three objectives and the specific device design techniques to accomplish them.

2.1.2 Channel design

Super-steep-retrograde channel doping (SSR) is used to suppress the device punchthrough [2, 52]. The designed SSR channel doping profiles are shown in Fig. 2-1, as obtained from SUPREM-III device simulator [22]. The SSR channel doping is carried out first by a shallow indium ion implant to adjust the threshold voltage and raise the sub-surface-bulk charge concentration, and then by a deep boron ion implant to raise the charge concentration of the sub-surface as well as deep bulk regions. As shown in Fig. 2-1, the three SSR doping profiles, SSR-I, II, and III, have different surface, peak, and bulk dopant concentrations for the purpose of comparing the device performance and short-channel effect. In addition, a STEP doping profile is also designed for the comparison to the SSR profiles. Ion implant parameters and the corresponding long-channel threshold voltages, V_t, for the SSR and STEP channel dopings are listed in Table 2.1.

All three SSR profiles and the STEP profile have a calculated surface impurity concentration lower than or equal to $N_s = 1.0 \times 10^7\ cm^{-3}$. This is to ensure that the surface impurity concentration is low enough so that the impurity scattering does not become the dominant scattering mechanism and the carrier mobility is not degraded due to high vertical electric field under certain bias conditions [28] (see Chapter 4 for more detail).

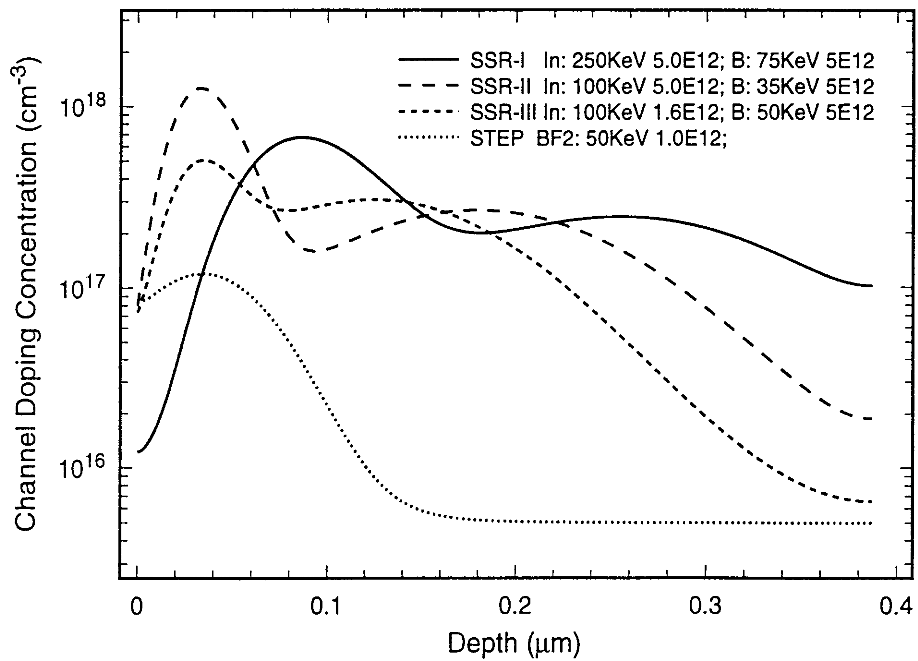

Figure 2-1: Super-Steep-Retrograde (SSR) and Step (STEP) channel doping profiles with ion-implant conditions summarized in Table 2.1.

PROFILE	Implant Parameters	V_t
SSR-I	Indium: $5.0 \times 10^{12}, 250\ KeV$; Boron: $5.0 \times 10^{12}, 75\ KeV$	$0.37\ V$
SSR-II	Indium: $5.0 \times 10^{12}, 100\ KeV$; Boron: $5.0 \times 10^{12}, 35\ KeV$	$0.43\ V$
SSR-III	Indium: $1.6 \times 10^{12}, 100\ KeV$; Boron: $5.0 \times 10^{12}, 50\ KeV$	$0.47\ V$
STEP	BF_2: $1.0 \times 10^{12}, 50\ KeV$;	$0.21\ V$

Table 2.1: Ion-implant parameters for Super-Steep Retrograde (SSR) and uniform (STEP) channel doping profiles shown in Fig. 2-1. The threshold voltage, V_t, is the measured value.

2.1.3 Source/drain design

A shallow and moderately long source/drain extension structure is used to improve the short-channel behavior of the SSR MOSFETs. Two crucial technological elements, indium pre-amorphization and "halo" substrate doping [53], and Ti/Co bimetallic $CoSi_2$ self-align silicide [59], are implemented to form the extension source/drain structure. Implanting indium before the arsenic extension implant amorphizes the crystal structure of the source/drain active area and thus significantly reduces the implant "channeling" effect of the subsequent ion implant so that the extension junction depth can be controlled within the sub-50 nm range, which is necessary for sub-0.1 μm scale MOSFETs according to the classical scaling rules [16, 5, 38]. It also provides a "halo" doping around the extension junction, which forms a p-n$^+$ type of extension rather than a conventional p$^-$-n$^+$ type. Fig. 2-2 shows vertical doping profiles of the indium and arsenic extension implants. As can be seen, the indium dopant profile has a longer tail than the arsenic profile at an impurity concentration on the order of 10^{19} cm^{-3} or below. This "halo" doping can greatly improve the device short-channel behavior for it makes the extension junction more abrupt and the depletion width less susceptible to the drain influence.

The final device cross-section is schematically shown in Fig. 2-3, where all the important technological elements mentioned above are indicated.

The designed device structures are simulated with MINIMOS-4 device simulator [51] to examine if the designed device indeed shows satisfactory short-channel behavior, i.e., drain-induced barrier lowering and device punchthrough, and device performance, i.e., transconductance. Fig. 2-4 shows the simulated subthreshold characteristics of a $L_{eff} = 0.1$ μm n-channel MOSFET with a SSR-III channel doping profile and a gate oxide thickness of $t_{ox} = 5.0$ nm. As it clearly shows, there is no sign of device punchthrough (see Chapter 3 and 4 for more detail), and the drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, measured by the amount of parallel shift in the $\log(I_d)$ $vs.$ V_{gs} curves from $V_{ds} = 0.05$ V to 1.5 V, is less than 80 mV/V.

Fig. 2-5 shows the simulated saturation transconductance, g_m, of n-channel MOSFETs with the same SSR-III channel doping profile, the same $t_{ox} = 50$ nm, and vari-

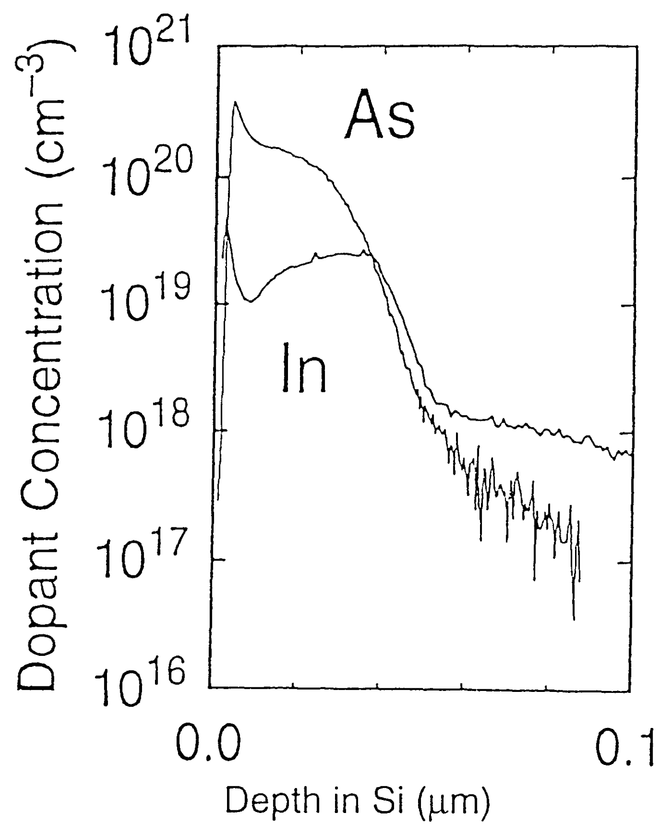

Figure 2-2: Vertical source/drain extension junction profiles obtained by SIMS (after G. G. Shahidi *et al.* [55]).

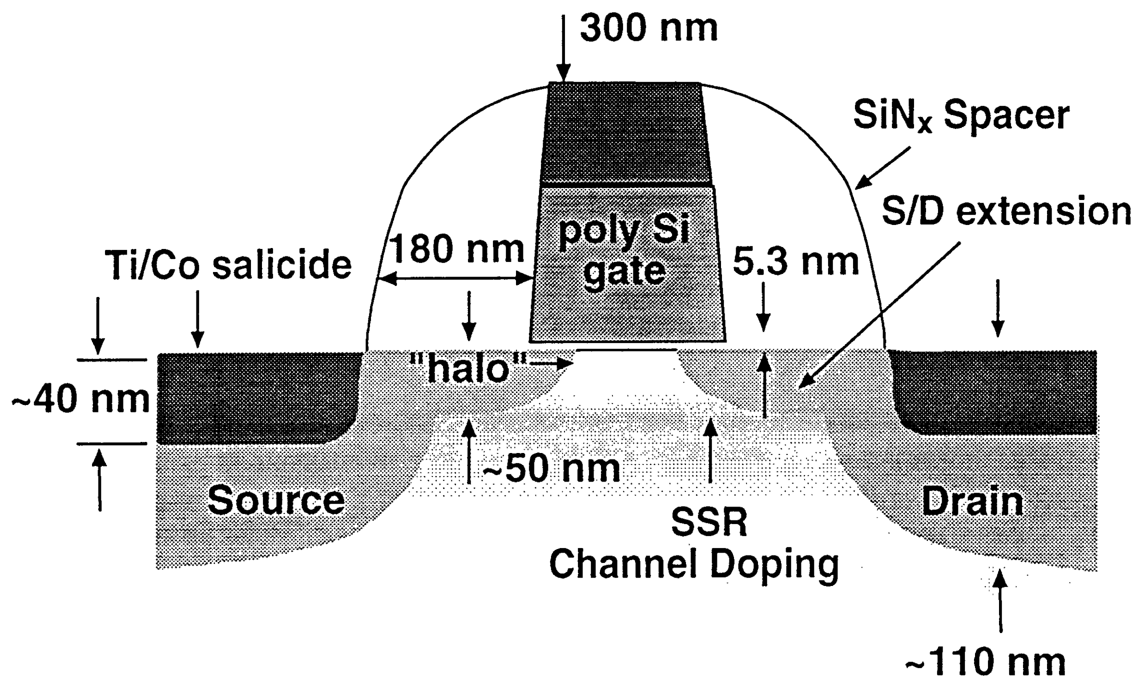

Figure 2-3: The cross-sectional schematic of a sub-0.1 μm SSR n-channel MOSFET device (not to scale). The "halo" doping is an approximately 10 nm wide p-type region (indium implant) around the n^+-type arsenic source/drain extension.

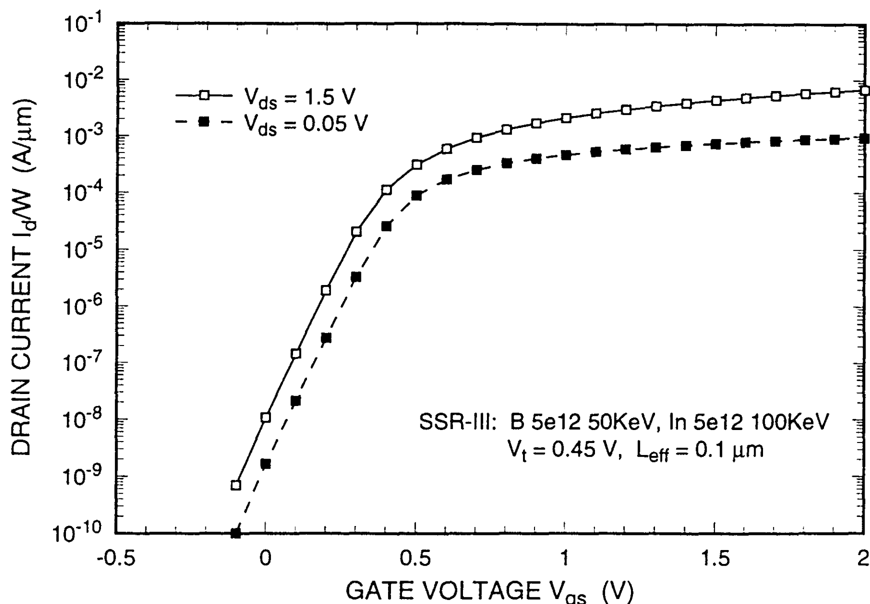

Figure 2-4: The simulated subthreshold characteristics of a $L_{eff} = 0.1\ \mu m$ n-channel MOSFET with $t_{ox} = 5.0\ nm$ and a SSR-III channel doping profile described in Table 2.1 and Fig. 2-1.

ous effective channel lengths from $L_{eff} = 0.1\ \mu m$ to $0.5\ \mu m$. The MINIMOS velocity-saturation model is used here. As clearly demonstrated, the saturation transconductance of the $L_{eff} = 0.1\ \mu m$ device reaches nearly $500\ mS/mm$ at $V_{ds} = 2.0\ V$, which corresponds to an average electron velocity of $7.2 \times 10^6\ cm/s$ with $t_{ox} = 5.0\ nm$.

The excellent $\frac{\delta V_t}{\delta V_{ds}}$ and g_m values obtained from the device simulations demonstrate that the specific device design techniques used here are effective in achieving the three design objectives outlined earlier.

2.2 Sub-0.1 μm n-Channel MOSFET Device Fabrication

Unique fabrication processes are needed to implement the technological elements mentioned in the previous sections in order to realize the design objectives for sub-0.1 μm MOSFET performance and short-channel effects. In order to achieve high

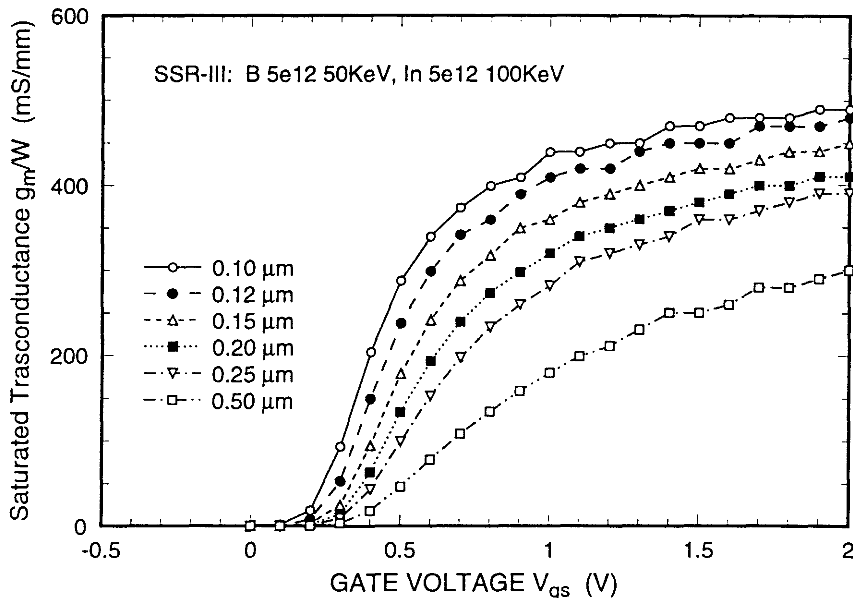

Figure 2-5: The simulated saturation transconductance, g_m, of n-channel MOSFETs with t_{ox} = 5.0 nm, L_{eff} = 0.1 μm ~ 0.5 μm, and SSR-III channel doping profiles described in Table 2.1 and Fig. 2-1. The saturation drain voltage V_{ds} = 2.0 V.

current drive and high transconductance, the MOSFET polysilicon gate has to be well defined in the sub-0.1 μm range, i.e., having smooth edges, minimal line-width variations, robust step coverages, and vertical side walls. The polysilicon gate has to have a smooth edge, since otherwise it is equivalent to many "sub-gates" with varying effective channel lengths in parallel, which is in turn equivalent to many "sub-devices" with different threshold voltages in parallel. X-ray lithography is used for device polysilicon gate definition throughout the device fabrication process in this thesis work for its high process latitude, high throughput (comparing to electron-beam lithography), and robust capability of defining resist patterns with minimum line-widths down to sub-0.1 μm range. In order to fabricated shallow and abrupt source/drain junctions, indium pre-amorphization followed by low-energy arsenic ion implant is used to reduce the junction depth. In order to minimize the device parasitic resistance, an elaborate Ti/Co bimetallic self-aligned $CoSi_2$ silicide is used for its low sheet resistance and thermal stability. The following sections address the detailed fabrication processes for the polysilicon gate definition and the source/drain

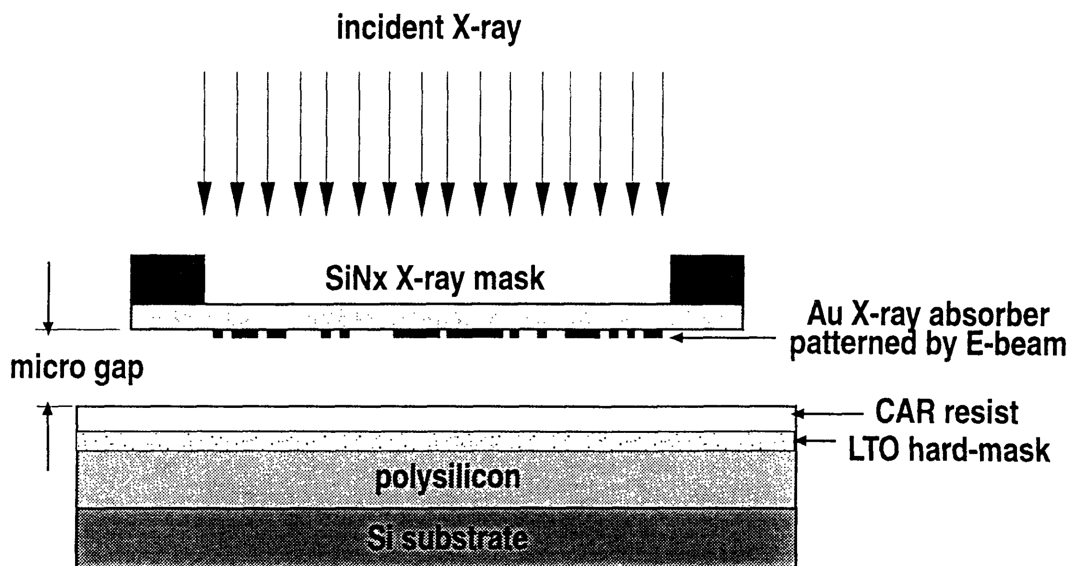

Figure 2-6: The schematic of the proximity X-ray lithography.

junction formation.

2.2.1 X-ray lithography

Fig. 2-6 shows the schematic of proximity X-ray lithography. The LOCOS-patterned silicon substrate is deposited with undoped polysilicon (typically 300 nm thick), low-temperature oxide (typically 50 nm thick), which acts as a hard mask for the pattern transfer from the resist layer to the polysilicon gate layer, and chemically-amplified resist (approximately 400 nm thick SAL-601 negative resist). The X-ray mask consists of a SiN_x membrane (typically about 1 μm thick) with Au X-ray absorbers (typically 200 nm thick) patterned by electron-beam lithography and electro-plated (see [37, 12] for more detail on X-ray mask fabrication). The minimum feature size on the SSR MOSFET X-ray masks is about 80 nm.

On the system level, the X-ray lithography process is carried out in a custom-built X-ray alignment/exposure system shown in Fig. 2-7. The substrate is placed onto an alignment stage under the X-ray mask with a micro-gap in-between (typically 3 $\mu m \sim 5$ μm). The micro-gap magnitude and uniformity are controlled by observing

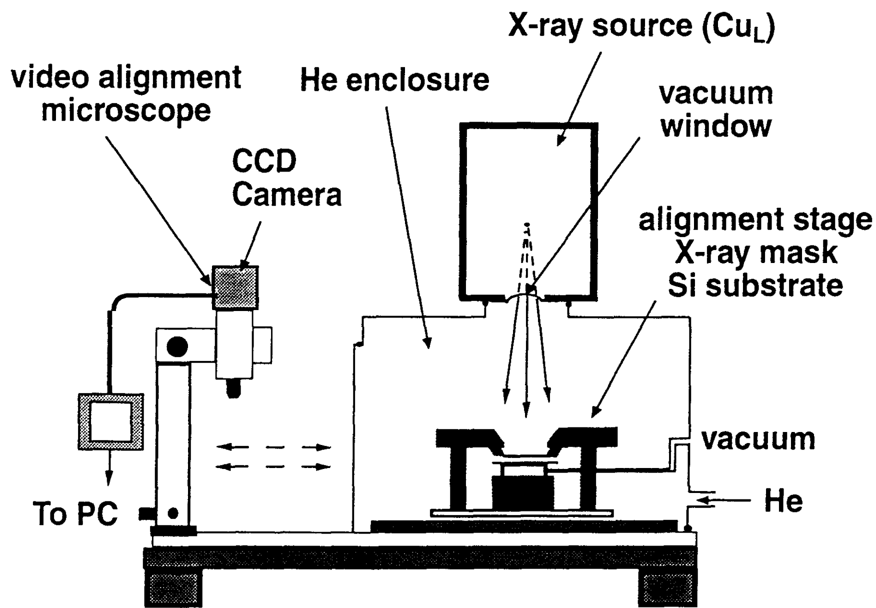

Figure 2-7: The custom-built X-ray lithography alignment and exposure system [37].

the interference patterns under monochromatic illumination (typically green light illumination in this process). The alignment stage is controlled by X-Y-Θ piezos for the fine alignment and micrometers for the coarse alignment. The X-ray source is a Cu_L point source with a wave length of 1.32 nm. The alignment is conducted outside the exposure chamber using a CCD camera and a microscope. The exposure is carried out in a helium environment at atmosphere pressure, which has an oxygen concentration of less than 250 ppm for minimizing the X-ray attenuation.

The final polysilicon-gate formation step involves a two-step dry-etching and pattern-transferring process. After the resist is patterned by X-ray lithography, the substrate is etched in a CHF_3 RIE environment to transfer the resist pattern onto the LTO hard mask. Then the polysilicon-gate layer is etched in a Cl_2 plasma environment to transfer the LTO pattern. These two etching processes are chosen for their high selectivity and capability of reproducing vertical side walls during pattern transfers. Fig. 2-8 shows a resist/LTO-mask gate structure defined by X-ray lithography and CHF_3 etching. Fig. 2-9 shows a final polysilicon gate structure defined

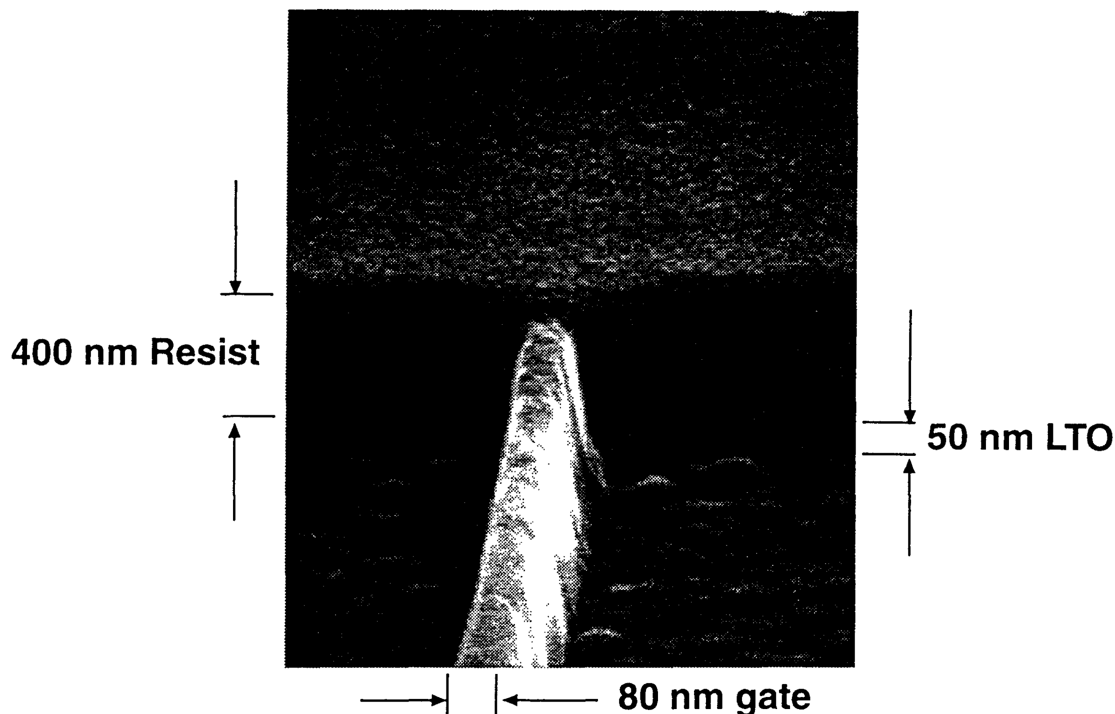

Figure 2-8: SEM micrograph of a resist(SAL-601)/LTO-mask gate structure defined by X-ray lithography and CHF_3 etching. The SAL-601 resist is about 400 nm thick, and the LTO hard-mask is about 50 nm thick.

by Cl_2 etching. The excellent step coverage and the smooth edges of the 0.1 μm-scale polysilicon gate pattern clearly demonstrate the high process latitude and the robustness of the X-ray lithography technology used in this thesis work.

2.2.2 Channel doping

After the LOCOS formation, a layer of dry oxide is thermally grown on the active area, which has the same thickness as the final gate oxide, $t_{ox} = 5.3\ nm$. This oxide layer acts as the "screening" oxide for the channel doping implant, during which the indium is implanted first and the boron is implanted afterwards. Then a 5.3 nm thick oxide layer is grown at 850^oC in a dry O_2 environment to form the gate dielectric, and a 300 nm thick undoped polysilicon is deposited via a CVD process. The channel dopants are activated by the thermal process later during the source/drain formation

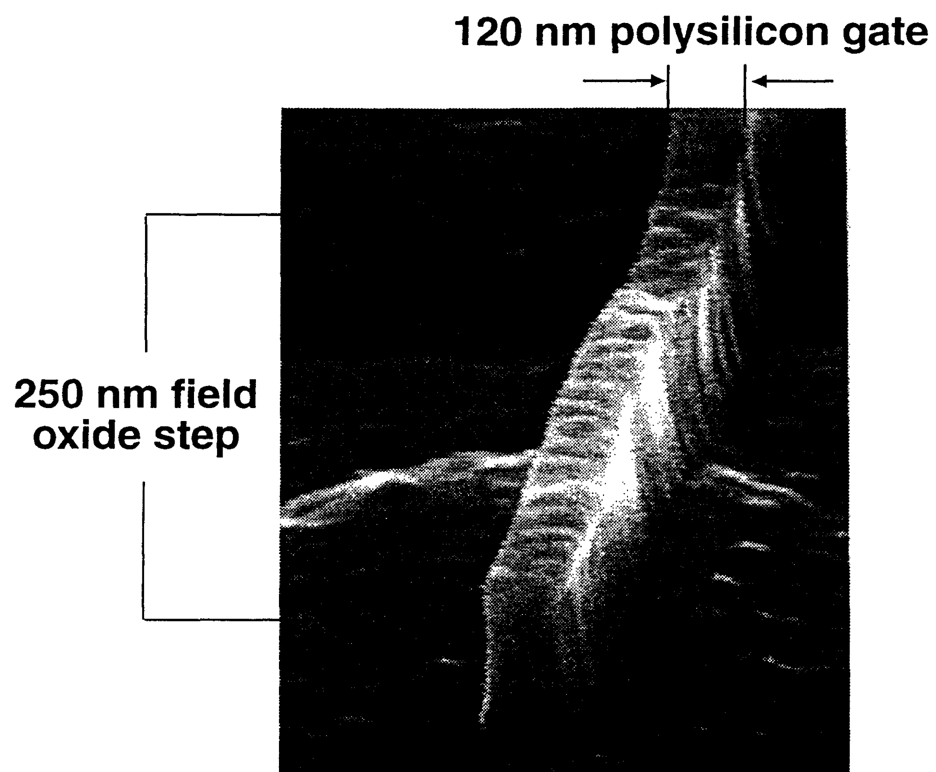

Figure 2-9: SEM micrograph of a polysilicon gate structure defined by LTO hard-mask and Cl_2 etching. The polysilicon gate is about 300 nm thick.

described in next section.

2.2.3 Source/drain junction formation

As shown in Fig. 2-3, the indium-arsenic extension source/drain structure is formed according to the following fabrication sequence. A thin layer of LTO (typically $10 \sim 12\ nm$) is deposited as a "screening" (i.e., randomizing) oxide layer for the subsequent ion implant. A moderate indium implant, at a dose of $1.0 \times 10^{12}\ cm^{-2}$ and energy of 40 KeV, is carried out to pre-amorphize the source/drain junction area. Then a low-energy arsenic implant is carried out at a dose of $9.0 \times 10^{14}\ cm^{-2}$ and energy of 10 KeV to form the shallow extension. After this two-step extension ion implant, a layer of 50 nm thick LTO and a layer of 200 nm thick silicon nitride are consecutively deposited, and subsequently etched in a low-power CF_4 plasma environment to form an approximately 180 nm thick spacer around the polysilicon-gate side wall. Then the second arsenic ion implant, at a dose of $5.0 \times 10^{15}\ cm^{-2}$ and energy of 20 KeV, is carried out to form the deep source/drain junctions. The added junction depth is an "insurance" against aluminum-contact spiking. The polysilicon gate and source/drain dopant activation is done strictly with a rapid thermal annealing (RTA) process at a temperature of at least 1000^oC for 20 s or longer. A novel self-aligned $CoSi_2$ technology using titanium($\sim 3.5\ nm$ thick)/cobalt($\sim 14\ nm$ thick) bimetallic laminate is then formed on the polysilicon gate and the deep source/drain junction area [59], which has a final sheet resistance of $8 \sim 9\ \Omega/\square$ on the n^+ source/drain and $10 \sim 15\ \Omega/\square$ on the n^+ polysilicon gate. Finally, aluminum contacts are formed on LTO passivation layers.

2.3 Sub-0.1 μm n-Channel MOSFET Characterization

A generalized form of the saturation drain current, I_d, can be written as [61],

$$I_d = \frac{W}{L_{eff}} C_{ox}\mu_{eff} \frac{(V_{gs} - V_t)^2}{2(1+\delta)} \frac{1}{1 + \frac{V_{dsat}}{L_{eff}\mathcal{E}_c}} \frac{1}{1 - \frac{\Delta L}{L_{eff}}} \tag{2.1}$$

where W and L are the device channel width and length, respectively, C_{ox} is the gate capacitance, μ_{eff} is the effective carrier mobility, V_{gs} is the gate voltage, V_t is the threshold voltage depending on the drain voltage V_{ds}, δ is a substrate depletion correction factor due to non-uniform depletion source/drain under finite drain voltage, V_{ds}, V_{dsat} is the drain voltage at which I_d saturates (i.e., becomes independent of drain voltage, V_{ds}), $\mathcal{E}_c$ is a "critical" electric field at which the carrier velocity saturates with increasing electric field, and ΔL the effective channel length modulation due to the finite pinch-off length in the saturation region. Experimental data show that $\mathcal{E}_c \approx 4.0 \times 10^4\ V/cm$ for electrons and $\mathcal{E}_c \approx 6.0 \times 10^4\ V/cm$ for holes [57]. δ, μ_{eff}, V_{dsat}, and ΔL are all functions of $V_{gs} - V_t$. This equation takes into account the effects of channel length modulation, velocity saturation, and mobility degradation.

The extrinsic transconductance, g_m, is obtained from the device $I - V$ characteristics,

$$g_m = \frac{\partial I_d}{\partial V_{gs}}\Big|_{V_{ds}}. \tag{2.2}$$

The gate oxide thickness, t_{ox}, is extracted from the gate-oxide capacitance, $C_{ox} = \epsilon_{Si}\epsilon_0/t_{ox}$, measured on a large-area MOS transistor in the inversion mode.

The effective channel length, L_{eff}, and the source-to-drain parasitic resistance, R_{sd}, are extracted simultaneously using the method described in [66, 9]. The effective channel width, W_{eff}, is extracted with the linear transconductance (g_{ml}) ratio method after effective channel length is known, according to

$$W_i - \Delta W = (W_{eff})_i = (g_{ml}L_{eff})_i = constant \qquad i = 1, 2, ..., n \tag{2.3}$$

where the index $i = 1, 2, ..., n$ represents individual devices with different L_{eff} or g_{ml}.

The short-channel effect is represented by the amount of drain-induced barrier lowering (DIBL), $\frac{\delta V_t}{\delta V_{ds}}$, defined by

$$\frac{\delta V_t}{\delta V_{ds}} = \frac{V_t(V_{ds0}) - V_t(V_{ds})}{V_{ds} - V_{ds0}} \tag{2.4}$$

which has units of mV/V with V_{ds} typically taken as $1.0 \sim 2.0\ V$ and V_{ds0} taken as $0.05\ V$. In practice, $\frac{\delta V_t}{\delta V_{ds}}$ is measured as the amount of parallel shift in the $\log I_d$ *vs.* V_{gs} curves from V_{ds0} to V_{ds} at a drain current of $I_d/W = 1\ \mu A/\mu m$.

The SSR n-channel MOSFETs are characterized with the methods mentioned above. Fig. 2-10 and 2-11 show the drain current per channel width, I_d/W_{eff}, *vs.* the drain voltage, V_{ds}, and the gate voltage, V_{gs}, characteristics, respectively, of a $L_{eff} = 0.085\ \mu m$ NMOSFET with a SSR-III channel doping profile.

As can be seen from the I_d/W_{eff} *vs.* V_{ds} characteristics, the device exhibits a saturation current of $0.74\ mA/\mu m$ at $V_{ds} = 2.0\ V$ and $V_{gs} = 2.0\ V$, while maintaining excellent subthreshold characteristics and short-channel behavior with a threshold voltage of $0.36\ V$, a subthreshold slope of $88.1\ mV/decade$, a DIBL value of about $100\ mV/V$, and little device punchthrough. The excellent current drive is undoubtedly attributable to the exceptionally low source-to-drain parasitic resistance, $R_{sd} = 200 \sim 230\ \Omega - \mu m$, and extremely short channel length, $L_{eff} = 0.085\ \mu m$. The excellent subthreshold and short-channel behavior are undoubtedly attributable to the well designed SSR channel doping profile and the well controlled source/drain junction structure (in particular, the indium "halo" substrate doping and the shallow source/drain extension). More SSR MOSFET characterizations on the device performance and the short-channel effects are demonstrated in more detail in Chapter 4.

2.4 Conclusion

High performance sub-0.1 μm MOSFET devices using X-ray lithography, self-aligned $CoSi_2$ silicide formed by Ti/Co laminates, super-steep retrograde channel doping,

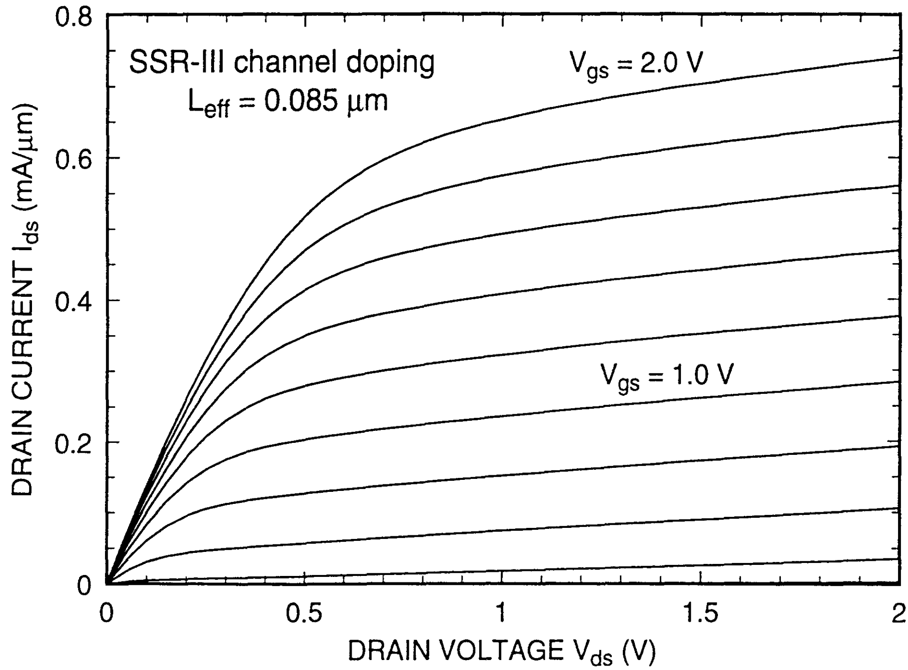

Figure 2-10: The drain current, I_d, *vs.* drain voltage, V_{ds}, characteristics of a $L_{eff} =$ 0.085 μm n-channel MOSFET with a SSR-III channel doping profile. V_{gs} steps of 0.2 V.

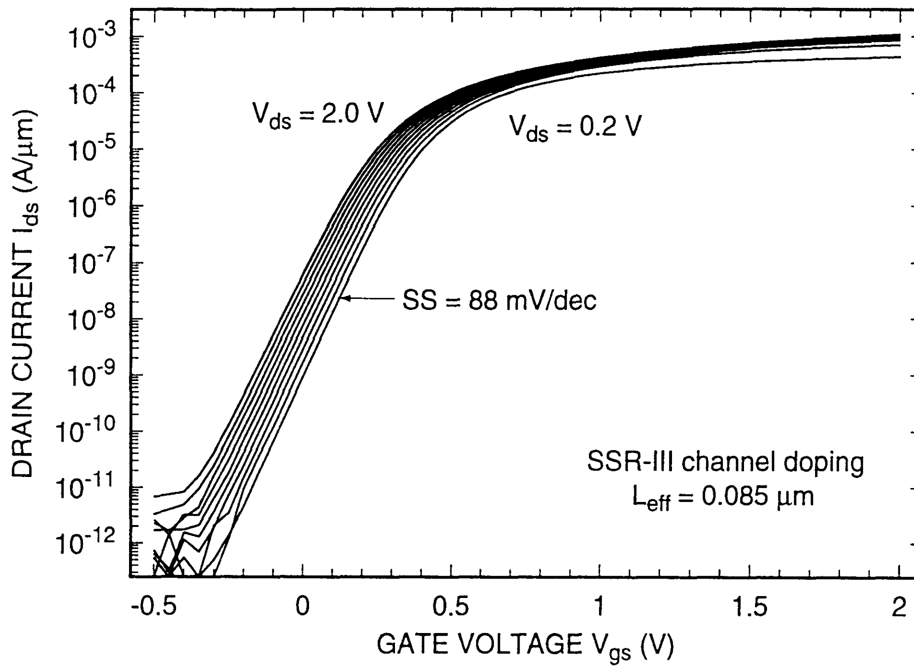

Figure 2-11: The drain current, I_d, *vs.* gate voltage, V_{gs}, characteristics of a $L_{eff} =$ 0.085 μm n-channel MOSFET with a SSR-III channel doping profile. V_{ds} steps of 0.2 V.

and ultra-shallow source/drain extension structure with "halo" doping are demonstrated. These SSR n-channel MOSFETs exhibit very high saturation current drive and transconductance with minimal short-channel effects. X-ray lithography is proven to be a highly promising lithography technology for deep-submicron MOSFET fabrication. The ultra-shallow source/drain extension structure coupled with Ti/Co bimetallic $CoSi_2$ silicide used in this thesis work is demonstrated to be highly effective in controlling short-channel effects and minimizing parasitic resistance. Super-steep retrograde channel doping is shown to be highly effective in preventing device punchthrough, while maintaining the device electrostatic integrity. The SSR n-channel MOSFETs demonstrated in this chapter exhibit excellent overall behavior, i.e, high performance, well-controlled short-channel effects, and minimal leakage currents. This is essential for unambiguous device measurements that provide accurate experimental data for the investigation of deep-submicron MOSFET physics.

Chapter 3

Physics of Velocity Overshoot in Perspective of MOSFET Scaling

In this chapter, the theoretical background of electron velocity overshoot in the framework of a hydrodynamic model is introduced. Then the experimental results from the past and present are presented and compared. Finally, the impact of electron velocity overshoot in extremely short-channel MOSFETs based on the new experimental observations from this thesis work is discussed from the perspective of MOSFET scaling theory which is discussed in more detail in Chapter 4.

3.1 Electron Velocity Overshoot in Si MOSFETs

3.1.1 Hydrodynamic transport model

Hydrodynamic theory has been widely used to model the electron velocity overshoot in deep-submicron MOSFETs for its theoretical simplicity and computational efficiency. The model is derived from the Boltzmann equation, which can be formally written as:

$$\frac{df(\mathbf{r}(t),\mathbf{k}(t))}{dt} = \frac{\partial f(\mathbf{r}(t),\mathbf{k}(t))}{\partial t} + (\frac{d\mathbf{k}}{dt}\cdot\nabla_{\mathbf{k}} + \frac{d\mathbf{r}}{dt}\cdot\nabla_{\mathbf{r}})f(\mathbf{r}(t),\mathbf{k}(t))$$

$$= \frac{\partial f(\mathbf{r}(t), \mathbf{k}(t))}{\partial t}\Big|_{col} \tag{3.1}$$

where $f(\mathbf{r}(t), \mathbf{k}(t))d^3\mathbf{r}d^3\mathbf{k}$ is interpreted as the number of electrons in the phase space element $d^3\mathbf{r}d^3\mathbf{k}$, and the equations of motion, within the semi-classical framework, are given by

$$\frac{d\mathbf{k}}{dt} = -\frac{e}{\hbar}(\mathcal{E} + \mathbf{v} \times \mathbf{B}) \tag{3.2}$$

$$\frac{d\mathbf{r}}{dt} = -\frac{1}{\hbar}\nabla_{\mathbf{k}}(E(\mathbf{k})) \tag{3.3}$$

where $E(\mathbf{k})$ is the electron energy.

The collision term can be written in the following general form:

$$\frac{\partial f(\mathbf{r}(t), \mathbf{k}(t))}{\partial t}\Big|_{col} = \sum_{\mathbf{k}'}[P(\mathbf{k}, \mathbf{k}')f(\mathbf{k}')(1 - f(\mathbf{k})) - P(\mathbf{k}', \mathbf{k})f(\mathbf{k})(1 - f(\mathbf{k}'))] \tag{3.4}$$

where $P(\mathbf{k}, \mathbf{k}')$ is the transitional probability for an electron to undergo a transition from state $\mathbf{k}$ to $\mathbf{k}'$, and,

$$f(\mathbf{k}) = \int_{\Omega_r} f(\mathbf{r}(t), \mathbf{k}(t))d^3\mathbf{r}. \tag{3.5}$$

Due to the extreme difficulty of obtaining an analytical closed-form expression for the collision term, $\frac{\partial f(\mathbf{r}(t),\mathbf{k}(t))}{\partial t}\Big|_{col}$, one of the common practices is to make the following "relaxation time" approximation:

$$\frac{\partial f(\mathbf{r}(t), \mathbf{k}(t))}{\partial t}\Big|_{col} = -\frac{f - f_0}{\tau} = -\frac{f_1}{\tau} \tag{3.6}$$

where f_1 is the "one-particle" joint electron density distribution taking into account only the one-particle interaction, and f_0 is the equilibrium electron density distribution which takes the Fermi-Dirac form,

$$f_0(E_{\mathbf{k}}) = -\frac{1}{1 + e^{(E_k - E_f)/k_B T_e}} \tag{3.7}$$

where T_e is the electron temperature, and E_f is the electron Fermi energy. τ is the effective relaxation time which measures the rate at which the electrons approach the

equilibrium state in an ensemble average sense after the external electromagnetic or thermal perturbation is turned off. τ is the Mathiessen's average of all the distinctive relaxation times associated with all the possible electron-phonon, electron-ionized impurity, electron-surface roughness, and electron-electron scattering processes. For the typical momentum relaxation, energy relaxation, or thermal transfer processes in semiconductors, the relaxation time τ is in general a function of electron crystal momentum, $\mathbf{k}$, and electron energy, which is usually of a complicated form and requires extensive resources to compute, and often its physical meanings are not intuitively instructive.

The simplest form of the Boltzmann transport theory, the "drift-diffusion" model, is obtained by taking the zeroth and the first moments of the Boltzmann equation, known as the continuity equation and the momentum balance equation, respectively. This model can be summarized in the following equation:

$$\frac{d\mathbf{r}}{dt} = \mu_e(\mathcal{E})\mathcal{E} + D_e(\mathcal{E})\frac{\nabla_{\mathbf{r}} n(\mathbf{r})}{n(\mathbf{r})} \tag{3.8}$$

where $n(\mathbf{r})$ is the number density in position space, $\mu_e(\mathcal{E})$ the electron mobility, and $D_e(\mathcal{E})$ the electron diffusion coefficient. The "drift-diffusion" model assumes that all the characteristic transport quantities are only functions of the local electric field, $\mathcal{E}$, and it fails to capture the higher order effects due to high electric-field gradients because it ignores the higher moments of the Boltzmann equation.

The hydrodynamic equations are obtained by taking up to the second moment of the Boltzmann equation with the assumption of a displaced Maxwellian distribution to describe the electron energy distribution:

$$f(\mathbf{r}(t), \mathbf{k}(t)) = \frac{\hbar^3}{(2\pi m^* k_B T_e)^{3/2}} \mathrm{e}^{-\frac{m^*(\mathbf{V}-\mathbf{V}_g)}{2k_B T_e}} \tag{3.9}$$

where $\mathbf{v}_g$ is the electron group velocity and T_e is the electron temperature. The dynamics of electron transport in the inversion layer of a silicon MOSFET can be described by the hydrodynamic equations coupled with the Poisson equation which relates the electrostatic variables. The collision terms of the first and second moment

equations are expressed in the relaxation time approximation with the characteristic time scales, τ_p and τ_e, respectively. The full set of the hydrodynamic equations can be expressed as follows:

Particle conservation equation:

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{v}) = G \tag{3.10}$$

Momentum conservation equation:

$$\frac{\partial m^*\mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla(m^*\mathbf{v}) = -q\mathcal{E} - \frac{k_B}{n}\nabla(nT_e) - \frac{m^*\mathbf{v}}{\tau_p} \tag{3.11}$$

Energy conservation equation:

$$\frac{\partial \epsilon}{\partial t} + \mathbf{v} \cdot \nabla(\epsilon) = -q\mathcal{E} \cdot \mathbf{v} - \frac{k_B}{n}\nabla \cdot (nT_e\mathbf{v}) - \frac{1}{n}\nabla \cdot (\kappa \nabla T_e) - \frac{\epsilon - \epsilon_0}{\tau_e} \tag{3.12}$$

Poisson equation:

$$\nabla^2_{\mathbf{r}}\psi(\mathbf{r}) = \frac{q}{\epsilon_{si}}(n(\mathbf{r}) - N_D^+(\mathbf{r}) + N_A^-(\mathbf{r})) \tag{3.13}$$

where $\psi(\mathbf{r})$ is the electrostatic potential, ϵ_{si} is the silicon dielectric constant, and $n(\mathbf{r})$ is the inversion layer conduction electron density, and $N_D^+(\mathbf{r})$ and $N_A^-(\mathbf{r})$ are the ionized donor and acceptor concentrations, respectively. G is the net carrier generation due to impact ionization and recombination. τ_p and τ_e are the momentum relaxation time and the energy relaxation time, respectively. They are in general complicated functions of electron energy [44, 41, 27]. ϵ is the total electron energy which is assumed to be composed of its random motion part and convective motion part:

$$\epsilon = \frac{3}{2}k_BT_e + \frac{m^*v^2}{2} \tag{3.14}$$

where $v = |\mathbf{v}|$. ϵ_0 is the equilibrium electron energy, $\frac{3}{2}k_BT_0$.

3.1.2 Dynamics of electron velocity overshoot in MOSFETs

Electron velocity overshoot in silicon occurs when electrons exhibit an average drift velocity higher than the saturation velocity under certain electric field distributions and geometric configurations in silicon. Electron velocity saturation is a consequence of electron-optical phonon scattering during which process the electrons gain energy from an external electric field and at the same time lose energy to optical phonons at a scattering rate, τ_{e-op}, that is a monotonically increasing function of electron energy [50]. In the high electric field regime ($\mathcal{E} \sim 10^4\ V/cm$ in silicon), the electrons eventually gain and lose energy at the same rate due to the nature of the optical-phonon scattering process so that their velocity ceases to increase with increasing external electric field. However, as is demonstrated later in this section, the dynamic nature of the hydrodynamic transport allows electron velocity overshoot under certain non-equilibrium conditions. A simplified hydrodynamic model presented below will capture that nature and demonstrate the possibility of electron velocity overshoot.

As schematically shown in Fig. 3-1, a one-dimensional electron velocity overshoot model within the hydrodynamic framework can be presented as follows: the electrons are in equilibrium with the lattice in the source of an MOSFET, $T_e(x=0)=T_0$, and they are constantly emitted at a steady rate from the source into the inverted channel, and accelerated by the lateral electric field, $\mathcal{E}(x)$, across the channel where their transport dynamics is governed by Eqs. 3.10, 3.11, and 3.12, and they are injected into the drain where they return to equilibrium with the lattice, $T_e(x=L_{eff})=T_0$, where L_{eff} is the effective channel length.

By assuming a uniformly distributed electric field, $\mathcal{E}(x)=\mathcal{E}$, the hydrodynamic equations can be rewritten as

$$\frac{d(nv)}{dx} = G \tag{3.15}$$

$$m^* \mathbf{v}\frac{d\mathbf{v}}{dx} = -q\mathcal{E} - \frac{k_B}{n}\frac{(nT_e)}{dx} - \frac{m^*\mathbf{v}}{\tau_p} \tag{3.16}$$

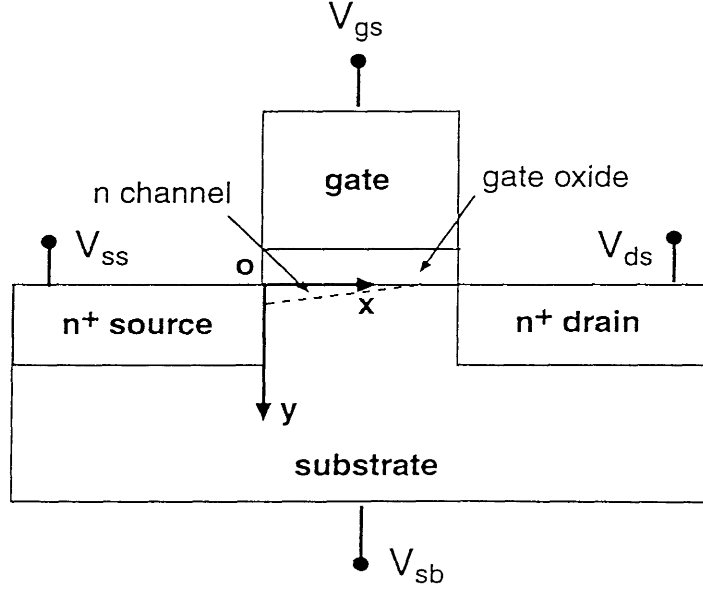

Figure 3-1: A schematic of the coordinate system used for demonstrating electron transport processes in a MOSFET.

$$\mathrm{v}\frac{d\epsilon}{dx} = -q\mathcal{E}\mathrm{v} - \frac{k_B}{n}\frac{d(nT_e\mathrm{v})}{dx} - \frac{1}{n}\frac{d}{dx}(\kappa\frac{dT_e}{dx}) - \frac{\epsilon - \epsilon_0}{\tau_e}. \tag{3.17}$$

The momentum and energy relaxation times in this model can be approximated by their homogeneous steady-state values which can be obtained by solving Eqs. 3.16 and 3.17 under steady state and homogeneous conditions, i.e. neglecting d/dx terms. The solutions are

$$\tau_p(T_e) = \frac{m^* v}{q|\mathcal{E}|} \tag{3.18}$$

and

$$\tau_e(T_e) = \frac{3}{2}k_B\frac{T_e - T_0}{q\mathcal{E}^2}\frac{m^*}{q\tau_p(T_e)} = \frac{3}{2}nk_B(T_e - T_0)\frac{1}{\sigma_0\mathcal{E}^2} \tag{3.19}$$

where $\sigma_0 = nq^2\tau_p/m^*$ is the conductivity due to electrons. It is assumed in obtaining Eq. 3.19 that the electron convective energy is much smaller than its random energy, $|\frac{1}{2}m^*v^2| \ll |\frac{3}{2}k_BT_e|$ [45, 41], so that the total electron energy is only a function of

electron temperature, $\epsilon \approx \frac{3}{2}k_B T_e$.

The relaxation times in Eqs. 3.18 and 3.19 can also be expressed in terms of the electron mobility, $\mu(\mathcal{E}) = |\mathbf{v}/\mathcal{E}|$, which is an experimentally observable quantity, rather than the lateral electric field, $\mathcal{E}$. A reasonable empirical relationship between $\mu(\mathcal{E})$ and $\mathcal{E}$ is given in [6],

$$\mu(\mathcal{E}) = \frac{\mu_0}{(1 + (\mu_0 \mathcal{E}/v_{sat})^2)^{1/2}} \tag{3.20}$$

where $\mu_0 = \mu(\mathcal{E} \to 0)$ is the low-field ($T_e = T_0$) electron mobility in bulk silicon, and v_{sat} is the electron saturation velocity in bulk silicon, which is believed to be $v_{sat} = 1.0 \times 10^7\ cm/s$ [19, 15, 14]. The generalized Einstein relation [58] gives

$$\mu(T_e) = (\frac{T_0}{T_e})^\gamma \mu_0 \tag{3.21}$$

where γ is an experimental fitting factor determining how fast $\mu(T_e)$ changes with T_e. Combining Eq. 3.20 and Eq. 3.21 yields

$$\tau_p(T_e) = \frac{m^* \mu_0}{q}(\frac{T_0}{T_e})^{-\gamma} \tag{3.22}$$

and

$$\tau_e(T_e) = \frac{3}{2}\frac{k_B T_0}{q}\frac{\mu_0}{v_{sat}^2}\frac{1}{(1 + (\frac{T_0}{T_e})^\gamma)}. \tag{3.23}$$

Thus in this simplified model, the relaxation times are only functions of electron temperature. The higher the low-field mobility, $\mu_0 = \mu(T_0)$, the higher are the momentum and energy relaxation times, and thus the more pronounced are the dynamical effects of electron transport under inhomogeneous conditions such as the high electric field gradient in an extremely short-channel MOSFET.

To provide a qualitative description of electron velocity overshoot, the hydrodynamic model presented in Eqs. 3.16 and 3.17 can be further simplified by omitting all the derivative terms involving T_e [26] and substituting in the relaxation times given

in Eqs. 3.22 and 3.23. The further simplified model becomes

$$m^* v \frac{dv}{dx} + \frac{qv}{\mu(T_e)} = -q\mathcal{E} \tag{3.24}$$

and

$$\frac{d(T_e - T_0)}{dx} + \frac{T_e - T_0}{\tau_e \mu(T_e)\mathcal{E}} = -\frac{2}{3}\frac{q}{k_B}\mathcal{E} \tag{3.25}$$

where $\mu(T_e)$ is given by Eq. 3.21.

Electron velocity overshoot can be clearly demonstrated as a result of the dynamic nature of Eqs. 3.24 and 3.25, when a "step" electric field, $\mathcal{E}(x = 0^-) = 0$ and $\mathcal{E}(x =^+) > 0$, is introduced. Fig. 3-2 shows the 1-D electron velocity distribution as a function of position, $v(x)$, calculated by solving Eqs. 3.24 and 3.25 in a one-dimensional silicon cathode-drift region-anode structure under a constant electric field of $\mathcal{E} = 50\ KV/cm$ with various values of low-field mobility, μ_0, and Fig. 3-3 shows the corresponding electron temperature distribution as a function of position, $T_e(x)$, all at room temperature, $T_0 = 300\ K$.

Electrons are emitted from the source, $x = 0^-$, into the bulk silicon, and are collected by the drain boundary, $x = 0.2\ \mu m$ in this example. When μ_0 exceeds about $500\ cm^2/V \cdot sec$, for a short distance from the source ($x = 0$), the electrons exhibit velocities higher than their steady-state saturation velocity, $v_{sat} = 1.0 \times 10^7\ cm/s$. The portion of the space where the electrons exhibit velocity overshoot corresponds to the portion where the electron temperature rises but has yet to reach its steady-state equilibrium value. As soon as the electron temperature reaches its steady-state value, the steady-state electron velocity falls back below the saturation velocity, $v_{sat} = 10^7\ cm/s$. Thus, based on this simplified hydrodynamic model, electron velocity overshoot can be viewed as the consequence of the slower increase in electron temperature such that the electrons acquire excessive amount of convective energy from the electric field in a short distance to accelerate over v_{sat}, while the energy balance between the electrons and the silicon lattice has yet to be established to convert electron convective energy into thermal random energy via electron-phonon scattering. This phenomenon is sometimes called a "non-local" effect or "non-stationary" effect.

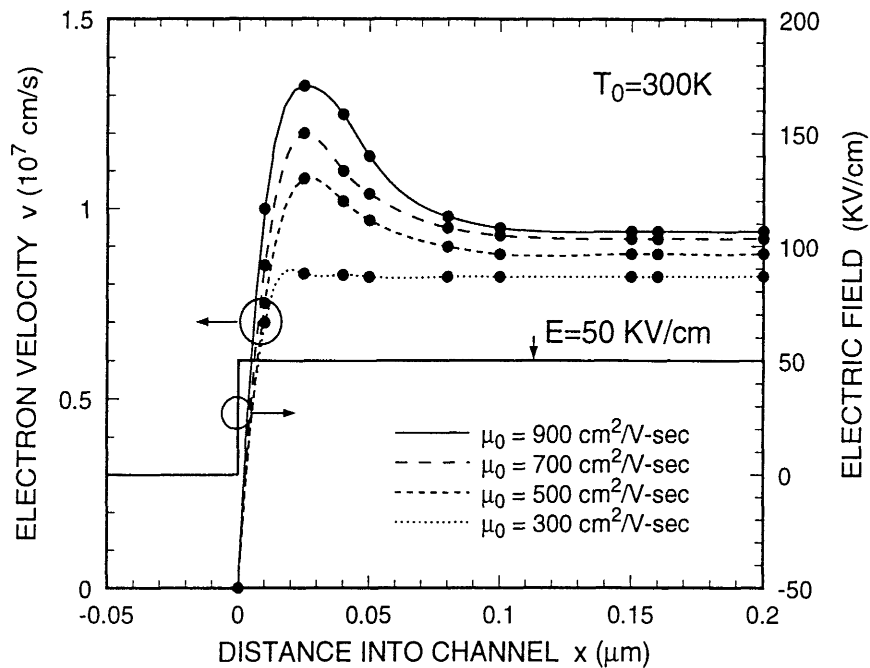

Figure 3-2: Electron velocity calculated by Eqs. 3.24 and 3.25 for various low-field mobilities, μ_0. The right-hand Y-axis shows the applied lateral electric field.

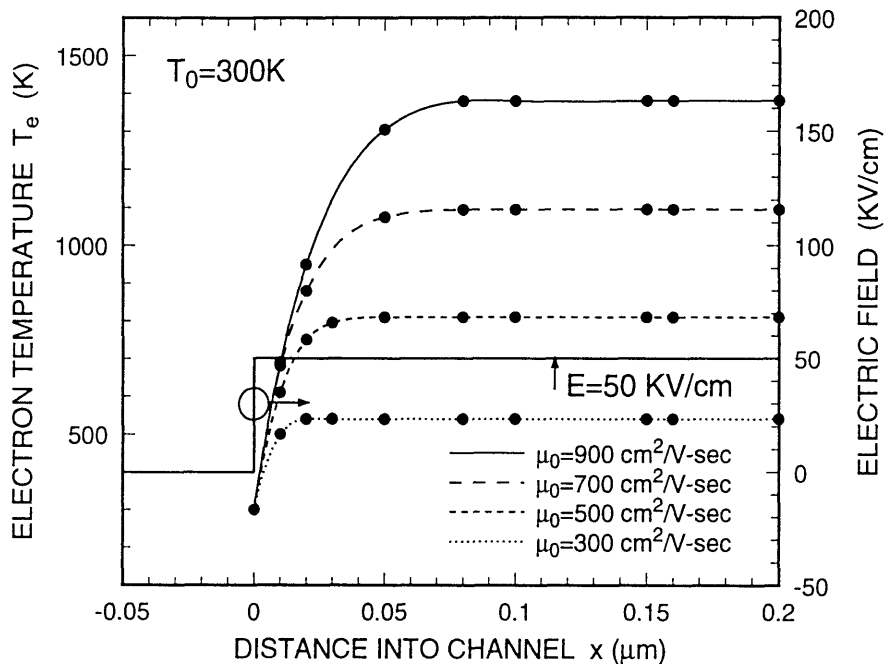

Figure 3-3: Electron temperature calculated by Eqs. 3.24 and 3.25 for various low-field mobilities, μ_0. The right-hand Y-axis shows the applied lateral electric field.

As can be seen in Fig. 3-2, it is the abrupt change in electric field from $x = 0^-$ to $x = 0^+$ that causes electron velocity overshoot near the source boundary, but not the uniform electric field in the channel region. This means that the higher the electric field gradient, the more pronounced is electron velocity overshoot.

Electron velocity overshoot can benefit the current and transconductance improvements in a MOSFET only when it happens over a significant portion of the device channel, and otherwise, the improvement is minimal [4, 30]. In an extremely short-channel MOSFET, improvements are possible because the lateral electric field can be fairly high even at the very beginning of the channel, i.e, the source side, under certain applied voltage configurations, as is shown by Monte Carlo simulations [32, 68, 20]. Fig. 3-4 and Fig. 3-5 show the electron velocity and electric field as a function of the distance from the source, respectively, along the channel Si-SiO_2 interface. For the $L_c = 233\ nm$ device, the portion of the channel where the electron velocity exceeds the saturation velocity, $v_{sat} = 10^7\ cm/s$, is about 20% of the channel, whereas for the $L_c = 43\ nm$ device, the portion is nearly 100% of the channel. This is because the electric field near the source side for the $L_c = 43\ nm$ device increases much more rapidly than that for the $L_c = 233\ nm$ device. The extra gain in the drain current and transconductance over what is predicted by the drift-diffusion theory is significantly greater for the much shorter $L_c = 43\ nm$ device.

3.2 Experimental Observations

3.2.1 Experimental techniques

The average electron velocity in the channel of a MOSFET can be extracted by its macroscopic observables, such as output current, output conductance and capacitance. As mentioned briefly in Chapter 2, the drain current of a MOSFET can be written as

$$I_d = WQ_I(x)v(x) = WQ_{I0}v_0 \tag{3.26}$$

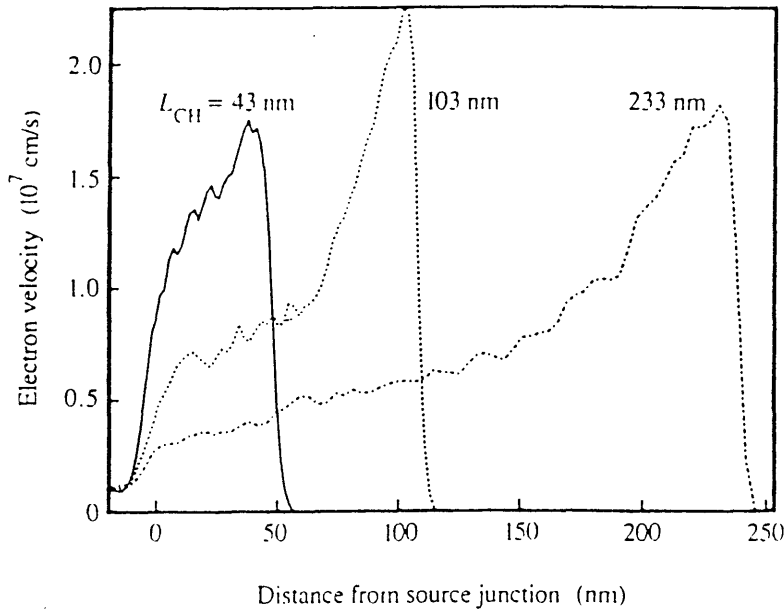

Figure 3-4: Calculated (Monte Carlo simulation) electron velocity for three silicon MOSFETs plotted against the distance from the source along the Si-SiO_2 interface. For the channel length $L_c = 43\ nm$ device, $V_{gs} = 0.7\ V$ and $V_{ds} = 0.6\ V$; for the channel length $L_c = 103\ nm$ device, $V_{gs} = 1.0\ V$ and $V_{ds} = 1.0\ V$; for the channel length $L_c = 233\ nm$ device, $V_{gs} = 2.5\ V$ and $V_{ds} = 2.5\ V$. All sources and substrates are grounded and $T_0 = 300\ K$. The velocity is in the source-to-drain direction, averaged over a depth of 10 nm below the interface.

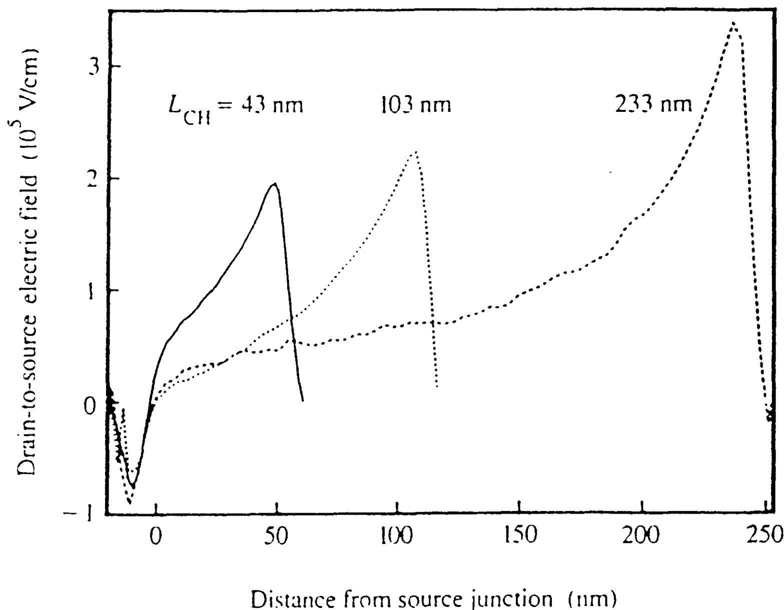

Figure 3-5: Calculated (Monte Carlo simulation) electric field for three silicon MOSFETs in Fig. 3-4 plotted against the distance from the source along the Si-SiO_2 interface. $T_0 = 300\ K$. The electric field is in the source-to-drain direction, averaged over a depth of 10 nm below the interface.

where W is the channel width, $Q_I(x)$ is the inverted channel charge density, $Q_{I0} = Q_I(x=0)$, $v(x)$ is the conduction electron velocity in the channel, and $v_0 = v(x=0)$. The drain current can also be written in terms of the voltage applied at the gate electrode, V_{gs}, as

$$I_d = WC_{gc}(V_{gs} - V_t)v_0 \approx WC_{ox}(V_{gs} - V_t)v_0 \leq WC_{ox}(V_{gs} - V_t)v_{sat} \tag{3.27}$$

where V_t is the threshold voltage for channel inversion, and C_{gs} is the gate-to-channel capacitance per unit area, which is only slightly smaller than C_{ox}, the static gate oxide capacitance per unit area, when $v_0 \approx v_{sat}$ [65, 53]. The above inequality holds if electron velocity overshoot does not occur at the source side ($x = 0$), $v_0 \leq v_{sat}$. However, if velocity overshoot does occur, such that the electron velocity exceeds saturation velocity at the source side, $v_0 \geq v_{sat}$, then the drain current I_d exceeds $WC_{ox}(V_{gs} - V_t)v_{sat}$ predicted by the conventional scaling model Eq. 3.27.

A convenient quantity to infer the electron velocity is the device intrinsic transconductance, g_{mi}, defined as

$$g_{mi} = \left.\frac{\partial I_d}{\partial V_{gs}}\right|_{V_{ds}} = WC_{ox}v_0 \tag{3.28}$$

where V_{ds} is the voltage applied at the drain. Hence,

$$v_0 = \frac{g_{mi}}{WC_{ox}} \tag{3.29}$$

The experimentally observable transconductance, g_m, is related to the intrinsic transconductance by [10]

$$g_{mi} = \frac{g_{m0}}{1 - R_{sd}g_d(1 + \frac{1}{2}R_{sd}g_{m0})} \tag{3.30}$$

where $g_{m0} = g_m/(1-\frac{1}{2}R_{sd}g_m)$, and R_{sd} is the total source-to-drain parasitic resistance, and $g_d = \left.\frac{\partial I_d}{\partial V_{ds}}\right|_{V_{gs}}$ is the drain output conductance. The gate capacitance, C_{ox}, can be approximated by the static capacitance measured on a large area MOS capacitor [53, 65]. The parasitic resistance, R_{sd}, can be extracted by the method described in

Chapter 2. The drain output conductance can be calculated from the device I_d *vs.* V_{ds} characteristics. Then the electron velocity can readily be calculated from Eqs. 3.29 and 3.30.

3.2.2 Experimental results

There were mainly three reports in the past on the experimental evidence of electron velocity overshoot in n-channel MOSFETs at various temperatures. Chou *et al.* [11] reported the first observation of electron velocity overshoot at 4.2 K. Shahidi *et al.* [53] later reported the observation of electron velocity overshoot at 77 K and 300 K. Both works used non self-aligned silicon n-channel MOSFETs which have large gate-to-drain overlaps and thus large parasitic resistance, R_{sd}, which makes the extraction of the intrinsic electron velocity in the channel rather questionable due to significant percentage-wise g_m corrections. In addition, their devices show severe punchthrough in saturation, as is described later in this section, and thus are not practically useful in real ULSI systems. Sai-Halasz *et al.* [48, 47] reported electron velocity overshoot observed in self-aligned n-channel MOSFETs operated at 77 K and 300 K. In the report, the measured g_m value shows that electron velocity overshoot does occur for a $L_{eff} = 0.07\ \mu m$ NMOSFET at 77 K if the electron saturation velocity, v_{sat}, is chosen to be $1.0 \times 10^7\ cm/s$ at 77 K (as opposed to the commonly-accepted $v_{sat} = 1.4 \times 10^7\ cm/s$ at 77 K), but does not occur at 300 K even if v_{sat} is chosen to be $0.8 \times 10^7\ cm/s$ at 300 K, which is 20% lower than the commonly-quoted v_{sat} value of $1.0 \times 10^7\ cm/s$ at 300 K. After correcting for the parasitic resistance of $R_{sd} = 440\ \Omega - \mu m$ at 300 K, which corresponds to about a 22% correction to the intrinsic transconductance at $L_{eff} = 0.07\ \mu m$ ($g_m \sim 580\ \mu S/\mu m$ and the corrected $g_{mi} \sim 710\ \mu S/\mu m$), the $L_{eff} = 0.07\ \mu m$ device shows an average electron velocity of $0.925 \times 10^7\ cm/s$ at 300 K, which still does not convincingly support the claimed room temperature velocity overshoot. Again, the reported NMOSFET also suffers from severe device punchthrough and cannot be turned off in a real ULSI system at room temperature. Since then, there have been numerous reports on various $0.1\ \mu m$ MOSFETs but none has showed any convincing evidence for electron velocity

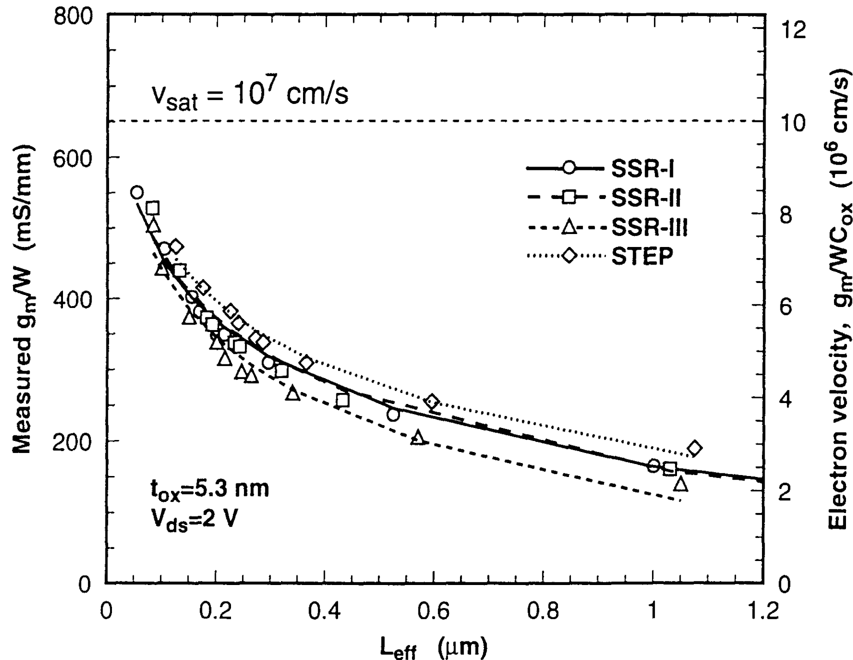

Figure 3-6: Measured saturated transconductance per unit width, g_m/W, *vs.* effective channel length, L_{eff}, for NMOSFETs with the four channel doping profiles described in Table. 2.1. The right-hand Y-axis shows the corresponding electron velocity, g_m/WC_{ox}, and $T = 300\ K$.

overshoot at room temperature in extremely short-channel MOSFETs [55, 64, 33, 67, 35, 42].

In order to investigate the possibility of room temperature velocity overshoot in silicon MOSFETs and its practical impact on improving MOSFET performance that can benefit actual ULSI systems, experiments were conducted on measuring the SSR NMOSFET devices described previously in Chapter 2. Fig. 3-6 shows the measured saturated transconductance, g_m, as a function of effective channel length, L_{eff}, for the SSR-I, SSR-II and SSR-III devices, as well as the STEP devices as described in Table 2.1 and Fig. 2-1. The highest electron velocity observed is about 0.85$\times$ 10^7 cm/s for a $L_{eff} = 0.055\ \mu m$ NMOSFET with SSR-I channel doping (Table. 2.1) at a drain voltage of $V_{ds} = 2.0\ V$, which is not yet in the velocity overshoot regime if v_{sat} is chosen to be 1.0$\times$ 10^7 cm/s.

After correcting for the source-to-drain parasitic resistance, $R_{sd} = 220\ \Omega - \mu m$, the intrinsic saturated transconductance, g_{mi}, is about 6.1% higher than the extrin-

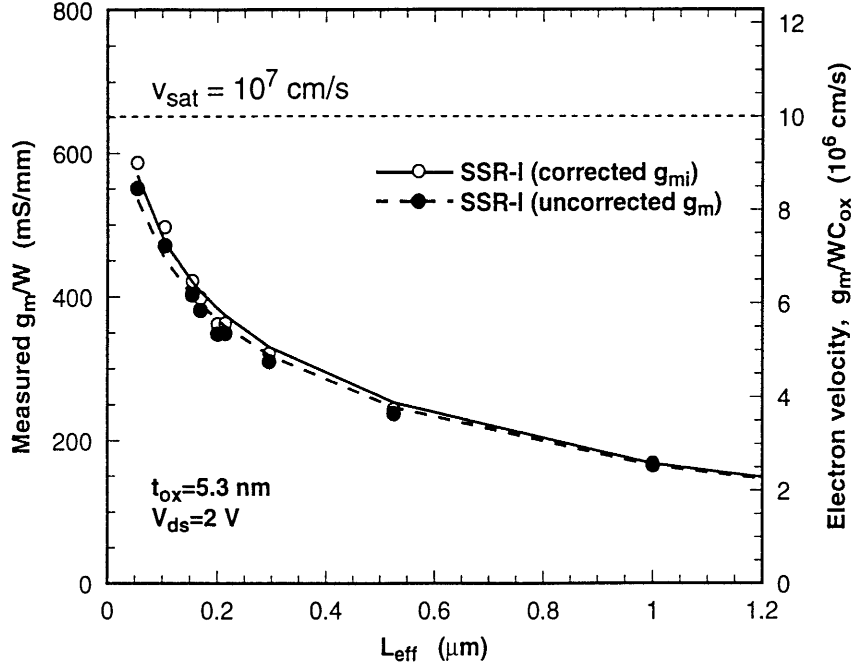

Figure 3-7: Comparison between the uncorrected (extrinsic) and corrected (intrinsic) saturated transconductance per unit width, g_m/W, *vs.* effective channel length, L_{eff}, for NMOSFETs with the SSR-I channel doping profile described in Table. 2.1. The right-hand Y-axis shows the corresponding electron velocity, g_m/WC_{ox}, and $T = 300\ K$.

sic saturated transconductance at $g_m = 550\ mS/mm$, as shown in Fig. 3-7 for the NMOSFETs with SSR-I channel doping. The corrected electron velocity is about $0.91\times\ 10^7\ cm/s$ for the same $L_{eff} = 0.055\ \mu m$ NMOSFET. Even though the upward trend of electron velocity increase with decreasing L_{eff} is towards the velocity overshoot regime, as clearly indicated by the experimental data in Fig. 3-6, the electron saturation velocity of $v_{sat} = 1.0 \times 10^7\ cm/s$ is still not quite reached for these NMOSFETs at room temperature.

One might then ask whether or not the electron velocity in a silicon MOSFET could eventually break the $v_{sat} = 1.0\times\ 10^7\ cm/s$ barrier if L_{eff} keeps decreasing, judging from the fact that there is still no sign of g_m/WC_{ox} saturation with shrinking L_{eff} even in the deep sub-0.1 μm regime. However, the question, though a legitimate one, is not a complete one, because the laws of conservation never allow something to keep improving without compromising something else, and the question fails to

address that. The something else that makes the law of conservation hold is the side-effect of increasing g_m/WC_{ox} due to decreasing L_{eff}, that is, the "short-channel effects". These effects are unwanted electrical characteristics that make a MOSFET deviate from its ideal behavior, namely a MOSFET with a long enough L_{eff}, such that the channel appears to be uniform, and the drain does not exert any influence on the device output characteristics, except draining current. A good quantitative measure of the degree of asymmetry due to short-channel effects is the so-called device punchthrough mentioned earlier. This concept can be illustrated with the series of four figures, Fig. 3-8, 3-9, 3-10, and 3-11, which show the subthreshold characteristics of four NMOSFETs with $L_{eff} = 0.21\ \mu m$, $0.15\ \mu m$, $0.10\ \mu m$ and $0.055\ \mu m$, respectively, and all with SSR-I channel doping profile described in Fig. 2-1. The figures show the drain current, I_d, as a function of gate voltage, V_{gs}, with various drain voltages, V_{ds}. An "ideal" MOSFET should have such subthreshold characteristics as the one shown in Fig. 3-8, namely a steep subthreshold slope, denoted by the so-called S-factor, $\frac{dV_{gs}}{d\log_{10} I_d}$, denoted by SS in the figures, which measures the rate at which a MOSFET is turned on and off, and a minimal current variation, or a horizontal shift of the subthreshold $\log I_d$ *vs.* V_{gs} curves, due to V_{ds} under small V_{gs}, denoted by $\frac{\delta V_t}{\delta V_{ds}}$, which measures the degree of unwanted drain influence on the device output characteristics. The shift can be thought of as composed of a parallel component and a non-parallel component. When L_{eff} is relatively large, as the cases shown in Fig. 3-8 and 3-9 for the $L_{eff} = 0.21\ \mu m$ and $0.15\ \mu m$ devices, the shift is mainly parallel. This parallel shift is caused by the so-called "drain-induced barrier lowering", or DIBL, as shall be described below. As L_{eff} becomes small enough, the non-parallel shift, which is caused by device punchthrough, becomes more pronounced and constitutes a larger portion of the total shift, while the parallel shift becomes larger also. The parallel shift is due to the increasing drain influence on the conduction-band edge on the source side when the channel is short enough so that the source can "feel" the drain pulling the conduction band down with increasing V_{ds}. The non-parallel shift is due to the increasing drain influence on the current flow in the bulk depletion region when the channel is short enough so that the depletion

regions around the source and the drain can "communicate" with each other, which allows the drain to exert more control of the current flow than the gate, and the current is more independent of the gate control, V_{gs}. Both of these effects are termed "short-channel effects" because they are more pronounced when the channel length gets smaller, and both effects should be minimized as much as possible because they both make it harder to turn the MOSFET off in digital ULSI systems.

To illustrate the effect of the parallel shift (DIBL), $\frac{\delta V_t}{\delta V_{ds}}$, one can plot this quantity against the calculated electron velocity, g_m/WC_{ox}, as shown in Fig. 3-12. Another direct measure of the MOSFET turn-off characteristics is the so-called off-state current, I_{off}, defined as $I_{off} = I_d(V_{ds} = V_{DD})|_{V_{gs}=0}$, where V_{DD} is the power supply voltage. I_{off} summarizes both DIBL and punchthrough effects in a single number. Fig. 3-13 plots the relationship between the calculated electron velocity, g_m/WC_{ox}, and the measured off-state current, I_{off}, for the SSR-I NMOSFETs with various L_{eff}s.

As both Fig. 3-12 and Fig. 3-13 clearly show, the electron velocity cannot keep increasing indefinitely without worsening the short-channel effects, $\frac{\delta V_t}{\delta V_{ds}}$, or I_{off}. As L_{eff} becomes shorter and shorter, the rate at which the short-channel effects increase becomes higher and higher. This means that it becomes increasingly difficult to keep the short-channel effects in control while increasing the electron velocity, and the higher the electron velocity, the harder it is to confine $\frac{\delta V_t}{\delta V_{ds}}$ and I_{off}.

It thus seems that even for the best performing silicon MOSFETs, such as the SSR devices [25], the trade-off between the electron velocity and the short-channel effects has set a limit such that the electron velocity overshoot cannot be realized in a practical ULSI system to provide the extra improvement on current gain or transconductance over what is predicted by the drift-diffusion theory. Why is this so? Of course every physicist can easily come up with an answer out of the law of conservation, and that is, there is never a "free lunch" and some magic product of the good and the bad is conserved. But perhaps the answer can be more specific, even for a complicated many-body system like MOSFETs. One rather philosophical way of looking at this problem is as follows.

As shown in Fig. 3-14, a MOSFET can be considered as an multiple-terminal

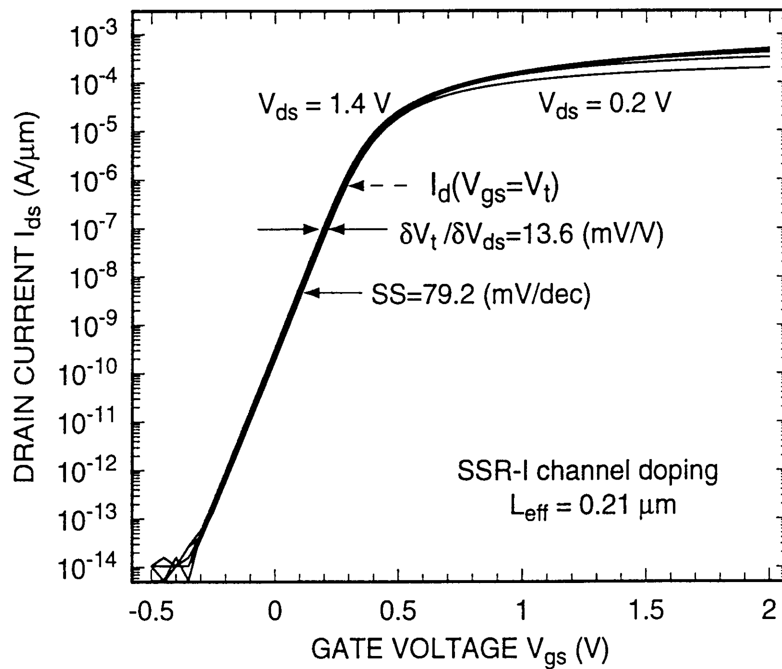

Figure 3-8: The $\log I_d$ *vs.* V_{gs} characteristics of a $L_{eff} = 0.21$ μm device with a SSR-I channel doping profile described in Table 2.1, and $T = 300$ K.

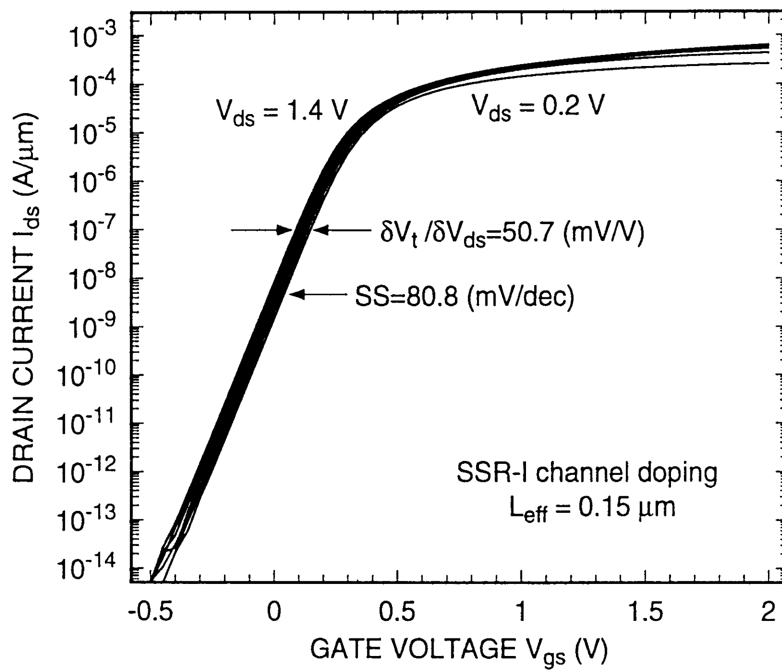

Figure 3-9: The $\log I_d$ *vs.* V_{gs} characteristics of a $L_{eff} = 0.15$ μm device with a SSR-I channel doping profile described in Table 2.1, and $T = 300$ K.

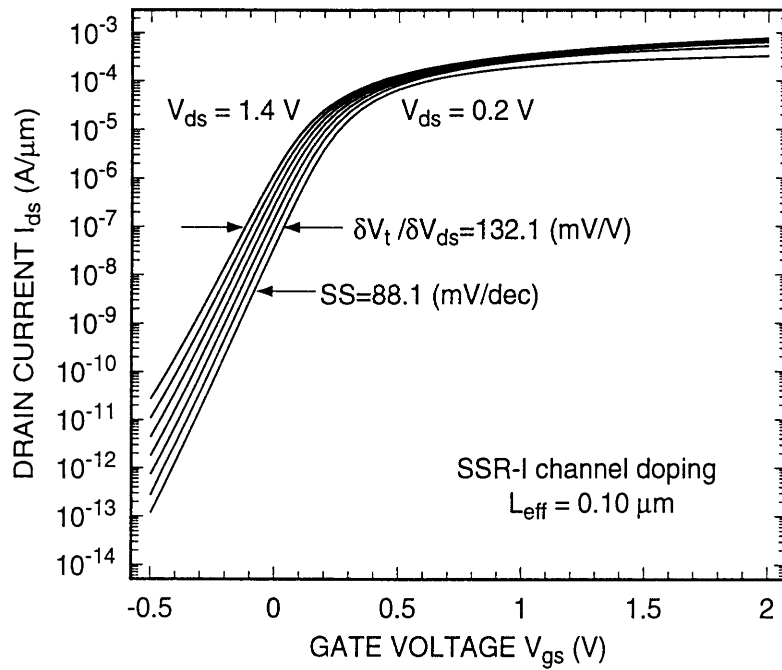

Figure 3-10: The $\log I_d$ *vs.* V_{gs} characteristics of a $L_{eff} = 0.10\ \mu m$ device with a SSR-I channel doping profile described in Table 2.1, and $T = 300\ K$.

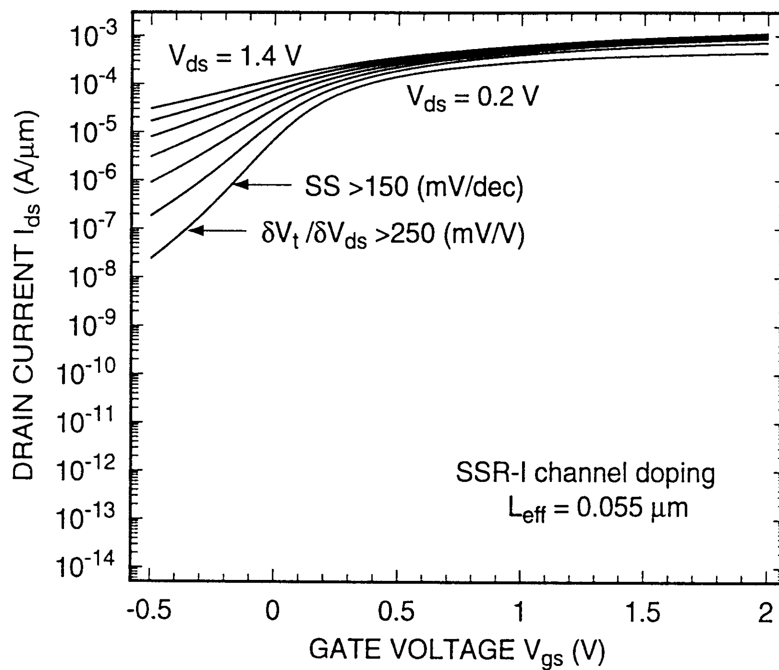

Figure 3-11: The $\log I_d$ *vs.* V_{gs} characteristics of a $L_{eff} = 0.055\ \mu m$ device with a SSR-I channel doping profile described in Table 2.1, and $T = 300\ K$.

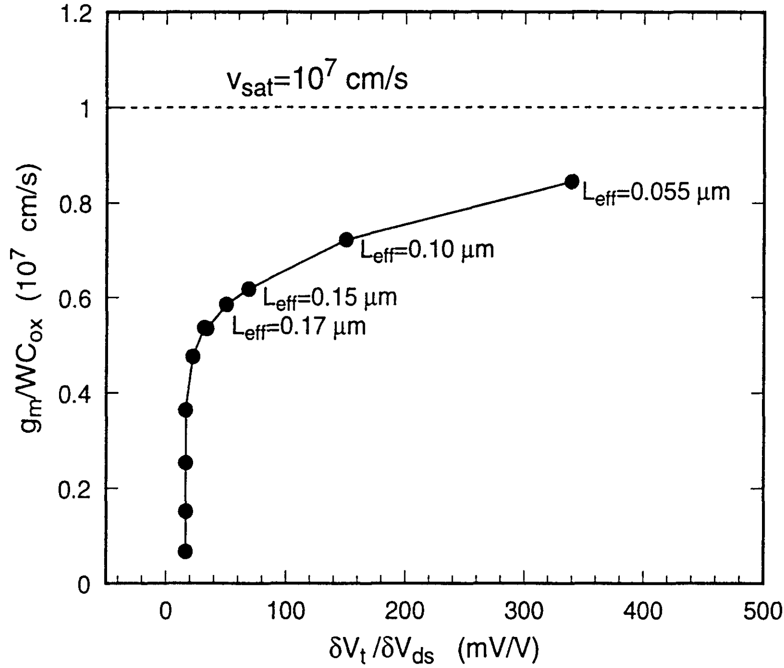

Figure 3-12: Electron velocity, g_m/WC_{ox}, as a function of drain-induced barrier lowering, $\delta V_t/\delta V_{ds}$, with L_{eff} as an implicit variable. All NMOSFETs have a SSR-I channel doping profile described in Table 2.1, $V_{ds} = 2.0\ V$, and $T = 300\ K$.

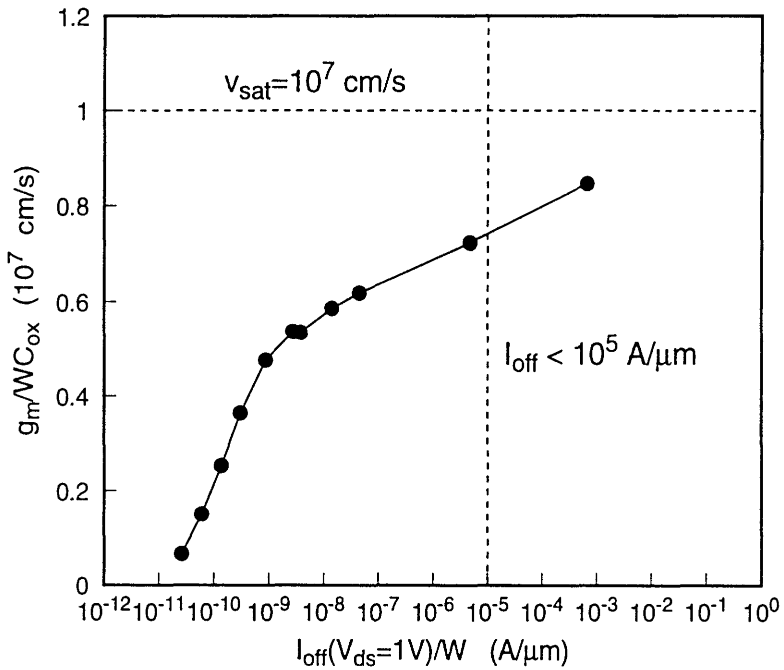

Figure 3-13: Electron velocity, g_m/WC_{ox}, as a function of off-current, I_{off}, with L_{eff} as an implicit variable. All NMOSFETs have SSR-I channel doping profiles described in Table 2.1, $V_{ds} = 1.4\ V$, and $T = 300\ K$.

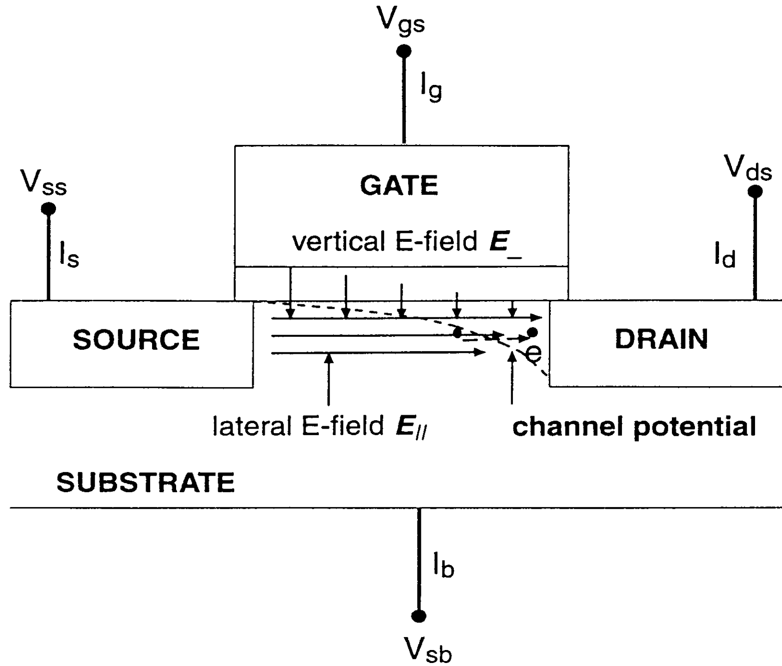

Figure 3-14: A schematic diagram of the electron transport under the vertical electric field control due to the gate and the lateral electric field control due to the drain in a MOSFET.

device with both the gate and the drain exerting control on the conduction electron flow in its inversion layer, by manipulating the gate-induced vertical electric field, $\mathcal{E}_{\perp}$, perpendicular to the electron flow, and the drain-induced lateral electric field, $\mathcal{E}_{\parallel}$, parallel to the electron flow. An "ideal" device should have the same sharp turn-on characteristics completely controlled by the gate voltage (thus the vertical electric field), as shown in Fig. 3-8, irrespective of the drain voltage (thus the lateral electric field), which is equivalent to a device without the short-channel effects, $\frac{\delta V_t}{\delta V_{ds}}$ and device punchthrough, as shown in Fig. 3-8. This can happen only when the magnitude of the vertical electric field far exceeds that of the lateral electric field, $|\mathcal{E}_{\perp}| \gg |\mathcal{E}_{\parallel}|$.

The effective electron mobility, μ_{eff}, is a universal function of the vertical effective electric field, $\mathcal{E}_{\perp}$, on a thermally oxidized silicon surface [46, 60]. This relationship can be summarized in the following equation:

$$\mu_{eff} = \mu_0 (\frac{\mathcal{E}_c}{\mathcal{E}_{\perp}})^C \tag{3.31}$$

where C and $\mathcal{E}_c$ are positive constants, and μ_0 is the low-field mobility defined in Eq. 3.20. The relationship between μ_{eff} and $\mathcal{E}_\perp$ is insensitive to the silicon surface impurity concentration, N_a, within a certain range [28], because surface roughness scattering on the $Si - SiO_2$ surface gives rise to the dominant electron mobility dependence on $\mathcal{E}_\perp$ in the inversion layer of a MOSFET. Eq. 3.31 indicates that the higher the gate control, the lower the effective electron velocity, i.e., an even higher lateral electric field is required to increase the electron velocity. It is this unique constraint, or trade-off, between the gate and the drain control, that imposes the limit on the electron velocity in a well-behaved MOSFET with an acceptable amount of short-channel effects. And it just so happens that the best trade-off frontier achieved so far by the SSR MOSFETs described in Chapter 2 still has not been pushed into the electron velocity overshoot regime at room temperature, as indicated in Fig. 3-12 and 3-13.

What would it take to break the barrier of $v_{sat} = 1.0 \times 10^7\ cm/s$ at room temperature? As is shown in more detail later in Chapter 4 (Fig. 4-11), for the best performing SSR MOSFETs described in Chapter 2, the empirical relationship among L_{eff}, g_m/WC_{ox}, and $\frac{\delta V_t}{\delta V_{ds}}$ is found to be (Eqs. 4.3, 4.4 and 4.5 in Chapter 4)

$$L_{eff} = \Theta(\frac{\delta V_t}{\delta V_{ds}})^{-\theta} = \Theta(\frac{\delta V_t}{\delta V_{ds}})^{-0.39} \tag{3.32}$$

$$g_m/WC_{ox} = \Lambda(L_{eff})^{-\lambda} = \Lambda(L_{eff})^{-0.42} \tag{3.33}$$

$$g_m/WC_{ox} = \Gamma(\frac{\delta V_t}{\delta V_{ds}})^{\gamma} = 1.01 \times 10^7(\frac{\delta V_t}{\delta V_{ds}})^{0.16}\ (cm/s) \tag{3.34}$$

where Θ, Λ, and Γ are constants fitted by experimental data.

Rewriting the drain current equation, Eq. 2.1 in Chapter 2, yields

$$\frac{I_d}{WC_{ox}} = \frac{1}{{L_{eff}}^{\alpha}}\mu_{eff}(V_{gs} - V_t)f(V_{gs} - V_t) \tag{3.35}$$

where $f(V_{gs} - V_t)$ is the product of all the $V_{gs} - V_t$ dependent factors except μ_{eff} and is independent of L_{eff}, and α summarizes all the channel length dependencies,

such as those due to velocity saturation and channel length modulation. Then, the electron velocity can be written as,

$$g_m/WC_{ox} = \frac{1}{{L_{eff}}^{\alpha}}(f(V_{gs}-V_t)\frac{d\mu_{eff}}{d(V_{gs}-V_t)} + \mu_{eff}(V_{gs}-V_t)\frac{df}{d(V_{gs}-V_t)}). \quad (3.36)$$

Substituting Eq. 3.31 into Eq. 3.36 yields

$$\begin{aligned} g_m/WC_{ox} &= \frac{\mu_{eff}(V_{gs}-V_t)}{{L_{eff}}^{\alpha}}(-f(V_{gs}-V_t)\frac{C}{\mathcal{E}_{\perp}(V_{gs}-V_t)} + \frac{df}{d(V_{gs}-V_t)}) \\ &= \frac{\mu_0}{{L_{eff}}^{\alpha}}F(V_{gs}-V_t). \end{aligned} \quad (3.37)$$

Experimental data show that the value of $V_{gs} - V_t$ where g_m/WC_{ox} reaches its maximum is independent of L_{eff} within the range from 0.5 μm down to sub-0.1 μm. In that case, if the electron velocity, g_m/WC_{ox}, is consistently calculated using the maximum g_m, then $F(V_{gs}-V_t) = F_0$ is independent of L_{eff}, and thus with the substitution of Eq. 3.33, one can rewrite Eq. 3.37 as,

$$g_m/WC_{ox} = \frac{\mu_0 F_0}{{L_{eff}}^{0.42}} \quad (3.38)$$

Substituting Eq. 3.32 into the above equation yields

$$g_m/WC_{ox} = \frac{\mu_0 F_0}{\Theta^{0.42}}(\frac{\delta V_t}{\delta V_{ds}})^{0.16} \quad (3.39)$$

This equation is equivalent to Eq. 3.34 which describes the trade-off relationship between the electron velocity and the short-channel effect shown in Fig. 3-12 and Fig. 4-11 in Chapter 4. It is therefore clear that the only solution to improve electron velocity without compromising on the short-channel effect is either to increase μ_0 or to improve the source/drain structure (see Chapter 4 for more details on this issue), as F_0 is fixed at a fixed temperature. The electron low-field mobility μ_0 in silicon is a well-studied quantity, and the commonly-believed value is somewhere between 550 $cm^2/(V \cdot sec)$ and 650 $cm^2/(V \cdot sec)$ at room temperature. For a given MOSFET structure and a given amount of "acceptable" short-channel effect, for

example, $\frac{\delta V_t}{\delta V_{ds}} = 100\ mV/V$, the best performing SSR NMOSFETs exhibit electron velocities $g_m/WC_{ox} = 0.70 \times 10^7\ cm/s$ according to Eq. 3.34. Thus μ_0 has to increase by 43% to somewhere between 785 $cm^2/(V \cdot sec)$ and 928 $cm^2/(V \cdot sec)$ for the electron velocity to reach $v_{sat} = 1.0 \times 10^7\ cm/s$ at room temperature.

Of course the electron saturation velocity in silicon, $v_{sat} = 10^7\ cm/s$, is not a magic elementary physical constant that imposes the limit on the future of MOSFET applications in ULSI systems. But with the existing MOSFET structure, for velocity overshoot to occur in a well-behaved device, one has to find a way to either improve the electron velocity without increasing the short-channel effects, or reduce the short-channel effects without degrading the electron velocity, or both. This issue will be addressed again in more detail in Chapter 4.

3.3 Conclusion

The electron velocity overshoot phenomenon in silicon inversion layers is experimentally investigated using high performance SSR n-channel MOSFETs (Chapter 2) with effective channel lengths down to sub-0.1 μm. It is found that the average electron velocity is not yet in the overshoot regime even for the best performing SSR MOSFET devices. From the perspective of deep-submicron MOSFET scaling, there exists a trade-off between the electron velocity and the device short-channel effects, such as the drain-induced barrier lowering effect and the punchthrough effect. The higher the electron velocity, the more pronounced is the short-channel effects, and the higher is the rate at which the short-channel effects increase with decreasing device effective channel lengths or increasing electron velocities. For the SSR MOSFET devices with an acceptable amount of drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}} = 100\ mV/V$, to break the barrier of the electron saturation velocity at room temperature, $v_{sat} = 1.0 \times 10^7\ cm/s$, the low-field electron mobility has to increase by 43%. This suggests the use of low-temperature (e.g., liquid-nitrogen temperature) silicon MOSFETs or compound semiconductor structures such as $Si - Ge$ FETs to push the frontier of the trade-off constraint between the electron velocity and the

short-channel effects. Chapter 4 will address the issue of the trade-off between device performance and short-channel effects again in more detail.

Chapter 4

Deep-Submicron MOSFET Scaling: Methodology and Analysis

4.1 Background

An empirical MOSFET scaling rule was first summarized in [5]. It states that for a MOSFET having a prescribed amount of short-channel effect, namely a drain-induced barrier lowering in this case, the trade-off among the four parameters, critical (i.e. minimum) channel length, L_c, gate oxide thickness, t_{ox}, channel doping, N_a, and junction depth, x_j, is given by

$$L_c = P(x_j t_{ox} W_{sd}^2)^{1/3} \tag{4.1}$$

where P is a constant, and W_{sd} is the sum of the depletion widths of the source and drain which is a function of channel doping N_a. This scaling rule was recently improved to correct the problem of resulting in zero L_c when $x_j = 0$ or $t_{ox} = 0$, and the inflexibility of not allowing a variable amount of short-channel effects [38]. The

improved trade-off rule for the parameters was given as

$$L_c = 2.2\ \mu m^2 (\frac{\delta V_t}{\delta V_{ds}})^{-0.37} (t_{ox} + 0.012\ \mu m)(W_{sd} + 0.15\ \mu m)(x_j + 2.9\ \mu m) \qquad (4.2)$$

where all length variables and constants are in units of μm, and $\frac{\delta V_t}{\delta V_{ds}}$, representing the drain-induced barrier lowering, or DIBL, is obtained from the parallel shift of the $\log(I_d) \sim V_{gs}$ curves at a given drain current level, corresponding approximately to $V_{gs} = V_t$, in the subthreshold regime. This improved scaling rule relates the critical channel length to the amount of short-channel effect represented by the DIBL effect, with the other three characteristic device dimensions, t_{ox}, W_{sd}, and x_j, as parameters. Eq. 4.2 was deduced entirely from numerical simulations on one set of homogeneous device structures centered around a conventional 0.25 μm technology with a uniform channel/substrate doping of $N_a = 4 \times 10^{17}\ cm^{-3}$ and a homogeneous source/drain profile. Its validity has not been confirmed with experimental data.

In the traditional practice of MOSFET scaling, the target values for L_c, t_{ox}, V_t, and maximum power supply voltage, V_{DD}, are typically defined at the outset. The resulting device must then meet three main criteria: (a) Electrostatic integrity in terms of short-channel behavior: acceptable threshold voltage roll-off, ΔV_t, *vs.* effective channel length, L_{eff}, acceptable drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, *vs.* effective channel length, L_{eff}, and negligible device punchthrough. (b) hot-carrier-induced degradation resistance in terms of the minimum time for a prescribed shift in threshold voltage, V_t, or in drain current, I_d, or in linear transconductance, g_{ml}, at maximum power supply voltage V_{DD}. (c) As high current drive, $I_d(V_{gs} = V_{DD}, V_{ds} = V_{DD})$, and as high saturated transconductance, g_m, as possible. The literature of scaling theories has so far concentrated only on the electrostatic integrity criterion with little or no attention given to hot-carrier-induced degradation and dynamic performance criteria.

In this chapter, a broader range of scaling relationships among fundamental MOSFET quantities is examined by including not only the two electrostatic quantities, effective channel length, L_{eff}, and short-channel effect (DIBL), $\frac{\delta V_t}{\delta V_{ds}}$, but also the

dynamic quantity, namely device speed or average carrier velocity, g_m/WC_{ox} (from here on, the terms average carrier velocity and device speed are used interchangeably). Both experimental and simulated data are included to cover a rather broad range of structural and electrical device parameters suitable for deep-submicron MOSFETs. First, the proper set of MOSFET parameters that are relevant to the scaling relationships among the three quantities is identified. Then, the scaling relationships in the form of power-laws are deduced via statistical nonlinear regression among these quantities from both experimental and simulation data. And finally, the sensitivities of these relationships on the particular set of parameters are investigated.

Hot-carrier-induced degradation criteria are not included in the scheme of the MOSFET scaling methodology presented in this chapter, but will be addressed in Chapter 5. It is reasonable to assume that the power supply voltage will be low enough for extreme submicron MOSFETs, so that hot-carrier effect may not impose severe restrictions on device scaling.

The experimental methodology developed here has proven to be very efficient, as shall be seen later in the chapter. For such a complicated physical system as a MOSFET, the macroscopic way of approaching the understanding of the physical laws behind the device operation is far more efficient than the microscopic way, such as complicated computational attempts to model realistic band structures and scattering rates. This is not to say that the microscopic way is not legitimate or not valuable, but simply not efficient enough to reveal the fundamental physical properties given the present supply of analytical and computational resources.

4.2 Methodology

The scaling relationships among the three MOSFET scaling variables are assumed to be in the following power-law form:

$$L_{eff} = \Theta(\frac{\delta V_t}{\delta V_{ds}})^{-\theta} \tag{4.3}$$

$$g_m/WC_{ox} = \Lambda(L_{eff})^{-\lambda} \tag{4.4}$$

$$g_m/WC_{ox} = \Gamma(\frac{\delta V_t}{\delta V_{ds}})^{\gamma} \tag{4.5}$$

where Θ, Λ, Γ and $\theta > 0, \lambda > 0, \gamma > 0$ are constants possibly dependent on the set of characteristic parameters which are identified as, (a) channel parameters: gate oxide thickness, t_{ox}, threshold voltage, V_t, and channel doping profile (e.g., surface impurity concentration N_s, bulk average impurity concentration N_a, and so on), and, (b) source/drain parameters: junction depth, x_j, parasitic resistance, R_{sd}, and junction abruptness (e.g., "halo" substrate doping). $\frac{\delta V_t}{\delta V_{ds}}$ is defined as the parallel shift of the $\log(I_d)$ $vs.$ V_{gs} curves at a given drain current level in the subthreshold regime, which excludes the non-parallel shift of $\log(I_d)$ $vs.$ V_{gs} curves due to punchthrough, and this issue shall be examined more closely later in the Discussion section of this chapter. g_m/WC_{ox} is the device speed, with g_m taken as the maximum saturated transconductance at a given drain bias, V_{ds}. Measurements on the three SSR-doped and the STEP-doped NMOSFETs described in Chapter 2 (Fig. 2-1 and Table. 2.1) and MINIMOS-4 (a 2-D MOS device simulator) [51] simulations are performed to examine the power-laws and their dependencies on the parameter set. Simulated NMOSFETs all have a uniformly doped channel profile with $N_a = 10^{17}\ cm^{-3}$ and a uniformly doped abrupt source/drain junction with various x_j from 30 nm to 150 nm and various t_{ox} from 3.3 nm to 10.3 nm. Eqs. 4.3, 4.4 and 4.5 are verified by nonlinear regressions on the data samples in the form of

$$\log(Y_i) = A + B\log(X_i) \qquad\qquad i = 1, 2, ...n \tag{4.6}$$

where X_i and Y_i are the samples of L_{eff}, $\frac{\delta V_t}{\delta V_{ds}}$ or g_m/WC_{ox}, A corresponds to the logarithm of the proportionality coefficient, and B corresponds to the power coefficient in Eqs.(3), (4) and (5). The statistical significance of the power-law hypothesis in Eqs.(3), (4) and (5) is estimated by the confidence factor r defined as

$$r = \frac{E[(\log(Y) - \log(\mu_Y))(\log(X) - \log(\mu_X))]}{\sqrt{E[(\log(Y) - \log(\mu_Y))^2]E[(\log(X) - \log(\mu_X))^2]}} \tag{4.7}$$

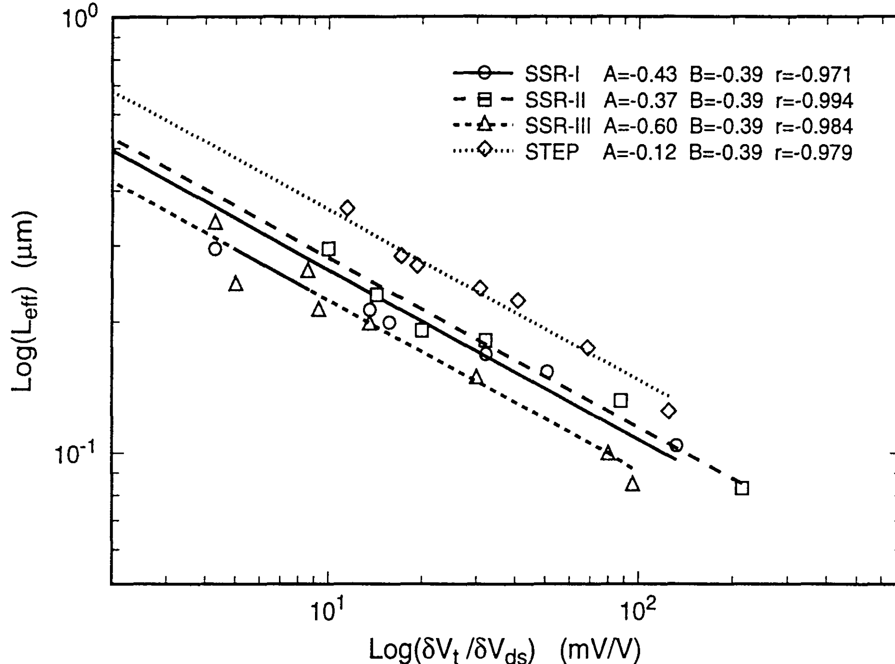

Figure 4-1: Measured drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.085 μm to 0.4 μm, $t_{ox} = 5.3\,nm$ and the four channel doping profiles (SSR-I, II, III and STEP) as shown in Fig. 2-1 and Table. 2.1. A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7.

where Y and X are sample vectors Y_i and X_i defined above, and $\mu_Y = E[Y]$ and $\mu_X = E[X]$ are sample means. r measures the covariance between the sample vectors.

4.3 Analysis

4.3.1 Channel parameters

The electrostatic relationship between L_{eff} and $\frac{\delta V_t}{\delta V_{ds}}$ (Eq. 4.3) suggested in [38] is verified against experimental data, as shown in Fig. 4-1, and simulations, as shown in Fig. 4-2 and Fig. 4-3.

The power-law relationship in Eq. 4.3 is indeed statistically significant on both experimental samples and simulation samples with an average $r = 0.982$ and $r = 0.995$, respectively. Fig. 4-1 shows that the power coefficient $\theta = B = -0.39$, which is rather close to $\theta = -0.37$ reported in [38], and to $\theta = -0.45$ reported here in Fig. 4-2

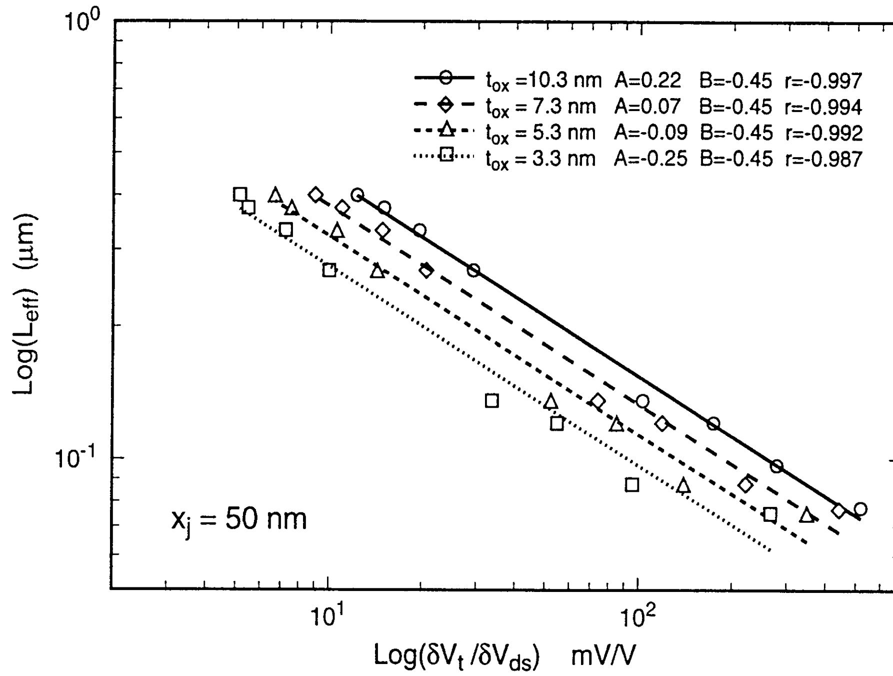

Figure 4-2: Simulated drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.07 μm to 0.4 μm, $t_{ox} = 3.3, 5.3, 7.3, 10.3$ nm, $x_j = 50$ nm and uniform channel doping profile ($N_a = 1 \times 10^{17}$ cm^{-3}). A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7.

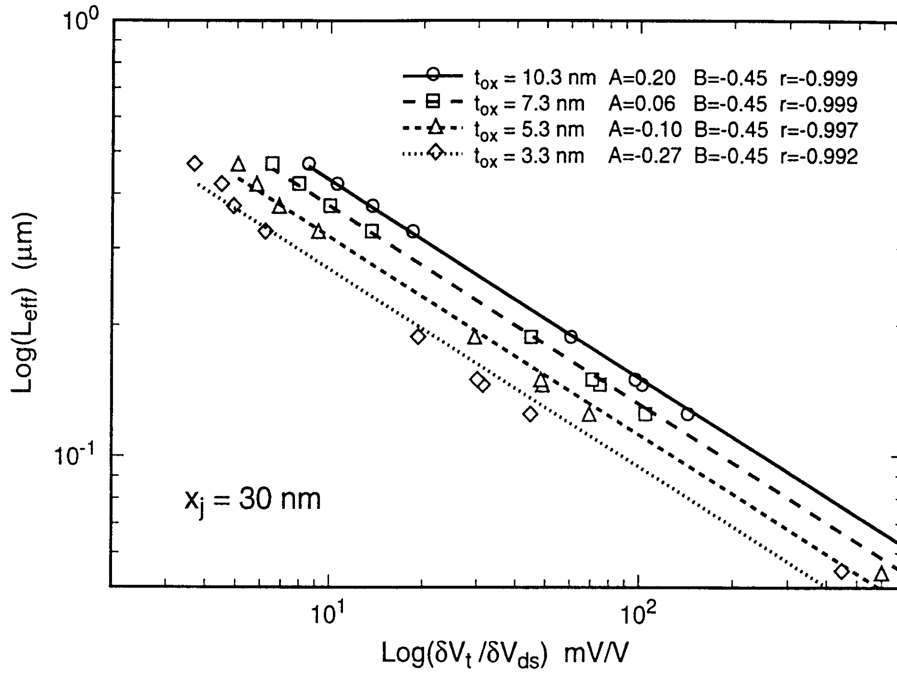

Figure 4-3: Simulated drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.07 μm to 0.4 μm, t_{ox} = 3.3, 5.3, 7.3, 10.3 nm, x_j = 30 nm and uniform channel doping profile ($N_a = 1 \times 10^{17}\ cm^{-3}$). A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7.

and Fig. 4-3. The slight difference between the experimental and simulated θ values is likely due to the difference in channel length definition between the metallurgical channel length in the device simulators and the electrically calibrated channel length for the actual devices. Furthermore, the experimental value of θ is independent of channel doping profile (SSR-I, II, III or STEP) and threshold voltage V_t within the range from 0.2 V to 0.47 V. Simulation shows that θ is also independent of t_{ox} at least within the range from 3.3 nm to 10.3 nm, and it remains the same for both $x_j = 50\ nm$ and $x_j = 30\ nm$ (Fig. 4-2 and Fig. 4-3). The difference in $A = \log(\Theta)$ coefficient reflects the difference in t_{ox}, channel doping profiles, and accordingly V_t. For example, devices having STEP channel doping profile and the lowest V_t have a given amount of $\frac{\delta V_t}{\delta V_{ds}}$ at the longest L_{eff}, whereas those having SSR-III channel doping and the highest V_t have the same amount of $\frac{\delta V_t}{\delta V_{ds}}$ at the shortest L_{eff}.

Fig. 4-4 shows the relationship between the measured intrinsic device speed and effective channel length for the four different channel doping profiles (SSR-I, II, III and STEP). Fig. 4-5 shows the same relationship for the simulated devices with four different t_{ox} values. Again the power-law relationship in Eq. 4.4 is statistically significant for both experimental data and simulation data with average $r = 0.997$ and 0.996, respectively. As L_{eff} decreases, the experimental results show that g_m/WC_{ox} rises with $\lambda = -0.42$ while the simulations show that g_m/WC_{ox} rises with $\lambda = -0.45$. Furthermore, λ is insensitive to t_{ox}, V_t, and channel doping profiles. The difference in the $A = \log(\Lambda)$ coefficient reflects the difference in V_t or t_{ox} when N_a is fixed as shown in Fig. 4-5, as one would expect. Both the experimental and simulated power coefficients are somewhat different from that in [43], which reported a much more rapid increase of saturated g_m with decreasing L_{eff} with a power coefficient of $\lambda = -0.67$. The difference could be explained if the hydrodynamic model incorporated in [43] overestimated the effect of electron velocity overshoot. The issue of velocity overshoot is addressed in more detail in Chapter 3.

As described above, the exponents θ and λ in Eqs. 4.3 and 4.4 are relatively insensitive to t_{ox}, V_t, and channel doping profile within their respective experimental range, whereas the proportionality coefficients Θ and Λ are distinct with distinct device pa-

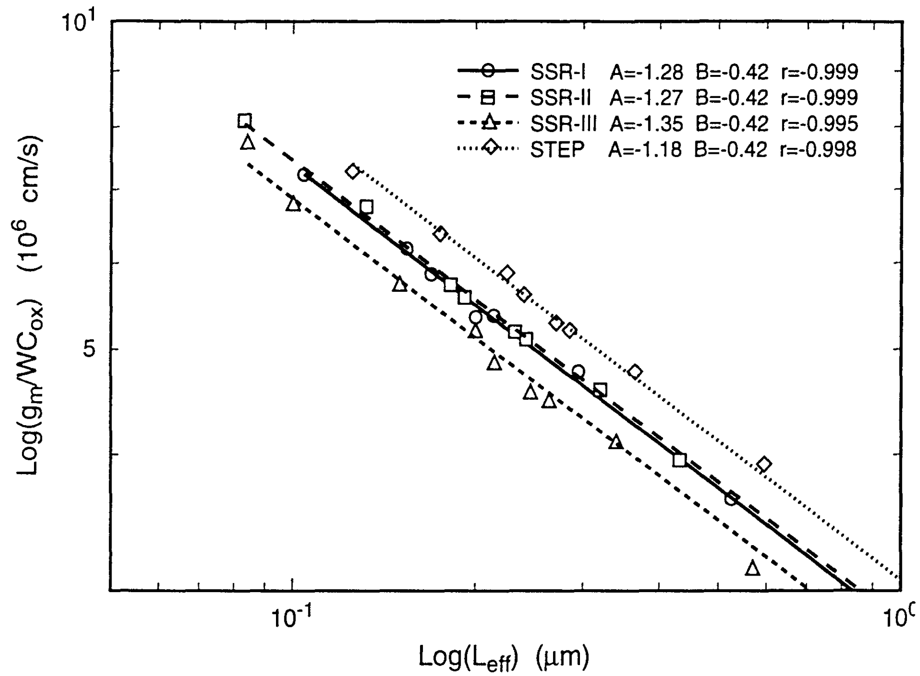

Figure 4-4: Measured device speed, g_m/WC_{ox}, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.085 μm to 0.4 μm, $t_{ox} = 5.3\ nm$ and the four channel doping profiles (SSR-I, II, III and STEP) as shown in Fig. 2-1 and Table. 2.1. A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7.

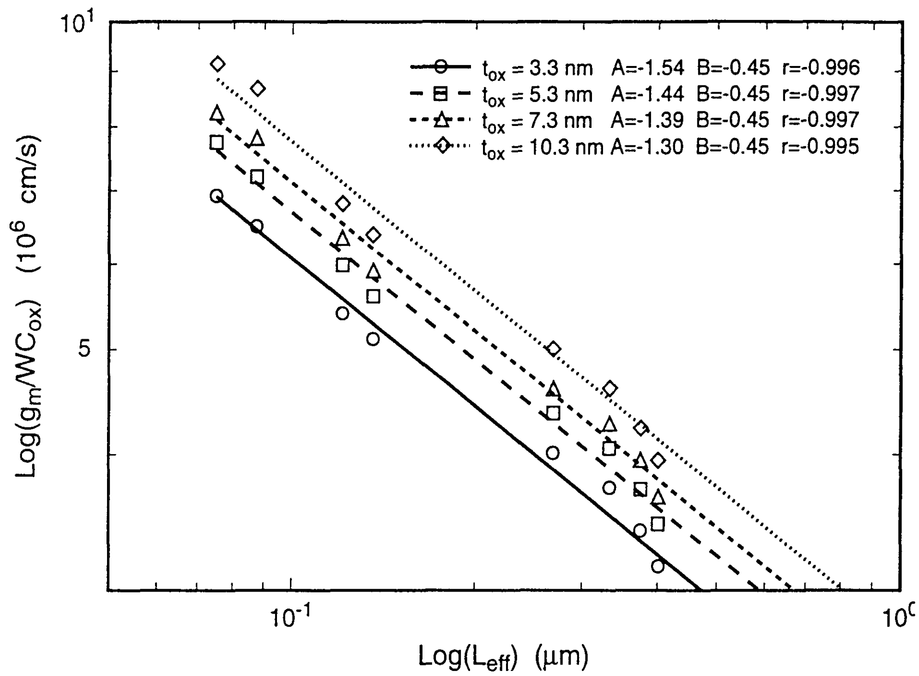

Figure 4-5: Simulated device speed, g_m/WC_{ox}, *vs.* effective channel length L_{eff} for NMOSFETs with L_{eff} ranging from 0.07 μm to 0.4 μm, $t_{ox} = 3.3, 5.3, 7.3, 10.3$ nm, $x_j = 50$ nm and uniform channel doping profile ($N_a = 1 \times 10^{17}$ cm^{-3}). A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7.

rameter sets. When the two power-law relationships are combined such that L_{eff} becomes an implicit variable in the relationship between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$, both the power coefficient γ and the proportionality coefficient Γ in Eq. 4.5 become rather insensitive to t_{ox}, V_t, and channel doping profile, as indicated by the rather consistent A and B regression coefficients in Fig. 4-6 and Fig. 4-7. Fig. 4-6 shows the measured g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ relationship for the four channel doping profiles (SSR-I, II, III and STEP). It clearly indicates that the relationship in Eq. 4.5 is rather independent of channel doping profile and V_t within the experimental range, as γ ($B = 0.16$) and $\Gamma(A = -1.05 \sim -1.08)$ remain nearly constant with average $r = 0.983$ for all profiles and V_ts. The reason for this "universal" trade-off relationship with respect to channel doping profile and V_t is not trivial, and it imposes an important question as to what device parameters are the dominant factors in determining the relationship between performance and short-channel effect. Fig. 4-5 shows the simulated g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ varying L_{eff} with t_{ox} from 3.3 nm to 10.3 nm while keeping channel doping and source/drain structure the same. Again, both the power coefficient γ ($B = 0.20$) and the proportionality coefficient Γ ($A = -1.40 \sim -1.44$) in Eq. 4.5 remain constant with average $r = 0.995$ for all values of t_{ox}, i.e., the relationship shows a universality with respect to t_{ox}. As t_{ox} decreases while other parameters remain constant, $\frac{\delta V_t}{\delta V_{ds}}$ decreases, while g_m/WC_{ox} also decreases due to lower mobility resulting from higher vertical electric field. The net effect of these two makes the t_{ox} dependence of the g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ relationship rather weak. This insensitivity with respect to t_{ox} is verified by further experimental data obtained from another set of NMOSFET devices (different process with non-halo source/drain) with two different $t_{ox} = 6.5$ nm and 9.0 nm values, as shown in Fig. 4-8.

We thus conclude that the channel parameters, as defined earlier, do not play a significant role in determining the trade-off relationship between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$, provided that those parameters are in a certain range that is appropriate for deep-submicron MOSFET design. This issue of "appropriateness" will be elaborated later in the Discussion section of this chapter.

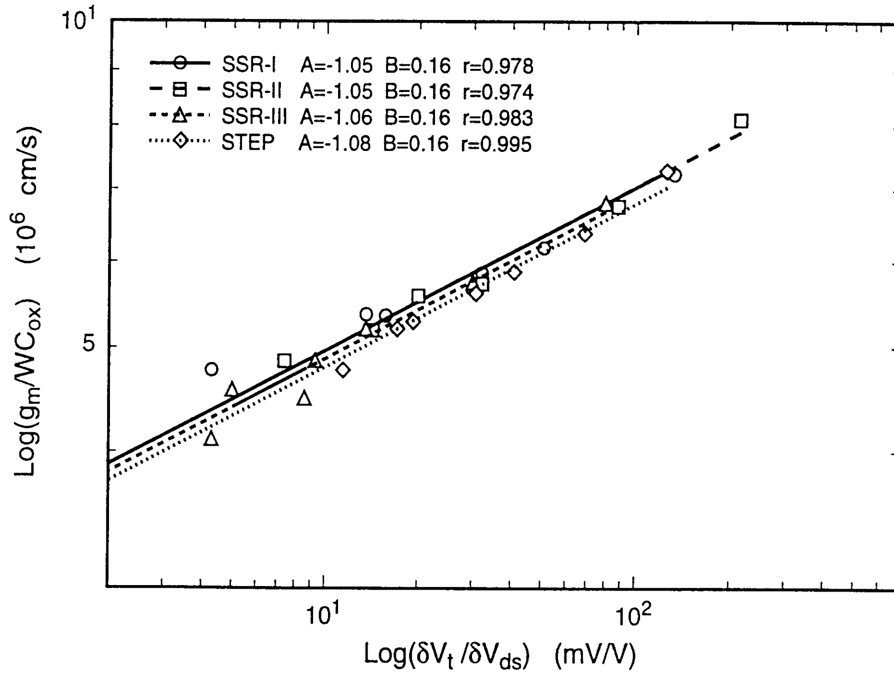

Figure 4-6: Measured device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for NMOSFETs with L_{eff} ranging from 0.085 μm to 0.4 μm, $t_{ox} = 5.3\ nm$ and the four channel doping profiles (SSR-I, II, III and STEP) as shown in Fig. 2-1 and Table. 2.1. A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7.

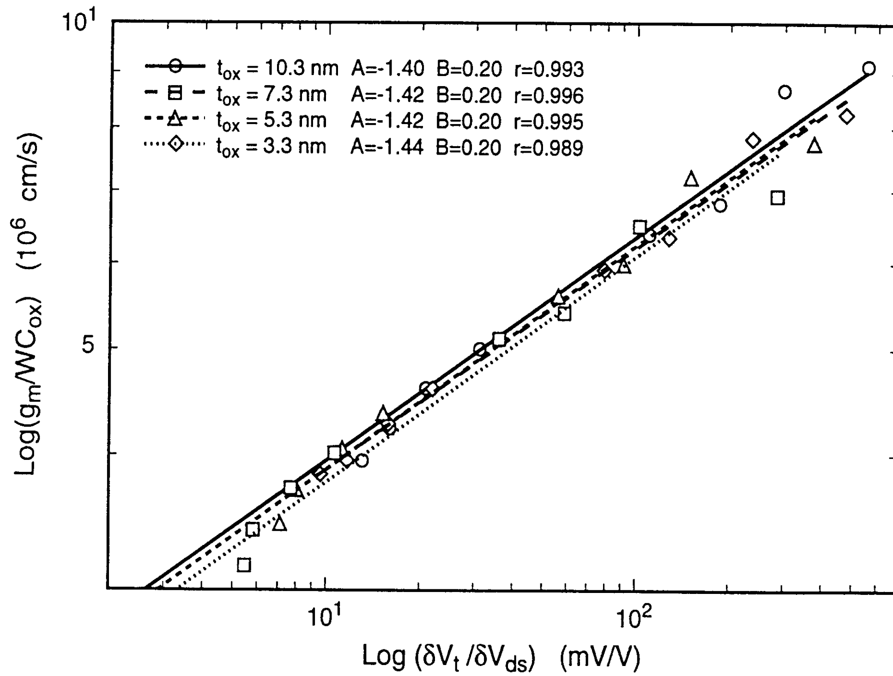

Figure 4-7: Simulated device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for NMOSFETs with L_{eff} ranging from 0.07 μm to 0.4 μm, t_{ox} = 3.3, 5.3, 7.3, 10.3 nm, $x_j = 50\ nm$ and uniform channel doping profile ($N_a = 1 \times 10^{17}\ cm^{-3}$). A and B are regression coefficients defined in Eq. 4.6. r is the statistical confidence factor defined in Eq. 4.7.

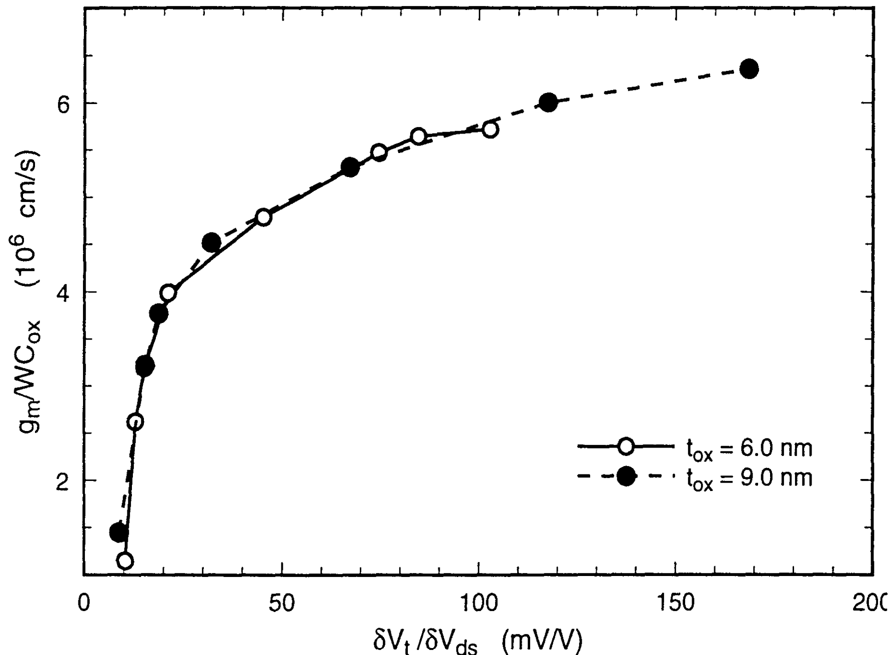

Figure 4-8: Measured device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for NMOSFETs with identical device structures except $t_{ox} = 6.5\ nm$ and $t_{ox} = 9.0\ nm$, respectively.

4.3.2 Source/drain parameters

In this section, the dependence of the g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ relationship on the source/drain parameters is examined. Those parameters include junction depth, x_j, parasitic resistance, R_{sd}, and junction abruptness which is determined by the particular type of "halo" substrate doping previously described. Fig. 4-9 shows the experimental data from two sets of NMOSFETs each containing "halo" and "non-halo" devices with the same long channel V_t.

For the lower long channel V_t set, the "halo" effect is more significant in shaping the trade-off between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$, as can be seen by comparing the sets with $V_t = 0.15\ V$ and $V_t = 0.57\ V$. This is of course expected as the lower V_t devices are more susceptible to DIBL, $\frac{\delta V_t}{\delta V_{ds}}$. Moreover, the devices with "halo" doping structure from the two sets with different V_t values show a nearly identical g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ relationship. This indicates that a proper "halo" structure can indeed improve the trade-off by containing the amount of DIBL while maintaining the device speed,

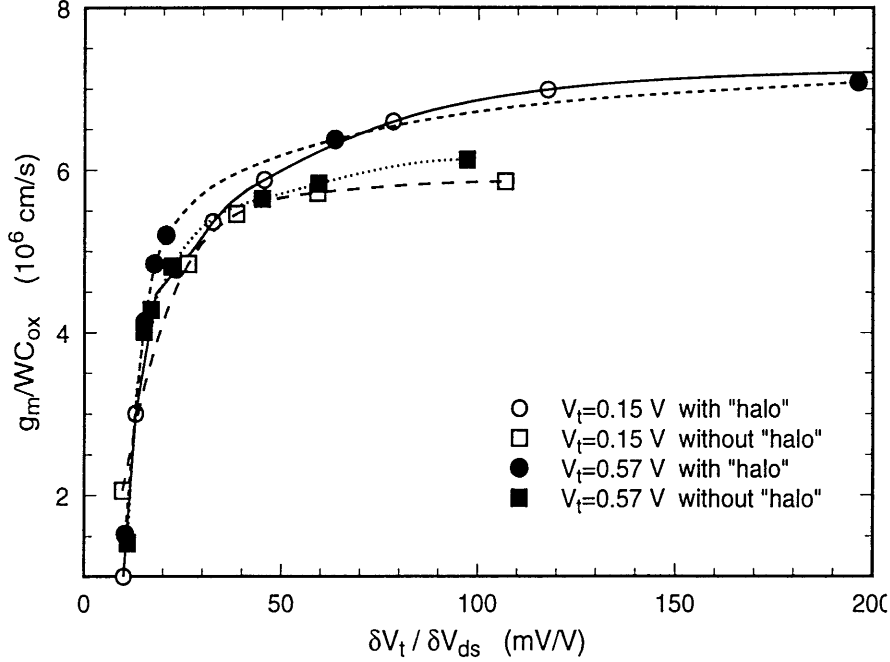

Figure 4-9: Experimental device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for two sets of NMOSFETs with and without an indium "halo" doping structure and with the same V_t. Empty symbols: $V_t = 0.15\ V$. Filled symbols: $V_t = 0.57\ V$.

without suffering from long channel V_t variation or channel mobility degradation, caused by the high level of impurity scattering near the source/drain junctions due to "halo" counter-doping.

Fig. 4-10 shows the simulated effect of x_j on the trade-off of g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ when all other parameters are kept the same.

As expected, the deeper junction worsens the DIBL without affecting the device speed, especially when L_{eff} is short and g_m/WC_{ox} is high.

The effect of R_{sd} on the trade-off between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$ is intuitively obvious as it lowers extrinsic g_m/WC_{ox} according to $g_m = g_{mi}/(1 + 0.5R_{sd}g_{mi})$, where g_{mi} is the intrinsic transconductance [10], while having no direct effect on DIBL, except through junction abruptness which is considered as a separate parameter.

To further strengthen the argument that it is the source/drain parameters that dominate the trade-off relationship between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$, it is worthwhile to compare several recent 0.1 μm technologies in the literature [55, 35, 64, 42, 67,

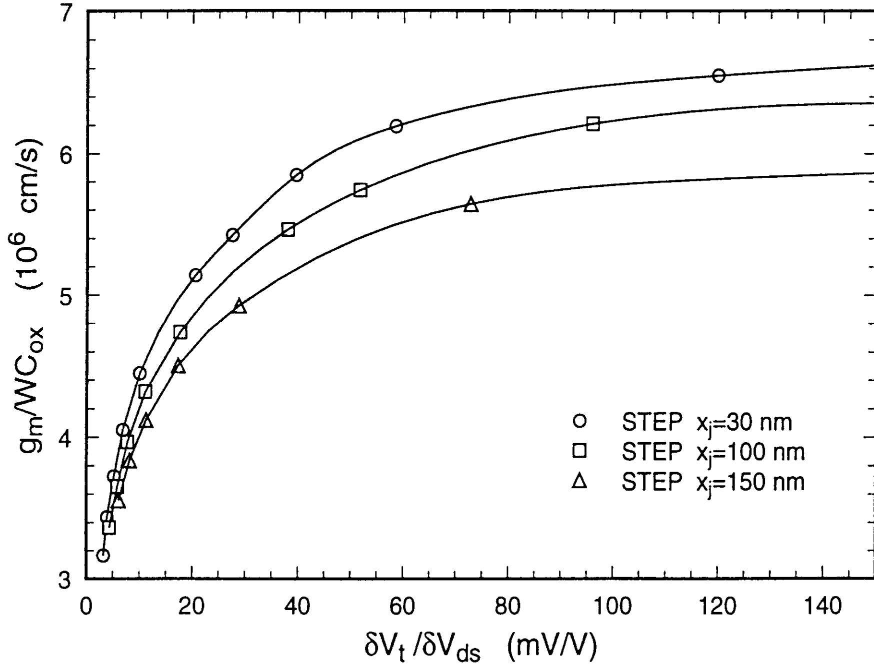

Figure 4-10: Simulated device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for NMOSFETs with $x_j = 30, 100, 150\ nm$ and uniform source/drain junctions.

33] which represent various approaches in designing deep-submicron MOSFETs. As shown in Fig. 4-11, different technologies produce different trade-off relationships.

A "better" source/drain technology can either enhance g_m/WC_{ox} without increasing $\frac{\delta V_t}{\delta V_{ds}}$, or reduce $\frac{\delta V_t}{\delta V_{ds}}$ without degrading g_m/WC_{ox} correspondingly. For example, by comparing MIT-SSR devices to AT&T Bell Lab devices [33], which have deeper x_j and have no "halo" doping, it can be seen that MIT-SSR devices show higher g_m/WC_{ox} at the same $\frac{\delta V_t}{\delta V_{ds}}$. On the other hand, the three IBM device sets [55, 64, 35] using a source/drain technology similar to that used in MIT-SSR devices exhibit a closer g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ relationship.

4.4 Discussion

From the above analysis, it is evident that the source/drain parameters, x_j, R_{sd} and the abruptness are the dominant factors in determining the trade-off between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$ compared to the channel parameters, t_{ox}, V_t and channel doping

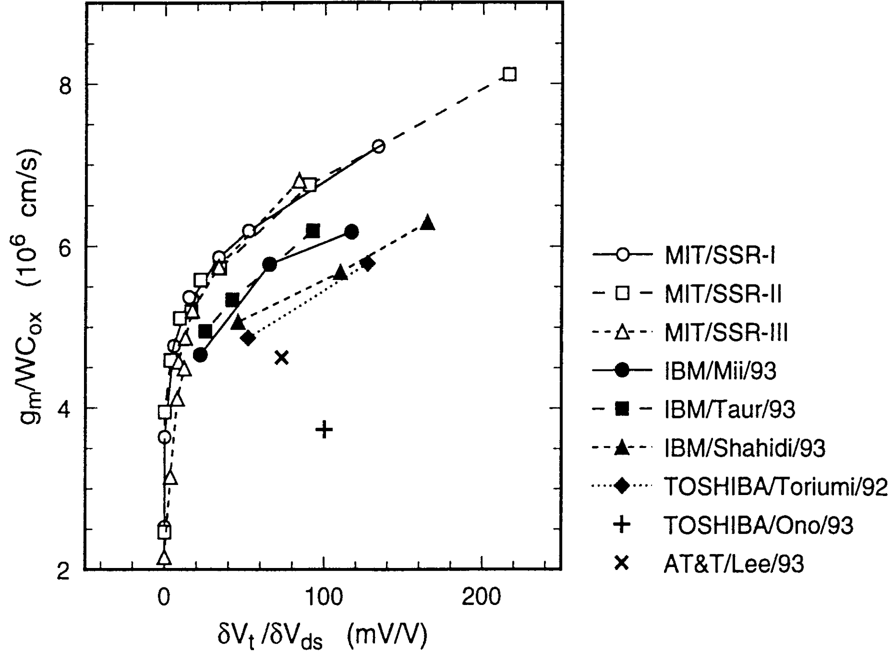

Figure 4-11: Comparison of measured device speed, g_m/WC_{ox}, *vs.* drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, for various recent 0.1 μm technologies given in the references.

profile, within a certain range appropriate for MOSFETs with L_{eff} from 0.5 μm down to sub-0.1 μm. Fig. 4-12 shows a hypothetical g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ trade-off relationship curve in the form of a power-law, as expressed in Eq. 4.5, which is fixed for a given set of source/drain parameters, as channel parameters do not alter that relationship.

A MOSFET with a specific set of L_{eff}, t_{ox}, and V_t is represented by a "device design point", $P(L_{eff}, t_{ox}, V_t)$ on this curve. When either L_{eff}, t_{ox} or V_t is varied independently, $P(L_{eff}, t_{ox}, V_t)$ moves approximately along the curve, since neither of the three alters the functionality between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$. Therefore the position of $P(L_{eff}, t_{ox}, V_t)$ is determined by any two of L_{eff}, t_{ox} and V_t, as only two of them are independent for a given $P(L_{eff}, t_{ox}, V_t)$ which is preset by the required "critical" amount of DIBL. Hence with a given source/drain technology, the MOSFET design rules are all constrained according to the fixed g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ trade-off curve.

It is a rather startling result that the g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ relationship is invariant with respect to t_{ox}, V_t, and channel doping profile, once the source/drain parameters

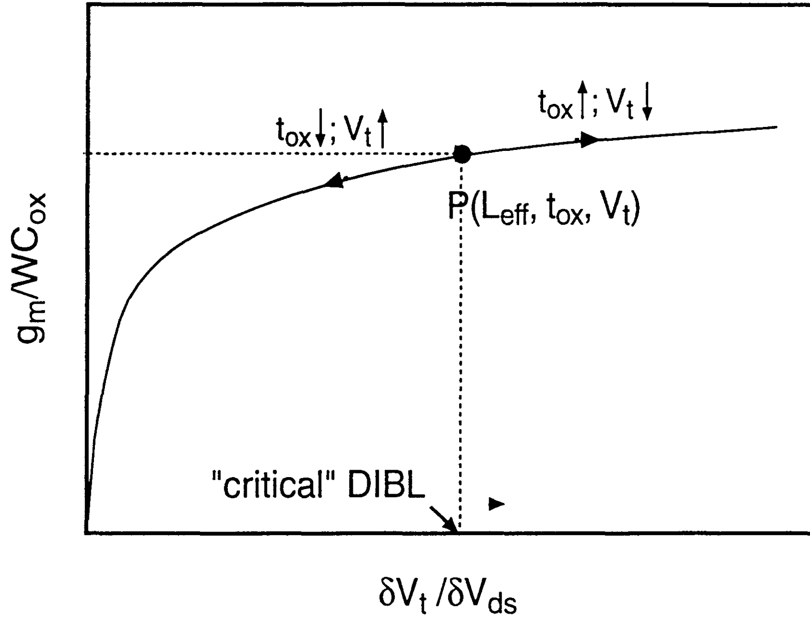

Figure 4-12: A hypothetical trade-off curve in the form of a power-law as expressed in Eq. 4.5 in g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ space. The arrows show the movement of the "device design point", $P(L_{eff}, t_{ox}, V_t)$, according to the changes in t_{ox} and V_t.

are set, which requires that the role of vertical channel doping engineering, and in particular SSR doping be clarified. By comparing the $V_t(L_{eff})$ characteristics in Fig. 4-13, it is clear that the SSR-III doping yields a more practical V_t value among all the dopings. It also exhibits, as shown in Fig. 4-13 and Fig. 4-14, the least amount of V_t roll-off, $\Delta V_t(V_{ds} = 0.05V)$, and $\frac{\delta V_t}{\delta V_{ds}}$ for a given L_{eff} among all the dopings, as expected.

Although the long channel V_ts of SSR-II and SSR-III are nearly the same, the heavier SSR-III doping provides better electrostatic integrity. However, it does pay the price of the lowest g_m/WC_{ox} for a given L_{eff} among all the dopings, according to Fig. 4-4. On the other hand, SSR-I and SSR-II have nearly identical $\Delta V_t(V_{ds} = 0.05V)$ and $\frac{\delta V_t}{\delta V_{ds}}$ for a given L_{eff} but quite different long channel V_t. So the two SSR dopings have allowed the tailoring of V_t while maintaining comparable performance and electrostatic integrity.

The above analysis of the three SSR doping profiles provides a consistent if not

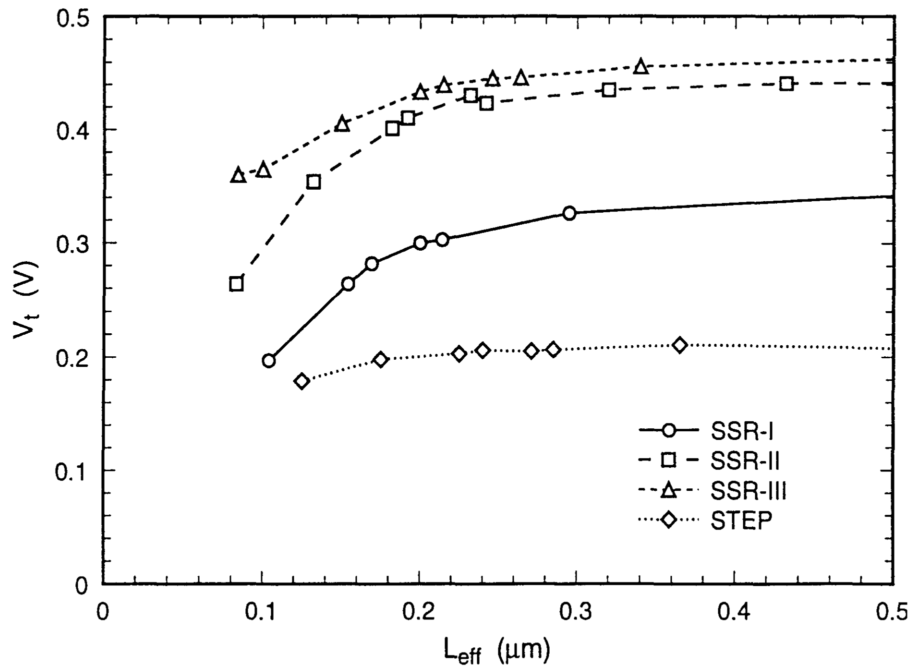

Figure 4-13: Threshold voltage roll-off behavior, V_t *vs.* L_{eff}, for NMOSFETs with SSR and STEP channel dopings, as shown in Fig. 2-1 and Table. 2.1.

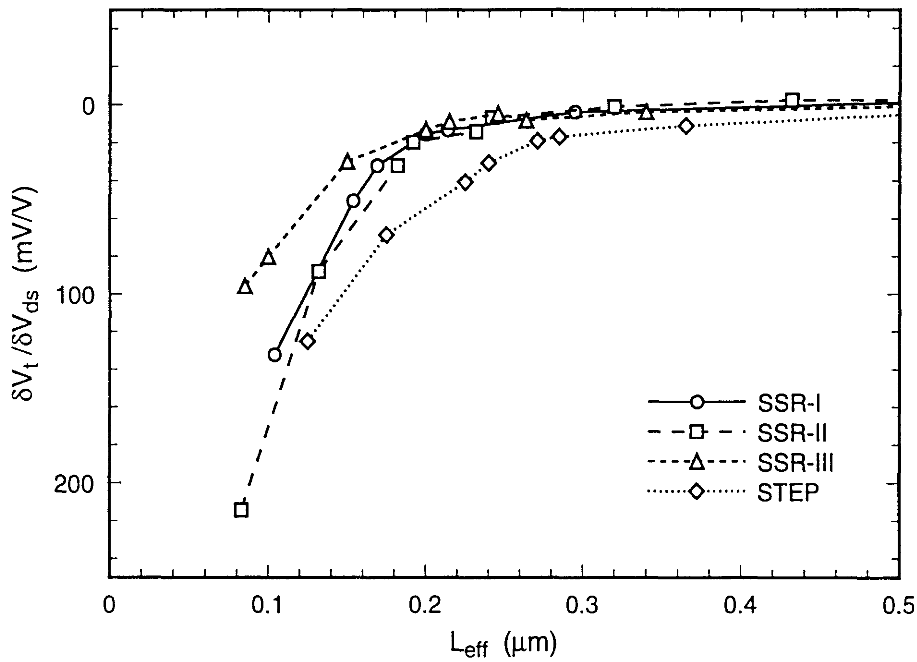

Figure 4-14: Drain-induced barrier lowering behavior, $\frac{\delta V_t}{\delta V_{ds}}$ *vs.* L_{eff}, for NMOSFETs with SSR and STEP channel dopings, as shown in Fig. 2-1 and Table. 2.1.

complete guildline for SSR channel profile engineering. However, it is also interesting to compare the conventional STEP doping profile with the SSR doping profiles. It is immediately obvious from Fig. 4-14 that the STEP-doped devices exhibit remarkable electrostatic robustness despite their low V_t. Clearly, this is a tribute to the source/drain technology used here. The question thus is whether the SSR channel profiles provide any performance advantages compared to the STEP channel profile.

As L_{eff} decreases to the sub-0.1 μm range, t_{ox} decreases to the 30 $\sim$ 50 nm range, and x_j decreases to 40 $\sim$ 60 nm range, the device channel doping, N_a, has to increase to over 5×10^{17} cm^{-3} to guarantee acceptable subthreshold characteristics, according to the classic scaling rule [16, 5]. For the STEP-doped devices with their V_t characteristics shown in Fig. 4-13, the channel doping level has to be raised from $N_a = 1.0 \times 10^{17}$ cm^{-3} to $N_a = 5.5 \times 10^{17}$ cm^{-3} to bring their long channel V_t from 0.21 V to that of SSR-III-doped devices, 0.47 V, which have a surface impurity concentration of only 1.0×10^{17} cm^{-3}. As the surface impurity concentration increases to well over 10^{17} cm^{-3}, devices with a STEP channel doping profile are expected to suffer from surface mobility degradation due to the fact that the effective channel mobility, μ_{eff}, deviates from the universal μ_{eff} *vs.* E_{eff} relationship (where E_{eff} is the effective vertical field), as ionized impurity scattering becomes increasingly significant, when t_{ox} is scaled down to a certain level [28, 62, 49]. Appropriate vertical channel engineering such as SSR doping can lower the surface impurity concentration so as to minimize μ_{eff} degradation, and therefore maximize g_m/WC_{ox}, while maintaining a reasonable V_t value necessary for the device subthreshold integrity. One other advantage of using SSR doping is that a lower surface impurity concentration is expected to reduce statistical threshold voltage fluctuation due to a random channel dopant distribution in sub-0.1 μm MOSFETs [40, 20].

The above analysis and conclusions are built upon a basic assumption that device punchthrough, or the non-parallel shift of $\log(I_d)$ *vs.* V_{gs} curves in the subthreshold regime, is not considered as a component of DIBL, $\frac{\delta V_t}{\delta V_{ds}}$, which only represents the parallel shift of $\log(I_d)$ *vs.* V_{gs} curves due to the conduction-band lowering near the source with increasing drain voltage. The scaling methodology and the data

analysis presented here are justified because the experimental devices with L_{eff} in the experimental range from 0.5 μm down to 0.085 μm used in this study show little signs of device punchthrough due to the use of "halo" doping structure, so that device punchthrough is decoupled from the short-channel effect. Undoubtedly SSR channel doping is another way of suppressing device punchthrough for it provides sufficient bulk charge to raise the bulk potential and contain the source/drain depletion [2, 55, 42, 25]. One way of quantifying device punchthrough is to measure the change in the subthreshold slope, or the S-factor denoted by $S(V_{ds})$, from low V_{ds} (e.g., $V_{ds} = 0.05\ V$) in the linear regime to high V_{ds} (e.g., $V_{ds} = V_{DD}$) in the saturation regime. Fig. 4-15 plots the percentage change in the S-factor, $\frac{\Delta S(V_{ds})}{S(V_{ds})}$, against L_{eff} for the SSR-doped and STEP-doped devices, where

$$\frac{\Delta S(V_{ds})}{S(V_{ds})}\ (\%) = (\frac{S(V_{ds} = 1.4\ V)}{S(V_{ds} = 0.05\ V)} - 1) \times 100\ \% \tag{4.8}$$

measures the degree of device punchthrough. An interesting observation can be made by comparing the $\frac{\delta V_t}{\delta V_{ds}}$ *vs.* L_{eff} characteristics, shown in Fig. 4-14, and the $\frac{\Delta S(V_{ds})}{S(V_{ds})}$ *vs.* L_{eff} characteristics, shown in Fig. 4-15, between the SSR-I-doped, SSR-II-doped and the STEP-doped devices. They both show similar DIBL at $L_{eff} = 0.12\ \mu m$ (within 20%), but yet very different subthreshold slope changes due to punchthrough (a factor of 4), which clearly demonstrates the decoupling of the short-channel effect and device punchthrough due to SSR doping.

Fig. 4-15 also shows clearly that SSR-III doping has the best punchthrough characteristics down to sub-0.1 μm because of its highest sub-surface bulk doping concentration, while STEP doping has the worst punchthrough characteristics because of its lowest sub-surface bulk doping concentration. It is also interesting to note that SSR-III doping is more punchthrough resistant than SSR-II, even though they both have nearly identical long channel V_ts, which is attributed to the heavier sub-channel doping concentration in SSR-III doping. This is another clear demonstration of the importance of the channel doping profile. All the experimental evidence supports the earlier argument that the channel doping profile does matter when punchthrough

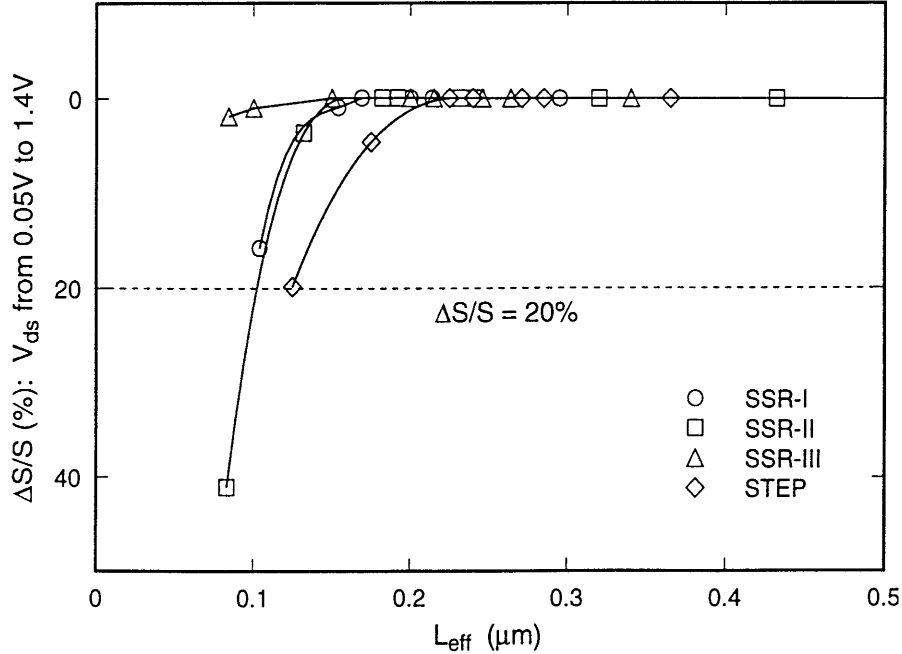

Figure 4-15: Punchthrough characteristics for NMOSFET with SSR-I,II,III and STEP channel doping profiles, as shown in Fig. 2-1 and Table. 2.1 The percentage change in the subthreshold slope, or the S-factor, $\frac{\Delta S(V_{ds})}{S(V_{ds})}$ is defined in Eq. 4.8.

cannot be prevented by other means, as L_{eff} decreases to a certain level with a given source/drain technology. The particular "halo" doping structure used in this work has been proven to be sufficient to suppress device punchthrough for L_{eff} down to about 0.12 μm, if the "acceptable" amount of punchthrough is set at $\frac{\Delta S(V_{ds})}{S(V_{ds})} \leq 20$ %, as indicated by the subthreshold characteristics of the STEP channel doping shown in Fig. 4-15. Therefore, both lateral channel profile engineering using more elaborate source/drain structures such as "halo" doping structure and SPI structure described in [23], and more elaborate vertical channel profile engineering such as SSR channel doping, are effective in preventing punchthrough in deep-submicron MOSFETs.

4.5 Conclusion

The scaling relationships among the three fundamental quantities of deep-submicron MOSFETs, device speed g_m/WC_{ox}, drain-induced barrier lowering (DIBL), $\frac{\delta V_t}{\delta V_{ds}}$,

and effective channel length, L_{eff}, were investigated with both device measurements and numerical simulations. The dependence of these relationships on the particular set of channel and source/drain parameters was also investigated experimentally and by numerical simulations in the deep-submicron L_{eff} regime from 0.5 μm down to sub-0.1 μm. The key findings are summarized as follows: (a) the scaling relationships can be expressed in appropriate power-law forms with excellent statistical significance for both experimental and simulation data samples; (b) the power coefficient relating $\frac{\delta V_t}{\delta V_{ds}}$ and L_{eff} is found to be insensitive to t_{ox}, V_t, and channel doping profile; (c) the relationship between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$ with L_{eff} as an implicit variable is insensitive to the channel parameters, t_{ox}, V_t, and channel doping within their respective experimental ranges; (d) the trade-off between device performance and the short-channel effect, i.e., g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$, is dominated by the source/drain parameters; and (e) the conclusions outlined above in (a) through (d) are justified with the absence of device punchthrough, and if device punchthrough is inevitable, and both lateral and vertical channel engineering are important in determining the trade-off between device performance and short-channel effects.

Chapter 5

Physics of Non-Equilibrium Hot-Carrier Effects

5.1 Theory of Hot-Carrier Current Generation in Si MOSFETs

In a silicon MOSFET, there are primarily two types of hot-carrier-current-generating mechanisms: impact ionization and channel hot-carrier injection. Impact ionization is caused by high-energy conduction carriers' transferring their momentum and energy to valence electrons or holes and creating electron-hole pairs during the collision process as they drift through the inverted channel. Hot-carrier injection is caused by high-energy conduction carriers' re-directing their momenta towards the Si-SiO_2 interface and subsequently being injected into the gate oxide from the inverted channel. The macroscopically measurable quantities associated with those two mechanisms are substrate current and gate current. As schematically shown in Fig. 5-1, the substrate current, I_b, is formed by the electrons or holes, generated through impact ionization, drifting to the substrate contact of a MOSFET under the electrostatic potential $V(x) - V_{sb}$, where $V(x)$ is the channel potential with respect to the source and V_{sb} is the substrate applied voltage. The gate current, I_g, is formed by the conduction electrons or holes with high enough energy to surmount the Si-SiO_2 barrier at the

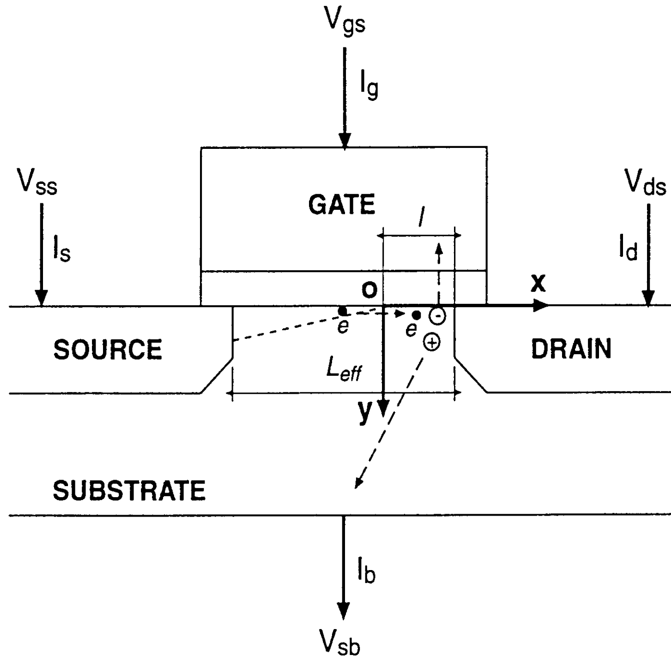

Figure 5-1: A schematic of hot-carrier-induced current generation in a n-channel MOSFET.

channel interface and to be injected into the gate oxide.

5.1.1 Channel electric field

The lateral electric field distribution, $E_{||}(x)$, plays a crucial role in modeling hot-carrier effects. A simplified model derived from the more elaborate ones described in [34, 29] is presented here for the purpose of deriving substrate and gate current models for hot-carrier effects. The channel of a MOSFET can be divided into two basic regions of operation, the inversion region and the saturation region where the electron velocity saturates (also called "pinch-off" region) because the lateral electric field in that region exceeds the saturation field, $\mathcal{E}_c = 4 \times 10^4\ V/cm$. It is well known [19, 29] that for conventional MOSFETs, hot-carrier effects are significant only when $|E_{||}(x)| \geq \mathcal{E}_c$. In the saturation region, the channel potential, $V(x)$, is a particular solution to the two-dimensional Poisson equation (Eq. 3.13 in Chapter 3) and is given

here by its one-dimensional approximation [18],

$$V(x) = V_{dsat} + V_0 \exp(\frac{x}{l}) \tag{5.1}$$

where l is the length of the saturation region, or pinch-off region, V_0 is a constant, and V_{dsat} is the potential at the pinch-off point in the channel which is taken as $x = 0$, as shown in Fig. 5-1. Thus, the lateral electric field can be modeled as

$$E_{||}(x) = \frac{V_0}{l} \exp(\frac{x}{l}). \tag{5.2}$$

The maximum electric field, E_m, occurs at the drain boundary, $x = l$, and is given by

$$E_m = \frac{V_{ds} - V_{dsat}}{l}. \tag{5.3}$$

The pinch-off length, l, can be assumed to be only dependent on the device structure, source/drain junction depth, x_j, and the gate oxide thickness, t_{ox}, as shown by experimental data [8] and numerical simulations [51].

A more accurate model introduced in [34, 29], which solves a pseudo two-dimensional Poisson equation, gives

$$E_{||}(x) = \frac{V_0}{2l}(\exp(\frac{x}{l}) + \exp(-\frac{x}{l})) \tag{5.4}$$

which reduces to Eq. 5.2 when x is large.

Generally speaking, $E_{||}(x)$ can be approximated with arbitrary precision according to the following hyperbolic expansion:

$$E_{||}(x) = \frac{V_0}{Kl} \sum_{k=0}^{\infty} [\exp(\frac{x}{(2k+1)l}) + \exp(-\frac{x}{(2k+1)l})] \tag{5.5}$$

where K is a normalization factor.

5.1.2 Substrate current generation

The electron-hole pair avalanche multiplication process due to impact ionization can be modeled as an exponential process with a probability distribution function given by $f_i(E_{||}(x)) = A_i \exp(-\Phi_i/q\lambda E_{||}(x))$, where $E_{||}(x)$ is the lateral electric field parallel to the carrier flow in the channel of a MOSFET, A_i and λ are constants typically fitted by experimental data, q is the elementary charge, and Φ_i is the threshold potential for an impact ionization event to take place. $f_i(E_{||}(x))$ is interpreted as the fraction of electron-hole pairs generated per conduction carrier per channel width, W. λ can be interpreted as the carrier mean-free path related to the carrier energy relaxation time, and $q\lambda E_{||}(x)$ is the carrier mean energy acquired from the lateral electric field, $E_{||}(x)$. In a n-channel MOSFET, the substrate current, I_b, resulting from the impact-ionization-generated holes collected at the substrate contact can then be written in terms of $f_i(E_{||}(x))$ as

$$I_b = \int A_i\ I_d\ \exp(-\frac{\Phi_i}{q\lambda E_{||}(x)})dx \tag{5.6}$$

$$= -\int_{E_s}^{E_m} A_i\ I_d\ \exp(-\frac{\Phi_i}{q\lambda E_{||}(x)})E_{||}(x)^2 \frac{dx}{dE_{||}} d(\frac{1}{E_{||}}) \tag{5.7}$$

where E_s is the lateral electric field at the source boundary and is usually much smaller than E_m. Note that

$$E_{||}(x)^2 \frac{dx}{dE_{||}} = lE_{||}(x) \tag{5.8}$$

according to Eq. 5.2. Thus,

$$I_b = A_i\ I_d\ l\ q\lambda E_m \int_{E_s}^{E_m} \exp(-\frac{\Phi_i}{q\lambda E_{||}})d(\frac{1}{q\lambda E_{||}}) \tag{5.9}$$

$$= \frac{A_i}{\Phi_i}\ I_d\ l\ q\lambda E_m \exp(-\frac{\Phi_i}{q\lambda E_{||}})\Big|_{E_s}^{E_m} \tag{5.10}$$

$$\approx \frac{A_i}{\Phi_i}\ I_d\ l\ q\lambda E_m \exp(-\frac{\Phi_i}{q\lambda E_m}). \tag{5.11}$$

I_b can also be expressed in terms of the reciprocal drain voltage, $1/(V_{ds} - V_{dsat})$, according to Eq. 5.3,

$$I_b = \frac{A_i}{\Phi_i} I_d \, q\lambda(V_{ds} - V_{dsat}) \exp(-\frac{l\Phi_i}{q\lambda(V_{ds} - V_{dsat})}) \quad (5.12)$$

$$\approx \frac{A_i'}{\Phi_i} I_d \, q\lambda \exp(-\frac{l\Phi_i}{q\lambda(V_{ds} - V_{dsat})}). \quad (5.13)$$

Typically for a MOSFET operating in the saturation mode, $V_{ds} - V_{dsat}$ is on the order of a few volts and is a much slower varying function of V_{ds} than $\exp(-\frac{l\Phi_i}{q\lambda(V_{ds}-V_{dsat})})$.

5.1.3 Gate current generation: the lucky-electron model

The gate current generation can be modeled, to first order, by the so-called lucky-electron model [56, 39, 63]. It is presented here from a slightly different approach. The gate current is caused by hot-electron injection from the channel to the gate oxide at the Si-SiO_2 interface in a MOSFET. Similar to the process assumed for the substrate current generation due to impact ionization, the probabilistic process for an ensemble of channel hot electrons to overcome the Si-SiO_2 potential barrier and to be injected into the gate oxide can also be assumed to follow an exponential process with its probability distribution function given by $f_b(E_{||}(x)) = C\exp(-\Phi_b/q\lambda E_{||}(x))$, where $E_{||}(x)$ is the lateral electric field parallel to the conduction carrier flow in the channel of the MOSFET, C and λ are constants typically fitted by experimental data, and Φ_b is the Si-SiO_2 potential barrier height for channel hot-electron injection. $f_b(E_{||}(x))$ is interpreted as the fraction of channel hot electrons injected into the gate oxide per conduction carrier per channel width, W. λ has the same physical meaning as that in the impact ionization process described earlier, and $q\lambda E_m$ is the electron mean energy acquired from the lateral electric field, $E_{||}(x)$, until the electron is injected into the gate oxide by a momentum re-directing collision. Different from the substrate current generation process, the gate current generation involves an additional process in which the electrons, after having their momenta re-directed, have to travel vertically to the Si-SiO_2 interface, without suffering other collisions, in order to be injected into the

gate oxide. In the lucky-electron model, the probability distribution associated with this process, denoted by $P(E_{ox})$, is assumed to be only dependent on the gate oxide field, E_{ox}, but independent of electron energy and lateral electric field [63]. Also, this process is assumed to be independent of the electron injection process mentioned earlier. Thus, the gate current can be written in terms of the product of the two probability distribution functions, associated with overcoming the gate oxide field and the barrier for hot-electron injection,

$$I_g = \int_0^{L_{eff}} I_d P(E_{ox})\, C \exp(-\frac{\Phi_b}{q\lambda E_{||}(x)}) dx \tag{5.14}$$

where L_{eff} is the effective channel length. Using the change of variables in Eq. 5.8 and assuming that the integral over $\exp(-\frac{\Phi_b}{q\lambda E_{||}(x)})$ can be approximated by its value at $x = L_{eff}$ where $E_{||}(L_{eff}) = E_m$, the above integral can be evaluated as

$$I_g = C I_d\, l\, (\frac{q\lambda E_m}{\Phi_b})^2 P(E_{ox}) \exp(-\frac{\Phi_b}{q\lambda E_m}). \tag{5.15}$$

I_g can also be expressed in terms of the reciprocal drain voltage, $1/(V_{ds} - V_{dsat})$, and according to Eq. 5.3,

$$I_g = \frac{C}{l\Phi_b^2} I_d\, (q\lambda)^2 (V_{ds} - V_{dsat})^2 P(E_{ox}) \exp(-\frac{l\Phi_b}{q\lambda(V_{ds} - V_{dsat})}) \tag{5.16}$$

$$\approx \frac{C'}{l\Phi_b^2} I_d\, (q\lambda)^2 \exp(-\frac{l\Phi_b}{q\lambda(V_{ds} - V_{dsat})}). \tag{5.17}$$

Again, $(V_{ds} - V_{dsat})^2$ is a much slower varying function of V_{ds} than $\exp(-\frac{l\Phi_b}{q\lambda(V_{ds} - V_{dsat})})$, so that the approximation used in Eq. 5.16 holds.

5.2 Hot-Electron Injection Barrier Lowering

The lucky-electron model [56, 63, 39] presented in the previous section has been widely used to model MOSFET gate and substrate currents as well as hot-carrier-induced device degradation to predict device reliability and device lifetime. The model is summarized again here in the following two equations based on the earlier

derivations (Eqs. 5.9 and 5.15):

$$\frac{I_g}{I_d} = C\ l\ (\frac{q\lambda E_m}{\Phi_b})^2\ P(E_{ox})\exp(-\frac{\Phi_b}{q\lambda E_m}) \tag{5.18}$$

and

$$\frac{I_b}{I_d} = A_i\ l\ (\frac{q\lambda E_m}{\Phi_i})\ \exp(-\frac{\Phi_i}{q\lambda E_m}) \tag{5.19}$$

where I_g, I_b and I_d are gate, substrate, and drain currents, respectively. C, A_i, λ and l are constants for a given device structure and a given applied voltage configuration. E_m is the peak lateral electric field in the direction of current flow in the channel, $E_m = E_{||}(x = x_0)$, where $x = x_0$ is the channel hot-electron injection point. $P(E_{ox})$ is a relatively weak function of the oxide field, E_{ox}, at the channel hot-electron injection point, $x = x_0$. Φ_b is the Si-SiO_2 potential barrier height for channel electrons at $x = x_0$, and Φ_i is the threshold potential for impact ionization. The lucky-electron model implies a linear correlation between the two hot-electron currents, I_g and I_b, because the current generating factor is assumed to be the same in both cases, namely E_m, and the hot-electron injection occurs at the same point where the impact ionization takes place. By combining Eqs. 5.18 and 5.19, this correlation can be expressed as

$$\Phi_i \log(\frac{I_g}{I_d}) - \Phi_b \log(\frac{I_b}{I_d}) = \Psi(E_{ox}, E_m) \tag{5.20}$$

where $\Psi(E_{ox}, E_m)$ is a slowly varying function of E_m, since $\Psi \sim \log(E_m)$, in comparison to the exponential dependence of I_b/I_d and I_g/I_d on E_m in Eqs. 5.18 and 5.19, $\exp(1/E_m)$. Eq. 5.20 indicates that the correlation coefficient relating I_g/I_d to I_b/I_d, Φ_b/Φ_i, is independent of device effective channel length, L_{eff}, as long as the rest of the device structure, such as the gate oxide thickness, t_{ox}, and the source/drain junction depth, x_j, is the same and the oxide field at the hot-electron injection point, $E_{ox}(x_0)$, is held constant. Previous works [39, 13] have confirmed this model in NMOSFETs with relatively long L_{eff}. In this section, for the first time, the dependence of the gate and substrate current correlation, Φ_b/Φ_i, on MOSFET effective channel length, L_{eff}, is demonstrated in the 0.1 μm regime. The physical mechanism that can possibly

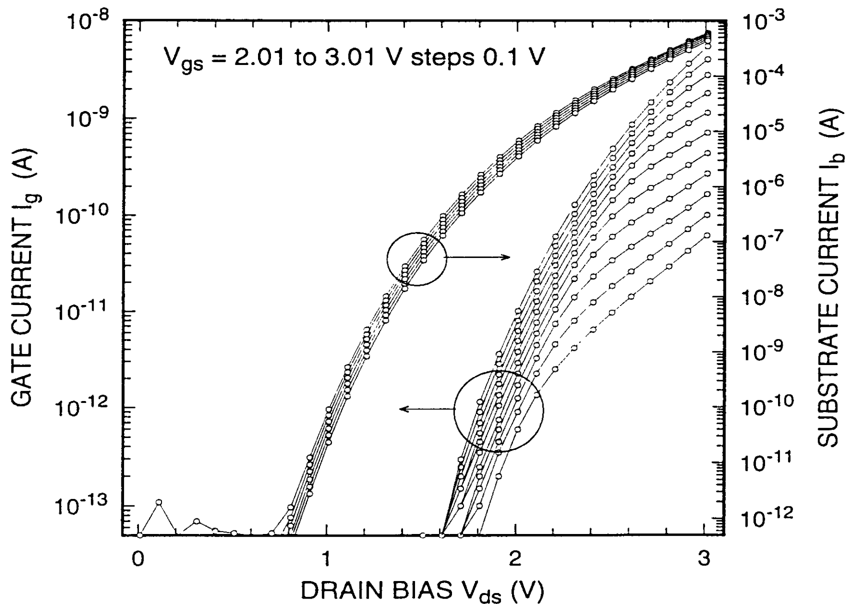

Figure 5-2: Gate current, I_g, and substrate current, I_b, characteristics as a function of drain voltage, V_{ds}, for the $L_{eff} = 0.1\ \mu m$ SSR-III MOSFET with V_{gs} steps of $0.1\ V$ from $2.01\ V$ to $3.01\ V$. The arrows indicate the axes which the curve sets are associated to.

explain this new effect is then discussed. This effect, termed as the hot-electron injection barrier lowering from here on, is of both theoretical and practical importance, because it is a direct indication of the difference in hot-electron transport dynamics between deep-submicron and long-channel MOSFETs, and it provides the first experimental evidence that the conventional lucky-electron model needs to be modified in the deep-submicron regime, in order to model the hot-carrier-induced currents and effects with higher precision.

5.2.1 Experimental observations

Gate current, I_g, and substrate current, I_b, were measured for four SSR NMOSFETs with $W_{eff} = 49.4\ \mu m$ and $L_{eff} = 0.1\ \mu m, 0.13\ \mu m, 0.18\ \mu m$, and $0.20\ \mu m$, as described in Chapter 2. Fig. 5-2 shows the I_g and I_b characteristics as a function of drain voltage, V_{ds}, for the $L_{eff} = 0.1\ \mu m$ SSR NMOSFET.

I_g was observed when $V_{ds} \geq 1.7\ V$ at $V_{gs} = 2.4\ V$, and I_b was observed when

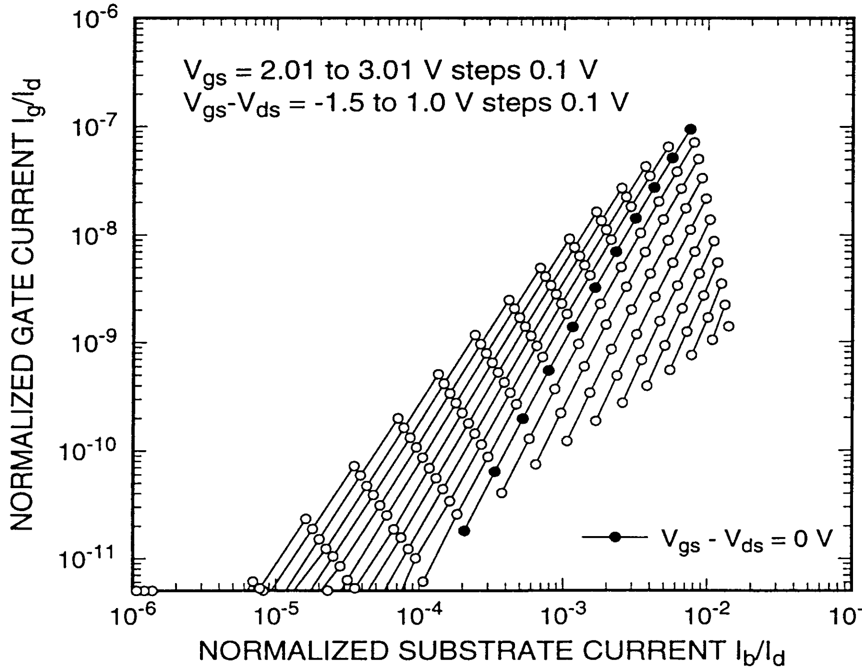

Figure 5-3: The correlation between normalized gate current, I_g/I_d, and normalized substrate current, I_b/I_d, for the $L_{eff} = 0.1\ \mu m$ SSR NMOSFET with constant $V_{gs} - V_{ds}$ steps of 0.1 V from -1.5 V to 1.0 V.

$V_{ds} \geq 0.7\ V$ at $V_{gs} = 2.0\ V$. The observation of NMOSFET gate current at such a low drain voltage, $V_{ds} = 1.7\ V$, is believed to be the first ever reported in the literature. There is no measurable I_g and I_b at low drain voltage, $V_{ds} \leq 1.0\ V$ in the strong inversion regime, which indicates that there is no appreciable gate-leakage current or reverse junction-leakage current even for NMOSFETs with L_{eff} down to 0.1 μm. The excellent behavior of these MOSFETs ensures the unambiguous interpretation of the experimental gate and substrate current data.

According to the lucky-electron model, the $\log(I_g/I_d)$ *vs.* $\log(I_b/I_d)$ relationship is a linear relationship, and the slope of this relationship gives the value of the correlation coefficient, $\Phi_b/\Phi_i = \frac{\Delta \log(I_g/I_d)}{\Delta \log(I_b/I_d)}$, at a given gate voltage V_{gs}, as shown in Fig. 5-3 for a set of fixed $V_{gs} - V_{ds}$ values. Note that the correlation coefficient has to be defined as above with fixed $V_{gs} - V_{ds}$ values to ensure fixed oxide field values, E_{ox}, for all applied voltage configurations. Fig. 5-4 shows the measured correlation coefficient, Φ_b/Φ_i, as a function of gate voltage V_{gs} with $V_{gs} = V_{ds}$ for the four NMOSFETs.

For a given V_{gs} value, Φ_b/Φ_i decreases as L_{eff} decreases. Furthermore, the Φ_b/Φ_i

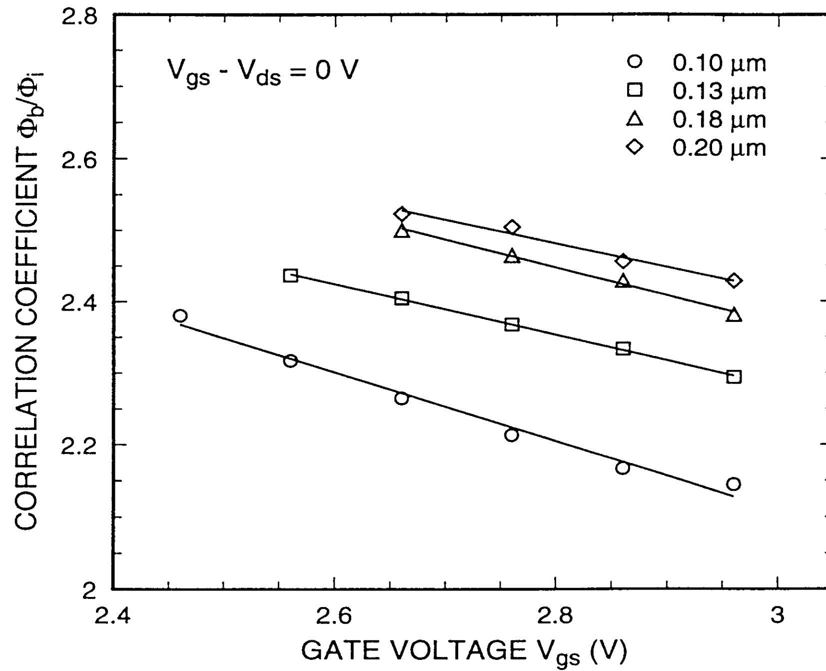

Figure 5-4: Measured correlation coefficient, Φ_b/Φ_i, *vs.* gate voltage, V_{gs}, with constant $V_{gs} - V_{ds} = 0\ V$ for the SSR NMOSFETs with $L_{eff} = 0.10\ \mu m$, $0.13\ \mu m$, $0.18\ \mu m$, and $0.20\ \mu m$. The straight lines are obtained from linear regressions on the data points.

dependence on V_{gs} becomes stronger as L_{eff} gets shorter. It is reasonable to assume that the effective impact ionization threshold, Φ_i, is independent of L_{eff} for a given device structure. Thus, the correlation coefficient lowering effect can be attributed to the reduction of the effective Si-SiO_2 barrier,Φ_i, for the channel hot-electron injection. MINIMOS-4 device simulations [51] show that when the four NMOSFETs with different L_{eff} are all biased at $V_{gs} - V_{ds} = 0\ V$, the oxide field, E_{ox}, at the maximum injection point $x = x_0$, where $E_{||}(x = x_0) = E_m$, is slightly positive (pointing downwards to the channel), and is nearly the same for all four NMOSFETs. This means that the Si-SiO_2 barrier for the channel hot-electrons at $x = x_0$ is nearly independent of L_{eff} under this particular bias configuration. The reason for the apparent Si-SiO_2 barrier lowering is likely that the contribution to the gate current due to hot-electron injection in the region closer to the source, $x < x_0$, becomes increasingly significant for shorter L_{eff}, because high energy electrons populate a wider region of the channel [32], and the lateral electric field, $E_{||}(x)$, is significantly higher in $x < x_0$ region, even though the peak lateral field, E_m, is nearly the same [20]. The oxide field at the Si-SiO_2 interface increases rapidly towards the source, and thus the Si-SiO_2 barrier decreases towards the source, according to $\Phi_b = [3.2 - \beta E_{ox}^{1/2} - \theta E_{ox}^{2/3}]$ (eV) [63], where β and θ are positive constants. Consequently, a greater fraction of the hot-electrons making up the gate current see a lower Si-SiO_2 barrier at $x < x_0$. In the lucky-electron framework, where only a single barrier height can be extracted from the gate and substrate current correlation, the ratio Φ_b/Φ_i is effectively lower for a shorter L_{eff} NMOSFET.

Fig. 5-5 shows the measured Φ_b/Φ_i ratio as a function of gate voltage V_{gs} with a set of fixed $V_{gs} - V_{ds}$ from -0.3 V to 0.3 V for the $L_{eff} = 0.1\ \mu m$ NMOSFET. For a given $V_{gs} - V_{ds}$ value, Φ_b/Φ_i decreases as V_{gs} increases. This can be explained by realizing that the vertical oxide field, $E_{ox}(x)$, across the region of channel hot-electron injection rises faster towards the source, i.e. for $x < x_0$, and hence the Si-SiO_2 potential barrier, Φ_b, is lowered in that channel region. In a shorter L_{eff} NMOSFET, this effect is stronger based on the analysis given above. The same argument can be applied to explain another feature indicated in Fig. 5-5 that for a

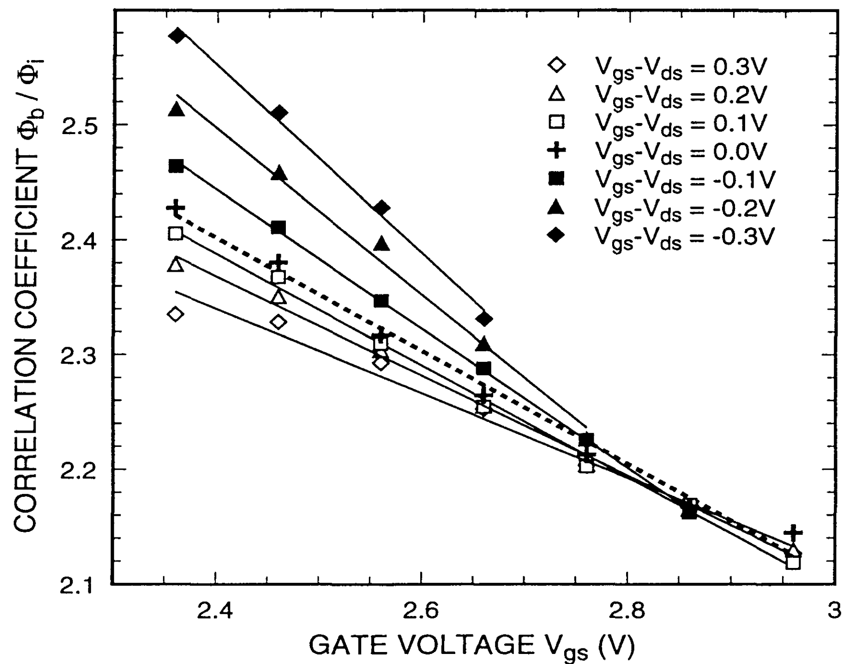

Figure 5-5: Measured Φ_b/Φ_i ratio as a function of gate voltage V_{gs} with a set of fixed $V_{gs} - V_{ds}$ steps 0.1 V from -0.3 V to 0.3 V for the $L_{eff} = 0.1$ μm SSR NMOSFET shown in Fig. 5-2. The straight lines are obtained from linear regressions on the data points.

given V_{gs}, Φ_b/Φ_i decreases as $V_{gs} - V_{ds}$ increases.

5.2.2 Analysis

A quantitative way of explaining the Φ_b/Φ_i dependence on L_{eff} in deep submicron NMOSFETs is to examine the specific assumptions made in deriving the lucky-electron model by comparing the model with the direct integration of a more realistic hot-electron energy distribution and a more realistic lateral electric field distribution across the channel from Monte-Carlo simulations, as demonstrated in [32, 21, 68].

Without loss of generality, the hot-electron current, I_h (e.g., I_g or I_b) can be expressed as

$$I_h \propto \int_{\epsilon_h}^{\infty} \epsilon^k d\epsilon \int_0^{L_y(x)} \int_0^{L_x} N(x,y) \exp[-\chi_h(x,y,\epsilon)\frac{\epsilon^m}{E_{||}(x,y)^n}]dxdy \tag{5.21}$$

where ϵ is the electron energy, ϵ_h is the effective threshold energy for an electron to cause hot-electron injection or impact ionization, $N(x,y)$ is the inversion layer electron density, $E_{||}(x,y)$ is the electric field in the direction of current flow, $\chi_h(x,y,\epsilon)$ is a fitting function depending on device structure and applied voltage configurations, and m, n, k are constant exponents with $k, m, n \geq 1.0$.

The hot-electron energy distribution function assumed in Eq. 5.21,

$$f_h(\epsilon) \propto \epsilon^k \exp(-\chi_h(x,y,\epsilon)\frac{\epsilon^m}{E_{||}(x,y)^n}) \tag{5.22}$$

accounts for the fact that the hot-electron energy distribution under high electric fields and electric field gradients in an extremely short-channel MOSFET exhibits kurtosis (high fourth moment or "fat" tail) and liptokurtosis (skew towards the high energy side) in comparison to a Maxwellian distribution, as shown by Monte-Carlo simulations [32, 68, 21].

The lucky-electron model can be derived by making the following approximations

to the above integral:

$$\begin{aligned} I_h &\propto \int_{\epsilon_h}^{\infty} \epsilon^k d\epsilon \; N(x_0, y_0) \exp[-\chi_h(x_0, y_0, \epsilon) \frac{\epsilon^m}{E_{||}(x_0, y_0)^n}] \\ &\propto \epsilon_h^k \; \exp[-\chi_h(x_0, y_0, \epsilon_h) \frac{\epsilon_h^m}{E_m^n}] \end{aligned} \quad (5.23)$$

where (x_0, y_0) is the position of the peak lateral field E_m.

It is easy to see that the lucky-electron model, Eqs. 5.18 and 5.19 are special cases of the above approximation in Eq. 5.23. Letting $m, n = 1$, $E_{||}(x_0, y_0) = E_m$, $\chi_h(x_0, y_0, \epsilon_h) = 1/q\lambda$, and $\epsilon_h = \Phi_b$ or $\epsilon_h = \Phi_i$, one obtains

$$I_g \propto \exp(-\frac{\Phi_b}{q\lambda E_m}) \quad (5.24)$$

and

$$I_b \propto \exp(-\frac{\Phi_i}{q\lambda E_m}) \quad (5.25)$$

respectively.

The approximation made in Eq. 5.23 is no longer valid when the integrand, dependent on both the hot-electron energy distribution and the lateral electric field distribution across the channel, significantly deviates from a δ-function. Such is the case in 0.1 μm-scale NMOSFETs at a relatively high drain voltage, as shown extensively by Monte-Carlo simulations [32, 68, 21]. Thus, the hot-electron injection barrier lowering effect should be predicted by the direct integration method shown in Eq. 5.21, if a realistic hot-electron energy distribution and a realistic lateral electric field distribution, $E_{||}(x)$, are implemented, such as those obtained by Monte-Carlo simulations which incorporate full multi-valley band structures and realistic scattering rates [32, 68]. The integral in Eq. 5.21 can be evaluated by numerical procedures. Then both gate and substrate currents can be calculated, and their correlation coefficient, Φ_b/Φ_i, can be obtained by calculating the slope of the I_g/I_d *vs.* I_b/I_d relationship with $V_{gs} - V_{ds} = 0$, such as that shown in Fig. 5-3.

Numerical simulations [1] were performed to evaluate the one-dimensional approx-

imation of the integral in Eq. 5.21 with $k = 3/2$, $m = 3$, $n = 3/2$, and $\chi_h = constant$,

$$I_h = -q\ A_i \int_0^{L_{eff}} n(x) \int_{\epsilon_h}^{\infty} \epsilon^{3/2} \exp[-\chi_h \frac{\epsilon^3}{E_{||}(x)^{3/2}}] d\epsilon\ dx \tag{5.26}$$

where A_i and χ_h are assumed to be constants fitted by experimental data. The values of the exponents, k, n, and m, are chosen such that the analytical distribution functions are as close to those obtained from Monte-Carlo simulations as possible. $n(x)$ is the one-dimensional inversion electron density calculated by the one-dimensional Poisson equation,

$$n(x) = \frac{1}{W_{eff}} C_{ox}(V_{gs} - V_t - V_x(x)) \tag{5.27}$$

in which W_{eff} is the effective channel width, C_{ox} is the static gate oxide capacitance, $C_{ox} = \epsilon_{Si}\epsilon_0/t_{ox}$, V_t is the long-channel threshold voltage, and $V_x(x)$ is the electron potential in the channel, as given by Eq. 5.1.

The simulation procedure is as follows. The integral in Eq. 5.26 is evaluated with the substitution of Eq. 5.27 under the bias conditions, $V_{gs} = V_{ds} = 2.4\ V, 2.6\ V, 2.8\ V$, and $3.0\ V$. χ_h is taken as $1.3 \times 10^8\ (V/cm)^{3/2}$ [1]. $E_{||}(x)$ is of the form in Eq. 5.4. $\epsilon_h = q\Phi_b$ is calculated according to

$$\Phi_b = 3.2 - \beta E_{ox}^{1/2} - \theta E_{ox}^{2/3}\ \ (eV) \tag{5.28}$$

where β and θ are positive constants with their default values given in [63], and the oxide field, E_{ox}, is calculated from the one-dimensional Poisson equation with given bias conditions and device structures. The gate current, I_g, calculated from the integral in Eq. 5.26, is compared to the experimental data such as the ones shown in Fig. 5-2, and the constant, A_i, in Eq. 5.26 is fitted accordingly. Then, substrate current, I_b, is calculated in a similar fashion under the same bias conditions [1]. Finally, Φ_b/Φ_i is calculated from the slope of the I_g/I_d *vs.* I_b/I_d relationship.

The simulation results are shown in Fig. 5-6 for two MOSFETs with $L_{eff} = 0.14\ \mu m$ and $0.20\ \mu m$, respectively. Although the overall magnitudes and the slopes of the simulated Φ_b/Φ_i *vs.* V_{gs} relationship do not match the experimental results

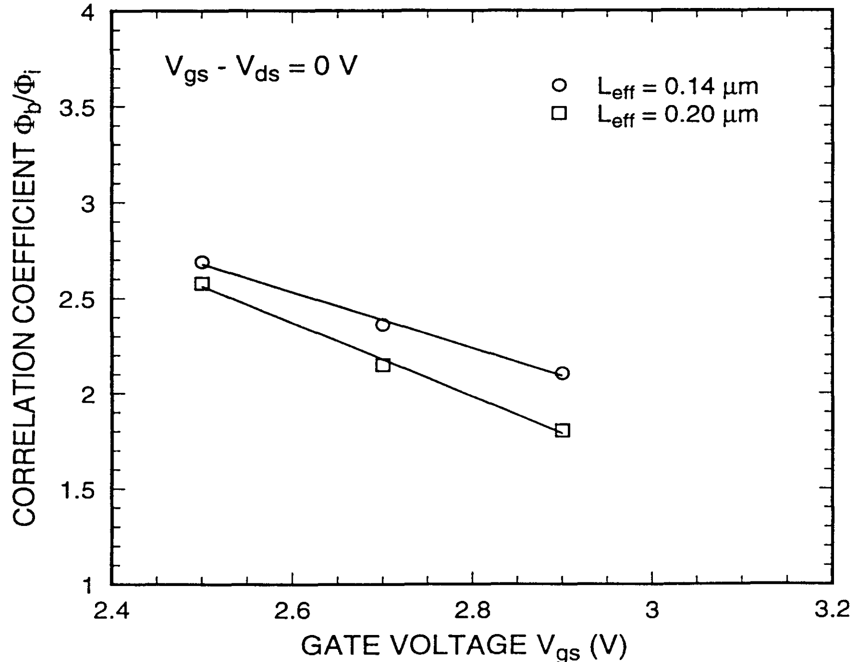

Figure 5-6: Simulated correlation coefficient, Φ_b/Φ_i, *vs.* gate voltage, V_{gs}, with constant $V_{gs} - V_{ds} = 0\ V$ for two NMOSFETs with $L_{eff} = 0.14\ \mu m$ and $0.20\ \mu m$. The straight lines are obtained from linear regressions on the data points.

precisely, the injection-barrier lowering effect is clearly demonstrated by the numerical procedure, as Φ_b/Φ_i decreases from $L_{eff} = 0.20\ \mu m$ to $0.14\ \mu m$. This is the consequence of the non-equilibrium nature of electron transport dynamics in the deep-submicron regime.

5.3 Hot-Carrier "Cooling" Effect

The previous section is an attempt to understand one of the hot-electron effects, namely, the hot-electron injection barrier lowering effect, associated with the non-equilibrium transport dynamics unique to extremely short-channel MOSFETs. The macroscopic characteristics of this effect carries the information that allows one to trace out the microscopic difference in the electron energy distribution and the lateral electric field between an extremely short-channel MOSFET, which shows the much more pronounced signature of non-equilibrium transport, and a relatively long-channel MOSFET, where the conventional "drift-diffusion" theory is accurate enough

to describe the transport properties. In this section, the tracing of that signature is viewed from a different angle. By investigating the dependence of the hot-carrier-induced currents on the effective channel length, an attempt can be made to probe the change in the hot-carrier energy distribution function in the inversion layers of MOSFETs with decreasing L_{eff} down to sub-0.1 μm. That change may reveal some unique properties of the non-equilibrium transport dynamics in extremely short-channel MOSFETs.

Specifically in this section, experiments are conducted to investigate an anonymous hot-carrier "cooling" effect first observed by [54], which reported that when L_{eff} becomes short enough, one of the primary indicators of hot-carrier effects, the normalized substrate current, I_b/I_d, reduces with decreasing L_{eff} at a constant maximum lateral electric field. This is contrary to what is predicted by the conventional "drift-diffusion" theory and the hydrodynamic models, which state that the hot-carrier-induced currents should keep increasing with decreasing L_{eff} because a larger fraction of the total conduction carrier population in the inverted channel is likely to have energy higher than the threshold level to cause either an impact ionization event resulting in substrate current generation, or a hot-carrier injection event resulting in gate current generation. This hot-carrier "cooling" effect is speculated to be one of the consequences of conduction carrier velocity overshoot, or quasi-ballistic transport, in the inversion layer of a deep-submicron MOSFET.

The existence or non-existence of the hot-carrier "cooling" effect in deep-submicron MOSFETs is of great theoretical interest and practical importance. As mentioned earlier, it is of great theoretical interest because it is a direct proof of hot-carrier quasi-ballistic transport in deep-submicron MOSFETs and it reveals some basic features of non-equilibrium transport dynamics under ultra-high electric fields and electric field gradients. It is of great practical importance because it imposes additional constraints or relaxes existing constraints on deep-submicron MOSFET scaling, depending on whether or not the effect indeed exists. If the hot-carriers are indeed "cooler" in deep-submicron MOSFETs, the hot-carrier-induced device degradation should be less severe in shorter channel devices due to substrate and gate current

reduction as opposed to longer channel devices. Then there should be even more incentive to go after shorter channel MOSFETs than just seeking higher speed and lower power consumption.

5.3.1 Hot-electron "cooling" effect and non-equilibrium transport dynamics

The hot-carrier-induced current generation process can be modeled as a "thermionic emission" process in which the carriers follow a displaced Gaussian distribution characterized by a single carrier temperature, T_e, and a group velocity, v_d, and only those carriers with energy higher than some threshold energy, ϵ_i, give rise to the hot-carrier-induced current, I_h, i.e.,

$$\begin{aligned} I_h &= \int_{\epsilon_i}^{\infty} d\epsilon \int_{\Omega} F(\epsilon, \Omega)\, d\Omega && (5.29) \\ &= \int_{\Omega} f_{\Omega}(\Omega)\, d\Omega \int_{\epsilon_i}^{\infty} f(\epsilon)\, d\epsilon && (5.30) \\ &\sim A(\epsilon_i) && (5.31) \end{aligned}$$

where Ω represents the momentum space, $F(\epsilon, \Omega)$ is the joint probability distribution in $\Omega \otimes \epsilon$ space, $A(\epsilon_i)$ is the area under the $f(\epsilon > \epsilon_i)$ curves in Fig. 5-7, and $A(\epsilon_i)$ is directly proportional to the hot-carrier-induced current in this type of thermionic-emission model . Note that in deriving the above equation, it is implicitly assumed that $F(\epsilon, \Omega) = f_{\Omega}(\Omega) f(\epsilon)$, that is, for this particular type of thermionic emission process, the joint probability distribution in the position and momentum space is orthogonal to that in the energy space. This assumption is justified because it is within the general framework of the hydrodynamic model (Chapter 3) where the translational relaxation, the momentum relaxation, and the energy relaxation processes are treated independently.

A conceptual picture of this hot-carrier "cooling" effect is shown in Fig. 5-7 where three displaced Gaussian distributions are sketched, representing (a gross approxima-

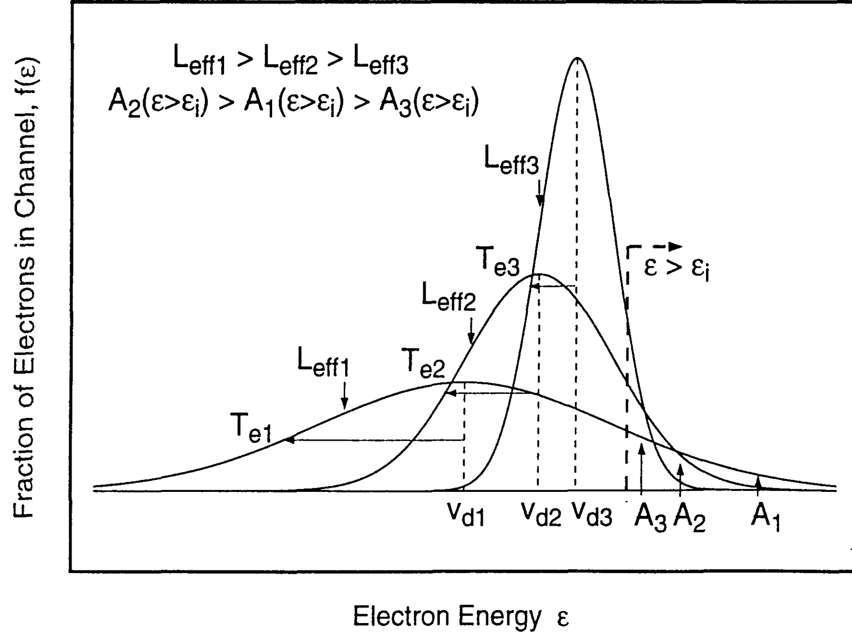

Figure 5-7: Three displaced Gaussian distributions representing the energy distribution of channel hot electrons in three MOSFETs with different L_{eff}s. $L_{eff1} > L_{eff2} > L_{eff3}$, $v_{d3} > v_{d2} > v_{d1}$, $T_{e1} > T_{e2} > T_{e3}$, and $A_2 > A_1 > A_3$.

tion, of course) the energy distribution of channel hot electrons with group velocities, v_{d1}, v_{d2}, and v_{d3}, and electron temperatures, T_{e1}, T_{e2}, and T_{e3}, in three MOSFETs with effective channel lengths, L_{eff1}, L_{eff2}, and L_{eff3}, and $L_{eff1} > L_{eff2} > L_{eff3}$. If hot-electron "cooling" indeed occurs to the extent that the electron temperature for the shortest channel MOSFET is much smaller than those for the longer channel MOSFETs, $T_{e3} \ll T_{e2}, T_{e1}$, it is possible that there is the least number of hot electrons with energy greater than the thermionic threshold, ϵ_i, to cause hot-electron-induced current, that is, $A_3(\epsilon_i) < A_2(\epsilon_i), A_1(\epsilon_i)$. In the limit of ballistic transport, the energy distribution, $f(\epsilon)$, at the source of a MOSFET, is a shifted replica of $f(\epsilon)$ at the drain. On the other hand, if the "cooling" effect is not strong enough to cause significant reduction in electron temperature, such as the case of L_{eff2} *vs.* L_{eff3}, $A_2(\epsilon_i)$ can still be greater than $A_1(\epsilon_i)$ even though $T_{e2} < T_{e1}$. Consequently, the hot-carrier-induced current, after being normalized to the total drain current, I_h/I_d, should correspond to the order in the probabilistic area such that $I_{h3} < I_{h1} < I_{h2}$.

In a MOSFET, both substrate and gate currents can be used to trace out the hot-

electron "cooling" effect, if it indeed exists and is strong enough to cause normalized substrate and gate current reduction with decreasing channel length. Traditionally, the substrate current was used to investigate the hot-electron "cooling" effect, because it is much easier to observe than the gate current, and its characteristics are well understood. The experimental results in the literature have been somewhat contradictory on this point, with some reports confirming and some reports rejecting the existence of the hot-carrier "cooling" effect in deep-submicron MOSFETs. Shahidi *et al.* first reported substrate current reduction at both 300K and 77K for NMOSFETs with L_{eff} below 0.15 μm [54]. Since the NMOSFETs used in this report for the substrate current measurements show severe device punchthrough, the implication of this on the interpretation of the experimental data remains to be justified on a firmer ground. Dutoit *et al.* later reported more direct experimental evidence of the reduction in electron temperature with decreasing channel length [17]. Their study was made by observing the increasing relative SdH oscillation amplitude, a monotonically decreasing function of electron temperature, with decreasing channel length at a given lateral electric field value. However, Mizuno *et al.* reported experimental evidence for continuously increasing substrate current with decreasing channel length down to 0.12 μm [36]. The NMOSFETs used in that study for substrate measurements are not short enough to convincingly reject the hot-electron "cooling" hypothesis. Generally speaking, the experimental data regarding the hot-electron "cooling" effect are difficult to interpret due to the fact that the reduction in substrate current, if there is any, results from some "appropriate" combinations of both electron temperature and electron group velocity values, as can be seen in the simple displaced Gaussian model shown in Fig. 5-7. What matters is whether there is less or more hot-carrier-induced current generation with decreasing channel length.

In this thesis work, great effort was made to fabricate sub-0.1 μm MOSFETs with as minimal device punchthrough and parasitic resistance as possible, so that the measurements of hot-carrier-induced currents can be as unambiguous as possible, and the experimental data can be easily interpreted. In addition, extensive gate current characterization was conducted along with that of the substrate current to provide

additional experimental evidence, and to minimize potential experimental errors in tracing the hot-carrier "cooling" effect.

5.3.2 Experimental observations at room temperature

According to the theories of substrate and gate current generation in the previous section, the logarithm of the normalized hot-carrier current, $\log(I_b/I_d)$, or $\log(I_g/I_d)$, *vs.* $1/(V_{ds} - V_{dsat})$ relationship is linear at a given V_{gs} (Eq. 5.9 and 5.15),

$$\log(\frac{I_b}{I_d}) = -l\ \frac{\Phi_i}{q\lambda}\frac{1}{V_{ds} - V_{dsat}} + \log(A_i\frac{q\lambda}{\Phi_i}) \qquad (5.32)$$

and

$$\log(\frac{I_g}{I_d}) = -l\ \frac{\Phi_b}{q\lambda}\frac{1}{V_{ds} - V_{dsat}} + \log(\frac{C}{l}(\frac{q\lambda}{\Phi_b})^2). \qquad (5.33)$$

If the hot-carrier "cooling" effect takes place, one should observe a reduction in the normalized gate or substrate current at a given $1/(V_{ds} - V_{dsat})$ with decreasing L_{eff}, since a constant $1/(V_{ds} - V_{dsat})$ ensures a constant peak lateral electric field, E_m. However, the actual observations were made at a given $1/V_{ds}$ rather than a given $1/(V_{ds} - V_{dsat})$ for the following reasons. (1) There is currently no way of accurately extracting the saturation drain voltage, V_{dsat}, in the deep-submicron regime where V_{dsat} is not well defined in the MOSFET $I - V$ characteristics. The commonly used method described in [7, 31] is in principle incorrect; (2) In a practical ULSI system, all transistors with various L_{eff}s are presumably biased at the same drain voltage, V_{ds}, rather than having constant $1/(V_{ds} - V_{dsat})$, and thus the reduction in the hot-carrier-induced currents only makes sense at a given V_{ds} rather than a given $1/(V_{ds} - V_{dsat})$.

Gate and substrate currents were measured on the SSR NMOSFETs described in Chapter 2 with channel width, $W = 49.4\ \mu m$ and various L_{eff}. Figs. 5-8, 5-10, 5-12, and 5-14 show the normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with gate voltage, $V_{gs} = 2\ V$ and $3\ V$, for SSR-I, SSR-II, SSR-III and STEP doped devices, respectively. Figs. 5-9, 5-11, 5-13, and 5-15 show the corresponding normalized substrate current, I_b/I_d, as a function of reciprocal drain

voltage $1/V_{ds}$ with the same gate voltages.

Experimental data for all SSR and STEP doped NMOSFETs are shown and analyzed because of an earlier presumption that the hot-carrier "cooling" effect might depend on the particular vertical electric field configuration, which is directly dependent on the channel doping profile. As the figures clearly show, there is no sign of impact ionization reduction due to hot-electron "cooling" at any V_{ds} in saturation, as the normalized substrate current, I_b/I_d, continuously increases with decreasing L_{eff} down to 0.1 μm regime in all cases at room temperature. And the same conclusion can be drawn from the gate current data that there is no sign of hot-electron injection reduction due to hot-electron "cooling" at any V_{ds} in saturation, as the normalized gate current, I_g/I_d, continuously increases with decreasing L_{eff} in all cases at room temperature. As a matter of fact, as shown in Figs. 5-19 and 5-20 in the next section, both I_b/I_d and I_g/I_d increase exponentially with decreasing L_{eff}.

It is interesting to note that the gate current increases much more dramatically than the substrate current with decreasing L_{eff}. This provides direct supporting evidence for the argument presented in the earlier section that in the deep-submicron regime, the driving factors for the two primary hot-carrier effects, impact ionization and hot-carrier injection, can no longer be considered in the same way, as assumed by the conventional lucky-electron model. The increasing fraction of the channel hot electrons injected into the gate oxide, due to lateral electric field broadening and perhaps other dynamic effects that are quasi-ballistic, enhances the gate current I_g/I_d dependence on the effect channel length, L_{eff}, while there is no analogy taking place in the impact ionization process so as to enhance the substrate current. In short, as the decoupling between the channel hot-electron injection and impact ionization increases with decreasing L_{eff} in the deep-submicron regime to the extent that the correlation between I_b/I_d and I_g/I_d no longer holds, the lucky-electron model needs to be revised to treat the two hot-carrier effects separately.

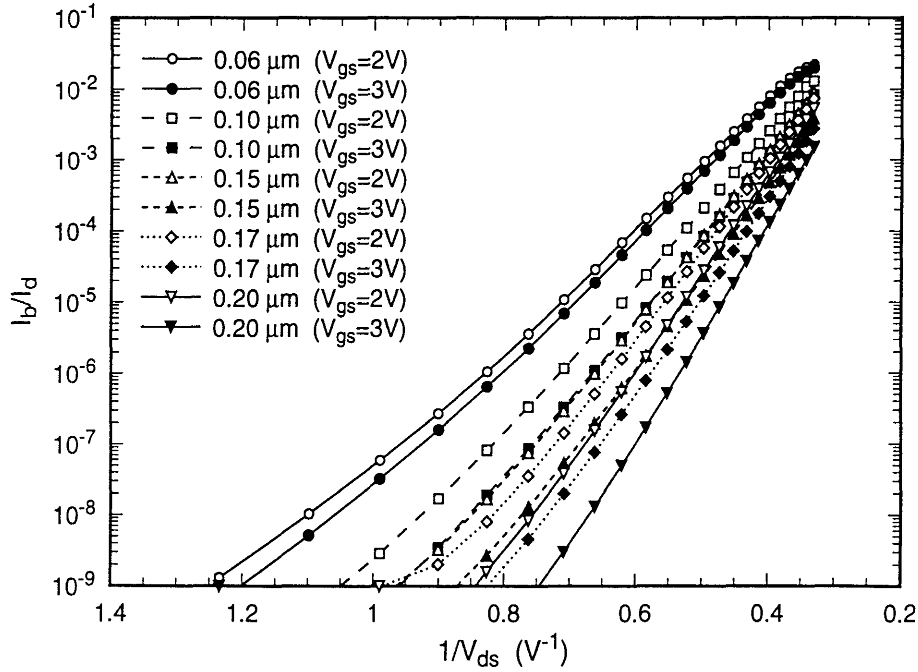

Figure 5-8: Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}s, and SSR-I channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2.

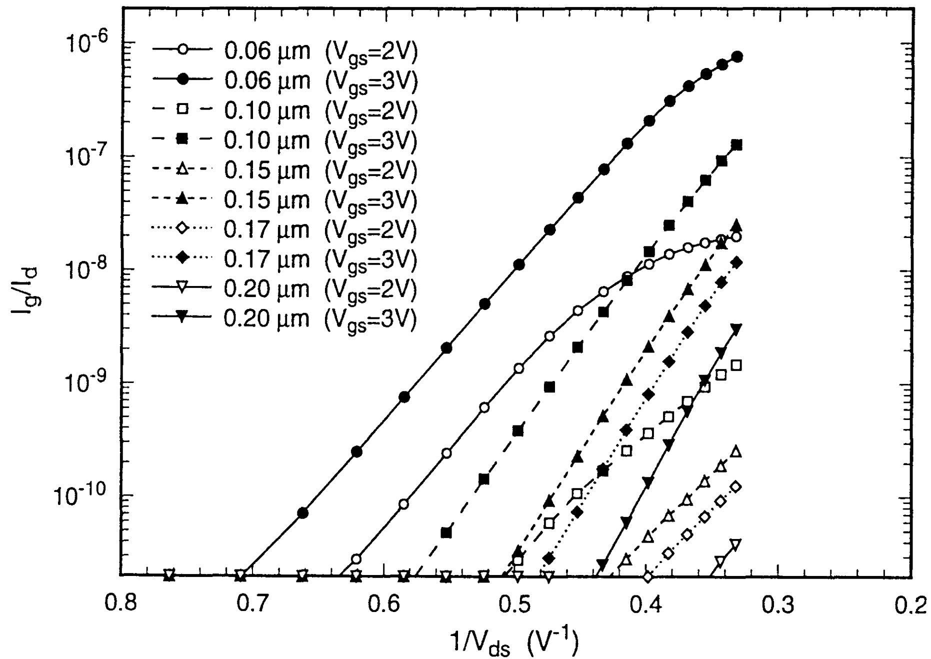

Figure 5-9: Measured normalized gate current, I_g/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-I channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2.

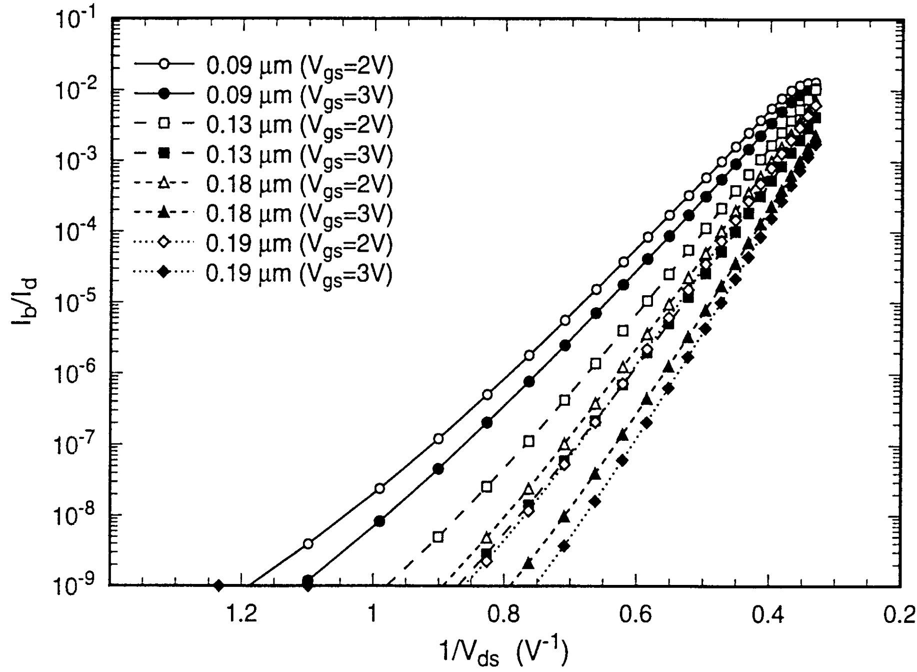

Figure 5-10: Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-II channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2.

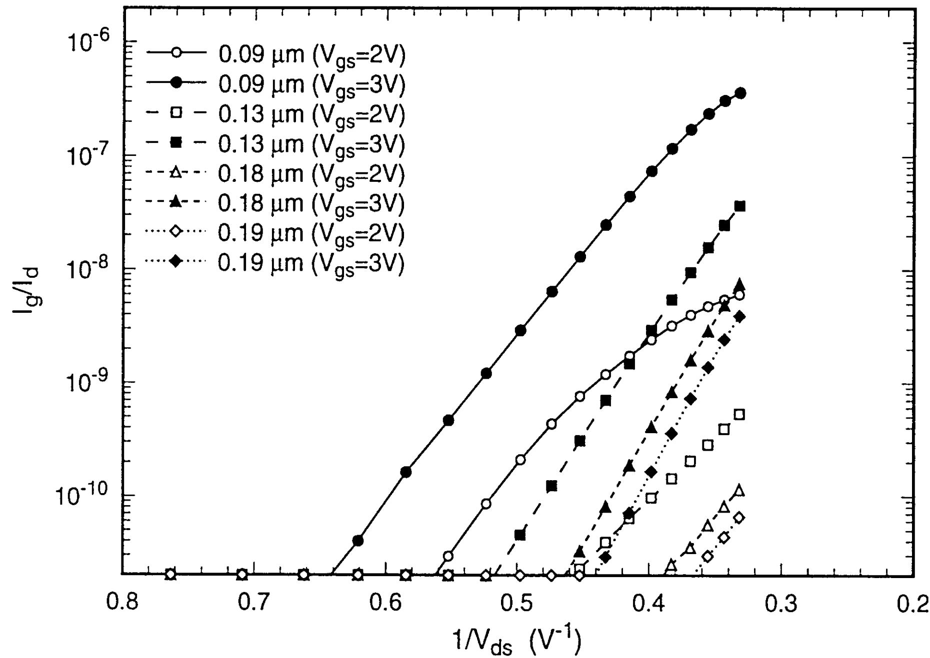

Figure 5-11: Measured normalized gate current, I_g/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-II channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2.

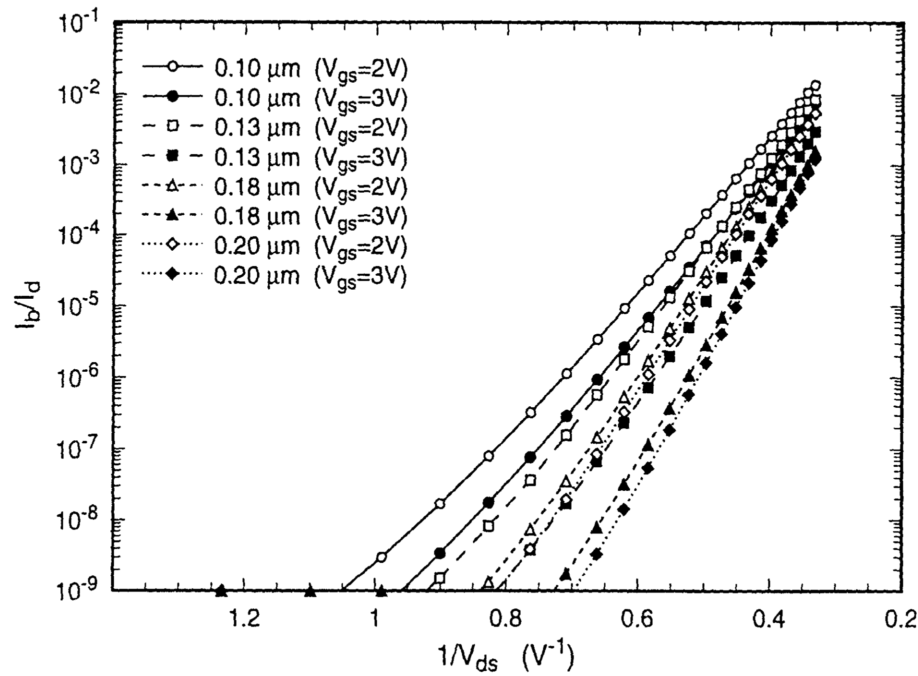

Figure 5-12: Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-III channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2.

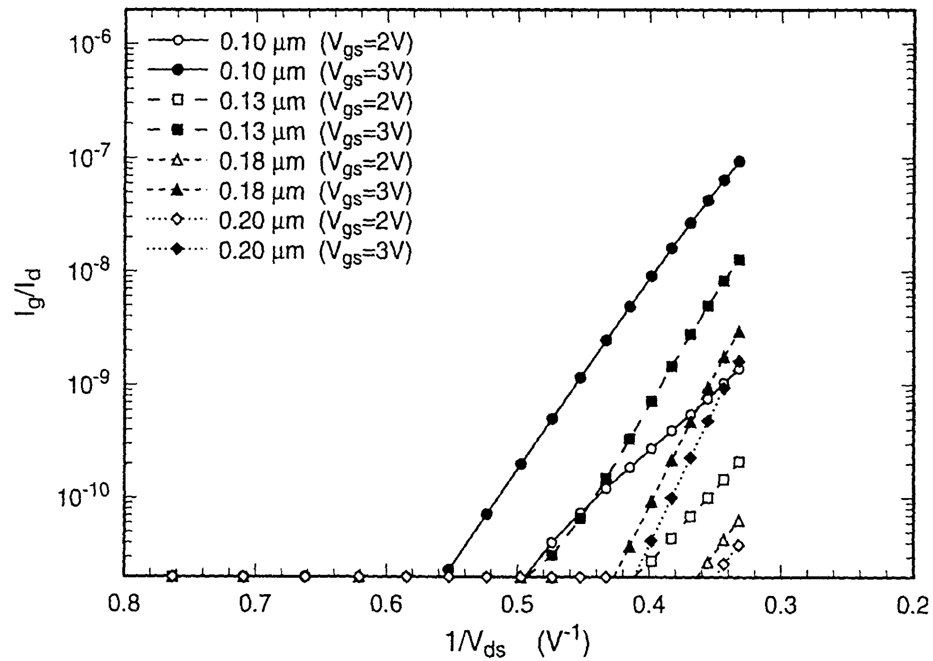

Figure 5-13: Measured normalized gate current, I_g/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and SSR-III channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2.

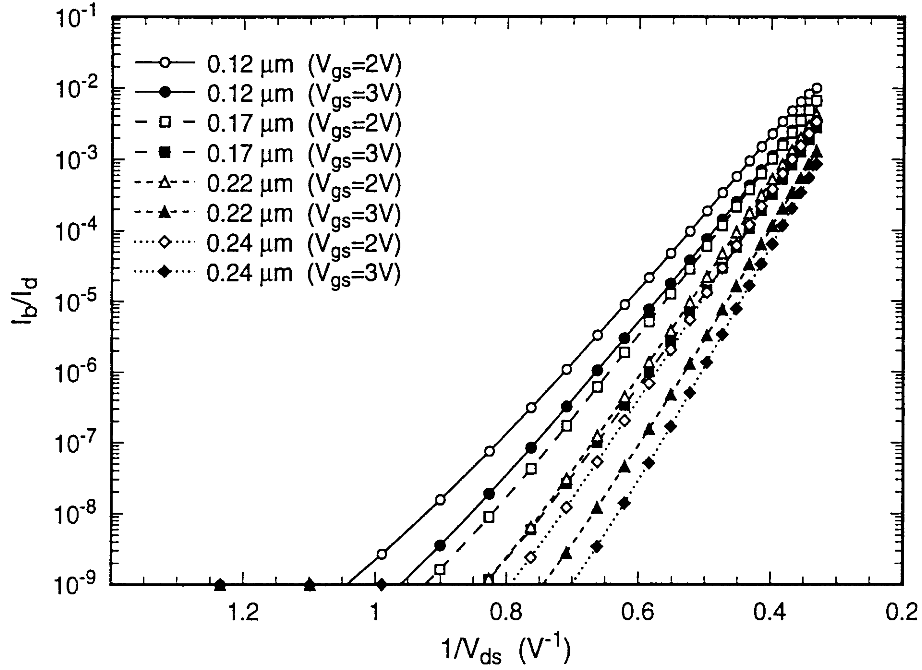

Figure 5-14: Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and STEP channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2.

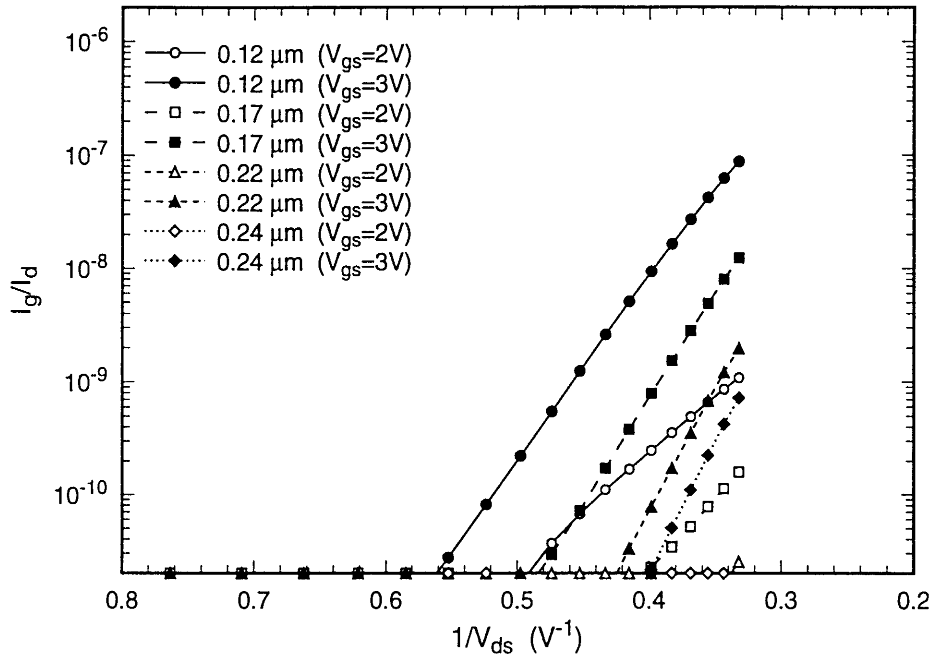

Figure 5-15: Measured normalized gate current, I_g/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 2\ V$ and $3\ V$ for NMOSFETs with various L_{eff}, and STEP channel doping described in Fig. 2-1 and Table. 2.1 in Chapter 2.

5.3.3 Experimental observations at low temperature

The substrate current was measured both at 77 K and 300 K on NMOSFETs fabricated by AT&T Bell Labs [33]. Figs. 5-16 and 5-17 show the measured I_b/I_d as a function of $1/V_{ds}$ with $V_{gs} = 1\ V$ at 77 K and 300 K, respectively, for NMOSFETs with various L_{eff}s ranging from 0.45 μm down to 0.11 μm. The actual values of the extracted L_{eff} may not be accurate, but their order is guaranteed by the order of their corresponding linear-transconductance values. In both cases, the I_b/I_d value for the $L_{eff} = 0.11\ \mu m$ device is slightly smaller than that for the $L_{eff} = 0.12\ \mu m$ device for a given V_{ds}, and the amount of reduction in I_b/I_d is nearly identical. This is rather peculiar for the following reasons. According to the non-equilibrium transport theory presented in Chapter 3, the low-field electron mobility, μ_0, should have a significant effect on the dynamics of the non-equilibrium transport. The higher the μ_0, the more pronounced is the quasi-ballistic nature of the hot-electron transport in the inversion layer of a MOSFET. For these silicon MOSFET devices with channel doping $N_a = 1.3 \times 10^{17}\ cm^{-3}$, the low-field mobility at 77 K, $\mu_0 \approx 3000\ cm^2/V-sec$, is at least five times greater than that at 300 K, $\mu_0 \approx 550\ cm^2/V - sec$, according to both the experimental and simulation data shown in [27]. Thus, if hot-electron "cooling" does exist at room temperature, it should be much more pronounced at 77 K due to the enhancement of quasi-ballistic transport. That is, there should be more significant reduction in I_b/I_d at 77 K than at 300 K, which is not the case with the experimental data shown in Fig. 5-16 and 5-17. The slight reduction in I_b/I_d is likely due to the fact that the shortest $L_{eff} = 0.11\ \mu m$ device has a disproportionate amount of deep punchthrough at high drain voltage, which enhances the drain current, I_d, but not the substrate current, I_b, because the deep-punchthrough electron flow is not confined to the particular region of the channel where the lateral electric field is high and impact ionization takes place.

Similar measurements and analysis were done on deep-submicron p-channel MOSFETs described in [24]. Fig. 5-18 shows the measured I_b/I_d as a function of $1/V_{ds}$ with $V_{gs} = -1\ V$ for these PMOSFETs. Again, there is no sign of a hot-hole "cooling" effect, since I_b/I_d keeps increasing as L_{eff} decreases down to 0.1 μm at both

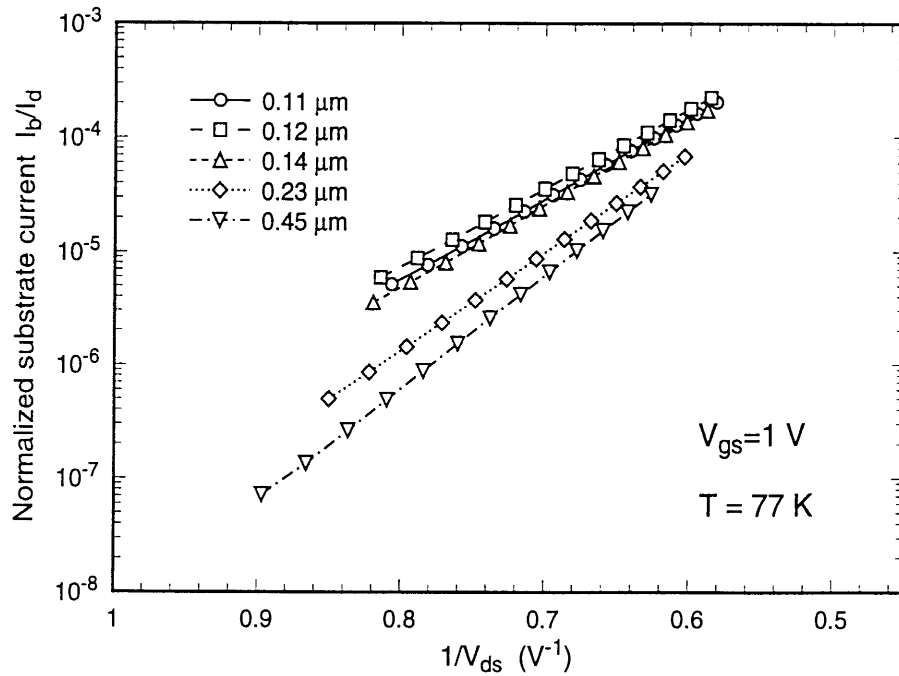

Figure 5-16: Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 1\ V$ for NMOSFETs with various L_{eff} at 77 K. Courtesy of AT&T Bell Labs for device fabrication [33].

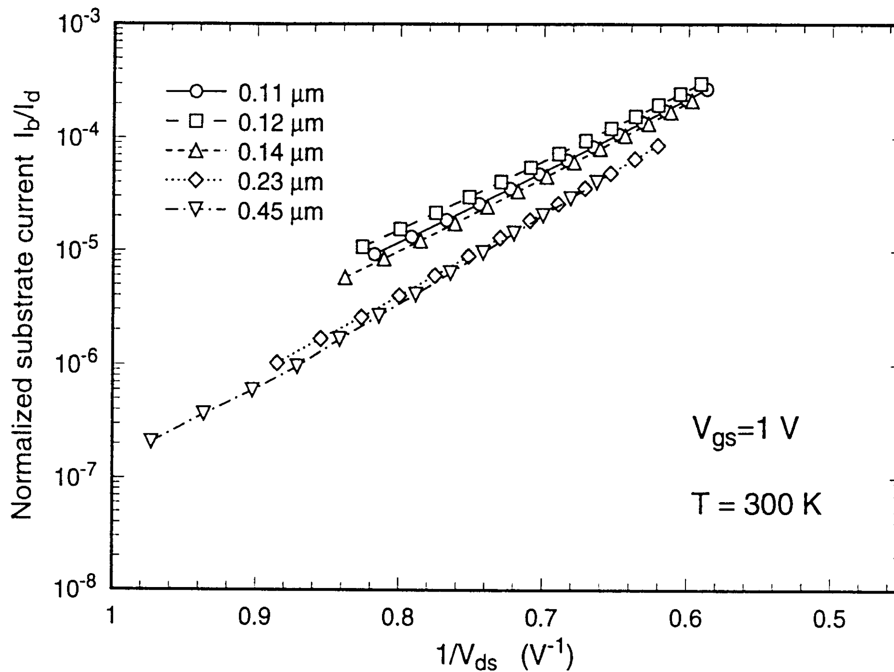

Figure 5-17: Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = 1\ V$ for NMOSFETs with various L_{eff} at 300 K. Courtesy of AT&T Bell Labs for device fabrication [33].

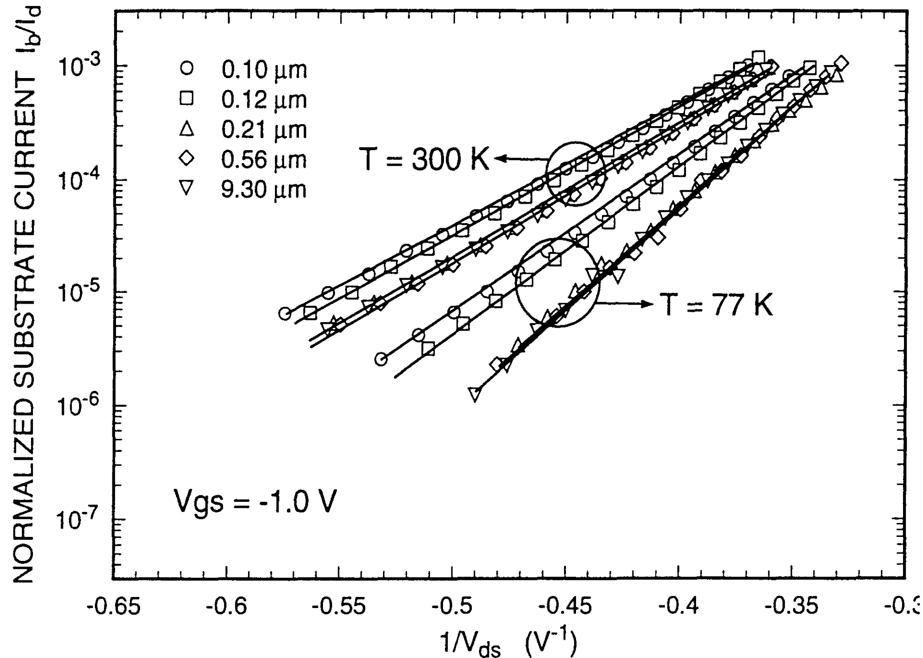

Figure 5-18: Measured normalized substrate current, I_b/I_d, as a function of reciprocal drain voltage $1/V_{ds}$ with $V_{gs} = -1\ V$ for PMOSFETs with various L_{eff}at 77 K and 300 K. The PMOSFETs used here are described in [24].

temperatures.

5.4 Hot-Carrier Effects in Scope of MOSFET Scaling

5.4.1 Hot-carrier effect: a fourth dimension to MOSFET scaling

Chapter 4 discusses MOSFET scaling among the three fundamental quantities: effective channel length, L_{eff}, drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, and device speed or electron velocity, g_m/WC_{ox}, in the deep-submicron regime with an unfinished note regarding the fourth fundamental quantity of a MOSFET, hot-carrier-induced device degradation. As mentioned in Chapter 4, the scaling criterion due to hot-carrier-induced device degradation, together with the dynamic performance criterion,

g_m/WC_{ox}, have not been investigated to any quantitative extent in the literature. Chapter 4 links L_{eff}, $\frac{\delta V_t}{\delta V_{ds}}$, and g_m/WC_{ox} together in a unified MOSFET scaling framework, and the study in this section attempts to add the fourth dimension to this unified framework, namely the hot-carrier effect. Strictly speaking, the appropriate quantity that should be introduced to this fourth dimension is the stress characteristics under hot-carrier-induced degradation, such as the shift in V_t, or in linear transconductance, g_{ml}, with time. However, it is reasonable to use the hot-carrier-induced currents, I_b and I_g, to represent the device stress characteristics, because they correlate remarkably well, as proven by numerous studies [18].

5.4.2 Methodology and analysis

In this section, the same methodology introduced in Chapter 4 is followed to deduce the scaling relationship between I_b/I_d, I_g/I_d, and the effective channel length, L_{eff}. Then the scaling relationships among I_b/I_d, I_g/I_d, $\frac{\delta V_t}{\delta V_{ds}}$, and g_m/WC_{ox} can be deduced from their dependence on L_{eff}. The dependence of those relationships on device parameters is discussed with the emphasis on the channel parameter, V_t and the channel doping profile.

The scaling relationship between I_b/I_d, I_g/I_d, and L_{eff} is again assumed to be of the following power-law form:

$$B_i(V_{ds}, V_{gs}) = I_b/I_d = B_0(V_{ds}, V_{gs})\ L_{eff}{}^{-\gamma_b(V_{ds}, V_{gs})} \tag{5.34}$$

and

$$G_i(V_{ds}, V_{gs}) = I_g/I_d = G_0(V_{ds}, V_{gs})\ L_{eff}{}^{-\gamma_g(V_{ds}, V_{gs})} \tag{5.35}$$

where $\gamma_b(V_{ds}, V_{gs}) > 0$ and $\gamma_g(V_{ds}, V_{gs}) > 0$ are constants, possibly dependent on channel parameters and source/drain parameters, as defined in Chapter 4. The power-law hypothesis in Eqs. 5.34 and 5.35 is tested by nonlinear regression on the data samples in the same form as that in Eq. 4.6 of Chapter 4, and the statistical significance is estimated by the same confidence factor defined in Eq. 4.7 of Chapter 4.

Fig. 5-19 shows the experimental relationship between I_b/I_d and L_{eff} for four

(V_{ds}, V_{gs}) (V)	(1.5, 2.0)	(3.0, 2.0)	(1.5, 3.0)	(3.0, 3.0)
SSR-I	34.5	9.4	47.7	17.3
SSR-II	34.8	9.2	49.9	16.2
SSR-III	35.2	9.0	50.2	17.4
STEP	34.8	9.1	46.5	16.7
Average	34.8	9.2	48.6	16.9
Std. Dev (%)	0.8%	1.9%	3.7%	3.3%

Table 5.1: $\gamma_b(V_{ds}, V_{gs})$ coefficient matrix according to Fig. 5-19 with index (V_{ds}, V_{gs}), as defined in Eq. 5.34, and channel doping profile, as shown in Fig. 2-1.

sets of V_{gs} and V_{ds} pairs. The average r factor for all four sets is greater than 0.98 in the deep-submicron range of interest, $L_{eff} < 0.25\ \mu m$, which indicates that the power-form hypothesis in Eqs. 5.34 is statistically significant.

Fig. 5-20 shows the experimental relationship between I_g/I_d and L_{eff} for two sets of V_{gs} and V_{ds} pairs with $V_{ds} = 3\ V$. The average r factor for the two sets is greater than 0.99 in the deep-submicron range of interest, $L_{eff} < 0.25\ \mu m$, which indicates that the power-form hypothesis in Eqs. 5.35 is also statistically significant.

Table 5.1 shows the matrix of $\gamma_b(V_{ds}, V_{gs})$ coefficients with index (V_{ds}, V_{gs}) and channel doping profiles (SSR-I, II, III, STEP). Table 5.2 shows the matrix of $\gamma_g(V_{ds}, V_{gs})$ coefficients with the same indices. As the statistics clearly shows, both $\gamma_b(V_{ds}, V_{gs})$ and $\gamma_g(V_{ds}, V_{gs})$ power coefficients, as defined in Eqs. 5.34 and 5.35, are insensitive to the channel doping profile and threshold voltage, V_t, within the experimental range indicated in Table 2.1. The standard deviations for all the matrix columns are less than 4%.

Nonetheless, as shown in Figs. 5-19 and 5-20, the magnitude of either I_b/I_d or I_g/I_d at a given L_{eff} is clearly dependent on the channel doping profiles and V_ts, which is reflected in the distinct B_0 and G_0 coefficients, as defined in Eqs. 5.34 and 5.35. Overall speaking, the characteristics of I_b/I_d *vs.* L_{eff} are similar to that of I_g/I_d *vs.* L_{eff} with respect to the channel doping profiles and V_t.

It is interesting to note that the trend of B_0 and G_0 dependencies on V_t and channel doping follows exactly that of the Λ coefficient, as defined in Eq. 4.4 in

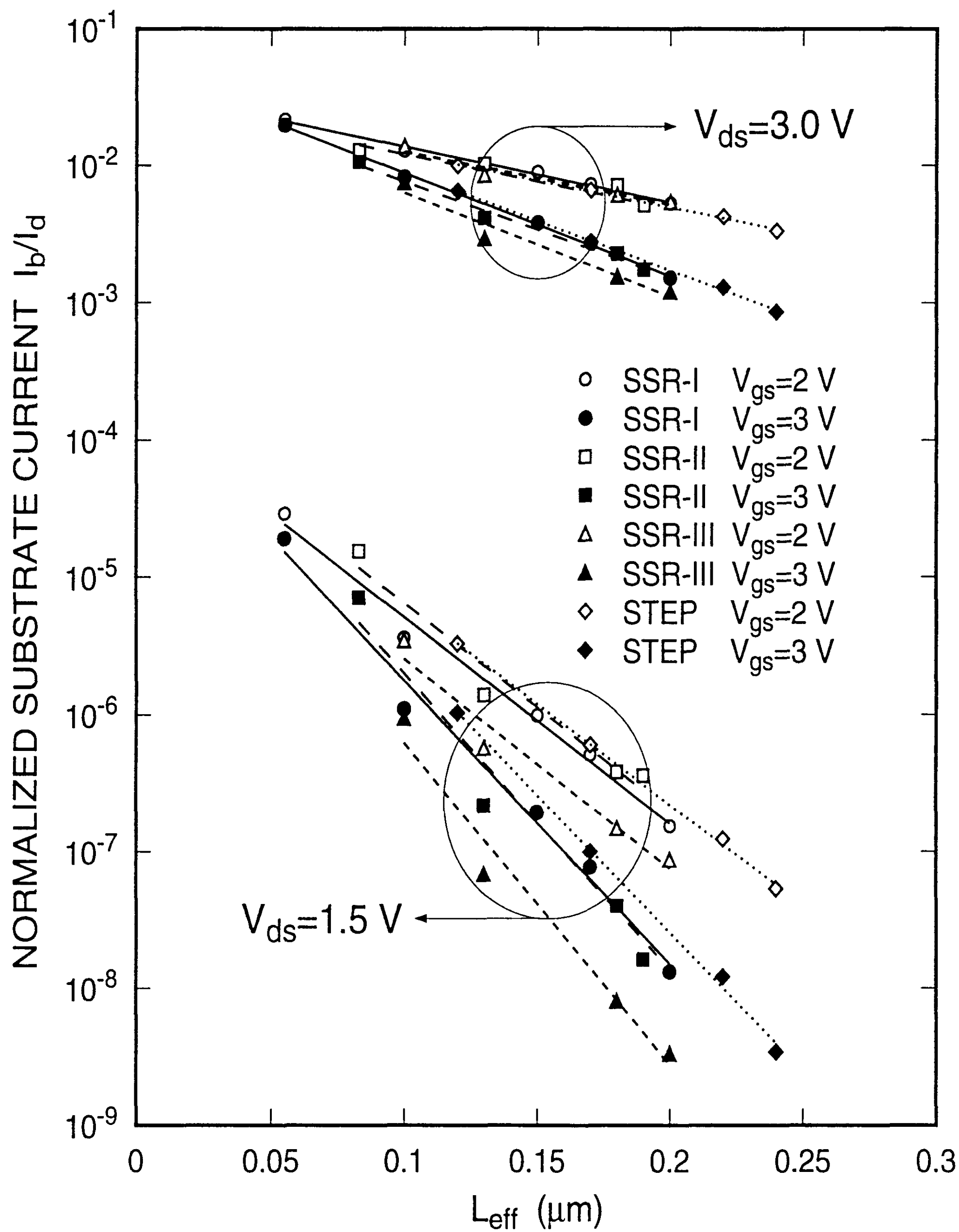

Figure 5-19: Measured normalized substrate current, I_b/I_d, as a function of effective channel length, L_{eff}, at a given drain voltage, $V_{ds} = 1.5\ V$ and $3\ V$, for NMOSFETs with the three SSR and STEP channel dopings described in Fig. 2-1 and Table. 2.1 in Chapter 2. $V_{gs} = 2\ V$ and $3\ V$.

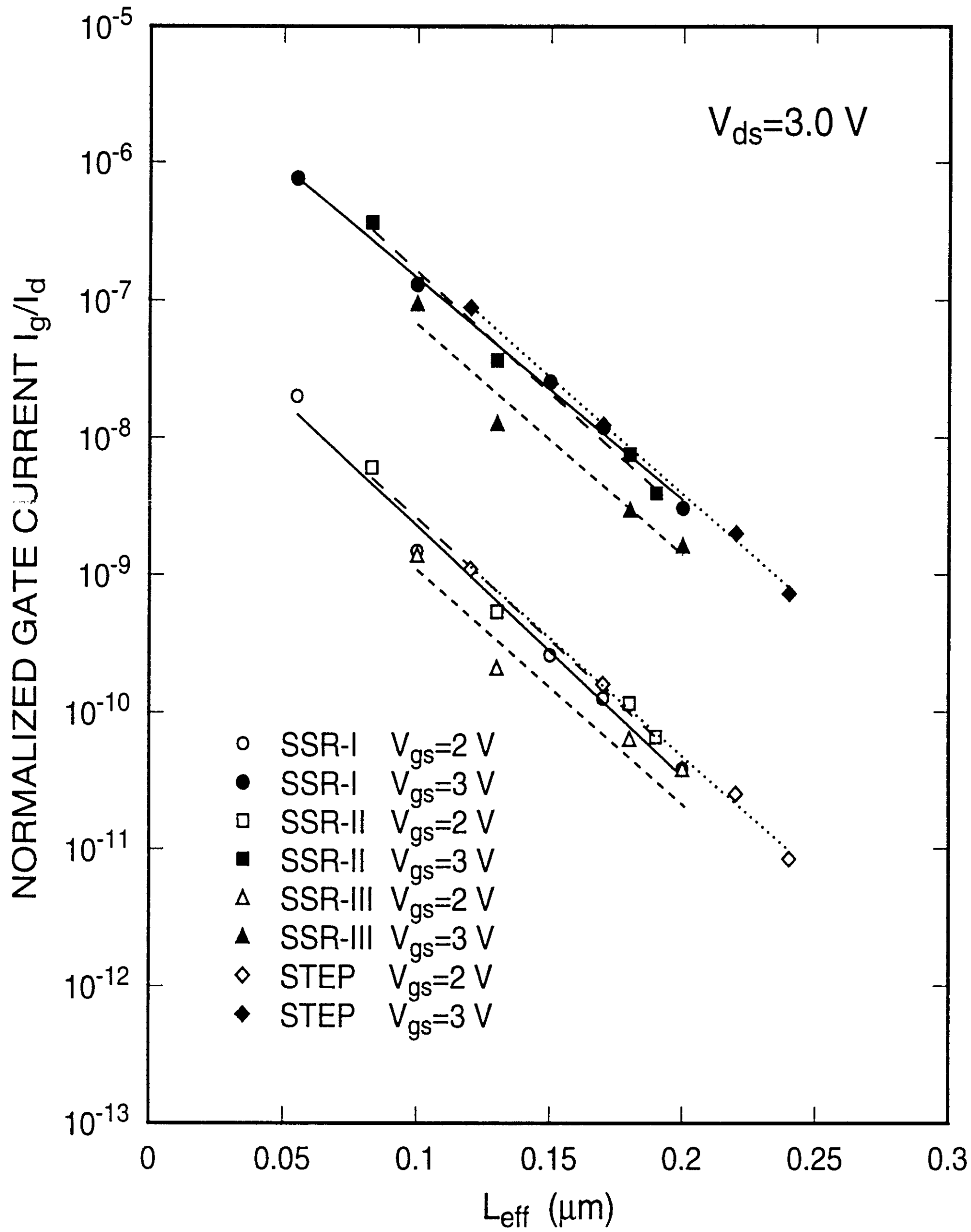

Figure 5-20: Measured normalized gate current, I_g/I_d, as a function of effective channel length, L_{eff}, at a given drain voltage, $V_{ds} = 3\ V$, for NMOSFETs with the three SSR and STEP channel dopings described in Fig. 2-1 and Table. 2.1 in Chapter 2. $V_{gs} = 2\ V$ and $3\ V$.

(V_{ds}, V_{gs}) (V)	(3.0, 2.0)	(3.0, 3.0)
SSR-I	41.9	37.2
SSR-II	40.7	40.6
SSR-III	39.3	38.8
STEP	39.7	39.3
Average (%)	40.4	39.0
Std. Dev (%)	2.9%	3.6%

Table 5.2: $\gamma_g(V_{ds}, V_{gs})$ coefficient matrix according to Fig. 5-20 with index (V_{ds}, V_{gs}), as defined in Eq. 5.35, and channel doping profile, as shown in Fig. 2-1.

Chapter 4. That is, the higher the long channel transconductance, g_m, the higher the I_b/I_d or I_g/I_d ratios. Also, the trend follows exactly that of the inverse $1/\Theta$ coefficient, as defined in Eq. 4.3 in Chapter 4. That is, the better (smaller) the drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}}$, the lower the I_b/I_d or I_g/I_d ratios. As discussed in Chapter 4, it is the trade-off between the two counter balanced trends of the g_m/WC_{ox} *vs.* L_{eff} dependence and the $\frac{\delta V_t}{\delta V_{ds}}$ *vs.* L_{eff} dependence on the channel parameters that enforces the "universality" in the g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$ relationship, as shown in Fig. 4-6 in Chapter 4. Thus, it is a logical extension of the universality between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$ to assume also the universality between I_b/I_d or I_g/I_d and $\frac{\delta V_t}{\delta V_{ds}}$ and between I_b/I_d or I_g/I_d and g_m/WC_{ox}, provided that the B_0 and G_0 coefficients are properly normalized.

This universality can be derived in a similar fashion to that in Chapter 4 as follows. By combining Eq. 4.3 and Eq. 5.34 or 5.35, one can hypothesize the following relation:

$$I_h/I_d = \Delta \; (\frac{\delta V_t}{\delta V_{ds}})^{\delta}, \tag{5.36}$$

and by combining Eq. 4.4 and Eq. 5.34 or 5.35, one can hypothesize the following relation:

$$g_m/WC_{ox} = \Xi \; (I_h/I_d)^{\xi} \tag{5.37}$$

where Δ, Ξ and $\delta > 0$, $\xi > 0$ are constants (presumably universal) with respect to V_t and the channel doping profile, but dependent on V_{gs} and V_{ds}. I_h/I_d is either I_b/I_d

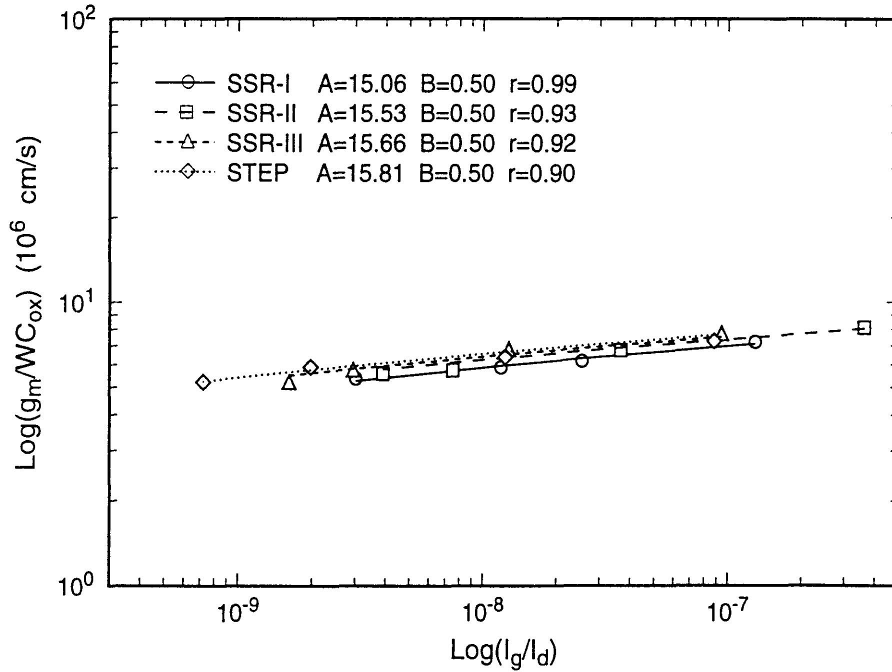

Figure 5-21: Measured normalized gate current, I_g/I_d, as a function of device speed, g_m/WC_{ox}, for NMOSFETs with the three SSR and STEP channel dopings described in Fig. 2-1 and Table. 2.1 in Chapter 2.

or I_g/I_d. Since it is shown in Fig. 5-19 and Fig. 5-20 that the dependencies of I_b/I_d *vs.* L_{eff} and I_g/I_d *vs.* L_{eff} are the same with respective to V_t and channel doping profile, the derivations for both should be the same.

As is indeed shown in Fig. 5-21, the relationship between g_m/WC_{ox} and I_h/I_d is universal or invariant with respect to V_t and channel doping profile. Not only the power coefficient, $B = \xi = 0.50$, remains a constant, but also the proportionality coefficient, $A = \log(\Xi) = 15.06 \sim 15.81$, remains rather invariant (the standard deviation = 2.02 %). The hypothesis in Eq. 5.37 is indeed statistically significant with the confidence factors, r, greater than 0.90 in all cases.

A similar derivation can be given to show that a universal relationship exists between I_h/I_d and $\frac{\delta V_t}{\delta V_{ds}}$, as shown in Eq. 5.36, by simply combining Eqs. 5.37 and 4.5 in Chapter 4.

5.5 Conclusion

The correlation between gate and substrate currents in NMOSFETs with effective channel lengths, L_{eff}, down to 0.1 μm is investigated within the general framework of the lucky-electron model. Experimental data suggest that the correlation coefficient, Φ_b/Φ_i, decreases with decreasing L_{eff} in the 0.1 μm regime. This hot-electron injection barrier lowering effect is confirmed by numerical simulations which incorporate the non-equilibrium dynamical effects associated with hot-electron transport in deep-submicron MOSFETs into the gate current evaluation. The observation made in comparing the much more rapid increase in I_g/I_d with respect to L_{eff} with the slower increase in I_b/I_d with respect to L_{eff}, as shown in Figs. 5-20 and 5-19, demonstrates the increasing decoupling between gate current generation and substrate current generation, or, between the channel hot-electron injection and the impact ionization, with decreasing L_{eff} in the deep-submicron regime. This new experimental evidence suggests the need of using gate current as an indicator to understand deep-submicron MOSFET degradation mechanisms, rather than using substrate current alone.

The anonymous hot-carrier "cooling" effect is investigated at both room temperature and liquid-nitrogen temperature. The reduction in the normalized substrate current, I_b/I_d, with decreasing effective channel length, L_{eff}, is not observed from the experimental data measured on the SSR n-channel MOSFET devices with L_{eff} down to sub-0.1 μm at room temperature. The slight reduction in I_b/I_d observed at both 300 K and 77 K from the AT&T Bell Labs MOSFETs is attributed to the device punchthrough rather than the hot-electron "cooling" effect. The same conclusion is drawn for the case of p-channel MOSFETs regarding the hot-hole "cooling" effect. Also, as a complementary test of the hot-electron "cooling" effect, the normalized gate current, I_g/I_d, is characterized with decreasing L_{eff}, and there is no indication of gate current reduction with L_{eff} down to sub-0.1 μm regime either. Thus, for well-behaved silicon MOSFETs, there is no convincing evidence of hot-carrier-induced current reduction as a manifestation of the hot-carrier "cooling" effect in the deep-submicron regime.

It is demonstrated that there exist universal trade-off relationships among the device performance, represented by g_m/WC_{ox}, the device short-channel effect, represented by $\frac{\delta V_t}{\delta V_{ds}}$, and the device hot carrier currents, represented by I_b/I_d or I_g/I_d, with respect to the channel parameters, the threshold voltage, V_t, and the channel doping profile. These universal relationships can be put into power-law forms, as defined in Eqs. 4.5, 5.37, and 5.36, with excellent statistical significance. With this scaling methodology, all fundamental quantities of MOSFET scaling are unified in the deep-submicron regime with L_{eff} down to sub-0.1 μm. The future direction on deep-submicron MOSFET scaling along those lines is also clear. It is important to investigate the source/drain parameter dependence of the universal relationships among all fundamental MOSFET scaling quantities. It is expected again that it is the source/drain parameters that determine the trade-offs among g_m/WC_{ox}, $\frac{\delta V_t}{\delta V_{ds}}$, and I_h/I_d.

Chapter 6

Conclusion

In this thesis work, physical effects associated with the non-equilibrium dynamics of the electron transport in the inversion layers of deep-submicron silicon MOSFET devices are investigated. These effects are the electron velocity overshoot, the hot-electron injection barrier lowering, and the hot-carrier "cooling" effect. On the theoretical side, these macroscopic effects reveal the microscopic mechanisms behind the unique transport dynamics of high-energy electrons and holes under non-equilibrium conditions such as high electric fields and electric-field gradients in deep-submicron MOSFETs. On the practical side, these effects have profound impact on one of the most important issues regarding the evaluation of silicon MOSFET devices and thus the ULSI industry, the deep-submicron MOSFET scaling. By pursuing the understanding of these physical effects from the perspective of MOSFET scaling, one can take a different view of the MOSFET scaling issues from the traditional one based on the classical scaling theory, and thus follow a new methodology that is theoretically plausible and practically efficient to unify all fundamental quantities in deep-submicron MOSFET scaling and providing the insight of deep-submicron MOSFET design.

Extensive effort is made to design and fabricate well-behaved silicon MOSFETs with effective channel lengths down to sub-0.1 μm in order to provide accurate experimental data for investigating the physical effects mentioned above. High performance sub-0.1 μm MOSFET devices (SSR MOSFETs) using X-ray lithography, self-aligned

$CoSi_2$ silicide formed by Ti/Co laminates, super-steep retrograde channel doping, and ultra-shallow source/drain extension structure with "halo" doping are demonstrated in this thesis work. These SSR n-channel MOSFETs exhibit very high saturation current drive and transconductance with minimal short-channel effects. They have achieved the best-to-date performance with a given amount of short-channel effect. X-ray lithography is proven to be a highly promising lithography technology for deep-submicron MOSFET fabrication. The ultra-shallow source/drain extension structure coupled with Ti/Co bimetallic $CoSi_2$ silicide used in this thesis work is demonstrated to be highly effective in controlling short-channel effects and minimizing parasitic resistance. Super-steep retrograde channel doping is shown to be highly effective in preventing device punchthrough, while maintaining the device electrostatic integrity. The excellent overall behavior of these SSR MOSFETs, i.e, high performance, well-controlled short-channel effects, and minimal leakage currents, is essential for unambiguous device measurements that ensure accurate experimental data for the investigation of deep-submicron MOSFET physics.

The electron velocity overshoot phenomenon in silicon inversion layers is experimentally investigated using high performance SSR n-channel MOSFETs with effective channel lengths down to sub-0.1 μm. It is found that the average electron velocity is not yet in the overshoot regime even for the best performing SSR MOSFET devices. From the perspective of deep-submicron MOSFET scaling, there exists a trade-off between the electron velocity and the device short-channel effects, such as the drain-induced barrier lowering effect and the punchthrough effect. The higher the electron velocity, the more pronounced is the short-channel effects, and the higher is the rate at which the short-channel effects increase with decreasing device effective channel lengths or increasing electron velocities. For the SSR MOSFET devices with an acceptable amount of drain-induced barrier lowering, $\frac{\delta V_t}{\delta V_{ds}} = 100\ mV/V$, to break the barrier of the electron saturation velocity at room temperature, $v_{sat} = 1.0 \times 10^7\ cm/s$, the low-field electron mobility has to increase by 43%. This suggests the use of low-temperature (e.g., liquid-nitrogen temperature) silicon MOSFETs or compound semiconductor structures such as $Si - Ge$ FETs to push the frontier of the trade-off

constraint between the electron velocity and the short-channel effects.

The correlation between gate and substrate currents in n-channel MOSFETs with effective channel lengths down to 0.1 μm is investigated within the general framework of the lucky-electron model. It is found for the first time that the correlation coefficient, Φ_b/Φ_i, decreases with decreasing L_{eff} in the 0.1 μm regime. This hot-electron injection barrier lowering effect is confirmed by numerical simulations which incorporate the non-equilibrium dynamical effects associated with hot-electron transport in deep-submicron MOSFETs into the gate current evaluation. By comparing the much more rapid increase in I_g/I_d with respect to L_{eff} with the slower increase in I_b/I_d with respect to L_{eff}, it is demonstrated that the decoupling between channel hot-electron injection and impact ionization increases rapidly with decreasing L_{eff} in the deep-submicron regime. This new experimental evidence suggests the need of using gate current as an indicator to investigate deep-submicron MOSFET degradation mechanisms, rather than using substrate current alone. Monte-Carlo simulations can greatly benefit the further investigation of this new-found hot-electron injection barrier lowering effect by providing the theoretical confirmation on the energy distribution and the lateral electric field distribution and thus predicting other possible effects as a result of the injection barrier lowering in the deep-submicron regime. It is also of great interest to investigate this effect at low temperatures, such as 77 K and 4.2 K, as this effect is expected to be more pronounced if it is indeed a consequence of hot-electron non-equilibrium transport dynamics.

The anonymous hot-carrier "cooling" effect is investigated at both room temperature and liquid-nitrogen temperature. The reduction in the normalized substrate current, I_b/I_d, with decreasing effective channel length, L_{eff}, is not observed from the experimental data obtained from the SSR n-channel MOSFET devices with L_{eff} down to sub-0.1 μm at room temperature. The slight reduction in I_b/I_d observed at both 300 K and 77 K from the AT&T Bell Labs MOSFETs is attributed to the device punchthrough rather than the hot-electron "cooling" effect. The same conclusion is drawn for the case of p-channel MOSFETs regarding the hot-hole "cooling" effect. Also, as a complementary test of the hot-electron "cooling" effect, the nor-

malized gate current, I_g/I_d, is characterized with decreasing L_{eff}. It is found that there is no indication of gate current reduction with L_{eff} down to sub-0.1 μm regime either. Thus, for well-behaved silicon MOSFETs, there is no convincing evidence of hot-carrier-induced current reduction as a manifestation of the hot-carrier "cooling" effect in the deep-submicron regime. Future work in this area could be to investigate the low-temperature behavior of the hot-carrier-induced currents in the deep-submicron regime in the anticipation of their reductions due to the enhancement of the quasi-ballistic transport.

The scaling relationships among all the fundamental quantities of deep-submicron MOSFETs, device speed g_m/WC_{ox}, drain-induced barrier lowering (DIBL), $\frac{\delta V_t}{\delta V_{ds}}$, effective channel length, L_{eff}, and hot-carrier-induced currents, I_b/I_d or I_g/I_d, are investigated with both device measurements and numerical simulations following a new methodology using nonlinear regressions. The dependence of these relationships on the particular set of channel and source/drain parameters is studied experimentally and by numerical simulations. With this new scaling methodology, all fundamental quantities of MOSFET scaling are unified in the deep-submicron regime with L_{eff} down to sub-0.1 μm. The key findings are summarized as follows: (a) the scaling relationships can be expressed in appropriate power-law forms with excellent statistical significance for both experimental and simulation data samples; (b) there exist universal trade-off relationships among the device performance, the device short-channel effect, and the device hot carrier currents, with respect to the channel parameters, the threshold voltage, V_t, and the channel doping profile. (c) the relationship between g_m/WC_{ox} and $\frac{\delta V_t}{\delta V_{ds}}$ with L_{eff} as an implicit variable is not only insensitive to V_t and channel doping profiles, but also gate-oxide thickness, t_{ox}, within their respective experimental ranges; (d) the trade-off between device performance and the short-channel effect, i.e., g_m/WC_{ox} *vs.* $\frac{\delta V_t}{\delta V_{ds}}$, is dominated by the source/drain parameters. The future direction on deep-submicron MOSFET scaling along those lines is also clear. It is important to investigate the source/drain parameter dependence of the universal relationships among all fundamental MOSFET scaling quantities. It is expected again that it is the source/drain parameters that determine the trade-offs

among g_m/WC_{ox}, $\frac{\delta V_t}{\delta V_{ds}}$, and I_h/I_d.

Bibliography

[1] *MEDICI User's Manual*, pages 2–63. TMA Associates, 1992.

[2] D. A. Antoniadis and J. E. Chung. Physics and Technology of Ultra Short Channel MOSFET Devices. In *Proc. IEDM Tech. Dig.*, pages 21–24, 1991.

[3] G. Baccarani, M. R. Wordeman, and R. H. Dennard. Generalized scaling theory and its application to 1/4 micron MOSFET design. *IEEE Trans. Electron Devices*, ED-31:452–462, 1984.

[4] G. Baccarani, M. R. Wordeman, and R. H. Dennard. An investigation of steady-state velocity overshoot in silicon. *Solid-State Electron.*, Vol-28:407, 1985.

[5] J. R. Brews, W. Fichtner, E. H. Nicollian, and S. M. Sze. Generalized guide for MOSFET miniaturization. *IEEE Electron Dev. Lett.*, EDL-1:2–5, 1980.

[6] D. M. Caughey and R. E. Thomas. Carrier mobilities in silicon empirically related to doping and field. In *Proc. IEEE*, volume Vol-55, pages 2192–2193, 1967.

[7] T. Y. Chan, P. K. Ko, and C. Hu. A simple method to characterize substrate current in MOSFET's. *IEEE Electron Dev. Lett.*, EDL-5(12):505–507, 1984.

[8] T. Y. Chan, P. K. Ko, and C. Hu. Dependence of channel electric field on device scaling. *IEEE Electron Dev. Lett.*, EDL-6(10):551–553, 1985.

[9] J. G. Chern, P. Chang, R. F. Motta, and N. Godinho. A new method to determine MOSFET channel length. *IEEE Electron Dev. Lett.*, EDL-1:170–178, 1980.

[10] S. Y. Chou and D. A. Antoniadis. Relationship between measured and intrinsic transconductance. *IEEE Trans. on Elec. Device*, ED-34:448, 1987.

[11] S. Y. Chou, D. A. Antoniadis, and H. I. Smith. Observation of Electron Velocity Overshoot in Sub-100-nm- channel MOSFET's in Silicon. *IEEE Elec. Dev. Lett.*, EDL-6:665, 1985.

[12] W. Chu. Inorganic X-Ray Mask Technology for Quantum-Effect Devices . In *Ph.D Thesis.* MIT, Cambridge, Massachusetts, 1993.

[13] J. E. Chung, M. Jeng, J. E. Moon, and P. Ko C. Hu. Low-voltage hot-electron currents and degradation in deep-submicrometer MOSFET's. *IEEE Trans. Electron Devices*, ED-37(7):1651–1657, 1990.

[14] R. Coen and R. S. Muller. Velocity of surface carriers in inversion layers on silicon. *Solid-State Electron*, Vol-23:35–40, 1980.

[15] J. A. Cooper and D. F. Nelson. Measurement of the high-field drift velocity of electron in inversion layer in silicon. *IEEE Electron Dev. Lett.*, EDL-2(7):169–173, 1983.

[16] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc. Design of ion-implanted MOSFETs with very small physical dimensions. *IEEE J. Solid-State Circuits*, SC-9:256–268, 1974.

[17] M. Dutoit, J. P. Mieville, Z. M. Shi, N. Revil, and S. Cristoloveanu. Experimental study of electron heating in 0.1 μm n-MOSFET's. In *Proc. IEDM Tech. Dig.*, page 35, 1992.

[18] N. G. Einspruch and G. S. Gildenblat. *Advanced MOS Device Physics*, pages 119–160. New York: Academic Press, Inc., 1989.

[19] F. F. Fang and A. B. Fowler. Hot-electron effects and saturation velocity in silicon inversion layer. *J. of Appl. Phys.*, Vol-41:1825, 1970.

[20] C. Fiegna, H. Iwai, T. Wada, M. Saito, E. Sangiorgi, and B. Ricco. Scaling the MOS transistor below 0.1 μm: methodology, device structures, and technology requirements. *IEEE Transactions on Electron Devices*, ED-41(6):941–951, 1993.

[21] C. Fiegna, F. Venturi, M. Melanotte, E. Sangiorgi, and B. Ricco. Simple and efficient modeling of EPROM writing. *IEEE Transactions on Electron Devices*, 38-3:603–610, 1991.

[22] C. P. Ho, J. D. Plummer, S. E. Hansen, and R. W. Dutton. VLSI process modeling — SUPREM-III. *IEEE Trans. Electron Devices*, ED-30(11):1438–1452, November 1983.

[23] A. Hori, M. Segawa, S. Kameyama, and M. Yasuhira. High-performance dual-gate CMOS utilizing a novel self-aligned pocket implantation (SPI) technology. *IEEE Trans. Electron Devices*, ED-40(9):1675–1681, 1993.

[24] H. Hu, H. Fang, and D. A. Antoniadis. Characterization of hole transport in 0.1 μm PMOSFETs at room and liquid nitrogen temperatures. *submitted to J. of Appl. Phys.*, 1994.

[25] H. Hu, L. T. Su, I. Y. Yang, D. A. Antoniadis, and H. I. Smith. Channel and source/drain engineering in high-performance sub-0.1 μm NMOSFETs using X-ray lithography. In *Proc. Int. Symp. VLSI Technology*, 1994.

[26] R. S. Huang and P. H. Ladbrooke. . *J. Appl. Phys.*, Vol-48:4791, 1977.

[27] C. Jacoboni, C. Canali, G. Ottaviani, and A. A. Quararta. A review of some charge transport properties in silicon. *Solid-State Electron.*, Vol-20:77, 1977.

[28] J. B. Jacobs and D. A. Antoniadis. Channel profile engineering for sub-100 nm MOSFETs. *submitted to the IEEE Transactions on Electron Devices*, 1994.

[29] P. K. Ko, R. S. Muller, and C. Hu. A unified model for hot-electron currents in MOSFETs. In *Proc. IEDM Tech. Dig.*, page 600, 1981.

[30] T. Kobayashi and K. Saito. . *IEEE Trans. Electron Devices*, ED-32:788, 1985.

[31] D. Lau, G. Gildenblat, C. G. Sodini, and D. E. Nelson. Low-temperature substrate characterization of n-channel MOSFETs. In *Proc. IEDM Tech. Dig.*, pages 565–568, 1985.

[32] S. E. Laux, M. V. Fischetti, and D. J. Frank. Monte Carlo analysis of semiconductor devices: The DAMOCLES program. *IBM Journal of Research and Development*, 34-4:466–494, 1990.

[33] K. F. Lee and et al. Room Temperature 0.1 um CMOS Technology with 11.8 ps Gate Delay. In *Proc. IEDM Tech. Dig.*, page 131, 1993.

[34] Y. A. El Mansy and A. R. Boothroyd. A simple two-dimensional model for IGFET operation in the saturation region. *IEEE Trans. Electron Devices*, ED-24(3):254–262, 1977.

[35] Y. Mii and et al. High Performance 0.1 um nMOSFET's with 10 ps/stage Delay (85 K) at 1.5 V Power Supply. In *Proc. Int. Symp. VLSI Technology*, page 91, 1993.

[36] T. Mizuno, A. Toriumi, M. Iwase, M. Takahashi, H. Niiyama, M. Fukumoto, and M. Yoshimi. Hot carrier effects in 0.1 μm gate length CMOS devices. In *Proc. IEDM Tech. Dig.*, pages 695–698, 1992.

[37] A. M. Moel. An Aligner for X-Ray Nanolithography. In *Ph.D Thesis*, pages 20–59. MIT, Cambridge, Massachusetts, 1993.

[38] K. K. Ng, S. A. Eshraghi, and T. D. Stanik. An improved generalized guide for MOSFET scaling. *IEEE Trans. Electron Devices*, ED-40(10):1895–1897, 1993.

[39] T. H. Ning, C. M. Osburn, and H. N. Yu. Emission probability of hot electrons from silicon into silicon dioxide. *J. Appl. Phys.*, 48:286–293, 1977.

[40] K. Nishinohara, N. Shigyo, and T. Wada. Effects of microscopic fluctuations in doping distributions on MOSFET threshold voltage. *IEEE Transactions on Electron Devices*, ED-39:634–639, 1992.

[41] J. P. Nougier, J. C. Vaissiere, D. Gasquet, J. Zimmermann, and C. Constant. Determination of transient regime of hot carriers in semiconductors, using the relaxation time approximation. *J. of Appl. Phys.*, Vol-52:825, 1981.

[42] M. Ono and et al. Sub-50 nm Gate Length N-MOSFETs with 10 nm Phosphorus Source and Drain Junctions. In *Proc. IEDM Tech. Dig.*, page 119, 1993.

[43] M. R. Pinto, E. Sangiorgi, and J. Bude. Silicon MOS transconductance scaling into the overshoot regime. *IEEE Electron Dev. Lett.*, EDL-14:375–378, 1993.

[44] E. Rosencher. A displaced Maxwellian approach to ballistic electron transport in semiconductors. *Solid State Comm.*, Vol-38:1293, 1981.

[45] J. G. Ruch. Electron Dynamics in Short Channel Field-Effect Transistors. *IEEE Trans. Electron Devices*, ED-19:652, 1972.

[46] A. G. Sabnis and J. T. Clemens. Characterization of the electron mobility in the inverted ¡100¿ Si surface. In *Proc. IEDM Tech. Dig.*, pages 18–21, 1979.

[47] G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, S. A. Rishton, and E. Ganin. High transconductance and velocity overshoot in NMOS devices at the 0.1 μm-gate-length level. *IEEE Electron Device Lett.*, EDL-9:464, 1988.

[48] G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, S. A. Rishton, E. Ganin, T. H. P. Chang, and R. H. Dennard. Experimental technology and performance of 0.1 μm-gate- length FETs operated at liquid-nitrogen temperature. *IBM J. of Research and Development*, Vol-34(4):452–465, 1990.

[49] E. Sangiorgi and M. R. Pinto. A semi-empirical model of surface scattering for Monte Carlo simulation of silicon n-MOSFET's. *IEEE Trans. Electron Devices*, ED-39(2):356–361, 1992.

[50] K. Seeger. *Semiconductor Physics*, pages 196–200. Berlin: Springer-Verlag, 5th Edition, 1991.

[51] S. Selberherr, A. Schutz, and H. W. Potzl. MINIMOS - A Two-Dimensional MOS Transistor Analyzer. *IEEE Trans. Electron Devices*, ED-27:1540, 1980.

[52] G. G. Shahidi. Non-Stationary Transport Effects in Deep-Sub-Micron Channel Si MOSFETs. In *Ph.D Thesis.* MIT, Cambridge, Massachusetts, 1986.

[53] G. G. Shahidi, D. A. Antoniadis, and H. I. Smith. Electron Velocity Overshoot at Room and Liquid Nitrogen Temperatures in Silicon Inversion Layers. *IEEE Elec. Dev. Lett.*, EDL-9:94, 1988.

[54] G. G. Shahidi, D. A. Antoniadis, and H. I. Smith. Reduction of channel hot-electron-generated substrate current in sub 150 nm channel length Si MOSFET's. *IEEE Elec. Dev. Lett.*, EDL-9:497–499, 1988.

[55] G. G. Shahidi and et al. A High Performance 0.15 μm CMOS. In *Proc. Int. Symp. VLSI Technology*, pages 93–94, 1993.

[56] W. Shockley. Problems related to p-n junctions in silicon. *Solid-State Electron*, 2:35–67, 1961.

[57] C. G. Sodini, P. K. Ko, and J. L. Moll. The effect of high fields on MOS device and circuit performance. *IEEE Trans. Electron Devices*, ED-31(10):1386, 1984.

[58] R. Stratton. Diffusion of hot and cold electrons in semiconductor barriers . *Phys. Rev.*, Vol-126:2002, 1962.

[59] L. T. Su, M. J. Sherony, H. Hu, J. E. Chung, and D. A. Antoniadis. Optimization of series resistance in sub-0.2 μm SOI MOSFETs. In *Proc. IEDM Tech. Dig.*, pages 723–726, 1993.

[60] S. C. Sun and J. D. Plummer. Electron mobility in inversion and accumulation layers on thermally oxidized silicon surface. *IEEE Trans. Electron Devices*, ED-27(8):1497, 1980.

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

[62] S. Takagi, M. Iwase, and A. Toriumi. On the universality of inversion-layer mobility in n- and p-channel MOSFET's. In *Proc. IEDM Tech. Dig.*, pages 398–401, 1988.

[63] S. Tam, P. K. Ko, and C. Hu. Lucky-electron model of electron injection in MOSFET's. *IEEE Trans. Electron Devices*, ED-31(9):1116–1125, 1984.

[64] Y. Taur and et al. High Performance 0.1 um CMOS Devices with 1.5 V Power Supply. In *Proc. IEDM Tech. Dig.*, page 127, 1993.

[65] G. W. Taylor. . *Bell Lab. Tech. J.*, Vol-63:1325, 1984.

[66] K. Terada and H. Muta. A new method to determine effective MOSFET channel length. *Japan J. Applied Physics*, Vol. 18:953–959, 1979.

[67] A. Toriumi and et al. High Speed 0.1 um CMOS Devices Operating at Room Temperature. In *Proc. SSDM Tech. Dig.*, page 487, 1992.

[68] F. Venturi, E. Sangiorgi, R. Brunetti, W. Quade, C. Jacoboni, and B. Ricco. Monte Carlo simulations of high energy electrons and holes in Si-n-MOSFET's. *IEEE Transactions on Computer-Aided Design*, 10-10:1276–1286, 1991.