COMPACT MODELING OF DEEP SUBMICRON CMOS TRANSISTOR WITH SHALLOW TRENCH ISOLATION MECHANICAL STRESS EFFECT

PHILIP TAN BEOW YEW

UNIVERSITI SAINS MALAYSIA

2008

COMPACT MODELING OF DEEP SUBMICRON CMOS TRANSISTOR WITH SHALLOW TRENCH ISOLATION MECHANICAL STRESS EFFECT

by

PHILIP TAN BEOW YEW

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

August 2008

ACKNOWLEDGEMENTS

I would like to express my gratitude to my Ph.D. project supervisor, Dr. Othman Sidek for his guidance and encouragement throughout the duration of this project. His support in determining the project that benefit to both my work and study is highly appreciated.

Special thanks go to my Silterra supervisor, Al Kordesch for his unconditional guidance and constructive suggestions, which enable me to overcome the obstacles in my Ph.D. project. I am also grateful to Silterra management team for the encouragement for further study and technical publications. Not forgetting to all my supportive colleagues in Technology Development and Device Modeling team, thanks for everything.

Finally, I would like to express my sincere thanks and utmost appreciation to my beloved family, especially to my wife, Christine. For their undying love, support and patient that enable me to complete my Ph.D. degree successfully.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

%	Percent
≥	Greater or equal
Δ	Delta
α	Power term for L dependence STI stress effect
π	pi (3.142)
μ	Carrier mobility
η	Ideality factor
$\mu_0(a)$	Active space carrier mobility
$\mu_0(a_{\_min})$	Minimum active space carrier mobility
μA	micro-Ampere
μ_{eff}	Effective carrier mobility
ϕ_f	Difference of Fermi potential and substrate intrinsic potential
μm	micrometer
δ_{NP}	Positive sign for NMOS and negative sign for PMOS
ε_{ox}	Oxide permittivity (12 x 8.85 x 10^{-12} F/m)
ε_s	Silicon permittivity (3.9 x 8.85 x 10^{-12} F/m)
<	Smaller
a	Active space
$a_{\_min}$	Minimum active space
A_{bulk}	Bulk charge effect
A_{gu0}	Gaussian distribution term for STI x-stress mismatch
C	Capacitance
C_{dep}	Depletion capacitance per unit area in the bulk
C_{ox}	Gate oxide capacitance
d	Distance

E	Electric Field
f_0, f_1, f_2	Representation of complicated functions
h	Planck's constant (6.6262×10^{-34} Js)
I	Current
I_d	Drain current
$I_{d,0}$	Ideal long channel I_d
I_{dlin}	Linear drain current
I_{dsat}	Saturation drain current
I_{off}	Leakage current (off current)
I_{sub}	Substrate current
J	Current density
k	p x ($h/2\pi$)
L	Channel length
L_{drawn}	Drawn channel length
L_{eff}	Effective channel length
L_{od}	Length of transistor active area (length of oxide definition)
L_{od}	Active length
m	meter
m*	Effective mass
mA	milli-Ampere
meV	milli-electron-Volt
mV	milli-Volt
N	Number of samples
n_c	Carrier concentration
nm	nanometer
p	Momentum
P_{a0}	L dependence bulk charge parameter

P_{ags}	Gate bias dependence bulk charge parameter
P_{aid}	STI x-stress mismatch magnitude parameter
P_{alpha0}	First SCBE effect on I_{sub} parameter
P_{alpha1}	L dependence SCBE effect on I_{sub} parameter
P_{b0}	Channel width bulk charge parameter
P_{b1}	Channel width offset bulk charge parameter
P_{beta0}	Second SCBE effect on I_{sub} parameter
P_{cdsc}	Drain and source to channel coupling capacitance parameter
P_{cdscb}	Body bias sensitivity of P_{cdsc} parameter
P_{cdscd}	Drain bias sensitivity of P_{cdsc} parameter
P_{cit}	Interface trap capacitance parameter
P_{dsub}	V_t L dependence exponent DIBL parameter
P_{dvt0}	First SCE parameter
P_{dvt0w}	First SS parameter
P_{dvt1}	L dependence exponent SCE parameter
P_{dvt1w}	L dependence exponent SS parameter
P_{dvt2}	Body bias SCE parameter
P_{dvt2w}	Body bias SS parameter
P_{dwb}	Body bias channel width offset parameter
P_{dwg}	Gate bias channel width offset parameter
P_{eta0}	V_t DIBL parameter
$P_{eta0\_original}$	P_{eta0} without STI stress effect
P_{etab}	V_t body bias DIBL parameter
P_{k1}	First-order body effect parameter
P_{k2}	Second-order body effect parameter
P_{k3}	NWE parameter
P_{k3b}	Body effect NWE parameter

P_{keta}	Body bias dependence bulk charge parameter
P_{ku0}	STI stress effect on P_{u0} parameter
P_{lint}	Channel length offset parameter
P_{lku0}, P_{wku0}, P_{pku0}	STI stress effect of P_{u0} binning parameters
P_{lku1}	L fitting parameters for STI stress effect
P_{ll}, P_{lw}, P_{lwl}, P_{lln}, P_{lwn}	Channel length geometry scaling parameters
P_{mobmod}	Mobility mode parameter
P_{mp}	STI x-stress mismatch sensitivity parameter
$P_{nfactor}$	Subthreshold turn-on swing parameter
P_{ngate}	Polysilicon gate doping concentration parameter
P_{nlx}	RSCE parameter
P_{nrd}	Number of drain square parameter
P_{nrs}	Number of source square parameter
P_{pclm}	Channel length modulation parameter
$P_{pdiblc1}$	First DIBL correction on V_a parameter
$P_{pdiblc2}$	Second DIBL correction on V_a parameter
$P_{pdiblcb}$	Body bias DIBL correction on V_a parameter
P_{prwb}	Body bias dependence of P_{rdsw} parameter
P_{prwg}	Gate bias dependence of P_{rdsw} parameter
P_{pscbe1}	First high V_d drain conductance modification parameter
P_{pscbe2}	Second high V_d drain conductance modification parameter
P_{pvag}	Gate bias dependence V_a parameter
P_{rdsw}	Drain and source resistance per μm of gate width parameter
P_{rsh}	Sheet resistance parameter
P_{saref}	s_a reference parameter
P_{sbref}	s_b reference parameter
P_{stxu01}	STI x-stress mobility parameter

$P_{stxvth01}$	STI x-stress V_t parameter
P_{stxwe}	STI x-stress width dependence parameter
P_{stxwep}	STI x-stress width dependence exponential parameter
P_{styu01}	First STI y-stress parameter
P_{styu02}	Second STI y-stress parameter
P_{tox}	Oxide thickness parameter
P_{u0}	Zero field universal mobility parameter
$P_{u0\_original}$	P_{u0} without STI stress effect or mismatch effect
P_{ua}	First order mobility degradation parameter
P_{ub}	Parabolic mobility degradation parameter
P_{uc}	Body bias mobility degradation parameter
P_{voff}	Subthreshold offset voltage parameter
P_{vsat}	Carrier saturation velocity parameter
$P_{vsat\_original}$	P_{vsat} without STI stress effect
P_{vth0}	Zero back bias threshold voltage parameter
$P_{vth0\_original}$	P_{vth0} without STI stress effect
P_{w0}	Channel width offset NWE parameter
P_{wint}	Channel width offset parameter
P_{wl}, P_{ww}, P_{wwl}, P_{wln}, P_{wwn}	Channel width geometry scaling parameters
P_{wr}	Exponent effective channel width of P_{rdsw} parameter
P_{wvth0}	Width binning parameter of P_{vth0}
P_{ww}	Channel width offset parameter
P_{xj}	Source/drain junction depth parameter
P_{xl}	Delta width
P_{xw}	Delta length
q	Electron charge (1.6e-19 Coulombs)
R_d	Drain resistance

R_{ds}	Drain and source resistance
R_{out}	Output resistance
R_s	Source resistance
S_{0l}, S_{1l}	L fitting parameters for STI stress effect
s_a, s_b	Distance between STI edge and gate edge (active space)
T	Temperature
T_{nom}	Nominal temperature
V	Voltage
V_a	Early voltage
V_b	Bulk voltage
V_{bi}	Built-in voltage
V_d	Drain voltage
$V_{d,eff}$	Effective drain voltage of linear to saturation smoothing function
V_{dd}	Operating voltage
V_{dsat}	Saturation drain voltage
V_g	Gate voltage
$V_{m\mu 0}$	Maximum $\mu_0(a)$ variation with respect to $\mu_0(a_{\_min})$
V_s	Source voltage
V_t	Threshold voltage
V_{tlin}	Linear threshold voltage
W	Channel width
W_{drawn}	Drawn channel width
W_{eff}	Effective channel width
x	Distance from the center of the channel
X_{dep}	Channel depletion thickness in the substrate
$X_{dep,0}$	Channel depletion thickness in the substrate at zero back bias
X_{poly}	Poly depletion thickness

LIST OF ABBREVIATION

2D	Two-dimensional
3D	Three-dimensional
AA	Active Area
BSIM3v3	Berkeley Short-Channel IGFET Model – Version 3.3
BSIM3v3.24	Berkeley Short-Channel IGFET Model – Version 3.3.24
BSIM3v3.3	Berkeley Short-Channel IGFET Model – Version 3.3.3
BSIM4	Berkeley Short-Channel IGFET Model – Version 4
BSIM4.3.0	Berkeley Short-Channel IGFET Model – Version 4.3.0
BSIM4.6.1	Berkeley Short-Channel IGFET Model – Version 4.6.1
BTBT	Band-To-Band Tunneling
CD	Critical Dimension
CLM	Channel Length Modulation
CMC	Compact Modeling Council
CMOS	Complementary Metal Oxide Semiconductor
CTE	Coefficient of Thermal Expansion
DC	Direct Current
DIBL	Drain Induced Barrier Lowering
DW	Delta Width
EKV	Bulk-referencing SPICE model
exp	Exponential
GCA	Gradual Channel Approximation
GIDL	Gate Induced Drain Leakage
HCE	Hot Carrier Effect
HISIM	Hiroshima University SPICE model
HSPICE	Circuit simulation engine created by Meta-Software

IC	Integrated Circuit
IEDM	International Electron Devices Meeting
IGFET	Insulated Gate Field Effect Transistor
INWE	Inverse Narrow Width Effect
LDD	Lightly Doped Drain
LOCOS	Localized-Oxidation-of-Silicon isolation
MOS	Metal Oxide Semiconductor
NMOS	N-type Metal Oxide Semiconductor
NWE	Narrow Width Effect
PMOS	P-type Metal Oxide Semiconductor
PSP	Penn State Philips SPICE model
PSS	Process-strained Silicon
RSCE	Reverse Short Channel Effect
SCBE	Substrate Current induced Body Effect
SCE	Short Channel Effect
SD	Source/drain
Si	Silicon
SiO_2	Silicon dioxide
SPICE	Simulation Program with Integrated Circuit Emphasis
sqrt	Square Root
SS	Small Size
stdev	Standard Deviation
STI	Shallow Trench Isolation
TCAD	Technology Computer Aided Design
TSMC	Taiwan Semiconductor Manufacturing Company
XSIM	National Technology University SPICE model

PEMODELAN PADAT UNTUK TRANSISTOR CMOS DI BAWAH SUBMIKRON DENGAN KESAN TEKANAN MEKANIKAL PENGASINGAN PEPARIT CETEK

ABSTRAK

Thesis ini memperkenalkan satu model padat, dua model berasaskan empirikal dan satu model berasaskan fizikal untuk kesan tekanan mekanikal Pengasingan Peparit Cetek (STI) ke atas transistor CMOS di bawah submikron. Model tekanan-x STI padat digunakan untuk menangkap tekanan dalam arah laluan transistor panjang. Model ini adalah lebih ringkas berbanding dengan model tekanan STI BSIM4, tetapi dapat mencapai ketepatan yang serupa. Dua ciri-ciri baru tekanan-x STI telah dikenalpastikan. Ciri yang pertama adalah fakta bahawa kesan tekanan-x STI bagi transistor CMOS berubah untuk transistor lebar yang berlainan. Kesan ini telah dikesan dalam teknologi-teknologi 0.18 μm dan 0.13 μm. Satu model kesan tekanan-x STI bergantung kepada lebar yang empirikal telah dicadangkan untuk menangkap kesan ini. Ciri baru yang kedua adalah fakta bahawa kesan tekanan-x STI mengubah ciri-ciri ketidaksamaan transistor CMOS. Satu model Monte Carlo yang empirikal telah dicadangkan untuk menangkap kesan ini. Satu lengkuk arus parit tepu (I_{dsat}) berupabentuk cangkul berlawan dengan transistor lebar yang baru telah dikenalpasti. Lengkuk ini tidak dapat dimodelkan dengan mengunakan model tekanan STI BSIM4. Dengan mengunakan satu kaedah bentangan yang baru, ciri-ciri fizikal lengkuk ini telah dikenalpasti. Lengkuk I_{dsat} berupabentuk cangkul ini disebabkan oleh gabungan kesan-kesan daripada tekanan-y STI (tekanan STI dalam arah laluan transistor lebar) yang menurunkan I_{dsat} dan kesan Perubahan Lebar (DW) yang meningkatkan I_{dsat}. Berdasarkan ciri-ciri fizikal ini, satu model tekanan-y STI yang fizikal diperkenalkan untuk menangkap kelakuan I_{dsat} berupabentuk cangkul ini. Ketepatan model-model yang dicadangkan di dalam thesis ini dipastikan dengan mengunakan silikon data benar yang difabrikasikan dengan mengunakan teknologi-teknologi CMOS Silterra 0.18 μm dan 0.13 μm yang berpiawaian industri. Model-model baru ini dibina dalam model

BSIM3v3 dengan mengunakan kaedah model makro (juga dikenali sebagai kaedah sublitar). Dua parameter SPICE, parameter voltan ambang tiada pincang belakang, P_{vth0} dan parameter kelincahan pembawa, P_{u0}, telah digunakan untuk membina model-model ini. Perbezaan masa simulasi antara model makro dan model konventional adalah tidak ketara (< 5 %).

COMPACT MODELING OF DEEP SUBMICRON CMOS TRANSISTOR WITH SHALLOW TRENCH ISOLATION MECHANICAL STRESS EFFECT

ABSTRACT

This thesis introduces a compact model, two empirical-based models and a physical-based model of Shallow Trench Isolation (STI) mechanical stress effect on deep submicron CMOS transistor. The compact STI x-stress model is used to capture the stress effect in the channel length direction. This model is simpler than the BSIM4 STI stress model, but able to achieve the similar accuracy. Two new characteristics of STI x-stress have been identified. The first characteristic is the fact that the STI x-stress effect on CMOS transistor varies for different transistor channel widths. An empirical width dependence of STI x-stress effect model has been proposed to capture this effect. The second new characteristic is the fact that STI x-stress effect changes the CMOS transistor mismatch characteristics. An empirical Monte Carlo model is proposed to capture this effect. A new hook shaped saturation drain current, I_{dsat} curve versus channel width has been identified. This curve cannot be modeled using the BSIM4 STI stress model. By using a new layout method, the physical characteristics of the curve are identified. The hook shaped I_{dsat} curve is caused by the combined effects of STI y-stress (stress in the channel width direction) that degrades the I_{dsat} and the Delta Width (DW) effect that increases the I_{dsat}. Based on the physical characteristics, a new physical-based STI y-stress model is proposed to capture the hook shaped I_{dsat} behavior. The accuracy of the models in this thesis is verified on actual silicon data fabricated using Silterra's industry-standard 0.18 μm and 0.13 μm CMOS technologies. These new models are incorporated into the BSIM3v3 model by using macro model method (also known as subcircuit method). The two SPICE parameters, the zero back bias threshold voltage parameter, P_{vth0} and the carrier mobility parameter, P_{u0}, are used for developing these models. The difference in simulation time between the macro model and the conventional model is insignificant (< 5 %).

CHAPTER 1
INTRODUCTION

1.0 Chapter Overview

This chapter introduces the compact modeling works in this thesis. First, the basic concept of compact modeling is introduced. Then, the history of compact modeling of CMOS transistor which started with the theory Gradual Channel Approximation (GCA) is reviewed. How the important deep submicron effects of CMOS transistor are captured in BSIM3v3 equations are then discussed.

After reviewing the theoretical concept of compact modeling, the reasons why this research focuses on SPICE modeling of the STI stress effect on CMOS transistor is discussed. The background and objective of this research are then given. Finally, the contribution of the proposed STI mechanical stress models to compact modeling research is discussed.

1.1 Introduction to Compact Modeling

Compact model is basically a set of simplified physics equations that describe the behaviors of semiconductor devices and able to run with SPICE simulator at minimum time. Compact modeling is finding the parameter values for the compact model equations to achieve a good fit between the simulation data and the measured data. The main objective of the compact modeling is to enable the accurate prediction of circuit performance before actual fabrication.

When transistor scaling goes to deep submicron, more and more significant physical phenomena need to be captured by compact model. The physical phenomena that are negligible in the CMOS micron technology have dominated the behavior of the CMOS deep submicron technology.

Many researchers have put in much effort in establishing the set of equations (Compact Model) that can provide the most accurate circuit simulation at the minimum

time. Berkeley Short-Channel IGFET models (BSIM3v3 and BSIM4 models) are currently widely used by wafer fabrication foundries and IC design houses. Other SPICE models such as PSP model, EKV model, HISIM model and XSIM model are also popular. The standardization of these models is performed by an independent organization called Compact Model Council (CMC).

All the modeling works in this thesis is based on BSIM3v3.24, where the final released version is BSIM3v3.3 model (Cheng *et al.*, 2005). In short, these BSIM3v3 family models, which built on similar source code, are known as BSIM3v3 model in this thesis. The discussion in this thesis also references to BSIM4.3.0 (Xi *et al.*, 2003) and BSIM4.6.1 (Dunga *et al.*, 2007) models for STI stress model discussion. In short, all the BSIM4 family models are known as BSIM4 models in this thesis. The BSIM4 model is built on a different source code compared to the BSIM3v3. Hence, the BSIM4 model is not backward compatible to the BSIM3v3 model. The technology nodes of the experimental transistors that are used in this thesis are 0.18 μm and 0.13 μm.

1.2 History of Compact Modeling

Compact modeling is originally driven by the need to express the CMOS transistor behavior in quantitative form. Specifically, to design circuits with CMOS transistors, how much current is carried in the “on” state of a transistor and how much leakage current flows in the “off” state must be calculated. In short, the current-voltage, I-V and capacitance-voltage, C-V characteristics of the CMOS transistor are needed, specifically the quantitative relationship between the drain current and terminal voltages.

In the beginning of the compact modeling era, there were four long-channel CMOS transistor models proposed for DC circuit simulation (or calculation) as explained in the reference book (Wolf, 1995). All the four models are based on the long-channel Gradual Channel Approximation (GCA), which assumes that the variation of the electric field in the x-direction (channel length, L direction) is much less than the

corresponding variation in the z-direction (perpendicular to channel plane direction). The four models are Bulk-Charge model, Square-Law model, Pao-Sah model and Charge-Sheet model.

Bulk-Charge model (Ihantola and Moll, 1964) and Square-Law model (simplified version of Bulk-Charge model) are only valid if the CMOS transistor is operated in strong inversion. This has limited the usage of these models.

Pao-Sah model (Pao and Sah, 1966) is the first model that able to cover the entire range of CMOS transistor operation. This model retains the GCA but the drain current, I_d equation requires numerical integration in two dimensions. Such numerical complexity has limited the role of this model for theoretical analysis of the CMOS transistor.

The limited practical usage of the Pao-Sah model motivated a search for an approximate advanced analytical model that is still accurate over a wider range of operating conditions. Charge-Sheet model (Brews, 1978) has become the most widely adopted long-channel CMOS model that is accurate in the inversion range.

Although the Charge-Sheet model allows I_d to be calculated with much less computational effort than that of the Pao-Sah model, the computational effort for determining I_d is still too great for most circuit simulation applications, and thus the Charge-Sheet model has not been incorporated into the popular circuit simulation engine, which is known as the SPICE simulator.

Since the Charge-Sheet model is too complex to be used in most circuit simulation applications, an alternate analytical expression (simplified from the Charge-Sheet model) is implemented to be used in subthreshold region. For the inversion region, the simpler Bulk-Charge and Square-Law models are used. The combination of these models enables the SPICE simulator to cover the entire range of CMOS transistor operation with a reasonable simulation time.

When comparing the usage of the Bulk-Charge and Square-Law models, the primary advantage of the Square-Law model is its simplicity. Most hand calculations for

circuit design with CMOS transistors make use of the Square-Law model. It is also the model used for the simplest (Level 1) analysis of circuits in SPICE simulator. While the more-accurate Level 2 SPICE for CMOS transistor is based on the Bulk-Charge model.

1.3 Compact Modeling of CMOS Transistor Deep Submicron Effects

This section discusses how SPICE model equations able to capture the CMOS transistors deep submicron effects. BSIM3v3 modeling equations are selected for the discussion in this section because the modeling equations are less complicated compared to BSIM4 but sufficient to describe most of the CMOS transistor deep submicron effects. One can easily catch up with BSIM4 or others SPICE models with the basic understanding of BSIM3v3.

The electrical characteristics of a CMOS transistor are threshold voltage, V_t, drain current, I_d, off current, I_{off} and substrate current, I_{sub}. How the deep submicron effects change the transistor electrical characteristics and how these effects can be captured in BSIM3v3 equations is discussed. The following discussions in this section are based on three SPICE modeling reference books (Taur and Ning, 1998, Cheng and Hu, 1999, and Liu, 2001), and two BSIM manuals; BSIM3v3 manual (Cheng *et al.*, 2005) and BSIM4 manual (Dunga *et al.*, 2007).

The physical phenomena of CMOS transistors that are discussed in this section, are (1) Body effect, (2) Short Channel V_t Effect, SCE due to charge sharing, (3) Reverse Short Channel Effect, RSCE due to lateral non-uniform doping, (4) Narrow Width Effect, NWE, due to LOCOS isolation, (5) Inverse Narrow Width Effect, INWE due to STI isolation, (6) Small Size effect, SS, (7) Drain Induced Barrier Lowering, DIBL, (8) Channel Length Modulation, CLM, (9) Velocity Saturation, (10) Subthreshold conduction, (11) Field dependent mobility, (12) Substrate Current induced Body Effect, SCBE, (13) Gate Induced Drain Leakage, GIDL, (14) Polysilicon gate depletion effect, (15) Velocity overshoot, (16) Source and drain resistance effect and (17) Effective length and effective width effect.

In BSIM3v3, there are six effects that change the threshold voltage, V_t as shown in Equation 1.

$$V_t = P_{vth0} + \delta_{np}\begin{pmatrix} \Delta V_{t,\,body\_effect} - \Delta V_{t,\,sce} + \Delta V_{t,\,rsce} + \Delta V_{t,\,nwe} \\ + \Delta V_{t,\,ss} - \Delta V_{t,\,dibl} \end{pmatrix} \qquad (1)$$

where P_{vth0} is the BSIM3v3 threshold voltage parameter of a long channel transistor at zero back bias, V_b = 0 V. δ_{np} is positive sign for NMOS and negative sign for PMOS. The four effects that increase V_t are body effect, RSCE, NWE and SS. The other two effects that decrease V_t are SCE and DIBL. These six effects on V_t are discussed in the following sections.

The simple Square-Law form of I_d equations (Brews, 1978) in linear region and saturation region are shown in Equation 2 and Equation 3 below:

$$I_d = \mu C_{ox}\frac{W}{L}\left[(V_g - V_t)V_d - \frac{1}{2}V_d{}^2\right] \qquad \text{when } V_d < V_{dsat} \qquad (2)$$

$$I_d = \mu C_{ox}\frac{W}{L}\left[\frac{(V_g - V_t)^2}{2}\right] \qquad \text{when } V_d \geq V_{dsat} \qquad (3)$$

where $V_{dsat} = (V_g - V_t)$, μ is the carrier mobility, C_{ox} is the gate oxide capacitance, W is the channel width and L is the channel length.

I_d modeling in BSIM3v3 equations is discussed based on seven effects. They are bulk charge effect (A_{bulk}), CLM, DIBL, field dependent carrier's mobility, μ, SCBE, velocity saturation and source drain resistance effect.

I_{off} is modeled using P_{voff}, which is the subthreshold offset voltage parameter. Subthreshold conduction is calculated based on ideality factor, η, where I_d is calculated when V_g is below V_t. These effects are further discussed in the following sections.

I_{sub}, which is caused by SCBE is model in BSIM3v3. The parameters that used to model the I_{sub} are actually P_{alpha0}, P_{alpha1} and P_{beta0}, but not P_{pscbe1} and P_{pscbe2}

(although these last two parameters' names have the term SCBE). SCBE is actually caused by impact ionization that produces extra electrons that sweep into the drain that increases I_d and also produces extra holes that sweep into the substrate that increases I_{sub}. The increase of I_d is modeled using P_{pscbe1} and P_{pscbe2} and the increase of I_{sub} is modeled using P_{alpha0}, P_{alpha1} and P_{beta0}.

1.3.1 Body effect and Bulk Charge Effect

The body effect refers to the effect of V_b on V_t. The body effect increases the V_t of a CMOS transistor, when a reverse bias V_b is applied. Larger reverse V_b increases the threshold voltage because the channel depletion layer becomes wider and hence higher gate voltage, V_g is needed to invert the channel.

Body effect on V_t is modeled in BSIM3v3 by using Equation 4,

$$\Delta V_{t,\,body\_effect} = P_{k1}\left(\sqrt{2\phi_f - V_b} - \sqrt{2\phi_f}\right) - P_{k2}V_b \tag{4}$$

where P_{k1} is the first-order body effect parameter and P_{k2} is the second-order body effect parameter. Default values of P_{k1} = 0.53 $V^{1/2}$ and P_{k2} = -0.0186. From Equation 4, the more positive value of P_{k1} and P_{k2}, the more significant the body effect is. ϕ_f is the difference between Fermi potential and the intrinsic potential of the substrate. $2\phi_f$ of a substrate doping of 1e17 cm^{-3} is approximately 0.81 V.

The bulk charge effect is closely related to the body bias effect and refers to the changing V_t along the channel when V_d > 0 V. V_t is not constant along the channel because the width of the depletion region along the channel is not uniform in the presence of a nonzero V_d, as shown in Figure 1.1.

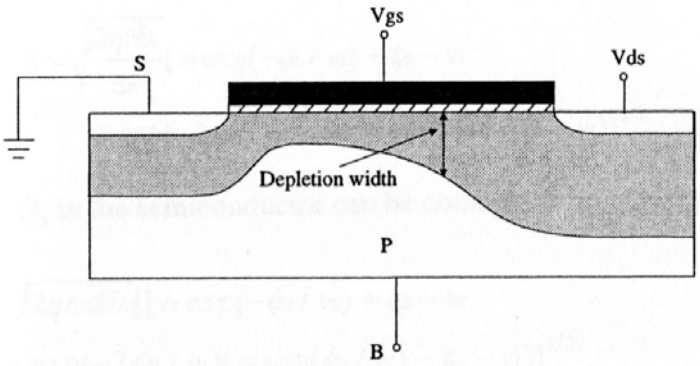

Figure 1.1: Non-uniform depletion width when V_d > 0 V (Cheng and Hu, 1999).

Body effect on I_d, is modeled in BSIM3v3 by using A_{bulk}, as shown in Equation 5.

$$A_{bulk} = \left[1 + \frac{P_{k1}}{2\sqrt{2\phi_f - V_b}}\left[\frac{P_{a0}L_{eff}}{L_{eff} + 2\sqrt{P_{xj}X_{dep}}}\left(1 - P_{ags}(V_g - V_t)\left(\frac{L_{eff}}{L_{eff} + 2\sqrt{P_{xj}X_{dep}}}\right)^2\right) + \frac{P_{b0}}{W_{eff} + P_{b1}}\right]\right]\frac{1}{1 + P_{keta}V_b} \quad (5)$$

where P_{a0} is the L dependence bulk charge parameter, P_{ags} is gate bias dependence bulk charge parameter, P_{keta} is body bias dependence bulk charge parameter, P_{b0} is channel width bulk charge parameter, P_{b1} is channel width offset bulk charge parameter. The default values of P_{a0} = 1, P_{ags} = 0, P_{keta} = -0.047 V^{-1}, P_{b0} = 0 m, P_{b1} = 0 m. P_{xj} is source/drain junction depth parameter and X_{dep} is the channel depletion thickness in the substrate. W_{eff} is the effective channel width. The W_{eff} is further discussed in Section 1.3.17.

Equation 6 and Equation 7 show the I_d equation with the term A_{bulk}.

$$I_d = \mu C_{ox}\frac{W}{L}\left[(V_g - V_t)V_d - \frac{1}{2}A_{bulk}\cdot V_d{}^2\right] \qquad \text{when } V_d < V_{dsat} \quad (6)$$

$$I_d = \mu C_{ox} \frac{W}{L} \left[\frac{(V_g - V_t)^2}{2 A_{bulk}} \right] \qquad \text{when } V_d \geq V_{dsat} \qquad (7)$$

where $V_{dsat} = \frac{(V_g - V_t)}{A_{bulk}}$

1.3.2 Short Channel Effect (SCE)

SCE is the decrease of transistor V_t as L is reduced, due to charge sharing. The SCE is especially pronounced when $V_d = V_{dd}$ (or at high drain bias). Figure 1.2(a) shows the long channel depletion region, where the effect from the source and drain can be ignored.

When L is decreased, there is more overlapping of the source and drain depletion region to the channel. This results in electric field pattern in two-dimensional, as shown in Figure 1.2(b). In other words, the depletion charge under the gate is actually induced by the gate together with the source and drain. The channel charge is considered to be “shared” by the gate, source and drain (Charge Sharing Model). Hence, less gate charge density (smaller gate voltage) is needed to induce inversion in short channel transistors than in long channel transistors. This means the gate voltage, V_g required to turn on the transistor become smaller as the L is reduced.

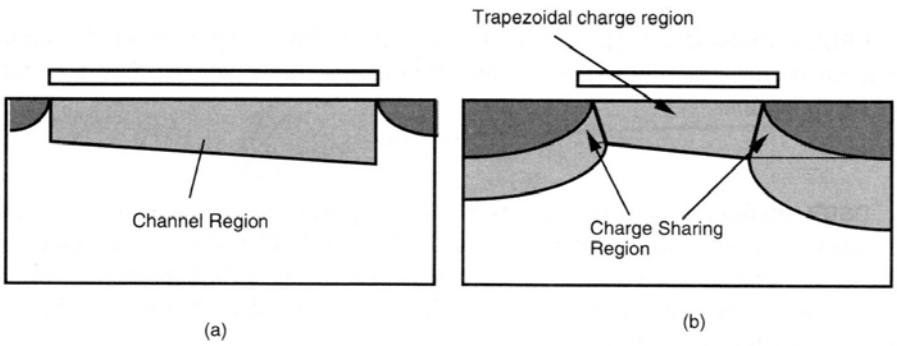

Figure 1.2: Charge regions: (a) in long channel; (b) in short channel (Liu, 2001).

SCE effect on V_t is model in BSIM3v3 by using the Equation 8,

$$\Delta V_{t,\,sce} = P_{dvt0}\left[e^{\left(-P_{dvt1}\frac{L_{eff}}{2L_t}\right)} + 2e^{\left(-P_{dvt1}\frac{L_{eff}}{L_t}\right)} \right](V_{bi} - 2\phi_f) \tag{8}$$

where $L_t = \sqrt{\frac{\varepsilon_s X_{dep}}{C_{ox}}}\left(1 + P_{dvt2}V_b\right)$ and

P_{dvt0} is the first SCE parameter, P_{dvt1} is the L dependence exponent SCE parameter and P_{dvt2} is the body bias SCE parameter. The default value of P_{dvt0} = 2.2, P_{dvt1} = 0.53 and P_{dvt2} = -0.032 V^{-1}. L_{eff} is the effective channel length. The L_{eff} is further discussed in Section 1.3.17. V_{bi} is the built-in voltage of the source-bulk/drain-bulk junction and ε_s is the silicon permittivity.

1.3.3 Reverse Short Channel Effect (RSCE)

RSCE causes the V_t to increase as L decreases. This is caused by the non-uniform lateral doping effect, due to pocket implantation, as shown in Figure 1.3. The channel doping concentration near the source and drain is higher than in the middle of the channel. The increased doping concentration in the two ends of the channel can result in an increase in V_t as the channel becomes shorter. The combined RSCE and SCE effects result in a 'hump' in the characteristics of V_t versus L, as shown in Figure 1.4.

RSCE effect on V_t is modeled in BSIM3v3 by using the Equation 9.

$$\Delta V_{t,\,rsce} = P_{k1}\left(\sqrt{1 + \frac{P_{nlx}}{L_{eff}}} - 1\right)\sqrt{2\phi_f} \tag{9}$$

where P_{nlx} is the RSCE parameter and L_{eff} is the effective channel length. The default value of P_{nlx} = 1.74 x 10^{-7} m. RSCE increases the V_t of short channel transistors and has negligible effect on V_t of long channel transistors.

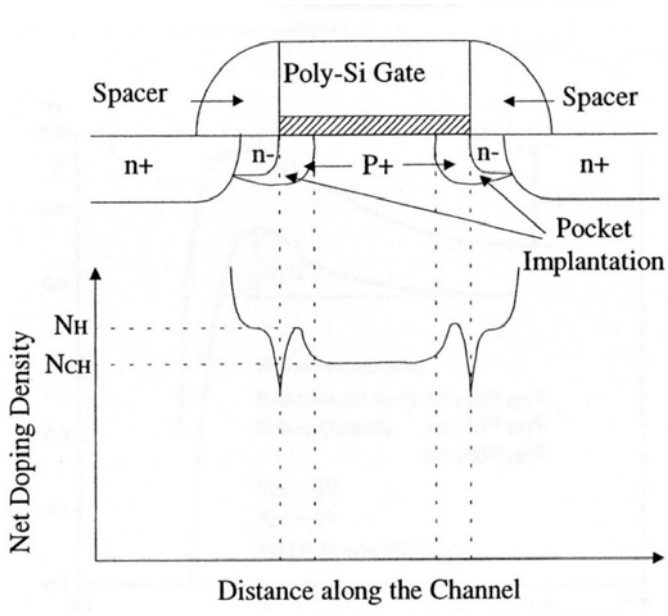

Figure 1.3: Lateral doping profile with pocket implantation (Cheng and Hu, 1999).

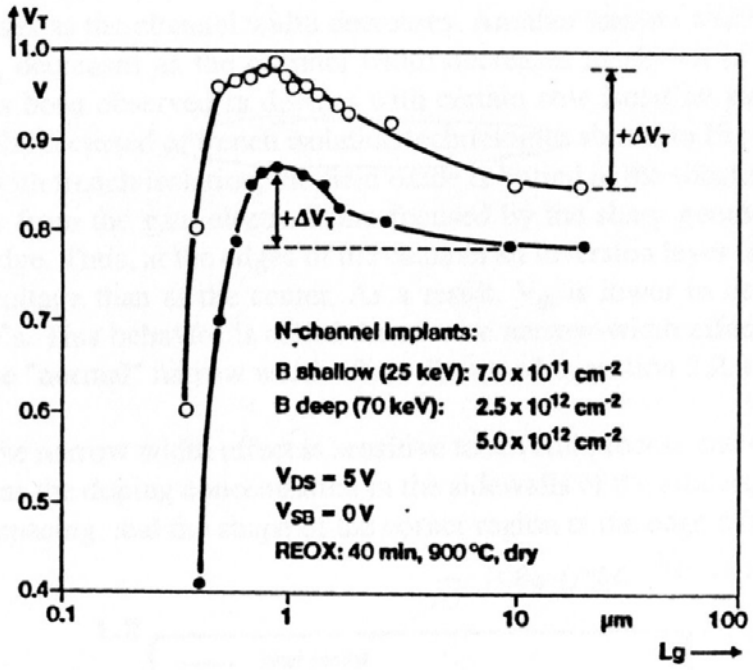

Figure 1.4: V_t versus L characteristics of a transistor with RSCE (Cheng and Hu, 1999).

1.3.4 Narrow Width Effect (NWE)

NWE happens when the Localized-oxidation-of-silicon (LOCOS) isolation technology is used. Figure 1.5 shows the transistor cross-section in the channel width, W direction with LOCOS isolation. Both ends of the gate oxide (in W direction) are known as “bird beak” due to their shape. The “bird beak” shape causes thicker gate oxide at the edge and hence results in higher V_t. As W reduces the “bird beak” effect becomes more severe. This increases the overall V_t.

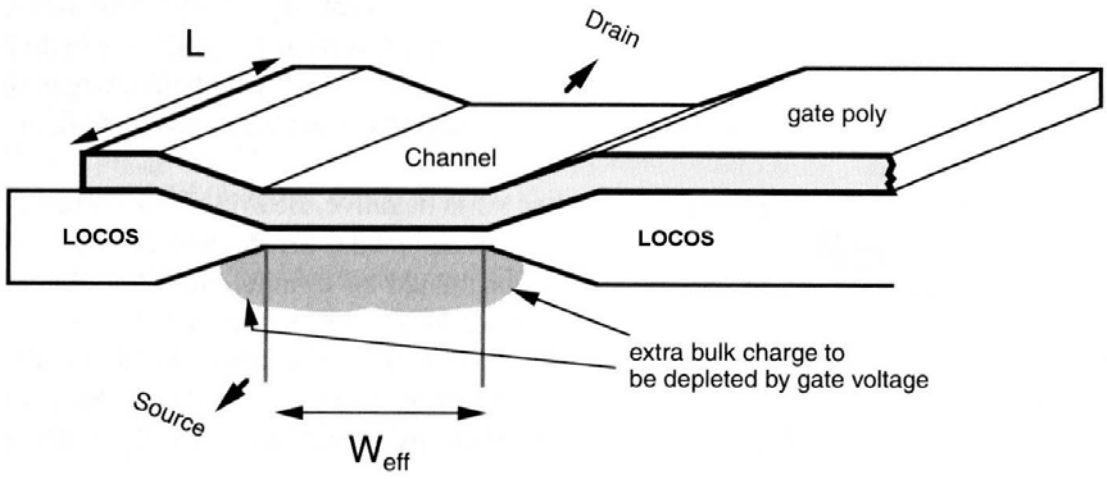

Figure 1.5: CMOS transistor with LOCOS isolation technology (Liu, 2001).

Also the fringing field from the gate that terminates at bulk charges outside the intrinsic portion of the transistor (as defined by W_{eff}), as shown in Figure 1.5, also causes higher V_t. This is because it takes a larger gate voltage to deplete the extra bulk charges before an inversion layer can be formed.

NWE effect on V_t is modeled in BSIM3v3 by using Equation 10,

$$\Delta V_{t,\,nwe} = \left(P_{k3} + P_{k3b} V_b\right) \frac{P_{tox}}{W_{eff} + P_{w0}} 2\phi_f \qquad (10)$$

where P_{k3} is the NWE parameter, P_{k3b} is the body effect NWE parameter, P_{tox} is the oxide thickness parameter, P_{w0} is the channel width offset NWE parameter and W_{eff} is

the effective channel width. The default value of P_{k3} = 80, P_{k3b} = 0 and P_{w0} = 2.5 x 10^{-6} m.

NWE only happens when LOCOS isolation is used. From quarter micron technology and below, the LOCOS isolation has been replaced by STI. STI process causes the opposite effect on V_t, which is known as INWE. Hence, P_{k3} should be set to a smaller value, such as 0.001 to make this term negligible in the final V_t calculation. Unfortunately, the INWE that decreases V_t for narrow width transistors is not captured in either BSIM3v3 or BSIM4.

1.3.5 Inverse Narrow Width Effect (INWE)

INWE causes the V_t to decrease as W decreases. INWE happens when STI technology is used. The STI profile causes thinner gate oxide at the edge of the channel and hence smaller V_t, as shown in Figure 1.6. When W reduces, the thinner gate oxide at the edge of the channel causes the entire channel to have lower V_t.

The field lines from the gate electrode that are focused by the sharp geometry of the channel edge also lower the V_t. This is because at the edges of the channel, an inversion layer is formed at a lower voltage than at the center. As a result, the average V_t of the channel is lower in the transistors with smaller W. As mentioned in the previous section, INWE is not modeled in both BSIM3v3 and BSIM4. Therefore, binning or global binning approach must be used.

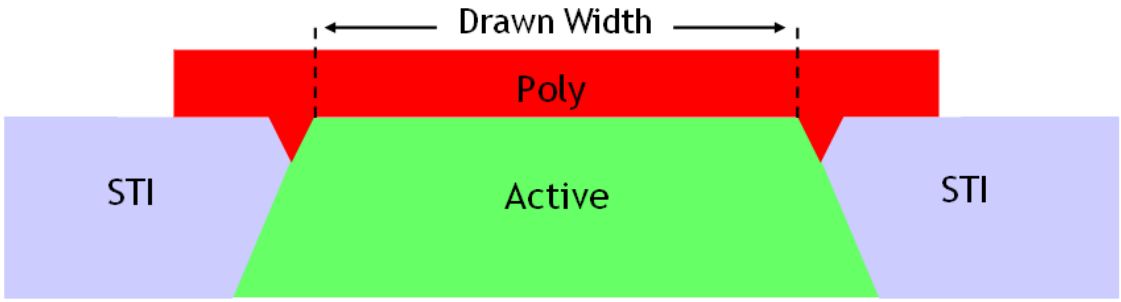

Figure 1.6: CMOS transistor with STI isolation technology.

1.3.6 Small Size Effect (SS) on Threshold Voltage

SS in CMOS transistors is basically similar to the SCE. Instead of decrease in the transistor V_t as L is reduced from SCE, the SS correct the short channel V_t for narrow width transistor by increasing the V_t.

SS on V_t is modeled in BSIM3v3 by using Equation 11,

$$\Delta V_{t,\,ss} = P_{dvt0w}\left[e^{\left(-P_{dvt1w}\frac{W_{eff}L_{eff}}{2L_{tw}} \right)} + 2e^{\left(-P_{dvt1w}\frac{W_{eff}L_{eff}}{L_{tw}} \right)} \right](V_{bi} - 2\phi_f) \qquad (11)$$

where $L_{tw} = \sqrt{\frac{\varepsilon_s X_{dep}}{C_{ox}}}\left(1 + P_{dvt2w}V_b\right)$ and

P_{dvt0w} is the first SS parameter, P_{dvt1w} is the L dependence exponent SS parameter and P_{dvt2w} is the body bias SS parameter. The default value of P_{dvt0w} = 0, P_{dvt1w} = 5.3e6 m^{-1} and P_{dvt2w} = -0.032 V^{-1}.

1.3.7 Drain Induced Barrier Lowering (DIBL)

The potential barrier at the surface between the source and drain for NMOS transistor is shown in Figure 1.7. Under off conditions, the potential barrier in the p-type region prevents electron from flowing to the drain. The surface potential is mainly controlled by the V_g.

When the V_g is below V_t, there are only a limited number of electrons injected from the source over the barrier and flow into the drain. This is called subthreshold current. In the long channel case, the potential barrier is flat over most of the transistor because the source and drain fields only affect the very ends of the channel.

As L become shorter, the source and drain fields penetrate deeply into the middle of the channel, which lowers the potential barrier between source and drain. This causes a substantial increase of the subthreshold current. In other words, the V_t

becomes lower than that of the long channel value. The region of maximum potential barrier also shrinks to a single point near the center of the transistor.

When a high V_d is applied to a short channel transistor, the barrier height is lowered even more, resulting in further V_t decrease. The maximum point of the barrier also shifts toward the source end as shown in Figure 1.7. This effect is referred to as DIBL. It explains the experimentally observed increase of subthreshold current with V_d in short channel transistors.

The conduction band profile of the DIBL effect on short channel transistor is shown in Figure 1.8. The electrons in the source region, which is at lower potential flow to the drain region (at higher potential) after overcoming the gate controlled barrier. Large V_d lowers the barrier and thus less gate voltage is needed to provide the electrons enough energy to overcome the barrier and flow to the drain region.

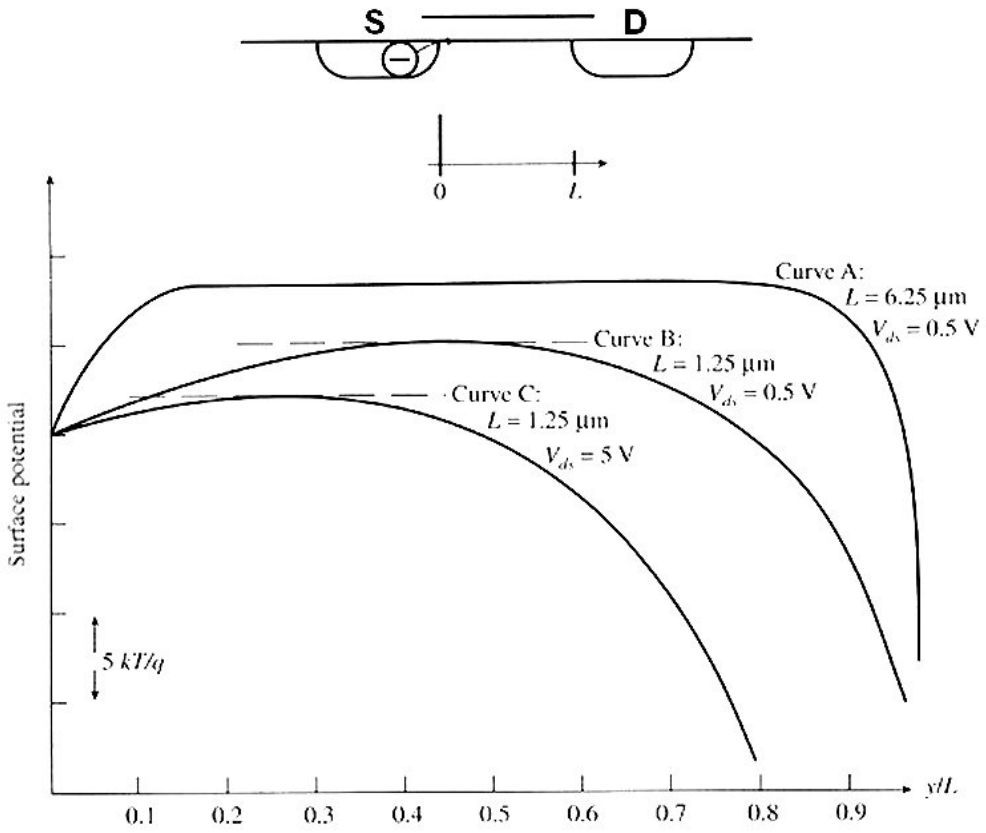

Figure 1.7: Surface potential versus L for (a) long channel transistor (b) short channel transistor, and (c) short channel transistor at high V_d (Taur and Ning, 1998).

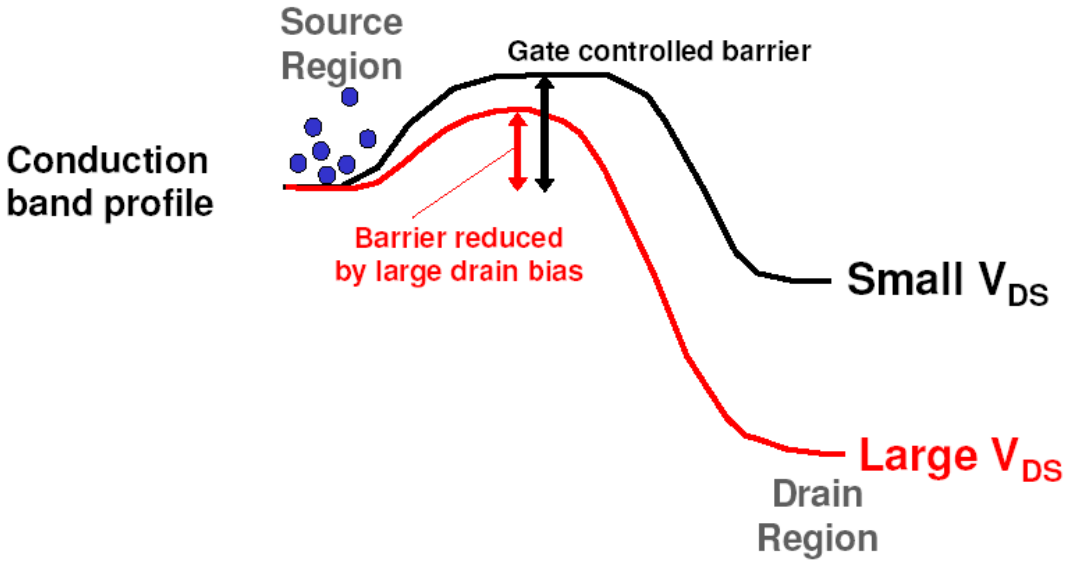

Figure 1.8: Conduction band profile showing DIBL effect in short channel transistor (Yeo, 2005).

DIBL effect on V_t is modeled in BSIM3v3 by using Equation 12,

$$\Delta V_{t,\,dibl} = \left[e^{\left(-P_{dsub}\frac{L_{eff}}{2L_{t0}}\right)} + 2e^{\left(-P_{dsub}\frac{L_{eff}}{L_{t0}}\right)} \right]\left(P_{eta0} + P_{etab}V_b\right)V_d \qquad (12)$$

where $L_{t0} = \sqrt{\frac{\varepsilon_s X_{dep,0}}{C_{ox}}}$ and

P_{dsub} is the V_t L dependence exponent DIBL parameter, P_{eta0} is the V_t DIBL parameter and P_{etab} is the V_t body bias DIBL parameter. The default value of P_{dsub} = 0.56, P_{eta0} = 0.08 and P_{etab} = -0.07 V^{-1}. $X_{dep,0}$ is the depletion thickness in the substrate at zero bulk bias.

1.3.8 Channel Length Modulation (CLM)

The saturation of I_d, I_{dsat} can be understood from the inversion charge density. When V_d is small (linear region), the inversion charge density at the drain end of the channel is only slightly lower than the source end, as shown in Figure 1.9(a).

As the V_d increases (for a fixed V_g), the I_d increases, but the inversion charge density at the drain decreases until finally it goes to zero when $V_d = V_{dsat} = V_g - V_t$. At this voltage, I_d reaches its maximum value. In other words, the surface channel vanishes at the drain end of the channel when saturation occurs. This is called pinch-off (indicated by Y) and is illustrated in Figure 1.9(b). This is a simplified diagram.

When V_d increases beyond the saturation, the pinch-off point moves toward the source, but the I_d remains essentially the same. This is because for $V_d > V_{dsat}$, the voltage at the pinch-off point remains at V_{dsat} and the current stays the same apart from a slight decrease in L (to L'), as shown in Figure 1.9(c). This phenomenon is called Channel Length Modulation (CLM) that makes the channel look as if it is shorter.

Since I_d increases when L reduces, the CLM increases the I_{dsat}. This effect can be seen from the plots of I_d-V_d curves. CLM and DIBL effect on I_d is modeled in BSIM3v3 by using Early Voltage, V_a as shown in Equation 13,

$$I_d = I_{d,0}\left(1 + \frac{V_d - V_{dsat}}{V_a}\right) \tag{13}$$

where $V_a = V_{a,sat} + V_{a,clm\_dibl}$.

$I_{d,0}$ is the ideal long channel I_d as discussed in Equation 3. Early Voltage, V_a is a variable that adjusts the slope of the variation of I_d with respect to V_d. Graphically, can be taken to be value on the –x axis for which the current extrapolates to 0 A, as shown in Figure 1.10.

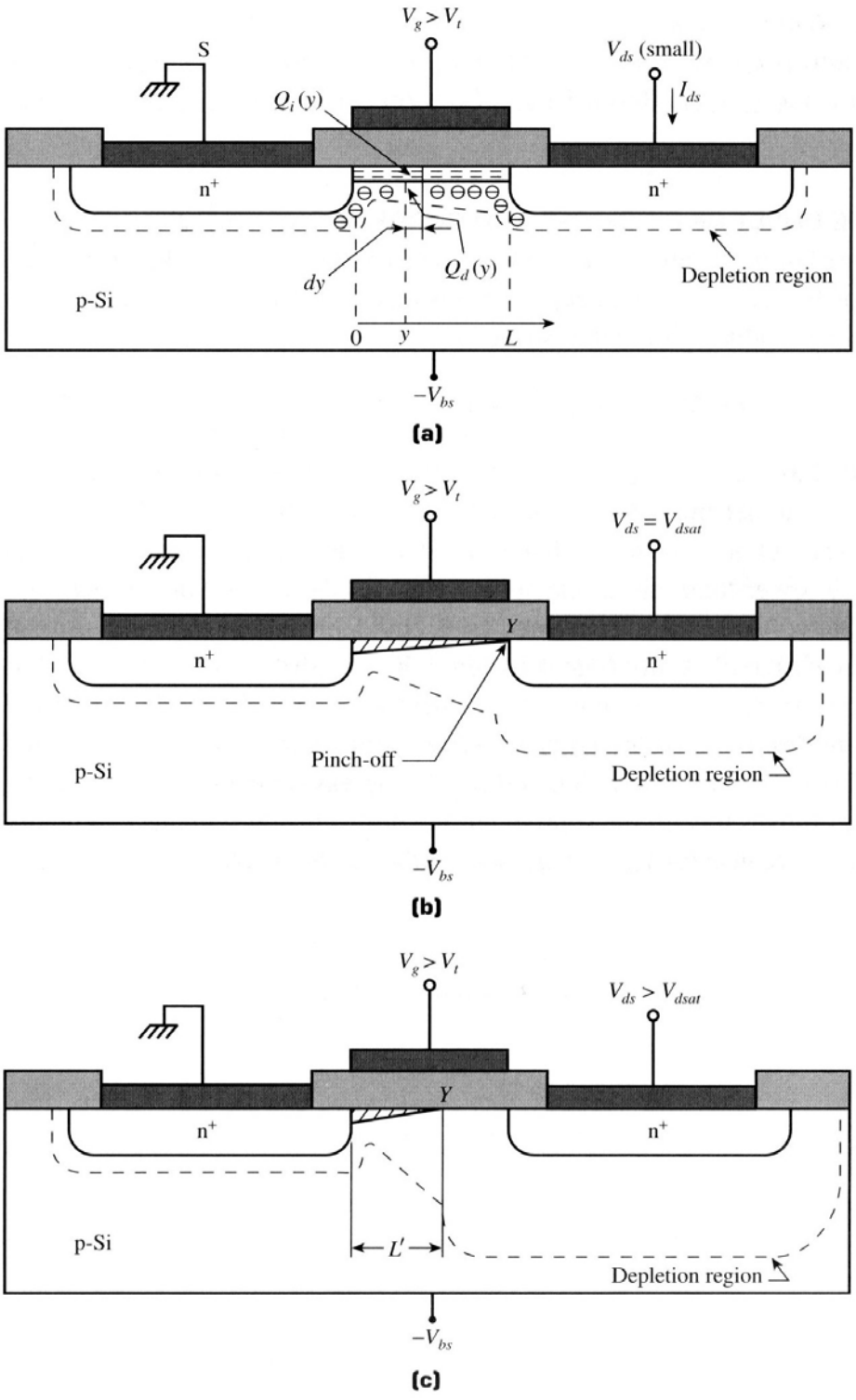

Figure 1.9: CMOS transistor operates in; (a) linear region, (b) onset of saturation, (c) beyond saturation where CLM occurs (Taur and Ning, 1998).

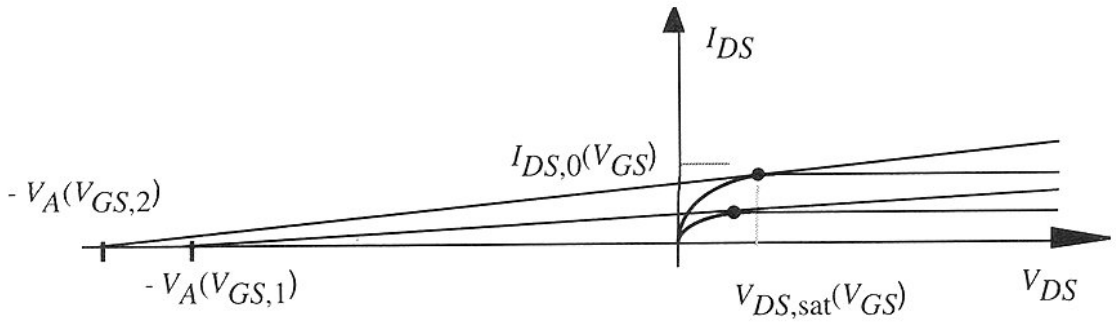

Figure 1.10: I_d-V_d curves with $V_{g,2} > V_{g,1}$, showing the relationship between V_a and the slope of I_d in saturation (Liu, 2001).

$V_{a,sat}$ and V_{a,clm_dibl} equations are shown in the following Equation 14 and Equation 15:

$$V_{a,\,sat} = f_0\left(L_{eff}, P_{vsat}, R_{ds}, \ldots\right) \tag{14}$$

$$V_{a,\,clm\_dibl} = \frac{1 + P_{pvag}\left(V_g - V_t\right)\dfrac{\mu_{eff}}{2P_{vsat}L_{eff}}}{P_{pclm}\,f_1 + \theta_{rout}\left(1 + P_{pdiblcb}V_b\right)f_2} \tag{15}$$

where $\theta_{rout} = P_{pdiblc1}\left[e^{\left(-P_{drout}\frac{L_{eff}}{2L_{t0}}\right)} + 2e^{\left(-P_{drout}\frac{L_{eff}}{L_{t0}}\right)}\right] + P_{pdiblc2}$

and $L_{t0} = \sqrt{\dfrac{\varepsilon_s X_{dep,0}}{C_{ox}}}$

f_0, f_1 and f_2 represent complicated functions and other variables that are covered in this discussion. P_{pvag} is gate bias dependence V_a parameter, P_{vsat} is carrier saturation velocity parameter. P_{pclm} is channel length modulation parameter, $P_{pdiblc1}$ is first DIBL correction on V_a parameter, $P_{pdiblc2}$ is the second DIBL correction on V_a parameter, $P_{pdiblcb}$ is the body bias DIBL correction on V_a parameter. The default value of P_{pvag} = 0, P_{vsat} = 8 x10^4 ms^{-1}, P_{pclm} = 1.3, $P_{pdiblc1}$ = 0.39, $P_{pdiblc2}$ = 0.0086 and $P_{pdiblcb}$ = 0.39.

1.3.9 Velocity Saturation

Carrier velocity saturation is another important parameter that affects the characteristics of short channel transistors. In long channel transistors, the I_{dsat} is explained by the concept of channel pinch-off. For short channel transistors, the I_{dsat} is explained by velocity saturation.

When the V_d increases in a long channel transistor, the I_d first increases, and then becomes saturated at a voltage with the onset of pinch-off at the drain. In a short channel transistor, the I_{dsat} occurs at a much lower voltage due to velocity saturation, as shown in the I_d-V_d curves in Figure 1.11. This is because at high (horizontal) field strength, the velocity of the carriers tends to saturate due to scattering effects (collisions suffered by the carriers). The transistor W/L is 9.5 μm/0.25 μm.

This velocity saturation effect is modeled in BSIM3v3 using P_{vsat} parameter. This parameter when used in NMOS, denotes the electron saturation velocity and when used in PMOS, denotes the hole saturation velocity. Therefore, P_{vsat} in NMOS should be slightly larger than the P_{vsat} in PMOS. P_{vsat} is a critical parameter determining the transistor current in short channel transistors, but has negligible impact on long channel transistors.

This agrees with the physical effects that in long channel transistors, the electric field parallel to the current conduction is small and the electron drift velocity is roughly equal to the mobility times the field. The drift velocity never reaches a magnitude comparable to P_{vsat}. In short channel transistors, in contrast, the field is large enough such that the carriers travel at the saturation velocity in a sizable portion of the channel. P_{vsat} also affects V_a, as shown in Equation 15 (in the previous section).

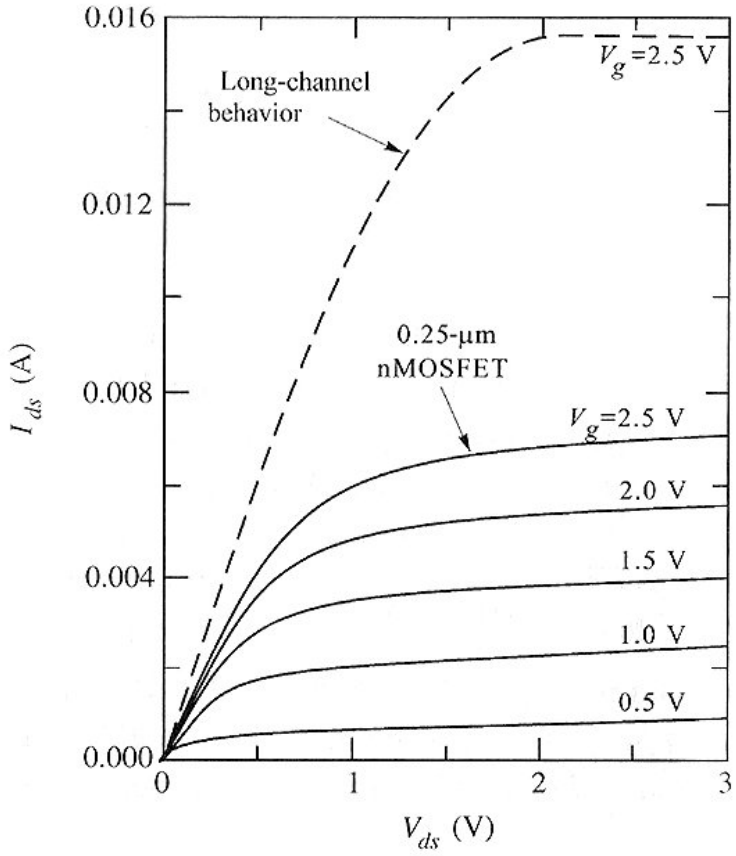

Figure 1.11: I_d-V_d characteristics with velocity saturation, solid lines and as if there were no velocity saturation, dashed lines (Taur and Ning, 1998).

1.3.10 Subthreshold Conduction

The subthreshold conduction in Figure 1.12 shows the transition of a CMOS transistor state from “on” to “off”. In an ideal case, the subthreshold slope should be infinity. Typical subthreshold slope is about 80 to 100 mV/dec.

When the subthreshold slope is degraded, this usually means that the transistor is harder to turn “off” because the gate loses control over the channel surface potential. The subthreshold slope degradation also reduces the V_t and causes higher I_{off}.

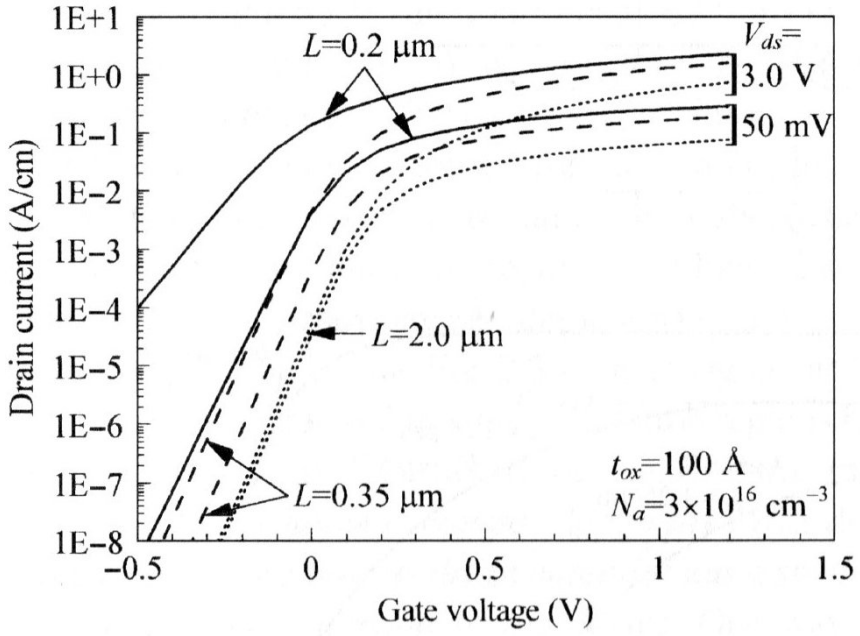

Figure 1.12: Subthreshold characteristics of short and long channel transistors at low and high V_d (Taur and Ning, 1998).

Subthreshold conduction effect on I_d at off condition ($V_g < V_t$), I_{off} is modeled in BSIM3v3 by using Equation 16,

$$I_{off} = \mu C_{ox} \frac{W}{L} \left(\frac{kT}{q} \right)^2 e^{\left(-\frac{qV_t}{\eta kT} \right)} e^{\left(-\frac{qP_{voff}}{\eta kT} \right)} \tag{16}$$

where P_{voff} is the subthreshold offset voltage parameter. The default value of P_{voff} = -0.08 V. q is the single electron charge (1.6e-19 C). From Equation 16, I_{off} is proportional to ideality factor, η (Equation 17),

$$I_{off} \propto e^{\left(\frac{qV_g}{\eta kT} \right)} \tag{17}$$

The ideality factor, η is given in Equation 18,

$$\eta = 1 + \left(P_{nfactor} \frac{C_{dep}}{C_{ox}} \right) + \frac{P_{cit}}{C_{ox}} + \left[\left(\frac{P_{cdsc} + P_{cdscd} V_d + P_{cdscb} V_b}{C_{ox}} \right) \left(e^{\left(-P_{dvt1} \frac{L_{eff}}{2 L_t} \right)} + 2e^{\left(-P_{dvt1} \frac{L_{eff}}{L_t} \right)} \right) \right] \quad (18)$$

where $L_t = \sqrt{\frac{\varepsilon_s X_{dep}}{C_{ox}}} \left(1 + P_{dvt2} V_b\right)$ and

$P_{nfactor}$ is subthreshold turn-on swing parameter, P_{cit} is the interface trap capacitance parameter (this parameter is purely a DC parameter which does not affect the C-V characteristics of the transistor), P_{cdsc} is drain and source to channel coupling capacitance parameter (a pure DC parameter), P_{cdscd} is drain bias sensitivity of P_{cdsc} parameter (a pure DC parameter), P_{cdscb} is body bias sensitivity of P_{cdsc} parameter (a pure DC parameter). The default value of $P_{nfactor}$ = 1, P_{cit} = 0 Fm^{-2}, P_{cdsc} = 2.4e-4 Fm^{-2}, P_{cdscd} = 0 $FV^{-1}m^{-2}$ and P_{cdscb} = 0 $FV^{-1}m^{-2}$. C_{dep} is the depletion capacitance per unit area in the bulk.

The last term is similar to the one used in SCE for V_t calculation. It is meant for the short channel transistors. Usually to model the subthreshold current, $P_{nfactor}$ and P_{voff} are used. Other parameters can set to their default values, rule of thumb is the ideality factor, η should be close to "1".

1.3.11 Field Dependent Mobility

Mobility is a key parameter in transistor modeling. It is a measure of the ease of carrier motion in semiconductor materials. Carrier transport in a MOS transistor mainly takes place along the interface between silicon, Si and silicon dioxide, SiO_2. The carrier mobility at the interface is lower than in the bulk and depends on both vertical and horizontal electric field.

In the early years of CMOS, the gate oxide was thick, and the vertical electric field induced by the gate bias was low and hence the influence of the vertical electric field could be ignored. But in today's short channel transistors, oxide thicknesses are very thin, for example in 0.13 μm technology the oxide thickness is about 2 nm. Thus, the influence of the vertical electric field is strong and the carrier mobility is not constant as the gate bias changes.

Three scattering mechanisms have been proposed to account for the dependence of mobility on the vertical component of the electric field. They are phonon scattering, coulomb scattering and surface roughness scattering. Each scattering mechanism is dominant under some specific conditions of the doping concentration, temperature and bias. For good quality interfaces, phonon scattering is generally the dominant scattering mechanism at room temperature, as shown in Figure 1.13.

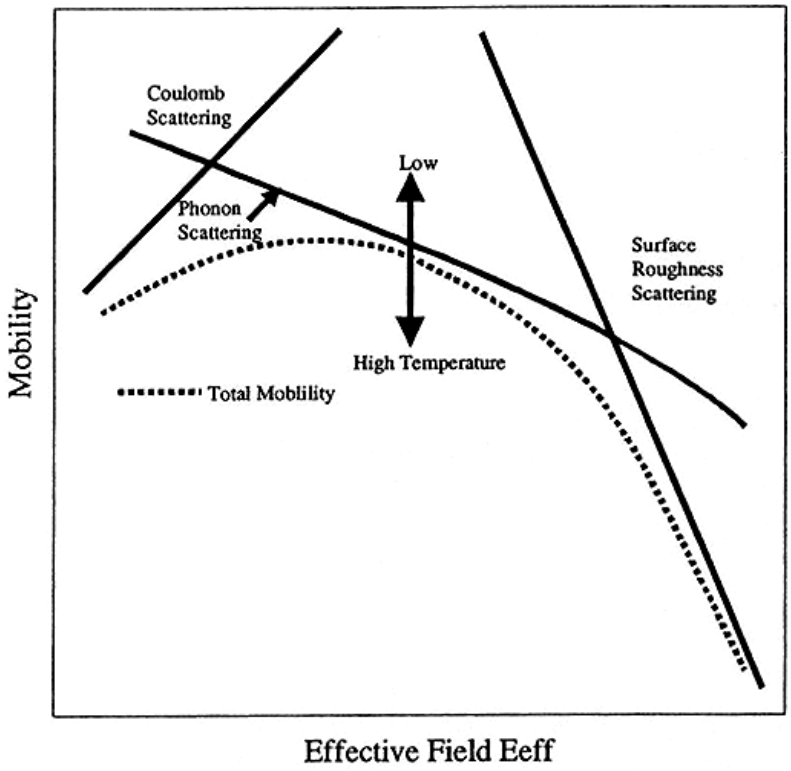

Figure 1.13: Mobility versus Vertical Effective Field to illustrate how the mobility behavior is dominated by different scattering mechanisms (Cheng and Hu, 1999).

In BSIM3v3, the effective carrier mobility, μ_{eff}, of mobility mode parameter, P_{mobmod} = 1, is given in Equation 19,

$$\mu_{eff} = \frac{P_{u0}}{1+\left(P_{ua}+P_{uc}V_b\right)\left(\frac{V_g+V_t}{P_{tox}}\right)+P_{ub}\left(\frac{V_g+V_t}{P_{tox}}\right)^2} \tag{19}$$

where P_{u0} is the zero field universal mobility parameter, P_{ua} is the first order mobility degradation parameter, P_{ub} is the parabolic mobility degradation parameter and P_{uc} is the body bias mobility degradation parameter. The default value of P_{u0} = 0.067 $m^2V^{-1}s^{-1}$ (NMOS) and 0.025 $m^2V^{-1}s^{-1}$ (PMOS), P_{ua} = 2.25 x 10^{-9} mV^{-1}, P_{ub} = 5.87 x 10^{-19} m^2V^{-2} and P_{uc} = -4.65 x 10^{-11} mV^{-2}.

1.3.12 Substrate Current induced Body Effect (SCBE)

At saturation mode, the high electric field near the drain region causes impact ionization of carriers. For NMOS transistor, the generated electrons are swept into the drain whereas the holes flow into the substrate. This phenomenon is known as SCBE. SCBE results in I_d increase that is many times larger than I_{sub}, as shown in Figure 1.14.

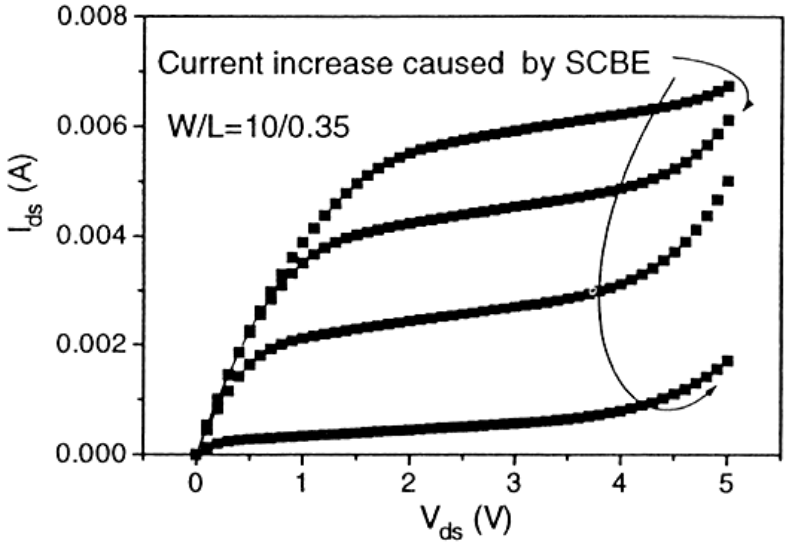

Figure 1.14: I_d-V_d characteristics due to SCBE (Cheng and Hu, 1999).