This document is downloaded from DR-NTU (https://dr.ntu.edu.sg)
Nanyang Technological University, Singapore.

Effects of mechanical stress on the performance of metal-oxide-semiconductor transistors

Yang, Peizhen

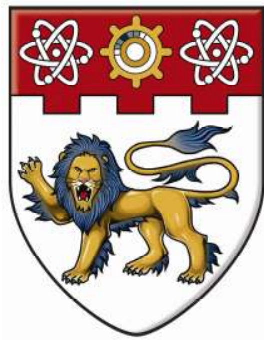
2012

Yang, P. (2012). Effects of mechanical stress on the performance of metal-oxide-semiconductor transistors. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/50306

https://doi.org/10.32657/10356/50306

Downloaded on 24 Aug 2022 21:10:27 SGT



NANYANG
TECHNOLOGICAL
UNIVERSITY

EFFECTS OF MECHANICAL STRESS ON THE PERFORMANCE OF METAL-OXIDE-SEMICONDUCTOR TRANSISTORS

YANG PEIZHEN

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

2012

EFFECTS OF MECHANICAL STRESS ON THE PERFORMANCE OF METAL-OXIDE-SEMICONDUCTOR TRANSISTORS

YANG PEIZHEN

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2012

ACKNOWLEDGEMENT

First and foremost, I would like to thank my thesis advisor, Assoc Prof Lau Wai Shing for his guidance, encouragement, and continuous support throughout my research. I am also indebted to Assoc Prof Chen Tupei for his assistance in the final stage of my postgraduate studies.

I would like to acknowledge the staff of GLOBALFOUNDRIES SINGAPORE Pte Ltd for providing the wafers for this project. In particular, I would like to thank the engineers and managers whom I have personally worked with: Siah Soh Yun, Lai Seow Wei, Patrick Lo, Tang Yee Ling, Lim Eng Hua, Eng Chee Wee, Vincent Ho, Loh Cheng Hou, Lim Boon Kiat, Toh Liew Feng, Liu Yang, Jacob Wang and See Kwang Seng.

My sincere gratitude goes to Asst Prof Sridhar Idapalapati and Mr Leo Cheng Seng from the School of Mechanical and Aerospace Engineering (MAE) for their assistance in the design and calibration of the four point rod bending fixture. My fellow coursemate, Teo Zhiqiang Andy was a great help with the construction and calibration of the four point rod bending fixture. I would like to specifically thank Mr Lim Meng Keong from the School of Materials Science and Engineering (MSE) for his assistance in the wafer dicing, which is essential for my four point rod bending experiments.

My research at NTU would not be possible without the funding from GLOBALFOUNDRIES SINGAPORE Pte Ltd and Singapore Economic Development Board (EDB). The weekly meeting has provided a unique opportunity to exchange ideas and receive timely comments and advice. Last, but not the least, my parents receive my deepest gratitude and love for their dedication and many years of support. My siblings, Sheryl and Zhiren were a continuous source of love and I wish them success in their endeavors and happiness throughout their life. I am also indebted to my husband, Justin for his extraordinary patience, understanding and love during the past few years.

ABSTRACT

Stress engineering is widely used in the microelectronics industry to improve the on-current (I_{on}) performance of the metal-oxide-semiconductor (MOS) transistors through the strain-induced mobility enhancement. However, there are still debates regarding the relevance of the low-field mobility in the saturation drain current of the nanoscale MOS transistors. Based on velocity saturation model, the high-field velocity is independent of the low-field mobility. In the other words, velocity saturation model predicts that mobility enhancement techniques will not improve I_{on} of the nanoscale MOS transistors. Ballistic transport model considers an ideal situation where the channel carriers do not experience any scattering when they transit from the source to the drain. Since mobility is a concept that involves channel scattering, ballistic transport regards mobility as irrelevant in the nanoscale MOS transistors. In quasi-ballistic transport model, channel carriers will undergo a number of channel scatterings before reaching the drain. Hence, quasi-ballistic transport model is able to account for the strain-induced I_{on} improvement in nanoscale MOS transistors. However, the saturation drain current equation of a transistor in the quasi-ballistic model comprises parameters that are not properly defined. Furthermore, some researchers managed to use velocity saturation model to fit the saturation current of the nanoscale MOS transistor. By improvising Lundstrom's 1997 theory on the quasi-ballistic transport and unifying the merits of existing transport models, we arrive at a simplified saturation drain current equation for nanoscale MOS transistors.

Most research in stress engineering focus on the strain-induced I_{on} improvement, but disregard the effects of the mechanical stress on the off-current (I_{off}). Using externally applied mechanical stress, we can isolate the effects of process variations from the strain-induced effects from the strain-induced effects on I_{off}. We studied the physics behind the strain-induced increase in the subthreshold I_{off} and found that the strain-induced change in the quantum mechanical confinement decreases the subthreshold swing (S_{ts}) of NMOS transistor but increases S_{ts} of p-channel MOS (PMOS) transistor. It is well-known that uniaxial tensile stress leads to electron mobility enhancement and a reduction in the threshold voltage of NMOS transistor. Since our experimental results show that uniaxial tensile stress can increase the subthreshold I_{off} of NMOS transistor, the effects of strain-induced reduction in threshold voltage and the strain-induced mobility enhancement will dominate over the strain-induced improvement in the subthreshold swing. To extend the study to the process-induced stress, we need to reduce the effects of process variation on I_{off}. In order to reduce the effects of wafer-to-wafer variation on I_{off}, we intentionally use

consecutive wafers with different amount of process-induced stress. In addition, we also identify NMOS transistors whose I_{off} is less sensitive to the die-to-die variation. Our experimental results show that CESL-induced tensile stress can increase both I_{on} and the subthreshold I_{off} of NMOS transistors. In fact, we observed that CESL-induced tensile stress has led to a bigger percentage increase in $\log I_{off}$ as compared to I_{on} in NMOS transistor. This makes us wonder if stress engineering can lead to an overall I_{on} improvement. The most straightforward approach to ascertain the overall strain-induced I_{on} improvement is to determine the strain-induced increase in $\log I_{off}$, and then obtain the corresponding increase in I_{on} from the $\log I_{off}$ versus I_{on} plot characteristics plot. However, this approach does not explain why there is an overall strain-induced improvement in I_{on} when stress engineering actually leads to a bigger percentage increase in $\log I_{off}$ compared to that of I_{on}. Hence, we introduce a third parameter ($V_{th,sat}$ or DIBL) in the $\log I_{off}$ versus I_{on} characteristics. We found that the subthreshold I_{off} is more sensitive to the change in $V_{th,sat}$ as compared to I_{on} and thus the strain-induced increase in I_{off} can be removed by a slight change in $V_{th,sat}$ without much effect on I_{on}. This leads to an overall improvement strain-induced improvement in I_{on} even though stress engineering actually leads to a bigger percentage increase in $\log I_{off}$ compared to that of I_{on}. Similar analysis can be applied by introducing DIBL as the third parameter in the $\log I_{off}$ versus I_{on} characteristics.

Apart from the conventional <110> channel orientation on (100) surface-oriented silicon wafer, transistors can also be fabricated on 45° rotated (100) surface-oriented wafer with <100> channel orientation. We have evaluated the impact of the change in channel orientation together with mechanical stress on the performance of NMOS transistors and PMOS transistors. We found that NMOS transistors are not improved by switching from the conventional <110> channel orientation to <100> channel orientation but they can be improved by tensile stress. PMOS transistors are improved by switching from the conventional <110> channel orientation to <100> channel orientation but PMOS transistors with <100> channel orientation are not sensitive to mechanical stress.

CONTENTS

ACKNOWLEDGEMENT i

ABSTRACT ii

LIST OF FIGURES vi

LIST OF TABLES xiv

1. Introduction 1

1.1 Background 1
1.2 Roles of carrier mobility in drain current transport 3
1.3 Mobility enhancement techniques 5
1.4 Organization of the Thesis 7

2. Literature review on Stress Engineering 8

2.1 Methods to introduce channel stress 8
2.1.1 Substrate-induced stress 8
2.1.2 Process-induced stress 9
2.1.3 Externally applied mechanical stress 15
2.2 Strain-induced mobility enhancement 18
2.2.1 First Brillouin zone 18
2.2.2 *E-k* diagram of unstrained bulk silicon 20
2.2.3 Relationship between the effective mass and the *E-k* diagram 23
2.2.4 Effects of mechanical stress on the conduction band minimum 24
2.2.5 Effects of mechanical stress on the valence band maximum 27
2.2.6 Low-field mobility: Drude model 30
2.3 Usage of Piezoresistance coefficients 32

3. Drain current transport in nanoscale MOS transistors 36

3.1 Velocity saturation model 36
3.2 Ballistic transport model 39
3.3 Quasi-ballistic transport 48
3.4 Semi-empirical model 51
3.5 Results and Discussion 53
3.5.1 Temperature dependency of the drain current 53
3.5.2 Saturation drain current equation 56
3.5.3 Physics behind the apparent velocity saturation 67

4. Effects of mechanical stress on the electrical parameters of MOS transistors 68

4.1 On-current (I_{on}) ...68
4.2 Threshold voltage...69
4.2.1 Strain-induced effects on silicon bandgap... 70
4.2.2 Equations for strain-induced change in the threshold voltage... 71
4.3 Gate leakage current...72
4.4 Gate-induced-drain leakage (GIDL) current ...74
4.5 Gate oxide thickness ...75
4.5.1 Poisson's ratio... 75
4.5.2 Quantum confinement ... 76
4.6 Subthreshold off-current (I_{off}) ...80
4.7 Results and Discussion...83
4.7.1 Physics behind the linearity in the log I_{off} versus I_{on} characteristics... 83
4.7.2 Four-point bending measurement ... 90
4.7.3 Tensile stressed contact etch stop layer (CESL)... 93
5. Impact of the change in channel orientation on performance...103
5.1 Impact of channel orientation on hole mobility ...104
5.2 Limited studies on the impact of channel orientation on electron mobility ...105
5.3 Experimental evidence of an increase in L_{eff} by channel orientation ...108
5.4 Mechanism behind the increase in L_{eff} by channel orientation...109
5.4.1 Impact of channel orientation on diffusion... 109
5.4.2 Impact of channel orientation on ion channeling... 111
5.5 Results and Discussion...114
5.5.1 Validity of the six degenerate ellipsoid model ... 114
5.5.2 Mechanism behind the increase in L_{eff} of <100> transistors... 118
5.5.3 Effects of channel orientation on I_{on} and I_{off} of MOS transistors... 121
6. Conclusions and Recommendation ...125
6.1 Main contributions of the thesis...125
6.2 Recommendations of future work ...126
List of publications ...128
Appendix: Effects of surface-orientation on electron mobility and hole mobility...131
Bibliography...134

LIST OF FIGURES

Figure 1-1 Moore's Law: CPU transistor count has increased by two times and feature size has decreased by 0.7 times in every two years [1.1]. 1

Figure 1-2 Excessive gate-oxide leakage occurs with further scaling of the conventional silicon oxynitride gate dielectric [1.2]. 2

Figure 1-3 Total chip dynamic and static power dissipation trends based on the International Technology Roadmap for Semiconductor (ITRS) [1.3]. 2

Figure 1-4 Experimental correlation between low-field mobility (μ_{eff}) and the high-field carrier velocity (v_{eff}) for the nanoscale NMOS transistors [1.29]. 4

Figure 1-5 Flow chart of the main mobility enhancement techniques [1.30]. 6

Figure 2-1 Cross-section of (a) NMOS transistor, (b) PMOS transistor fabricated on the strained-Si on relaxed Si_{1-x} Ge_y virtual substrate. 8

Figure 2-2 Schematics of (a) the ordinary S/D contacts and (b) the borderless S/D contacts. 10

Figure 2-3 The importance of CESL to prevent the electrical shorting between n^+ S/D and p-well when there is a mask misalignment: (a) Without CESL, (b) With CESL. 11

Figure 2-4 Process flow of CESL as a mechanical stressor. 11

Figure 2-5 Structural model for local bonding arrangement in PECVD SiN_x film for (a) Without UV curing, (b) With UV curing [2.14]. 11

Figure 2-6 Process flow of $Si_{1-y}Ge_y$ S/D stressor for PMOS transistor. 12

Figure 2-7 Process flow of Stress Memorization Technique (SMT). 14

Figure 2-8 Process flow of e-Si:C S/D as a mechanical stressor. 15

Figure 2-9 The use of statistics to compare the hole mobility enhancement by the Diamond-like Carbon (DLC) liner [2.37]. 16

Figure 2-10 Different types of externally applied mechanical stress. 17

Figure 2-11 Schematics of the curvature of the rectangular wafer strips under: (a) uniaxial tensile stress and (b) uniaxial compressive stress. Note that the transistor is on the top surface of the wafer strip. 17

Figure 2-12 Differences between uniaxial tensile stress and biaxial tensile tensile in terms of wafer displacement from equilibrium position [2.43]. 18

Figure 2-13 Schematics of a diamond crystal structure [2.47]. 19

Figure 2-14 Reciprocal lattice of a face-centred cubic lattice with its first Brillouin zone [2.48]. 19

Figure 2-15 Schematics of energy band diagram of silicon [2.61]. 21

Figure 2-16 Schematics of the conduction band valleys of silicon 21

Figure 2-17 Schematics of valence band maximum of silicon [2.62] 21

Figure 2-18 Constant energy surface and cross section of valence band maximum at k_z =0 plane [2.63]. 22

Figure 2-19 Splitting of the conduction band edge (ΔE_c) as a function of (a) biaxial stress, (b) uniaxial <100> stress, (c) uniaxial <110> stress [1.14]. 25
Figure 2-20 Electron mobility enhancement ($\Delta\mu_e/\Delta\mu_e$) as a function of the vertical electric field under (a) biaxial stress, (b) uniaxial <100> stress, (c) uniaxial <110> stress [1.14]. 26
Figure 2-21 Energy contours of the out-of-plane conduction band valleys (Δ_5, Δ_6) on (100) silicon plane under (a) uniaxial <100> tensile stress, (b) uniaxial <110> tensile stress [1.14]. 26
Figure 2-22 The effects of <110> uniaxial tensile stress on: (a) the longitudinal effective mass ($m_{t,//}$) and the transverse effective mass ($m_{t,\perp}$) of the out-of-plane conduction valleys, (b) the effective conductivity mass (m^*) [2.66]. 27
Figure 2-23 Electron mobility enhancement as a function of mechanical stress [1.14]. 27
Figure 2-24 Constant energy contours of (a) the light hole band and (b) heavy hole band at Γ for unstressed bulk silicon [1.15, 2.63]. k_x, k_y and k_z are the wave vectors along x, y and z directions. ..28
Figure 2-25 Constant energy contours separated by 25 meV for the lowest hole energy band of (100) bulk silicon at Γ under (a) the unstressed condition, (b) the uniaxial compressive stress along <110> direction, and (c) the uniaxial tensile stress along <110> direction [1.15]. 29
Figure 2-26 Constant energy contours of the lowest hole energy band at Γ of silicon inversion layer in (100) plane with vertical confinement field of 1 MV/cm under (a) the unstrained condition, (b) the uniaxial compressive stress along <110> channel direction (as indicated by the arrow) [1.15]. .29
Figure 2-27 Phonon dispersion spectrum of silicon along the (100) direction of the first Brillouin ($\Gamma \rightarrow \Delta \rightarrow X$). The f and g phonons participate in the intervalley scattering of electrons [2.71]. 31
Figure 2-28 First Heavy-hole subband equivalent energy lines of (100) surface-oriented silicon under uniaxial <100> tensile stress of 1.5 GPa (Courtesy of Dr Pham from BST, TU Braunschweig, Germany, Ref. [2.44]) 34
Figure 2-29 Non-linearity of piezoresistance coefficient at higher uniaxial stress [2.85]. C-R conversion model is represented by eqn. (2.16). $\Delta\mu/\mu \approx \Delta\rho/\rho \times \pi$ refers to the piezoresistance model, which is given by eqn. (2.13) 35
Figure 3-1 Classification of carrier transport in MOS transistors [1.21]. 36
Figure 3-2 Time-of-flight measurement suggests that the low-field mobility (μ_{eff}) is independent of carrier velocity at high lateral electric field [1.9]. 37
Figure 3-3 Electron velocity overshoot has been observed at room temperature in bulk NMOS transistor with gate length (L) of 32 nm [1.19]. 37
Figure 3-4 The average velocity versus position for electrons injected into a short slab of silicon with a high electric field [3.4]. Dotted line refers to the average velocity of electrons. Solid line refers to the kinetic energy of the electrons. 38
Figure 3-5 Equal spacing in the I_{ds} versus V_{DS} characteristics at high V_{DS} is observed in (a) long-channel NMOS transistor, (b) short-channel NMOS transistor. 39
Figure 3-6 Injection velocity (v_{inj}) as a function of the inversion carrier density [1.21]. 40

Figure 3-7 Temperature dependency of the injection velocity (v_{inj}) and the saturation drain current of the nanoscale NMOS transistors [1.20]. 41
Figure 3-8 Illustration of $V_{T, 2\phi F}$ and $V_{T,extrapol}$ in the I_{ds} versus V_{GS} characteristics[3.14]. 42
Figure 3-9 As opposed to Natori's 1994 theory on ballistic transport, the saturation I_{ds} of the nanoscale NMOS transistor (L = 60 nm) does not follow a $(V_{GS} - V_{th,sat})^{3/2}$ relationship. The symbols indicate the data points that are used for the linear extrapolation. The thicker lines indicate the experimental values of I_{ds}. 42
Figure 3-10 Simulation results of the drain current as a function of the drain-to-source voltage (V_{DS}) for a double-gate NMOS transistor with several channel lengths [3.15]. Solid line for ballistic transport and dashed line for classical diffusive transport. 43
Figure 3-11 Impact of the S/D series resistance (R_{sd}) on the potential profile along the channel of NMOS transistor [3.15]. 43
Figure 3-12 Impact of the S/D series resistance (R_{sd}) on the I_{ds} versus V_{DS} characteristics of a NMOS transistor [3.15]...44
Figure 3-13 Mean free path (λ) as a function of L_{eff} for NMOS transistor [3.18]. Constant gate overdrive of 1.1 V is used. The drain bias is 10 mV. 46
Figure 3-14 Mobility degradation with L_{eff} is observed in NMOS transistors without halo implants [3.20]. 46
Figure 3-15(a) Schematics of neutral defects near S/D regions , (b) Neutral defects recovery by increasing the rapid thermal anneal (RTA) temperature [3.20]. 47
Figure 3-16 Role of long-range Coulomb scattering in nanoscale MOS transistor: (a) Region of influence, (b) Potential fluctuations in the S/D regions [3.22]. Note that r is the distance between the channel carrier and the scattering ion. 47
Figure 3-17 The effective electron temperature in the S/D regions is at 800 K rather than the ambient temperature of 300 K: (a) the carrier velocity distribution, (b) the kinetic energy distribution at location close to the source and drain contacts [3.22]. 47
Figure 3-18 Effects of the long-range Coulomb scattering on the I_{ds} versus V_{DS} characteristics of a nanoscale NMOS transistor [3.22]. 48
Figure 3-19 Definition of the critical length (ℓ) for NMOS transistor [1.25]. 49
Figure 3-20 The average electric field, $\varepsilon(0^+)$ is a function of (a) V_{GS} and (b)V_{DS} [1.25]....................49
Figure 3-21 Schematics of the potential profile along the channel length of a NMOS transistor under quasi-ballistic transport [3.26]. $\Delta V_{overlap}$ is the drop in potential in the source series resistance and the accumulation region under the source-to-gate overlap. $(V_s)_{intrinsic}$ and $(V_D)_{intrinsic}$ are the intrinsic source bias and the intrinsic drain bias, respectively. $\Phi_{barrier}$ is the potential barrier at the source-channel...49

Figure 3-22 The thermal injection velocity ($\tilde{v}_T$) versus the inversion layer density (n_s) [1.25]. For V_{GS} below threshold voltage, $\tilde{v}_T \approx 1.2 \times 10^7$ cm/s. For V_{GS} above threshold voltage, the channel carriers become degenerate and $\tilde{v}_T$ increases....................50
Figure 3-23 Illustration of the virtual source point (x_o) in the NMOS transistor. Q_{ixo} and v_{xo} are defined at the top of the conduction band profile along the channel direction. [1.28]....................52
Figure 3-24 Extracted virtual source velocity (v_{xo}) as a function of the gate length for (a) NMOS transistors, (b) PMOS transistors that are fabricated on (100) Si wafer with <110> channel orientation [1.27]. Stress engineering can increase v_{xo} to beyond v_{sat}....................52
Figure 3-25 Extracted virtual source velocity (v_{xo}) and effective carrier velocity (v) versus DIBL for the nanoscale NMOS transistors and PMOS transistors [1.27]. 52
Figure 3-26 Temperature independent point (TIP) of the long-channel PMOS transistor with nominal gate length of 1 μm (a, b), and the short-channel PMOS transistor with nominal gate length of 60 nm (c,d). V_{DD} is -1.2 54
Figure 3-27 Gate length dependency of temperature independent point (TIP) for (a) NMOS transistor, (b) PMOS transistor. V_{DD} is 1.2 V. 54
Figure 3-28 NMOS transistors (fabricated by 65 nm CMOS technology) in the SCE regime: (a) TIP decreases with increasing V_{DS}, (b) V_{th} decreases with increasing V_{DS}. Note that DIBL = $V_{th,lin}$ -$V_{th,sat}$. 55
Figure 3-29 Temperature sensitivity in I_{on} and I_{eff} of NMOS transistors as a function of L. 56
Figure 3-30 Simulation shows that threshold adjustment implant can increase TIP [3.34]....................56
Figure 3-31 Effects of R_{sd} on E_c of a NMOS transistor in saturation operation. 59
Figure 3-32 Effects of CESL-induced tensile stress on: (a) μ_{eff} versus V_{GS} characteristics, (b) v_{eff} versus V_{GS} characteristics of a NMOS transistor with nominal gate length of 60 nm. The drain bias is 1.2 V. 59
Figure 3-33 Effects of temperature on v_{sat_eff} and v_{eff} of a NMOS transistor with nominal gate length of 60 nm. Note that v_{sat_eff} refers to the average value of v_{eff} when V_{GS} is close to V_{DD}. 60
Figure 3-34 Effects of the scattering mechanism on the μ_{eff} versus V_{GS} characteristics....................63
Figure 3-35 Schematic diagram showing the relationship of v_1, v_2 and v_{eff} with V_{GS}....................63
Figure 3-36 Extraction of V_{th,sat_IV} from the saturation I_{ds} versus V_{GS} characteristics. 65
Figure 3-37 V_{th,sat_IV} includes a component to overcome the Coulombic scattering by "screening": (a) μ_{eff} versus V_{GS} characteristics, (b) v_{eff} versus V_{GS} characteristics of a NMOS transistor with a nominal gate length of 60 nm. 65
Figure 3-38 Effects of R_{sd} correction on v_{sat_eff} of a nanoscale NMOS transistor at room temperature. 66

Figure 3-39 Effects of transistor scaling on (a) the μ_{eff} versus L characteristics and (b) the v_{sat_eff} versus L characteristics of NMOS transistors. R_{sd} = 0 Ω-µm refers to the case where the S/D series resistance correction is not performed. 67
Figure 4-1 The use of the log I_{off} versus I_{on} plot of NMOS transistors to assess the amount of I_{on} improvement by stress engineering [1.42]. 68
Figure 4-2 Validity of the linearity assumption in the log I_{off} versus I_{on} plot for NMOS transistor [1.39]. 68
Figure 4-3 Biaxial tensile stress leads to negative threshold voltage shifts in<110> MOS transistors on (100) Si [4.1].... 69
Figure 4-4 Uniaxial tensile stress and biaxial tensile stress lead to negative threshold voltage shift in <110> NMOS transistors on (100) Si [4.2].. 70
Figure 4-5 Effects of mechanical stress on the silicon bandgap of (100) Si[4.4]. 70
Figure 4-6 The effects of uniaxial stress on gate tunneling current of <110> MOS transistors on (100) Si surface[4.7,4.8]. 73
Figure 4-7 Valence band edge energy as a function of the uniaxial <110> stress [2.86]. 73
Figure 4-8 (a) Classical GIDL mechanism, (b) New GIDL mechanism[4.9] 74
Figure 4-9 Effects of threshold adjustment implant on the GIDL current of a nanoscale NMOS transistor (L = 40 nm)[4.9]. HVT: High dose, RVT: Regular dose, LVT: Low dose. 74
Figure 4-10 Effects of Poisson's ratio under (a) tensile stress, (b) compressive stress [4.11]. 75
Figure 4-11 Triangular potential well approximation. 77
Figure 4-12 Graphical representation of Airy functions, Ai(x) and Bi(x) [4.19]. 78
Figure 4-13 Effects of compressive stress on the C-V characteristics of PMOS capacitor [4.25].....79
Figure 4-14 Subthreshold current is taken to be the linear portion of log I_{ds} versus V_{GS} characteristics of NMOS transistor. 80
Figure 4-15 Magnetoresistance (MR) mobility measurements show that there is a correlation between μ_{sub} and μ_{eff} [4.29]. 81
Figure 4-16 Effects of externally applied tensile stress on the I_{ds} versus V_{GS} characteristics of a relatively long NMOS transistor with 4 nm conventional gate dielectric [4.30]. Inset shows the threshold voltage shift in the subthreshold regime. 81
Figure 4-17 Effects of CESL-induced compressive stress on the subthreshold current of <110> PMOS transistor with L = 0.55 µm and W = 10 µm [4.25]. 81
Figure 4-18 Linearity of logI_{off} versus I_{on} plot is valid for the short channel effect (SCE) regime of transistors fabricated by 0.11 µm CMOS technology. (a) & (b) refer to NMOS transistors. (c) & (d) refer to PMOS transistors. 84
Figure 4-19 For NMOS transistors in the SCE regime: (a) v_{sat_eff} has a good linearity with DIBL, (b) V_{th,sat_IV} has a good linearity with DIBL, (c) v_{sat_eff} has a good linearity with $V_{th,sat}$,(d) V_{th,sat_IV} has a

good linearity with $V_{th,sat}$, (e) v_{sat_eff} has a good linearity with $V_{th,lin}$, (f) V_{th,sat_IV} has a good linearity with $V_{th,lin}$. 85

Figure 4-20 Third parameter in the log I_{off} versus I_{on} characteristics of NMOS transistors in SCE regime: (a) and (b) show that DIBL has a linear relationship with log I_{off} and I_{on}. (c) and (d) show that $V_{th,sat}$ has a linear relationship with log I_{off} and I_{on}. (e) and (f) show that $V_{th,lin}$ has a linear relationship with log I_{off} and I_{on}. 86

Figure 4-21 Best-fit line of the log I_{off} versus I_{on} plot of NMOS transistors in the SCE regime. 87

Figure 4-22 Linearity of log I_{off} versus I_{on} characteristics of NMOS transistors in SCE regime (65 nm CMOS technology). 87

Figure 4-23 Linearity of the log I_{off} versus I_{on} characteristics may not be valid for PMOS transistors that are fabricated by 45 nm CMOS technology: (a) Anomalous increase in the threshold voltage, (b) Kink in the I_{ds_on} versus L characteristics, (c) Kink in the I_{ds_off} versus L characteristics, (d) Poor linearity in SCE regime. 89

Figure 4-24 Mechanism responsible for the poor linearity in log I_{off} versus I_{on} plot of PMOS transistors fabricated by 40nm CMOS technology. (a) & (c) refer to the case for the onset of the cross-over of halo implants. (b) & (d) refer to the case for the cross over of halo implants..............89

Figure 4-25 The effects of uniaxial stress on (a) the silicon bandgap and (b) the effective electron mass along the channel direction (m_x). 91

Figure 4-26 Effects of a "mechanical stress bias" in MOS transistor on the sensitivity of the silicon bandgap to a small amount of externally applied tensile stress: (a) highly compressively stressed CESL, (b) low compressive stressed CESL, (c) highly tensile stressed CESL. 92

Figure 4-27 Effects of the externally applied uniaxial tensile stress on <110> channel NMOS transistor fabricated on (100) Si substrate with 0.7 GPa tensile stressed CESL. 92

Figure 4-28 Effects of the externally applied uniaxial tensile stress on <110> channel PMOS transistor on (100) Si substrate with 0.7 GPa tensile stressed CESL. 93

Figure 4-29 Region of nearly constant I_{off} for NMOS transistors fabricated by 65 nm CMOS technology: (a) Maximum $V_{th,sat}$, (b) Minimum DIBL. 95

Figure 4-30 Region of nearly constant I_{off} for PMOS transistors fabricated by 65 nm CMOS technology: (a) Maximum $V_{th,sat}$, (b) Minimum DIBL 96

Figure 4-31 Overall strain-induced I_{on} improvement in <110> NMOS transistors. 97

Figure 4-32 The effects of CESL-induced tensile stress on: (a) $V_{th,sat}$, (b) μ_{eff} , (c) S_{ts} of NMOS transistors in the transition between SCE regime and RSCE regime..98

Figure 4-33 Overall performance improvement in <110> NMOS transistors by the CESL-induced tensile stress: (a) log I_{off} versus I_{on} characteristics, (b) I_{on} versus $V_{th,sat}$ characteristics, (c) log I_{off} versus $V_{th,sat}$ characteristics. 100

Figure 4-34 Overall strain-induced I_{on} improvement in <100> NMOS transistors. 101

Figure 4-35 Overall performance improvement in <100> NMOS transistors by CESL-induced tensile stress: (a) the log I_{off} versus I_{on} characteristics, (b) the I_{on} versus $V_{th,sat}$ characteristics, and (c) the log I_{off} versus $V_{th,sat}$ characteristics. 102
Figure 5-1 Switching in channel orientation increases hole mobility[1.32]. 103
Figure 5-2 Schematics of the (a) conventional (100)Si, (b) 45° rotated (100)Si 103
Figure 5-3 Effects of channel orientation on the effective hole mass of (100)Si[1.34]. 104
Figure 5-4 Effects of channel orientation increases I_{on} of PMOS transistors on (100)Si [1.33]. 104
Figure 5-5 Calculations of the effective conductivity electron mass (m^*): (a) 45° rotated (100)Si, (b) conventional (100) Si [1.34], (c) conduction band valleys on (100)Si plane............105
Figure 5-6 The effects of channel orientation on I_{g_off} of: (a) PMOS transistors, (b) NMOS transistors on (100) Si [1.33]. S/D extension implant dose: Ref = Thin Offset = SDE1 < SDE2. Thickness of the gate offset spacer: Ref = SDE1 = SDE2 > Thin Offset. 109
Figure 5-7 For cubic crystal structure such as silicon, (a) [100] direction is perpendicular to (100) plane, (b) [110] direction is perpendicular to (110) plane [5.15]............110
Figure 5-8 Various types of ion channeling : (a) planar channeling, (b) axial channeling, (c) indirect channeling [5.18]. 111
Figure 5-9 Monte Carlo simulation of the two-dimensional implanted boron distribution into (100) Si at a mask edge [5.19]............111
Figure 5-10 Critical angle for planar channeling along {110} plane and {100} plane as a function of the incident ion energy [5.22]. 113
Figure 5-11 Arrangements of the silicon atoms in (100) Si and (110) Si that are viewed along the direction normal to the plane [5.20, 5.21]............113
Figure 5-12 $\Re_{E\leftarrow D}$ of (100) Si for a generalized channel orientation. 116
Figure 5-13 (a) Schematics of the six conduction band valleys of (100) Si, (b) Validity of the six degenerate ellipsoid model for the conduction band energy minimum of silicon at the interface between SiO_2/Si interface. 116
Figure 5-14-Difference between lateral ion channeling and axial channeling. 119
Figure 5-15 Evidence of the increase in L_{eff} of <100> MOS transistors on (100) Si after a change in the channel orientation. Note that DIBL = $V_{th,lin} - V_{th,sat}$. 120
Figure 5-16 The ease of lateral ion channeling for arsenic and boron along <100> direction and <110> direction of (100)Si. 120
Figure 5-17 Effects of the change in channel orientation on the log I_{off} versus I_{on} plot of NMOS transistors (b) PMOS transistors on (100)Si. (L = 60 nm , W =1 μm). 121
Figure 5-18 Effects of the change in the channel orientation on μ_{eff} and v_{sat_eff} of NMOS transistors on (100) Si. 122
Figure 5-19 Effects of the change in the channel orientation on μ_{eff} and v_{sat_eff} of PMOS transistors on (100) Si. 122

Figure 5-20 Effects of switching from the conventional <110> channel orientation to <100> channel orientation on the I_{on} and I_{off} of NMOS transistors on (100)Si. 123
Figure 5-21 Effects of switching from the conventional <110> channel orientation to <100> channel orientation on the I_{on} and I_{off} of PMOS transistors on (100)Si. 124
Figure A-1 Higher hole mobility (μ_h) as a function of the effective vertical field (E_{eff}) for <110> PMOS transistors fabricated on (100)Si and (110)Si[A.1]...132
Figure A-2 Energy split (ΔE_{HH-LH}) in PMOS transistors on (100)Si and (110)Si [A.1].............132
Figure A-3 Electron mobility (μ_e) as a function of the surface carrier concentration (N_s) for NMOS transistors on (100)Si and (110)Si [A.1]...132
Figure A-4 Conduction band valleys on (100) plane and (110) plane [A.1].........................133

LIST OF TABLES

Table 2-1 Components of the piezoresistive tensor of bulk silicon (units: 10^{-11} Pa^{-1}) [2.74].............32

Table 2-2 Device-level piezoresistance coefficients of the MOS transistors fabricated on (100) surface-oriented silicon wafer (units: 10^{-11} Pa^{-1})...33

Table 3-1 Merits and limitations of various drain current transport theories for the nanoscale MOS transistors..57

Table 4-1 Evaluation of the suitability of the third parameter in the log I_{off} versus I_{on} plot. 88

Table 4-2 Comparison between various stress engineering techniques...94

Table 4-3 Method 1 to ascertain the overall strain-induced I_{on} improvement of <110> NMOS transistors..99

Table 4-4 Method 2 to ascertain the overall strain-induced I_{on} improvement of <110> NMOS transistor using $V_{th,sat}$ as a third parameter in the logI_{off} versus I_{on} plot. 99

Table 4-5 Overall performance improvement for <100> channel NMOS transistor by stress engineering when we consider $V_{th,sat}$ as the third parameter in the log I_{off} versus I_{on} characteristics. 102

Table 5-1 The area of open space (A_{open_space}) in the silicon crystal lattice for various surface orientation and crystal directions [5.20]. 112

Table A-1: Effects of surface orientation and channel orientation on the effective electron masses along the quantum confinement direction (m_z) [5.2]..133

Table A-2 Device-level piezoresistance coefficients of the <110> NMOS transistors fabricated on (100)Si and (110)Si (units: 10^{-11} Pa^{-1}) [2.41]..133

1. Introduction

1.1 Background

For the past 40 years, relentless focus on Moore's law scaling has provided ever-increasing switching speed and packing density in the metal-oxide-semiconductor (MOS) transistor, as illustrated in Fig.1-1 [1.1]. Traditionally, the tradeoffs between the three main indices of transistor performance, which are the on-current (I_{on}), the short-channel effect (SCE) and the power consumption, can be achieved by simply scaling the thickness of the silicon oxynitride-based gate dielectrics. However, it has been recognized that this conventional scaling has confronted difficulties in the sub-100 nm regime because the gate-oxide leakage current will become excessively large when the physical gate oxide thickness reaches 1.5 nm, as shown in Fig.1-2 [1.2]. From the perspective of the total chip power consumption, the static power dissipation is expected to dominate over the dynamic power dissipation in the nanoscale MOS transistors, as shown in Fig.1-3 [1.3]. Owing to the large increase in the gate leakage for very thin silicon oxynitride-based gate dielectrics, there will be significant increase static power dissipation, leading to a large increase in the total chip power dissipation. Hence, performance enhancement techniques such as mobility enhancement techniques, high-k dielectrics and metal gates [1.4] have been explored to sustain the performance scaling trends. This work will focus on the mobility enhancement techniques rather than the high-*k* gate dielectric with metal gates.

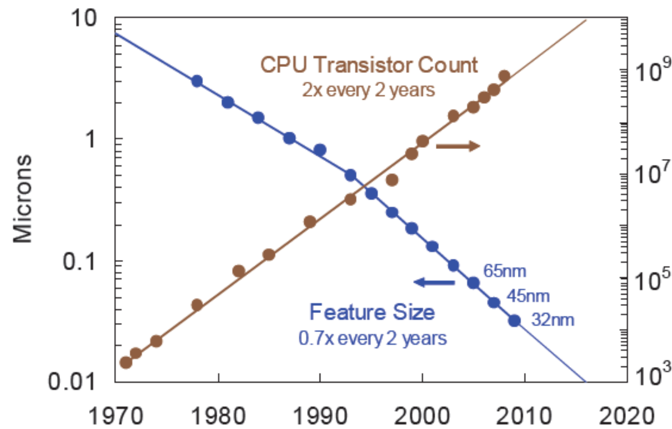


Figure 1-1 Moore's Law: CPU transistor count has increased by two times and feature size has decreased by 0.7 times in every two years [1.1].

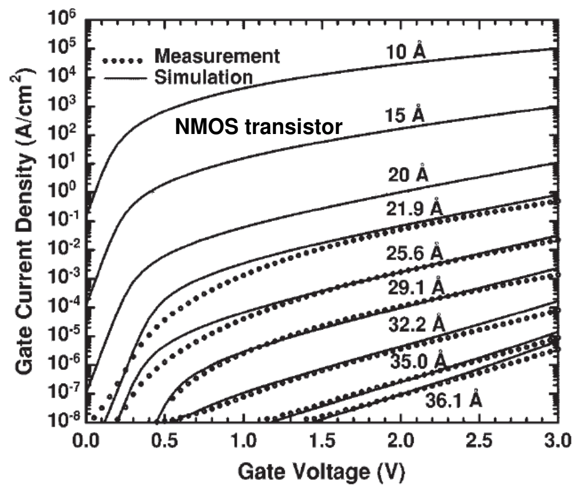


Figure 1-2 Excessive gate-oxide leakage occurs with further scaling of the conventional silicon oxynitride gate dielectric [1.2].

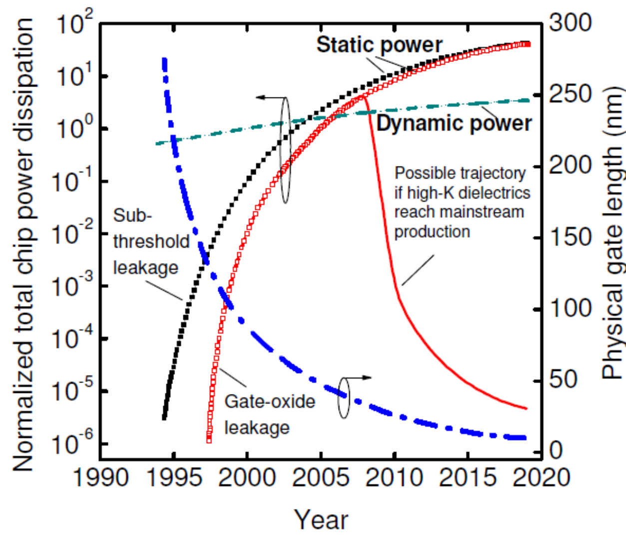


Figure 1-3 Total chip dynamic and static power dissipation trends based on the International Technology Roadmap for Semiconductor (ITRS) [1.3].

1.2 Roles of carrier mobility in drain current transport

The drift-diffusion model is typically associated with the long-channel MOS transistors. The drain current (I_{ds}) comprises the diffusion current (I_{diff}) and the drift current (I_{drift}). When the gate bias (V_{GS}) is smaller than the threshold voltage, I_{ds} is dominated by I_{diff}. When the V_{GS} is bigger than the threshold voltage, I_{ds} is dominated by I_{drift}. The equation for the saturation drain current (I_{ds}) of a long-channel MOS transistor [1.5],

$$I_{ds} = \mu_{eff} C_{ox,inv} \frac{W}{2L} \left(V_{GS} - V_{th,sat}\right)^2 \tag{1.1}$$

where μ_{eff} is the low-field mobility. $C_{ox,inv}$ is the gate oxide capacitance at strong inversion per unit area. W is the gate width. L is the gate length. $V_{th,sat}$ is the saturation threshold voltage.

On the other hand, velocity saturation [1.6-1.10] is often associated with the nanoscale MOS transistors. It is based on the balance of energy equation for channel carriers: the rate of energy gained from the lateral electric field is equal to the rate of energy dissipation through the emission of longitudinal optical phonons [1.7]. For nanoscale MOS transistor in velocity saturation [1.6],

$$I_{ds} = v_{sat} C_{ox,inv} W \left(V_{GS} - V_{th,sat}\right) \tag{1.2}$$

where v_{sat} is the saturation velocity. Based on the time-of-flight measurement, v_{sat} for electrons in silicon is 10^7 cm/s while v_{sat} for holes in silicon is 6×10^6 cm/s at a temperature of 300 K [1.9]. According to the theoretical predictions, v_{sat} is independent of μ_{eff} [1.10]. From eqn. (1.2), the velocity saturation model predicts that mobility enhancement will not improve I_{on} of nanoscale MOS transistor.

However, this is contradictory to the experimental observations of strained-induced I_{on} improvement in the nanoscale MOS transistors [1.11-1.13]. Furthermore, Monte Carlo simulation by Ruch [1.16] and Mizuno et al. [1.17] showed that the velocity overshoot can occur in the nanoscale MOS transistors. In addition, the Monte Carlo simulation by Miyata et al. [1.18] showed that velocity overshoot can be further increased by the application of mechanical stress. In fact, Kim et al. has reported their experimental observation of electron velocity overshoot in bulk n-channel MOS (NMOS) transistors at room temperature [1.19].

Ballistic transport [1.20-1.23], which is associated with nanoscale MOS transistors, considers an ideal situation where the channel carrier does not undergo any scattering.

Since mobility is a concept that involved scattering, ballistic transport actually disregards μ_{eff} in the nanoscale MOS transistors [1.20-1.23] and thus cannot account for the strain-induced I_{on} improvement in nanoscale MOS transistors.

In the quasi-ballistic transport [1.24, 1.25], the channel carriers will experience some scatterings when it moves from the source to the drain but the number of channel scatterings is much smaller than that of the velocity saturation model. Since mobility is a concept that involved scattering, quasi-ballistic transport considers the effects of μ_{eff} and thus can account for the strain-induced I_{on} improvement in nanoscale MOS transistor [1.11-1.13]. However, there are some terms in the drain current equation for the quasi-ballistic transport that are not properly defined.

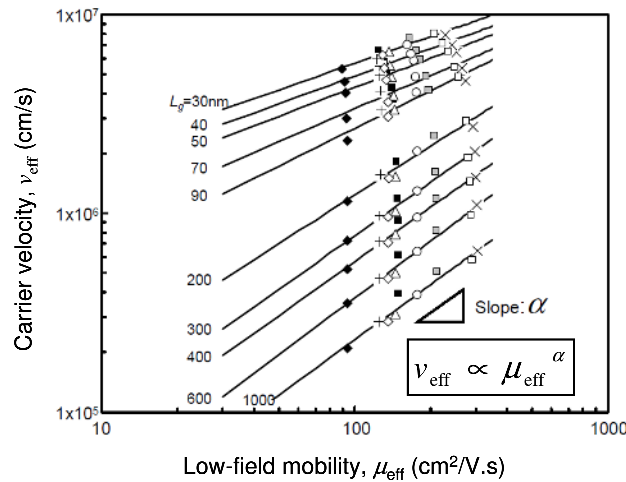


Figure 1-4 Experimental correlation between low-field mobility (μ_{eff}) and the high-field carrier velocity (v_{eff}) for the nanoscale NMOS transistors [1.29].

At this point, it may seem that quasi-ballistic transport is the most appropriate model for nanoscale MOS transistor. However, some researchers have managed to use the conventional velocity saturation model to fit the experimental saturation drain current (I_{ds}) of the nanoscale MOS transistors. In the physics-based model for MOS transistors developed by Hauser [1.26], v_{sat} is treated as a fitting parameter that can be increased to 2.06×10^7 cm/s so as to fit the experimental I_{ds} versus the drain-to-source voltage (V_{DS}) characteristics of the nanoscale NMOS transistor. Although this approach is conceptually wrong, this can avoid detailed discussion in velocity overshoot and quasi-ballistic transport. Furthermore, Khakifirooz et al. introduced a semi-empirical model for the saturation drain current of the nanoscale MOS transistor that is based on the sheet charge approximation [1.27, 1.28]. In addition, Tatsumura et al. [1.29] has performed an extensive

experimental study to show that there is indeed a correlation between the low-field mobility and the high-field carrier velocity, as shown in Fig. 1-4. Hence, there is a need to further clarify the drain current transport of the nanoscale MOS transistor.

1.3 Mobility enhancement techniques

In view of the increasing cost and complexities involved in the scaling of MOS transistors, the semiconductor industry uses mobility enhancement techniques to improve transistor performance. From Fig. 1-5, the mobility enhancement techniques can be categorized into the substrate-induced stress and the process-induced stress [1.30]. The use of biaxial tensile stressed silicon on relaxed $Si_{1-y}Ge_y$ virtual substrate [1.11, 1.31], will increase the mobilities of both NMOS transistors and p-channel MOS (PMOS) transistors. The use of 45° rotated (100) surface-oriented silicon instead of the conventional (100) surface-oriented silicon increases the hole mobility [1.32, 1.33] but does not affect the electron mobility [1.34]. Since the change in the channel orientation results in the mechanical insensitivity of <100> PMOS transistors [1.35, 1.36], stress engineering techniques cannot be used to improve I_{on} of <100> PMOS transistors. However, this unique property of <100> PMOS transistor will allow us to use a single tensile contact etch stop layer (CESL) over both <100> NMOS transistors and <100> PMOS transistors without degrading I_{on} of PMOS transistors. For Hybrid-Orientation Technology (HOT), PMOS transistors are fabricated on (110) surface-oriented silicon with <110> channel orientation while NMOS transistors are fabricated on (100) surface-oriented silicon with <110> channel orientation [1.37]. As opposed to 45° rotated (100) surface-oriented silicon, hole mobility is enhanced by HOT and can be enhanced by uniaxial compressive stress [1.37]. For process-induced stress, mechanical stress can be introduced into the channel through silicon nitride liners, S/D stressors, gate and contact. For liners, mechanical stress in CESL can be modulated such that compressive stressed CESL is deposited over PMOS transistors while tensile stressed CESL is deposited over NMOS transistors [1.38]. Stress Memorization technique (SMT), which involves the deposition of the tensile stressed silicon nitride liner over the NMOS transistor after Ge pre-amorphization implantation and S/D implantation, will induce a beneficial compressive vertical channel stress in NMOS transistor after S/D annealing [1.39, 1.40]. Embedded $Si_{1-y}Ge_y$ (e-SiGe) S/D stressor will induce a longitudinal compressive stress for PMOS transistor [1.41]. Embedded carbon-doped silicon (e-Si:C) S/D stressor will induce a longitudinal tensile stress for NMOS transistor [1.42]. Recently, Intel patented a novel way to increase I_{on} of NMOS transistors

by introducing longitudinal tensile stress to NMOS transistor using trench contact with tensile fill [1.43]. The stress engineering techniques are applicable to the silicon oxynitride gate dielectric/ poly-Si gate stack as well as the high-*k* gate dielectric/ metal gate stack. In fact, the channel strain of Gate-last high-k/ metal gate scheme is higher than Gate-first High-*k*/ metal gate and the conventional $SiON_x$/ poly-Si gate schemes [1.44, 1.45] because it does not experience any reaction strain by gate.

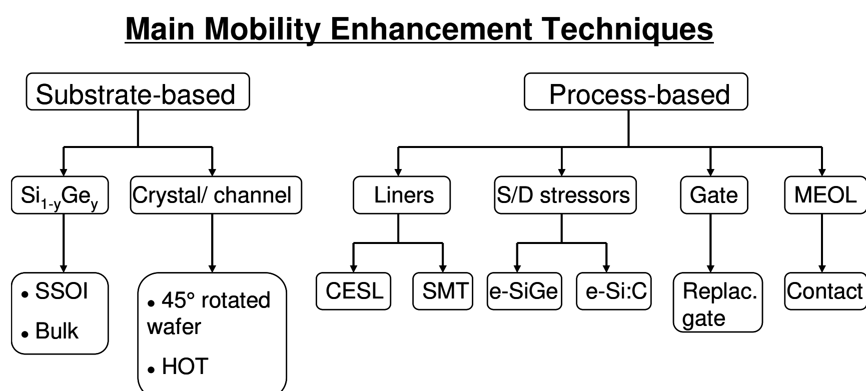


Figure 1-5 Flow chart of the main mobility enhancement techniques [1.30].

Objectives

(1) To clarify the drain current transport in the nanoscale MOS transistor

There are some conflicting theories regarding the effects of mobility enhancement on the I_{on} of nanoscale MOS transistors: (i) velocity saturation model, (ii) quasi-ballistic model, (iii) ballistic model, and (iv) empirical model. However, experimental data has shown that stress engineering can improve I_{on} of the nanoscale MOS transistors through mobility enhancement. This work will clarify the merits and limitations of the theories and then come up with a simplified drain current equation that can reconcile the existing theories on drain current transport.

(2) To determine if CESL-induced tensile stress can bring about an overall I_{on} improvement in NMOS transistor

It is intuitive that a higher mobility will increase both I_{on} and subthreshold I_{off}. This work will investigate the physics behind the overall strain-induced I_{on} improvement in NMOS transistors even though stress engineering leads to a bigger percentage increase in log I_{off} than I_{on}.

(3) To determine the effects of switching from the conventional <110> channel orientation to <100> channel orientation on I_{on} and I_{off} of MOS transistors that are fabricated on (100)Si

Apart from the conventional (100) Si, 45° rotated (100)Si is also widely used in the semiconductor industry owing to its higher hole mobility. However, the effects of the change in channel orientation on I_{on} of NMOS transistor are not well studied. The physics behind the increase in effective channel length (L_{eff}) is also left unexplored. This work will first study the effects of the change in channel orientation on electron mobility of NMOS transistors as well as the physics behind the increase in L_{eff}. Then, we discuss the effects of the change in channel orientation on I_{on} and I_{off} of NMOS transistor as well as the strain-induced effects on I_{on} and I_{off} of <100> channel NMOS transistors.

1.4 Organization of the Thesis

Chapter 2 gives an overview of the stress engineering techniques and its effects on the conduction band minimum and valence band maximum. The usage of piezoresistance coefficients to predict the strain-induced change in mobility is also discussed.

Chapter 3 explains the relevance of the low-field mobility (μ_{eff}) in the nanoscale MOS transistors. This work will unify the merits of velocity saturation model, the ballistic transport and the quasi-ballistic transport and then come up with a simplified saturation drain current equation for nanoscale MOS transistor.

Chapter 4 discusses the effects of mechanical stress on the electrical parameters of MOS transistors. Unlike most publication, this work will discuss the physics behind the strain-induced increase in the I_{off} of NMOS transistors. This work will address why CESL-induced tensile stress will lead to an overall I_{on} improvement in NMOS transistors even though the percentage increase in log I_{off} is bigger than the percentage increase in I_{on}.

Chapter 5 discusses the effects of the switching from the conventional (100)Si to 45° rotated (100)Si on I_{on} and I_{off} of MOS transistors. This work shows that the effective conductivity electron mass of (100)Si is independent of channel orientation and discusses the mechanism behind the increase in L_{eff}.

Chapter 6 concludes the thesis by highlighting the main contributions of this project and providing some suggestions for future work.

2. Literature review on Stress Engineering

2.1 Methods to introduce channel stress

2.1.1 Substrate-induced stress

Fig. 2-1 shows the cross-section of MOS transistors fabricated on the strained-Si on relaxed $Si_{1-y}Ge_y$ virtual substrate. The fabrication process can be found in Ref.[1.11]. For NMOS transistor, the silicon epitaxial layer is in-situ doped with boron. For PMOS transistor, the silicon epitaxial layer is in-situ doped with arsenic. Owing to the lattice mismatch between the epitaxial Si and the relaxed $Si_{1-y}Ge_y$ virtual substrate, the transistors will experience biaxial tensile stress, leading to an increase in electron mobility and hole mobility [2.1, 2.2].

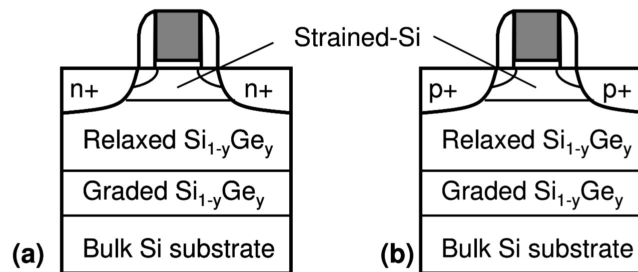


Figure 2-1 Cross-section of (a) NMOS transistor, (b) PMOS transistor fabricated on the strained-Si on relaxed Si_{1-x} Ge_y virtual substrate.

A decrease in the thickness of the strained-Si layer causes a decrease in the carrier confinement because of quantum mechanical effects at the SiO_2/ Si interface [2.3]. This leads to carrier conduction through the low-mobility relaxed $Si_{1-y}Ge_y$ underlayer and thus degrading the mobility. When the thickness of the strained-Si layer is bigger than the equilibrium critical thickness (t_{crit}), strain relaxation will occur through the formation of misfit dislocations. Although a higher Ge content will increase the amount of biaxial tensile stress, t_{crit} will decrease with increasing Ge content [2.4]. Fiorenza et al. reported that there is mobility enhancement in MOS transistors fabricated on the strained-Si on relaxed $Si_{1-y}Ge_y$ virtual substrate even though the strained-Si layer is significantly bigger than t_{crit} [2.5]. However, there will be a significant increase in the off-state leakage current of the MOS transistors whose thickness of the strained-Si layer is significantly bigger than t_{crit} [2.5]. From the photo-emission microscopy, the increase in the leakage current is caused by the enhanced dopant diffusion near misfit locations, which leads to the S/D electrical shorting [2.5].

2.1.2 Process-induced stress

2.1.2.1 Shallow trench isolation (STI)

The purpose of shallow trench isolation (STI) is to electrically isolate the adjacent MOS transistors. It involves the deposition of High Density Plasma (HDP) oxide in silicon trenches, followed by chemical mechanical polish (CMP) to planarize the silicon surface. Owing to the difference in thermal expansion coefficient between silicon and silicon dioxide, STI will induce a lateral compressive stress in the channel [2.6]. Based on piezoresistance coefficient, this is undesirable for NMOS transistors and PMOS transistors that are fabricated on (100) surface-oriented silicon. Furthermore, the mechanical stress-induced dopant diffusion model has shown that STI-induced compressive stress will retard the diffusion of boron, arsenic and phosphorus [2.7]. This is consistent with the results obtained using the density function theory calculations [2.8]. In addition, Hsieh et al. showed that a reduction in S/D overhang can decrease the gate-to-S/D overlap of NMOS transistor owing to STI-induced retardation of the lateral diffusion of the n^+ S/D extension implants [2.9].

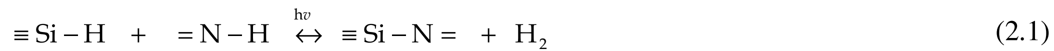
2.1.2.2 Contact etch stop layer (CESL)

In order to achieve a higher packing density, interconnects have migrated from the ordinary source/drain (S/D) contacts to the borderless S/D contacts (see Fig. 2-2). In the absence of the contact etch stop layer (CESL), the contact misalignment can result in a short-circuit between the n^+ S/D and p-well (or p^+ S/D and n-well) owing to the poor etch selectivity between STI and the inter-level dielectric (ILD). Fortunately, this problem can be alleviated by the use of CESL, as shown in Fig. 2-3 [2.10]. Etch selectivity is improved because the etching of the contact holes can be carried out in two steps: (i) etch ILD until CESL, and (ii) etch the CESL to expose the silicide. Subsequently, TiN is deposited over the contact holes for adhesion purposes and then the contact holes are filled with tungsten.

Nowadays, CESL also serves as a mechanical stressor to enhance the channel carrier mobility. Experimental evidence shows that the presence of tensile stressed CESL will induce lateral tensile stress in the channel while compressive stressed CESL will induce lateral compressive stress [2.11]. There are various schemes for this stress engineering technique. For <110> MOS transistors fabricated on (100) surface-oriented silicon, the application of a single tensile stressed CESL will improve the I_{on} of NMOS transistor but degrade I_{on} of PMOS transistor. On the other hand, application of a single compressive

stressed CESL will improve the I_{on} of PMOS transistor but degrade I_{on} of NMOS transistor. This leads to the development of the dual-CESL technique, which involves tensile stressed CESL over the NMOS transistor and compressive stressed CESL on PMOS transistor [1.38] (see Fig. 2-4).

Mechanical stress of plasma-enhanced chemical vapor deposition (PECVD) SiN_x films is directly related to the evolution of hydrogen content and plasma ion bombardment during the deposition [2.12]. A smaller hydrogen content will yield a more tensile SiN_x film. A bigger plasma ion bombardment will yield a more compressive SiN_x film. Moreover, the tensile stress of PECVD SiN_x film can be further increased by ultra-violet (UV) curing [2.13].



$$\equiv \mathrm{Si-H} \ + \ =\mathrm{N-H} \ \overset{h\upsilon}{\leftrightarrow} \ \equiv \mathrm{Si-N=} \ + \ \mathrm{H_2} \tag{2.1}$$

Since this reaction has a negative enthalpy (ΔH) of -1.86 eV, $\equiv \mathrm{Si-N=}$ bonds and H_2 release are more energetically favored. From Fig. 2-5, UV curing promotes Si-N-Si cross-linking due to dehydrogenization, leading to a near stoichiometric Si_3N_4 network containing smaller vacancies due to the remaining Si-H and N-H bonds [2.14]. The blanket CESL film stress (σ) can be measured by Tencor FLX-2320, which is based on the wafer-bowing measurements and the Stoney formula [2.11, 2.15].

$$\sigma = \frac{E_{\mathrm{Si}} t_{\mathrm{substrate}}^{\ 2}}{6(1-\nu_{\mathrm{Si}}) R t_{\mathrm{film}}} \tag{2.2}$$

where E_{Si} is the Young's modulus of silicon. ν_{Si} is the Poisson's ratio of silicon. R is the radius of curvature of the substrate. $t_{\mathrm{substrate}}$ is the thickness of the substrate. t_{film} is the thickness of the film.

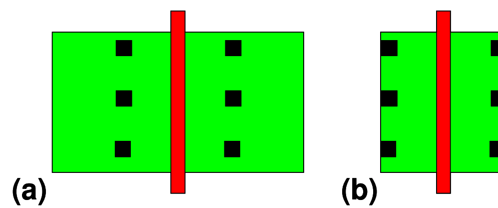


Figure 2-2 Schematics of (a) the ordinary S/D contacts and (b) the borderless S/D contacts.

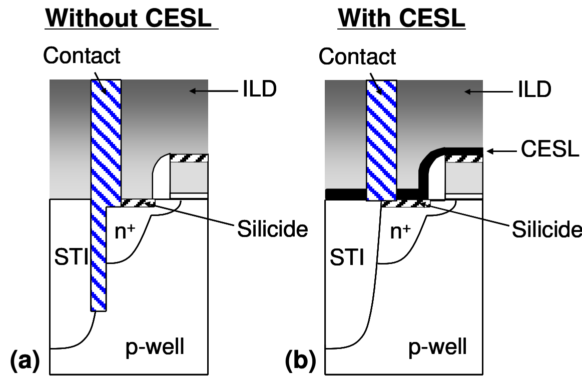


Figure 2-3 The importance of CESL to prevent the electrical shorting between n^+ S/D and p-well when there is a mask misalignment: (a) Without CESL, (b) With CESL.

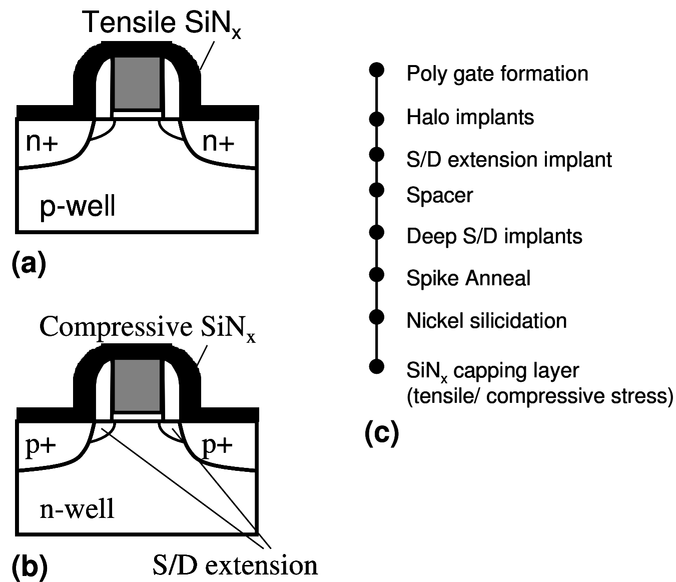


Figure 2-4 Process flow of CESL as a mechanical stressor.

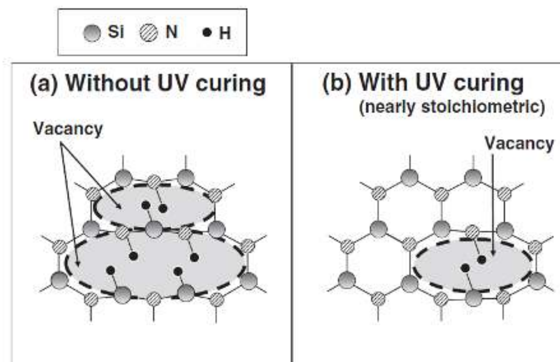


Figure 2-5 Structural model for local bonding arrangement in PECVD SiN_x film for (a) Without UV curing, (b) With UV curing [2.14].

2.1.2.1 SiGe S/D for PMOS transistor

Fig. 2-6 shows the fabrication process of the embedded $Si_{1-y}Ge_y$ (e-SiGe) S/D stressor. After halo implantation and S/D extension implantation, the deep S/D regions are anisotropically etched. Next, e-SiGe S/D process is carried out by either (i) in-situ boron doping in the epitaxial SiGe growth process, or (ii) non-doped epitaxial SiGe growth followed by boron ion implantation [1.41]. Finally, the front-end process is completed by a spike anneal and a nickel silicidation process. Finite-element study and experimental measurement of the strain distribution show that the $Si_{1-y}Ge_y$ S/D stressor will induce a longitudinal compressive stress in the silicon channel [2.16, 2.17]. When the Ge mole fraction (y) is increased, the longitudinal compressive stress will also be increased [1.43]. In addition, studies have shown that the parasitic S/D series resistance (R_{sd}) in $Si_{1-y}Ge_y$ S/D stressor is lower than that in the conventional Si S/D because of the reduction in Schottky barrier height caused by Ge incorporation [2.18] and the pronounced boron activation in $Si_{1-y}Ge_y$ [2.18, 2.19]. However, the increase of Ge content must be approached with care to avoid strain relaxation through the formation of misfit and threading dislocations [2.20]. Apart from dislocations, elastic strain relaxation can also occur at the edges of the recessed S/D close to the spacer [2.21]. Gonzalez et al. [2.22] has proposed a semi-empirical model to describe the effects of Ge mole fraction (y) in the $Si_{1-y}Ge_y$ S/D stressor on the increase in the junction leakage current but he did not consider the effects of Ge content on R_{sd}.

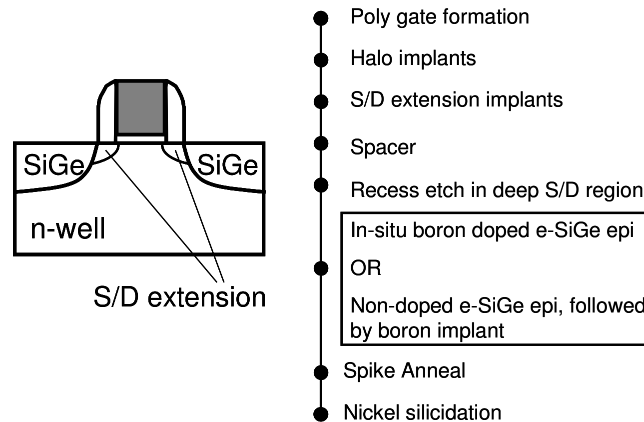


Figure 2-6 Process flow of $Si_{1-y}Ge_y$ S/D stressor for PMOS transistor.

2.1.2.1 Stress Memorization Technique (SMT)

Fig. 2-7 illustrates the fabrication process of Stress Memorization Technique (SMT) [1.39]. After the spacer formation and S/D implantation, a Poly Amorphization Implantation (PAI) of the poly-Si gate by Ge is performed on the NMOS transistor. A high tensile stressed SiN_x layer (also known as SMT cap) is selectively deposited over the NMOS **t**ransistor. The selective removal of tensile stressed SiN_x layer over the PMOS transistor will alleviate the strain-induced degradation of hole mobility. Subsequently, S/D annealing is done to activate the S/D dopants. Next, the SiN_x layer is removed and nickel silicidation is done. There are two distinct SMT phenomena: (i) S/D component that occurs at lower annealing temperature (around 500 °C), and (ii) poly component that occurs at the final high temperature anneal (around 1000 °C) [2.23]. The longitudinal tensile stress induced in the S/D regions is probably caused by volume expansion during amorphization, and the inability of the film to reduce to proper volume in the presence of a SMT cap [2.23]. When gate length, the S/D component of SMT is expected to increase [2.23]. On the other hand, Miyashita et al. found that the channel compressive stress in the vertical direction originates from the volume expansion of the poly-Si gate, which is associated with the grain growth and the highly concentrated dopants implanted into the poly-Si gates [1.40]. There are several uncertainties regarding the strain-induced effects of SMT on the gate leakage current. Miyashita et al. reported that the gate leakage current of NMOS transistor is increased when a SMT cap with a bigger tensile stress is used [1.40]. Liu et al. reported that the increase in the gate leakage current of NMOS transistor is caused by the strain-induced increase in the diffusion coefficient of the S/D extension implants of NMOS transistor [2.24]. This is consistent with the experimental observation of enhanced arsenic diffusion by tensile stress during S/D annealing [2.25]. Furthermore, the density function calculations also show that interstitial-mediated arsenic diffusion is enhanced by tensile stress at temperature of 1000 °C [2.8]. However, Morifuji et al. reported that the gate leakage current is reduced when a SMT cap with a bigger tensile stress owing to a reduction in the grain-size of the poly-Si gate [2.26]. This contradicts with Miyashita et al. who reported that the grain-size of the poly-Si gate is increased by SMT [1.40].

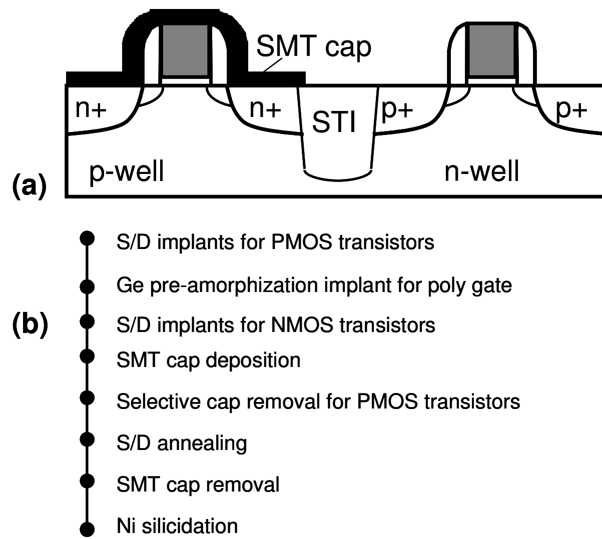


Figure 2-7 Process flow of Stress Memorization Technique (SMT).

2.1.2.2 e-Si:C S/D for NMOS transistor

Fig. 2-8 shows the fabrication process of the embedded Si:C (e-Si:C) S/D stressor for NMOS transistor. A straightforward approach is to recess the deep S/D region and then deposit e-Si:C using a selective epitaxy process [2.27]. A more recent approach is to use C ion implantation and solid-state epitaxy anneal to form e-Si:C in the S/D regions [1.42]. The second approach is able to achieve a higher substitutional C concentration in the S/D regions. The finite-element study and the experimental measurement of the strain distribution show that the $Si_{1-y}C_y$ S/D stressor will induce a lateral tensile stress in the silicon channel [2.28-2.31]. Although a higher substitutional carbon concentration is desirable for stress engineering, the dopant activation in S/D regions decreases with increasing carbon concentration [2.30]. Despite the reduction in Schottky barrier height caused by carbon incorporation [2.32], the increase in carbon concentration degrades the S/D series resistance (R_{sd}) [1.42, 2.30] and thus compromises the I_{on} performance gain expected from stress engineering. The stress relaxation mechanism of $Si_{1-y}C_y$ layers involves the formation of carbon-containing interstitial complexes [2.33], leading to carbon atom precipitation out from the substitutional sites and thus the longitudinal tensile stress is reduced. Simoen et al. proposed that the diffusion of the boron halo implants is retarded by the vertical compressive stress induced by $Si_{1-y}C_y$ stressor, leading to a higher electric field at the depletion region of the S/D regions and thus junction leakage current is increased [2.34]. However, it has been reported that the presence of carbon in silicon can suppress the interstitial-enhanced boron diffusion [2.35]. Hence, there is a possibility that

the increased junction leakage current in e-Si:C S/D regions is caused by the ability of carbon to suppress the diffusion of the boron halo implants rather than the effects of the pure mechanical stress to retard the diffusion of the boron halo implants. Nevertheless, the important message is that $Si_{1-y}C_y$ S/D stressor can increase both junction leakage current and I_{on} of NMOS transistor.

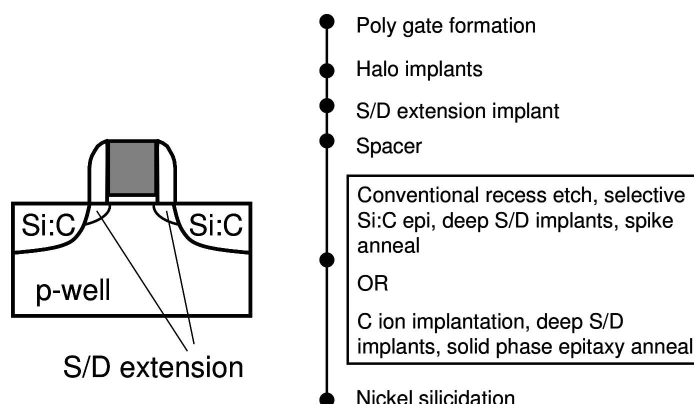


Figure 2-8 Process flow of e-Si:C S/D as a mechanical stressor.

2.1.3 Externally applied mechanical stress

Externally applied mechanical stress allows us to decouple the effects of pure mechanical stress from the chemical nature of the mechanical stressor. For the $Si_{1-y}Ge_y$ S/D stressor, the amount of longitudinal compressive stress induced in the PMOS transistor can be varied by changing the Ge concentration but this will decrease R_{sd} owing to the reduction in Schottky barrier height caused by Ge incorporation [2.18] and the pronounced boron activation in $Si_{1-y}Ge_y$ S/D regions [2.18, 2.19]. The situation is more complicated for e-Si:C S/D stressor. Although a higher carbon concentration is desirable for tensile stress, the stress relaxation will cause carbon atom precipitation out from substitutional sites and thus reduce tensile stress [2.33]. Furthermore, an increase in carbon concentration will degrade R_{sd} [1.42, 2.30] and thus compromise the I_{on} performance gain expected from stress engineering.

According to Kanno et al. [2.36], the wafer-to-wafer variation within a lot are attributed to the wafer position in the equipment and chamber conditions, while the die-to-die variation with a wafer are caused by lithography and the wafer-edge effects. From Fig. 2-9, statistics can be used to minimize the effects of the above variations and demonstrate that the hole mobility enhancement by a new type of stress engineering technique, Diamond-like Carbon (DLC) liner [2.37, 2.38]. However, externally applied

mechanical stress can eliminate the influence of the above variations because the same transistor is used to study the effects of the mechanical stress. Hence, externally applied mechanical stress is often used to explain the effects of mechanical stress on the conduction band minimum of silicon [1.14] and valence band maximum of silicon [1.15, 2.39]. From Fig. 2-10, Cantilever method [2.40, 2.41], three-point rod method [2.42] and four-point rod method [2.43] can be used to apply uniaxial stress, whereas o-rings [2.40, 2.41, 2.43] can be used to apply biaxial stress. To apply uniaxial stress along the channel direction, the wafer strips are cut along the direction that is perpendicular to the poly-Si gate alignment. To apply uniaxial stress perpendicular to the channel direction, the wafer strips are cut along the direction that is parallel to the poly-Si gate alignment. For the externally applied biaxial stress, we can either use the entire wafer or cut the wafer into a square. From Fig. 2-11, downward force at the edges of the wafer strip induces uniaxial tensile stress, whereas upward force at the edges of the wafer strip induces uniaxial compressive stress. The application of upward force in the centre of the wafer will result in biaxial tensile stress while the application of a downward force in the centre of the wafer will result in biaxial compressive stress. From Fig. 2-12, uniaxial tensile stress results in a vertical displacement along the stress direction, whereas biaxial tensile stress results in an equal amount of vertical displacement for two perpendicular directions on the wafer surface [2.43]. Similar analysis can be done for compressive stress.

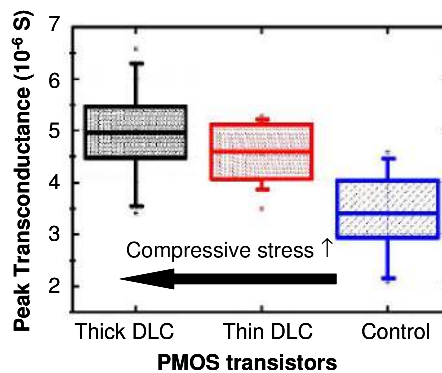


Figure 2-9 The use of statistics to compare the hole mobility enhancement by the Diamond-like Carbon (DLC) liner [2.37].

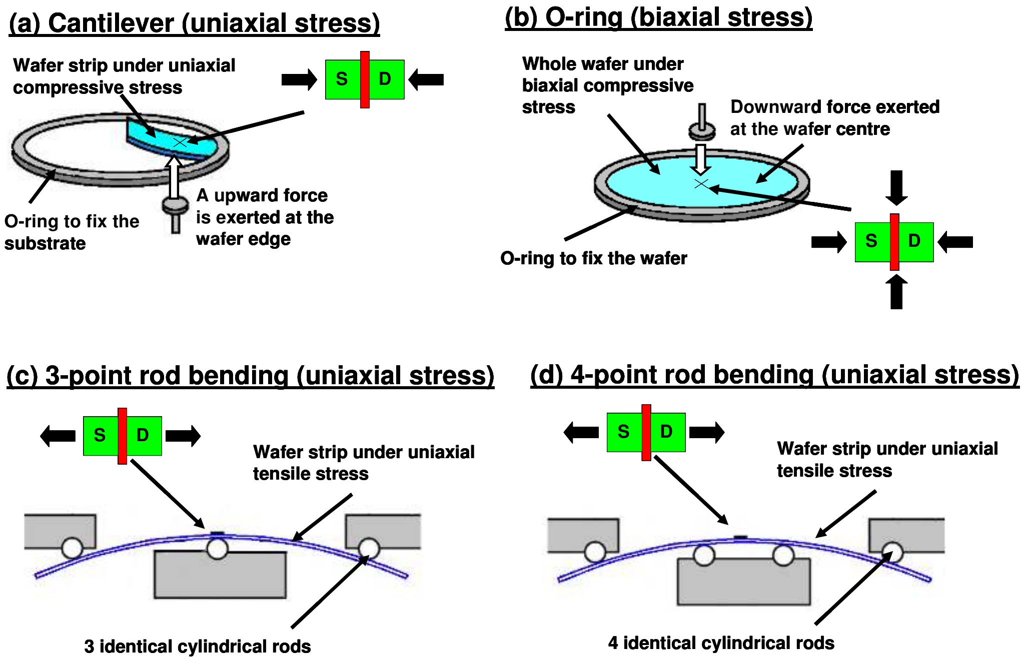


Figure 2-10 Different types of externally applied mechanical stress.

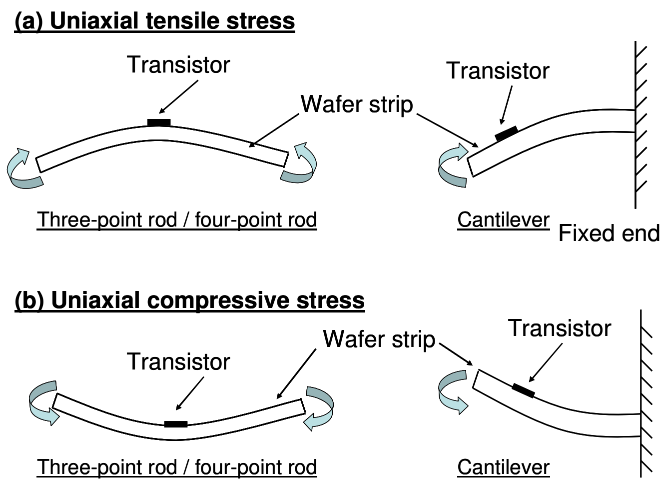


Figure 2-11 Schematics of the curvature of the rectangular wafer strips under: (a) uniaxial tensile stress and (b) uniaxial compressive stress. Note that the transistor is on the top surface of the wafer strip.

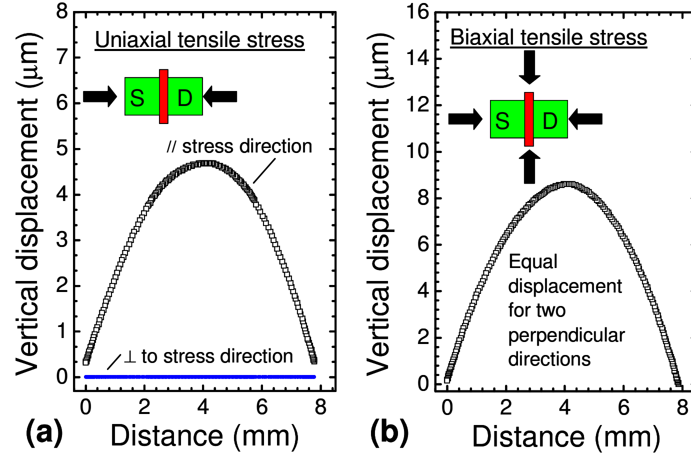


Figure 2-12 Differences between uniaxial tensile stress and biaxial tensile tensile in terms of wafer displacement from equilibrium position [2.43].

2.2 Strain-induced mobility enhancement

It is well-established that the stress engineering techniques [1.11- 1.13] and the externally applied mechanical stress [1.14-1.15] can increase the electron mobility and hole mobility. This is consistent with hole mobility model [2.44, 2.45] and electron mobility model [2.45, 2.46]. This section will explain the effects of mechanical stress on the conduction band minimum and the valence band maximum as well as the usage of piezoresistance coefficients.

2.2.1 First Brillouin zone

To simplify the discussion of the conduction band minimum and valence band maximum, it is a common practice to consider the effects of electron diffraction in a perfect crystal structure. As shown in Fig. 2-13, silicon has a diamond crystal structure [2.47]. This type of structure consists of two inter-penetrating face-centered cubic (FCC) structures that are displaced relative to one another by 1/4 of the lattice constant (a_0). Primitive vectors of FCC are given by,

$\boldsymbol{a_1} = 0.5a_0(0,1,1)^{\mathrm{T}}$, $\boldsymbol{a}_2 = 0.5a_0(1,0,1)^{\mathrm{T}}$ and $\boldsymbol{a}_3 = 0.5a_0(1,1,0)^{\mathrm{T}}$

Primitive vectors of the reciprocal lattice ($\boldsymbol{g_1}$, $\boldsymbol{g_2}$ and $\boldsymbol{g_3}$) are given by [2.48, 2.49],

$\boldsymbol{g_1} = \dfrac{2\pi}{a_0}(-1,1,1)^{\mathrm{T}}$, $\boldsymbol{g_2} = \dfrac{2\pi}{a_0}(1,-1,1)^{\mathrm{T}}$ and $\boldsymbol{g_3} = \dfrac{2\pi}{a_0}(1,1,-1)^{\mathrm{T}}$.

The first Brillouin zone in three dimensions can be defined as the smallest polyhedron confined by planes perpendicularly bisecting the reciprocal lattice vectors. Fig. 2-14 shows the reciprocal lattice for FCC is a body-centred cubic (BCC) lattice [2.48]. Since the reciprocal lattice vectors are obtained from the crystal lattice vectors, the symmetry of the Brillouin zone is determined by the symmetry of the crystal lattice. The center of Brillouin zone is denoted by, Γ = (0,0,0) ($2\pi/a_o$). The three high-symmetry directions in the Brillouin zone are [2.49],

[100] direction: $\Gamma \rightarrow \Delta \rightarrow X$

[111] direction: $\Gamma \rightarrow \Lambda \rightarrow L$

[110] direction: $\Gamma \rightarrow \Sigma \rightarrow K$

where X = (1,0,0) ($2\pi/a_o$), L = (1,1,1) ($2\pi/a_o$), K= U = (1,1,0) ($3\pi/2a_o$),

Δ = (ζ,0,0) ($2\pi/a_o$), Λ = (ζ,ζ,ζ) ($2\pi/a_o$), Σ = (ζ,ζ,ζ) ($3\pi/2a_o$) with $0 < \zeta < 1$.

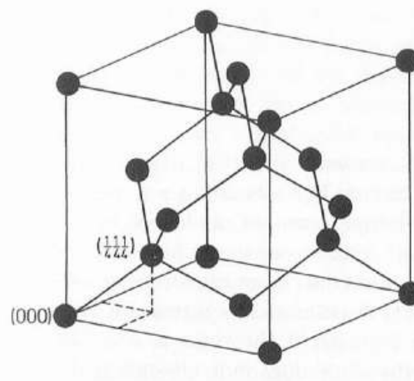


Figure 2-13 Schematics of a diamond crystal structure [2.47].

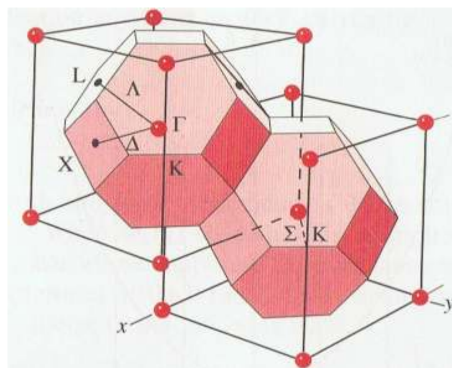


Figure 2-14 Reciprocal lattice of a face-centred cubic lattice with its first Brillouin zone [2.48].

2.2.2 *E-k* diagram of unstrained bulk silicon

*E-**k*** diagram allows us to study the interactions with photons and phonons where energy (*E*) and momentum (***k***) have to be conserved. The interactions with electrons and holes lead to the concept of bandgap. $\hbar\boldsymbol{k}$ is regarded as the crystal momentum rather than true electron momentum because it contains the effects of the internal crystal potential. *E-**k*** diagram is obtained by solving the Schrödinger equation of an approximate one-electron problem. The Bloch theorem states that if a potential function, $V(\boldsymbol{r})$ is periodic in the crystal lattice space, then the solutions for the wavefunction $\psi(\boldsymbol{r}, \boldsymbol{k})$ of the Schrödinger equation are of the form of a Bloch function.

$$\left[-\frac{\hbar^2}{2m^*}\nabla^2 + V(\boldsymbol{r})\right]\psi(\boldsymbol{r},\boldsymbol{k}) = E(\boldsymbol{k})\psi(\boldsymbol{r},\boldsymbol{k}) \tag{2.3}$$

$$\psi(\boldsymbol{r},\boldsymbol{k}) = \exp(i\,\boldsymbol{k}\cdot\boldsymbol{r})U_{\mathrm{b}}(\boldsymbol{r},\boldsymbol{k}) \tag{2.4}$$

where $U_{\mathrm{b}}(\boldsymbol{r},\boldsymbol{k})$ is a periodic function with the same periodicity as $V(\boldsymbol{r})$. Since $E(\boldsymbol{k})$ is periodic in the reciprocal lattice, it is sufficient to study *E-**k*** diagram of the first Brillouin zone. There are two main categories of bandstructure calculations: (i) methods that describe the entire valence band and conduction bands, and (ii) methods that describe the near bandedge bandstructures. Pseudopotential method [2.50] assumes that the core electrons of an atom are tightly bound to its nucleus by an attractive core potential, and thus are treated as if they are frozen in an atomic-like configuration. The repulsive potential of core electrons will keep the valence electrons out of the nucleus of the atom. Hence, the net effect of the two opposing potentials on the valence electrons is a weak potential, which is also known as the pseudopotential.

Tight-binding method has a realistic description of the structural and the dielectric properties in terms of the chemical bonds. Using a minimal sp^3 basis and the interactions between the nearest neighbours, this method can describe the valence bands but fails to reproduce the conduction band of indirect bandgap semiconductor such as silicon [2.51, 2.52] because it omits the essential physics of the interactions between the excited atomic states and the anti-bonding p-like conduction states of Si near the *X* point of the Brillouin zone [2.53]. To mimic the influence of the excited state, s^* orbital is added to the sp^3 basis [2.53] but the transverse masses of the conduction band valleys are in poor agreement with the experimental values [2.54]. Following the recognition of the importance of *d* states in the pseudopotential calculations [2.55], $sp^3d^5s^*$ TB model is proposed [2.56].

The $\boldsymbol{k \cdot p}$ method uses a parameterization of the bands at a given point in the Brillouin zone as a starting point, and then uses the perturbation theory to obtain the other points in the Brillouin zone. It is known to be very efficient to accurately describe either the conduction band [2.57] or the valence band [2.58] in the vicinity of a given point of the Brillouin zone. By disregarding the effects of spin-orbit coupling, $\boldsymbol{k \cdot p}$ method with 15 bands can be used to generate silicon bandstructure [2.59, 2.60]. Although the accuracy will be increased when the split-off band is considered, the simulation time will be significantly increased [2.61].

Figure 2-15 Schematics of energy band diagram of silicon [2.61].

Figure 2-16 Schematics of the conduction band valleys of silicon

Figure 2-17 Schematics of valence band maximum of silicon [2.62].

Fig. 2-15 shows the E-$\boldsymbol{k}$ diagram of silicon. The conduction band minimum occurs at wave vector $\boldsymbol{k} \approx 0.8\ X = 0.8\ (1,0,0)\ (2\pi/a_o)$ [2.60]. From Fig. 2-16, there are four in-plane valleys (Δ_1, Δ_2, Δ_3, Δ_4) and two out-of-plane valleys (Δ_5, Δ_6). The valence band maximum occurs at Γ. From Fig. 2-17, the heavy hole (HH) band and the light hole (LH) band are degenerate at Γ but the hole energy of the split-off (SO) band is 0.044 eV higher than the HH band and LH band [2.60, 2.62]. This splitting of the bands is caused by the spin-orbit interaction. Fig. 2-18 shows that LH band is relatively isotropic while the HH band is anisotropic [2.63]. This will affect the hole mobility along <110> channel direction and <100> channel direction on (100) surface-oriented silicon wafer. (refer to Chapter 5).

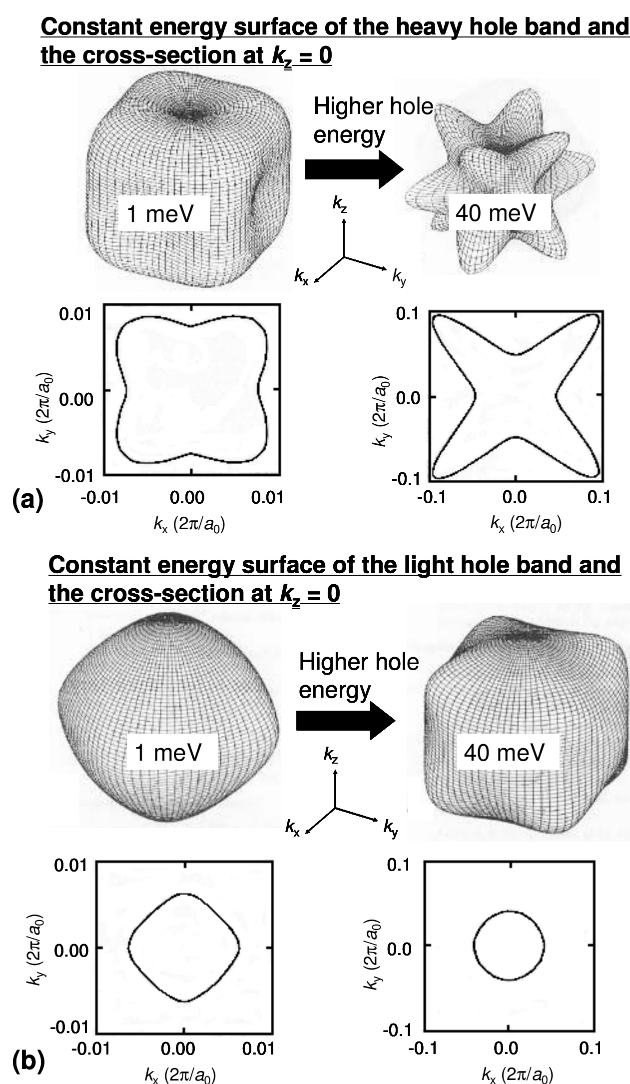


Figure 2-18 Constant energy surface and cross section of valence band maximum at k_z =0 plane [2.63].

2.2.3 Relationship between the effective mass and the *E-k* diagram

Since the solutions of Schrödinger equation can be approximated by the Bloch function, it is sufficient to study the localized. Using our understanding on waves, we can define the group velocity of this wavepacket as [2.64],

$$v_g = \frac{d\omega}{d\boldsymbol{k}} = \frac{1}{\hbar}\frac{dE}{dk} \tag{2.5}$$

where the electron energy (E) is the product of the angular frequency (ω) and the reduced Planck's constant ($\hbar$). In the presence of an external electric field, the work done on the electron during a time interval (δt) by an external force (F_{ext}) is given by [2.64],

$$\delta E = \boldsymbol{F}_{ext} \cdot \boldsymbol{v}_g \delta t = \left(\frac{dE}{d\boldsymbol{k}}\right)\delta \boldsymbol{k} = \hbar \boldsymbol{v}_g \cdot \delta \boldsymbol{k} \tag{2.6}$$

Hence, $\hbar \dfrac{d\boldsymbol{k}}{dt} = \boldsymbol{F}_{ext}$ (2.7)

$$\frac{dv_g}{dt} = \frac{1}{\hbar}\frac{d^2E}{dkdt} = \frac{1}{\hbar}\frac{d^2E}{dk^2} \times \frac{dk}{dt} \tag{2.8}$$

Substitute eqn.(2.7) into eqn.(2.8), and then re-arranging,

$$\boldsymbol{F}_{ext} = \frac{\hbar^2}{\left(\dfrac{d^2E}{dk^2}\right)}\frac{dv_g}{dt} \tag{2.9}$$

Eqn.(2.9) takes the form of the Newton's equation of motion. Hence, the effective conductivity mass (m^*) can be expressed as [2.64],

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{d^2E}{d\boldsymbol{k}^2} \tag{2.10}$$

From eqn.(2.10), a bigger curvature in the ***E-k*** diagram means a smaller m^*. Here, it is important to note that the electron have positive m^* at conduction band (E_c) minimum and negative m^*at the valence band (E_v) maximum [2.65]. Positive m^* refers to the effective conductivity mass of electron, whereas negative m^* refers to the effective conductivity mass of hole.

2.2.4 Effects of mechanical stress on the conduction band minimum

For unstrained silicon, the conduction band minimum is made up of six degenerate conduction band valleys. From Fig. 2-19, mechanical stress will cause the splitting of the conduction band edge (ΔE_c) and the resulting electron energy level of the conduction band valleys (Δ_1 to Δ_6) will depend on the type of mechanical stress [1.14]. Considering (100) Si, uniaxial <110> tensile stress causes the electron energy level of the out-of-plane conduction band valleys (Δ_5, Δ_6) to be lower than the electron energy level of the in-plane conduction band valleys (Δ_1 to Δ_4). Since electrons will preferentially occupy the lowest electron energy level, the in-plane effective electron mass will be smaller than the out-of-plane effective electron mass. However, the situation is different for uniaxial tensile stress along <100> channel direction. Taking the uniaxial <100> tensile stress to be along the direction of Δ_1 and Δ_2, the electron energy level of the in-plane conduction band valleys (Δ_1, Δ_2) will be bigger than the other conduction band valleys (Δ_3 to Δ_6), and thus the in-plane effective electron mass along the channel direction will be reduced.

From Fig. 2-20, the splitting of the conduction band edge (ΔE_c) for biaxial tensile stress and uniaxial <100> tensile stress can reproduce the experimental low-field mobility enhancement. However, this is not the case for uniaxial <110> tensile stress. If we only consider the effects of ΔE_c, the electron mobility enhancement of the calculated $\sigma_{//}$ and the calculated $\sigma_{\perp}$ will deviate from that of the experimental $\sigma_{//}$ and the experimental $\sigma_{\perp}$, as shown in Fig. 2-20(c). Note that $\sigma_{//}$ and $\sigma_{\perp}$ refer to the cases where the mechanical stress is applied along the channel direction and perpendicular to the channel direction, respectively. This discrepancy can be explained as follows. Unlike <100> uniaxial stress, the energy surface of the out-of-plane valleys (Δ_5, Δ_6) is warped under uniaxial <110> tensile stress, as shown in Fig. 2-21 [1.14]. The transverse mass of the out-of-plane conduction band ellipsoid that is perpendicular to the direction of the uniaxial <110> tensile stress ($m_{t,\perp}$) is bigger than the transverse mass of the out-of-plane conduction band ellipsoid under the unstrained condition (m_t). On the other hand, the transverse mass of the out-of-plane conduction band ellipsoid that is parallel to the direction of the uniaxial <110> tensile stress ($m_{t,//}$) is smaller than m_t. From Fig. 2-22, when uniaxial <110> tensile stress increases, $m_{t,//}$ decreases but $m_{t,\perp}$ increases [2.66]. This leads to a decrease in the effective conductivity mass (m^*) under uniaxial <110> tensile stress along the channel direction (also known as <110> longitudinal stress). If band warping is not considered, which is the case for uniaxial <100> tensile stress, the strain-induced reduction in m^* will

saturate at higher tensile stress [2.66] because the change of electron population and the suppression of the intervalley phonon scattering saturate at larger splitting of the conduction band edges [1.14]. On the other hand, electron mobility under uniaxial <110> tensile stress does not saturates at higher stress because of the warping of the out-of-plane conduction band valleys, as shown in Fig. 2-23 [1.14]. As a result, uniaxial <110> tensile stress is more advantageous than biaxial tensile and uniaxial <100> tensile stress in terms of electron mobility enhancement, particularly at higher tensile stress.

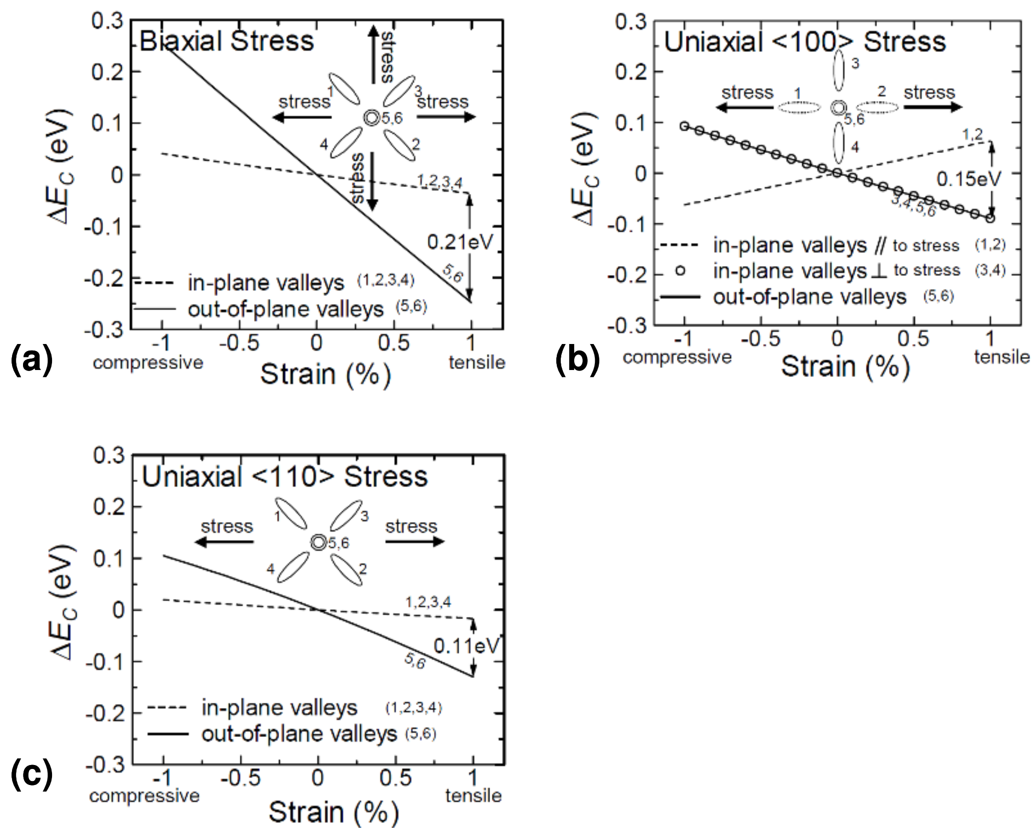


Figure 2-19 Splitting of the conduction band edge (ΔE_c) as a function of (a) biaxial stress, (b) uniaxial <100> stress, (c) uniaxial <110> stress [1.14].

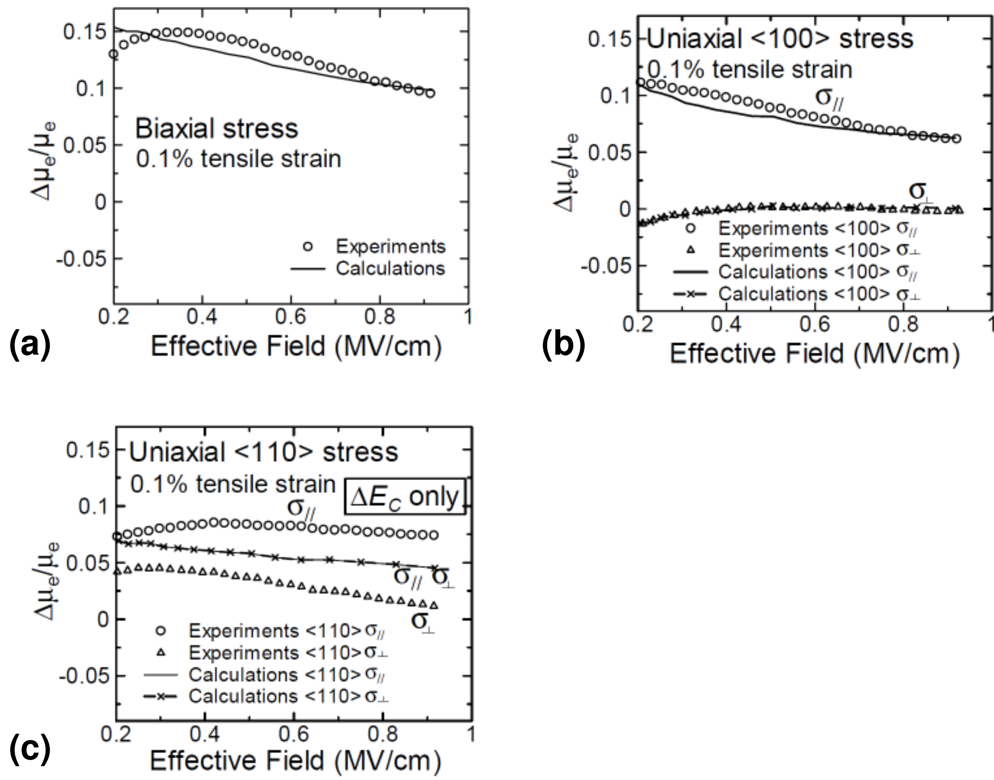


Figure 2-20 Electron mobility enhancement ($\Delta\mu_e/\Delta\mu_e$) as a function of the vertical electric field under (a) biaxial stress, (b) uniaxial <100> stress, (c) uniaxial <110> stress [1.14].

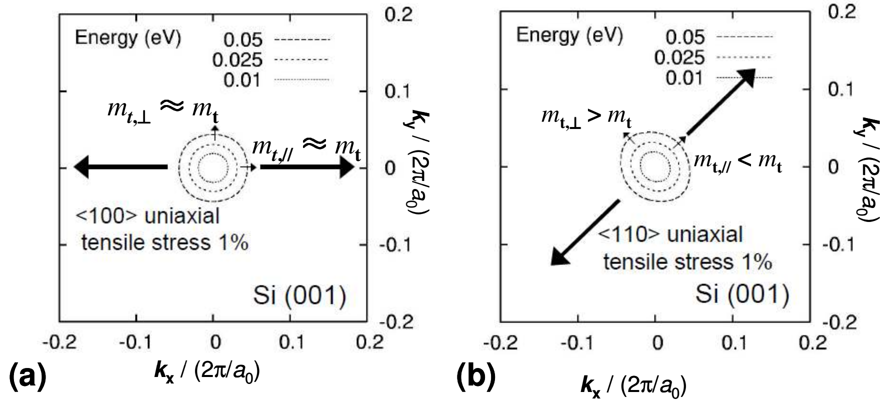


Figure 2-21 Energy contours of the out-of-plane conduction band valleys (Δ_5, Δ_6) on (100) silicon plane under (a) uniaxial <100> tensile stress, (b) uniaxial <110> tensile stress [1.14].

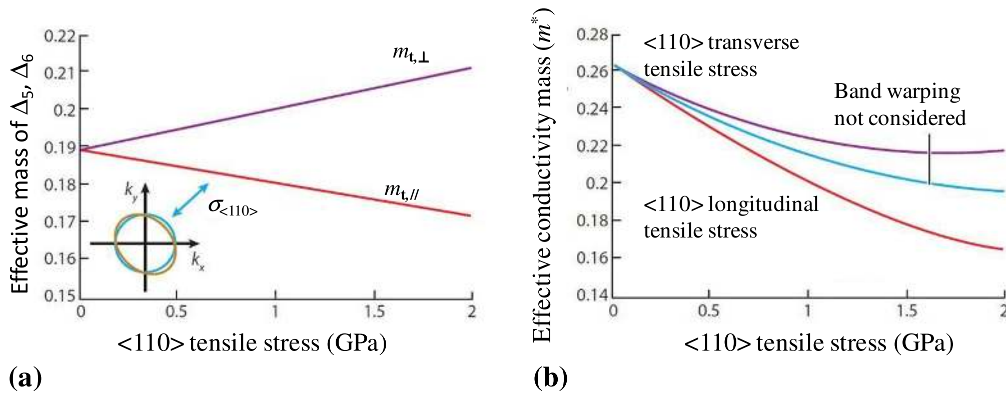


Figure 2-22 The effects of <110> uniaxial tensile stress on: (a) the longitudinal effective mass ($m_{t,//}$) and the transverse effective mass ($m_{t,\perp}$) of the out-of-plane conduction valleys, (b) the effective conductivity mass (m^*) [2.66].

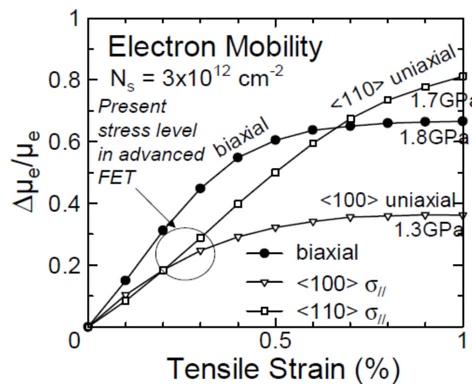


Figure 2-23 Electron mobility enhancement as a function of mechanical stress [1.14].

2.2.5 Effects of mechanical stress on the valence band maximum

Fig. 2-24 shows the constant energy contours of the light hole band maximum and heavy hole band maximum for the unstrained bulk silicon [1.15, 2.63]. The constant energy contours of the heavy hole band of the unstrained silicon are characterized by twelve low energy branches referred to as "wings". k_x, k_y and k_z are the wave vectors along *x*, *y* and *z* directions, respectively. a_o is the lattice constant of the unstrained silicon. Considering (100) plane with $k_z = 0$, there are eight out-of-plane wings (O1, O2, O3, O4, O5, O6, O7, O8) and the four in-plane wings (I1, I2, I3, I4). Figs. 2-25 shows the effects of mechanical stress on the constant energy contours of the lowest hole energy. The application of 1 GPa compressive stress along <110> channel direction leads to a lowering in the hole energy of I1 and I3 wings but a rise in the hole energy of I2 and I4 wings. Hence, there will be a carrier repopulation from I2 and I4 wings to I1 and I3 wings. Since

the <110> channel direction is along the direction of I2 and I4 wings, there will be a hole mobility enhancement along the channel direction. The application of 1 GPa tensile stress along <110> channel direction leads to the opposite conclusion: the hole carriers are redistributed from I1 and I3 wings to I2 and I4 wings, leading to a hole mobility degradation along <110> direction. From Fig. 2-24, it is worth noticing that the eight out-of-plane wings (O1 to O8) are aligned 45° away from the $\boldsymbol{k_z}$ axis [1.15]. The presence of a vertical confinement field in the *z*-direction will reorder $\boldsymbol{k_z}$ so that the lowest hole energy is on the first subbands of the out-of-plane wings, and thus project the non-zero $\boldsymbol{k_z}$ out-of-plane wings into the 2-D Brillouin zone.

Fig. 2-26 shows the effects of the vertical confinement field on the constant energy contours of the lowest hole energy subband of (100) silicon plane [1.15]. It is quite similar to the case of the unstrained bulk silicon (see Fig. 2-25) except for the presence of the out-of-plane wings. Note that O5 through O8 wings cannot be seen because O5 through O8 wings have the same energy as O1 through O4 wings. From Fig. 2-26(b), the application of uniaxial <110> compressive stress to the hole inversion layer will lower the hole energy of I1 and I3 wings but increase in the hole energy of I2 and I4 wings, resulting in hole mobility enhancement along <110> direction [1.15]. From Fig. 2-25(b), the application of uniaxial <110> tensile stress to bulk silicon will lower the hole energy of I2 and I4 wings but increase in the hole energy of I1 and I3 wings, leading to a hole mobility degradation along <110> direction. Hence, we would expect that the application of uniaxial <110> tensile stress to the hole inversion layer will degrade the hole mobility along <110> direction.

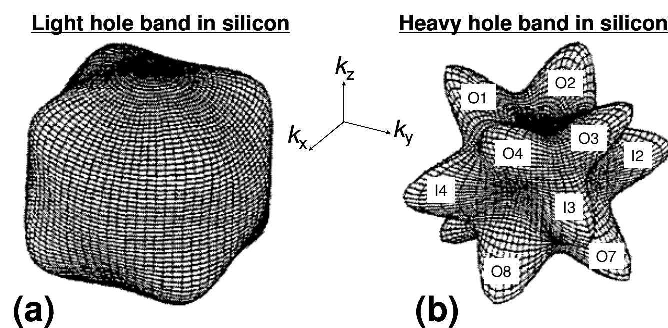


Figure 2-24 Constant energy contours of (a) the light hole band and (b) heavy hole band at Γ for unstressed bulk silicon [1.15, 2.63]. k_x, k_y and k_z are the wave vectors along x, y and z directions.

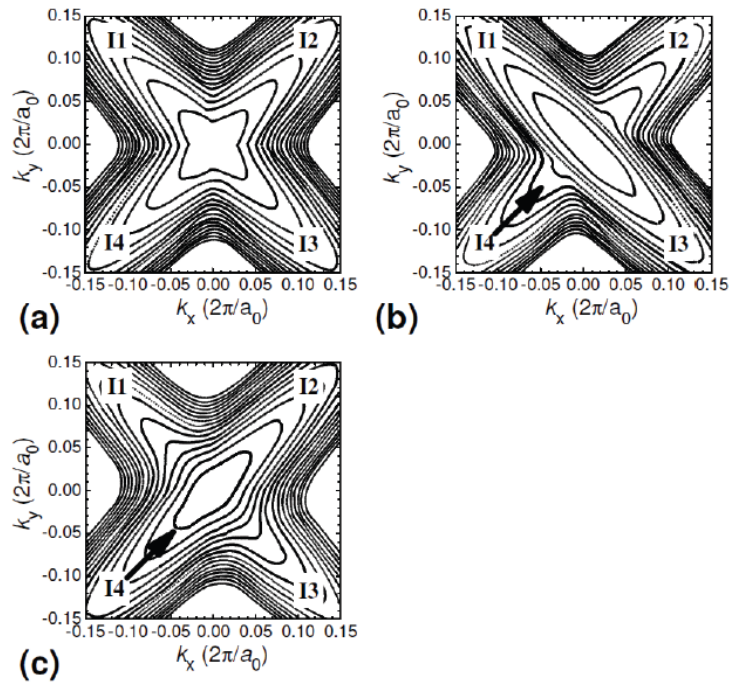


Figure 2-25 Constant energy contours separated by 25 meV for the lowest hole energy band of (100) bulk silicon at Γ under (a) the unstressed condition, (b) the uniaxial compressive stress along <110> direction, and (c) the uniaxial tensile stress along <110> direction [1.15].

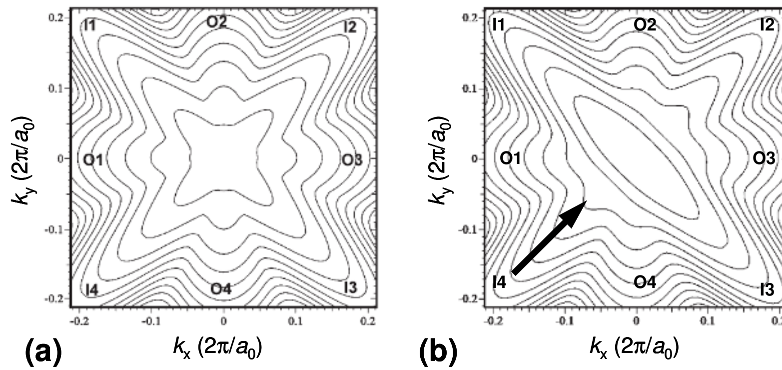


Figure 2-26 Constant energy contours of the lowest hole energy band at Γ of silicon inversion layer in (100) plane with vertical confinement field of 1 MV/cm under (a) the unstrained condition, (b) the uniaxial compressive stress along <110> channel direction (as indicated by the arrow) [1.15].

2.2.6 Low-field mobility: Drude model

In semiconductors, the collisions experienced by the free electrons or holes are mostly caused by phonons and doping impurities. Electron-electron collisions are much less important. The simplest transport model was developed by Drude. Its main assumption is that the collisions act as a damping force on the electrons and thus prevents the electrons from being indefinitely accelerated by the applied electric field. The expression for the low-field mobility (μ_{eff}) is given by [2.66],

$$\mu_{\text{eff}} = \frac{q\tau_{\text{m}}}{m^*} \tag{2.11}$$

where τ_{m} is the momentum relaxation time. The low-field mobility can be improved by: (i) a smaller m^*, (ii) a larger energy split between subbands to suppress phonon scattering , and (iii) a smaller density-of-state effective mass (m_{DOS}) has a smaller number of final states and thus reduce the scattering frequency. According to Guillaume and Mouis [2.67], the application of uniaxial compressive stress along <110> direction on (100) Si plane will decrease the m_{DOS} of the valence subband with the lowest hole energy. For (100) surface-oriented silicon, $m^*$is given by $3(m_\ell^{-1} + 2\, m_{\text{t}}^{-1})^{-1}$, whereas m_{DOS} is given by $\sqrt[3]{m_\ell m_{\text{t}} m_{\text{t}}}$ [2.68]. From cyclotron measurements, m_ℓ is $0.97m_0$ while m_{t} is $0.19m_{\text{o}}$ where m_{o} is the free electron mass [2.69].

2.2.6.1 Phonon scattering

Atoms in crystals vibrate naturally around their equilibrium lattice positions because of heat transfer. The positional disturbance of atoms will be transmitted in the manner as if the atoms were connected to each other by massless springs. For silicon, there are two atoms in a primitive cell and thus there are two vibrational modes, namely the optical phonon and the acoustic phonon. In addition, phonons can be further classified based on the oscillation direction and wave propagation direction: longitudinal phonons and transverse phonons. In short, the phonons are also classified by their range of frequencies in the phonon dispersion spectrum: the higher frequency mode consists of the longitudinal optical (LO) phonons and the transverse optical (TO) phonons while the lower frequency mode consists of the longitudinal acoustic (LA) phonons and the transverse acoustic (TA) phonons. For acoustic phonon, the two atoms in the unit cell oscillate in phase [2.70]. For optical phonon, the two atoms in the unit cell oscillate out of phase [2.70].

Fig. 2-27 shows the phonon dispersion spectrum of silicon along the <100> direction of the first Brillouin zone ($\Gamma \rightarrow \Delta \rightarrow X$) [2.71]. Intra-valley scattering refers to the scattering within the same conduction band valley and usually involves only acoustic phonons [2.71]. Since mechanical stress removes the degeneracy of the six conduction band valleys, intervalley scattering would be affected by mechanical stress. The intervalley scattering is of *g*-type when the electrons scatter between conduction band valleys on the same *k* axis (e.g. from Δ_5 to Δ_6 as defined in Fig. 2-16). The intervalley scattering is of *f*-type when electrons scatter between conduction band valleys on the perpendicular *k* axis (e.g. from Δ_1 to Δ_5 as defined in Fig. 2-16). As shown in Fig. 2-15, the conduction band minimum occurs near to *X* of the first Brillouin zone. From Fig. 2-26, the phonon scattering can be fully suppressed if the energy splitting of the conduction band valleys can be bigger than 60 meV. Similar analysis can be done for valence band maximum. According to Sun et al. [2.72] and Thomson et al. [2.73], the process-induced compressive stress , which are typically around 1 GPa, results in valence band splitting of about 20 – 30 meV. Since the optical phonon energy of silicon is over 60 meV, the phonon scattering rate is only slightly reduced. Hence, the strain-induced hole mobility enhancement is mainly caused by the warping of the valence subbands rather than the strain-induced reduction in the phonon scattering.

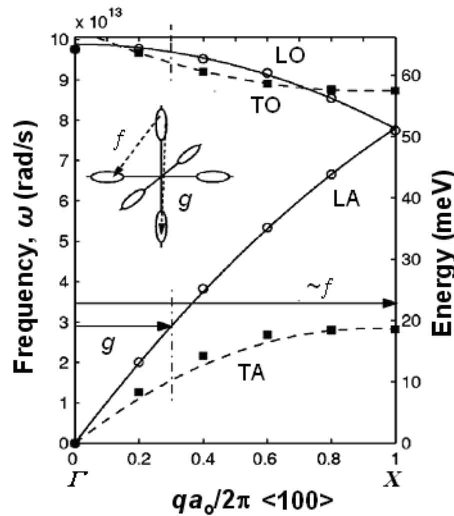


Figure 2-27 Phonon dispersion spectrum of silicon along the (100) direction of the first Brillouin ($\Gamma \rightarrow \Delta \rightarrow X$). The *f* and *g* phonons participate in the intervalley scattering of electrons [2.71].

2.3 Usage of Piezoresistance coefficients

The most accurate approach to estimate the strain-induced change in the low-field mobility ($\Delta\mu/\mu$) is to use the bandstructure calculation together with the Monte Carlo simulation of mobility [1.14]. However, an approximate piezoresistance model is often used to organize the experimental data owing to its simplicity. For silicon, the piezoresistance coefficient (π_{ij}) is related to the change in piezoresistivity ($\Delta\rho_\kappa/\rho_\kappa$) and the mechanical stress (σ_κ) where the subscript i, j = 1, 2, 3, 4, 5, 6 and κ = 1, 2, 3, 4, 5, 6 [2.74, 2.75]. For silicon, there are only three non-zero independent components of piezoresistive tensor [2.76], as shown in Table 2-1.

Table 2-1 Components of the piezoresistive tensor of bulk silicon (units: 10^{-11} Pa^{-1}) [2.74].

	π_{11}	π_{12}	π_{44}
p-Si	6.6	-1.1	138.1
n-Si	-102.2	53.4	-13.6

$$\begin{bmatrix} \Delta\mu_1/\mu \\ \Delta\mu_2/\mu \\ \Delta\mu_3/\mu \\ \Delta\mu_4/\mu \\ \Delta\mu_5/\mu \\ \Delta\mu_6/\mu \end{bmatrix} \approx - \begin{bmatrix} \Delta\rho_1/\rho \\ \Delta\rho_2/\rho \\ \Delta\rho_3/\rho \\ \Delta\rho_4/\rho \\ \Delta\rho_5/\rho \\ \Delta\rho_6/\rho \end{bmatrix} = - \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{11} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{12} & \pi_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \pi_{44} \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix} \quad (2.12)$$

Using matrix manipulations [2.75, 2.76, 2.40, 2.41], eqn. (2.12) can be expressed as $\Delta\mu/\mu \approx -(\pi_{//}\sigma_{//} + \pi_{\perp}\sigma_{\perp} + \pi_{out}\sigma_{out})$. $\sigma_{//}$, $\sigma_{\perp}$ and σ_{out} are the mechanical stress that are parallel to L, perpendicular to L, and along the quantum confinement direction, respectively. $\pi_{//}$, $\pi_{\perp}$ and π_{out} are the piezoresistance coefficients that are parallel to L, perpendicular to L, and along the quantum confinement direction, respectively. Tensile stress is taken to be positive stress while compressive stress is taken to be negative stress. From eqn.(1.1), the saturation drain current of long-channel MOS transistor is related to the low-field mobility. Hence, the piezoresistance coefficient can be found from the slope of the $\Delta\mu/\mu$ versus the uniaxial stress plot and $\Delta I_{ds}/I_{ds}$ versus uniaxial stress plot [2.41, 2.78]. Table 2-2 shows the device-level piezoresistance coefficients for MOS transistors fabricated on (100) Si wafer with <110> channel and <100> channel. A negative piezoresistance coefficient

means that tensile stress will improve the mobility. On the other hand, a positive piezoresistance coefficient means that compressive stress will improve the mobility. The differences in extracted π-coefficients [2.41, 2.77, 2.78, 2.79, 2.80] can be attributed to the variation in the channel doping concentration of the transistors [2.81]. Uniaxial tensile stress along the channel direction is beneficial to both <110> NMOS transistor and <100> NMOS transistor. The application of the uniaxial compressive stress along the channel direction is beneficial to <110> PMOS transistor. However, piezoresistance coefficient predicts that <100> PMOS transistor is virtually insensitive to mechanical stress. This is consistent with experimental results [1.35, 2.82]. Unlike <100> mechanical stress, <110> mechanical stress contains shear-strain components $\delta_{xy} = \delta_{yx}$. Since shear strain deformation potential is higher than the other strain deformation potentials, the warping of valence subband structure is more significant for <110> mechanical stress (see Fig. 2-26 and Fig. 2-28) [1.15, 2.44]. In fact, the valence subband structure for unstrained silicon and <100> uniaxial stress are quite similar.

Table 2-2 Device-level piezoresistance coefficients of the MOS transistors fabricated on (100) surface-oriented silicon wafer (units: 10^{-11} Pa^{-1}).

	<110> channel direction		<100> channel direction	
	NMOS	PMOS	NMOS	PMOS
$\pi_{//}$	(-49)[a]	(90)[a]	(-38.6)[c]	(9.1)[c,d]
	(-48.5)[b]	(60)[b]	(-42.6)[d]	(-15)[e]
	(-35.5)[c,d]	(71.7)[c,d]	(-47.7)[e]	
	(-32)[e]	(71)[e]		
$\pi_{\perp}$	(-16)[a]	(-46)[a]	(-18.7)[c]	(-6.2)[c,d]
	(-21.2)[b]	(-38.8)[b]	(-20.7)[d]	(9)[e]
	(-14.5)[c,d]	(-33.8)[c,d]	(-22)[e]	
	(-15)[e]	(-32)[e]		
π_{out}	(87)[a]	(-44)[a]		
	(27)[e]	(-20)[e]		

[a]after [2.41],[b] after [2.77], [c] after [2.78], [d] after [2.79], [e] after [2.80].

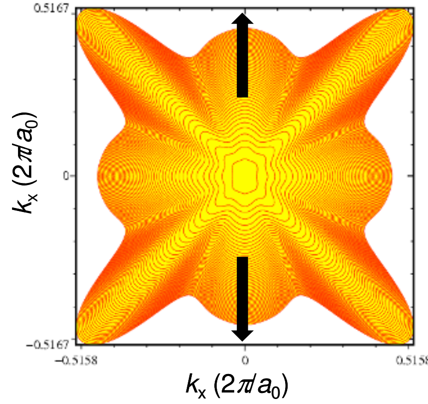


Figure 2-28 First Heavy-hole subband equivalent energy lines of (100) surface-oriented silicon under uniaxial <100> tensile stress of 1.5 GPa (Courtesy of Dr Pham from BST, TU Braunschweig, Germany, Ref. [2.44]).

From Table 2-2, for <110> NMOS transistor fabricated on (100) Si, π_{out} is positive, whereas $\pi_{//}$ is negative. This means that uniaxial tensile stress will decrease the in-plane effective electron mass along the channel direction (m_x) but increase the out-of-plane effective electron mass (m_z). This is consistent with Nakabayashi et al. [2.83] who reported that the channel carriers with lighter m_x will have heavier m_z. On the other hand, the uniaxial compressive stress will decrease hole m_x but increase the hole m_z [2.84].

However, the accuracy of the approximate piezoresistance model in eqn. (2.13) is only restricted to low mechanical stress [2.78]. This can be understood using the relation: resistivity, ρ is the reciprocal of the product of the electron charge, low-field mobility and carrier concentration.

$$\frac{\Delta\rho}{\rho_u} = \frac{\rho_s - \rho_n}{\rho_u} = \frac{\mu_u - \mu_s}{\mu_s} \tag{2.13a}$$

$$\frac{\Delta\mu}{\mu_u} = \frac{\mu_s - \mu_u}{\mu_u} \tag{2.13b}$$

where ρ_s , ρ_u , μ_s and μ_u are the resistivity of the stress-engineered MOS transistor, the resistivity of the unstrained MOS transistor, the low-field mobility of the stress-engineered MOS transistor, and the low-field mobility of the unstrained MOS transistor, respectively. At low mechanical stress, $\mu_s \approx \mu_u$,

$$\frac{\Delta\mu}{\mu_u} \approx \frac{\Delta\rho}{\rho_u} \tag{2.14}$$

However, μ_s and μ_u differ significantly at higher level of mechanical stress. Tsang et al.[2.85] developed a *C-R* conversion by combining eqn.(2.12) , eqn.(2.13a) and eqn.(2.13b)

$$\frac{\Delta\mu_\kappa}{\mu} = \frac{1}{\left(\frac{\Delta\rho_\kappa}{\rho}+1\right)} - 1 = \frac{1}{\left(\sum_\kappa \pi_{ij}\sigma_\kappa\right)+1} - 1 \qquad (2.15)$$

From Fig. 2-29, *C-R* conversion model (eqn.2.15) has a better fit to the experimental $\Delta\mu/\mu$ versus the externally applied uniaxial stress as compared to the approximate piezoresistance model (eqn. 2.12).

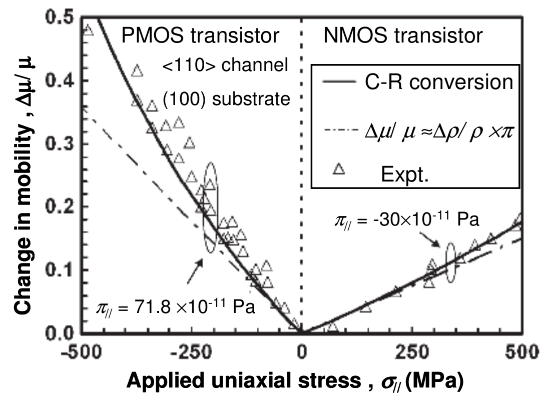


Figure 2-29 Non-linearity of piezoresistance coefficient at higher uniaxial stress [2.85]. C-R conversion model is represented by eqn. (2.16). $\Delta\mu/\mu \approx \Delta\rho/\rho \times \pi$ refers to the piezoresistance model, which is given by eqn. (2.13).

3. Drain current transport in nanoscale MOS transistors

This chapter will evaluate the merits and limitations of various drain current transport models. According to Natori [1.2.1], the carrier transport in MOS transistors can be classified based on the relative dimension between the gate length (L) and the mean free path (λ), as illustrated in Fig. 3-1. Qualitatively, λ is the average distance covered by the channel carrier between the successive collisions. λ is estimated to be in the 10 nm range [3.1]. Hence, the probability of a carrier encountering scattering events within the channel is expected to decrease when the transistor scaling intensifies. Since mobility is a concept involving channel scattering, mobility becomes irrelevant for ballistic transport.

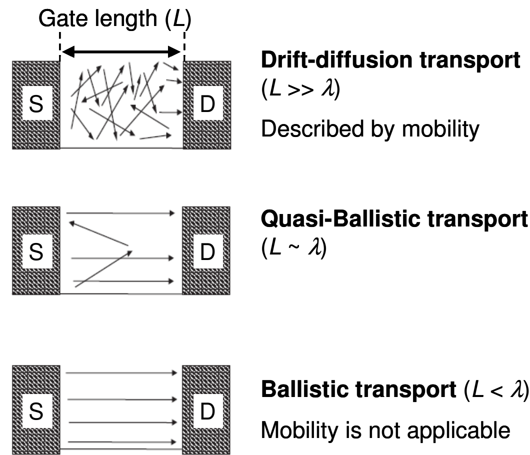


Figure 3-1 Classification of carrier transport in MOS transistors [1.21].

3.1 Velocity saturation model

The main advantage of this model is the simplicity of its saturation drain current equation for the nanoscale MOS transistor: $I_{ds} = v_{sat} C_{ox,inv} W (V_{GS} - V_{th,sat})$. From Fig. 3-2, v_{sat} for electrons is 10^7 cm/s , whereas v_{sat} for holes is 6×10^6 cm/s at temperature of 300 K [1.9]. Based on the theoretical predictions [1.10], v_{sat} is independent of μ_{eff}. In the other words, the velocity saturation model predicts that the I_{on} of the nanoscale MOS transistors cannot be increased by mobility enhancement techniques. This is contradictory to the experimental results of strain-induced I_{on} improvement in nanoscale MOS transistors [1.11-1.13]. In addition, Tatsumura et al. [1.29] and Khakifirooz [3.2] have experimentally proven that the high-field carrier velocity is indeed related to the low-field mobility.

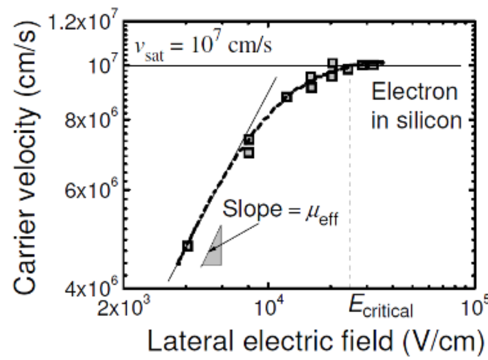


Figure 3-2 Time-of-flight measurement suggests that the low-field mobility (μ_{eff}) is independent of carrier velocity at high lateral electric field [1.9].

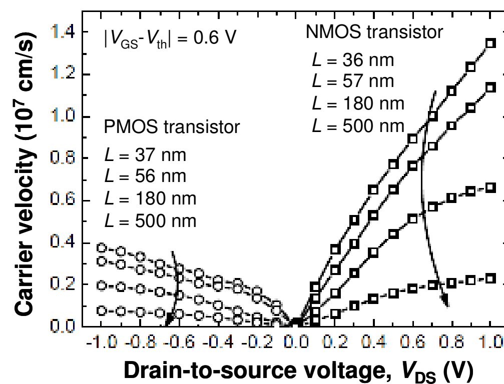


Figure 3-3 Electron velocity overshoot has been observed at room temperature in bulk NMOS transistor with gate length (*L*) of 32 nm [1.19].

Monte Carlo simulation [1.16-1.18] and experimental data [1.19] showed that velocity overshoot can occur in the nanoscale MOS transistors. Fig. 3-3 shows that electron velocity overshoot can be observed at room temperature in bulk NMOS transistor with nominal gate length (L) of 37 nm [1.19]. The physics behind the observation of velocity overshoot can be understood as follows. Lateral electric fields well-above 10^4 V/cm are present in the channel of a nanoscale MOS transistor. This is certainly high enough to cause velocity saturation in bulk silicon, but transients can occur in short, high electric field regions. This can be understood as follows. Electrons are first injected into the channel and then are accelerated by the lateral electric field. Although several phonon collisions are required to reduce a carrier's energy, a single large angle scattering event can remove all of its directed momentum, and this causes the energy relaxation times to be longer than the momentum relaxation times [3.3]. As a consequence, the mobility is initially high, so the velocity can be very high. As the energy increases, scattering increases and thus mobility decreases, and the velocity eventually decreases to 10^7 cm/s

(saturation velocity of electrons in silicon). From Fig.3-4, the spatial width of the transient is approximately 100 nm [3.4]. Since the gate length of the state-of-the-art MOS transistors is below 100 nm, strong velocity overshoot can be expected in the state-of-the-art MOS transistors.

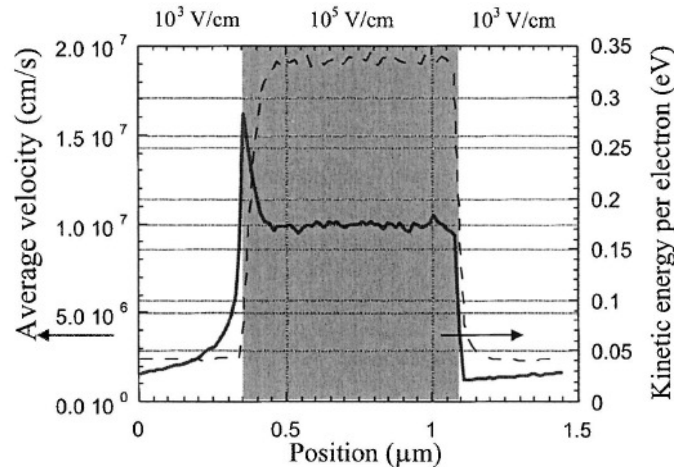


Figure 3-4 The average velocity versus position for electrons injected into a short slab of silicon with a high electric field [3.4]. Dotted line refers to the average velocity of electrons. Solid line refers to the kinetic energy of the electrons.

Another common misconception about velocity saturation is related to the drain current saturation in the drain current (I_{ds}) versus the drain-to-source (V_{DS}) characteristics at high V_{DS}. From eqn. (1.2), the velocity saturation model predicts that the saturation I_{ds} of the short channel MOS transistor has a linear relationship with V_{GS}, and thus the I_{ds} versus V_{DS} characteristics of a nanoscale transistor is expected to have constant spacing for equal V_{GS} step at high V_{DS} [3.5]. On the other hand, the saturation I_{ds} of the long channel MOS transistor is controlled by pinchoff [3.6]. Based on the constant mobility assumption, eqn. (1.1) predicts that the saturation I_{ds} of long channel MOS transistor has a quadratic relationship with V_{GS} and thus the saturation I_{ds} versus V_{DS} characteristics is expected to have increasing spacing for equal V_{GS} step [3.5]. However, the constant spacing for equal V_{GS} step is often observed in the experimental I_{ds} versus V_{DS} characteristics of the long channel MOS transistor, as shown in Fig. 3-5. Hence, the constant spacing in the I_{ds} versus V_{DS} characteristics at high V_{DS} cannot be taken as an indication of the onset of velocity saturation. Using the concept of velocity saturation, Suzuki and Usuki [3.7] proposed an equation for the saturation drain voltage (V_{Dsat}) that can account for the disparity between the experimental V_{Dsat} and the V_{Dsat} that is predicted by the classical pinchoff theory [3.8] : $V_{Dsat} = V_{GS} - V_{th,sat}$.

$$V_{\mathrm{Dsat}} = \frac{V_{\mathrm{GS}} - V_{\mathrm{th,sat}}}{0.5 + \sqrt{0.25 + \frac{\mu_{\mathrm{e}}\left(V_{\mathrm{GS}} - V_{\mathrm{th,sat}}\right)}{v_{\mathrm{sat}} L_{eff}}}} \tag{3.1}$$

where μ_{e} is the low-field mobility. v_{sat} is the saturation velocity. Based on the conventional MOS transistor theory, V_{Dsat} is given by $(V_{\mathrm{GS}} - V_{\mathrm{th,sat}})/m$ where $1.1 \leq m \leq 1.4$ [3.9].

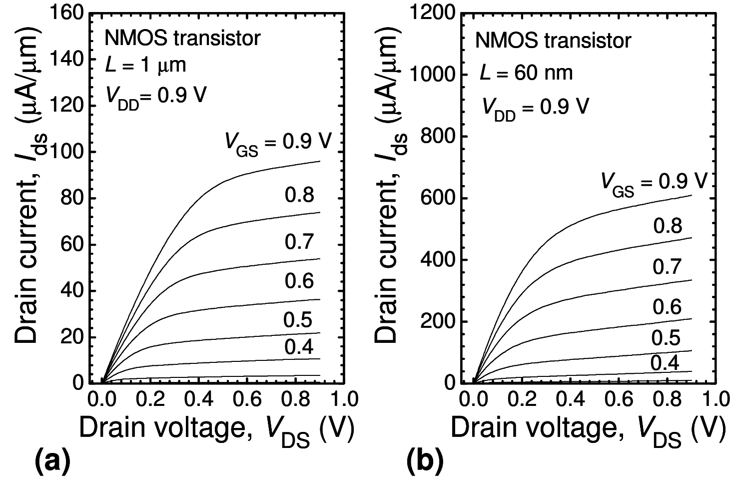


Figure 3-5 Equal spacing in the I_{ds} versus V_{DS} characteristics at high V_{DS} is observed in (a) long-channel NMOS transistor, (b) short-channel NMOS transistor.

3.2 Ballistic transport model

According to Natori [1.20], the saturation drain current of a nanoscale MOS transistor under ballistic transport is,

$$I_{\mathrm{ds}} = v_{\mathrm{inj}} W C_{\mathrm{ox}} \left(V_{\mathrm{GS}} - V_{\mathrm{th,sat}}\right) \tag{3.2}$$

where C_{ox} is the gate oxide capacitance per unit area. The injection velocity (v_{inj}) is given by [1.20],

$$v_{\mathrm{inj}} = \frac{8\hbar\sqrt{C_{\mathrm{ox}}\left(V_{\mathrm{GS}} - V_{\mathrm{th,sat}}\right)}}{3m_{\mathrm{t}}\sqrt{q\pi M_{\mathrm{v}}}} \tag{3.3}$$

Substituting eqn.(3.3) into eqn. (3.2),

$$I_{\mathrm{ds}} = \frac{8\hbar W\left[C_{\mathrm{ox}}(V_{\mathrm{GS}} - V_{\mathrm{th,sat}})\right]^{3/2}}{3m_{\mathrm{t}}\sqrt{q\pi M_{\mathrm{v}}}} \tag{3.4}$$

where M_v is the effective lowest level degeneracy in the conduction subband that is modified to include the effects of the population in upper subband levels. When V_{GS} increases, the potential barrier at the source-channel region is lowered and thus the carrier density is increased. According to Pauli principle, the carriers are expected to populate higher energy levels and the increase in kinetic energy leads to an increase in the mean velocity [3.1]. Hence, v_{inj} increases with an increase in V_{GS}, as shown in Fig. 3-6 [1.21].

From Fig. 3-7, Natori's ballistic theory [1.20] predicts that an increase in temperature will increase v_{inj}, and thus the saturation drain current is expected to increase with increasing temperature in accordance with eqn. (3.2). However, this theoretical prediction of the temperature behaviour of I_{ds} is contradictory to the experimental data. According to Park et al. [3.10], the experimental I_{ds} will experience a reversal in temperature dependency across the temperature independent point (TIP) owing to the opposing behaviour of the low-field mobility (μ_{eff}) and the threshold voltage with temperature. When V_{GS} is smaller than TIP, I_{ds} will increase with increasing temperature. This can be associated with v_{inj} of Natori's 1994 theory [1.20]. When V_{GS} is bigger than TIP, I_{ds} will decrease with increasing temperature owing to the mobility degradation at elevated temperature [3.11-3.13].

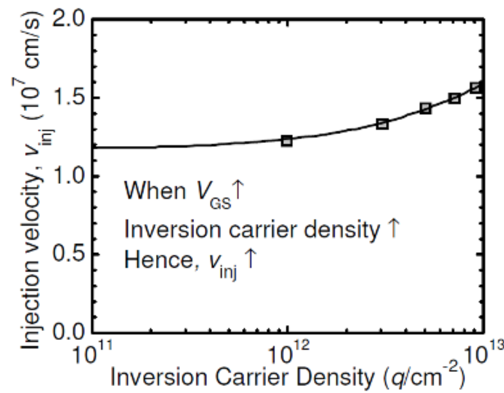


Figure 3-6 Injection velocity (v_{inj}) as a function of the inversion carrier density [1.21].

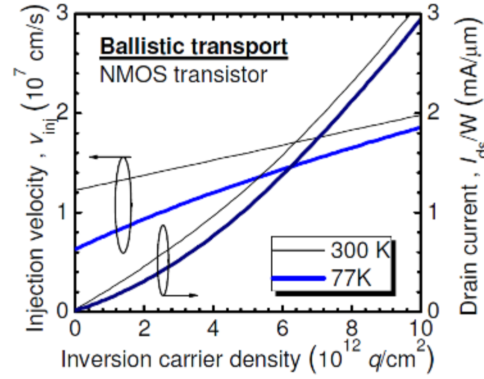


Figure 3-7 Temperature dependency of the injection velocity (v_{inj}) and the saturation drain current of the nanoscale NMOS transistors [1.20].

Another problem with Natori's 1994 theory [1.20] is that his theory predicts that the saturation I_{ds} of the nanoscale MOS transistor will follow a $(V_{GS} - V_{th,sat})^{3/2}$ relationship. From eqn. (3.4), $V_{th,sat}$ can be found by extrapolating the $I_{ds}^{2/3}$ versus V_{GS} characteristics. In this context, we regard this threshold voltage as the extrapolated threshold voltage ($V_{T,\ extrapol}$). First of all, there is clearly no unique gate voltage at which drain current begins to flow. Considering a large geometry NMOS transistor with no short-channel effects and no narrow-channel effects, the threshold voltage of conduction ($V_{T,\ 2\phi F}$) is given by [3.14], $V_{T,2\phi_F} = V_{FB} + \phi_{2F} + \frac{\sqrt{2q\varepsilon_{si}\varepsilon_o N_A(2\phi_F - V_{BS})}}{C_{ox}}$ where V_{BS} is the substrate-source voltage. V_{FB} is the flatband voltage. ϕ_F is Fermi-level of bulk . From Fig.3-8, $V_{T,extrapol}$ is always bigger than $V_{T,\ 2\ \phi F.}$

From Fig. 3-9(a), the measured $I_{ds}^{2/3}$ versus V_{GS} characteristics is smaller than the extrapolated line and $V_{T,extrapol}$ is smaller than $V_{T,2\phi F.}$ Hence, it is unlikely that the saturation I_{ds} of nanoscale MOS transistor will follow a $(V_{GS} - V_{th,sat})^{3/2}$ relationship. On the other hand, eqn. (1.2), which is based on the concepts of velocity saturation, predicts that I_{ds} will follow a linear relationship with V_{GS}. In the other words, $V_{th,sat}$ can be found by extrapolating the I_{ds} versus V_{GS} characteristics. From Fig. 3-9(b), the measured I_{ds} versus V_{GS} characteristics is quite close to the extrapolated line and $V_{T,extrapol}$ is bigger than $V_{T,\ 2\ \phi F.}$ Hence, it is more likely that the saturation I_{ds} of nanoscale MOS transistor will follow a $(V_{GS} - V_{th,sat})$ relationship.

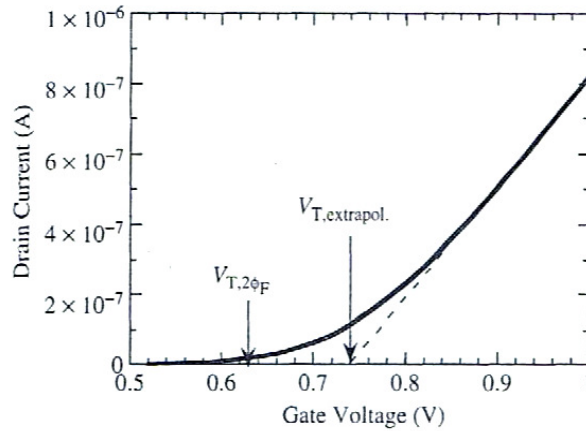


Figure 3-8 Illustration of $V_{T,\,2\phi F}$ and $V_{T,extrapol}$ in the I_{ds} versus V_{GS} characteristics[3.14].

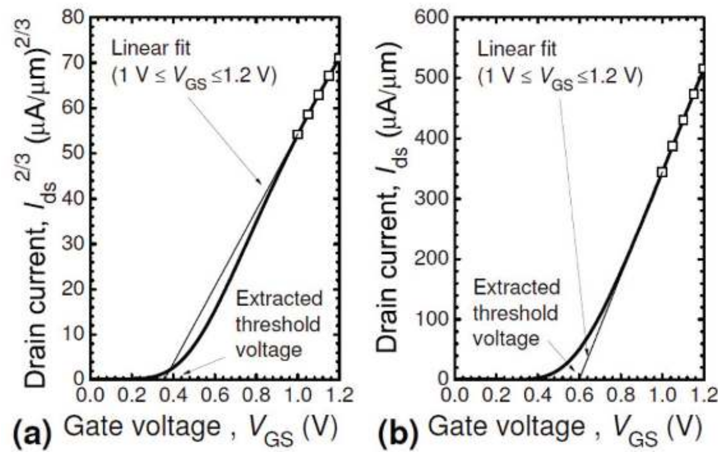


Figure 3-9 As opposed to Natori's 1994 theory on ballistic transport, the saturation I_{ds} of the nanoscale NMOS transistor (L = 60 nm) does not follow a $(V_{GS} - V_{th,sat})^{3/2}$ relationship. The symbols indicate the data points that are used for the linear extrapolation. The thicker lines indicate the experimental values of I_{ds}.

In Natori's 2008 paper [3.1], he acknowledged that his equation for the saturation drain current of nanoscale MOS transistor in ballistic transport is independent of the gate length. From Fig. 3-10, the simulation results show that the drain current of a double-gate NMOS transistor in ballistic transport is found independent of the channel length, except for the shortest gate length (10 nm) for which the short channel effects (SCEs) are exacerbated, leading to the strong reduction of the threshold voltage [3.15]. This is contradictory to the experimental data, which clearly show that the drain-induced barrier lowering (DIBL) and SCEs will increase in the saturation I_{ds} of the nanoscale MOS transistor.

Furthermore, Natori also pointed out that his theory is derived by Landeur's formula whereby the source and drain are assumed to be ideal reservoirs that inject sufficient carriers into the channel [3.1]. In the other words, Natori's theory does not consider the effects of the S/D series resistance (R_{sd}) on the carrier transport. With reference to Fig. 3-11, the potential along the channel direction varies linearly in the source and drain regions when R_{sd} is considered but the potential profile in the channel is the same with or without R_{sd} [3.15]. From Fig. 3-12, when R_{sd} is considered, the simulated I_{ds} of a NMOS transistor in the ballistic transport will become comparable to a NMOS transistor in the drift-diffusion transport [3.15].

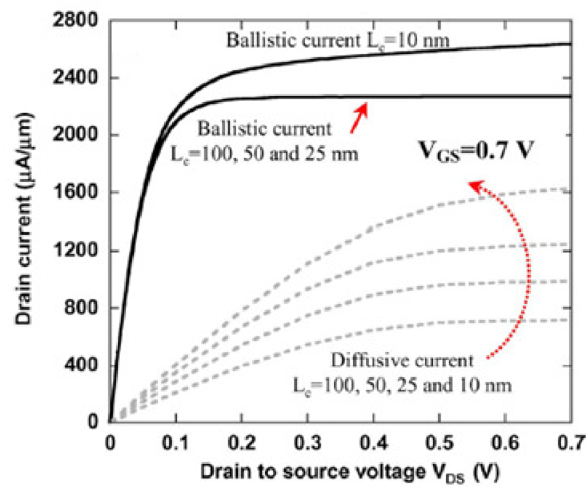


Figure 3-10 Simulation results of the drain current as a function of V_{DS} for a double-gate NMOS transistor with several channel lengths [3.15]. Solid line for ballistic transport and dashed line for classical diffusive transport.

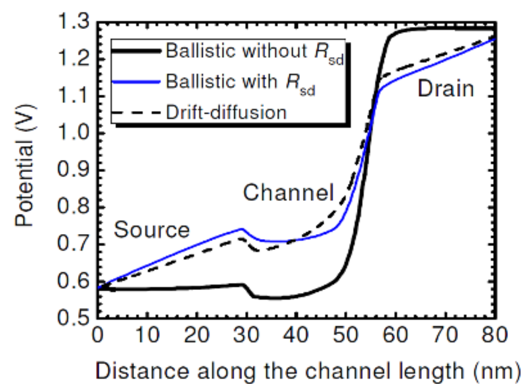


Figure 3-11 Impact of the S/D series resistance (R_{sd}) on the potential profile along the channel of NMOS transistor [3.15].

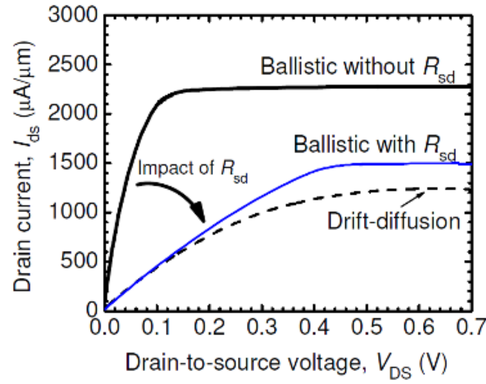


Figure 3-12 Impact of the S/D series resistance (R_{sd}) on the I_{ds} versus V_{DS} characteristics of a NMOS transistor [3.15].

Using the concepts of ballistic transport, Saad et al. proposed that the equation of V_{Dsat} can be expressed as follows [3.16],

$$V_{Dsat} = \frac{1}{(2\alpha - 1)}\left[(s-\alpha)V_c - (1-\alpha)(V_{GS} - V_{th,sat})\right] \tag{3.5}$$

with $$s = \sqrt{\left[\alpha + (1-\alpha)\frac{V_{GS} - V_{th,sat}}{V_c}\right] + 2\alpha(2\alpha - 1)\frac{V_{GS} - V_{th,sat}}{V_c}} \tag{3.6}$$

and $$V_c = \frac{v_{sat}}{\mu_{\ell f}} L_{eff} \tag{3.7}$$

and $$\mu_{\ell f} = 0.0700\exp(-V_{GS}/1.33) \qquad \text{(units: m}^2\text{/V.s)} \tag{3.8}$$

where α is less than 1. According to Saad et al. [3.16], α is around 0.7 when V_{DS} is 0.9 V. Despite the complexity of their theory, the important message is that V_{Dsat} of long-channel MOS transistor is bigger than V_{Dsat} of nanoscale MOS transistor [3.17].

Another fundamental problem of ballistic transport is the definition of ballistic transport based on the relative dimension between L and λ. From Fig. 3-13, the experimental results show a dramatic decrease of λ (down to 2.6 nm) and a reduction in μ_{eff} when gate length is reduced [3.18]. This means that ballistic transport is unlikely to occur in the state-of-the-art MOS transistors with $L \geq 32$ nm because $L > \lambda$. This strong reduction in mobility is typically observed in MOS transistors when the gate length is further reduced [3.19-3.23]. The exact mechanism of the mobility degradation is still not clearly understood. It is first attributed to the presence of halo implants because the contribution of halo implants to the channel doping concentration increases with decreasing gate length

[3.19]. However, the mobility degradation also occurs in MOS transistors without halo implants, as shown in Fig. 3-14 [3.20]. Furthermore, this mobility degradation is also observed in the undoped double gate MOS transistors [3.20] and the undoped fully-depleted silicon-on-insulator (FD-SOI) MOS transistors [3.23]. This indicates that the halo implant is not the dominant factor involved in the mobility degradation at shorter gate length.

Another limiting transport mechanism expected to be non-negligible in the short-channel MOS transistor is the presence of neutral defects induced by S/D extension implants [3.20]. From Fig. 3-15, the increase in the activation anneal temperature from 1050 °C to 1080 °C improves the low-field mobility of NMOS transistors at short gate length without any effect on the threshold voltage (V_{th}), evidencing neutral defects recovery. Another explanation of the mobility degradation at smaller gate length is that the increase in the long-range Coulomb scattering interactions between the high-density electron gases in the S/D regions and the channel electrons for very short channel MOS transistors [3.21, 3.22]. As shown in Fig. 3-16, the short-range Coulomb scattering, which occurs between the electron-electron interactions as well as the electron–ion interactions, will cease to exist when the distance between the electron and ion (r) becomes bigger than the screening length [3.22]. On the other hand, the plasma fluctuations present in the high electron density regions can penetrate into the channel of the MOS transistor through distances in the order of tens of nanometers [3.21, 3.22]. Fig. 3-17 shows that the long-range Coulomb scattering thermalizes the electron distribution in the S/D regions from the ambient temperature of 300 K to elevated S/D effective temperature of 800 K [3.22]. Monte Carlo simulation of I_{ds} versus V_{DS} characteristics shows that the long-range Coulomb scattering is essential to properly model the I_{ds} versus V_{DS} characteristics of a nanoscale MOS transistor, as shown in Fig. 3-18.

According to Heiblum & Eastman [3.24], the electrons will experience ballistic transport if there is no obstacle to scatter the electrons. In the other words, the electrons will move under the influence of the electric field in accordance to Newton's 2nd law of motion when it is in a vacuum environment or there is no obstacle. In Year 1928, Bloch postulated that the wave-particle duality of electron allows it to move without scattering in the densely packed atoms of a crystalline solid if (i) the crystal lattice is perfect, and (ii) there is no lattice vibration [3.25]. However, doping impurities such as boron, arsenic and phosphorus are added to the silicon crystal so as to tune the electrical parameters such as the threshold voltage and I_{off}. These dopants will disrupt the periodic arrangement of the

crystal lattice, leading to collisions with the impurity ions and the crystalline defects. Moreover, the atoms in crystals are always in constant motion according to the Particle Theory of Matter. These thermal vibrations cause waves of compression and expansion to move through the crystal and thus scatter the electrons [3.25]. Therefore, achieving ballistic transport in MOS transistors is only an ideal situation [3.1].

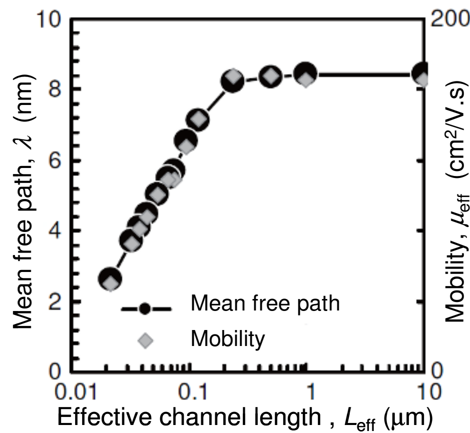


Figure 3-13 Mean free path (λ) as a function of L_{eff} for NMOS transistor [3.18]. Constant gate overdrive of 1.1 V is used. The drain bias is 10 mV.

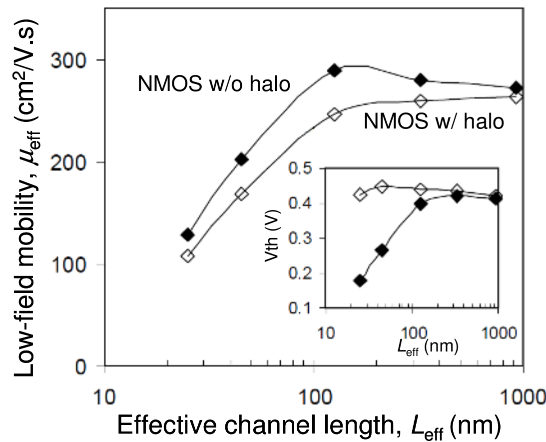


Figure 3-14 Mobility degradation with L_{eff} is observed in NMOS transistors without halo implants [3.20].

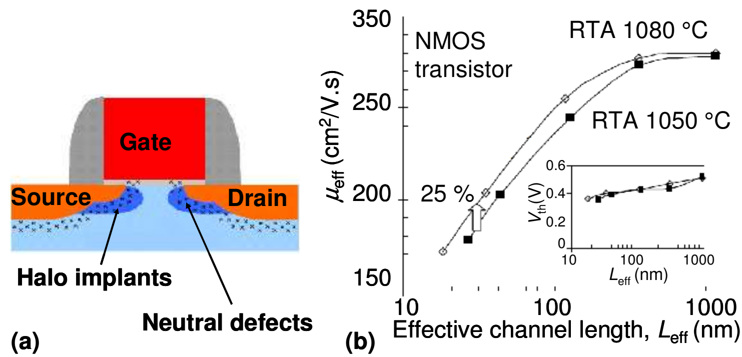


Figure 3-15(a) Schematics of neutral defects near S/D regions , (b) Neutral defects recovery by increasing the rapid thermal anneal (RTA) temperature [3.20].

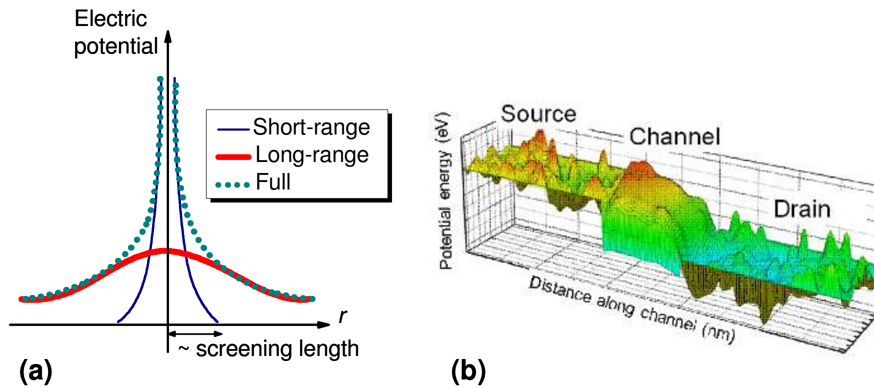


Figure 3-16 Role of long-range Coulomb scattering in nanoscale MOS transistor: (a) Region of influence, (b) Potential fluctuations in the S/D regions [3.22]. Note that *r* is the distance between the channel carrier and the scattering ion.

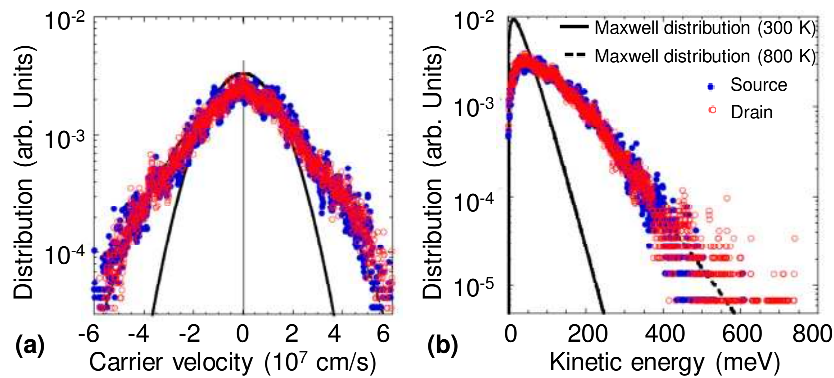


Figure 3-17 The effective electron temperature in the S/D regions is at 800 K rather than the ambient temperature of 300 K: (a) the carrier velocity distribution, (b) the kinetic energy distribution at location close to the source and drain contacts [3.22].

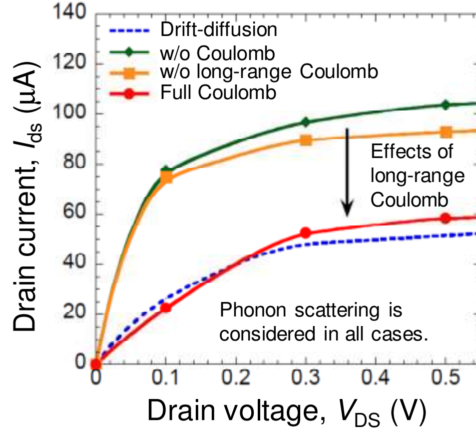


Figure 3-18 Effects of the long-range Coulomb scattering on the I_{ds} versus V_{DS} characteristics of a nanoscale NMOS transistor [3.22].

3.3 Quasi-ballistic transport

Based on Lundstrom's 1997 theory on quasi-ballistic transport, the saturation drain current (I_{ds}) of a nanoscale MOS transistor is related to the low-field mobility (μ_{eff}). In the other words, quasi-ballistic transport is able to account for the strain-induced I_{on} improvement in the nanoscale MOS transistor [1.11-1.13].

$$I_{ds} = \left[\frac{C_{ox} W}{\frac{1}{v_T} + \frac{1}{\mu_{eff}\, \varepsilon(0)^+}} \right] \left(V_{GS} - V_{th,sat} \right) \tag{3.9}$$

with the thermal velocity (v_T) is given by [1.24],

$$v_T = \sqrt{2 k_B T / (\pi m_t)} \tag{3.10}$$

where k_B is the Boltzmann's constant. T is the absolute temperature. Using eqn. (3.10), v_T is approximately equal to 1.2×10^7 cm/s at temperature of 25 °C.

According to Lundstrom [1.25], if a carrier backscatters beyond the critical length (ℓ), it is likely to exit from the drain and is unlikely to return back to the source, as shown in Fig. 3-19. For NMOS transistor, ℓ is defined as the distance between the top of the conduction band edge and the point along the channel where channel potential drops by $k_B T/q$. From Fig. 3-20, $\varepsilon(0^+)$, which is defined as the average electric field within the length ℓ ,increases with increasing V_{DS} [1.25]. When pinch-off occurs, V_{DS} is fixed at V_{Dsat} and thus $\varepsilon(0^+)$ becomes independent of V_{DS} (see Fig. 3-21) [3.26]. According to Lee et

al.[3.27], $\varepsilon(0^{+})$ of a nanoscale PMOS transistor with a nominal gate length of 50 nm is between 8×10^{4} V/cm and 3×10^{5} V/cm .

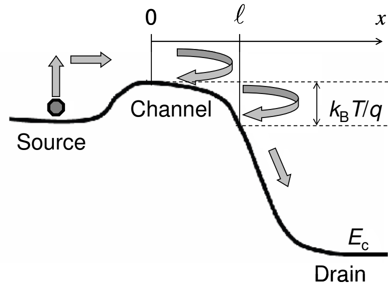


Figure 3-19 Definition of the critical length (ℓ) for NMOS transistor [1.25].

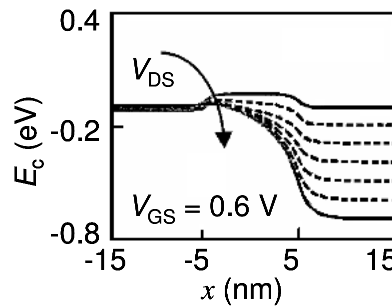


Figure 3-20 The average electric field, $\varepsilon(0^{+})$ is a function of V_{DS} [1.25].

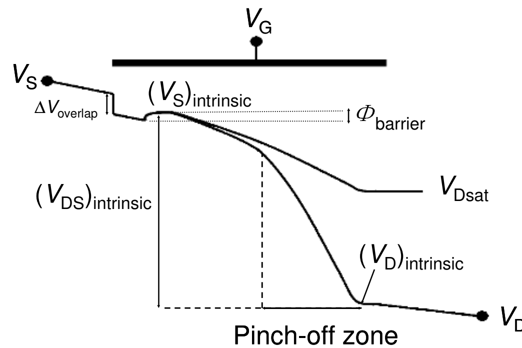


Figure 3-21 Schematics of the potential profile along the channel length of a NMOS transistor under quasi-ballistic transport [3.26]. $\Delta V_{overlap}$ is the drop in potential in the source series resistance and the accumulation region under the source-to-gate overlap. $(V_s)_{intrinsic}$ and $(V_D)_{intrinsic}$ are the intrinsic source bias and the intrinsic drain bias, respectively. $\Phi_{barrier}$ is the potential barrier at the source-channel.

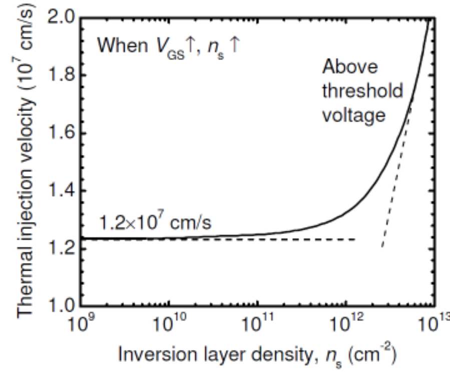


Figure 3-22 The thermal injection velocity ($\tilde{v}_T$) versus the inversion layer density (n_s) [1.25]. For V_{GS} below threshold voltage, $\tilde{v}_T \approx 1.2 \times 10^7$ cm/s. For V_{GS} above threshold voltage, the channel carriers become degenerate and $\tilde{v}_T$ increases.

However, there is actually a major flaw in eqn. (3.10). At very low temperature such as close to 0 K, v_T tends to zero and thus eqn. (3.9) predicts that the saturation I_{ds} will approach zero. However, there are numerous reports that MOS transistors and CMOS integrated circuits can function quite well at very low temperature such as liquid helium temperature [3.28-3.30]. In Year 2000, Lundstrom and Ren [1.25] made an attempt to incorporate Natori's 1994 theory into their theory,

$$I_{ds} = Q_i(V_{GS})\tilde{v}_T \left[\frac{\dfrac{1-\Im_{1/2}(\eta - U_{DS})}{\Im_{1/2}(\eta)}}{\dfrac{1+\ln\left(1+e^{\eta - U_{DS}}\right)}{\ln\left(1+e^{\eta}\right)}} \right] \tag{3.11}$$

where the inversion layer, $Q_i(V_{GS})$ is approximately equal to $2C_{ox}(V_{GS} - V_{th,sat})$ when V_{GS} is bigger than the threshold voltage. $\eta = (E_F - E_c)/ (k_B T)$ where E_F is the Fermi level, and E_c is the conduction band edge. U_{DS} is V_{DS} normalized to $k_B T/q$. The Fermi-Dirac integral, $\Im_{1/2}(u) = \int_0^\infty \frac{\sqrt{y}}{1+\exp(y-u)} dy$ [1.21]. The equation for the thermal injection velocity ($\tilde{v}_T$) is given by [1.25],

$$\tilde{v}_T = \sqrt{\frac{2k_B T}{\pi m_t} \left[\frac{\Im_{1/2}(\eta)}{\ln\left(1+e^{\eta}\right)} \right]} \tag{3.12}$$

As opposed to v_T in eqn. (3.10), $\tilde{v}_T$ in eqn. (3.12) is a function of T and V_{GS}. From Fig. 3-22, $\tilde{v}_T$ actually increases with increasing V_{GS} [1.25].

3.4 Semi-empirical model

Khakifirooz et al. proposed a semi-empirical model for the saturation drain current of the nanoscale MOS transistor [1.28]. This model is based on the "virtual source", which is located at the top of the conduction band profile for NMOS transistor, as shown in Fig. 3-23. Based on the "charge-sheet" approximation, the saturation I_{ds} of the nanoscale MOS transistor can be described by the product of the local charge density (Q_{ixo}) and the carrier velocity (v_{xo}) [1.27,1.28].

$$I_{ds} = W Q_{ixo} v_{xo} \tag{3.13}$$

where Q_{ixo} is given by [1.28],

$$Q_{ixo} = C_{ox} \frac{m k_B T}{q} \ln\left[1 + \exp\left(\frac{V_{GS} - I_{ds} R_s - V_{th,sat}}{m k_B T / q}\right)\right] \tag{3.14}$$

where R_s is the source series resistance, and m is the body-effect coefficient.
v_{xo} is given by [1.27, 1.28] is,

$$v_{xo} = \frac{v}{1 - C_{ox} R_s W (1 + 2\delta) v} \tag{3.15}$$

where δ is the drain-induced-barrier lowering (DIBL) with units of V/V. The carrier velocity can be extracted as follows [1.27],

$$v = \frac{I_{ds} / W}{C_{ox} (V_{GS} - V_{th,sat})} \tag{3.16}$$

The semi-empirical model is able to account for the strain-induced I_{on} enhancement in the nanoscale MOS transistor. From Fig. 3-24, stress engineering can increase v_{xo} of the nanoscale MOS transistors to beyond v_{sat} [1.27]. In addition, semi-empirical method considers the effects of R_s on the saturation I_{ds} of the nanoscale MOS transistors. Furthermore, the semi-empirical model predicts that the saturation I_{ds} of the nanoscale MOS transistors is dependent on DIBL and SCE (see Fig. 3-25).

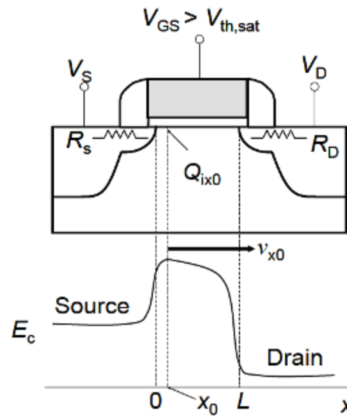


Figure 3-23 Illustration of the virtual source point (x_o) in the NMOS transistor. Q_{ixo} and v_{xo} are defined at the top of the conduction band profile along the channel direction. [1.28].

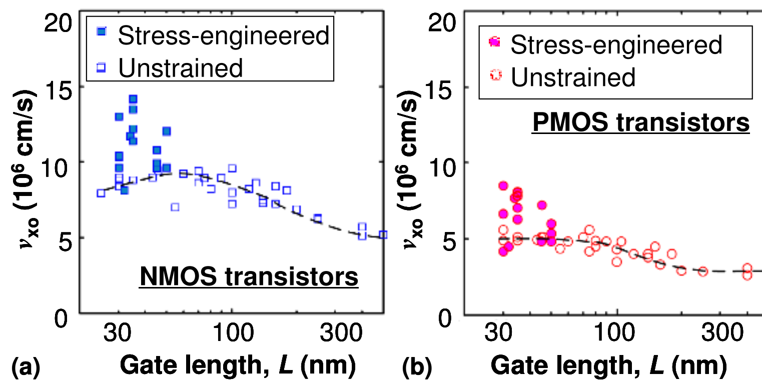


Figure 3-24 Extracted virtual source velocity (v_{xo}) as a function of the gate length for (a) NMOS transistors, (b) PMOS transistors that are fabricated on (100) Si wafer with <110> channel orientation [1.27]. Stress engineering can increase v_{xo} to beyond v_{sat}.

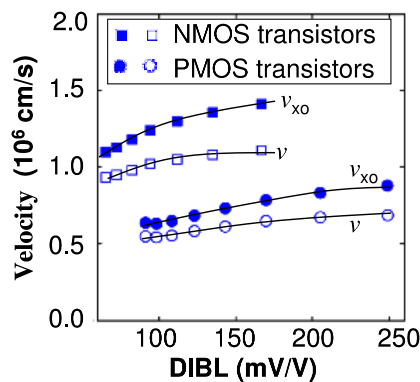


Figure 3-25 Extracted virtual source velocity (v_{xo}) and effective carrier velocity (v) versus DIBL for the nanoscale NMOS transistors and PMOS transistors [1.27].

3.5 Results and Discussion

3.5.1 Temperature dependency of the drain current

In Section 3.1.2, we have pointed out that Natori's 1994 theory on ballistic transport [1.20] predicts that the saturation drain current of nanoscale MOS transistor will increase when temperature increases. However, experimental results have shown that the I_{on} of nanoscale MOS transistor actually decreases with increasing temperature. In this chapter, we will look into the V_{GS} range where the prediction of Natori's 1994 theory is valid. Another motivation of this study is to identify the gate length of MOS transistors and V_{GS} biasing condition in which I_{ds} is less sensitive to temperature.

From Fig. 3-26, we observed that the temperature independent point (TIP) is dependent on the gate length of the transistors. For 1 μm long PMOS transistor, TIP is about -0.6 V. For 60 nm long PMOS transistor, TIP is about -0.95 V. Since the power supply voltage (V_{DD}) of these PMOS transistors is -1.2 V, I_{on} of the short-channel PMOS transistor is less sensitive to a temperature increase compared to that of the long-channel PMOS transistor. In the other words, the temperature dependency of I_{ds} will obey Natori's 1994 theory only if the transistor is biased with V_{GS} below TIP. The practical implication of this work is that if we can operate the MOS transistors at V_{GS} below TIP, I_{ds} will not be degraded by joule heating in the integrated circuits [3.31].

To further investigate the validity of Natori 1994 theory on the temperature dependency of I_{ds}, we studied the gate length dependency of TIP. Fig. 3-27 shows that TIP versus L characteristics of MOS transistors is related to the saturation threshold voltage ($V_{th,sat}$) versus L characteristics. $V_{th,sat}$ is extracted using the constant current method with reference current, I_{ref} = 0.1 μA W/L. When the MOS transistors are in the reverse short channel effect (RSCE) regime (0.1 μm ≤ L ≤ 10 μm), both $V_{th,sat}$ and TIP increase with decreasing L. When the MOS transistors are in the short-channel effect (SCE) regime (L < 60 nm), both $V_{th,sat}$ and TIP decrease with decreasing L. It is interesting to note that both TIP and $V_{th,sat}$ reach their maximum point in the transition between RSCE and SCE regimes. In addition, we studied the effects of the drain bias (V_{DS}) on the TIP versus L characteristics of the NMOS transistors. From Fig. 3-28, both TIP and $V_{th,sat}$ of NMOS transistors in the SCE regime decrease when V_{DS} is increased from 0.6 V to 1.2 V. This shows that the roll-off of TIP at smaller gate length is indeed caused by SCE.

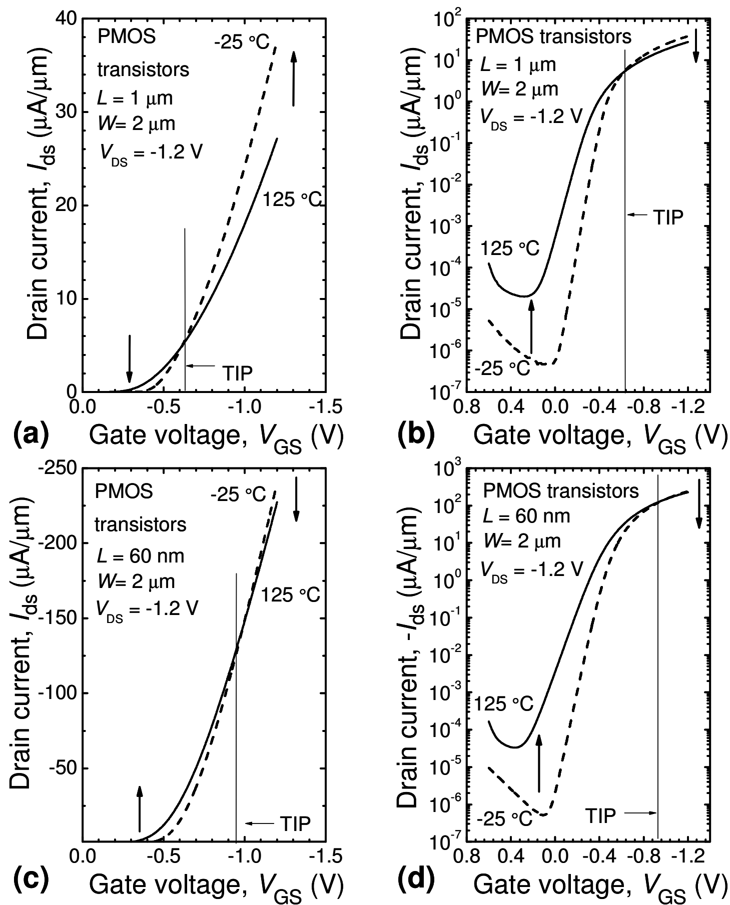


Figure 3-26 Temperature independent point (TIP) of the long-channel PMOS transistor with nominal gate length of 1 μm (a, b), and the short-channel PMOS transistor with nominal gate length of 60 nm (c,d). V_{DD} is -1.2

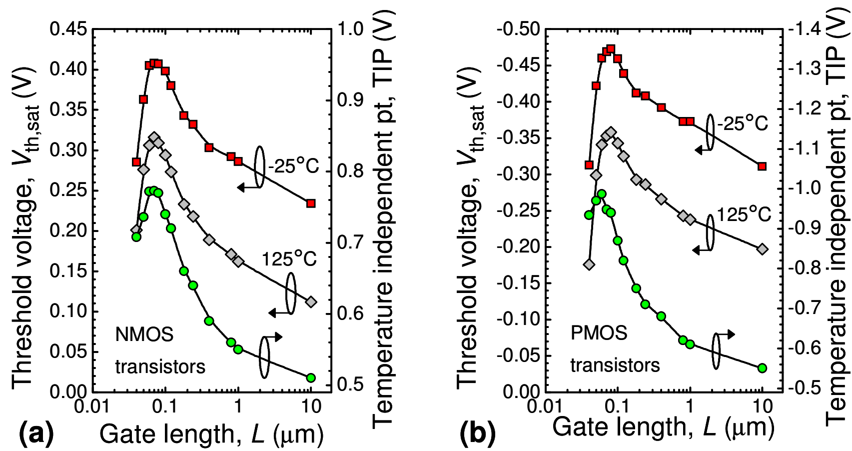


Figure 3-27 Gate length dependency of temperature independent point (TIP) for (a) NMOS transistor, (b) PMOS transistor. V_{DD} is 1.2 V.

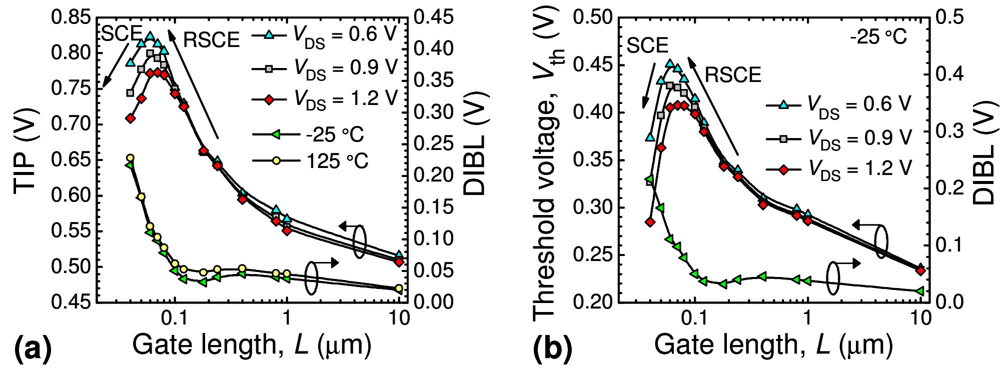


Figure 3-28 NMOS transistors (fabricated by 65 nm CMOS technology) in the SCE regime: (a) TIP decreases with increasing V_{DS}, (b) V_{th} decreases with increasing V_{DS}. Note that DIBL = $V_{th,lin}$ - $V_{th,sat}$.

Previous work has shown that it is more appropriate to use the effective drive current (I_{eff}) in the gate delay metric compared to I_{on} because I_{on} is never reached during inverter switching [3.32]. Using the four-point model for I_{eff} extraction [3.33], we extracted the percentage change in I_{eff} (ΔI_{eff}) when the temperature is increased from -25°C to 125°C. Moreover, we also measured the percentage change in I_{on} (ΔI_{on}) when the temperature is increased from -25°C to 125°C. From Fig. 3-29, we observed that ΔI_{eff} and ΔI_{on} are the smallest when the MOS transistors are in the transition between the SCE regime and RSCE regime. This corresponds to a maximum in the TIP versus L characteristics. Hence, I_{on} and I_{eff} of transistors in the transition regime are less sensitive to temperature variation.

In short, we have studied the gate length dependency of I_{on}, I_{eff}, $V_{th,sat}$ and TIP. We found that the gate length dependency of TIP in the nanoscale MOS transistors is related to the $V_{th,sat}$ versus L characteristics. Based on the simulation results by Hisamitsu et al. [3.34], an increase in the threshold adjustment implant dose will increase the magnitude of TIP, as shown in Fig. 3-30. Since the temperature sensitivity of I_{ds} will be smaller when TIP is closer to the operating V_{GS}, we can tune the dose of the threshold adjustment implant such that TIP of the MOS transistor will match the operating V_{GS} of the transistors. However, this approach will degrade I_{on} and I_{eff} because an increase in the dose of the threshold adjustment will increase in Coulomb scattering [3.11]. From Fig. 3-29, maximum TIP occurs at the transition regime between RSCE regime and SCE regime. Hence, a better approach is to tune the halo implant such that the nanoscale MOS transistor with the target gate length will lie in the transition between RSCE regime and SCE regime.

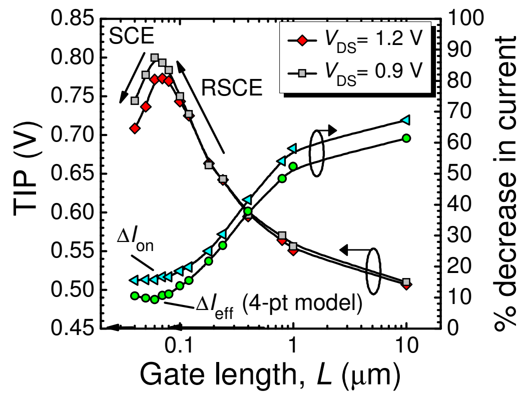


Figure 3-29 Temperature sensitivity in I_{on} and I_{eff} of NMOS transistors as a function of L.

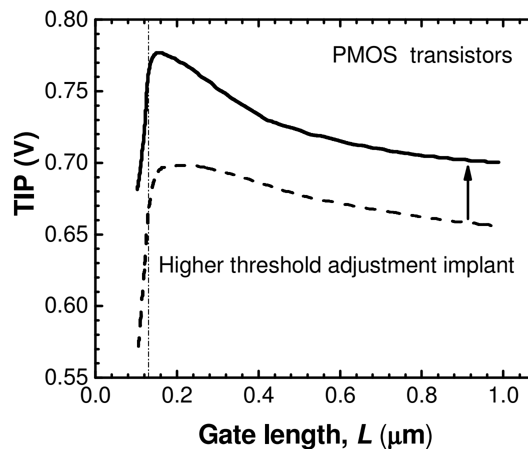


Figure 3-30 Simulation shows that threshold adjustment implant can increase TIP [3.34].

3.5.2 Saturation drain current equation

Table 3-1 summarizes the merits and limitations of various drain current transport theories for the nanoscale MOS transistors. The quasi-ballistic transport and the semi-empirical model are able to account for the strain-induced I_{on} improvement in the nanoscale MOS transistor [1.11-1.13]. Only the semi-empirical model can account for the effects of the parasitic S/D series resistance (R_{sd}) on the drain current of the nanoscale MOS transistor. From Fig. 3-12, the simulated I_{ds} of a nanoscale MOS transistor in ballistic transport will become comparable to that in drift-diffusion transport when R_{sd} is considered [3.15]. Hence, it is important to consider the effects of R_{sd} in the drain current transport of the nanoscale MOS transistor. On the other hand, the velocity saturation model and the semi-empirical model consist of parameters that can be easily obtained from the standard electrical measurements. However, ballistic transport and quasi-ballistic transport provide a better physical picture of the carrier transport as compared to the semi-empirical model.

Hence, there is a need to unify the merits of the above transport models and then come up with a simplified equation for the saturation drain current of nanoscale MOS transistor.

Table 3-1 Merits and limitations of various drain current transport theories for the nanoscale MOS transistors.

	Velocity saturation model[a]	Ballistic transport[b]	Quasi-ballistic transport[c]	Semi-empirical model[d]
I_{on} is related to μ_{eff}	No	No	Yes	Yes
Velocity overshoot	No	Yes	Yes	Yes
L dependency	No	No	Yes	Yes
Ease of parameter extraction	Yes	No	No	Yes
R_{sd} dependency	No	No	No	Yes
Better I_{on} performance at very low temp ($T \rightarrow 0$ K)	Yes	No	No	Yes

[a] after [1.6-1.10], [b] after [1.20-1.23], [c] after [1.24,1.25], [d] after [1.27,1.28].

By improvising Lundstrom's 1997 theory, we proposed that the equation of the saturation drain current of nanoscale MOS transistor can be expressed as follows,

$$I_{ds} = v_{eff}(\mu_{eff}, V_{GS}, T, R_{sd}) W C_{ox} (V_{GS} - V_{th,sat}) \tag{3.17a}$$

Furthermore, effective carrier velocity (v_{eff}) is also related to v_1 and v_2 terms.

$$v_{eff}(\mu_{eff}, V_{GS}, T, R_{sd}) = \left[\frac{1}{v_1(V_{GS}, T, R_{sd})} + \frac{1}{v_2(\mu_{eff}, V_{GS}, T)} \right]^{-1} \tag{3.17b}$$

$$v_1(V_{GS}, T, R_{sd}) = v_{inj}(V_{GS}, T, R_{sd}) \tag{3.17c}$$

$$v_2(V_{GS}, T) = \mu_{eff}(V_{GS}, T) \varepsilon(0)^{+} \tag{3.17d}$$

where v_{eff} can be experimentally extracted at high V_{DS} [3.35]. Table 3-2 shows the differences among our proposed theory, Natori's 1994 theory and Lundstrom's 1997 theory. Although Lundstrom's 1997 theory is able to account for the strain-induced I_{on} improvement, his theory is unable to explain the I_{on} improvement at very low temperature such as liquid helium temperature. In Lundstrom's 1997 theory, v_1 is equal to v_T, which is only dependent on temperature. From the perspective of gaining thermal energy from the ambient, when temperature is very low, the kinetic energy of the channel carriers are expected to be low, and thus leading to a smaller I_{on}. In Natori's 1994 theory, v_1 is equal to

v_{inj} term, which is a function of both V_{GS} and temperature. In this case, even when the temperature is very low, an increase in V_{GS} can lower the potential barrier at the source-channel region, leading to an increase in the carrier density. Since the drain current is a product of the carrier density and the carrier velocity, the introduction of v_{inj} into Lundstrom's 1997 theory may be able to account for the experimental observation of I_{on} enhancement at very low temperature. Hence, we proposed to replace v_1 of Lundstrom's 1997 theory by v_1 of Natori's 1994 theory.

Table 3-2 Comparison between Natori's 1994 theory, Lundstrom's 1997 theory and our proposed theory on drain current transport of the nanoscale MOS transistor.

	v_1	v_2	Remarks
Natori's theory	v_{inj} (V_{GS}, T)	------	Cannot account for • Strain-induced I_{on} improvement • I_{on} degradation at elevated temp.
Lundstrom's theory	$v_T = \sqrt{2k_B T/(\pi m_t)}$	$\mu_{eff}(V_{GS}, T)\varepsilon(0)^+$	Cannot account for • I_{on} improvement at liquid helium temp.
Our proposed theory	v_{inj} (V_{GS}, T, R_{sd})	$\mu_{eff}(V_{GS}, T)\varepsilon(0)^+$	Can account for • Strain-induced I_{on} improvement • I_{on} degradation at elevated temp. • I_{on} improvement at liquid helium temp.

In addition, we also modify the definition of v_{inj} to include the effects of R_{sd}. Our rationale can be understood as follows. From Fig. 3-12, the presence of R_{sd} will not affect the conduction band edge (E_c) profile in the n-channel [3.15]. However, the presence of R_{sd} will cause a potential drop in the S/D regions, resulting in a built-in electric field with the S/D regions, as shown in Fig. 3-31. This electric field in the source region will accelerate the electrons. Hence, the presence of R_{sd} allows the electrons to attain higher energy prior to thermionic emission from the source into the channel. According to Chen et al. [3.36], R_s is about 75 Ω-μm and I_{on} of the NMOS transistor (L = 68 nm, T_{ox} = 1.65 nm) is about 800 μA/μm. Hence, the voltage drop in the source side due to R_s is about 800 μA/μm × 75 Ω-μm = 60 mV. (Note that v_T is approximately 26 meV at room temperature). We proposed that the electrons are "heated" up by this 60 meV energy due to R_{sd} and thus their velocities can be significantly bigger than 1.2×10^7 cm/s. Moreover, this extra energy from

electron heating in the R_{sd} region is expected to increase with increasing V_{GS} because a higher V_{GS} implies a bigger I_{ds}. According to Nakanishi et al. [3.22], the long-range Coulomb scattering thermalizes the electron distribution in the S/D regions from the ambient temperature of 300 K to elevated S/D effective temperature of 800K, and thus the maximum carrier velocity achievable in the S/D region can increase to 6×10^7 cm/s, as shown in Fig. 3-17. By considering the electron heating in the R_{sd} regions, our simplified saturation drain current of nanoscale MOS transistor can overcome the weakness of Lundstrom's 1997 theory [1.24] and account for I_{on} improvement of MOS transistors and CMOS integrated circuits at very low temperature such as liquid helium temperature.

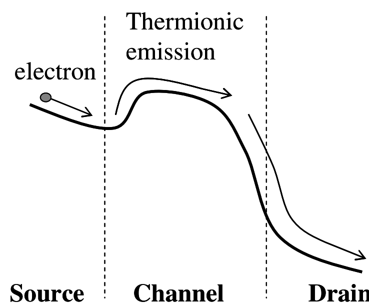


Figure 3-31 Effects of R_{sd} on E_c of a NMOS transistor in saturation operation.

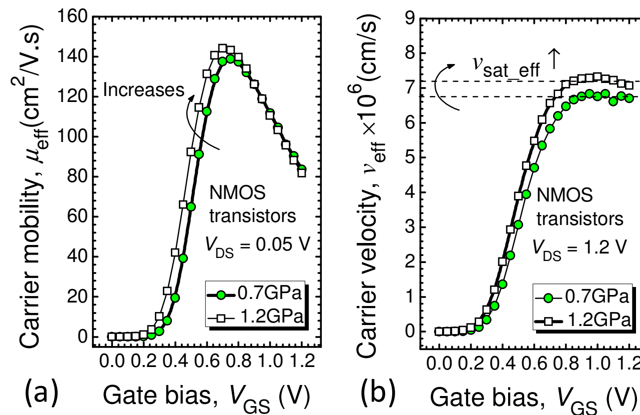


Figure 3-32 Effects of CESL-induced tensile stress on: (a) μ_{eff} versus V_{GS} characteristics, (b) v_{eff} versus V_{GS} characteristics of a NMOS transistor (L = 60 nm).

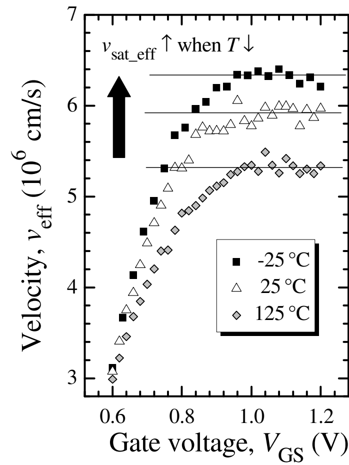


Figure 3-33 Effects of temperature on v_{sat_eff} and v_{eff} of a NMOS transistor with nominal gate length of 60 nm. Note that v_{sat_eff} refers to the average value of v_{eff} when V_{GS} is close to V_{DD}.

As shown in Fig. 3-32 and Fig. 3-33, v_{eff} is a function of V_{GS} and T at a fixed V_{DS}. The relationship between v_{eff} and μ_{eff} can be understood as follows. From eqn.(3.17d), a higher μ_{eff} will lead to a higher v_2 should give a higher effective velocity. An inspection of Figs. 3-32 shows that this is case at low V_{GS} (i.e. before the peak in μ_{eff}). At high V_{GS} (i.e. after the peak in μ_{eff}), Fig. 3-32(c) shows that μ_{eff} becomes insensitive to the strain, but Fig.3-32 (d) shows that v_{sat_eff} clearly is increased by strain even at high V_{GS}. The mechanism behind the strain-induced change in μ_{eff} depends on the V_{GS} range. For low V_{GS}, μ_{eff} is affected by phonon scattering, which is decreased by uniaxial tensile stress owing to the strain-induced suppression of the intervalley photon scattering. On the other hand, at high V_{GS}, μ_{eff} is affected by surface roughness scattering. Strain effects on surface roughness limited mobility is controlled by two competing factors. The first factor is related to the strain effects on the proximity of the carrier centroid to the roughness induced potential perturbations at the Si/SiO_2 interface. Since uniaxial tensile stress moves the centroid of the inversion charge closer to the Si/SiO_2 interface [3.37], the carrier/roughness interaction increases, and thus the surface roughness limited mobility decreases. The second factor is related to the morphological change in surface roughness by strain. Atomic force microscopy measurements on epitaxially 0.8% strained SOI devices have shown that tensile stress can indeed reduce the surface roughness [3.38], and thus leads to the electron mobility enhancement. Owing to the competing factors in the surface-limited mobility, it appears that mobility is not affected by strain. Unlike μ_{eff} that is extracted at low drain bias, v_{sat_eff} is extracted at high drain bias. For the same gate bias, the electric field across the Si/SiO_2 interface will be smaller at high drain bias compared to low

drain bias. Hence, the carrier centroid will be further away from the Si/SiO_2 interface at high drain bias. In the other words, the strain effects on the proximity of the carrier centroid to the surface roughness induced potential perturbations at the Si/SiO_2 interface is expected to be smaller for high drain bias compared to low drain bias. Considering the same amount of tensile stress, the morphological change in surface roughness by strain is the same for high drain bias and low drain bias. Therefore, at high V_{GS}, there is a strain-induced improvement at high V_{DS} (i.e. v_{sat_eff}) but not at low V_{GS} (i.e. μ_{eff}).

The significance of v_2 term is that it establishes a link between I_{on} of the nanoscale MOS transistor and μ_{eff}. With reference to Fig. 1-4, there is an experimental correlation between the low-field mobility and the high-field carrier velocity [1.29]. Furthermore, the v_2 term also provides a better compatibility between theory and the experimental results of the strain-induced I_{on} improvement in the nanoscale MOS transistor [1.11-1.13]. As explained in Section 3.1.2, one of the weaknesses of Natori's 1994 theory on ballistic transport [1.20] is that it cannot account for the experimental observation of the strain-induced I_{on} improvement. By incorporating the v_2 term from Lundstrom's 1997 theory [1.24] into Natori's 1994 theory [1.20], our simplified saturation drain current of nanoscale MOS transistor can overcome the weakness of Natori's 1994 theory [1.20] and account for the strain-induced I_{on} performance of MOS transistors [1.11-1.13].

Another weakness in Lundstrom's 1997 theory is that there is no equation for $\varepsilon(0^+)$. With reference to Fig. 3-20 and Fig. 3-21, we deduce that $\varepsilon(0^+)$ is a function of both V_{GS} and V_{DS} such that $\varepsilon(0^+, V_{DS} = V_{DD})$ is approximately equal to $\varepsilon(0^+, V_{DS} = V_{Dsat})$ where V_{DD} is the power supply voltage and V_{Dsat} is the saturation drain voltage. Therefore, we propose that $\varepsilon(0^+)$ can be expressed as follows,

$$\varepsilon(0)^+ = \frac{\alpha_1 V_{Dsat}}{L_{eff}} \tag{3.18a}$$

where the correction factor (α_1) is smaller than 1. Based on the conventional MOS transistor theory [3.9], V_{Dsat} is given by $(V_{GS} - V_{th,sat})/m$ where $1.1 \le m \le 1.4$. Furthermore, Suzuki and Usuki [3.7] also proposed that V_{Dsat} is smaller than $(V_{GS} - V_{th,sat})$ for the short-channel MOS transistors, as shown in eqn. (3.1). This shows that the relationship of $V_{Dsat} = (V_{GS} - V_{th,sat})/m$ is still reasonably correct for very short MOS transistors. Therefore, $\varepsilon(0^+)$ can also be expressed as,

$$\varepsilon(0)^+ = \frac{\alpha_2 (V_{GS} - V_{th,sat})}{L_{eff}} \tag{3.18b}$$

where correction factor (α_2) is smaller than 1. Using the L_{eff} extraction method proposed by Guo et al. [3.39], we found that the L_{eff} of our 60 nm NMOS transistor) is 0.030 μm. Using the constant current method with reference current, $I_{\text{ref}} = 0.1$ μA W/L, the extracted $V_{\text{th,sat}}$ is about 0.3 V. By substituting $L_{\text{eff}} = 3 \times 10^{-6}$ cm, $V_{\text{GS}} = 1.2$ V, $V_{\text{th,sat}} = 0.3$ V into eqn. (3.18b),

$$\varepsilon(0)^{+} = 3\times10^{5}\times\alpha_{2} \qquad \text{(units: V/cm)} \tag{3.18c}$$

Since the low-field mobility can be taken as the proportionality constant between the high-field carrier velocity and the lateral electric field, we can write $v_{\text{sat_eff}} \approx \mu_{\text{eff}} \times \varepsilon(0^{+})$.

$$\varepsilon(0)^{+} = \frac{v_{\text{sat_eff}}}{\mu_{\text{eff}}} \tag{3.18d}$$

The value of α_2 can be estimated from the v_{eff} versus V_{GS} characteristics and the μ_{eff} versus V_{GS} characteristics. From Fig. 3-32, for the CESL with a tensile stress of 1.2 GPa, $v_{\text{sat_eff}}$ of the NMOS transistor with mask gate length of 60 nm is 7.3×10^{6} cm/s. From Fig. 3-33(c), μ_{eff} is about 85 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ at $V_{\text{GS}} = 1.2$ V. Substituting $v_{\text{sat_eff}} = 7.3\times10^{6}$ cm/s and $\mu_{\text{eff}} = 85$ $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ into eqn. (3.18d),

$$\varepsilon(0)^{+} = \frac{v_{\text{sat_eff}}}{\mu_{\text{eff}}} = \frac{7.3\times10^{6}}{85} = 8.588\times10^{4} \text{ V/cm} \tag{3.18e}$$

According to Lee et al.[3.27], the simulated $\varepsilon(0^{+})$ of a PMOS transistor with a nominal gate length of 50 nm is between 8×10^{4} V/cm and 3×10^{5} V/cm for gate overdrives between 0.4 V and 0.8 V. This shows that our estimation of $\varepsilon(0^{+})$ based on experimental data is reasonably correct. By solving eqn. (3.18c) and eqn. (3.18e), α_2 is equal to 0.29.

When temperature decreases, v_{inj} decreases according to Natori's 1994 theory on ballistic transport [1.20] (see Fig. 3-7). Since v_1 is related to v_{inj} (see eqn. 3.17c), v_1 is expected to decrease with decreasing temperature. From eqn. (3.17d), v_2 is related to μ_{eff}. From Fig. 3-34, based on the general understanding of the inversion layer mobility, the μ_{eff} versus V_{GS} characteristics can be categorized into three dominant scattering mechanisms: (i) Coulomb-limited mobility, (ii) Phonon-limited mobility, and (iii) Surface-roughness limited mobility. Here, we will like to point out that the Coulomb-limited mobility has a different temperature dependency as compared to phonon-limited mobility [3.12] and surface-roughness limited mobility [3.13]. According to Chen et al. [3.40], when temperature is decreased, the thermal velocity of carriers are reduced and thus the

interaction time between channel carriers and the ionized impurity charges becomes bigger, leading to a higher Coulomb scattering probability and lower Coulomb mobility. Since Coulomb-limited mobility exhibits very weak dependency on the process-induced stress, it is recommended that we avoid Coulomb mobility domination in the carrier transport by using a lower channel doping concentration [3.40]. From Fig. 3-32, μ_{eff} of our MOS transistor at V_{GS} that is close to V_{DD} of 1.2 V will be dominated by surface roughness scattering rather than Coulomb scattering. Since v_2 is related to μ_{eff} , we expect v_2 to increase when temperature decreases. From Fig. 3-33, we observed that the experimental v_{eff} increases when temperature decreases, and thus we believe that v_2 dominates over v_1.

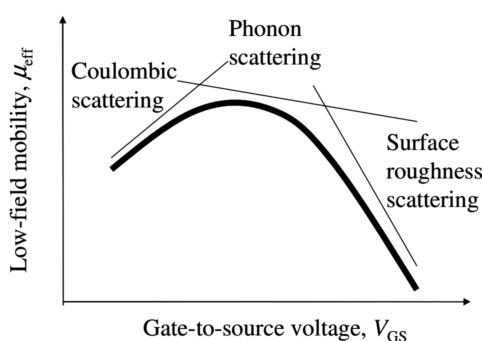


Figure 3-34 Effects of the scattering mechanism on the μ_{eff} versus V_{GS} characteristics.

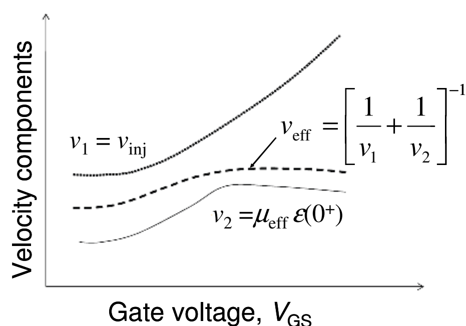


Figure 3-35 Schematic diagram showing the relationship of v_1, v_2 and v_{eff} with V_{GS}

Another evidence to show the relative importance of v_1 and v_2 is through their V_{GS} dependency. Fig. 3-35 shows the schematics of the behaviour of v_1, v_2 and v_{eff} with V_{GS}. As discussed in Section 3.1.2, an increase in V_{GS} will lower the potential barrier at the source-to-channel region and thus the carrier density is increased. According to Pauli principle, the carriers are expected to populate higher energy levels and the increase in kinetic energy leads to an increase in the mean velocity [3.1]. Hence, v_{inj} is expected to increase with an increase in V_{GS}. Since v_1 is related to v_{inj}, v_1 is expected to increase with increasing V_{GS} as shown in Fig. 3-6 [1.21]. At V_{GS} close to V_{DD}, μ_{eff} is dominated by

surface-roughness scattering and thus μ_{eff} is expected to decrease when V_{GS} increases. From eqn. (3.17d), $v_2 = \mu_{eff}\varepsilon(0^+)$. From eqn.(3.18b), $\varepsilon(0^+)$ increases with increasing V_{GS}. Owing to the opposing effects of μ_{eff} and $\varepsilon(0^+)$ with increasing V_{GS}, v_2 is expected to approach a constant when V_{GS} increases. This behavior of v_2 resembles that of v_{eff} and thus we believe that v_2 dominates over v_1.

Since v_{eff} approaches a constant when V_{GS} is close to V_{DD} (see Fig. 3-32 and Fig. 3-33), it is more appropriate to replace v_{eff} in eqn. (3.17a) by v_{sat_eff}, which is the average value of v_{eff} when V_{GS} is close to V_{DD}.

$$I_{ds} = v_{sat_eff}(\mu_{eff}, T)WC_{ox,inv}(V_{GS} - V_{th,sat_IV}) \quad (3.19)$$

From Fig. 3-32, v_{sat_eff} of NMOS transistor increases when tensile stress increases. This shows that eqn. (3.19) is able to account for the strain-induced I_{on} improvement in the short-channel MOS transistors by stress engineering [1.11-1.13]. As shown in Fig. 3-33, v_{sat_eff} increases when temperature decreases, resulting in a better I_{on} performance at lower temperature. This shows that eqn. (3.19) is able to account for the I_{on} improvement in MOS transistor at very low temperature such as the liquid helium temperature [3.28-3.30]. Moreover, $V_{th,sat}$ in eqn. (3.17a) needs to be replaced by V_{th,sat_IV}, which is a fitting parameter in the saturation I_{ds} versus V_{GS} characteristics. V_{th,sat_IV} is extracted using the linear extrapolation of the saturation I_{ds} versus V_{GS} characteristics when v_{eff} approaches a constant in the v_{eff} versus V_{GS} characteristics. Since v_{eff} of our NMOS transistor approaches a constant when 1 V ≤ V_{GS} ≤ 1.2 V (see Fig. 3-32), we performed a best-fit line of the saturation I_{ds} versus V_{GS} characteristics when V_{GS} is close to 1.2 V. On the other hand, both the linear threshold voltage ($V_{th,lin}$) and the saturation threshold voltage ($V_{th,sat}$) are extracted using the Constant Current method with reference current (I_{ref}) defined as 0.1 μA*W*/*L*. Typically, the saturation threshold voltage of nanoscale MOS transistor will be smaller than the linear threshold voltage owing to the short-channel effects. For our nanoscale NMOS transistor, the extracted $V_{th,sat}$ is 0.351 V and $V_{th,lin}$ is 0.484 V. From Fig. 3-36, V_{th,sat_IV} of our nanoscale NMOS transistor is 0.603 V. Our explanation is that V_{th,sat_IV} accounts for additional V_{GS} that is required to produce electrons to screen the Coulombic scattering centre, and thus V_{th,sat_IV} is bigger than the linear threshold voltage. Furthermore, polysilicon depletion and quantum mechanical effects will make the gate oxide appears thicker, and thus C_{ox} in eqn. (3.17a) has to be replaced by $C_{ox,inv}$, which is the gate oxide capacitance at inversion per unit area.

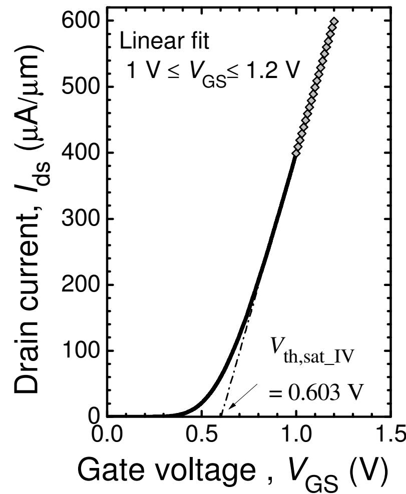


Figure 3-36 Extraction of V_{th,sat_IV} from the saturation I_{ds} versus V_{GS} characteristics.

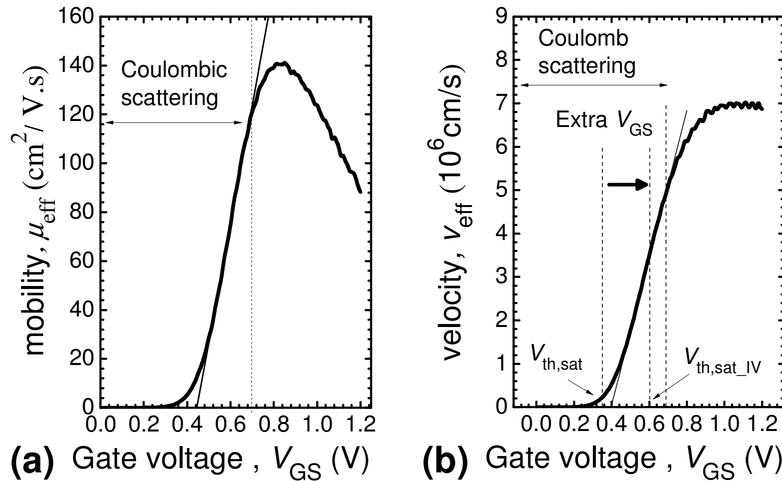


Figure 3-37 V_{th,sat_IV} includes a component to overcome the Coulombic scattering by "screening": (a) μ_{eff} versus V_{GS} characteristics, (b) v_{eff} versus V_{GS} characteristics of a NMOS transistor with a nominal gate length of 60 nm.

At this point, it may seem that eqn. (3.19) is quite similar to the equation for velocity saturation, as shown in eqn. (1.2). Here, we would like to highlight the differences between the effective saturation velocity (v_{sat_eff}) in eqn. (3.19) and v_{sat} in eqn. (1.2). First of all, v_{sat} is the equilibrium value of carrier velocity in silicon, whereas v_{sat_eff} is taken to be the average v_{eff} when V_{GS} is close to V_{DD}. From eqn. (3.17b) and eqn. (3.17c), v_{eff} is a function of v_{inj}. According to the Monte Carlo simulation by Natori [1.21], v_{inj} ranges between 1.2×10^7 cm/s to 1.6×10^7 cm/s. Hence, v_{sat_eff} can reach values that are bigger than v_{sat} for

electrons in silicon [1.9]. v_{eff} can be experimentally extracted at high V_{DS} (e.g. $V_{\mathrm{DS}} = 1.2$ V) as follows [3.35],

$$v_{\mathrm{eff}} = \frac{g_{\mathrm{m}}}{WC_{\mathrm{ox,inv}}} \tag{3.20a}$$

where g_{m} is the measured transconductance. To account for R_{sd}, we should use the intrinsic transconductance (g_{mi}) instead of g_{m} [3.41].

$$v_{\mathrm{eff}} = \frac{g_{\mathrm{mi}}}{WC_{\mathrm{ox,inv}}} \tag{3.20b}$$

where $g_{\mathrm{mi}} = \dfrac{g_{\mathrm{m}}^{0}}{1 - R_{\mathrm{sd}} g_{\mathrm{d}} \left(1 + R_{\mathrm{s}} g_{\mathrm{m}}^{0}\right)}$, $g_{\mathrm{m}}^{0} = \dfrac{g_{\mathrm{m}}}{1 - R_{\mathrm{s}} g_{\mathrm{m}}}$, $R_{\mathrm{s}} = 0.5\, R_{\mathrm{sd}}$,

g_{d} is the measured drain conductance. R_{sd} is extracted using a method proposed by Chern et al. [3.42]. From Fig. 3-38, $v_{\mathrm{sat_eff}}$ of a NMOS transistor with nominal gate length of 32 nm is bigger than 10^7 cm/s at temperature of 300 K even before R_{sd} correction. This shows that the accuracy of R_{sd} extraction will not affect our conclusion.

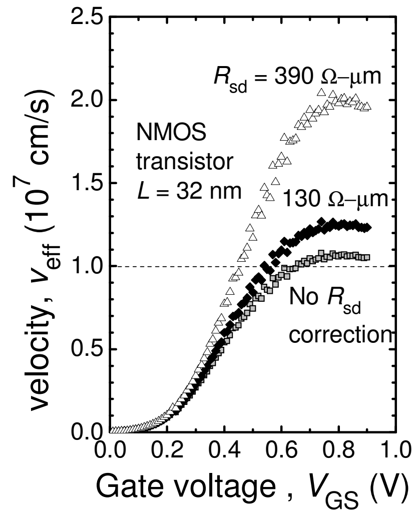


Figure 3-38 Effects of R_{sd} correction on $v_{\mathrm{sat_eff}}$ of a nanoscale NMOS transistor at room temperature.

3.5.3 Physics behind the apparent velocity saturation

We observed that the aggressive transistor scaling brings about another phenomenon that can be easily confused with velocity saturation. As explained in Section 1.2, velocity saturation occurs when there is no net gain in energy of the channel carriers [1.7]. From Fig. 3-39, a strong reduction in mobility is observed when L decreases. From Section 3.1.2, this strong reduction in mobility observed during transistor scaling has been attributed to various mechanisms: (i) the increase in halo implant contribution to the channel doping concentration [3.21], (ii) the presence of neutral defects induced by S/D extension implants [3.22], and (iii) the increase in the long-range Coulomb scattering interactions between the high-density electron gases in the S/D regions and the channel electrons for very short channel MOS transistors [3.23, 3.24]. Since v_{sat_eff} is the average value of v_{eff} when V_{GS} is close to V_{DD}, we can understand the physics behind the apparent velocity saturation by studying the v_{eff} versus V_{GS} characteristics. From eqn. (3.17b), v_{eff} is a function of both v_1 and v_2. From the temperature dependency and the V_{GS} dependency, we deduced that v_2 dominates over v_1 such that $v_{eff} \approx v_2 = \mu_{eff}(V_{GS},T)\varepsilon(0)^+$ (see Fig. 3-31, Fig. 3-34, Fig. 3-35). From eqn. (3.18b), $\varepsilon(0)^+$ is expected to increase with decreasing L. Hence, v_{sat_eff} will tend to saturate when L decreases owing to the opposing effects of μ_{eff} and $\varepsilon(0)^+$ with decreasing L. This explains the observation of the apparent velocity saturation during the gate length scaling of MOS transistors within the same CMOS technology node. If we consider the carrier velocity of individual MOS transistor, there is no velocity saturation because v_{sat_eff} of NMOS transistor can go beyond 10^7 cm/s.

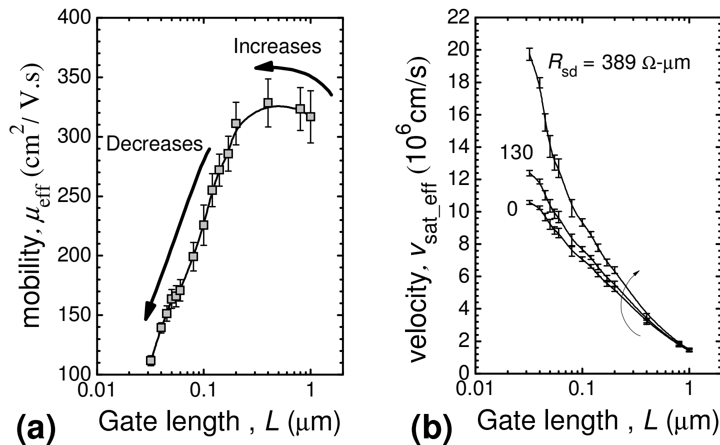


Figure 3-39 Effects of transistor scaling on (a) the μ_{eff} versus L characteristics and (b) the v_{sat_eff} versus L characteristics of NMOS transistors.

4. Effects of mechanical stress on the electrical parameters

4.1 On-current (I_{on})

The typical approach to assess the amount of strain-induced I_{on} improvement is to compare the log I_{off} versus I_{on} plot of the stress-engineered MOS transistors with that of the unstressed MOS transistors. Subsequently, the amount of strain-induced I_{on} improvement can be found by considering the amount of horizontal shift in the log I_{off} versus I_{on} plot for a fixed I_{off}, as shown in Fig. 4-1 [1.42]. The assumptions are: (i) I_{off} is not affected by stress engineering, and (ii) the linearity of the log I_{off} versus I_{on} plot. However, several researchers have reported on strain-induced I_{off} increase [2.5, 2.20, 2.22, 2.34]. From Fig. 4-2, the linearity assumption of the log I_{off} versus I_{on} plot is only valid for a specified range of I_{on} and I_{off}, which is different for each CMOS process fabrication flow [1.39]. However, there is no published report that discusses the physics behind the linearity of the log I_{off} versus I_{on} plot.

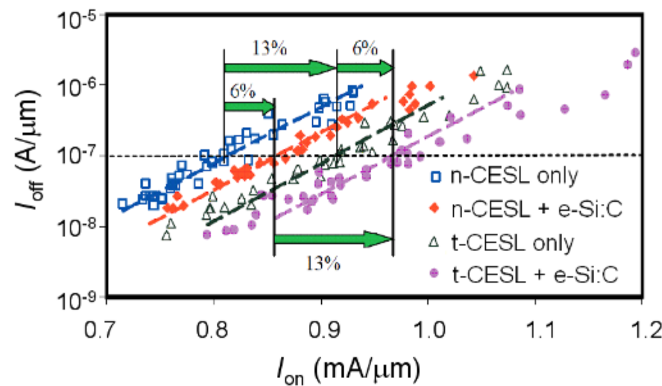


Figure 4-1 log I_{off} versus I_{on} plot to assess I_{on} improvement by stress engineering [1.42].

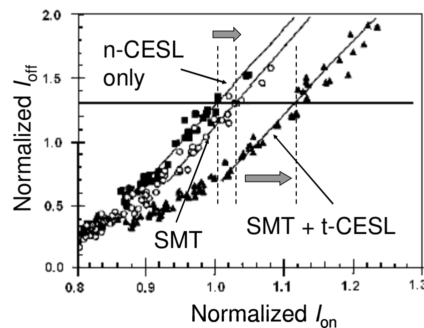


Figure 4-2 Validity of the linearity assumption in the log I_{off} versus I_{on} plot for NMOS transistor [1.39].

4.2 Threshold voltage

Fig. 4-3 shows that the biaxial tensile stress leads to negative threshold voltage shifts for both NMOS transistors and PMOS transistors [4.1]. The threshold voltage shift, $\Delta V_{th}(\sigma) = V_{th}(\sigma) - V_{th}(0)$ where $V_{th}(\sigma)$ is the threshold voltage of the stress engineered MOS transistor while $V_{th}(0)$ is the threshold voltage of the unstrained MOS transistor. Since the threshold voltage of enhancement-type NMOS transistor is positive, a negative $\Delta V_{th}(\sigma)$ means that the biaxial tensile stress reduces the magnitude of the threshold voltage of the NMOS transistors. Since the threshold voltage of enhancement-type PMOS transistor is negative, a negative $\Delta V_{th}(\sigma)$ means that the biaxial tensile stress increases the magnitude of the threshold voltage of the PMOS transistor. From Fig. 4-4, uniaxial tensile stress and biaxial tensile stress reduce the magnitude of the threshold voltage of NMOS transistors. According to Zhang et al. [4.1] and Lim et al. [4.2], the strain-induced effects on the threshold voltage (V_{th}) is mainly caused by the strain-induced change in the silicon bandgap. Hence, it is important to clarify the effects of mechanical stress on silicon bandgap before we discuss the equation for the strain-induced change in threshold voltage.

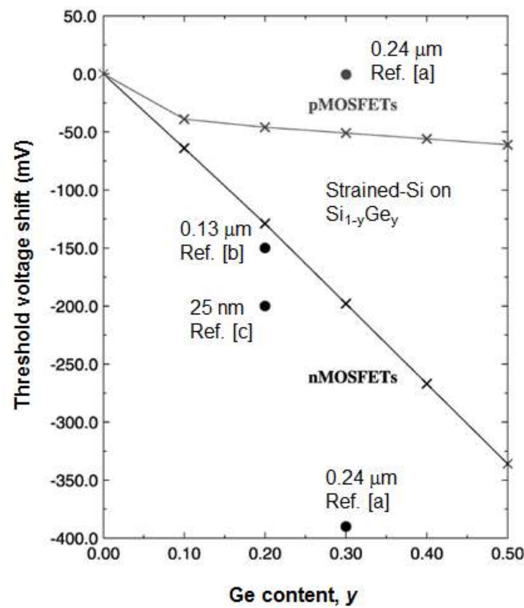


Figure 4-3 Biaxial tensile stress leads to negative threshold voltage shifts in<110> MOS transistors on (100) Si [4.1]. Ref.[a]: N. Sugii et al., TED, 49(12), pp. 2237-2243 (2002). Ref. [b]: K. Rim et al., TED, 47(7), pp. 1406-1414 (2000). Ref. [c]: Q. Xiang et al., VLSI Symp. 2002, pp. 101-102.

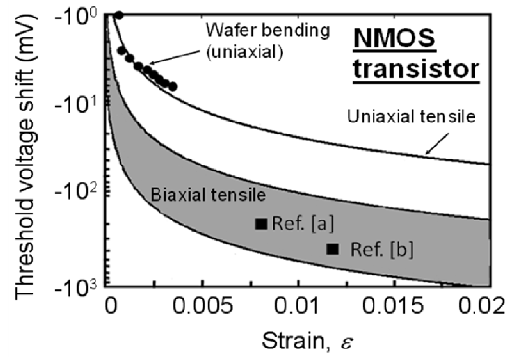


Figure 4-4 Uniaxial tensile stress and biaxial tensile stress lead to negative threshold voltage shift in <110> NMOS transistors on (100) Si [4.2]. Ref.[a]: Q. Xiang et al., VLSI Symp. 2003, pp.101-102. Ref. [b]: N. Sugii et al., TED, Vol. 49, pp. 2237-2243 (2002).

4.2.1 Strain-induced effects on silicon bandgap

There are several conflicting reports regarding the effects of uniaxial stress on the silicon bandgap. Lim et al. showed that the uniaxial tensile stress will narrow the silicon bandgap while the uniaxial compressive stress will broaden the silicon bandgap [4.2]. However, Yang et al. arrived at an opposite conclusion [4.3]. Using tight-binding bandstructure calculation, Lu et al. [4.4] showed that the silicon bandgap is reduced for both tensile stress and compressive stress. From Fig. 4-5, compressive stress results in a bigger bandgap narrowing compared to tensile stress. Moreover, the silicon bandgap narrowing induced by <100> uniaxial stress and biaxial stress is larger than that induced by <110> uniaxial stress. Since tensile stress-induced change in V_{th} [4.2] and compressive stress-induced change on the gate-induced-drain-leakage (GIDL) current [4.5] have been attributed to the reduction in silicon bandgap, it is more reasonable to say that silicon bandgap is reduced by both uniaxial tensile stress and uniaxial compressive stress.

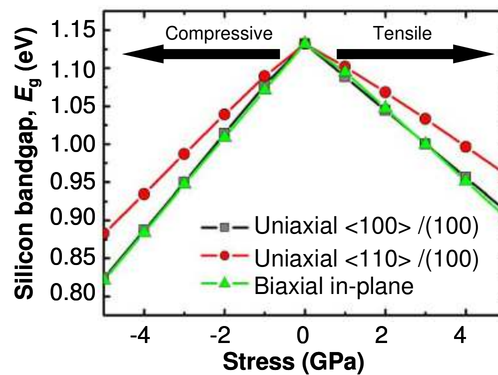


Figure 4-5 Effects of mechanical stress on the silicon bandgap of (100) Si[4.4].

4.2.2 Equations for strain-induced change in the threshold voltage

4.2.2.1 Biaxially tensile stressed MOS transistor

By considering equal inversion charge density at threshold voltage, Zhang and Fossum proposed that the equation for the strain-induced change in the V_{th} of a biaxial tensile stressed <110> NMOS transistor on the relaxed $Si_{1-y}Ge_y$ virtual substrate can be expressed as follows [4.1],

$$\Delta V_{th}(\sigma) \approx -\Delta E_c - (m-1)[E_G(0) - E_G(\sigma)] - m\frac{k_B T}{q}\ln\left(\frac{\mu_e(\sigma)}{\mu_e(0)}\right) \tag{4.1}$$

Similar expression can be obtained for a biaxial tensile stressed <110> PMOS transistor on the relaxed $Si_{1-y}Ge_y$virtual substrate [4.1].

$$\Delta V_{th}(\sigma) = -\Delta E_c + m[E_G(0) - E_G(\sigma)] + m\frac{k_B T}{q}\ln\left(\frac{\mu_h(\sigma)}{\mu_h(0)}\right) \tag{4.2}$$

where the body-effect coefficient (m) typically lies between 1.1 and 1.4 [3.8]. $E_G(0)$, $\mu_e(0)$ and $\mu_h(0)$ are the silicon bandgap, the electron mobility and the hole mobility of unstrained MOS transistor, respectively. $E_G(\sigma)$ $\mu_e(\sigma)$ and $\mu_h(\sigma)$ are the silicon bandgap, the electron mobility and the hole mobility of the stress-engineered MOS transistor, respectively.. $\Delta E_c = \chi_{ss} - \chi_{Si}$ where χ_{Si} is the electron affinity of the poly-Si gate and χ_{ss} is the electron affinity of the biaxial tensile stressed silicon channel.

4.2.2.2 Uniaxially tensile stressed <110> MOS transistor

For uniaxial tensile stress, ΔE_c term is zero because the electron affinity for n^+ poly-Si gate is the same as the electron affinity of the silicon channel [4.2]. Hence, the equation for the strain-induced change in the V_{th} of a uniaxially stressed <110> NMOS transistor can be expressed as [4.2],

$$\Delta V_{th}(\sigma) \approx -(m-1)[E_G(0) - E_G(\sigma)] - m\frac{k_B T}{q}\ln\left(\frac{\mu_e(\sigma)}{\mu_e(0)}\right) \tag{4.3}$$

Similarly, the equation of strain-induced change in V_{th} of a unaxially stressed <110> PMOS transistor can be expressed as,

$$\Delta V_{th}(\sigma) \approx m[E_G(0) - E_G(\sigma)] + m\frac{k_B T}{q}\ln\left(\frac{\mu_h(\sigma)}{\mu_h(0)}\right) \tag{4.4}$$

From Fig. 4-5, uniaxial tensile stress leads to silicon bandgap narrowing and thus $[E_G(0) - E_G(\sigma)]$ is positive. From Section 2.2.4, the application of uniaxial tensile stress leads to $\mu_e(\sigma) > \mu_e(0)$ and $\mu_h(\sigma) < \mu_h(0)$ [1.14, 1.15]. From eqn. (4.3), uniaxial tensile stress results in a negative $\Delta V_{th}(\sigma)$ in NMOS transistor. Since the threshold voltage of enhancement type NMOS transistor is a positive value, tensile stress is expected to reduce the magnitude of the threshold voltage of NMOS transistor. From eqn. (4.4), the strain-induced effects on the threshold voltage of PMOS transistor is not clear because the first term is positive whereas the second term is negative. Hence, there is a need to verify the effects of uniaxial stress on V_{th} of PMOS transistor.

4.3 Gate leakage current

Fig. 4-6 shows the strain-induced change in gate tunneling current ($\Delta I_g/I_g$) for <110> MOS transistors fabricated on (100) Si surface [4.6, 4.7, 4.8]. For <110> NMOS transistor, uniaxial tensile stress decreases $\Delta I_g/I_g$ but uniaxial compressive stress increases $\Delta I_g/I_g$. For <110> PMOS transistor, uniaxial compressive stress decreases $\Delta I_g/I_g$ but uniaxial tensile stress increases $\Delta I_g/I_g$. The strain-induced change in gate current can be understood from the strain-induced change in the out-of-plane effective mass (m_z), the subband splitting and the carrier re-population. As explained in Chapter 2.2.4, uniaxial <110> tensile stress lowers the electron energy level of out-of-plane conduction band valleys (Δ_5, Δ_6) while uniaxial <110> compressive stress increases the electron energy level of out-of-plane conduction band valleys (Δ_5, Δ_6). In the other words, the tunneling barrier height (ϕ_B) for electrons increases under tensile stress but decreases under compressive stress. From Table 2-2, π_{out} for <110> NMOS transistor is positive, and thus tensile stress will increase m_z but compressive stress will decrease m_z. Since the tunneling probability decreases with increasing ϕ_B and increasing m_z [2.75], the gate tunneling current for NMOS transistor is expected to decrease under tensile stress but increase under compressive stress.

Owing to the strain-induced degeneracy splitting of the heavy-hole band and the light-hole band, the term "heavy hole" and "light hole" is ambiguous. The upper band refers to the valence subband that is nearer to the conduction band, whereas the lower band refers to the valence subband that is further away from the conduction band [2.86]. Since holes will preferentially occupy the lowest hole energy level, most of the holes will reside in the upper band. From Fig. 4-7, the hole energy lowering by tensile stress is comparable to that of compressive stress [2.86]. Hence, the strain-induced change in the gate tunneling current for PMOS transistor is unlikely to be caused by the change in the ϕ_B for holes. From Table 2-2, π_{out} for <110> PMOS transistor is negative, and thus tensile stress will decrease m_z but compressive stress will increase m_z. Since the tunneling probability decreases with increasing m_z [2.75], the gate tunneling current for PMOS transistor is expected to increase under tensile stress but decrease under compressive stress [4.8].

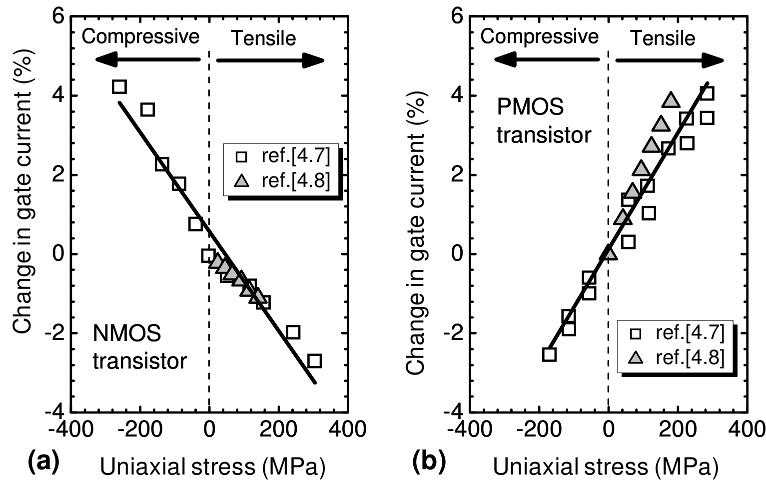


Figure 4-6 The effects of uniaxial stress on gate tunneling current of <110> MOS transistors on (100) Si surface[4.7,4.8].

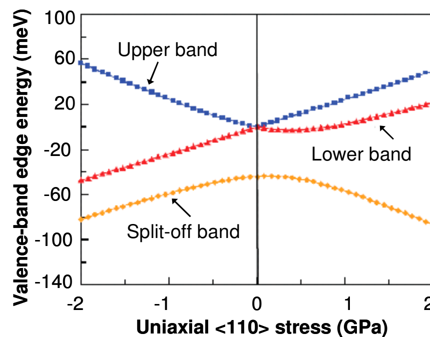


Figure 4-7 Valence band edge energy as a function of the uniaxial <110> stress [2.86].

4.4 Gate-induced-drain leakage (GIDL) current

From Fig. 4-8, the new GIDL current mechanism will require a smaller gate-to-drain bias for band-to-band tunnelling (BTBT) to occur because of the built-in potential barrier at the p-body and n^+ drain junction [4.9]. From Fig. 4-9, NMOS transistor with the highest threshold adjustment implant has the largest GIDL current [4.9]. This shows that the GIDL current in nanoscale MOS transistor is indeed dominated by the BTBT current in the reverse-biased p-n junction formed between the body and drain.

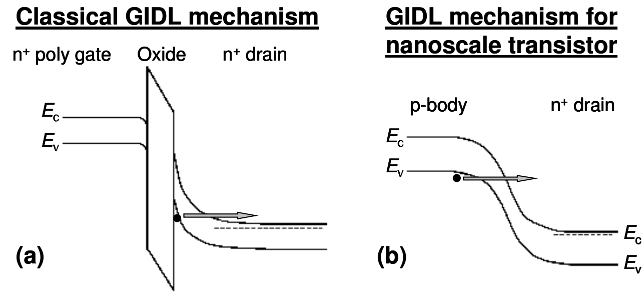


Figure 4-8 (a) Classical GIDL mechanism, (b) New GIDL mechanism[4.9]

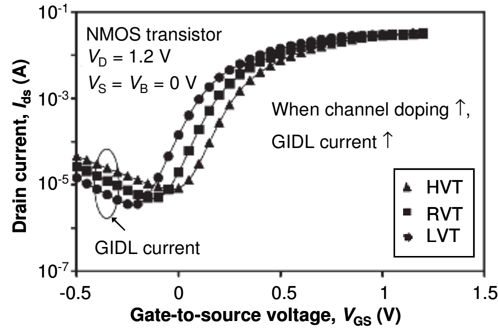


Figure 4-9 Effects of threshold adjustment implant on the GIDL current of a nanoscale NMOS transistor (L = 40 nm)[4.9]. HVT: High dose, RVT: Regular dose, LVT: Low dose.

The tunneling current density for the BTBT process in the reverse-biased p-n junction is expressed in the following [4.9],

$$J_{\mathrm{BTBT}} = \frac{\sqrt{2m_{\mathrm{x}}}\, q^3 E_{\mathrm{j}} V_{\mathrm{app}}}{4\pi^3 \hbar^2 \sqrt{E_{\mathrm{G}}}} \exp\left(-\frac{4\sqrt{2m_{\mathrm{x}}}\, E_{\mathrm{G}}^{3/2}}{3qE_{\mathrm{j}}\hbar} \right) \tag{4.5}$$

where E_j is the maximum electric field in the p-n junction. V_{app} is the applied reverse bias on the p-n junction. TCAD simulation [4.10] and experimental results [4.5] show that STI-induced compressive stress will increase GIDL current owing to the strain-induced bandgap narrowing.

4.5 Gate oxide thickness

4.5.1 Poisson's ratio

Poisson's ratio (v) is defined as the ratio of the transverse strain (δ_y) to the longitudinal strain (δ_x) [4.11]: $v = - \delta_y / \delta_x$. From Fig. 4-10, the application of uniaxial tensile stress along x direction will lead to elongation along x direction (positive δ_x) and a contraction along y direction (negative δ_y). On the other hand, the application of uniaxial compressive stress along x direction will lead to a negative δ_x and a positive δ_y. By convention, σ_x is positive for tensile stress and negative for compressive stress.

Unlike the substrate-induced stress, the process-induced stress such as the tensile stressed CESL will transfer mechanical stress to the channel, the gate electrode and the gate dielectric. Experiment data show that the Poisson's ratio of SiO_2 (v_{oxide}) is between 0.17 and 0.25 [4.12] and Poisson's ratio of Si_3N_4 is about 0.26 [4.13]. Hence, the Poisson's ratio of silicon oxynitride is expected to be a positive number. Therefore, the application of uniaxial tensile stress along the channel direction will lead to the physical thinning of the gate dielectric. Considering a MOS transistor fabricated on (100) surface-oriented silicon with <110> channel orientation, the Young's modulus of silicon (E_{si}) is 169 GPa [4.14]. By Hooke's law, the mechanical stress experienced by silicon along x direction ($\sigma_{silicon}$) is given by the product of the mechanical strain experienced by silicon along x direction ($\delta_{silicon}$) and E_{si}. Since the physical gate oxide thickness (~20 Å) is much thinner than the Si substrate, the gate oxide will tend to follow the strain at the Si/ SiO_2 interface [4.15]. In the other words, the strain experienced by the gate oxide along x direction $(\delta_{oxide})_x$ is approximately equal to ($\delta_{silicon}$). Using v_{oxide} of 0.17 and E_{si} of 169 GPa, the percentage increase in T_{ox} is estimated to be around 0.03 % when uniaxial compressive stress of 300 MPa is applied along <110> channel direction. This indicates that the change in the physical gate oxide thickness owing to Poisson's ratio is almost negligible.

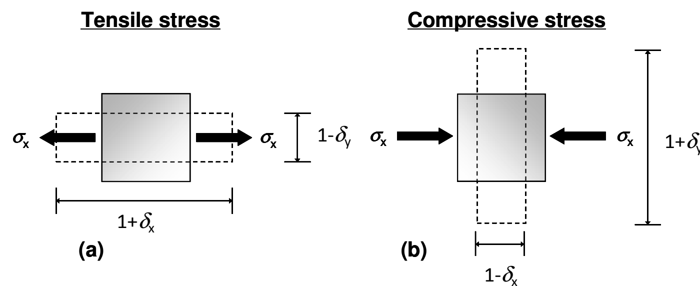


Figure 4-10 Effects of Poisson's ratio under (a) tensile stress, (b) compressive stress [4.11].

4.5.2 Quantum confinement

For the state-of-the-art MOS transistors with ultra-thin gate oxide and high substrate doping, the electrons/holes in the potential well at Si-SiO_2 interface create a two-dimensional electron gas (2-DEG)/ two-dimensional hole gas (2-DHG) [4.16]. In the other words, their potential energy in the direction perpendicular to the silicon surface (z) is quantized, whereas the energy of their motion in the plane parallel to the semiconductor surface can take arbitrary values. The electrostatics of the 2-DEG or 2-DHG can be described by the system of 1D Schrödinger's equation and Poisson's equation, as follows[4.16].

$$\left[-\frac{\hbar^2}{2m_z}\frac{\partial^2}{\partial z^2}-qV(z)\right]\psi_j(z)=E_j\,\psi_j(z) \tag{4.6}$$

where j = 1,2,3 … Note that j = 1 corresponds to the ground state.

$$\frac{d^2V}{dz^2}=-\left[\rho_{dep1}(z)-q\sum_j N_j\psi_j^2(z)\right]/(\varepsilon_o\varepsilon_{Si}) \tag{4.7}$$

where m_z is the effective mass describing the dynamics of the electron motion in the z direction. E_j refers to the discrete energy level. $V(z)$ is the electrostatic potential that describes the band bending at Si-SiO_2 interface. ρ_{dep1} represents the charge density of the depletion layer. N_j is the carrier concentration in the jth subband. Since the self-consistent calculations of 1D Schrödinger's equation and Poisson's equation [4.16, 4.17] are very time consuming, the approximate inversion layer centroid (z_I) can be obtained by using the triangular potential well approximation. From Fig. 4-11, the triangular potential well approximation is closer to the actual E_c bending when j is small. This corresponds to the case of the subthreshold conduction.

$$V(z)=\begin{cases}0 & \text{for } z<0\\ -F_s z & \text{for } z>0\end{cases} \tag{4.8}$$

where z = 0 corresponds to the Si-SiO_2 interface. z > 0 refers to the depth into silicon substrate. F_s is the effective electric field at the Si-SiO_2 interface. Since $\frac{d^2V}{dz^2}=0$, we can decouple the 1D Schrödinger's equation and Poisson's equation.

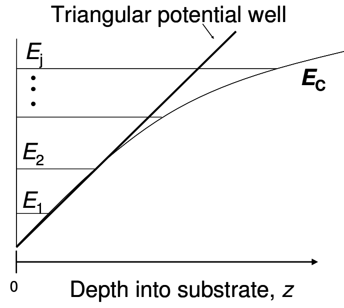


Figure 4-11 Triangular potential well approximation.

Substituting $V(z) = -F_s z$ into eqn. (4.6) and re-arranging,

$$\left(\frac{2m_z qF_s}{\hbar^2}\right)^{-2/3}\frac{d^2\psi_j}{dz^2} - \left(\frac{2m_z qF_s}{\hbar^2}\right)^{1/3}\left(z - \frac{E_j}{qF_s}\right)\psi_j = 0 \quad (4.9)$$

Next, we introduce a new variable, $s = \left(\frac{2m_z qF_s}{\hbar^2}\right)^{1/3}\left(z - \frac{E_j}{qF_s}\right)$ (4.10)

Using Faà di Bruno's formula, which generalizes chain rule to higher derivatives [4.18],

$$\frac{d^2\psi_j}{ds^2} = \frac{d^2\psi_j}{dz^2}\left(\frac{dz}{ds}\right)^2 + \frac{d\psi_j}{ds}\left(\frac{d^2 s}{dz^2}\right) \quad (4.11)$$

Differentiating eqn. (4.10) with respect to z,

$$\frac{ds}{dz} = \left(\frac{2m_z qF_s}{\hbar^2}\right)^{1/3} \quad , \quad \frac{d^2 s}{dz^2} = 0 \quad (4.12)$$

Substituting eqn. (4.12) into eqn. (4.11),

$$\frac{d^2\psi_j}{ds^2} = \left(\frac{2m_z qF_s}{\hbar^2}\right)^{-2/3}\frac{d^2\psi_j}{dz^2} \quad (4.13)$$

Substituting eqn. (4.10) and eqn. (4.13) into eqn. (4.9),

$$\frac{d^2\psi_j}{ds^2} - s\psi_j = 0 \quad (4.14)$$

This is in the form of the Airy differential equation [4.19]: $f''(x) - xf(x) = 0$ with two linearly independent solutions are labeled Ai(x) and Bi(x). From Fig. 4-12, Bi(x) diverges for $x \rightarrow \infty$ while Ai(x) decays monotonically to zero for $x > 0$ and oscillates for $x < 0$.

From Handbook of mathematical functions [4.20],

$$\mathrm{Ai}(-x) \approx \frac{1}{\sqrt{\pi}\, x^{0.5}} \left[\sin\left(\frac{2}{3} x^{3/2} + \frac{\pi}{4} \right) \right] \tag{4.15}$$

Considering the boundary conditions of ψ_j,

When z = 0, $s = \left(\frac{2m_z qF_s}{\hbar^2} \right)^{1/3} \left(-\frac{E_j}{qF_s} \right) = -\bar{S}$

$$\psi_j(s = -\bar{S}) = 0 \tag{4.16a}$$

When $z = \infty$, $s = \infty$,

$$\psi_j(s = \infty) = 0 \tag{4.16b}$$

Hence, the solution of eqn.(4.14) can be expressed as,

$$\psi_j = \alpha\, \mathrm{Ai}(s) \tag{4.17}$$

where α is a constant. E_j is chosen such that that Ai(s) has j roots, and E_j is taken to be zero for all s below this j^{th} root.

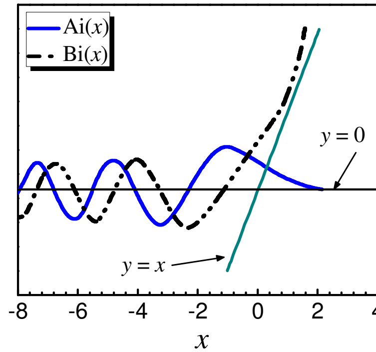


Figure 4-12 Graphical representation of Airy functions, Ai(x) and Bi(x) [4.19].

Solving eqn.(4.15), eqn.(4.16a) and eqn. (4.17),

$$E_j = \left(\frac{\hbar^2}{2m_z} \right)^{1/3} \left[\frac{3\pi}{2} qF_s \left(j - \frac{1}{4} \right) \right]^{2/3} \tag{4.18}$$

Inversion charge centroid (z_I) is given by [4.16],

$$z_I = 3\left(\frac{\varepsilon_0 \varepsilon_{SI} \hbar^2}{12 m_z q^2 N^*} \right)^{1/3} \tag{4.19}$$

where $N^{*} = N_{\text{depl}} + \frac{11}{32} N_{\text{inv}}$. Experimental results show that z_{I} of silicon inversion layer about 1 nm [4.22]. Considering quantum mechanical effects, the effective gate oxide capacitance (C_{ox}^{*}) is given by [4.23, 4.24].

$$C_{\text{ox}}^{*} = \frac{\varepsilon_0 \varepsilon_{\text{ox}}}{T_{\text{ox}} + \frac{\varepsilon_{\text{ox}}}{\varepsilon_{\text{Si}}}(z_{\text{I}} + z_{\text{D}})} \tag{4.20}$$

where z_{D} is the poly-Si depletion charge centroid. Unlike $C_{\text{ox,inv}}$ in eqn.(3.19) that refers to gate oxide capacitance in strong inversion, C_{ox}^{*} is a general term for gate oxide capacitance for depletion mode, subthreshold mode as well as strong inversion. According to Nakabayashi et al. [2.83], a smaller m_{x} leads to a bigger m_{z}, resulting in a smaller z_{I}. From Table 2-2, uniaxial tensile stress decreases m_{x} of NMOS transistor, whereas uniaxial compressive stress decreases m_{x} of PMOS transistor. Hence, tensile stress will increase C_{ox}^{*} of NMOS transistor but decrease C_{ox}^{*} of PMOS transistor. On the other hand, compressive stress will increase C_{ox}^{*} of PMOS transistor but decrease C_{ox}^{*} of NMOS transistor. From Fig. 4-13, experimental data shows that the application of compressive stress to PMOS transistor has minimal effects on the physical gate oxide thickness (T_{ox}) but decreases the electrical oxide thickness (EOT) [4.25]. Similarly, the application of uniaxial stress to NMOS transistor has minimal effects on T_{ox} but decreases EOT.

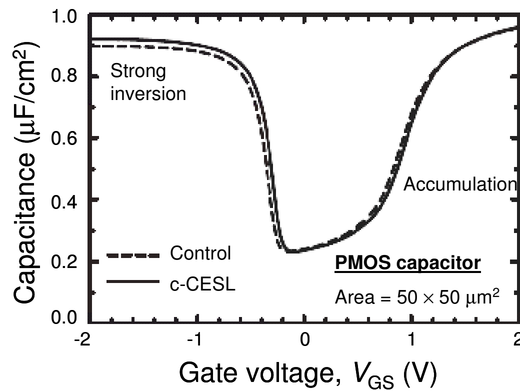


Figure 4-13 Effects of compressive stress on the C-V characteristics of PMOS capacitor [4.25].

4.6 Subthreshold off-current (I_{off})

The subthreshold conduction occurs when V_{GS} is below the threshold voltage. From Fig. 4-14, the subthreshold current appears as a linear portion of the log I_{ds} versus V_{GS} characteristics of a NMOS transistor. Unlike the strong inversion region in which the drift current dominates, the subthreshold current is dominated by diffusion current. The equation of the subthreshold current of MOS transistor can be expressed as [4.26],

$$I_{ds} = \mu_{sub} C_{ox} \frac{W}{L}(m-1)\left(\frac{k_B T}{q}\right)^2 \exp\left[\frac{q(V_{GS} - V_{th,sat})}{m k_B T}\right]\left[1 - \exp\left(\frac{-qV_{DS}}{k_B T}\right)\right] \tag{4.21}$$

Where μ_{sub} is the low-field mobility in the subthreshold regime [4.27-4.29]. Since Einstein relation states that the diffusion coefficient is related to μ_{eff}, we will expect μ_{sub} to increase when μ_{eff} increases. From Fig. 4-15, magnetoresistance (MR) mobility measurements shows that there is indeed a correlation between μ_{eff} and μ_{sub} [4.29]. From Fig. 4-16, uniaxial tensile stress increases I_{on} and subthreshold I_{off} of <110> NMOS transistor [4.30]. From Fig. 4-17, CESL-induced compressive stress increases the subthreshold current of <110> PMOS transistors [4.25].

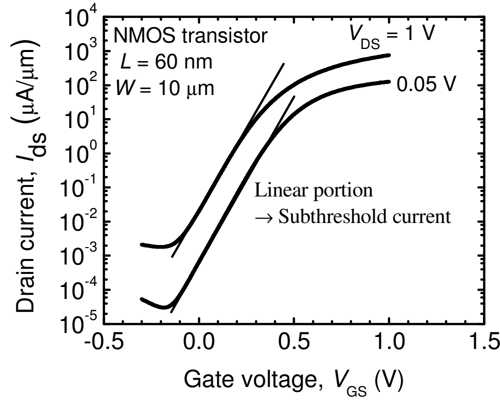


Figure 4-14 Subthreshold current is taken to be the linear portion of log I_{ds} versus V_{GS} characteristics of NMOS transistor.

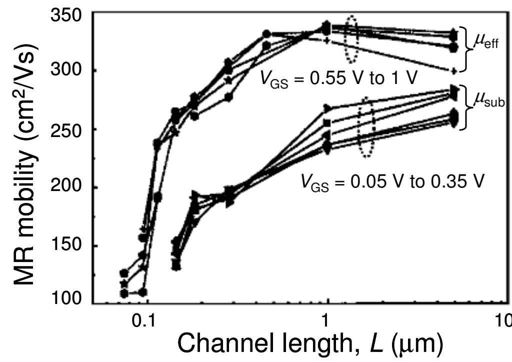


Figure 4-15 Magnetoresistance (MR) mobility measurements show that there is a correlation between μ_{sub} and μ_{eff} [4.29].

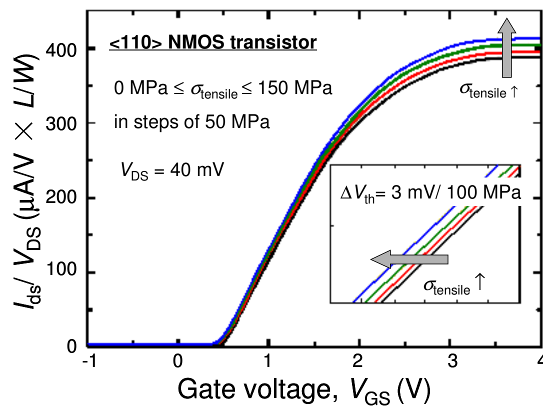


Figure 4-16 Effects of externally applied tensile stress on the I_{ds} versus V_{GS} characteristics of a relatively long NMOS transistor with 4 nm conventional gate dielectric [4.30]. Inset shows the threshold voltage shift in the subthreshold regime.

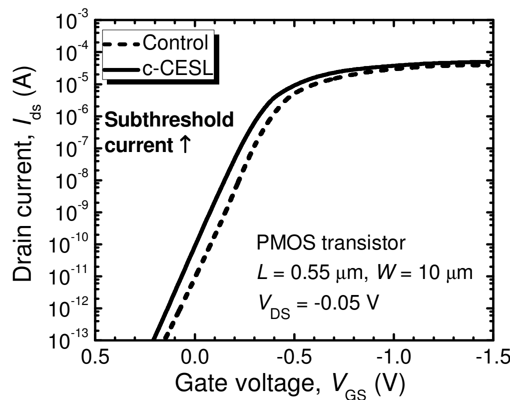


Figure 4-17 Effects of CESL-induced compressive stress on the subthreshold current of <110> PMOS transistor with $L = 0.55$ μm and $W = 10$ μm [4.25].

The subthreshold swing (S_{ts}) is defined as the inverse of the logI_{ds} versus V_{GS} characteristics of a MOS transistor. In general, the equation of S_{ts} can be expressed as follows [4.26],

$$S_{ts} = 2.3\frac{mk_{B}T}{q} \tag{4.22}$$

where the body-effect coefficient (m) can be expressed as [4.26],

$$m = 1 + \frac{1}{C_{ox}}\left(\sqrt{\frac{\varepsilon_{o}\varepsilon_{Si}qN_{ch}}{4\psi_{B}}}\right) \tag{4.23}$$

where N_{ch} is the channel doping concentration. ψ_{B} is the difference between the Fermi level in the channel region and the intrinsic Fermi level. For very short MOS transistors, the equation of S_{ts} has to be modified to account for the degradation of S_{ts} when the short channel effects are strong [4.31] but qualitatively S_{ts} still decreases when C_{ox} increases. For MOS transistors with thick gate oxide and low channel doping, $C_{ox} = \varepsilon_{o}\varepsilon_{ox}/T_{ox}$. For the state-of-the-art MOS transistors with ultra-thin gate oxide and high channel doping, quantum confinement must be considered. Hence, C_{ox} in eqn. (4.23) must be replaced by C_{ox}^{*} in eqn.(4.35).

$$m = 1 + \frac{1}{C_{ox}^{*}}\left(\sqrt{\frac{\varepsilon_{o}\varepsilon_{Si}qN_{ch}}{4\psi_{B}}}\right) \tag{4.24}$$

As discussed in Section 4.5.2, the application of uniaxial tensile stress along the channel direction will increase C_{ox}^{*} of NMOS transistor but decrease C_{ox}^{*} of PMOS transistor owing to the strain-induced change in the quantum confinement. Based on eqn. (4.22) and eqn.(4.24), the application of uniaxial tensile stress is expected to improve S_{ts} of NMOS transistor but degrade S_{ts} of PMOS transistor.

4.7 Results and Discussion

4.7.1 Physics behind the linearity in the log I_{off} versus I_{on} characteristics

One of the assumptions used to determine the amount of strain-induced I_{on} improvement is the linearity of the log I_{off} versus I_{on} plot. From Fig. 4-18, we observed that the linearity is only valid for transistors in the short channel effect (SCE) regime, which correspond to $L = 0.12$ µm, 0.13 µm, 0.14 µm, 0.15 µm in this context. The physics behind the linearity can be understood as follows. In Chapter 3, we have improvised Lundstrom's quasi-ballistic theory and arrived at a simplified drain current equation for the nanoscale MOS transistor. Since I_{on} is taken to be the saturation drain current when $V_{GS} = V_{DD}$, we can rewrite eqn. (3.19) as follows,

$$I_{on} = v_{sat_eff}(\mu_{eff}, T)WC_{ox,inv}(V_{DD} - V_{th,sat_IV}) \tag{4.25}$$

From Fig.4-19, v_{sat_eff} and V_{th,sat_IV} have good linearity with DIBL, $V_{th,lin}$ and $V_{th,sat}$. Note that Constant current method with reference current of 0.1µA W/L is used to extract the saturation threshold voltage ($V_{th,sat}$) and the linear threshold voltage ($V_{th,lin}$). In this context, DIBL = $V_{th,lin} - V_{th,sat}$. In the other words, eqn. (4.25) can be expressed in terms of DIBL, $V_{th,sat}$ and $V_{th,lin}$.

$$I_{on} \approx A_1 \times \text{DIBL} + A_2 \tag{4.26a}$$

$$I_{on} \approx A_3 \times V_{th,sat} + A_4 \tag{4.26b}$$

$$I_{on} \approx A_5 \times V_{th,lin} + A_6 \tag{4.26c}$$

where A_1, A_2, A_3, A_4, A_5 and A_6 are constants. From Fig. 4-19, DIBL, $V_{th,lin}$ and $V_{th,sat}$ are less than 1, and thus we can ignore the second order terms. During the subthreshold conduction, MOS transistor behaves like a bipolar transistor.

$$I_{off} = I_{BJTO} W \exp\left(\frac{qV_{BE_eq}}{k_B T}\right) \tag{4.27}$$

where I_{BJTO} is the saturation current constant. During normal bipolar transistor operation, a base-emitter voltage (V_{BE_eq}) is applied to reduce the energy barrier at the emitter-base junction. Since the emitter-base energy barrier of the parasitic bipolar transistor is equivalent to the energy barrier ($\phi_{barrier}$) at the source-channel, V_{BE_eq} will be proportional to

DIBL because the reduction of ϕ_{barrier} is proportional to the change in the threshold voltage when V_{DS} is varied [4.32]. From Fig. 4-20, $\log I_{off}$ has a good linearity with DIBL, $V_{th,sat}$ and $V_{th,lin}$.

$$\log I_{off} = A_7 \times \text{DIBL} + A_8 \tag{4.28a}$$

$$\log I_{off} = A_9 \times V_{th,sat} + A_{10} \tag{4.28b}$$

$$\log I_{off} = A_{11} \times V_{th,lin} + A_{12} \tag{4.28c}$$

where A_7, A_8, A_9, A_{10}, A_{11} and A_{12} are constants. From Fig. 4-21, the equation of best-fit line is $\log I_{off} = 8.333\times10^{-3} \times I_{on} - 1.122$. This is comparable with the $\log I_{off}$ versus I_{on} equations that are obtained by mathematical elimination of the third parameter (see Table 4-1). Based on MOS transistors fabricated by 0.11 μm CMOS technology, our experimental results show that DIBL, $V_{th,sat}$ and $V_{th,lin}$ can be used as the third parameter in the $\log I_{off}$ versus I_{on} plot. This conclusion is also applicable to MOS transistors that are fabricated by 65 nm CMOS technology, as shown in Fig. 4-28.

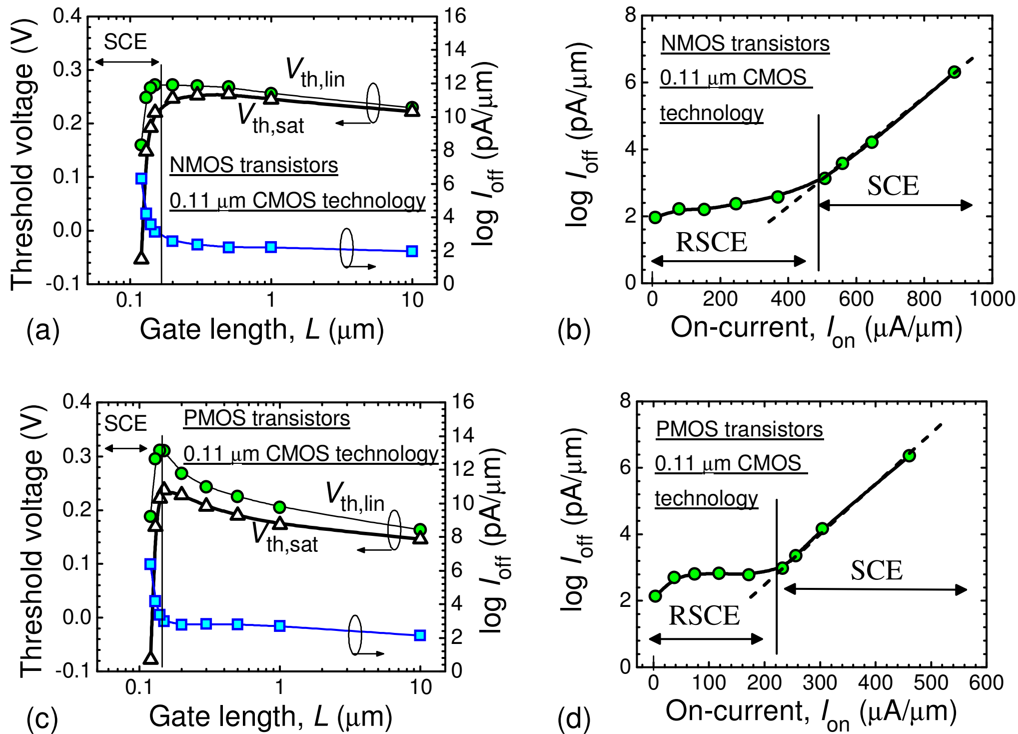


Figure 4-18 Linearity of $\log I_{off}$ versus I_{on} plot is valid for the short channel effect (SCE) regime of transistors fabricated by 0.11 μm CMOS technology. (a) & (b) refer to NMOS transistors. (c) & (d) refer to PMOS transistors.

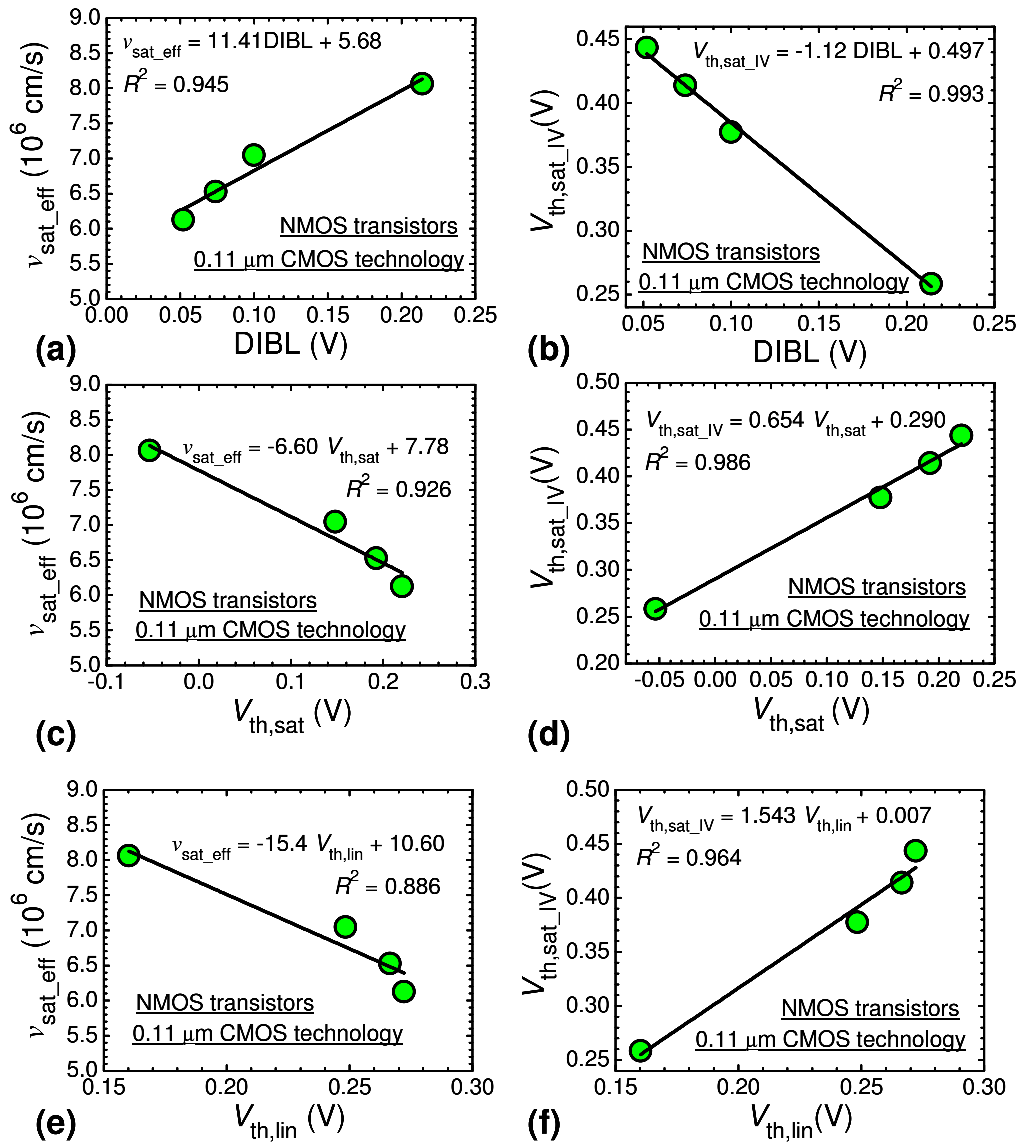


Figure 4-19 For NMOS transistors in the SCE regime: (a) v_{sat_eff} has a good linearity with DIBL, (b) V_{th,sat_IV} has a good linearity with DIBL, (c) v_{sat_eff} has a good linearity with $V_{th,sat}$,(d) V_{th,sat_IV} has a good linearity with $V_{th,sat}$, (e) v_{sat_eff} has a good linearity with $V_{th,lin}$, (f) V_{th,sat_IV} has a good linearity with $V_{th,lin}$.

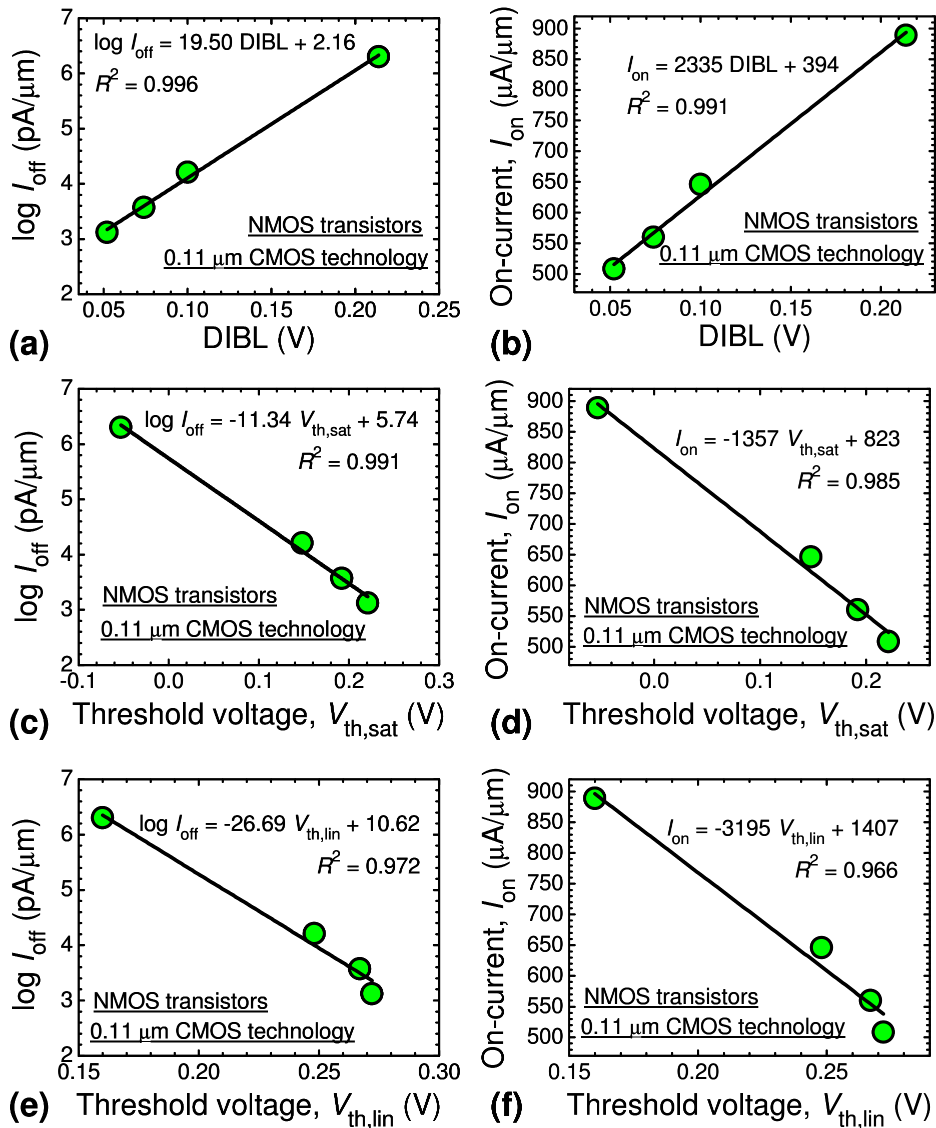


Figure 4-20 Third parameter in the log I_{off} versus I_{on} characteristics of NMOS transistors in SCE regime: (a) and (b) show that DIBL has a linear relationship with log I_{off} and I_{on}. (c) and (d) show that $V_{th,sat}$ has a linear relationship with log I_{off} and I_{on}. (e) and (f) show that $V_{th,lin}$ has a linear relationship with log I_{off} and I_{on}.

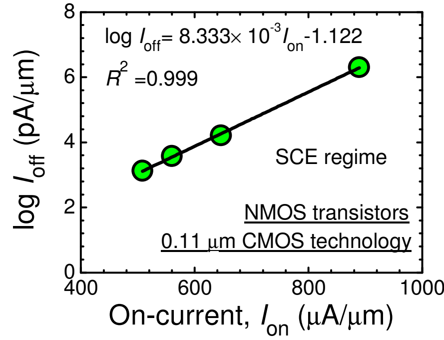


Figure 4-21 Best-fit line of the log I_{off} versus I_{on} plot of NMOS transistors in the SCE regime.

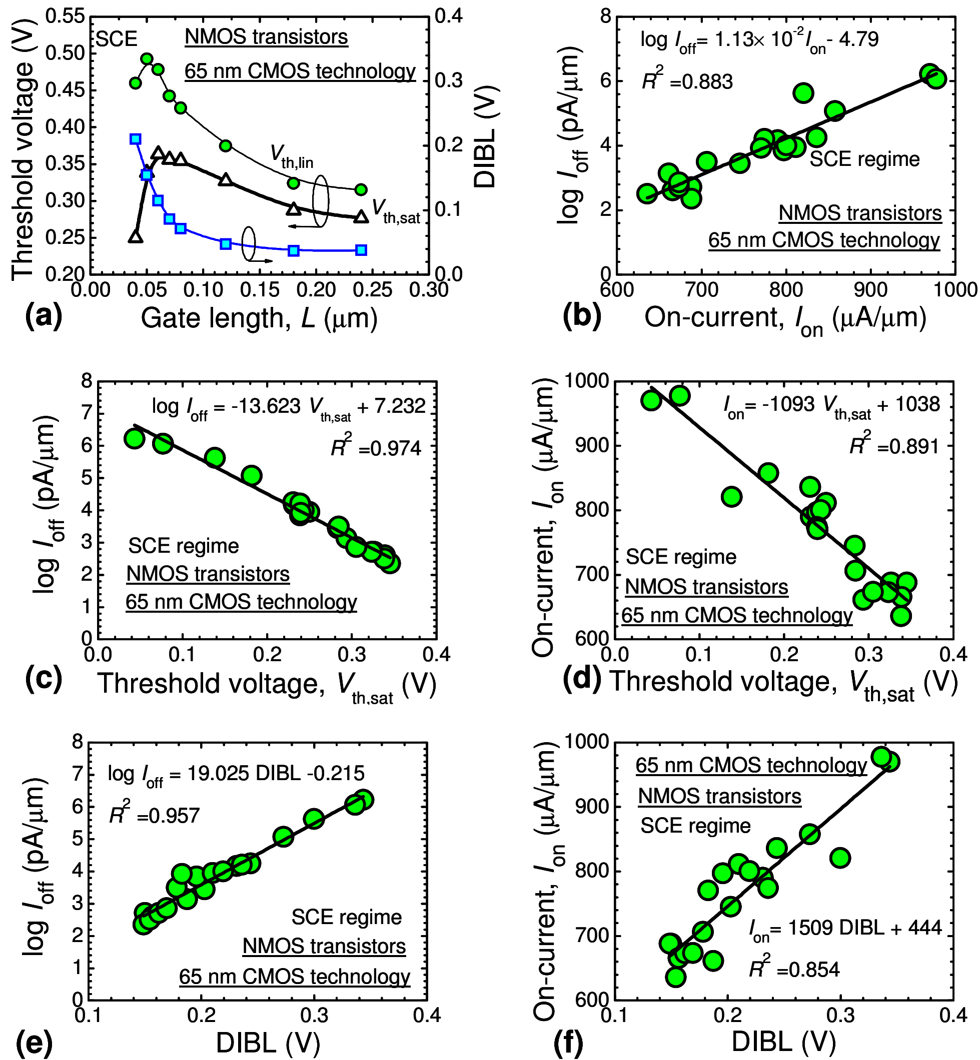


Figure 4-22 Linearity of log I_{off} versus I_{on} characteristics of NMOS transistors in SCE regime (65 nm CMOS technology).

Table 4-1 Evaluation of the suitability of the third parameter in the log I_{off} versus I_{on} plot.

Best-fit line of I_{on}	Best-fit line of logI_{off}	Mathematical elimination
I_{on} = 2335×DIBL + 394	logI_{off} = 19.50×DIBL + 2.16	logI_{off} = 8.35×10^{-3}×I_{on}-1.13
I_{on} = -1357×$V_{th,sat}$ + 823	logI_{off} = -11.34×$V_{th,sat}$ + 5.74	logI_{off} = 8.36×10^{-3}×I_{on}-1.14
I_{on} = -3195×$V_{th,lin}$+ 1407	logI_{off} = -26.69×$V_{th,lin}$+ 10.62	logI_{off} = 8.35×10^{-3}×I_{on}-1.13

However, the linearity of the log I_{off} versus I_{on} characteristics may not be valid for PMOS transistors fabricated by 40 nm CMOS technology even though the transistors are in the SCE regime. From Fig. 4-23, there are two peaks in the V_{th} versus L characteristics. The first peak occurs at L around 0.15 µm because of the pile up of channel dopants near the S/D regions due to the defects generated during the deep S/D implantation [4.33 – 4.35]. The second peak occurs around 55 nm because of the high-angle halo implant that is used to suppress the short-channel effects [4.36, 4.37]. Typically, the threshold voltage of transistors in SCE regime will decrease with a reduction in L. This is valid for our PMOS transistors with L = 45 nm and L = 50 nm. However, we observed an anomalous increase in the threshold voltage when L is further reduced to 32 nm and 40 nm. When L is further reduced to 32 nm and 40 nm, there are kinks in the on-state current (I_{ds_on}) versus L characteristics and off-state current (I_{ds_off}) versus L characteristics. Our theory is that the cross-over of the halo implants is responsible for the above observations in the look-ahead transistor test structure. From Fig. 4-24, scenario 1 refers to the the onset of the cross-over of halo implants occurs at mask gate length (L_1). Scenario 2 refers to the cross-over of halo implants occurs at a slightly smaller mask gate length (L_2). If we were to disregard the effects of the halo implants on the gate-to-S/D overlap, the metallurgical channel length is expected to be bigger for scenario 1 compared to scenario 2 (i.e. $L_{met,1} > L_{met,2}$). However, experimental data have shown that halo implants can decrease the gate-to-S/D overlap [4.38], resulting in a longer L_{eff} [4.39]. When the effects of halo implants on the gate-to-S/D overlap are considered, the reduction in the gate-to-S/D overlap by the halo implants is expected to be bigger for scenario 2 compared to scenario 1. The difference in the reduction of the gate-to-S/D overlap can offset a small difference in the gate electrode length, resulting in an interesting phenomenon whereby $L_{met,1}$ is comparable to $L_{met,2}$ even though they have different mask gate lengths. This results in kinks in the I_{off} and I_{on} versus L characteristics, leading to a poor linearity in logI_{off} versus I_{on} plot even though the transistors are in SCE regime.

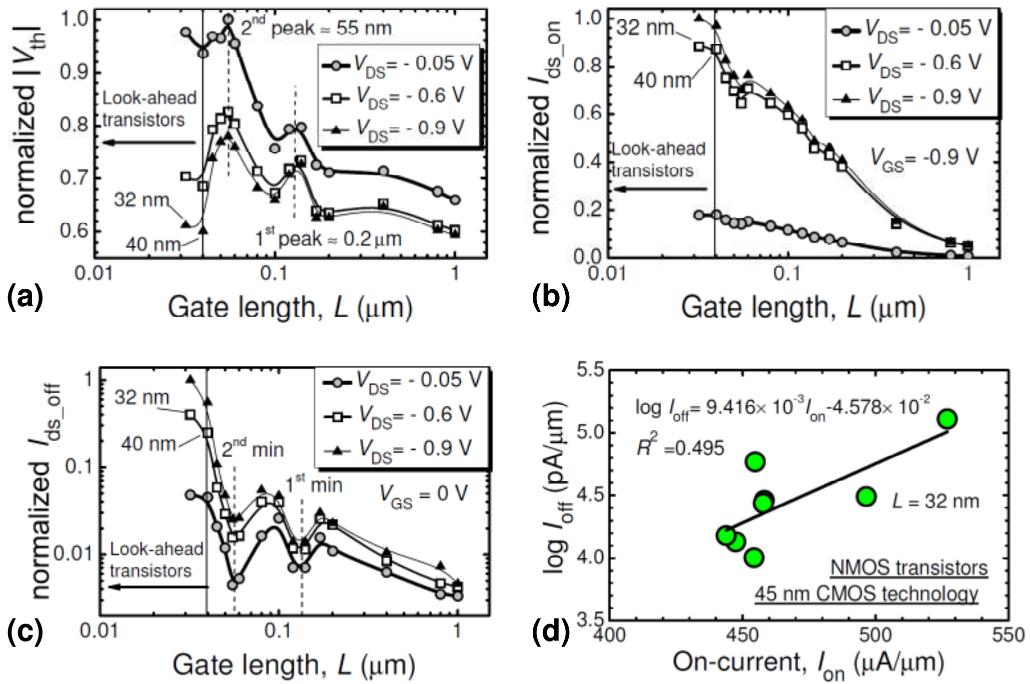


Figure 4-23 Linearity of the log I_{off} versus I_{on} characteristics may not be valid for PMOS transistors that are fabricated by 45 nm CMOS technology: (a) Anomalous increase in the threshold voltage, (b) Kink in the I_{ds_on} versus L characteristics, (c) Kink in the I_{ds_off} versus L characteristics, (d) Poor linearity in SCE regime. I_{ds_on} is I_{ds}@ V_{GS} = V_{DD} = -0.9 V. I_{on} is I_{ds}@ V_{GS} = V_{DS} =V_{DD} = -0.9 V. I_{ds_off} is I_{ds}@ V_{GS} = 0 V . I_{off} is I_{ds}@V_{GS} = 0 V & V_{DS} =V_{DD} = -0.9 V.

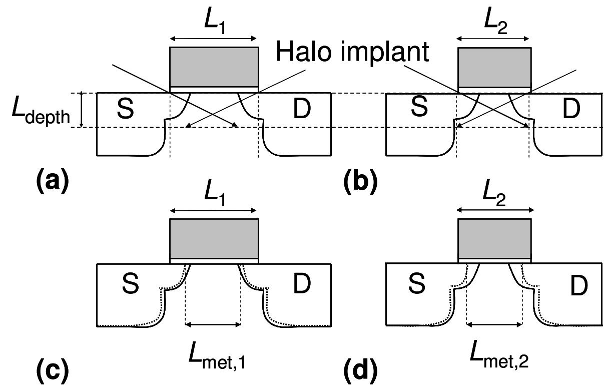


Figure 4-24 Mechanism responsible for the poor linearity in log I_{off} versus I_{on} plot of PMOS transistors fabricated by 40nm CMOS technology. (a) & (c) refer to the case for the onset of the cross-over of halo implants. (b) & (d) refer to the case for the cross over of halo implants.

4.7.2 Four-point bending measurement

Experimental data has shown that the externally applied tensilel stress can increase the subthreshold current of NMOS transistors [4.30]. This work will explain the physics behind the strain-induced increase in the subthreshold current. From Fig. 4-25, the strain-induced change in silicon bandgap (E_G) is a non-linear function of mechanical stress, whereas the strain-induced change in the effective electron mass (m_x) is a linear function of mechanical stress. Hence, we would expect the inherent mechanical stress in the MOS transistors to affect the mechanical stress sensitivity of the silicon bandgap. From Fig. 4-26, the effects of externally applied tensile stress on E_G is expected to be the bigger for highly tensile stressed CESL and highly compressive stressed CESL, as compared to neutral CESL and low compressive stressed CESL. On the other hand, we would expect the strain-induced effects on I_{on} to be less influenced by the inherent mechanical stress in the MOS transistor because m_x is a linear function of mechanical stress.

From Section 4.2 and Section 4.4, the strain-induced change in $V_{th,sat}$ and GIDL current are functions of the silicon bandgap. Hence, we would expect the effects of externally applied tensile stress on the threshold voltage and the GIDL current to be more obvious when the transistor has a highly tensile CESL. From Fig. 4-27 and Fig. 4-28, the externally applied tensile stress has resulted in an increase in GIDL current of NMOS transistors and PMOS transistors, as predicted by eqn. (4.5). We observed that uniaxial tensile stress increases the subthreshold current of <110> NMOS transistor but decreases the subthreshold current of <110> PMOS transistor. The factors involved in the strain-induced change in the subthreshold current are: (i) the strain-induced change in the low-field mobility (ii) the strain-induced change in threshold voltage, (iii) strain-induced change in the gate oxide thickness. From Section 4.6, experimental data has shown that μ_{sub} is expected to increase when μ_{eff} increases. Since tensile stress is expected to increase μ_{eff} of NMOS transistor [1.14] but decrease μ_{eff} of PMOS transistor [1.15], we would expect tensile stress to increase μ_{sub} of NMOS transistor but decrease μ_{sub} of PMOS transistor. Considering the strain-induced effects on μ_{sub}, tensile stress can increase the subthreshold current of NMOS transistor but decrease the subthreshold current of PMOS transistor. From Fig. 4-27, tensile stress has reduced $V_{th,sat}$ of <110> NMOS transistor from 0.366 V to 0.320 V, as predicted by eqn. (4.5). From Fig. 4-28, tensile stress has increased $V_{th,sat}$ of <110> PMOS transistor from 0.407V to 0.449 V. This shows that the mobility term of eqn. (4.3) actually dominates over the bandgap term of eqn. (4.3). A

decrease in $V_{th,sat}$ of <110> NMOS transistor leads to an increase in the subthreshold current whereas an increase in the magnitude of $V_{th,sat}$ of <110> PMOS transistor leads to a decrease in the subthreshold current. From Fig. 4-27 and Fig. 4-28, we observed that the externally applied uniaxial tensile stress leads to a small improvement in S_{ts} of <110> NMOS transistor but a small degradation in S_{ts} of <110> PMOS transistor. As discussed in Section 4.5, this shows that the strain-induced change in z_I dominate over the strain-induced change in T_{ox}. An improvement in S_{ts} is expected to decrease the subthreshold current of NMOS transistor, whereas degradation in S_{ts} is expected to increase the subthreshold current of PMOS transistor. Since there is an experimental observation of the strain-induced I_{off} decrease in <110> PMOS transistor and strain-induced I_{off} increase in <110> NMOS transistor, the effects of the strain-induced mobility degradation and the strain-induced increase in the magnitude of $V_{th,sat}$ dominate over the effects of the strain-induced S_{ts} degradation. From Ref. [4.25], the presence of CESL-induced stress will lead to a 3 % increase in $C_{ox,inv}$ and a 21 % increase in μ_{eff}. From eqn. (3.19), the saturation drain current of the nanoscale MOS transistor is a linear function of both $C_{ox,inv}$ and μ_{eff}. Hence, we believe that the strain-induced increase in μ_{eff} will dominate over the strain-induced increase in $C_{ox,inv}$ for MOS transistors that are in strong inversion.

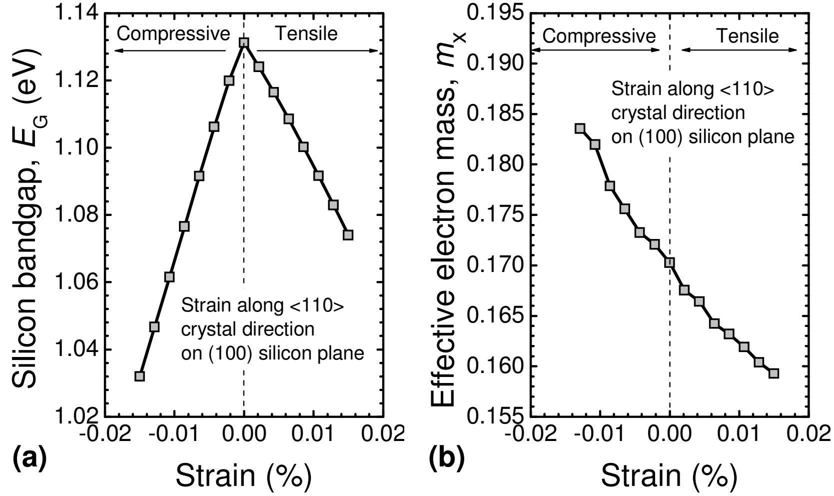


Figure 4-25 The effects of uniaxial stress on (a) the silicon bandgap and (b) the effective electron mass along the channel direction (m_x).

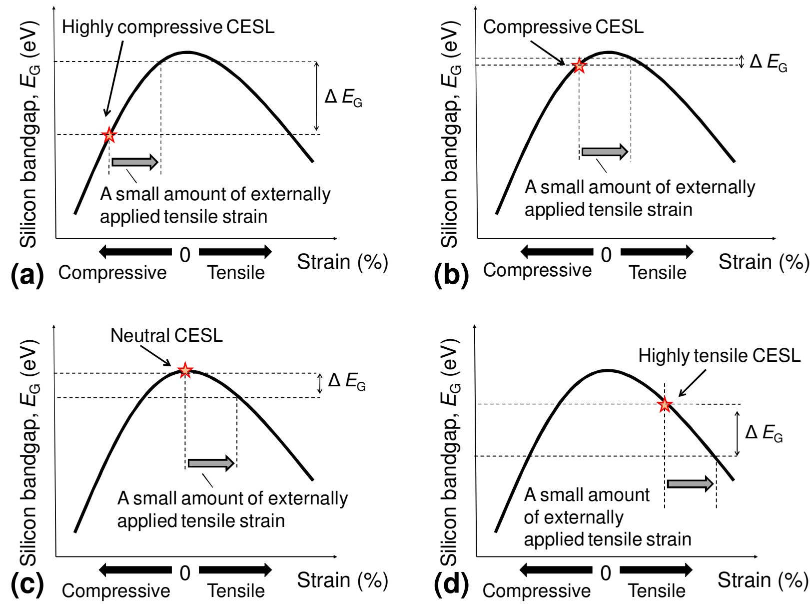
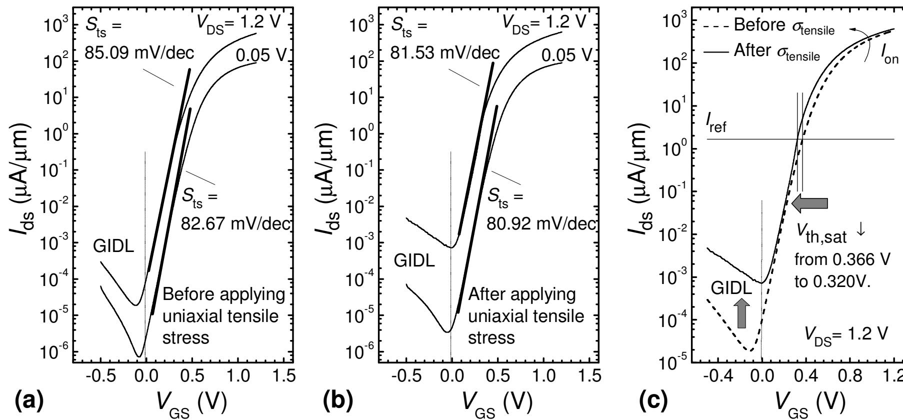


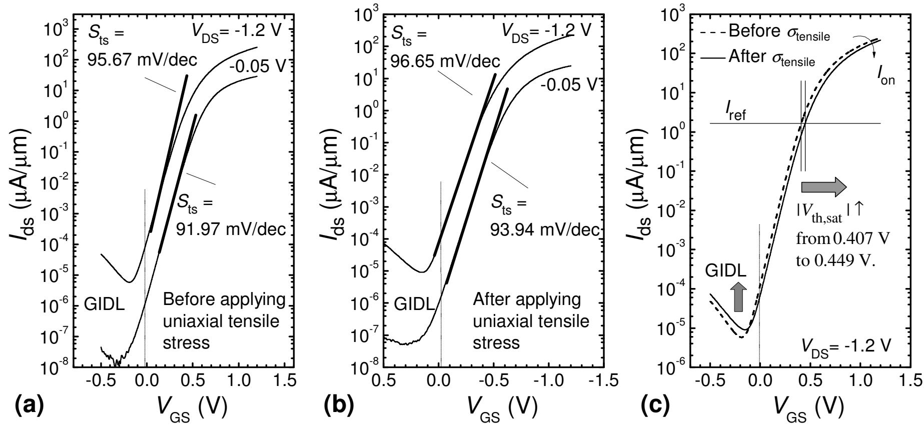
Figure 4-26 Effects of a "mechanical stress bias" in MOS transistor on the sensitivity of the silicon bandgap to a small amount of externally applied tensile stress: (a) highly compressively stressed CESL, (b) low compressive stressed CESL, (c) highly tensile stressed CESL.



	Before applying tensile stress	After applying tensile stress	Percentage change
I_{on} (measured from drain)	577 μA/μm	628 μA/μm	8.98 % ↑
I_{on} (measured from source)	576 μA/μm	628 μA/μm	9.01 % ↑
I_{off} (measured from drain)	109 pA/μm	717 pA/μm	557 % ↑
I_{off} (measured from source)	102 pA/μm	151 pA/μm	48 % ↑

(d)

Figure 4-27 Effects of the externally applied uniaxial tensile stress on <110> channel NMOS transistor fabricated on (100) Si substrate with 0.7 GPa tensile stressed CESL.



	Before applying tensile stress	After applying tensile stress	Percentage change
I_{on} (measured from drain)	252 μA/μm	217 μA/μm	-14 % ↓
I_{on} (measured from source)	252 μA/μm	218 μA/μm	-14 % ↓
I_{off} (measured from drain)	134 pA/μm	70 pA/μm	-47 % ↓
I_{off} (measured from source)	131 pA/μm	66 pA/μm	-50 % ↓

(d)

Figure 4-28 Effects of the externally applied uniaxial tensile stress on <110> channel PMOS transistor on (100) Si substrate with 0.7 GPa tensile stressed CESL.

4.7.3 Tensile stressed contact etch stop layer (CESL)

Table 4-2 summarizes the impact of the material nature of the mechanical stressor on the MOS transistor. In terms of thermal budget, the strained-Si on relaxed $Si_{1-y}Ge_y$ virtual substrate, STI, e-$Si_{1-y}Ge_y$ S/D stressor, e-Si:C S/D stressor and SMT have to undergo the high temperature annealing (around 900 °C) for S/D dopant activation. Since CESL is deposited after nickel salicidation, the maximum thermal budget for CESL should be relatively lower (around 500 °C) to avoid a phase transition from the low sheet resistivity NiSi to the high resistivity $NiSi_2$ [4.40]. Since the strain-induced change in dopant diffusion is highly dependent on temperature, CESL-induced tensile stress will be the least affected by the strain-induced dopant diffusion as compared to the other stress engineering techniques. Considering the strain-induced diffusion of the S/D extension implants, CESL-induced tensile stress will lead to a minimal change in the effective channel length (ΔL_{eff}). Furthermore, studies have shown that the parasitic S/D series [2.18, 2.19]. On the other hand, studies show that an increase in carbon concentration in the e-Si:C S/D stressor degrades R_{sd} [1.42, 2.30]. Fiorenza et al. has shown that the use of the strained-Si on relaxed $Si_{1-y}Ge_y$ virtual substrate can increase the S/D leakage current owing

to the enhanced dopant diffusion near the misfit dislocation [2.4, 2.5]. Simoen et al. reported that the implementation of e-$Si_{1-y}Ge_y$ S/D stressor can increase the junction leakage current owing to the strain relaxation through the formation of misfit and dislocations [2.20]. Osten et al. reported that the strain relaxation mechanism for e-Si:C S/D stressor is through the precipitation of carbon atom from the substitutional sites of the silicon lattice [2.33] and thus the channel stress decreases. The strain relaxation mechanism for SMT is through the recrystallization of the poly-Si gate [2.26] and thus the resistance of the poly-Si gate may be different compared to an unstrained MOS transistor. In contrast to the other stress engineering techniques, the stress relaxation of CESL does not occur at the S/D junctions or within the poly-Si gate. Although cracks may be generated in the tensile stressed CESL and delamination may occur in the compressive stressed CESL [4.41], severe cracking and delamination issues should be resolved during the development of the CESL recipe. In view of the thermal budget, the change in R_{sd}, and the strain-relaxation mechanism, CESL has the least complications compared to the other stress engineering techniques. Therefore, we studied the effects of tensile stress on the performance of NMOS transistor using tensile stressed CESL.

Table 4-2 Comparison between various stress engineering techniques.

Stress engineering techniques	Thermal budget (ΔL_{eff})	Change in R_{sd}	Strain-relaxation mechanism
CESL	Max. temp limited by nickel silicide	No	Cracking of CESL Delamination of CESL
Strained-Si on $Si_{1-y}Ge_y$ virtual substrate	High temp. S/D anneal	No	Misfit and dislocations (S/D electrical shorting)
STI	High temp. S/D anneal	No	Misfit and dislocations (S/D electrical shorting)
e-$Si_{1-y}Ge_y$ S/D stressor	High temp. S/D anneal	Decrease R_{sd}	Misfit and dislocations (High junction leakage)
e-Si:C S/D stressor	High temp. S/D anneal	Increase R_{sd}	Precipitation of carbon atoms
SMT	High temp. S/D anneal	No	Grain size of the poly-Si gate changes

4.7.3.1 Selection of MOS transistors

In order to minimize the effects of inter-wafer variation and intra-wafer variation on I_{off}, we need to select transistors whose gate lengths are less sensitive to the statistical variation in gate length (ΔL). Hueting and Heringa [4.42] have proposed an analytical model for NMOS transistors with halo implants, which shows that there is a minimum point in the log I_{off} versus L characteristics. According to basic principles of calculus, I_{off} will be the least sensitive to the statistical variation in gate length at the minimum in the log I_{off} versus L characteristics. Using MOS transistors that are fabricated by 65 nm low-power CMOS technology, we found that the actual situation is more complicated. From Fig. 4-29, there are two minimum in the log I_{off} versus L characteristics of NMOS transistors. The first minimum corresponds a $V_{th,sat}$ maximum, whereas the second minimum corresponds to a DIBL minimum. From Fig. 4-30, there is a region of nearly constant in the log I_{off} versus L characteristics of PMOS transistors. Unlike the NMOS transistors whose maximum $V_{th,sat}$ and minimum DIBL are positioned at about 0.33 μm apart, the PMOS transistors have its maximum $V_{th,sat}$ and its minimum DIBL positioned at about 0.07 μm apart. We believe that the interaction between the I_{off} minimum at maximum $V_{th,sat}$ and the I_{off} minimum at minimum DIBL.

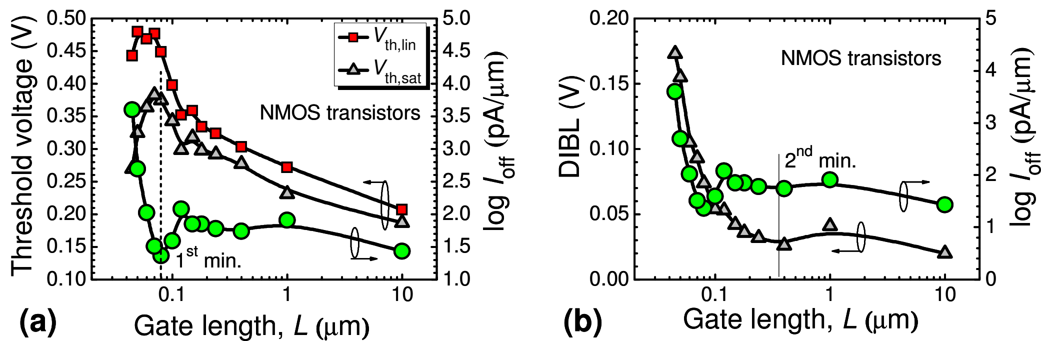


Figure 4-29 Region of nearly constant I_{off} for NMOS transistors fabricated by 65 nm CMOS technology: (a) Maximum $V_{th,sat}$, (b) Minimum DIBL.

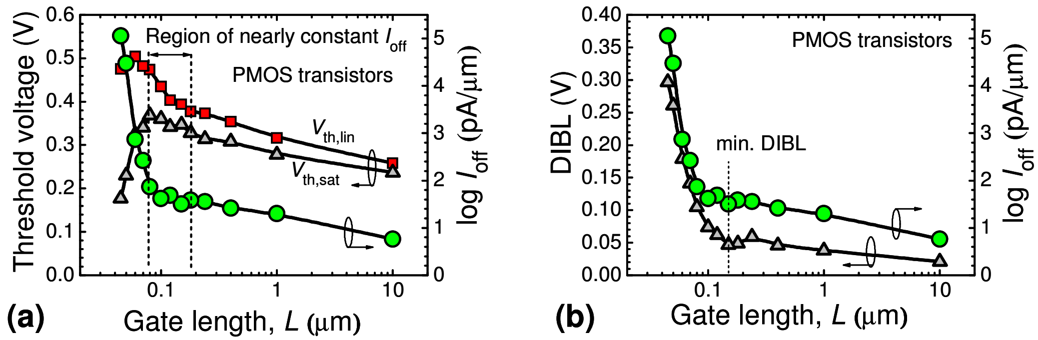
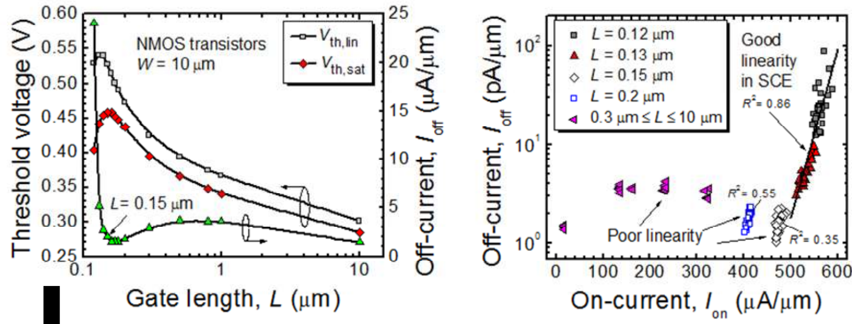


Figure 4-30 Region of nearly constant I_{off} for PMOS transistors fabricated by 65 nm CMOS technology: (a) Maximum $V_{th,sat}$, (b) Minimum DIBL

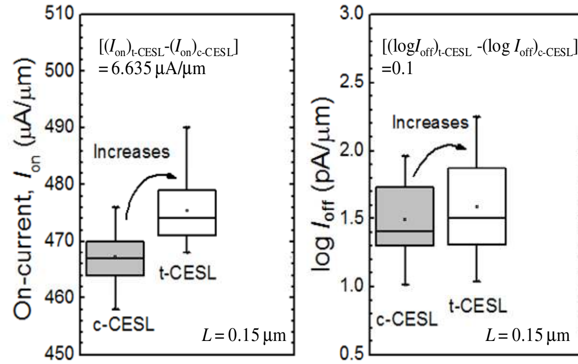
4.7.3.2 Overall performance improvement of <110> NMOS transistor

This study was performed using NMOS transistors with 125 MPa compressive stressed CESL (c-CESL) and 500 MPa tensile stressed CESL (t-CESL). Our approach is to use transistors whose I_{off} is less sensitive to ΔL to study the effects of CESL-induced tensile stress on I_{off}. Subsequently, we will use transistors that have a good linearity in $\log I_{off}$ versus I_{on} characteristics to assess the overall strain-induced I_{on} improvement. From Fig. 4-31, NMOS transistors with nominal gate length of 0.15 μm are in the transition between SCE regime and RSCE regime, and thus I_{off} of these transistors will be less sensitive to ΔL. This is evident in the poor linearity in the $\log I_{off}$ versus I_{on} characteristics of NMOS transistors with L = 0.15 μm. From Fig.4-31, $[(I_{on})_{t\text{-CESL}}-(I_{on})_{c\text{-CESL}}]$ is 6.635 μA/μm (1.4 % increase), whereas $[(\log I_{off})_{t\text{-CESL}} - (\log I_{off})_{c\text{-CESL}}]$ is 0.1 (6.67 % increases). This observation of the strain-induced I_{off} increase in NMOS transistor is consistent with our four-point bending results (see Fig. 4-27). Fig. 4-32 shows that CESL-induced tensile stress decreases $V_{th,sat}$ of the NMOS transistors, which can be explained by eqn. (4.5). At the same time, the CESL-induced tensile stress leads to a slight improvement in S_{ts}, which has been explained in Section 4.5 & Section 4.6. Since our experimental results show that CESL-induced tensile stress actually increases the subthreshold current of NMOS transistors, we believe that the effects of the strain-induced reduction in $V_{th,sat}$ and the strain-induced increase in μ_{sub} dominate over the effects of the strain-induced improvement in S_{ts}.



Step 1: Transition regime

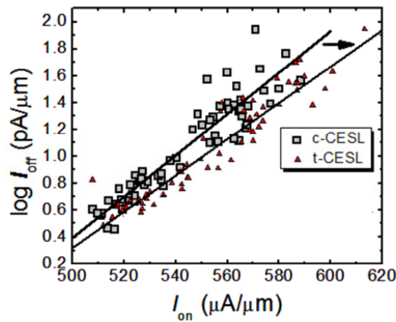
I_{off} is less sensitive to ΔL variation.



Step 2: SCE regime

Good linearity in $\log I_{off}$ versus I_{on} plot

Overall strain-induced I_{on} improvement



Method 1: Without introducing a third parameter

Best-fit line: $\log I_{off} = C_1 \times I_{on} + C_2$

Increase in I_{on} owing to a strain-induced increase in I_{off},

$I_{on_increase_by_offcurrent} = [(\log I_{off})_{t\text{-CESL}} - (\log I_{off})_{c\text{-CESL}}] / C_1$

Overall strain-induced increase in I_{on},

$I_{on_improvement} = [(I_{on})_{t\text{-CESL}} - (I_{on})_{c\text{-CESL}}] - I_{on_increase_by_offcurrent}$

Method 2: Introducing a third parameter

Best-fit line: $\log I_{off} = C_1 \times V_{th,sat} + C_2$

Best-fit line: $I_{on} = C_3 \times V_{th,sat} + C_4$

To offset the strain-induced I_{off} increase ,

$V_{th_sat_offset_off} = [(\log I_{off})_{t\text{-CESL}} - (\log I_{off})_{c\text{-CESL}}] / C_1$

To offset the strain-induced I_{on} increase ,

$V_{th_sat_offset_on} = [(I_{on})_{t\text{-CESL}} - (I_{on})_{c\text{-CESL}}] / C_3$

Overall strain-induced increase in I_{on},

$I_{on_improvement} = C_3(V_{th_sat_offset_on} - V_{th_sat_offset_off})$

Figure 4-31 Overall strain-induced I_{on} improvement in <110> NMOS transistors.

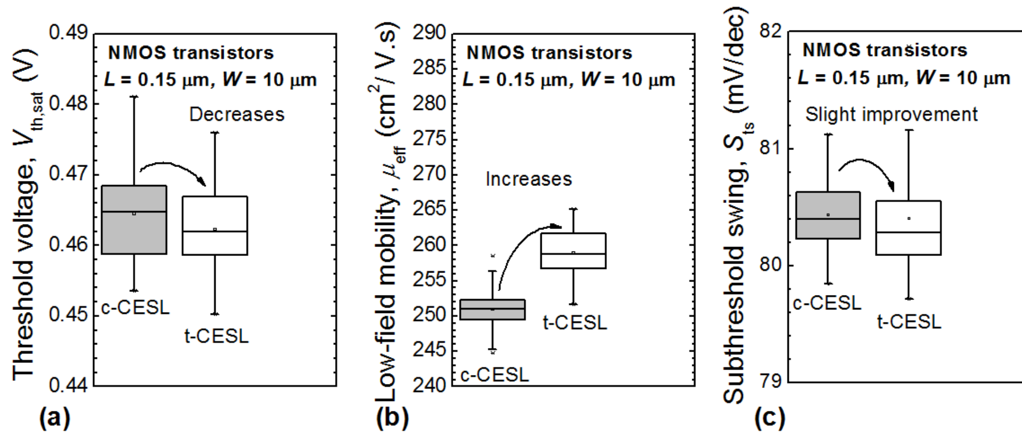


Figure 4-32 The effects of CESL-induced tensile stress on: (a) $V_{th,sat}$, (b) μ_{eff}, (c) S_{ts} of NMOS transistors in the transition between SCE regime and RSCE regime.

To assess the overall strain-induced I_{on} improvement in NMOS transistors, we need to use transistors with a good linearity in the logI_{off} versus I_{on} characteristics. This corresponds to transistors in SCE regime (with L = 0.12 µm and L = 0.13 µm). Fig. 4-31 shows the differences between the two methods that can be used to ascertain the overall strain-induced I_{on} improvement. Unlike Method 1, Method 2 introduces $V_{th,sat}$ as a a third parameter in the logI_{off} versus I_{on} plot. In this manner, Method 2 can address the physics behind how the bigger percentage increase in the strain-induced logI_{off} compared to the percentage increase in strain-induced I_{on} can lead to an overall strain-induced I_{on} improvement. The details of Method 1 and Method 2 can be found in Table 4-3 and Table 4-4 respectively. For Method 1, $I_{on_increase_by_offcurrent}$ refers to the corresponding increase in I_{on} owing to the strain-induced increase in logI_{off}. For Method 2, $V_{th_sat_offset_on}$ and $V_{th_sat_offset_off}$ refer to the $V_{th,sat}$ increase to offset the strain-induced I_{on} increase and the strain-induced logI_{off} increase, respectively. From Table 4-4 and Fig. 4-33, we observed that $V_{th_sat_offset_on}$ is bigger than $V_{th_sat_offset_off}$ but stress engineering actually leads to a bigger percentage increase in logI_{off} as compared to the percentage increase in I_{on}. This indicates that logI_{off} is more sensitive to the change in $V_{th,sat}$ compared to I_{on} such that the strain-induced increase in logI_{off} can be removed by a slight change in $V_{th,sat}$ without much effect on I_{on}. Therefore, stress engineering can lead to an overall strain-induced I_{on} improvement in <110> NMOS transistors on (100) p-Si substrate even though stress engineering actually leads to a bigger percentage increase in logI_{off} as compared to the percentage increase in I_{on}.

Table 4-3 Method 1 to ascertain the overall strain-induced I_{on} improvement of <110> NMOS transistors.

	t-CESL	c-CESL
Best-fit line $\log I_{off} = C_1 \times I_{on} + C_2$	$\log I_{off} = 0.0135 I_{on} - 6.4307$	$\log I_{off} = 0.0154\ I_{on} - 7.3079$
Effects of strain-induced increase in off-current, $I_{on_increase_by_offcurrent} = [(\log I_{off})_{t\text{-}CESL} - (\log I_{off})_{c\text{-}CESL}] / C_1$	1.926 μA/μm	1.688 μA/μm
$I_{on_improvement} = [(I_{on})_{t\text{-}CESL} - (I_{on})_{c\text{-}CESL}] - I_{on_increase_by_offcurrent}$	4.709 μA/μm	4.947 μA/μm

Table 4-4 Method 2 to ascertain the overall strain-induced I_{on} improvement of <110> NMOS transistor using $V_{th,sat}$ as a third parameter in the $\log I_{off}$ versus I_{on} plot.

	t-CESL	c-CESL
Best-fit line $\log I_{off} = C_1 \times V_{th,sat} + C_2$	$\log I_{off} = -16.941 \times V_{th,sat} + 8.1941$	$\log I_{off} = -18.509 \times V_{th,sat} + 8.8968$
Best-fit line $I_{on} = C_3 \times V_{th,sat} + C_4$	$I_{on} = -1179.7\ V_{th,sat} + 1050.7$	$I_{on} = -1118.4 V_{th,sat} + 1017.8$
$V_{th_sat_offset_off} = [(\log I_{off})_{t\text{-}CESL} - (\log I_{off})_{c\text{-}CESL}] / C_1$	1.53×10^{-3} V	1.41×10^{-3} V
$V_{th_sat_offset_on} = [(I_{on})_{t\text{-}CESL} - (I_{on})_{c\text{-}CESL}] / C_3$	5.62×10^{-3} V	5.93×10^{-3} V
$I_{on_improvement} = C_1(V_{th_sat_offset_on} - V_{th_sat_offset_off})$	4.83 μA/μm	5.06 μA/μm

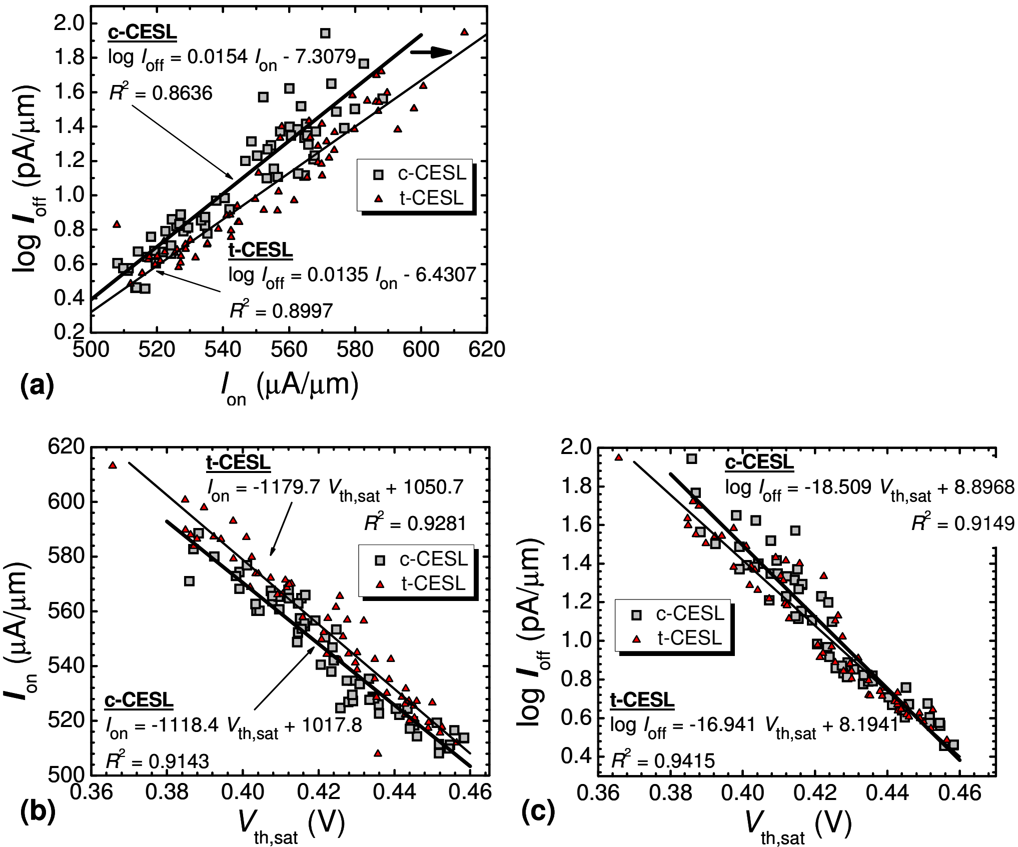


Figure 4-33 Overall performance improvement in <110> NMOS transistors by the CESL-induced tensile stress: (a) log I_{off} versus I_{on} characteristics, (b) I_{on} versus $V_{th,sat}$ characteristics, (c) log I_{off} versus $V_{th,sat}$ characteristics.

4.7.3.3 Overall performance improvement of <100> NMOS transistor

Similar analysis can be done for <100> NMOS transistors on (100)Si. From Fig. 4-34, NMOS transistors with nominal gate length of 70 nm are in the transition between SCE regime and RSCE regime. Based on 30 transistors, $[(I_{on})_{t\text{-CESL}}-(I_{on})_{n\text{-CESL}}]$ is 115.34 μA/μm (22% increase), whereas $[(\log I_{off})_{t\text{-CESL}} -(\log I_{off})_{n\text{-CESL}}]$ is 0.71 (355% increase). In order to assess the overall strain-induced I_{on} improvement, we used transistors in the SCE regime (L = 32 nm) because of the good linearity in the logI_{off} versus I_{on} plot. By introducing $V_{th,sat}$ as the third parameter, we observed that $V_{th_sat_offset_on}$ is bigger than $V_{th_sat_offset_off}$ but stress engineering actually leads to a bigger percentage increase in logI_{off} as compared to the percentage increase in I_{on}. The details can be found in Table 4-5 and Fig. 4-35. This indicates that logI_{off} is more sensitive to the change in $V_{th,sat}$ compared to I_{on} such that the strain-induced increase in logI_{off} can be removed by a slight change in $V_{th,sat}$ without much effect on I_{on}. Therefore, stress engineering can lead to an overall strain-

induced I_{on} improvement in <100> NMOS transistors on (100) p-Si substrate stress engineering actually leads to a bigger percentage increase in $\log I_{off}$ as compared to the percentage increase in I_{on}.

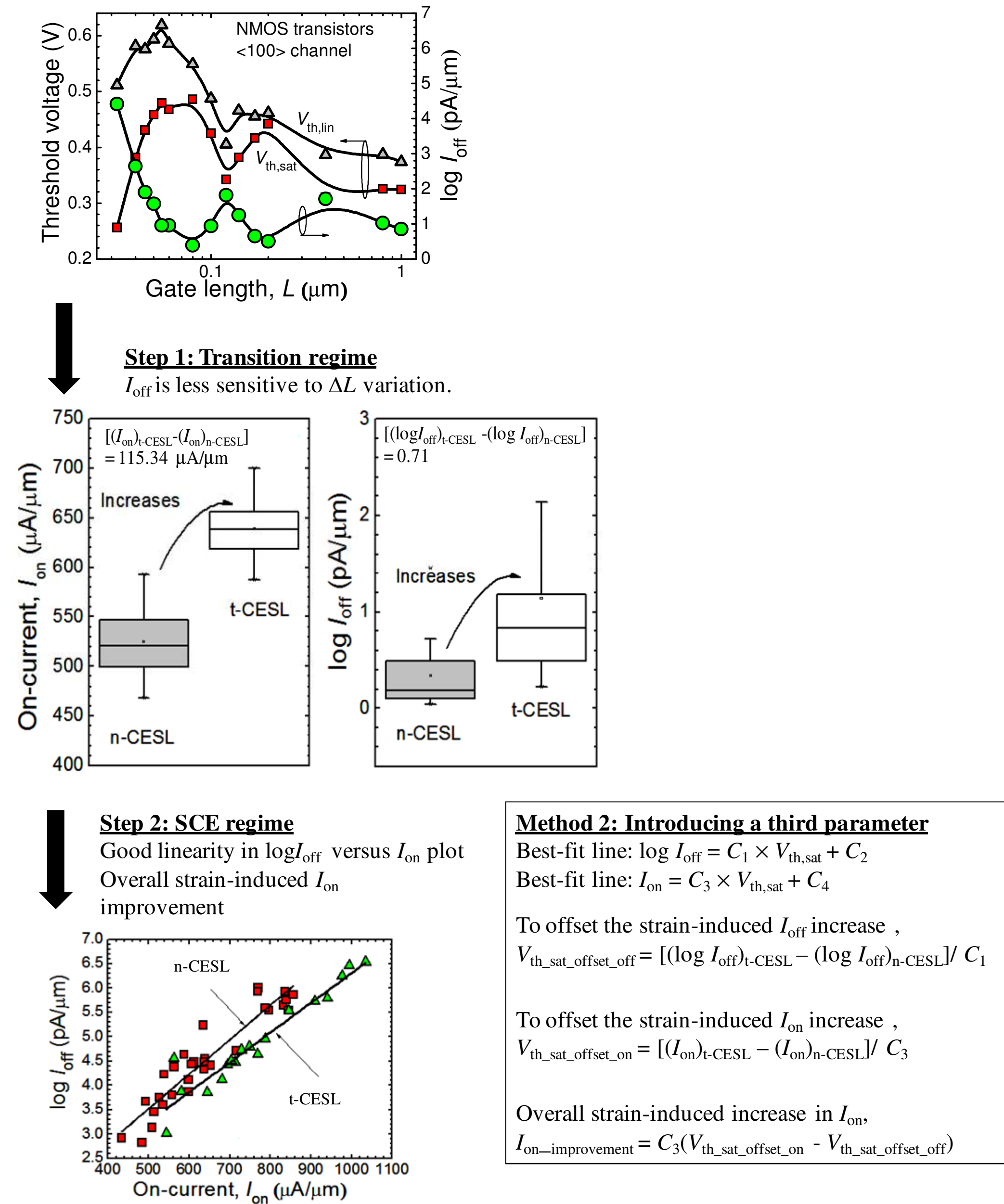


Figure 4-34 Overall strain-induced I_{on} improvement in <100> NMOS transistors.

Table 4-5 Overall performance improvement for <100> channel NMOS transistor by stress engineering when we consider $V_{th,sat}$ as the third parameter in the log I_{off} versus I_{on} characteristics.

	t-CESL	n-CESL
Best-fit line $I_{on} = C_9 \times V_{th,sat} + C_{10}$	I_{on} = -1352.5 $V_{th,sat}$ + 1037.9	I_{on} = -1259 $V_{th,sat}$ + 962.6
Best-fit line log $I_{off} = C_{11} \times V_{th,sat} + C_{12}$	log I_{off} = -8.922 $V_{th,sat}$ + 6.669	log I_{off} = -9.85 $V_{th,sat}$ + 7.03
$V_{th_sat_offset_on}$ = $[(I_{on})_{t\text{-}CESL} - (I_{on})_{n\text{-}CESL}]/ C_9$	8.53×10^{-2} V	9.16×10^{-2} V
$V_{th_sat_offset_off}$ = $[(\log I_{off})_{t\text{-}CESL} - (\log I_{off})_{n\text{-}CESL}]/ C_{11}$	7.02×10^{-2} V	6.355×10^{-2} V
$I_{on_improvement}$ = $C_9(V_{th_sat_offset_on} - V_{th_sat_offset_off})$	20.4 μA/μm	35.31 μA/μm

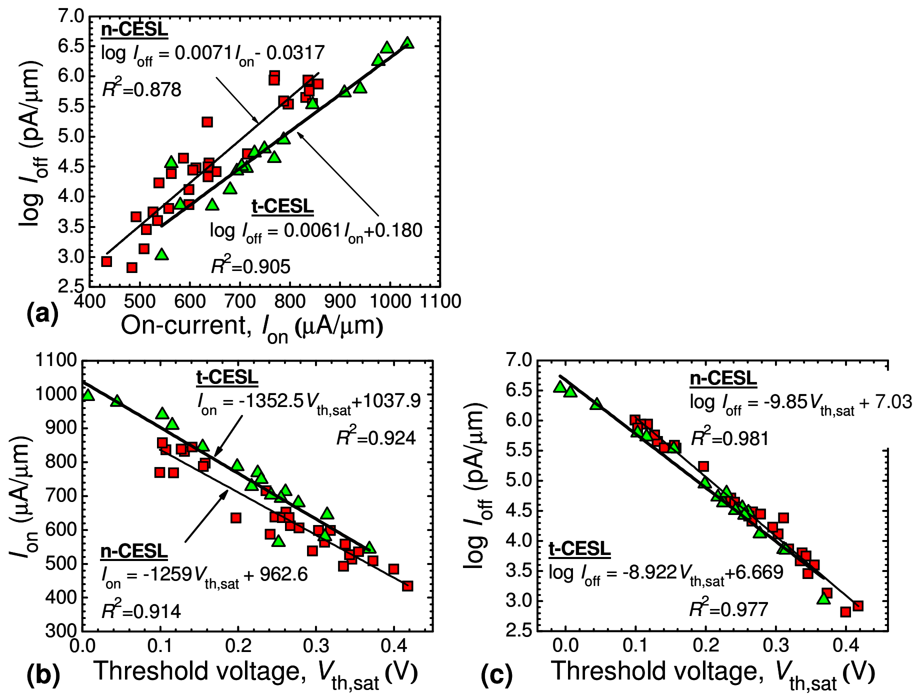


Figure 4-35 Overall performance improvement in <100> NMOS transistors by CESL-induced tensile stress: (a) the log I_{off} versus I_{on} characteristics, (b) the I_{on} versus $V_{th,sat}$ characteristics, and (c) the log I_{off} versus $V_{th,sat}$ characteristics.

5. Impact of the change in channel orientation on performance

From Fig. 5-1, the switching from <110> to <100> channel orientation can lead to an increase in I_{on} of PMOS transistor fabricated on (100) Si owing to a hole mobility enhancement [1.32]. Historically, wafer dicing was done by sawing through the scribe lines and hence the wafer notch needs to be formed along the easily cleaved crystal direction, which is <110> direction. However, laser system is now readily available for wafer dicing, and thus it is no longer a technological requirement to use <110> direction. Orienting the channel direction along <100> direction is a simple process change in the starting wafer, as shown in Fig. 5-2. Since the poly-Si gate of the transistor is aligned along the wafer notch, the channel direction of the conventional (100) wafer is <110> while the channel direction of 45° rotated (100) wafer is <100>.

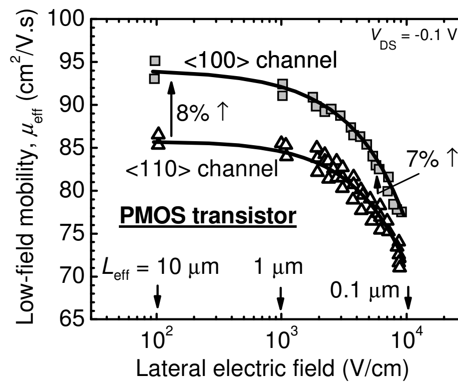


Figure 5-1 Switching in channel orientation increases hole mobility[1.32].

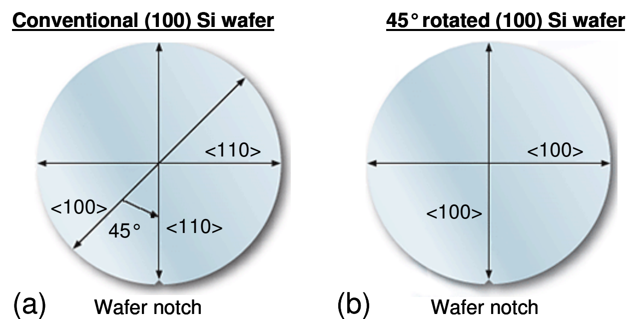


Figure 5-2 Schematics of the (a) conventional (100)Si, (b) 45° rotated (100) Si

5.1 Impact of channel orientation on hole mobility

From Fig. 5-3, the LH band is relatively isotropic on (100) plane but HH band is anisotropic on (100) plane. We observed that the curvature of the ***E-k*** diagram of HH band is bigger for <100> direction compared to <110> direction. Based on eqn. (2.10), <100> direction would have a smaller effective conductivity mass (m^*), and thus a bigger low-field hole mobility. From eqn. (3.19), an increase in the hole mobility will lead to an increase in the I_{on} of the nanoscale PMOS transistor. Hence, the switching from the conventional <110> channel orientation to <100> channel orientation will increase I_{on} of PMOS transistors fabricated on (100) surface-oriented silicon wafers, as shown in Fig. 5-4.

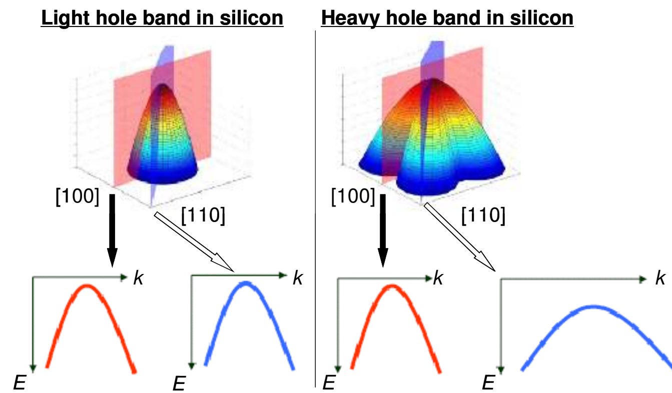


Figure 5-3 Effects of channel orientation on the effective hole mass of (100)Si[1.34].

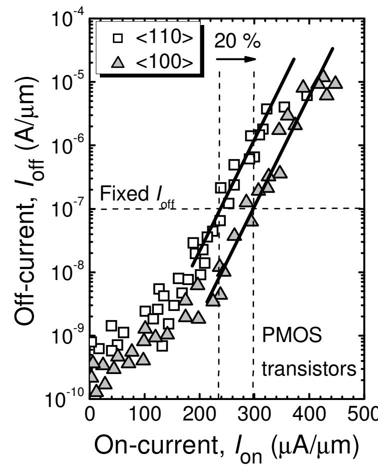


Figure 5-4 Effects of channel orientation increases I_{on} of PMOS transistors on (100)Si [1.33].

5.2 Limited studies on the impact of channel orientation on electron mobility

Unlike hole mobility, the effects of the switching from the conventional <110> channel orientation to <100> channel orientation on the electron mobility is less studied. Fig. 5-5 shows the calculations of the effective conductivity mass (m^*) for <110> NMOS transistor and <100> NMOS transistor on (100)Si [1.34]. This has been attributed to the axial symmetry of the unstrained conduction band valleys on (100) plane [5.1]. However, Skotnicki did not explain how he arrived at the effective mass of the in-plane conduction band valleys for <110> channel [1.34]. Based on eqn. (2.11), m^* of NMOS transistor should be unchanged after channel orientation if the momentum relaxation time (τ_m) is independent of channel direction.

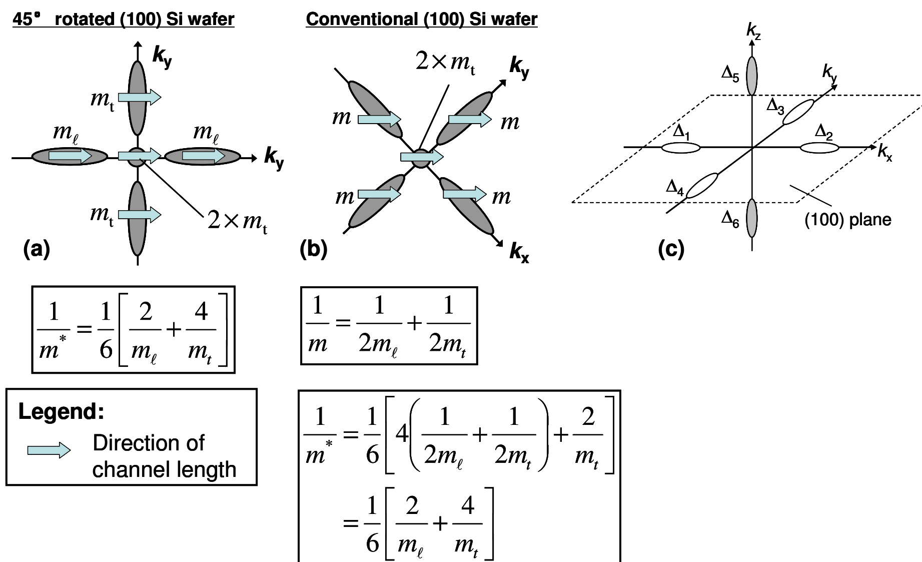


Figure 5-5 Calculations of the effective conductivity electron mass (m^*): (a) 45° rotated (100)Si, (b) conventional (100) Si [1.34], (c) conduction band valleys on (100)Si plane.

The calculations of m^* for NMOS transistors can be understood as follows. The generalized effective mass equation can be defined by the three orthogonal coordinate systems: (i) device coordinate system (DCS), (ii) crystal coordinate system (CCS) and (iii) ellipsoid coordinate system (ECS) [5.2]. For silicon, the ***k***-space close to the conduction band minimum can be described by an ellipsoid with m_ℓ along its major axis and m_t along its two minor axes. The constant energy ellipsoid can be expressed as in ECS,

$$E(k_{//},k_{\perp 1},k_{\perp 2}) = \frac{\hbar^2 k_{//}^{\ 2}}{2m_\ell} + \frac{\hbar^2 k_{\perp 1}^{\ 2}}{2m_t} + \frac{\hbar^2 k_{\perp 2}^{\ 2}}{2m_t} \tag{5.1}$$

where $k_{//}$ is the direction along the major axis of the conduction band ellipsoid. $k_{\perp 1}$ and $k_{\perp 2}$ are the directions along the minor axes of the conduction band ellipsoid. In compact vector notation, eqn. (5.1) can be written as [5.2],

$$E(k_{//},k_{\perp 1},k_{\perp 2}) = \frac{\hbar^2}{2} \vec{k}_{\mathrm{E}}^{\ \mathrm{T}} \left(M_{\mathrm{E}}^{\ -1}\right) \vec{k}_{\mathrm{E}} \tag{5.2}$$

where $\vec{k}_{\mathrm{E}} = (k_{//},k_{\perp 1},k_{\perp 2})^{\mathrm{T}}$.The inverse effective mass tensor of the conduction band valley (M_{E}^{-1}) in ECS is a 3×3 diagonal matrix with diagonal elements of m_ℓ, m_t and m_t [5.3]. From Fig. 2-16, M_{E}^{-1} will be different for each conduction band ellipsoid.

For Δ_1 and Δ_2 (bulk silicon),

$$\left(M_{\mathrm{E}}^{\ -1}\right) = \begin{bmatrix} \frac{1}{m_\ell} & 0 & 0 \\ 0 & \frac{1}{m_t} & 0 \\ 0 & 0 & \frac{1}{m_t} \end{bmatrix} \tag{5.3 a}$$

For Δ_3 and Δ_4 (bulk silicon),

$$\left(M_{\mathrm{E}}^{\ -1}\right) = \begin{bmatrix} \frac{1}{m_t} & 0 & 0 \\ 0 & \frac{1}{m_\ell} & 0 \\ 0 & 0 & \frac{1}{m_t} \end{bmatrix} \tag{5.3 b}$$

For Δ_5 and Δ_6 (bulk silicon), $$\left(M_{\mathrm{E}}^{-1}\right)=\begin{bmatrix} \frac{1}{m_t} & 0 & 0 \\ 0 & \frac{1}{m_t} & 0 \\ 0 & 0 & \frac{1}{m_\ell} \end{bmatrix} \quad (5.3\text{ c})$$

To calculate m^* of NMOS transistor, we need to express the constant energy ellipsoid in DCS. First of all, the rotation matrix $\Re_{\mathrm{C}\leftarrow\mathrm{D}}$ will transforms an arbitrary vector $\vec{k}_{\mathrm{D}}=\left(k_{\mathrm{x}},k_{\mathrm{y}},k_{z}\right)^{\mathrm{T}}$ in DCS to a vector $\vec{k}_{\mathrm{C}}$ in CCS. Subsequently, the rotation matrix $\Re_{\mathrm{E}\leftarrow\mathrm{C}}$ will transform vector $\vec{k}_{\mathrm{C}}$ to $\vec{k}_{\mathrm{E}}$.

$$\vec{k}_{\mathrm{E}}=\Re_{\mathrm{E}\leftarrow\mathrm{C}}\left(\Re_{\mathrm{C}\leftarrow\mathrm{D}}\vec{k}_{\mathrm{D}}\right) \quad (5.4)$$

For (100) Si, the principal axes of the six-fold degenerate conduction band ellipsoids are along the crystal coordinate axes, and thus $\Re_{\mathrm{E}\leftarrow\mathrm{C}}$ will be an identity matrix. The columns of $\Re_{\mathrm{C}\leftarrow\mathrm{D}}$ are the components of the unit vectors along *L*, *W*, and the quantum confinement direction. Considering <100> NMOS transistor on (100) Si, $\Re_{\mathrm{C}\leftarrow\mathrm{D}}$ is an identity matrix [5.2]. Substituting eqn.(5.4) into eqn. (5.2),

$$E\left(k_x,k_y,k_z\right)=\frac{\hbar^2}{2}\vec{k}_{\mathrm{D}}^{\mathrm{T}}\left(M_{\mathrm{D}}^{-1}\right)\vec{k}_{\mathrm{D}} \quad (5.5)$$

where k_{x}, k_{y} and k_{z} are the direction along *L*, *W*, and the quantum confinement direction, respectively. The inverse effective mass (M_{D}^{-1}) in the DCS is given by [5.2],

$$\left(M_{\mathrm{D}}^{-1}\right)=\Re_{\mathrm{E}\leftarrow\mathrm{D}}^{\mathrm{T}}\left(M_{\mathrm{E}}^{-1}\right)\Re_{\mathrm{E}\leftarrow\mathrm{D}} \quad (5.6)$$

For unstrained silicon, the six conduction band valleys are degenerate, and thus there is an equal probability that an electron will reside in one of the six conduction band valleys. By considering the average conductivity mass of the six conduction band valleys and using eqn. (2.11),

$$\frac{1}{m^*}=\frac{1}{6}\left[\frac{1}{\hbar^2}\frac{d^2}{dk_x^{\,2}}\sum_i E_{\Delta i}\right] \quad (5.7)$$

where $E_{\Delta i}$ is the constant energy surface of the ith conduction band valley where i = 1 to 6 . Solving eqn. (5.5) and eqn. (5.7), we arrive at $\frac{1}{m^*} = \frac{1}{3}\left(\frac{1}{m_\ell} + \frac{2}{m_t}\right)$ for <100> NMOS transistor on (100) Si.

5.3 Experimental evidence of an increase in L_{eff} by channel orientation

From Fig. 5-6, the change in channel orientation results in a reduction in I_{g_off} of PMOS transistor and NMOS transistor [1.33]. I_{g_off} is the gate current when the body, source and gate terminals are grounded while the drain terminal is tied to V_{DD}. Considering the same S/D implantation dose, the use of a thin-offset spacer is expected to increase the S/D-to-gate overlap and thus I_{g_off} of MOS transistors with thin offset spacer will be bigger than that of reference. Considering the same offset spacer but vary the S/D implantation dose, I_{g_off} of MOS transistor with a bigger implant dose will be bigger than that of reference. In the other words, the value of I_{g_off} in ascending order is SDE2 > SDE1 > Ref. This shows that a reduction in I_{g_off} is indicative of a smaller gate-to-drain overlap length [1.33]. In the other words, <100> MOS transistor has a bigger L_{eff} as compared to <110> MOS transistor.

From Fig. 5-6, we observed that I_{g_off} of NMOS transistor with thin offset spacer is smaller than that of SCE1 and SCE2. This is contradictory to that of PMOS transistor. This can be understood as follows. Since arsenic has a bigger atomic mass than boron, the arsenic implantation damage is expected to be bigger than boron implantation damage. If the S/D annealing is not sufficient to remove the crystalline defects in silicon caused by arsenic implantation, transient enhanced diffusion of arsenic can occur [5.4] and thus lead an even bigger increase in S/D-to-gate overlap length. This can account for the bigger I_{g_off} of NMOS transistor with SDE1 and SDE2 compared to that of thin offset spacer.

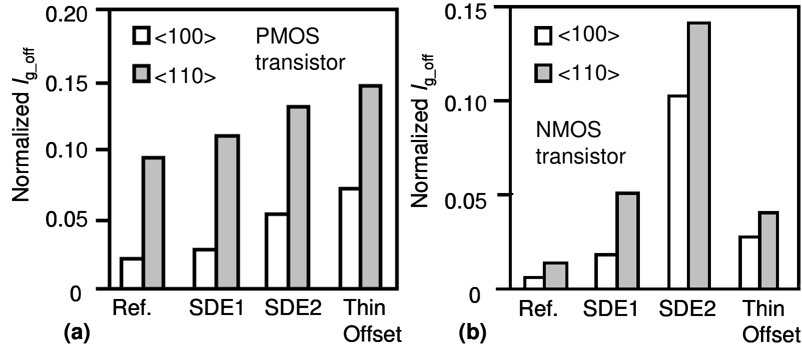


Figure 5-6 The effects of channel orientation on I_{g_off} of: (a) PMOS transistors, (b) NMOS transistors on (100) Si [1.33]. S/D extension implant dose: Ref = Thin Offset = SDE1 < SDE2. Thickness of the gate offset spacer: Ref = SDE1 = SDE2 > Thin Offset.

5.4 Mechanism behind the increase in L_{eff} by channel orientation

The existing work did not provide any reference or experimental results for their claim that the boron diffusion coefficient is smaller in <100> channel orientation as compared to <110> channel orientation [1.32, 1.33]. Hence, this work will study the mechanism behind the increase in L_{eff} owing to the change in channel orientation. For PMOS transistor, the S/D extension implant can be either boron or BF_2 , and arsenic is used as the halo implant. For NMOS transistor, arsenic is used as the S/D extension implant and boron is used as the halo implant. Since the dose of the S/D extension implant is typically one or two orders bigger than the dose of the halo implant [5.5], the impact of channel orientation on the S/D extension implants is likely to dominate over that of halo implants.

5.4.1 Impact of channel orientation on diffusion

Although boron can interact with silicon interstitials to form big boron interstitials clusters (BICs) during high temperature diffusion after ion implantation [5.6], BICs cannot go through <100> channel and <110> channel easily because BICs are much bigger than boron. Hence, it is sufficient to focus on boron diffusion. In an oxidizing ambient, the boron diffusion coefficient and the boron diffusion depth tend to be larger for (100) Si compared to (110) Si [5.7, 5.8]. Aleksandrov and Afonin used the difference in the interface-state density (ΔD_{it}) for gate oxide on various Si planes to explain the orientation dependent boron diffusion in an oxidizing ambient [5.9]. This can be understood as follows. It is well-known that ΔD_{it} at SiO_2/Si interface acts as a sink for the excess silicon interstitials [5.10]. Since ΔD_{it} of (111)Si > ΔD_{it} of (110)Si > ΔD_{it} of (100)Si [5.11], the

effective rate of generation for the intrinsic interstitials is highest for (100)Si and the lowest for (111)Si [5.9]. Hence, the supersaturation by intrinsic interstitials (a_{int}) is the biggest for (100)Si and the smallest for (111)Si [5.9]. Since a decrease in a_{int} will lead to an increase in boron diffusion [5.12], boron diffusion into (111)Si > (110)Si > (100)Si [5.9]. For the state-of-the-art PMOS transistors, transient enhanced diffusion (TED) of boron and boron enhanced diffusion (BED) are also important. Since OED, TED and BED are related phenomena involving boron diffusion enhanced by silicon interstitials [5.13, 5.14], the crystal orientation dependence trend for OED [5.7, 5.8] can be applied to TED and BED. Hence, the boron diffusion studies reported from the 1970s up to now [5.7-5.14] show that the boron diffusion is faster for (100)Si compared to (110)Si. For silicon, [100] direction is perpendicular to (100) plane, whereas [110] direction is perpendicular to (110) plane, as shown in Fig. 5-7 [5.15]. In the other words, boron diffusion is faster along <100> crystal direction compared to <110> crystal direction. Therefore, the increase in L_{eff} of <100> channel PMOS transistors on (100) Si is unlikely to be caused by a change in boron diffusion coefficient of p^+ S/D extension implants.

Since arsenic TED and boron TED are mediated by silicon interstitials [5.16], we can expect arsenic diffusion of (100)Si > arsenic diffusion of (110)Si > arsenic diffusion of (111)Si. Experimental studies show that there is negligible difference in arsenic diffusion for (100)Si and (111)Si [5.17]. Hence, there will be minimal change in the arsenic diffusion coefficient after the change in channel orientation. Therefore, the increase in L_{eff} of NMOS transistor is unlikely to be caused by the change in arsenic diffusion coefficient of the n^+ S/D extension implants.

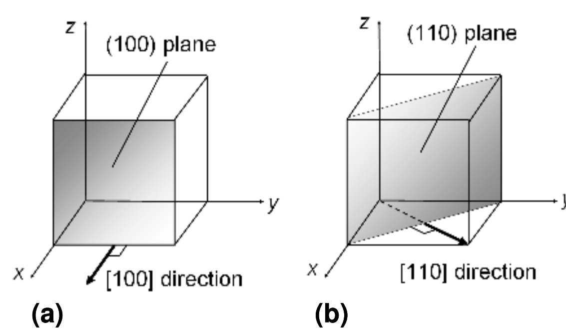


Figure 5-7 For cubic crystal structure such as silicon, (a) [100] direction is perpendicular to (100) plane, (b) [110] direction is perpendicular to (110) plane [5.15].

5.4.2 Impact of channel orientation on ion channeling

It is well-known that the ion motion into a crystalline solid target can become constrained into the relatively open spaces in the crystal between the adjacent rows of atoms along the low Miller index crystallographic directions. This is known as channelling. From Fig. 5-8, the types of channelling can be catergorized according to the critical angle [5.18]. The critical angle (θ_c) is defined as the maximum angle at which an ion can approach the "potential wall" of open space before it is reflected back into the channel. When the angle of incidence of the implanted ion (θ_{impl}) is smaller than θ_c, the ions will experience direct channelling. When $\theta_{impl} > \theta_c$, the ions will experience random movement into the silicon lattice. However, the random nuclear collisions of the ion may result in a trajectory that is aligned with a major crystallographic axis. This is known as the indirect channelling.

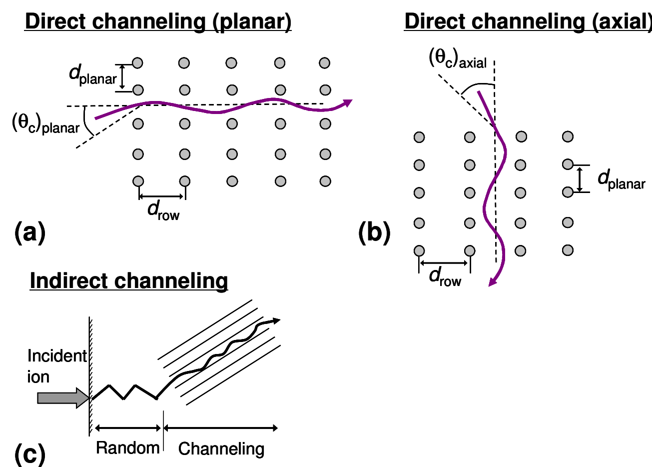


Figure 5-8 Various types of ion channeling : (a) planar channeling, (b) axial channeling, (c) indirect channeling [5.18].

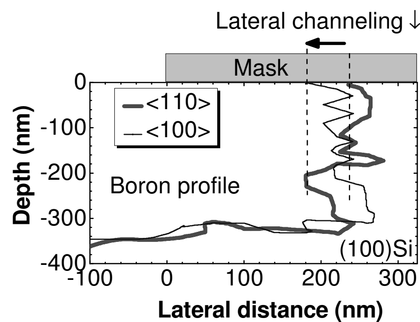


Figure 5-9 Monte Carlo simulation of the two-dimensional implanted boron distribution into (100)Si at a mask edge [5.19].

Since <100> transistors and <110> transistors are fabricated on (100)Si, they will have the same probability of experiencing the axial channelling and the planar channelling. Here, we are referring to a special case of indirect channeling whereby the S/D extension implants can channel beneath the poly-Si gate and thus affects L_{eff}. We called this phenomenon as the lateral ion channelling. From Fig. 5-9, Monte Carlo simulation of the two-dimensional implanted boron distribution into (100) Si at a mask edge shows that the lateral channelling for <100> channel is smaller than that of <110> channel [5.19]. As there is no equation for lateral ion channelling, we have to modify the equations for planar channelling and axial channelling. The equation of the critical angle for planar channelling can be expressed as follows [5.18].

$$(\theta_{\mathrm{c}})_{\mathrm{planar}} = K_1 \sqrt{\frac{Z_{\mathrm{ion}} Z_{\mathrm{Si}} q^2 N d_{\mathrm{planar}}}{E_{\mathrm{ion}}}} \tag{5.8}$$

where K_1 is a constant. Z_{ion} and Z_{Si} are the atomic numbers of the incident ions and silicon, respectively. E_{ion} is the incident ion energy. N is the density of silicon. d_{planar} is the interplanar spacing that defines a planar channel for ion channeling. From Fig. 5-10, $(\theta_{\mathrm{c}})_{\mathrm{planar}}$ is bigger for (110)Si compared to (100)Si [5.22]. For lateral ion channelling, we need to define the open space in the silicon lattice ($A_{\mathrm{open_space}}$). From Fig. 5-11 and Table 5-1, $A_{\mathrm{open_space}}$ along <110> direction is bigger for <110> channel as compared to <100> channel [5.20, 5.21]. The effective radius of the open space in the silicon crystal lattice (r_{eff}) can be expressed as follows,

$$r_{\mathrm{eff}} = \left(A_{\mathrm{open_space}} / \pi\right)^{1/2} \tag{5.9}$$

Hence, the equation of the critical angle for planar channelling can be expressed as follows,

$$(\theta_{\mathrm{c}})_{\mathrm{lateral}} = K_2 \sqrt{\frac{2 Z_{\mathrm{ion}} Z_{\mathrm{Si}} q^2 N r_{\mathrm{eff}}}{E_{\mathrm{ion}}}} \tag{5.10}$$

Table 5-1 The area of open space ($A_{\mathrm{open_space}}$) in the silicon crystal lattice for various surface orientation and crystal directions [5.20].

Surface orientation/ crystal direction	Area of the open space, $A_{\mathrm{open_space}}$ (nm^2)
(110)/ <110>	0.1053
(100)/<100>	0.0369
(111)/ <111>	0.0213

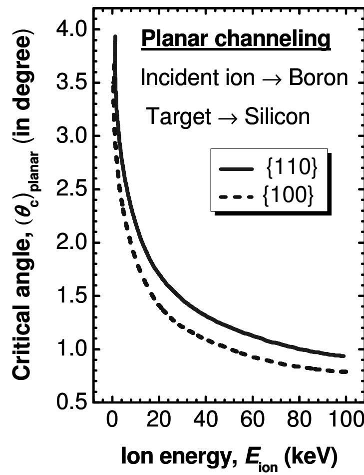


Figure 5-10 Critical angle for planar channeling along {110} plane and {100} plane as a function of the incident ion energy [5.22].

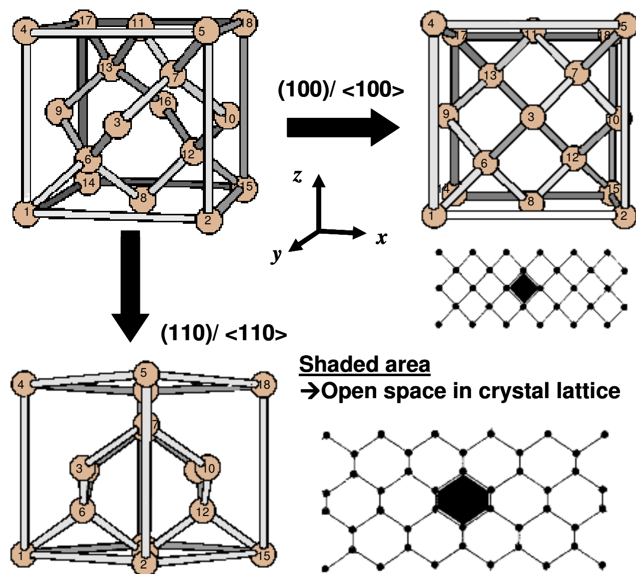


Figure 5-11 Arrangements of the silicon atoms in (100) Si and (110) Si that are viewed along the direction normal to the plane [5.20, 5.21].

5.5 Results and Discussion

5.5.1 Validity of the six degenerate ellipsoid model

This work will show that m^* of NMOS transistor on (100)Si is independent of channel orientation even though the actual conduction band minimum at the SiO_2/Si interface may be slightly different from the six degenerate ellipsoid model. First, we will assume the validity of the six degenerate ellipsoid model and then calculate m^* of (100) Si with an arbitrary channel orientation. From Fig. 5-12, $\Re_{C\leftarrow D}$ for (100)Si with a generalized channel orientation can be expressed as follows,

$$\Re_{C\leftarrow D} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.11}$$

where $\theta = 0$ corresponds to <100> channel orientation. Since $\Re_{E\leftarrow C}$ is an identity matrix for (100)Si, $\Re_{E\leftarrow D} = \Re_{C\leftarrow D}$. By substituting eqn. (5.3) into eqn.(5.6), we can obtain M_D^{-1}.

For Δ_1 and Δ_2 (bulk silicon),

$$M_D^{-1} = \begin{bmatrix} \left(\dfrac{\cos^2\theta}{m_\ell} + \dfrac{\sin^2\theta}{m_t}\right) & \left(\dfrac{\cos\theta\sin\theta}{m_\ell} - \dfrac{\cos\theta\sin\theta}{m_t}\right) & 0 \\ \left(\dfrac{\cos\theta\sin\theta}{m_\ell} - \dfrac{\cos\theta\sin\theta}{m_t}\right) & \left(\dfrac{\sin^2\theta}{m_\ell} + \dfrac{\cos^2\theta}{m_t}\right) & 0 \\ 0 & 0 & \dfrac{1}{m_t} \end{bmatrix} \tag{5.12}$$

For Δ_3 and Δ_4 (bulk silicon),

$$M_D^{-1} = \begin{bmatrix} \left(\dfrac{\cos^2\theta}{m_t} + \dfrac{\sin^2\theta}{m_\ell}\right) & \left(\dfrac{\cos\theta\sin\theta}{m_t} - \dfrac{\cos\theta\sin\theta}{m_\ell}\right) & 0 \\ \left(\dfrac{\cos\theta\sin\theta}{m_t} - \dfrac{\cos\theta\sin\theta}{m_\ell}\right) & \left(\dfrac{\sin^2\theta}{m_t} + \dfrac{\cos^2\theta}{m_\ell}\right) & 0 \\ 0 & 0 & \dfrac{1}{m_t} \end{bmatrix} \tag{5.13}$$

For Δ_5 and Δ_6 (bulk silicon),

$$M_D^{-1} = \begin{bmatrix} \frac{1}{m_t} & 0 & 0 \\ 0 & \frac{1}{m_t} & 0 \\ 0 & 0 & \frac{1}{m_\ell} \end{bmatrix} \tag{5.14}$$

By substituting eqn.(5.12), eqn. (5.13) and eqn.(5.14) into eqn.(5.5),

$$E_{\Delta 1} = E_{\Delta 2} = \frac{\hbar^2}{2} k_x^2 \left(\frac{\cos^2\theta}{m_\ell} + \frac{\sin^2\theta}{m_t} \right) + \hbar^2 k_x k_y \left(\frac{\cos\theta \sin\theta}{m_\ell} - \frac{\cos\theta \sin\theta}{m_t} \right) + \frac{\hbar^2}{2} k_y^2 \left(\frac{\sin^2\theta}{m_\ell} + \frac{\cos^2\theta}{m_t} \right) + \frac{\hbar^2}{2} k_z^2 \left(\frac{1}{m_t} \right) \tag{5.15}$$

$$E_{\Delta 3} = E_{\Delta 4} = \frac{\hbar^2}{2} k_x^2 \left(\frac{\cos^2\theta}{m_t} + \frac{\sin^2\theta}{m_\ell} \right) + \hbar^2 k_x k_y \left(\frac{\cos\theta \sin\theta}{m_t} - \frac{\cos\theta \sin\theta}{m_\ell} \right) + \frac{\hbar^2}{2} k_y^2 \left(\frac{\sin^2\theta}{m_t} + \frac{\cos^2\theta}{m_\ell} \right) + \frac{\hbar^2}{2} k_z^2 \left(\frac{1}{m_t} \right) \tag{5.16}$$

$$E_{\Delta 5} = E_{\Delta 6} = \frac{\hbar^2}{2} k_x^2 \left(\frac{1}{m_t} \right) + \frac{\hbar^2}{2} k_y^2 \left(\frac{1}{m_t} \right) + \frac{\hbar^2}{2} k_z^2 \left(\frac{1}{m_\ell} \right) \tag{5.17}$$

Substituting eqn.(5.15), eqn.(5.16) and eqn.(5.17) into eqn. (5.7),

$$\frac{1}{m^*} = \frac{1}{3} \left(\frac{1}{m_\ell} + \frac{2}{m_t} \right) \tag{5.18}$$

Hence, the six degenerate ellipsoid model predicts that m^* of (100) Si is independent of the channel orientation. As opposed to bulk silicon, the electron carrier concentration in the inversion-layer silicon has a charge centroid about 1 nm away from the interface between silicon and the gate dielectric owing to quantum confinement [4.22]. The out-of-plane mass is a function of the distance of the inversion-layer charge centroid from the SiO_2/Si interface [2.83]. Hence, we believe that the conduction band minimum near the SiO_2/Si interface may be slightly different from that of bulk silicon (see Fig. 5-13). Considering (100)Si, M_E^{-1} of the in-plane valleys (Δ_1 to Δ_4) near the SiO_2/Si interface will be the same as that of bulk silicon, as shown in eqn. (5.3a) and eqn. (5.3b). Our theory is that the out-of-plane conduction valleys (Δ_5 and Δ_6) that are near the SiO_2/Si interface may be slightly different from that of bulk silicon.

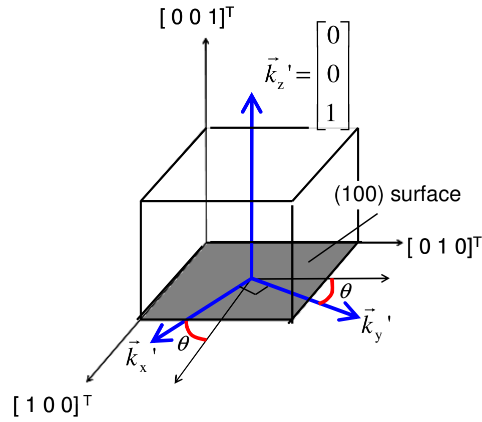


Figure 5-12 $\Re_{E \leftarrow D}$ **of (100) Si for a generalized channel orientation.**

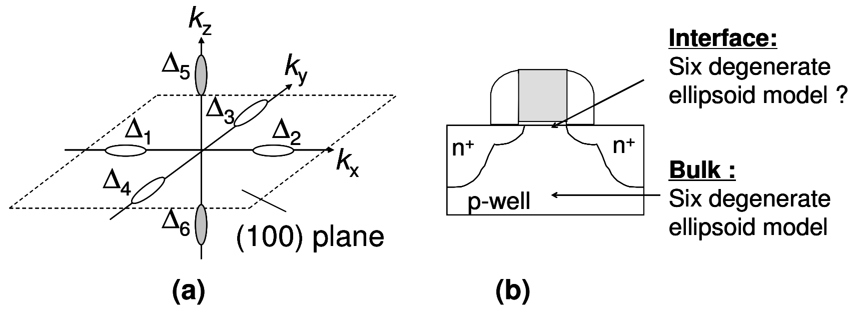


Figure 5-13 (a) Schematics of the six conduction band valleys of (100) Si, (b) Validity of the six degenerate ellipsoid model for the conduction band energy minimum of silicon at the interface between SiO_2/Si interface.

For Δ_5 (near the SiO_2/Si interface),
$$M_E^{-1} = \begin{bmatrix} \frac{1}{m_t'} & 0 & 0 \\ 0 & \frac{1}{m_t'} & 0 \\ 0 & 0 & \frac{1}{m_\ell'} \end{bmatrix} \tag{5.19}$$

where m_t' is the transverse mass of Δ_5 near the SiO_2/Si interface. m_ℓ' is the longitudinal mass of Δ_5 near the SiO_2/Si interface.

For Δ_6 (near the SiO_2/Si interface),
$$M_E^{-1} = \begin{bmatrix} \frac{1}{m_t''} & 0 & 0 \\ 0 & \frac{1}{m_t''} & 0 \\ 0 & 0 & \frac{1}{m_\ell''} \end{bmatrix} \tag{5.20}$$

where m_t'' is the transverse mass of Δ_6 near the SiO_2/Si interface. m_ℓ'' is the longitudinal mass of Δ_6 near the SiO_2/Si interface. Since we do not know which conduction band valleys near the SiO_2/Si interface will have the lowest energy, we will explore the various possibilities.

Case 1: The six conduction band valleys (Δ_1 to Δ_6) have the same electron energy but M_E^{-1} for the out-of-plane valleys (Δ_5 and Δ_6) near the SiO_2/Si interface are different from that of bulk silicon. Since there is equal probability that an electron can reside in one of the six conduction band valleys, we have to consider the average of the six conduction band valleys when we calculate m^*.

$$m^* = \frac{6}{2\left(\frac{1}{m_t} + \frac{1}{m_\ell}\right) + \frac{1}{m_t'} + \frac{1}{m_t''}} \tag{5.21}$$

Case 2: The electron energy of $\Delta_1$to Δ_4 is lower than that of Δ_5 and Δ_6. Since electrons will preferentially occupy the lowest electron energy level, the four in-plane conduction band valleys will be filled with electrons but the out-of-plane conduction band valleys will be empty. In this case, there is an equal probability that an electron can reside in one of the four in-plane conduction band valleys, and thus we have to take the average of the four in-plane conduction band valleys when we calculate m^*.

$$m^* = \frac{4}{2\left(\frac{1}{m_t} + \frac{1}{m_\ell}\right)} = \frac{2}{\left(\frac{1}{m_t} + \frac{1}{m_\ell}\right)} \tag{5.22}$$

Case 3: If Δ_5 has the lowest electron energy, $m^* = m_t'$ (5.23)

Case 4: If Δ_6 has the lowest electron energy, $m^* = m_t''$ (5.24)

As opposed to the six degenerate ellipsoid model, our theory shows that the electron mobility of NMOS transistor on (100) Si will be independent of the channel orientation if the momentum relaxation time (τ_m) is independent of the channel orientation.

5.5.2 Mechanism behind the increase in L_{eff} of <100> transistors

Our theory is that (100) Si is first amorphorized by the S/D extension implants and thus some of the implanted S/D extension implants may experience lateral ion channeling into either <110> crystal direction or <100> crystal direction. Fig. 5-14 shows the schematics of lateral ion channeling and axial ion channeling. Axial channeling is associated with the junction depth whereas lateral ion channeling is associated with the L_{eff}. The probability for lateral ion channeling ($Pr_{lateral}$) can be expressed as a product of three different probabilities.

$$Pr_{lateral} = Pr_1 Pr_2 Pr_3 \qquad (5.25)$$

where Pr_1 is the probability for vertically implanted ions to be scattered laterally. Pr_2 is the probability for laterally scattered ions to enter <110> or <100> channel. Pr_3 is the probability for the ions that have entered <110> or <100> channel to be retained in the channel. Since <110> transistors and <100> transistors are fabricated on (100)Si, Pr_1 is the same for both transistors. From Table 5-1 and eqn. (5.9), r_{eff} along <100> direction ($r_{eff,100}$) is 0.1084 nm, whereas r_{eff} along <110> direction ($r_{eff,110}$) is 0.183 nm. Hence, Pr_2 is smaller for <100> transistor compared to <110> transistor. From eqn.(5.9), $(\theta_c)_{lateral}$ of <100> direction is smaller than that of <110> direction. Since Pr_3 decreases with a decrease in the critical angle [5.18], Pr_3 is smaller for <100> transistor compared to <110> transistor. Consequently, $Pr_{lateral}$ of <100> transistor is smaller than that of <110> transistor. This shows that the increase in L_{eff} of <100> MOS transistors can be explained by the reduction in the lateral ion channeling of the S/D extension implants.

From Fig. 5-15, the change in the channel orientation decreases I_{g_off} of PMOS transistors by 0.130 pA and decreases I_{g_off} of NMOS transistors by 0.489 pA. Similarly, DIBL of PMOS transistors is reduced by 2.62 mV and DIBL of NMOS transistors is reduced by 21.13 mV. Since a smaller I_{g_off} is indicative of a smaller gate-to-drain overlap [1.33] and a smaller DIBL is also indicative of a bigger L_{eff}, this shows that change in channel orientation has indeed increased the L_{eff} of PMOS transistors and NMOS transistors on (100)Si. The bigger increase in L_{eff} of NMOS transistor can be understood as follows. From Fig. 5-16, the ease of lateral ion channeling depends on the relative dimension between the atomic radius of the channeling ions and the effective radius of the open space in the crystal lattice. From Table 5-1 and eqn. (5.9), $r_{eff,100}$ is 0.1084 nm, whereas $r_{eff,110}$ is 0.183 nm. The atomic radius of arsenic ($r_{arsenic}$) is 0.115 nm, whereas the atomic radius of boron (r_{boron}) is 0.085 nm. Since r_{boron} is smaller than $r_{eff,100}$ and $r_{eff,110}$, it is

possible for lateral boron channeling to occur along <100> direction and <110> direction. However, the situation is different for arsenic. Since $r_{eff,100} < r_{arsenic} < r_{eff,110}$, it is harder for lateral arsenic channeling to occur along <100> direction. Hence, <100> NMOS transistor will experience a bigger reduction in the lateral ion channeling of the S/D extension implants as compared to PMOS transistor. Another contributing factor to the bigger increase in L_{eff} of NMOS transistors can be understood as follows. As discussed in Section 5.4.1 and Section 5.4.2, the S/D extension implants of NMOS transistor is only affected by the reduction in the lateral arsenic ion channeling as there is negligible change in the arsenic diffusion. On the other hand, the S/D extension implants of PMOS transistor experience two competing factors: (i) the reduction in lateral boron channeling along <100> direction, and (ii) the increase in boron diffusion along <100> direction. This accounts for the bigger increase in L_{eff} of <100> NMOS transistors as compared to <100> PMOS transistors.

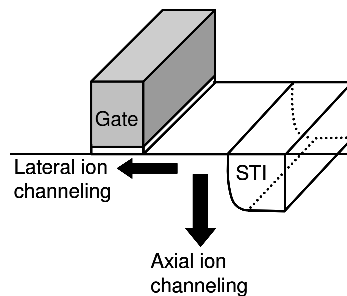


Figure 5-14 Difference between lateral ion channeling and axial channeling.

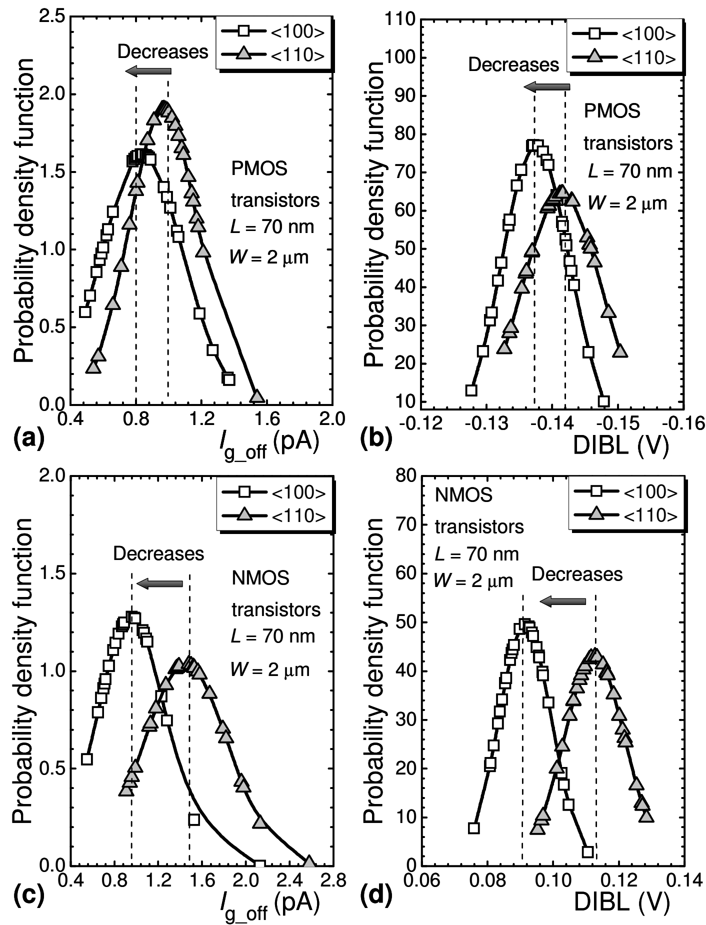


Figure 5-15 Evidence of the increase in L_{eff} of <100> MOS transistors on (100) Si after a change in the channel orientation. Note that DIBL = $V_{th,lin} - V_{th,sat}$.

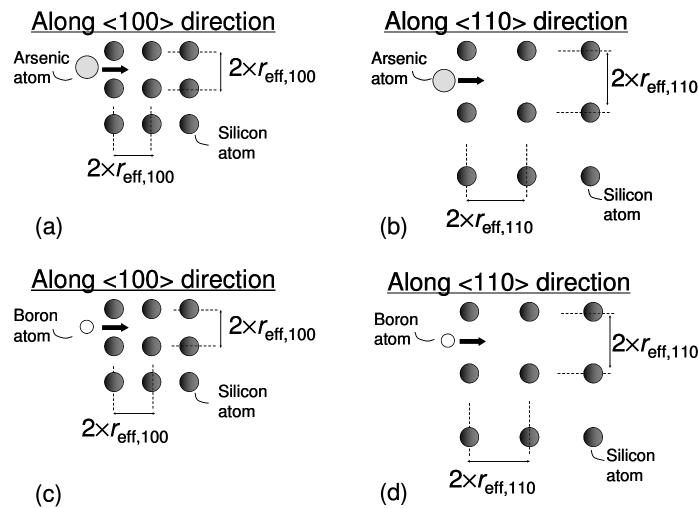


Figure 5-16 The ease of lateral ion channeling for arsenic and boron along <100> direction and <110> direction of (100)Si.

5.5.3 Effects of channel orientation on I_{on} and I_{off} of MOS transistors

In Section 5.5.1, we have shown that the electron mobility of (100)Si will be independent of the channel orientation if τ_m is independent of the channel orientation. Another approach to investigate the effects of the change in channel orientation on the electron mobility is to use the log I_{off} versus I_{on} plot. From Fig. 5-17, log I_{off} versus I_{on} characteristics of <110> PMOS transistors follow a different distribution as that of <100> PMOS transistors. On the other hand, the log I_{off} versus I_{on} characteristics of <110> NMOS transistors follow the same distribution as that of <100> NMOS transistors. This shows that the change in channel orientation has increased the hole mobility of PMOS transistors but has minimal effects of the electron mobility of NMOS transistors. This is consistent with our experimental results, as shown in Fig. 5-18 and Fig. 5-19. In Section 5.6.2, we have shown that the change in channel orientation can lead to a slight increase in L_{eff} of PMOS transistors but a significant increase in L_{eff} of NMOS transistors. This results in a decrease the subthreshold I_{off} for both long-channel <100>NMOS transistors and short-channel NMOS transistors. This is consistent with our experimental data, as shown in Fig. 5-17. From eqn. (1.1), I_{on} of long-channel NMOS transistors is expected to decrease when L_{eff} is increased. From eqn. (3.19) and Fig. 3-39, I_{on} of the nanoscale MOS transistors is a function of v_{sat_eff} , which is dependent on L_{eff}. From Fig. 5-20, the switching from the conventional <110> channel orientation to <100> channel orientation has indeed increased I_{on} and I_{off} of both long-channel NMOS transistors and short-channel NMOS transistors that are fabricated on (100)Si.

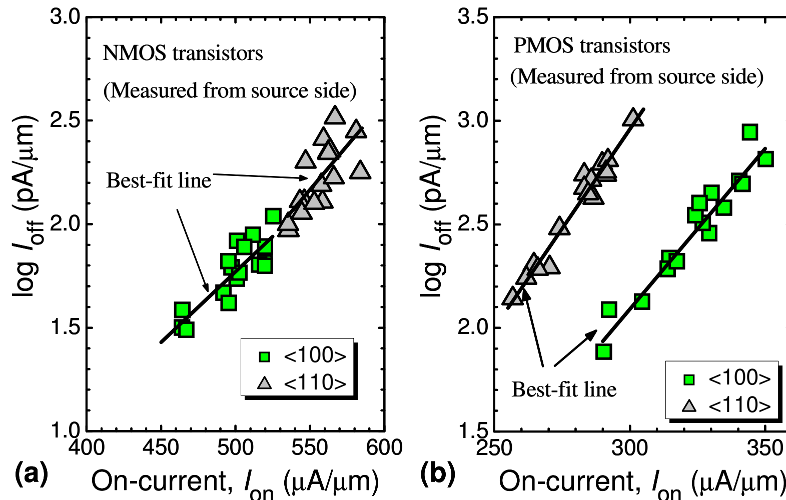


Figure 5-17 Effects of the change in channel orientation on the log I_{off} versus I_{on} plot of NMOS transistors (b) PMOS transistors on (100)Si. (L = 60 nm , W =1 µm).

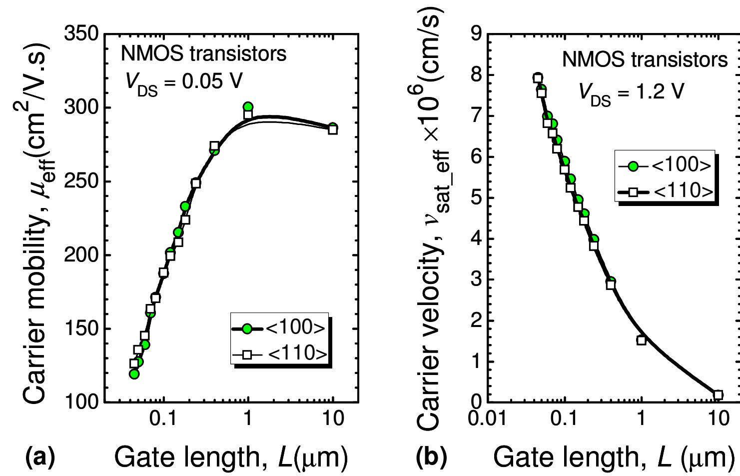


Figure 5-18 Effects of the change in the channel orientation on μ_{eff} and v_{sat_eff} of NMOS transistors on (100) Si.

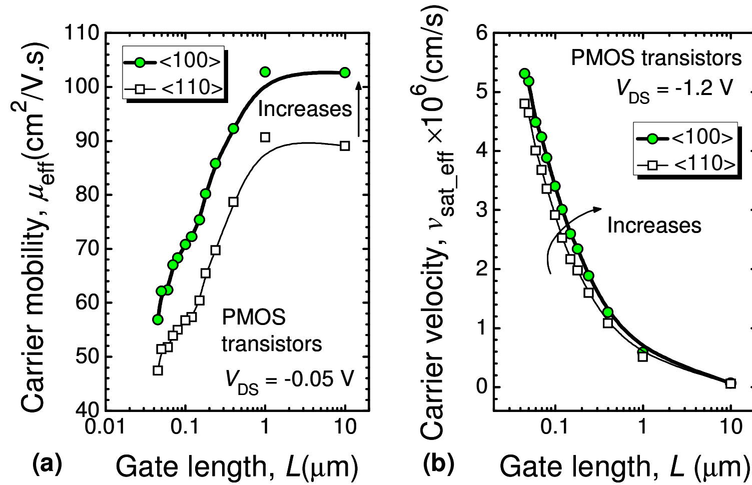


Figure 5-19 Effects of the change in the channel orientation on μ_{eff} and v_{sat_eff} of PMOS transistors on (100) Si.

As discussed in Section 5.1, the switching from the conventional <110> channel orientation to <100> channel orientation increases the hole mobility (μ_{eff}). Based on eqn. (1.1) and eqn (3.19), I_{on} of long-channel PMOS transistor and nanoscale PMOS transistor are expected to increase after the change in the channel orientation. This is consistent with our experimental results as shown in Fig. 5-21. With reference to eqn. (4.36), the equation of the subthreshold I_{off} is a function of the low-field mobility in the subthreshold regime (μ_{sub}) and L_{eff}. From Section 4.6, experimental data has shown that there is indeed a relationship between μ_{sub} and μ_{eff} such that μ_{sub} is expected to increase when μ_{eff} increases. Hence, we would expect μ_{sub} of <100> PMOS transistor to be bigger

than that of <110> PMOS transistor. In Section 5.5.2, we have shown that the change in the channel orientation can lead to a slight increase in L_{eff} of PMOS transistor. In short, the subthreshold I_{off} of <100> PMOS transistor is subjected to two competing factors: (i) the increase in μ_{sub} can lead to an increase in the subthreshold current, and (ii) the slight increase in L_{eff} can lead to a decrease in the subthreshold current. For long-channel PMOS transistors, the effects of the increase in μ_{sub} is likely to dominate over the effects of the slight increase in L_{eff}, and thus the subthreshold I_{off} of <100> PMOS transistor is bigger than that of <110> PMOS transistor. For short-channel PMOS transistors, the effects of the slight increase in L_{eff} is likely to dominate over the effects of the increase in μ_{sub}, and thus the subthreshold I_{off} of <100> PMOS transistor is smaller than that of <110> PMOS transistor.

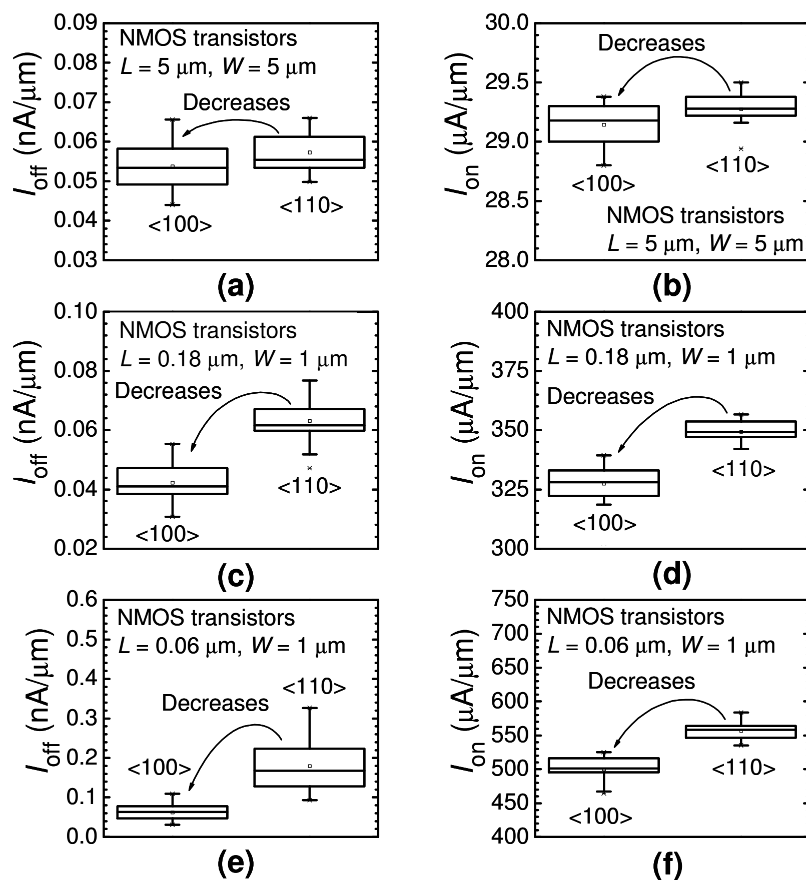


Figure 5-20 Effects of switching from the conventional <110> channel orientation to <100> channel orientation on the I_{on} and I_{off} of NMOS transistors on (100)Si.

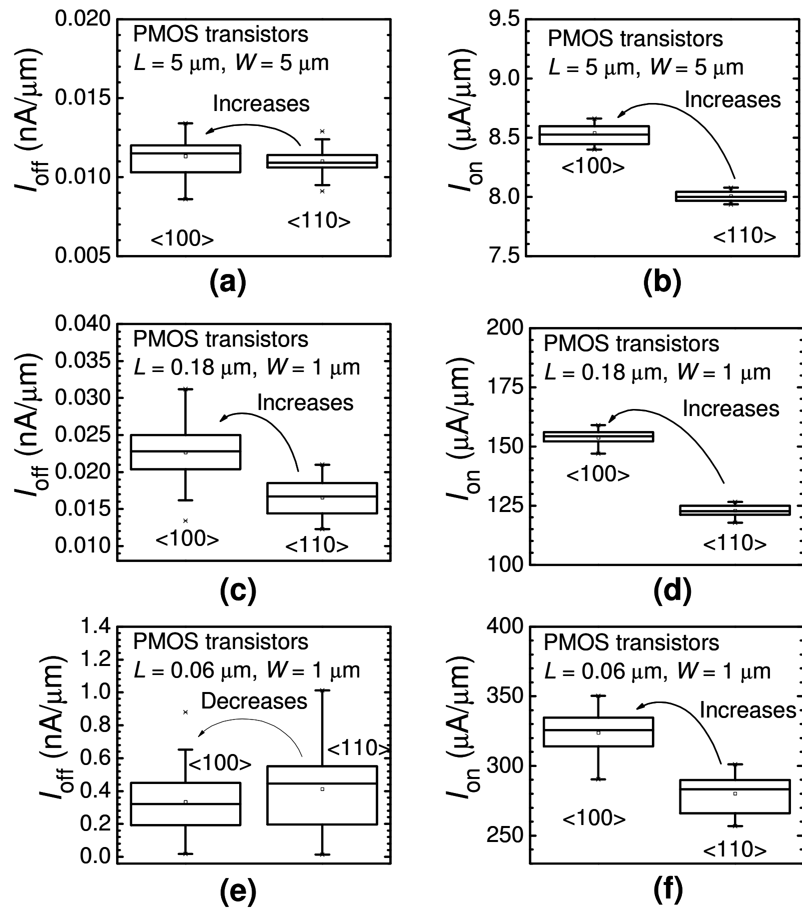


Figure 5-21 Effects of switching from the conventional <110> channel orientation to <100> channel orientation on the I_{on} and I_{off} of PMOS transistors on (100)Si.

6. Conclusions and Recommendation

6.1 Main contributions of the thesis

(1) To clarify the drain current transport mechanism in nanoscale MOS transistor

Using the backbone of the quasi-ballistic theory proposed by Lundstrom, we will unify the merits of velocity saturation model, the ballistic transport and the quasi-ballistic transport and then come up with a simplified saturation drain current equation for nanoscale MOS transistor. This equation can account for the strain-induced I_{on} improvement but yet comprises of parameters that can be easily obtained from the standard electrical measurements.

(2) To determine if there is an overall on-current improvement in NMOS transistor by CESL-induced tensile stress

We observed that the application of CESL-induced tensile stress to NMOS transistors has led to a bigger percentage increase in $\log I_{off}$ as compared to I_{on}. This makes us wonder if stress engineering can lead to an overall I_{on} improvement. The most straightforward approach to ascertain the overall strain-induced I_{on} improvement is to first determine the strain-induced increase in $\log I_{off}$, and then read the corresponding increase in I_{on} from the $\log I_{off}$ versus I_{on} plot characteristics plot. However, this approach does not explain why there is an overall strain-induced improvement in I_{on} when stress engineering actually leads to a bigger percentage increase in $\log I_{off}$ compared to that of I_{on}. Hence, we introduce a third parameter ($V_{th,sat}$ or DIBL) in the $\log I_{off}$ versus I_{on} characteristics. We found that the subthreshold I_{off} is more sensitive to the change in $V_{th,sat}$ as compared to I_{on} and thus the strain-induced increase in I_{off} can be removed by a slight change in $V_{th,sat}$ without much effect on I_{on}. Therefore, there is an overall improvement strain-induced improvement in I_{on} even though stress engineering actually leads to a bigger percentage increase in $\log I_{off}$ compared to that of I_{on}.

(3) To determine the effects of switching from <110> to <100> channel orientation on the on-current and off-current of MOS transistors fabricated on (100) surface-oriented silicon wafer

We have shown that the effective conductivity electron mass of (100) surface-oriented silicon is independent of the channel orientation even though the six degenerate ellipsoid model may not be applicable near the SiO_2/Si interface. In the other words, the switching

from the conventional <110> channel orientation to <100> channel orientation will not change the electron mobility. Experimental results have shown that there is an increase in L_{eff} in NMOS transistors and PMOS transistors after the change in channel orientation. We found that the mechanism behind the increase in L_{eff} is caused by a reduction in the lateral ion channeling of the S/D extension implants rather than a change in the diffusion coefficient of the S/D extension implants. Furthermore, we observed that the increase in L_{eff} of NMOS transistor is more significant as compared to that of PMOS transistor. We attributed this observation to the ease of ion channeling through the open space in the silicon crystal lattice. These findings provide an insight to the behavior of I_{on} and I_{off} of MOS transistors after the change in the channel orientation. We believe that the decrease in I_{on} and I_{off} of long-channel <100> NMOS transistors and short-channel <100> NMOS transistors owing to an increase in L_{eff}. However, the situation is more complicated for PMOS transistors because the subthreshold I_{off} of <100> PMOS transistor is subjected to two competing factors: (i) an increase in μ_{sub} that leads to an increase in the subthreshold current, and (ii) a slight increase in L_{eff} that leads to a decrease in the subthreshold current. Hence, the subthreshold I_{off} of long-channel PMOS transistor is increased by the change in channel orientation owing to the increase in μ_{sub}. On the other hand, the subthreshold I_{off} of short-channel PMOS transistor is decreased by the change in channel orientation owing to the increase in L_{eff}.

6.2 Recommendations of future work

[1] To investigate if the subthreshold mobility is increased by the externally applied mechanical stress

Suggestions: Use four point rod bending and magnetoresistance mobility measurement to show that subthreshold mobility is indeed increased by stress engineering. Note that the four point rod bending fixture cannot be made up of ferromagnetic material as it may interfere with the magnetoresistance mobility measurement.

[2] To investigate the effects of mechanical stress bias in MOS transistor on the sensitivity of the silicon bandgap to a small externally applied mechanical stress

Suggestions: Perform a CESL split comprising of highly compressive stressed CESL, low compressive stressed CESL, neutral CESL, low tensile stressed CESL and highly tensile stressed CESL. Use a sophisticated four point rod bending to monitor the change in

subthreshold swing and threshold voltage at small incremental increase in the externally applied mechanical stress.

[3] To investigate if uniaxial stress will affect the GIDL current of PMOS transistors fabricated on (100) surface-oriented silicon wafer with <100> channel orientation

Rationale: Experimental results have shown that <100> PMOS transistor is virtually insensitive to mechanical stress [1.35, 2.84]. However, the tight-binding calculations show that the strain-induced silicon bandgap narrowing is bigger for <100> channel compared to <110> channel [4.4]. Since GIDL current is dependent on the silicon bandgap, we would expect GIDL current of <100> PMOS transistor to be increased by tensile stress.

[4] To investigate why strained NMOS transistors on (110) surface-oriented silicon substrate can have comparable performance to the strained NMOS transistors on (100) surface-oriented silicon substrate

Rationale: Many studies have demonstrated that the PMOS transistors fabricated on (110) surface-oriented silicon wafer have higher hole mobility than that on (100) surface-oriented silicon wafer, and the strained NMOS transistors on (110) surface-oriented silicon wafer can have comparable performance to that on (100) surface-oriented silicon (See Appendix).

List of publications

Book Chapter

[1] **Peizhen Yang**, W.S. Lau, Seow Wei Lai, V.L. Lo, S.Y. Siah and L. Chan (2010), "The Evolution of Theory on Drain Current Saturation Mechanism of MOSFETs from the Early Days to the Present Days," in *Solid State Circuits Technologies*, edited by J.W. Swart, Intech: Croatia , pp. 49-82.

Journal papers

[1] **Peizhen Yang,** W.S. Lau, V. Ho, C.H. Loh, S.Y. Siah and L. Chan, "Effect of tensile stress on the various components of the off current of n-channel metal-oxide-semiconductor transistors," Appl. Phys. Lett., Vol. 91, No.7, 073514 (2007).

[2] **Peizhen Yang**, W.S. Lau, Tiong Liang Ng, V. Ho, C.H. Loh, L.F. Toh, Y. Liu, S.Y. Siah and L. Chan, " Improvement of n-channel metal-oxide-semiconductor transistors by tensile stress despite increase in both on and subthreshold off currents," J. Appl. Phys, Vol. 103, article 094518 (2008).

[3] **Peizhen Yang**, W.S. Lau, V. Ho, B.K. Lim, S.Y. Siah and L. Chan, "Switching from <110> to <100> orientation increases both the on current and off current of p-channel metal-oxide semiconductor transistors," Appl. Phys. Lett., Vol. 93, No. 3, 033501 (2008).

[4] **Peizhen Yang** , W.S. Lau, Seow Wei, V. L. Lo, S.Y. Siah and L. Chan, "Effects of switching from <110> to <100> channel orientation and tensile stress on n-channel and p-channel metal-oxide-semiconductor transistors," Solid-State Electronics, Vol. 54, No. 4, pp. 461-474 (2010).

[5] **Peizhen Yang**, W.S. Lau, Seow Wei Lai, V. L. Lo, L.F. Toh, Jacob Wang, S.Y. Siah and L. Chan, "Mechanism for improvement of n-channel metal-oxide-semiconductor transistors by tensile stress," J. Appl. Phys., Vol. 108, article 034506 (2010).

[6] **Peizhen Yang**, W.S. Lau, Seow Wei Lai, V. L. Lo, , S.Y. Siah and L. Chan, "Selection of gate length and gate bias to make nanoscale metal-oxide-semiconductor transistors less sensitive to both statistical gate length variation and temperature variation," Solid-State Electronics, Vol. 54, No. 11, pp. 1304–1311(2010).

[7] W.S. Lau, **Peizhen Yang**, C.W. Eng, V. Ho, C.H. Loh, S.Y. Siah, D. Vigar and L. Chan, “ A study of the linearity between I_{on} and log I_{off} of modern MOS transistors and its application to stress engineering,” Microelectron. Reliab., Vol. 48, No.4, pp. 497-503 (2008).

[8] W.S. Lau, **Peizhen Yang** , V. Ho, L.F. Toh, Y. Liu, S.Y. Siah and L. Chan, "An explanation of the dependence of the effective saturation velocity on gate voltage in sub-0.1 μm metal-oxide-semiconductor transistors by quasi-ballistic transport theory," Microelectron. Reliab., Vol. 48, No. 10, pp. 1641-1648 (2008).

[9] W.S. Lau, **Peizhen Yang**, V. Ho, B.K. Lim, S.Y. Siah and L.Chan," Reduced boron lateral ion channeling in very short p-channel transistors by switching from <110> to <100> channel orientation," Appl. Phys. Lett.,Vol. 93,No. 23, 233505 (2008).

[10] W.S. Lau, **Peizhen Yang**, Seow Wei Lai, V.L. Lo, S.Y. Siah and L. Chan, “ Improvement of subthreshold swing of n-channel transistor by uniaxial tensile stress due to a quantum mechanical mechanism instead of physical thinning,” Appl. Phys. Lett., Vol. 94, No. 17, article 173503 (2009).

[11] W.S. Lau, **Peizhen Yang**, Eng Hua Lim, Yee Ling Tang, Seow Wei Lai, V. L. Lo, S.Y. Siah and L. Chan, “Observation of halo implant from the drain side reaching the source side and vice versa in extremely short p-channel transistors,” Microelectron. Reliab., Vol. 50, No. 3, pp. 346-350 (2010).

[12] W.S. Lau, **Peizhen Yang**, T.P. Chen, S.Y. Siah and L. Chan, “Physics of electron mobility independent of channel orientation in n-channel transistors based on (100) silicon wafers and its experimental verification,” Appl. Phys. Lett, Vol. 97, No. 13, article 133508 (2010).

Conference paper

[1] **Peizhen Yang**, W.S. Lau, V. Ho, C.H. Loh, S.Y. Siah and L. Chan, “ A comparison between the quasi-ballistic transport model and the conventional velocity saturation model for sub-0.1 μm MOS transistors,” IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC 2007) , Tainan, Taiwan, pp. 99-102.

[2] W.S. Lau, **Peizhen Yang**, V. Ho, B.K. Lim, S.Y. Siah and L. Chan, “ Effective channel length increased due to switching from <110> to <100> orientation for PMOS transistors fabricated by 65 nm CMOS technology,” IEEE International

Conference on Electron Devices and Solid-State Circuits (EDSSC 2008) , Hongkong, China, pp. 1-4.

[3] W.S. Lau, **Peizhen Yang**, Edwin T.L. Ng, Z.W. Chian, V. Ho, S.Y. Siah and L. Chan, " Region of nearly constant off current versus gate length characteristics for sub-0.1 μm low power CMOS technology," IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC 2008) , Hongkong, China, pp. 5-9 .

Appendix: Effects of surface-orientation on electron mobility and hole mobility

From Fig.A-1, PMOS transistors on (110)Si have higher hole mobility than PMOS transistor on (100)Si [A.1]. This can be understood as follows. The energy split between HH and LH band in (110) PMOS transistor, $\Delta E_{HH\text{-}LH}{}^{(110)}$ is bigger compared to that of (100) PMOS transistors, $\Delta E_{HH\text{-}LH}{}^{(100)}$. From Fig.A-2, $\Delta E_{HH\text{-}LH}{}^{(110)}$ is larger than the optical phonon energy when the substrate impurity concentration (N_{sub}) or the surface carrier concentration (N_s) is high [A.1]. As a result, optical phonon scattering, which is the dominant scattering mechanism in PMOS transistors, can be suppressed in (110) PMOS transistors but not in (100) PMOS transistor. This accounts for the higher hole mobility in PMOS transistors on (110)Si compared to PMOS transistors on (100)Si.

In the absence of mechanical stress, the electron mobility is higher for NMOS transistor on (100)Si compared to NMOS transistor in (110)Si, as shown in Fig.A-3 [A.1]. This can be understood as follows. When the six degenerate conduction band valleys are projected onto (100)Si, there are two out-of-plane valleys (Δ_5, Δ_6) and four in-plane valleys (Δ_1, Δ_2, Δ_3, Δ_4). On the other hand, when the 6-fold valleys are projected onto (110)Si, there are four out-of-plane valleys (Δ_1, Δ_2, Δ_3, Δ_4) and two in-plane valleys (Δ_5, Δ_6), as shown in Fig. A-4. Owing to the electrostatic confinement at the Si/SiO_2 interface, the six degenerate conduction band valleys are split into two-fold valleys and four-fold valleys. From Table A-1, for (100)Si, the confinement mass (m_z) of the two out-of-plane valleys (Δ_5, Δ_6) is 0.97 m_o whereas m_z of the four in-plane valleys (Δ_1, Δ_2, Δ_3, Δ_4) is 0.19 m_o [5.2]. For (110)Si , m_z of the four out-of-plane valleys (Δ_1, Δ_2, Δ_3, Δ_4) is 0.317 m_o whereas m_z of the two in-plane valleys (Δ_5, Δ_6) is 0.19 m_o. From eqn. (4.18), a bigger m_z leads to a lower energy level of the conduction subband in the potential well at the Si/SiO_2 interface. Since the difference in m_z between the 2-fold valleys and 4-fold valleys is larger in (100)Si, there is a bigger energy split in the conduction band valleys for (100)Si. This leads to a higher electron mobility in NMOS transistor on (100)Si under unstrained condition.

However, Saitoh et al. reported that the saturation drain current of scaled (110) NMOS transistor approaches (100) NMOS transistor owing to STI induced compressive stress along the channel width [A.2]. This can be understood as follows. From Table A-2, the device-level piezoresistance coefficient that is perpendicular to the channel length ($\pi_\perp$) is negative for (100) NMOS transistor but positive for (110) NMOS transistor [2.41]. In the other words, the transverse STI-induced compressive stress is expected to degrade the electron mobility of (100) NMOS transistor but enhance the electron mobility of (110)

NMOS transistor. In view of the above discussion, there is a tradeoff between the two competing factors: (i) a smaller energy split in the conduction band valleys in (110) NMOS transistor compared to (100) NMOS transistor, (ii) positive $\pi_{\perp}$ in (110) NMOS transistor and negative $\pi_{\perp}$ in (10) NMOS transistor. This explains why the strained (110) NMOS transistors can get comparable performance to the (100) NMOS transistor.

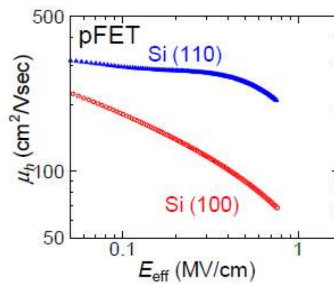


Figure A-1 Higher hole mobility (μ_h) as a function of the effective vertical field (E_{eff}) for <110> PMOS transistors fabricated on (100)Si and (110)Si[A.1].

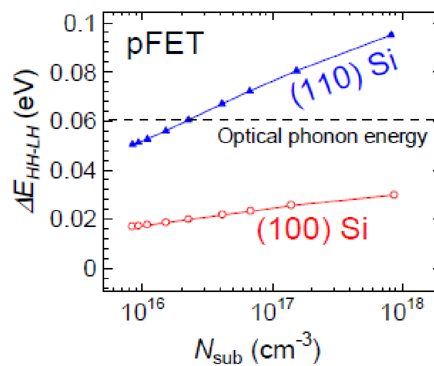


Figure A-2 Energy split ($\Delta E_{HH\text{-}LH}$) in PMOS transistors on (100)Si and (110)Si [A.1].

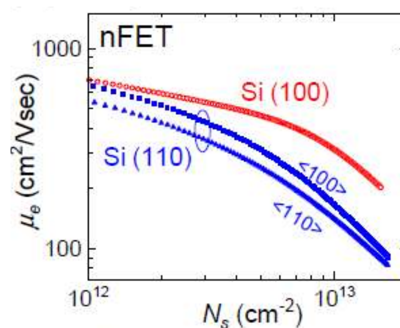


Figure A-3 Electron mobility (μ_e) as a function of the surface carrier concentration (N_s) for NMOS transistors on (100)Si and (110)Si [A.1].

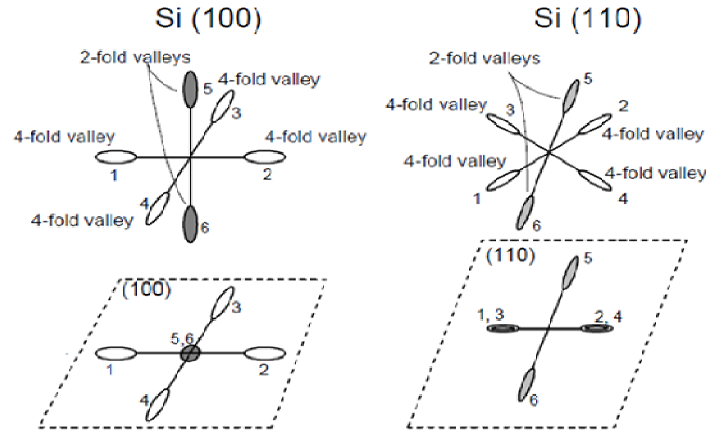


Figure A-4 Conduction band valleys on (100) plane and (110) plane [A.1].

Table A-1: Effects of surface orientation and channel orientation on the effective electron masses along the quantum confinement direction (m_z) [5.2].

Surface, Channel	Quantum confinement	Conduction band valleys	Effective mass along m_x	m_y	m_z	Degeneracy
(100), [100] /[110]	[001]	Δ_5 , Δ_6	m_t	m_t	m_ℓ	2
		Δ_3 , Δ_4	m_ℓ	m_t	m_t	2
		Δ_1 , Δ_2	m_t	m_ℓ	m_t	2
(110), [001]	[110]	Δ_5 , Δ_6	m_ℓ	m_t	m_t	2
		Δ_1 - Δ_4	m_t	$\frac{m_t + m_\ell}{2}$	$\frac{2m_t m_\ell}{m_t + m_\ell}$	4
(110), $[1\bar{1}0]$	[110]	Δ_5 , Δ_6	m_t	m_ℓ	m_t	2
		Δ_1 - Δ_4	$\frac{m_t + m_\ell}{2}$	m_t	$\frac{2m_t m_\ell}{m_t + m_\ell}$	4

Table A-2 Device-level piezoresistance coefficients of the <110> NMOS transistors fabricated on (100)Si and (110)Si (units: 10^{-11} Pa^{-1}) [2.41].

Piezoresistance coefficient	(100) surface-oriented silicon	(110) surface-oriented silicon
$\pi_{//}$	-49	-27
$\pi_{\perp}$	-16	+53

Bibliography

[1.1] K.J. Kuhn, "Moore's law past 32 nm: future challenges in device scaling," IEEE Int. Workshop on Computational Electronics (IWCE) 2009, pp. 1-4.

[1.2] D.J. Frank, R.H. Dennard, E. Nowak, P.M. Solomon and Y. Taur, "Device scaling limits of Si MOSFETs and their application dependencies," Proceedings of the IEEE, Vol. 89, No. 3, pp. 259-288 (2001).

[1.3] N.S. Kim, T. Austin, D. Blaauw, T. Mudge, K. Flautner, et al., "Leakage Current: Moore's Law Meets Static Power," Computer, Vol. 36, No. 12, pp. 68-75 (2003).

[1.4] M.M. Frank, S. Kim, S.L. Brown, J. Bruley, M. Copel, et al., "Scaling the MOSFET gate dielectric: From high-*k* to higher-*k* ?," Microelectronics Engineering, Vol. 86, pp. 1603-1608 (2009).

[1.5] C.T. Sah (1991). *Fundamentals of Solid-State Electronics*, 1st ed., World Scientific, Singapore, p.555.

[1.6] Y. Taur and T. Ning (1998). *Fundamentals of modern VLSI devices*. 1st ed, Cambridge University Press, New York, p. 152.

[1.7] C-T. Sah (1991). *Fundamentals of Solid-State Electronics*, 1st ed., World Scientific, Singapore, p.251.

[1.8] A.B. Bhattacharyya (2009). *Compact MOSFET models for VLSI Design*, John Wiley & Sons (Asia) Singapore , pp. 27-29.

[1.9] C.B. Jr. Norris and J.F. Gibbons, "Measurement of high-field carrier drift velocities in silicon by a time-of-flight technique," IEEE Trans. Electron Dev., Vol. 14, No. 1, pp. 38-42 (1967).

[1.10] K.K. Thornber, "Relation of drift velocity to low-field mobility and high-field saturation velocity," J. Appl. Phys., Vol. 51, No. 4, pp. 2127-2136 (1980).

[1.11] K. Rim, J.L. Hoyt and J.F. Gibbons, "Fabrication and Analysis of Deep Submicron Strained-Si N-MOSFET's," IEEE Trans. Electron Dev., Vol. 47, No.7, pp. 1406- 1415 (2000).

[1.12] S. Pidin, T. Mori, R. Nakamura, T. Saiki, R. Tanabe, et al., "MOSFET Current Drive Optimization Using Silicon Nitride Capping Layer for 65-nm Technology Node," IEEE Symp. VLSI Tech. Dig. 2004, pp. 54-55.

[1.13] M. Nishikawa, K. Okabe, K. Ikeda, N. Tamura, H. Maekawa, et al., "Successful Integration Scheme of Cost Effective Dual Embedded Stressor Featuring Carbon Implant and Solid Phase Epitaxy for High Performance CMOS," IEEE Symp. VLSI Tech. System. And Applications (VLSI-TSA) 2009, pp. 26-27.

[1.14] K. Uchida, T. Krishnamohan, K. C. Saraswat and Y. Nishi, "Physical Mechanisms of Electron Mobility Enhancement in Uniaxial Stressed MOSFETs and Impact of Uniaxial Stress Engineering in Ballistic Regime," IEEE Electron Device Meet. (IEDM) 2005, pp. 129-132 (2005).

[1.15] E.X. Wang, P. Matagne, L. Shifren, B. Obradovie, R. Kotlyar, et al., " Physics of Hole Transport in Strained Silicon MOSFET Inversion Layers," IEEE Trans. Electron Dev., Vol. 53, No. 8, pp. 1840-1851 (2006).

[1.16] J.G. Ruch, "Electron Dynamics in Short Channel Field-Effect Transistors," IEEE Trans. Electron Dev., Vol. 19, No. 5, pp. 652-654 (1972).

[1.17] T. Mizuno, "New Channel Engineering for Sub-100 nm MOS Devices Considering Both Carrier Velocity Overshoot and Statistical Performance Fluctuations," IEEE Trans. Electron Dev., Vol. 47, No. 4, pp.756-761 (2000).

[1.18] H. Miyata, T. Yamada and D.K. Ferry, "Electron transport properties of a strained Si layer on a relaxed $Si_{1-x}Ge_x$ substrate by Monte Carlo simulation," Appl. Phys. Lett.,Vol. 62, No. 21, pp. 2661- 2663 (1993).

[1.19] J. Kim, J. Lee, Y. Yun, B-G. Park, J.D. Lee, et al., "Extraction of Effective Carrier Velocity and Observation of Velocity Overshoot in Sub-40 nm MOSFETs," Journal of Semiconductor Technology and Science, Vol. 8, No. 2, pp. 115-120 (2008). Note that this is a Korean journal.

[1.20] K. Natori, "Ballistic metal-oxide-semiconductor field effect transistor," J. Appl. Phys., Vol. 76, No. 8, pp. 4879- 4890 (1994).

[1.21] K. Natori, T. Shimizu and T. Ikenobe, "Multi-subband effects on performance limit of nanoscale MOSFETs," Jpn. J. Appl. Phys., Vol. 42, No. 4B, pp. 2063-2066 (2003).

[1.22] K. Natori, N. Wada and T. Kurusu, "New Monte Carlo simulation technique for quasi-ballistic transport in ultrasmall metal oxide semiconductor field-effect transistors," Jpn. J. Appl. Phys., Vol. 44, No. 9A, pp.6463-6470 (2005).

[1.23] K. Natori, "New Solution to High-Field Transport in Semiconductors: II. Velocity Saturation and Ballistic Transmission," Jpn. J. Appl. Phys., Vol. 48, article 034504 (2009).

[1.24] M. Lundstrom, "Elementary Scattering Theory of the Si MOSFET," IEEE Electron Dev. Lett., Vol. 18, No. 7, pp. 361-363 (1997).

[1.25] M. Lundstrom and Z. Ren, "Essential physics of carrier transport in nanoscale MOSFETs," IEEE Trans. Electron Dev., Vol. 49, No. 1, pp. 133 -141 (2002).

[1.26] J.R. Hauser, "A New and Improved Physics-Based Model for MOS Transistors," IEEE Trans. Electron Dev., Vol. 52, No. 12, pp. 2640- 2647 (2005).

[1.27] A. Khakifirooz and D.A. Antoniadis, "MOSFET Performance Scaling – Part I: Historical Trends," IEEE Trans. Electron Dev., Vol. 55, No. 6 , pp. 1391-1400 (2008).

[1.28] A. Khakifirooz, O.M. Nayfeh and D.A. Antoniadis, " A simple semiempirical short channel MOSFET current: voltage model continuous across all regions of operation and employing only physical parameters," IEEE Trans. Electron Dev., Vol. 56, No. 8, pp. 1674-1680 (2009).

[1.29] K. Tatsumura, M. Goto, S. Kawanaka and A. Kinoshita," Correlation between Low-Field Mobility and High-Field Carrier Velocity in Quasi-Ballistic-Transport MISFETs scaled down to L_g = 30 nm," IEEE Electron Device Meet. (IEDM) 2009, pp. 465- 468.

[1.30] T. Skotnicki, “Low Power/ Low Energy Circuits: From Device to System Aspects,” IEEE Electron Device Meet. (IEDM) 2009 Short Course, p. 19.

[1.31] T. Mizuno, N. Sugiyama, T. Tezuka, T. Numata and S. Takagi, “High-Performance Strained-SOI CMOS Devices Using Thin-Film SiGe-on-Insulator Technology,” IEEE Trans. Electron Dev., Vol. 50, No. 4, pp. 988-994 (2003).

[1.32] H. Sayama, Y. Nishida, H. Oda, T. Oishi, S. Shimizu, et al., “Effect of <100> Channel Direction for high performance SCE immune pMOSFET with less than 0.15 μm gate length,” IEEE Electron Device. Meet. (IEDM) 1999, pp. 657-660.

[1.33] T. Komoda, A. Oishi, T. Sanuki, K. Kasai, H. Yoshimura, “Mobility Improvement for 45 nm Node by Combination of Optimized Stress Control and Channel Orientation Design,” IEEE Electron Device Meet. (IEDM) 2004, pp. 217-220.

[1.34] T. Skotnicki, “Low Power/ Low Energy Circuits: From Device to System Aspects,” IEEE Electron Device Meet. (IEDM) 2009 Short Course, pp. 79-82.

[1.35] F. Nouri, P. Verheyen, L. Washington, V. Moro, I. De Wolf, et al.,“ A systematic study of the trade-offs in engineering a locally strained pMOSFET, ” IEEE Electron Device Meet. (IEDM) 2004, pp. 1055-1058.

[1.36] C. Ortolland, Pierre Morin, F. Arnaud, S. Orain, C. Reddy, et al., “Mobility Enhancement by Strained Nitride Liners for 65 nm CMOS Logic Design Features,” Mater. Res. Soc. Symp. Proc., Vol. 913, pp. 52-58 (2006).

[1.37] M. Yang, V.W.C. Chan, K.K. Chan, L. Shi, O.M. Fried, et al., “Hybrid-Orientation Technology (HOT): Opportunities and Challenges,” IEEE Trans. Electron Dev., Vol. 53, No. 5, pp. 965-978 (2006).

[1.38] H.S. Yang, R. Malik, S. Narasimha, Y. Li, R. Divakaruni, et al., “Dual Stress Liner for High Performance sub-45nm Gate Length SOI CMOS Manufacturing,” IEEE Electron Device Meet. (IEDM) 2004, pp. 1075-1077.

[1.39] C.-H. Chen, T.L. Lee, T.H. Hou, C.L. Chen, C.C. Chen, et al., “Stress Memorization Technique (SMT) by Selectively Strained-Nitride Capping for sub-65 nm High-Performance Strained-Si Device Application,” IEEE Symp. VLSI Tech. Dig 2004, pp. 56-57.

[1.40] T. Miyashita, T. Owada, A. Hatada, Y. Hayami, K. Ookoshi, et al., “Physical and Electrical Analysis of the Stress Memorization Technique (SMT) using Poly-Gates and its Optimization for Beyond 45-nm High-Performance Applications,” IEEE Electron Device Meet. (IEDM) 2008, pp. 55-58.

[1.41] H. Okamoto, A. Hokazono, K. Adachi, N. Yasutake, H. Itokawa, et al. “ In situ Doped Embedded-SiGe Source/Drain Technique for 32 nm Node p-channel Metal-Oxide-Semiconductor Field-Effect Transistor,” Jpn. J. Appl. Phys., Vol. 47, No. 4 , pp. 2564-2568 (2008).

[1.42] Y. Liu, O. Gluschenkov, J. Li, A. Madan, A. Ozcan, et al., “Strained Si Channel MOSFETs with Embedded Silicon Carbon Formed by Solid Phase Epitaxy,” IEEE Symp. VLSI Tech. Dig. 2007, pp. 44-45.

[1.43] C. Auth, A. Cappellani, J.-S. Chun, A. Dalis, A. Davis, et al., "45nm High-k + Metal Gate Strain-Enhanced Transistors," IEEE Symp. VLSI Tech. Dig. 2008, pp. 128-129.

[1.44] S. Mayuzumi, S. Yamakawa, Y. Tateshita, M. Tsukamoto, H. Wakabayashi, et al., "Stress-Enhancement Technique in Narrowing NMOSFETs with Damascene-Gate Process and Tensile Liner," Int. Symp. On VLSI Technology, Systems and Applications (VLSI-TSA) 2009, pp. 20-21.

[1.45] S. Yamakawa, S. Mayuzumi, J. Wang, Y. Tateshita, H. Wakabayashi, et al. "Study of Stress Effect on Replacement Gate Technology with Compressive Stress Liner and eSiGe for pFETs," Int. Conference on Simulation of Semiconductor Processes and Devices (SISPAD) 2008, pp. 109-112.

[2.1] J. Welser, J.L. Hoyt and J.F. Gibbons, "Electron Mobility Enhancement in Strained-Si N-Type Metal-Oxide-Semiconductor Field-Effect Transistors," IEEE Electron Dev. Lett., Vol. 15, No. 3, pp. 100-102 (1994).

[2.2] K. Rim, J. Welser, J.L. Hoyt and J.F. Gibbons, "Enhanced Hole Mobilities in Surface-channel Strained-Si p-MOSFETs," IEEE Electron Device Meet. (IEDM) 1995, pp. 517- 520.

[2.3] M.T. Currie, C.W. Leitz, T.A. Langdo, E.A. Fitzgerald and D.A. Antoniadis, "Carrier mobilities and process stability of strained Si n- and p-MOSFETs on SiGe virtual substrates," J. Vac. Sci. Technol. B, Vol. 19, No. 6, pp. 2268-2279 (2001).

[2.4] J.G. Fiorenza, G. Braithwaite, C. Leitz, M.T. Currie, Z.Y. Cheng, et al., "Investigation of misfit dislocation leakage in supercritical strained silicon MOSFET," IEEE Symp. Int. Reliab. Phys. Symp. (IRPS) 2004, pp. 493-497.

[2.5] J.G. Fiorenza, G. Braithwaite, C.W. Leitz,, M.T. Currie, J. Yap, et al., "Film thickness constraints for manufacturable strained silicon CMOS," Semicond. Sci. Technol., Vol. 19, No. 1, pp. L4-L8 (2004).

[2.6] M. Eberlein, S. Escoubas, M. Gailhanou, O. Thomas, P. Rohr, et al.,"Influence of crystallographic orientation on local strains in silicon: A combined high-resolution X-ray diffraction and finite element modeling investigation," Thin Solid Films, Vol. 516, pp. 8042-8048 (2008).

[2.7] Y-M. Sheu, S-J. Yang, C-C. Wang, C-S. Chang, L-P. Huang, et al., "Modeling Mechanical Stress Effect on Dopant Diffusion in Scaled MOSFETs," IEEE Trans. Electron Dev., Vol. 52, No. 1, pp. 30-38 (2005).

[2.8] S.T. Dunham, M. Diebel, C. Ahn and C.L. Shih, "Calculations of effect of anisotropic stress/strain on dopant diffusion in silicon under equilibrium and nonequilibrium conditions," J. Vac. Sci. Technol. B, Vol. 24, No. 1, pp. 456- 461 (2006).

[2.9] C-Y. Hsieh and M-J. Chen, "Electrical Measurement of Local Stress and Lateral Diffusion Near Source/Drain Extension Corner of Uniaxially Stressed n-MOSFETs," IEEE Trans. Electron Dev., Vol. 55, No. 3, pp. 844-849 (2008).

[2.10] H. Liao, P.S. Lee. L.N.L. Goh, H. Liu, J.L. Sudijono, et al.," The impact of etch-stop layer for borderless contacts on deep submicron CMOS device performance – a comparative study," Thin Solid Films, Vol. 462-463, pp. 29-33 (2004).

[2.11] D. Kosemura, M. Takei, K. Nagata, H. Akamatsu, M. Kohno, et al., "Study of Strain Induction for Metal-Oxide-Semiconductor Field-Effect Transistors using Transparent Dummy Gates and Stress Liner," Jpn. J. Appl. Phys., Vol. 48, pp. 066508-1 to 066508-7 (2009).

[2.12] M. Balseanu, L.Q. Xia, V. Zubkov, M. Le, J. Lee, et al.,"Stress modulation of PECVD silicon nitride," ECS Trans., Vol.1, No. 1, pp. 115-124 (2005).

[2.13] P. Morin, D. Benoit and J. Metz, "Modeling the stress enhancement of plasma enhanced chemical vapor deposited silicon nitride films by UV post treatment – impact of the film density," Eur. Phys. J. Appl., Vol. 43, pp. 315- 320 (2008).

[2.14] Y. Miyagawa, T. Murata, Y. Nishida, T. Nakai, A. Uedono, et al., "Local Bonding Structure of High-Stress Silicon Nitride Film Modified by UV Curing for Strained Silicon Technology beyond 45 nm Node SoC Devices," Jpn. J. Appl. Phys., Vol. 46, No. 4B, pp. 1984-1988 (2007).

[2.15] M. Finot, I. A. Blech, S. Suresh and H. Fujimoto, "Large deformation and geometric instability of substrates with thin-film deposits," J. Appl. Phys., Vol. 81, pp. 3457- 3464 (1997).

[2.16] Y-C. Yeo and J. Sun, "Finite-element study of strain distribution in transistor with silicon-germanium source and drain regions," Appl. Phys. Lett., Vol. 86, article 023103 (2005).

[2.17] F. Hüe, M. Hÿtch, F. Houdellier, E. Snoeck and A. Claverie, "Strain mapping in MOSFETs by high-resolution electron microscopy and electron holography ," Mater. Sci. Eng. B, Vol. 154-155, pp. 221-224 (2008).

[2.18] J. Liu and M.C. Ozturk, "Nickel Germanosilicide Contacts Formed on Heavily Boron Doped $Si_{1-x}Ge_x$ Source/Drain Junctions for Nanoscale CMOS," IEEE Trans. Electron Dev., Vol. 52, No. 7, pp. 1535-1540 (2005).

[2.19] S. Gannavaram, N. Pesovic and M.C. Öztürk, "Low Temperature (≤ 800 °C) Recessed Junction Selective Silicon-Germanium Source/Drain Technology for sub-70 nm CMOS," IEEE Electron Dev. Meet. (IEDM) 2000, pp. 437- 440.

[2.20] E.R. Simoen, M.B. Gonzalez, G. Eneman, P. Verheyen, A. Benedetti, et al., " Germanium content dependence of the leakage current of recessed SiGe source/drain junctions," J. Mater. Sci: Mater. Electron., Vol. 18, pp. 787-791 (2007).

[2.21] A. Hikavyy, N. Bhouri, R. Loo, O. Verheyen, F. Clemente, et al., "pMOS transistor with embedded SiGe: Elastic and plastic relaxation issues," Thin solid Films, Vol. 51, pp. 113-116 (2008).

[2.22] M.B. Gonzalez, E. Simoen, B. Vissouvanadin, P. Verheyen, R. Loo, et al., "Impact of the Ge Content on the Bandgap-Narrowing Induced Leakage Current of Recessed $Si_{1-x}Ge_x$ Source/Drain Junctions," IEEE Trans. Electron Dev., Vol. 56, No. 7, pp. 1418-1423 (2009).

[2.23] A. Wei, M. Wiatr, A. Mowry, A. Gehring, R. Boschke, et al., "Multiple Stress Memorization In Advanced SOI CMOS Technologies," IEEE Symp. VLSI Tech. Dig. 2007, pp. 216-217.

[2.24] P-T. Liu, C-S. Huang, P-S. Lim, D-Y. Lee, S-W. Tsao, "Anomalous Gate-Edge Leakage Induced by High Tensile Stress in NMOSFET," IEEE Electron Dev. Lett., Vol. 29, No. 11, pp. 1249-1251 (2008).

[2.25] G.D.M. Dilliway, A.J. Smith, J.J. Hamilton, J. Benson, L. Xu, et al., "Transient enhanced diffusion and deactivation of ion-implanted As in strained Si," Nucl. Instrum. Methods Phys. Res. B, Beam Interact. Mater., Vol. 237, No. 1-2, pp. 131-135 (2005).

[2.26] E. Morifuji, A. Eiho, T. Sanuki, M. Iwai and F. Matsuoka, "Optimization of Stress Memorization Technique for 45 nm Complementary Metal-Oxide-Semiconductor Technology," Jpn. J. Appl. Phys., Vol. 48, pp. 031203-1 to 031203-5 (2009).

[2.27] K.W. Ang, K.J. Chui, V. Bliznetsov, A Du, N. Balasubramanian, et al., "Enhanced Performance in 50 nm N-MOSFETs with Silicon-Carbon Source/Drain Regions," IEEE Electron Device Meet. (IEDM) 2004, pp. 1069-1071.

[2.28] K-J. Chui, K-W. Ang, N. Balasubramanian, M-F. Li, G.S. Samudra, et al., "n-MOSFET With Silicon-Carbon Source/Drain for Enhancement of Carrier Transport," IEEE Trans. Electron Dev., Vol. 54, No. 2, pp. 249-256 (2007).

[2.29] Y. Liu, O. Gluschenkov, J. Li, A. Madau, A. Ozcan, et al., " Strained Si Channel MOSFETs with Embedded Silicon Carbon Formed by Solid Phase Epitaxy," IEEE Symp. VLSI Tech. Dig. 2007, pp. 44-45.

[2.30] H. Itokawa, N. Yasutake, N. Kusunoki, S. Okamoto, N. Aoki, et al., "Source/drain engineering for MOSFETs with embedded-Si:C technology," Appl. Surf. Sci., Vol. 254, No. 19, pp. 6135-6139 (2008).

[2.31] F. Hüe, M. Hÿtch, F. Houdellier, H. Bender and A. Claverie, "Strain mapping of tensiley strained silicon transistors with embedded $Si_{1-y}C_y$ source and drain by dark-field electron holography ," Appl. Phys. Lett., Vol. 95, pp. 073103-1 to 073103-3 (2009).

[2.32] E. Alptekin and M.C. Ozturk, "Schottky Barrier Height of Nickel Silicide Contacts Formed on $Si_{1-x}C_x$ Epitaxial Layers," IEEE Electron Dev. Lett., Vol. 30, No. 12, pp. 1320-1322 (2009).

[2.33] H.J. Osten, D. Endisch, E. Bugiel, B. Dietrich, G.G. Fischer, et al., "Strain relaxation in tensile-strained $Si_{1-y}C_y$ layers on Si (001)," Semicond. Sci. Technol., Vol. 11, pp. 1678-1687 (1996).

[2.34] E. Simoen, B. Vissouvanadin, N. Taleb, M.B. Gonzalez, P. Verheyen, et al., " Leakage current study of $Si_{1-x}C_x$ embedded source/drain junctions," Appl. Surf. Sci., Vol. 254, pp. 6140-6143 (2008).

[2.35] P. A. Stolk, D. J. Eaglesham, H.-J. Gossmann and J.M. Poate, "Carbon incorporation in silicon for suppressing interstitial-enhanced boron diffusion," Appl. Phys. Lett., Vol. 66, No. 11, pp. 1370 – 1372 (1995).

[2.36] M. Kanno, A. Shibuya, M. Matsumura, K. Tamura and H. Tsuno, "Empirical Characteristics and Extraction of Overall Variations for 65-nm MOSFETs and Beyond," IEEE Symp. VLSI Tech. Dig. 2007, pp. 88-89.

[2.37] K-M. Tian, M. Zhu, W-W. Fang, M. Yang, T-Y. Liow, et al., "A New Liner Stressor with Very High Intrinsic Stress (> 6 GPa) and Low Permittivity Comprising Diamond-Like Carbon (DLC) for Strained P-Channel Transistors," IEEE Electron Device Meet. (IEDM) 2007, pp. 127-130 (2007).

[2.38] D.K. Schroder (2009). *Semiconductor material and device characterization.* 3rd ed. John Wiley & Sons: New Jersey. pp. 501.

[2.39] K. Shimizu, T. Saraya and T. Hiramoto, "Physical Understanding of Si (110) Hole Mobility in Ultra-Thin Body pFETs by <110> and <111> Uniaxial Compressive Strain," IEEE Electron Device Meet. (IEDM) 2009, pp. 473-476 (2009).

[2.40] J.W. Pan, P.W. Liu, T.Y. Chang, W.T. Chiang, C.H. Tsai, et al., "Mobility and strain effects on <110>/(110) SiGe channel pMOSFETs for high current enhancement," IEEE Electron Dev. Meet. (IEDM) 2006, pp. 1-4.

[2.41] W.T. Chiang, J.W. Pan, P.W. Liu, C.H. Tsai and G.H. Ma, "Strain Effects of Si and SiGe Channel on (100) and (110) Si Surfaces for Advanced CMOS Applications,"Int. Symp. VLSI Tech. Systems and Applications (VLSI-TSA) 2007, pp. 84-85.

[2.42] W. Zhao, J. He, R.E. Belford, L.E. Wernerson and A. Seabaugh, "Partially Depleted SOI MOSFETs Under Uniaxial Tensile Strain," IEEE Trans. Electron Dev., Vol. 51, No. 3, pp. 317- 323 (2004).

[2.43] K. Uchida, R. Zednik, C-H. Lu, H. Jagannathan, J. McVittie, et al., "Experimental Study of Biaxial and Uniaxial Strain Effects on Carrier Mobility in Bulk and Ultrathin-body SOI MOSFETs," IEEE Electron Dev. Meet. (IEDM) 2004, pp. 229-232.

[2.44] A.T. Pham, C. Jungemann and B. Meinerzhagen, "Microscopic modeling of the hole inversion layer mobility in unstrained and uniaxially stressed Si on arbitrarily oriented substrates," Solid-State Electronics, Vol. 52, No. 9, pp. 1437-1442 (2008).

[2.45] W-C. Wang, S-T. Chang, J. Huang and S-J. Kwang, "3D TCAD simulations of unstrained Si CMOS devices with silicon-based alloy stressors and stressed CESL,"Solid-State Electronics, Vol. 53, pp. 880-887 (2009).

[2.46] S. Dhar, E. Ungersböck, H. Kosina, T. Grasser and S. Selberherr,"Electron Mobility Model for <110> Stressed Silicon Including Strain-Dependent Mass,"IEEE Trans. Nanotechnology, Vol. 6, No. 1, pp. 97-100 (2007).

[2.47] H. Ibach and H. Lüth (2003). *Solid-State Physics: An Introduction to Principles of Materials Science*, Springer-Verlag Berlin Heidelberg New York, p. 35.

[2.48] H. Ibach and H. Lüth (2003). *Solid-State Physics: An introduction to Principles of Materials Science*, 3rd ed., Springer-Verlog Berlin Heidelberg New York, pp. 57-58.

[2.49] P.Y. Yu and M. Cardona (2001). *Fundamentals of Semiconductors: Physics and Materials Properties*. 3rd ed. Springer-Verlag Berlin Heidelberg New York, pp. 20-24.

[2.50] M.L. Cohen and J.R. Chelikowsky (1989). *Electronic Structure and Optical Properties of Semiconductors*, Solid-State Sciences 75, 2nd ed., Springer-Verlag New York Berlin Heidelberg, pp. 16-24.

[2.51] D. J. Chadi and M. L. Cohen,"Tight-binding calculations of the valence bands of diamond and zincblende crystals**,** " Phys. Status Solidi B ,Vol. 68, No. 1 , pp. 405-419 (1975).

[2.52] W.A. Harrison (1980). *Electronic Structure and Properties of Solids*, Freeman, San Francisco,. pp. 142-145.

[2.53] P. Vogl, H. P. Hjalmarson and J. D. Dow, “A semi-empirical tight-binding theory of the electronic structure of semiconductor,” J. Phys. Chem, Vol. 44, No. 5, pp 365-378 (1983).

[2.54] Y-C. Chang and D.E. Aspnes, “Theory of dielectric-function anisotropies of (001) GaAs (2 1) surfaces,” Phys. Rev. B, Vol. 41, No. 17, pp. 12002- 12012 (1990).

[2.55] S.L. Richardson, M.L. Cohen, S.G. Louie and J.R. Chelikowsky ,“ Electron charge densities at conduction-band edges of semiconductor,” Phys. Rev. B, Vol. 33, No. 2, pp. 1177- 1182 (1986).

[2.56] J-M. Jancu, R. Scholz, F. Beltram and F. Bassani, “ Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters,” Phys. Rev. B, Vol. 57, No. 11, pp. 6493 – 6507 (1998).

[2.57] G. Dresselhaus, “Spin-Orbit Coupling Effects in Zinc Blende Structures,” Phys. Rev. , Vol. 100, No. 2, pp. 580- 586 (1955).

[2.58] J.M. Luttinger and W. Kohn, “Motion of Electrons and Holes in Perturbed Periodic Fields,” Phys. Rev., Vol. 97, No.4, pp. 869- 883(1955).

[2.59] M. Cardona and F.H. Pollak, “Energy-Band Structure of Germanium and Silicon: The k.p Method,” Phys. Rev., Vol. 142, No. 2, pp.530- 543 (1966).

[2.60] R.F. Pierret (2003). *Advanced Semiconductor Fundamentals*, Modular Series on Solid State Devices vol.VI, 2nd ed., Prentice Hall New Jersey, pp.72-74.

[2.61] S. Richard, F. Aniel and G. Fishman, “Energy-band structure of Ge, Si and GaAs: A thirty-band k.p method, ” Phys. Rev.B, Vol. 70, article 235204 (2004).

[2.62] H. Ibach and H. Lüth (2003). S*olid-State Physics: An Introduction to Principles of Materials Science*, Springer-Verlag Berlin Heidelberg New York, p. 395.

[2.63] J. Singh (1993). *Physics of Semiconductor and their heterostructures*. McGraw-Hill, pp. 160-165.

[2.64] J. Singh (1993). *Physics of Semiconductor and their heterostructures*. McGraw-Hill, pp. 98-109.

[2.65] C. Kittel (1996). *Introduction to solid state physics*. 7th ed. John Wiley & Sons, USA, pp. 209 - 211.

[2.66] M. Chu, Y. Sun, U. Aghoram and S.E. Thompson, “Strain: A Solution for Higher Carrier Mobility in Nanoscale MOSFETs,” Annu. Rev. Mater. Res., Vol. 39, pp. 203-229 (2009).

[2.67] T. Guillaume and M. Mouis, “Calculations of hole mass in [110]-uniaxially strained silicon for the stress-engineering of p-MOS transistors,” Solid-State Electronics, Vol. 50, pp. 701-708 (2006).

[2.68] K. Seeger (2004). *Semiconductor Physics*. 9th ed. Springer-Verlag Berlin Heidelberg New York, pp.226-230.

[2.69] G. Dresselhaus, A.F. Kip and C. Kittel, “Cyclotron Resonance of Electron and Holes in Silicon and Germanium Crystals,” Phys. Rev., Vol. 98, No. 2, pp.368- 384 (1955).

[2.70] P.Y. Yu and M. Cardona (2001). *Fundamentals of Semiconductors- Physics and Materials Properties*. Springer. New York, pp. 110-113.

[2.71] E. Pop, R.W. Dutton and K. E. Goodson, " Analytic band Monte Carlo model for electron transport in Si including acoustic and optical phonon dispersion," J. Applied Physics, Vol. 96, No. 9, pp. 4998 – 5005 (2004).

[2.72] Y. Sun, S.E. Thompson and T. Nishida, "Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors," J. Appl. Phys., Vol. 101, article 104503 (2007).

[2.73] S.E. Thompson, G. Sun, K. Wu, J. Lim and T. Nishida, "Key Differences For Process-induced Uniaxial vs. Substrate-induced Biaxial Stressed Si and Ge Channel MOSFETs," IEEE Electron Dev. Meet. (IEDM) 2004, pp. 221-224.

[2.74] M-H. Bao (2000). *Micro mechanical transducers: pressure sensors, accelerometers, and gyroscopes*. Elsevier , New York, pp. 199-239.

[2.75]Y. Kanda,"A Graphical Representation of the Piezoresistance Coefficients in Silicon," IEEE Trans. Electron Dev., Vol. 29, No. 1, pp. 64-70 (1982).

[2.76] C.S. Smith, "Piezoresistance Effect in Germanium and Silicon," Phys. Rev., Vol. 94, No. 1, pp. 42-49 (1954).

[2.77] C. Gallon, G. Reimbold, G. Ghibaudo, R.A. Bianchi and R. Gwoziecki, "Electrical analysis of external mechanical stress effects in short channel MOSFETs on (001) silicon," Solid State Electron., Vol. 48, No. 4, pp. 561-566 (2004).

[2.78] S.E. Thompson, S. Suthram, Y. Sun, G. Sun, S. Parthasarathy, et al., "Future of strained Si/semiconductors in nanoscale MOSFETs," IEEE Electron Dev. Meet. (IEDM) 2006, pp. 681-684.

[2.79] S.E. Thompson, G. Sun, Y.S. Choi and T. Nishida, "Uniaxial-process-induced strained-Si: Extending the CMOS roadmap,"IEEE Trans. Electron Dev. , Vol. 53, No. 5, pp. 1010-1020 (2006).

[2.80] M. Chu, T. Nishida, X. Lv, N. Mohta and S.E. Thompson, "Comparison between high-field piezoresistance coefficients of Si metal-oxide-semiconductor field-effect transistors and bulk Si under uniaxial and biaxial stress,"J. Appl. Phys., Vol. 102, No. 11, article 113704 (2008).

[2.81] T. Toriyama and S. Sugiyama, "Analysis of Piezoresistance in p-Type Silicon for Mechanical Sensors," J. Microelectromechanical Systems, Vol. 11, No. 5, pp. 598-604 (2002).

[2.82] K. Ishimaru, "45 nm/32 nm CMOS – Challenge and perspective," Solid-State Electronics, Vol. 52, No. 9, pp. 1266-1273 (2008).

[2.83] Y. Nakabayashi, T. Ishihara, T. Numata, K. Uchida and S. Takagi, "Experimental Evaluation of Coulomb-Scattering-Limited Inversion-Layer Mobility of n-type Metal-Oxide-Semiconductor Field-Effect Transistors on Si (100), (110), and (111)-Surfaces: Impact of Correlation between Conductivity Mass and Normal Mass," Jpn. J. Appl. Phys., Vol. 49, article 04DC21(2010).

[2.84] T. Guillaume and M. Mouis, "Calculations of hole mass in [110]-uniaxially strained silicon for the stress-engineering for p-MOS transistors," Solid-State Electronics, Vol. 50, pp. 701-708 (2006).

[2.85] Y.L. Tsang, A.G.O' Neill, B.J. Gallacher and S.H. Olsen., "Using Piezoresistance Model With *C-R.* Conversion for Modeling of Strain-Induced Mobility," IEEE Electron Dev. Lett., Vol. 29, No. 9, pp.1062- 1064 (2008).

[3.1] K. Natori, "Ballistic/ quasi-ballistic transport in nanoscale transistors,"Appl. Surf. Sci., Vol. 254, pp. 6194-6198 (2008).

[3.2] A. Khakifirooz and D.A. Antoniadis, "Transistor Performance Scaling: The Role of Virtual Source Velocity and Its Mobility Dependence," IEEE Electron Dev. Meet. (IEDM) 2006, pp. 1-4.

[3.3] M. Lundstrom (2000) Fundamentals of carrier transport. 2nd ed. New York : Cambridge University Press, pp. 332-334.

[3.4] M.S. Lundstrom and J. Guo (2006). Nanoscale Transistors: Device Physics, Modeling and Simulation. New York, USA, Springer, pp. 30-31.

[3.5] S.M. Sze and K.K. Ng (2007). *Physics of Semiconductor Devices*, 3rd ed., Wiley-Interscience, Hoboken, N.J., p. 309.

[3.6] S.R. Hofstein and F.P. Heiman," Silicon insulated-gate field-effect transistor, " IEEE Proceedings, Vol. 51, No. 9, pp. 1190- 1202 (1963).

[3.7] K. Suzuki and T. Usuki, "Metal oxide semiconductor field effect transistor (MOSFET) based on a physical high-field carrier-velocity model," Jpn. J. Appl. Phys.,Vol. 43, No. 1, pp.77 – 81 (2004).

[3.8] C.T. Sah (1991). *Fundamentals of Solid-State Electronics*. 1st ed., World Scientific, p.554.

[3.9] Y. Taur and T. Ning (1998). *Fundamentals of modern VLSI devices*. 1st ed. New York: Cambridge University Press, New York, pp. 120 - 121.

[3.10] C. Park, J.P. John, K. Klein, J. Teplik, J. Caravella, et al., "Reversal of Temperature Dependence of Integrated Circuits Operating at Very Low Voltages," IEEE Electron Dev. Meet. (IEDM) 1995, pp. 71- 74.

[3.11] S Takagi, A.Toriumi, M. Iwase and H. Tango," On the Universality of Inversion Layer Mobility in Si MOSFET's Part I – Effects of Substrate Impurity Concentration," IEEE Trans. Electron Dev.,Vol. 41, No. 12, pp. 2357-23 62 (1994).

[3.12] K. Chain, J-H. Huang, J. Duster, P.K. Ko and C. Hu, "A MOSFET electron mobility model of wide temperature range (77-400 K) for IC simulation," Semicond. Sci. Technol., Vol. 12, pp. 355-358 (1997).

[3.13] G. Mazzoni, A.L. Lacaita, L.M. Perron, and A. Pirovano, "On surface roughness-limited mobility in highly doped n-MOSFET's," IEEE Trans. Electron Dev., Vol. 46 , No. 7, pp. 1423-1428 (1999).

[3.14] D.K. Schroder (2006) *Semiconductor Material and Device Characterization*. 3rd ed. John Wiley & Sons, Inc., Hoboken, New Jersey. pp. 222 – 223.

[3.15] S. Martinie, G.L. Carval, D. Munteanu, S. Soliveres and J-L. Autran, "Impact of Ballistic and Quasi-Ballistic Transport on Performances of Double-Gate MOSFET-Based Circuits," IEEE Trans. Electron Dev., Vol. 55, No. 9, pp. 2443 – 2453 (2008).

[3.16] I. Saad, R.M.A. Lee, R. Ismail and V.K. Arora., “Physics-based modeling of ballistic transport in nanoscale transistor,” IEEE Proceeding Conference on Modeling and Simulation (AMS) 2009, pp, 729-34.

[3.17] M.L.P. Tan, V.K. Arora, I. Saad, M.T. Ahmadi and R. Ismail, “The drain velocity overshoot in an 80 nm metal-oxide-semiconductor field-effect transistor,” J. Appl. Phys., Vol. 105, article 074503 (2009).

[3.18] D. Fleury, G. Bidal, A. Cros, F. Boeuf, T. Skotnicki, et al., “New Experimental Insight into Ballisticity of Transport in Strained Bulk MOSFETs,” IEEE VLSI Symp. Tech. Dig. 2009, pp. 16-17.

[3.19] K. Romanjek, F. Andrieu, T. Ernst and G. Ghibaudo, “Improved Split C-V Method for Effective Mobility Extraction in sub-0.1 µm Si MOSFETs,” IEEE Elec. Dev. Lett., Vol. 25, No. 8, pp. 583-585 (2004).

[3.20] A. Cros, K. Romanjek, D. Fleury, S. Harrison, R. Cerutti, et al., “Unexpected mobility degradation for very short devices: A new challenge for CMOS scaling,” IEEE Electron Dev. Meet. (IEDM) 2006, pp. 1- 4.

[3.21] M.V. Fischetti and S.E. Laux, “Long-range Coulomb interactions in small Si devices. Part I: Performance and reliability,” J. Appl. Phys., Vol. 89, No. 2, pp. 1205-1231 (2001).

[3.22] K. Nakanishi, T. Uechi and N. Sano, “Self-Consistent Monte Carlo Device Simulations under Nano-Scale Device Structures: Role of Coulomb interaction, Degeneracy, and Boundary Condition,” IEEE Electron Dev. Meet. (IEDM) 2009, pp. 79-82.

[3.23] M. Cassé, F. Rochette, L. Thevenod, N. Bhouri, F. Andrieu, et al., “A comprehensive study of magnetoresistance mobility in short channel transistors: Application to strained and unstrained silicon-on-insulator field-effect transistors. J. Appl. Phys., Vol. 105, No. 8, article 084503 (2009).

[3.24] M. Heiblum and L.F. Eastman, “Ballistic Electrons in Semiconductors,” Scientific American, Vol. 256, No. 2, pp. 64- 73, (1987).

[3.25] F. Bloch, “Uber die Quantenmechanik der Elektronen in Kristalgittern,” Zeitschrift fur Physik, Vol. 52, No. 7-8, pp. 555-600 (1928). (Note: This paper was in German. The title after translation into English is “Quantum mechanics of electrons in crystal lattices”.)

[3.26] E. Fuchs, P. Dollfus, G.L. Carval, S. Barraud, D. Villanueva, et al.,“A New Backscattering Model Giving a Description of the Quasi-Ballistic Transport in Nano-MOSFET,” IEEE Trans, Electron Dev., Vol. 52, No. 10, pp. 2280-2289 (2005).

[3.27] W. Lee, J.J.-Y. Kuo, W. P.-N. Chen, P. Su and M-C. Jeng,“ Impact of Uniaxial Strain on Channel Backscattering Characteristics and Drain Current Variation for Nanoscale PMOSFETs ,” IEEE Symp. VLSI Tech. Dig. 2009, pp. 112-113.

[3.28] 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 Electron Dev. Lett. , Vol. 6, No. 12, pp. 665-667 (1985).

[3.29] G. Ghibaudo and F. Balestra, "Low temperature characterization of silicon CMOS devices," Microelectron. Reliab., Vol. 37, No. 9, pp. 1353-1366 (1997).

[3.30] N. Yoshikawa, T. Tomida, M. Tokuda, Q. Liu, X. Meng, et al., "Characterization of 4K CMOS devices and circuits for hybrid Josephson CMOS systems," IEEE Trans. Appl. Supercond., Vol. 15, No. 2, pp. 267-271 (2005).

[3.31] X. Li, J. Tong and J. Mao, "Temperature-dependent Device Behavior in Advanced CMOS Technologies," Proceedings of the 2010 International Symp. on Signals, Systems and Electronics (ISSE), pp. 452-455, 2010.

[3.32] K.K. Ng, C.S. Rafferty and H.I. Cong, "Effective ON-current of MOSFETs for large-signal speed considerations," IEEE Electron Device Meet. (IEDM) 2002, pp. 121-124.

[3.33] J. Deng and H.-S. P. Wong, "Metrics for Performance Benchmarking of Nanoscale Si and Carbon Nanotube FETs Including Device Non-idealities," IEEE Trans. Electron Dev., Vol. 53, No. 6, pp. 1317 - 1322 (2006).

[3.34] K. Hisamitsu, H. Ueno, M. Tanaka, D. Kitamaru, M. Miura-Mattausch, et al., "Temperature-independence-point properties for 0.1 μm-scale pocket-implant technologies and the impact on circuit design," Proceedings of the Asia and South Pacific Design Automation Conference (ASP-DAC), pp. 179- 183 (2003).

[3.35] A. Lochtefeld, I.J. Djomehri, G. Samudra and D.A. Antoniadis, " New insights into carrier transport in n-MOSFETs," IBM J. Res. & Dev., Vol. 46, No. 2-3, pp. 347-357 (2002).

[3.36] M-J. Chen, H-T. Huang, Y-C. Chou, R.T. Chen, Y-Y. Tseng, et al., "Separation of channel backscattering coefficients in nanoscale MOSFETs," IEEE Trans. Electron Dev., Vol. 51, No. 9, pp. 1409-1415 (2004).

[3.37] M.O. Baykan, S.E. Thompson and T. Nishida, "Strain effects on three-dimensional, two-dimensional, and one-dimensional silicon logic devices: Predicting the future of strained silicon ," J. Appl. Phys., Vol. 108, 093716 (2010).

[3.38] O. Bonno, S. Barraud, D. Marioller and F. Andrieu, "Effects of strain on the electron effective mobility in biaxially strained silicon inversion layers: An experimental and theoretical analysis via atomic force microscopy measurements and Kubo-Greenwood mobility calculations," J. Appl. Phys., Vol. 103, 063715 (2008).

[3.39] J.-C. Guo, S.S.-S. Chung, C.C-H. Hsu, "A new approach to determine the effective channel length and the drain-to-source series resistance of miniaturized MOSFET's," IEEE Trans. Electron Dev., Vol. 41, No. 10, pp. 1811-1818 (1994).

[3.40] W. P-N. Chen, S. Pin and K-I. Goto, "Investigation of Coulomb mobility in nanoscale strained PMOSFETs, " IEEE Trans. on Nanotechnology, Vol. 7, No.5, pp. 538-543 (2008).

[3.41] S.Y. Chou and D.A. Antoniadis, "Relationship between measured and intrinsic transconductance of FET's," IEEE Trans. Electron Dev., Vol. 34, No. 2, pp. 448-450 (1987).

[3.42] J.G.J. Chern, P. Chang, R.F. Motta and N. Godinho, "A new method to determine MOSFET channel length," IEEE Electron Dev. Lett., Vol. 1, No. 9, pp. 170-173 (1980).

[4.1] W. Zhang and J.G. Fossum, "On the Threshold Voltage of Strained-Si-$Si_{1-x}Ge_x$ MOSFETs,"IEEE Trans. Electron Dev., Vol. 52, No. 2, pp. 263-268 (2005).

[4.2] J-S. Lim, S.E. Thompson and J.G. Fossum, "Comparison of Threshold-Voltage Shifts for Uniaxial and Biaxial Tensile-Stressed n-MOSFETs," IEEE Electron Dev. Lett., Vol. 25, No. 11, pp. 731-733 (2004).

[4.3] W. Yang, G. Qin, X. Shao, Z. Yu and L. Tian, "Analysis of GIDL Dependence on STI-induced Mechanical Stress," IEEE Conference Electron Dev. Solid-State Circ. (EDSSC) 2005, pp. 769-772.

[4.4] J. Lu, C. Duvvury, H. Gossner and K. Banerjee, "Impact of Strain Engineering and Channel Orientation on the ESD Performance of Nanometer Scale CMOS Devices," IEEE Electron Dev. Meet. (IEDM) 2009, pp. 413-416.

[4.5] C.Y. Cheng, Y.K. Fang, J.C. Liao, T.J. Wang, Y.T. Hou, et al., "The effects of STI induced mechanical strain on GIDL current in Hf-based and SiON MOSFETs," Solid-State Electronics, Vol. 53, pp. 892-896 (2009).

[4.6] X. Yang, Y. Choi, T. Nishida and S.E. Thompson, "Gate Direct Tunneling Currents in Uniaxial Stressed MOSFETs," IEEE Proceedings Int. Workshop on Electron Dev. and Semiconductor Tech. (IEDST) 2007, pp. 149-152.

[4.7] W. Zhao, A. Seabaugh, V. Adams, D. Jovanoic and B. Winstead, "Opposing Dependence of the Electron and Hole Gate Currents in SOI MOSFETs Under Uniaxial Strain," IEEE Electron Dev. Lett. , Vol. 26, No. 6, pp. 410- 412 (2005).

[4.8] X. Yang, J. Lim, G. Sun, K. Wu, T. Nishida, et al., "Strain-induced changes in the gate tunneling currents in p-channel metal-oxide-semiconductor field-effect transistors," Appl. Phys. Lett., Vol. 88, No. 5 , article 052108 (2006).

[4.9] X. Yuan, J-E. Park, J. Wang, E. Zhao, D.C. Ahlgren, et al., "Gate-Induced-Drain-Leakage Current in 45-nm CMOS Technology," IEEE Trans. Dev. Mater. Reliab., Vol. 8, No. 3, pp. 501-508 (2008).

[4.10] W. Yang, G. Qin, X. Shao, Z. Yu and L. Tian. "Analysis of GIDL Dependence on STI-induced Mechanical Stress," IEEE Conference on Electron Devices and Solid-State Circuiy (EDSSC) 2005, pp. 769-772.

[4.11] D.W. Snoke (2009). *Solid state physics: essential concepts*. Pearson Education. Addison-Welsey , USA, pp. 144-145.

[4.12] J-H. Zhao, R. Todd, P.S. Ho, A.J. Mckerrow and W-Y. Shih, "Measurement of elastic modulus, Poisson ratio, and coefficient of thermal expansion of on-wafer submicron," J. Appl. Phys., Vol. 85, No. 9, pp. 6421-6424 (1999).

[4.13] C-L. Dai and Y-M. Chang, "A resonant method for determining mechanical properties of Si_3N_4 and SiO_2 thin films," Mater. Lett., Vol. 61, pp. 3089-3092 (2007).

[4.14] M.A. Hopcroft, W.D. Nix and T.W. Kenny, "What is the Young's Modulus of Silicon?," J. Microelectromechanical Systems, Vol. 19, No. 2 , pp. 229-238 (2010).

[4.15] S.M. Sze and Kwok K. Ng (2007). *Physics of semiconductor devices*, 3rd ed., Wiley-Interscience, Hoboken, New Jersey, p. 56.

[4.16] F. Stern, "Self-Consistent Results for n-Type Si inversion Layers," Phys. Rev. B, Vol. 6, No. 12, pp. 4891-4899 (1972).

[4.17] C. Moglestue, "Self-consistent calculation of electron and hole inversion charges at silicon-silicon dioxide interfaces," J. Appl. Phys., Vol. 59, No. 9, pp. 3175-3183 (1986).

[4.18] Johnson, W.P, "The Curious History of Faà di Bruno's Formula". *American Mathematical Monthly* , Vol. 109, No. 3, pp. 217–234 (2009).

[4.19] R.W. Robinett (2006). *Quantum Mechanics: classical results, modern systems, and visualized examples*, 2nd edition, Oxford University Press Inc, US, New York, pp. 667 - 668.

[4.20] M. Abramowitz and I.A. Stegun (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover, pp. 446-448.

[4.21] T. Ando, A.B. Fowler and F. Stern, "Electronic properties of two-dimensional systems," Rev. Mod. Phys., Vol. 54, No. 2, pp. 437-672 (1982).

[4.22] A. Lubow, S. Ismail-Beigi and T.P. Ma, "Comparison of drive currents in metal-oxide-semiconductor field-effect transistors made of Si, Ge, GaAs, InGaAs, and InAs channels," Appl. Phys. Lett., Vol. 96, No. 12, article 122105 (2010).

[4.23] J.A. López-Villanueva, P. Cartujo-Casinello, J. Banquer, F. Gámiz and S. Rodríguez, "Effects of the Inversion Layer Centroid on MOSFET Behavior," IEEE Trans. Electron Dev., Vol. 44, No. 11, pp. 1915 –1922 (1997).

[4.24] N. Rodriguez, F. Gámiz and J.B. Roldán, "Modeling of Inversion Layer Centroid and Polysilicon Depletion Effects on Ultrathin-Gate-Oxide MOSFET Behavior: The Influence of Crystalline Orientation," IEEE Trans. Electron Dev., Vol. 54, No. 4, pp. 723-732 (2007).

[4.25] C-Y. Lu, H-C. Lin, Y-F. Chang and T-Y. Huang, "Device Characteristics and Aggravated Negative Bias Temperature Instability in p-Channel Metal-Oxide-Semiconductor Field-Effect Transistors with Uniaxial Compressive Strain," Jpn. J. Appl. Phys., Vol. 45, No. 4B, pp. 3064-3069 (2006).

[4.26] Y. Taur and T. Ning (1998). *Fundamentals of modern VLSI devices*. 1st ed, Cambridge University Press, New York, pp. 126-129.

[4.27] T. Khan, D. Vasileska and T.J. Thornton, "Subthreshold Electron Mobility in SOI MOSFETs and MESFETs," IEEE Trans. Electron Dev., Vol. 52, No. 7, pp. 1622-1626 (2005).

[4.28] C.W. Eng, W.S. Lau, D. Vigar, S.S. Tan and L. Chan, "Effective channel length measurement of metal-oxide-semiconductor transistors with pocket implant using the subthreshold current-voltage characteristics based on remote Coulomb scattering," Appl. Phys. Lett. , Vol. 87, No. 15, article 153510 (2005).

[4.29] Y.M. Meziani, J. Łusakowski, W. Knap, N. Dyakonova, F. Teppe, et al., "Magnetoresistance characterization of nanometer Si metal-oxide-semiconductor transistors," J. Appl. Phys., Vol. 96, No. 10, pp. 5761-5765 (2004).

[4.30] F. Rochette, M. Cassé, M. Mouis, G. Reimbold, D. Blachier, et al.,“ Experimental evidence and extraction of the electron mass variation in [110] uniaxially strained MOSFETs,” Solid-State Electronics, Vol. 51, pp. 1458-1465 (2007).

[4.31] A. Pouydebasque, C. Charbuillet, R. Gwoziecki, and T. Skotnicki, “Refinement of the Subthreshold Slope Modeling for Advanced Bulk CMOS Devices, ”IEEE Trans. Electron Devices, Vol. 54, No. 10, pp. 2723-2729 (2007).

[4.32] A.A. Mutlu and M. Rahman, “Two-dimensional analytical model for drain induced barrier lowering (DIBL) in short channel MOSFETs,” IEEE Conference Proceedings SoutheastCon 2000. `Preparing for The New Millennium', pp. 340-344.

[4.33] S.W. Crowder, P.M. Rousseau, J.P. Scott, P.B. Griffin, J.D. Plummer, “The effect of source/drain processing on the reverse short-channel effect of deep sub-micron bulk and SOI NMOSFETS,” IEEE Electron Device Meet. (IEDM) 1995, pp. 427- 430.

[4.34] C.S. Rafferty, H-H. Vuong, S.A. Eshraghi, M.D. Giles, M.R. Pinto, S.J. Hillenius, “Explanation of reverse short channel effect by defect gradients,” IEEE Electron Device Meet. (IEDM) 1993, pp. 311-314.

[4.35] R. Gwoziecki, T. Skotnicki, P. Bouillon, P. Gentil, “Optimization of Vth roll-off in MOSFET’s with advanced channel architecture-retrograde doping and pockets,” IEEE Trans. Electron Dev., Vol. 46, No. 7, pp. 1551- 1561 (1999).

[4.36] B. Yin, C.H.J. Wann, E.D. Nowak, K. Noda, C. Hu, “Short-channel effect improved by lateral channel-engineering in deep submicrometer MOSFET’s,” IEEE Trans. Electron Dev., Vol. 44, No. 4, pp. 627-634 (1997).

[4.37] T. Hori, “A 0.1 µm CMOS technology with tilt-implanted punchthrough stopper (TIPS),” IEEE Electron Device Meet. (IEDM) 1994, pp. 75-78.

[4.38] Y. Naitou, N. Ookubo, “Halo ion implantation effect on extension profile studied by scanning capacitance microscopy using all-metal probe under FM control,” Jpn. J. Appl. Phys., Vol. 43, No. 4B, pp. 1848- 1851 (2004).

[4.39] D. Fleury, A. Cros, K. Romanjek, D. Roy, F. Perrier, B. Dumont, et al., “Automatic extraction methodology for accurate measurements of effective channel length on 65 nm MOSFET technology and below,” IEEE Trans. Semi. Manuf., Vol. 21, No. 4, pp. 504-512 (2008).

[4.40] H. Iwai, T. Ohguro and S. Ohmi, “NiSi salicide technology for scaled CMOS,” Microelectronic Eng., Vol. 60, pp. 157-169 (2002).

[4.41] M.P. Hughey and R.F. Cook, “Massive stress changes in plasma-enhanced chemical vapor deposited silicon nitride films on thermal cycling,” Thin Solid Films, Vol. 460, pp. 7-16 (2004).

[4.42] R. J.E. Hueting and A. Heringa, “Analysis of the Subthreshold Current of Pocket or Halo-Implanted nMOSFETs,” IEEE Trans. Electron Dev., Vol. 53, No. 7, pp. 1641- 1646 (2006).

[4.43] N.G. Tarr, D.J. Wallkey, M.B. Rowlandson, S.B. Hewitt and T.W. MacElwee, “Short-channel effects on MOSFET subthreshold swing,” Solid-State Electron., Vol. 38, pp. 697 (1995).

[4.44] R.R. Troutman, "VLSI limitation from drain-induced barrier lowering," IEEE Journal of Solid-State Circuits, Vol. 14, p. 383 (1979),

[4.45] S. Orain, V. Fiori, D. Villanueva, A. Dray and C. Ortolland, "Method for Managing the Stress Due to the Strained Nitride Capping Layer in MOS Transistors," IEEE Trans. Electron Dev., Vol. 54, No. 4, pp. 814 - 821 (2007).

[5.1] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon, F. Boeuf, S. Monfray, et al., "Innovative Materials, Devices, and CMOS Technologies for Low-Power Mobile Multimedia," IEEE Trans. Electron Dev., Vol. 55, No. 1, pp. 96- 130 (2008).

[5.2] A. Rahman, M.S. Lundstrom and A.W. Ghosh, "Generalized effective-mass approach for n-type metal-oxide-semiconductor field-effect transistors on arbitrarily oriented wafers," J. Appl. Phys., Vol. 97, No. 5, 053702 (2005).

[5.3] R.F. Pierret (2003). *Advanced Semiconductor Fundamentals*, Modular Series on Solid State Devices vol. VI, 2nd ed., Prentice Hall New Jersey, pp.77-79.

[5.4]S. Solmi, M. Ferri, M. Bersani, D. Giubertoni and V. Sonicini, "Transient enhanced diffusion of arsenic in silicon, " J. Appl. Phys. , Vol. 94, No. 8, pp. 4950-4955 (2003).

[5.5] K. Inoue, F. Yano, A. Nishida, H. Takamizawa, T. Tsunomura, et al., "Dopant distribution in n-MOSFET structure observed by atom probe tomography," Ultramicroscopy, Vol. 109, No. 12, pp. 1479-1484 (2009).

[5.6] S. Boninelli, S. Mirabella, E. Bruno, F. Priolo, F. Cristiano, et al., "Evolution of boron-interstitial cluster in crystalline Si studied by transmission electron microscopy," Appl. Phys. Lett. , Vol. 91, No. 3, article 031905 (2007).

[5.7] M. Okamura, "The Orientation Dependence of Boron Diffusion," Jpn. J. Appl. Phys., Vol. 9, pp. 848-849 (1970).

[5.8] G. Masetti, S. Solmi and G. Soncini, "Temperature Dependence of boron diffusion in (111), (110) and (100) silicon," Solid-State Electron, Vol. 19, pp. 545-546 (1976).

[5.9] O.V. Aleksandrov and N.N. Afonin, "Effect of Oxidizing Environments on the Diffusion-Segregation Boron Distribution in the Thermal Silicon Dioxide-Silicon System," Technical Physics, Vol. 48, No.5, pp. 580-586 (2003). Translated from Zhurnal Tecknichekoĭ Fiziki, Vol. 73, No.5, 2003, pp. 57-63. Original Russian Text Copyright © 2003 by Aleksandrov, Afonin.

[5.10] S.M. Hu, "Formation of stacking faults and enhanced diffusion in the oxidation of silicon," JAP, Vol. 45, no.4, pp. 1567-1569 (1974).

[5.11] Y. Mitani and A. Toriumi, "Re-consideration of Influence of Silicon Wafer Surface Orientation on Gate Oxide Reliability from TDDG Statistics Point of View," IRPS 2010, pp.299-305.

[5.12] M. Miyake and H. Harada,",Diffusion and Segregation of Low-Dose Implanted Boron in Silicon under Dry 0_2 Ambient," J. Electrochem. Soc., Vol. 129, No.5, pp. 1097-1103 (1982).

[5.13] L. Shao, J. Liu, Q.Y. Chen and W-K. Chen, "Boron diffusion in silicon: the anomalies and control by point defect engineering," Mater. Sci. Eng. R., Vol. 42, pp. 65-114 (2003).

[5.14] S. C. Jain, W. Schoenmaker, R. Lindsay, P.A. Stolk, S. Decoutere, et al., “Transient enhanced diffusion of boron in Si,” J. Appl. Phys., Vol. 91, pp.8919-8941 (2002).

[5.15] R. Simonton and Al. F. Tasch (1992.) “Channeling effects in ion implantation,” In *Handbook of Ion Implantation Technology*, edited by J.F. Ziegler, Elsevier Science Publisher, North Holland, New York, p. 138.

[5.16] R. Kim, T. Hirose, T. Shano, H. Tsuji and K. Taniguchi, “Influences of point and extended defects on As diffusion in Si,”Jpn. J. Appl. Phys., Vol. 41, pp. 227-231 (2002).

[5.17] T.C. Chan and C.C. Mai,” Diffusion of boron, phosphorus, arsenic and antimony into (100) and (111) silicon slices,” IEEE Proceedings, Vol. 58, No. 4, pp. 588-589 (1970).

[5.18] R. Simonton and Al. F. Tasch (1992.) “Channeling effects in ion implantation.” In *Handbook of Ion Implantation Technology*, edited by J.F. Ziegler, Elsevier Science Publisher, North Holland, New York, pp. 122-134.

[5.19] G. Hobler, “Monte Carlo simulation of two-dimensional implanted dopant distributions at mask edges,”Nucl. Instr. Meth. B, Vol. 96, pp. 155-162 (1995).

[5.20] C. Tian, S. Gara, G. Hobler and G. Stingeder, “Dependence of Boron Axial Channelling in Silicon on Crystal Orientation”, Surf. Interface Anal., Vol.19, pp. 369-373 (1992).

[5.21] A. Paul, V. Savikhin, G. Klimeck, K. Hoganson (2007), "Crystal Viewer Tool," DOI: 10254/nanohub-r3741.12. (DOI: 10254/nanohub-r3741.12).

[5.22] C. Park, K.M. Klein and Al F. Tasch, “Critical Angles for Channeling of Boron Ions Implanted into Single-Crystal Silicon”, J. Electrochem. Soc., Vol. 138, No. 7, pp. 2107- 2115 (1991).

[A.1] K. Uchida, M. Saitoh and S. Kobayashi, “Carrier Transport and Stress Engineering in Advanced Nanoscale Transistors From (100) and (110) Transistors to Carbon Nanotube FETs and Beyond,” IEEE Electron Device Meet. (IEDM) 2006.

[A.2] S. Masumi, Y. Nobuaki, N. Yukio, N. Toshinori and U. Ken, “Comprehensive performance assessment of scaled (110) CMOSFETs based on understanding of STI stress effect and velocity saturation, ” IEEE Electron Dev. Meet. (IEDM) 2008, pp. 1- 4.