Copyright

by

Xiaofeng Fan

2006

The Dissertation Committee for Xiaofeng Fan certifies that this is the approved version of the following dissertation:

Quantum Corrected Full-band Semiclassical Monte Carlo Simulation Research of Charge Transport in Si, Stressed-Si, and SiGe MOSFETs

Committee

____

Sanjay K. Banerjee, Supervisor

____

Leonard F. Register, Co-Supervisor

____

Ananth Dodabalapur

____

Joe C. Campbell

____

Graham F. Carey

____

Peter Zeitzoff

____

Quantum Corrected Full-band Semiclassical Monte Carlo Simulation Research of Charge Transport in Si, Stressed-Si, and SiGe MOSFETs

by

Xiaofeng Fan, B.S., M.S.E

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2006

To my wife Yanjie

and parents, Jianxin Ma and Jifu Fan

Acknowledgements

I would like to thank my supervisor, Dr. Sanjay K. Banerjee, for his encouragement, patience and stimulating guidance. In the past five years, I have been impressed by his wide knowledge, strong curiosity to acquire new knowledge and enthusiasm for research. He gave me great academic freedom in my research, which helped me gradually establish independent research ability. It is very important for my academic training as a graduate student.

I also like to thank my Co-Supervisor, Dr. Leonard F. Register. I am very impressed by his knowledge and experience in research, enthusiasm towards research and generous, friendly relationship with students. His supervision played an irreplaceable role in my work. I gained a lot of knowledge and experience from stimulating and fruitful discussions with him.

I would like to express my sincere appreciation to some special colleagues and mentors, Dr. Xin Wang, Dr. Brian A. Winstead, and Dr. Mark C. Foisy. The invaluable guidance and discussions in my graduate study finally leads to pleasing and fruitful collaboration.

Lots of colleagues and friends have given me help and shown me a good time during my time at the Microelectronics Research Center. My thanks go to Wanqiang Chen, Yang-yu Fan, Fei Li, Li Lin, Xin Zheng, Ken Liu, Yueran Liu, Joseph Donnelly, Bahniman Ghosh, Ning Kong, Mehedi, Ningyu Shi, and Matthew Gilbert.

My gratitude also goes to my doctoral committee. Finally, I am deeply indebted to my parents and wife. Their consideration, support and understanding are the most powerful motivations for my research.

Xiaofeng Fan

The University of Texas at Austin
October 2006

Quantum Corrected Full-band Semiclassical Monte Carlo Simulation Research of Charge Transport in Si, Stressed-Si, and SiGe MOSFETs

Publication No. _____________

Xiaofeng Fan, Ph.D.
The University of Texas at Austin, 2006

Supervisor: Sanjay K. Banerjee
Co-Supervisor: Leonard F. Register

This Ph.D. research is centered around a full-band Monte Carlo device simulator ("Monte Carlo at the University of Texas", MCUT) with quantum corrections (based on one-dimensional Schrödinger equation solver). The code itself was based on a solid infrastructure of a Monte Carlo simulator, "MoCa" from the University of Illinois at Urbana-Champaign. To that there were added new methods and features during my Ph.D. program, including strained band structures, alternative (to conventional $\langle 100 \rangle$) surface orientations, full-band scattering mechanisms, and valley-dependent quantum correction. These

features enable “MCUT” to be used to model various strained and/or alloyed silicon MOSFETs, as well as the MOSFETs composed of alternative materials such as Ge, in sub-100 nm regime. Monte Carlo simulation, itself, handles short channel effects and hot carriers in ultra small device well; full-band structure replaces the inaccurate and unknown (for new/strained materials) analytical formulae; and the quantum corrections approximate quantum-confinement effects on device performance. The goal is to understand and predict the device behavior of the so called “non-classical” CMOS — beyond bulk Si based CMOS — in the sub-100 nm regime.

Table of Contents

1 Research Background 1

1.1 Full-band Semiclassical Monte Carlo Simulation 1

1.2 Strained Materials for Semiconductor Devices 3

1.3 Quantum Correction in Semiclassical Monte Carlo Method 5

1.4 Monte Carlo at the University of Texas (MCUT) 6

2 Tensily Biaxially stressed Si nMOS 7

2.1 Motivation 7

2.2 Strain Effects 9

2.3 Valley-edge Quantum Corrections 13

2.4 Full-band Quantum Corrections via “Effective Strain” 18

2.5 Transport within a Spatially-varying Band Structure 21

2.6 Model Devices 22

2.7 Mobility Calculations 24

2.8 Illustrative High-field and Deeply Scaled Device Results 29

2.9 Scalability of TSSi nMOS on Technology Roadmap 32

2.10 Conclusion 38

3 Uniaxially Stressed Si CMOS 39

3.1 Motivation 39

3.2 Band Structures 40

3.3 Bulk Mobility 47

3.4 Discussion 50

4 PMOS Wafer Crystal Orientation 52

4.1 Motivation 52

4.2 Band Structure Change under Confinement 54

4.3 Confinement to Stress Mapping 56

5 Optimum Combination of Stress and Crystal Orientation 61

5.1 Position Dependent Stress 61

5.2 Simulations with Position-Dependent Stress.. 63

5.3 Combination of Stress and Quantum Confinement............................. 65

6 Dual Gate NMOS Simulation 69

6.1 Alternative quantum correction scheme.. 69

6.2 MCUT simulations with dual gate Si nMOS 73

7 Injection Efficiency with Bandgap Engineering 76

7.1 Motivation 76

7.2 Simulation Results.. 77

8 Conclusions and recommendations 79

8.1 Conclusions 79

8.2 Recommendations for future research.. 82

References 84

Vita 92

List of Figures

Figure 1-1: The lattice constant of SiGe is higher than that of Si. When Si is grown epitaxially on SiGe, it stretches to conform to the atomic spacing of SiGe which puts it under biaxial tensile strain. 3

Figure 1-2: A cross-sectional electron micrograph of a strained Si transistor fabricated by UMC. The fabrication sequence used 130 nm node Si technology, and the physical gate length for the transistor shown is 90 nm. Credit: UMC. Image from www.compoundsemiconductor.net. 4

Figure 1-3: TEM micrograph of 45 nm p-type and n-type MOSFET. Uniaxial stress is introduced in the channel, by either $Si_{1-\gamma}Ge_{\gamma}$ epitaxially grown source/drain or high stress overlayer film. Image from [Tho04]. 5

Figure 2-1: (a) Schematic equi-energy surfaces of the six split valleys in the first conduction band in TS-Si. The minimum energies of the dark valleys are lower than those of the open ellipsoids. (b) Schematic energy diagrams of heavy hole (HH) and light hole (LH) bands in TS-Si. 8

Figure 2-2: Δ_2 and Δ_4 valley splitting vs. Ge mole fraction in $Si_{1-\gamma}Ge_{\gamma}$ buffer. Note the essentially linear dependence. 9

Figure 2-3: Drift velocity of electrons in bulk Si as a function of electric field. For comparison, the other measurements and simulation results are also shown. [Jac83][Can75][Fis88] 11

Figure 2-4: Average energy of electrons in bulk Si as a function of electric field. For comparison, the other measurements and simulation results are also shown. [Jac83][Fis88] 12

Figure 2-5: Low-field mobility vs. Ge mole fraction in $Si_{1-x}Ge_x$ buffer. Note the saturation of mobility for increasing strain as the valley-splitting-induced transfer to the Δ_2 valleys is completed for these low-energy carriers. 12

Figure 2-6: Eigenvalues of the energy well for longitudinal and transverse valleys 17

Figure 2-7: Quantum mechanical electron distribution density per valley of Δ_2 and Δ_4 valleys as a function of depth below interface. 18

Figure 2-8: Effects of quantum corrections alone on the Δ_2 and Δ_4 valley-minimum energies for unstrained Si exhibiting position dependent additional valley splitting. 19

Figure 2-9: An illustration of the calculation of the effective strain/mole fraction γ_{eff} (at one position) from the quantum-corrected valley-minimum splitting, using the strain vs. valley splitting results of Figure 2-2. 20

Figure 2-10: Effects of both quantum correction and tensile strain on the Δ_2 and Δ_4 valley-minimum energies for TSSi grown on $Si_{1-\gamma}Ge_\gamma$ buffer. 20

Figure 2-11: An illustration of the calculation of the effective strain/mole fraction γ_{eff} (at one position) starting with the real strain-induced valley splitting and then adding quantum-correction induced valley splitting to obtain a total valley splitting and corresponding new effective strain. 21

Figure 2-12: 2D device doping profile. 23

Figure 2-13: Doping density along cross-section normal to gate at mid-channel. 24

Figure 2-14: Drain current vs. gate voltage vs. normal effective field in MOSFET with TSSi channel on relaxed $Si_{1-\gamma}Ge_{\gamma}$ buffer. γ is the Ge mole fraction in the buffer layer such that $\gamma = 0$ denotes unstrained silicon. 26

Figure 2-15: Channel electron mobility vs. normal effective field in MOSFET with TSSi channel on relaxed $Si_{1-\gamma}Ge_{\gamma}$ buffer. γ is the Ge mole fraction in the buffer layer such that $\gamma = 0$ denotes unstrained silicon. For TSSi with buffer of $Si_{0.8}Ge_{0.2}$ and $Si_{0.7}Ge_{0.3}$, the mobility enhancements are very close as the carrier transfer to the Δ_2 valleys saturates. 27

Figure 2-16: Results of MCUT device simulation are compared with other simulation results [Rol96] and the universal mobility curve [Tag94]. Open triangles: MCUT MOSFET simulation on unstrained Si; solid squares: MCUT MOSFET simulation on TSSi. 28

Figure 2-17: MCUT TSSi MOSFET (2D) simulations with and without a reduced rms amplitude of the Si/SiO_2 interface roughness are compared to experiment data [Wel94][Rim00]. Unstrained Si universal mobility data [Tag94] are plotted as reference. Solid circle: MCUT TSSi MOSFET simulation with reduced surface roughness scattering rate; solid square: MCUT TSSi MOSFET simulation with fixed surface roughness scattering rate; open triangle: MCUT unstrained Si MOSFET simulation. 29

Figure 2-18: I_d-V_d curve. Drain currents as a function of drain voltage for nMOSFETs with TSSi ($\gamma = 0.2$) channel and an unstrained Si ($\gamma = 0.0$) channel. Closed circles: unstrained Si; closed squares: TSSi with

fixed surface roughness parameters; open diamonds: TSSi with reduced surface roughness scattering consistent with Figure 2-17. For the same $V_G - V_T$, enhancement of the drive current is observed in the TSSi. 30

Figure 2-19: Average carrier transport velocity along the channel for fixed V_D and $V_G - V_T$. 31

Figure 2-20: Average carrier kinetic energy distribution along the channel for fixed V_D and $V_G - V_T$. Comparison is made between a TSSi channel with reduced surface roughness scattering and a relaxed Si channel. Increases are observed in both average energy and velocity within the channel for TSSi as compared to relaxed Si. 31

Figure 2-21: I_d-V_d curve for a 50 nm channel length well-tempered device after Ref. [Mit00]. Drain currents as a function of drain voltage for nMOSFETs with TSSi ($\gamma = 0.2$) channel and an unstrained Si ($\gamma = 0.0$) channel. Closed circles: unstrained Si; closed squares: TSSi with fixed surface roughness parameters; open diamonds: TSSi with reduced surface roughness scattering consistent with Figure 2-17. For the same $V_G - V_T$, (V_T's are different in these three cases) enhancement of the drive current is observed in the TSSi, although perhaps a bit less than the longer channel devices for this isolated example. 32

Figure 2-22: Channel surface doping in strained Si and bulk Si nMOSFETs. 35

Figure 2-23: Average gate field in strained Si and bulk Si nMOSFETs. 35

Figure 2-24: Energy splitting between Δ_2 and Δ_4 valleys in SSi and bulk Si

channels. 36

Figure 2-25: Low-field bulk mobility plotted as a function of energy splitting between Δ_2 and Δ_4 valleys. 36

Figure 2-26: Drain current comparison between Si and SSi nMOSFETs. Lines represent MCUT simulations in "V_T-matched" and "process-matched" conditions. Experimental results come from: a. Ref [Xia03], b. Ref [Rim02a], c. Ref [Lee02], d. Ref [Hwa03], the rest from Ref [Goo03]. 37

Figure 3-1: Heavy hole and light hole band splitting in uniaxially-stressed Si. Here and for subsequent figures the positive stress axis corresponds to tensile stress, and the negative stress axis to compressive stress. 41

Figure 3-2: Heavy-hole and light-hole energy-K dispersion in [110] and $[1\bar{1}0]$ *directions with compressive [110] uniaxial-stress of 500 MPa.* 43

Figure 3-3: 2-D heavy-hole iso-energy contours (10 meV, 50 meV and 100 meV) for [110] uniaxially-stressed Si in the (001) plane with a compressive stress of 500 MPa in the [110] direction. 43

Figure 3-4: Percentile change in the thermal average of the inverse effective mass in the [110] and $[1\bar{1}0]$ *directions for [110] uniaxially-stressed Si as function of stress. The zero stress value is 2.79 times the reciprocal of the free space mass. For reference, the average inverse effective mass value in the [001] direction in unstrained Si is 2.80.* 45

Figure 3-5: Percentile change of the thermal velocity in the [110] and $[1\bar{1}0]$ *directions for [110] uniaxially-stressed Si as function of stress. The*

unstrained Si thermal velocity is 8.9×10^6 cm/s at room temperature. 46

Figure 3-6: Percentile change of the thermal average of inverse scattering rates (proportional to free flight time FFT) for [110] uniaxially-stressed Si as function of stress. The zero stress FFT (phonon scattering only) is ~0.15 ps at room temperature. 47

Figure 3-7: Low field mobility of uniaxially-stressed Si vs. stress strength. Dashed lines follow Smith's piezoresistance coefficient [Smi54]. Squares and diamonds plot MCUT bulk simulation results for transport parallel and perpendicular to stress directions, respectively. Experimental results from [Gil04][Tho04][Shi04] and simulation results from [Wan04] are plotted as triangles for comparison. Results are shown on two scales with the lower stress region shown in detail in right hand side. 50

Figure 4-1: The hole subband energy levels, as measured from the surface potential, formed in a triangular well in (left): (100) and (right): (110) surface orientations. The symbols are calculated value; the connecting lines only a guide to the eye. 55

Figure 4-2: The first subband E-k dispersion in $k_{//}$ plane for (left): (100) and (right): (110) surface orientations separately. 56

*Figure 4-3: The first band E-**k** dispersion in $k_{//}$ plane. (left) [100] uniaxial tensile stress, (100) plane; (right) [110] uniaxial tensile stress, (110) plane, spanned by [001] and $\left[1\bar{1}0\right]$ axes.* 58

Figure 4-4: Band splitting between the first and second hole bands induced by (left) [001] gate confinement field; and (right) [001] uniaxial stress.

59

Figure 4-5: [110] Hole transport mobility in [100] and [110] directions in (001) surface orientation. Curves show percentile change with [001] uniaxial tensile stress. 59

Figure 4-6: Band splitting between the first and second hole bands induced by (left) [110] gate confinement field; and (right) [110] uniaxial stress. 60

Figure 4-7: [110] transport mobility percentile change with [110] uniaxial stress. 60

Figure 5-1: Stress along the channel in a strained Si pMOS transistor, modeled by ISE04 [Ran05] 62

Figure 5-2: Drain current enhancement as a function of channel maximum stress. "High current" simulation was achieved with gate bias -1.5 V and drain bias -0.5 V; "medium current" simulation was with gate bias -1.2 V and drain bias -0.2 V; and "low current" simulation was with gate bias of -1.2 V and drain bias -50 mV. Source and substrate were grounded. 64

Figure 5-3: 2-D energy contours of heavy-hole band (or in other term, first subband),for a [001] confinement field of 0.1 MV/cm, 0.5 MV/cm, and 1.0 MV/cm separately, from left to right, in energy separation of 5 meV, from band edge upto 100 meV. The range of k is from -0.1 $Å^{-1}$ to 0.1 $Å^{-1}$. At different confinement field (0.1, 0.5 and 1.0 MV/cm), the band shape changes but a little. 66

Figure 5-4: This plotting shows the above figure, the E-k dispersion, in

[110] *and* $[1\bar{1}0]$ *direction, with individual lines represent different* [001] *confinement fields of 0.1, 0.5 and 1.0 MV/cm. The range of k is from -0.1 Å^{-1} to 0.1 Å^{-1}, with energy unit in eV. The change is rather small, which implies that uniaxial stressed Si behaves similarly in strong confinement as in bulk.* 66

Figure 5-5: Comparison between (110) orientation confinement effect and uniaxial stress effect. In the left hand side case, there is no stress; a field of 1.0 MV/cm is applied in gate (110) direction. In the right hand side, the confinement field is small 0.1 MV/cm, but a compressive uniaxial stress of 3.0 GPa is applied on transport direction. The contours show the first sub-band (heavy-hole) from band edge to 100 meV, in steps of 5 meV. 67

Figure 6-1: Phonon scattering rates as function of carrier energy in confined channels. Scattering rates form steps with sub-band levels, while they increase with confinement. 70

Figure 6-2: Replot scattering rates with energy reference fixed to bulk band edge at no confinement. 71

Figure 6-3: With confinement, when valleys/bands become non-degenerate, intra- and inter-valley phonon scattering rates are function of "total energy", while some inter-valley scattering becomes energy-prohibited. This effectively alternates valley/band concentration in simulations. 72

Figure 6-4: Valley population in Δ_2 and Δ_4 valleys in Si dual gate MOSFET body. t_{body} shrinks from 30 nm down to 2 nm. The solid red lines represent Δ_2 valley concentration and the green dashed lines for Δ_4

valleys. 74

Figure 6-5: Δ_2 and Δ_4 valley subband level in Si dual gate MOSFETs, with the energy denoted by $\Delta E1$ and $\Delta E2$. 75

Figure 6-6: Phonon-limited electron mobility in Si dual gate MOSFETs. The confinement enhanced surface roughness scattering is not included here. 75

Figure 7-1: Vertical pMOS structure with strained-SiGe grown on top of Si. Note the Ge mole fraction in SiGe forms steps from source to drain. 77

Figure 7-2: Injection efficiency comparison between a device with Ge mole fraction steps and a control device. Note the peaks represent the Ge mole fraction steps in channel. 78

1 Research Background

1.1 Full-band Semiclassical Monte Carlo Simulation

The Monte Carlo method is a widely used stochastic approach for studying charge transport in semiconductors and semiconductor devices. It involves the simulation of the motion of charge carriers subject to a wide variety of scattering mechanisms and external forces. Although it requires appreciable computer resources, the simplicity of implementation and relatively complete description of semi-classical transport make the Monte Carlo method appealing for semiconductor simulation. It is the most accurate approach to treat transport at high fields and in short, non-classical devices. The Monte Carlo simulation consists of iterations of free flight, scattering-type selection and post-scattering momentum selection for an ensemble of particles. According to the band structure used in the Monte Carlo simulation of electron transport, there are three types of implementation. The most computationally efficient is a low-energy model using a single parabolic or non-parabolic band [Jac83] that is only applicable to those electrons with energy below 0.5 eV. The second class uses a hybrid model [Vog92] [Wan93] to reduce the computation time. The third class uses a full band structure [Fis88] [Fan04], which is the most time-consuming and has the least amount of parameter tuning. However, the high costs and risks of novel silicon technology development make it imperative to employ simulation [Jac83]. Among the existing classical device simulation methods, the full-band semiclassical Monte Carlo (below noted as “MC” for full-band semiclassical

Monte Carlo simulation) technique is more predictive and can provide a more complete understanding of carrier transport than drift-diffusion or hydrodynamic transport models [Jac83]. MC has the ability to accurately model the transport of high energy carriers and to consider non-local field effects such as ballistic transport and velocity overshoot that have become increasingly important as devices are scaled down below 0.1 μm channel lengths for production devices. MC also offers a combination of computational efficiency and ability for incorporation of diverse scattering processes not (yet at least) available in (full) quantum transport simulators. Semi-classical transport along the channel within sub-bands could be explicitly considered, as for the calculation of mobilities in [Fis02]. However, unlike for near-equilibrium channel mobility calculations, for far from equilibrium transport calculations the associated computational burden would be much greater still for re-calculating the transition rates among the quantum confined states defined by sub-band and interface parallel momentum for each position along the channel and each time step/iteration self consistently with the scattering-dependent charge distribution [Kri06]. Furthermore, again unlike for mobility calculations, device simulations must often be repeated for varying *drain* voltages. And the extension for hot-carrier transport would still be questionable. Therefore, the exploration of future MOSFET designs demands, or at least would greatly benefit from, a full-band semiclassical MC simulation tool capable of — in addition to the many capabilities of conventional full-band MC — considering multi-material systems and the *combined* effects of strain and quantum confinement on degeneracy breaking among energy valleys, spatial charge re-distribution and the associated effects on transport.

1.2 Strained Materials for Semiconductor Devices

Strain and bandgap engineering of strained materials has emerged as an important technique for improving the device performance due to impediments to conventional scaling method. For example, Tensile-strained silicon (TSSi) on a relaxed SiGe buffer (*Figure 1-1*) has been studied extensively because of its enhanced electron and hole mobility which makes it a good candidate for n and p MOSFETs [Mii91] [Nel93] (*Figure 1-2*).

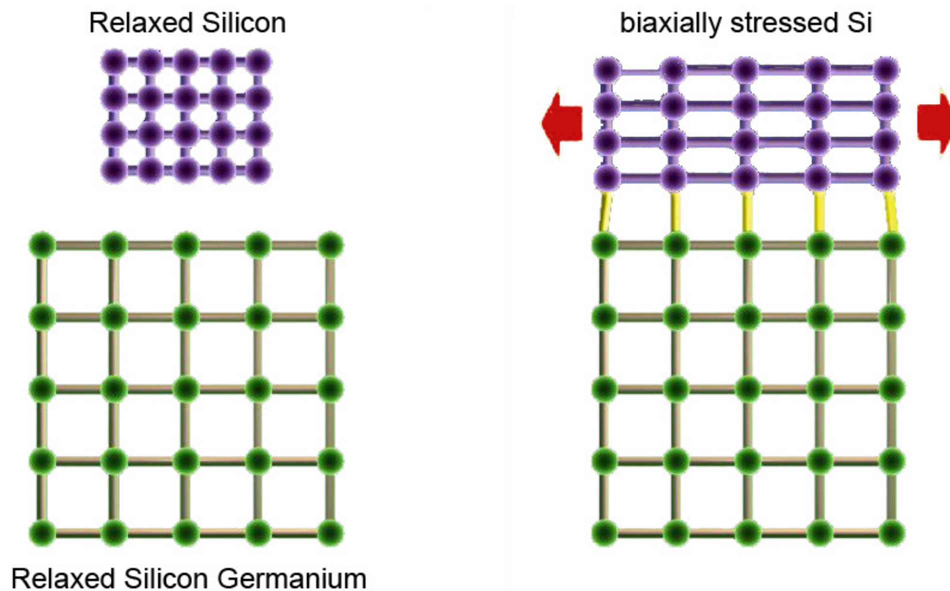

Figure 1-1: The lattice constant of SiGe is higher than that of Si. When Si is grown epitaxially on SiGe, it stretches to conform to the atomic spacing of SiGe which puts it under biaxial tensile strain.

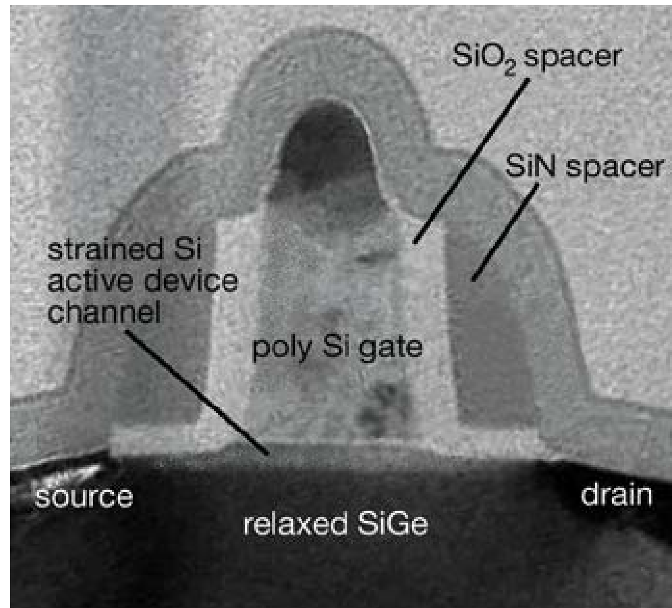

Figure 1-2: A cross-sectional electron micrograph of a strained Si transistor fabricated by UMC. The fabrication sequence used 130 nm node Si technology, and the physical gate length for the transistor shown is 90 nm. Credit: UMC. Image from www.compoundsemiconductor.net.

Recent experimental and theoretical work points to uniaxially-stressed Si as a possible channel material for next generation CMOS [Tho05] (*Figure 1-3*). Experimentally, uniaxial compressive stress has been applied to the channel area via source/drain hetero-epitaxial deposition of SiGe [Tho04, Gha03, Chia04], or by tensile/compressive capping layers on top of the gate [Arg04], while four-point or ring wafer bending is the main technique for quantitatively experimentally measuring piezoresistance [Bea92]. Greater than 50% stress-induced hole channel mobility improvement and 10% to 25% drain current enhancement has been demonstrated in strained Si devices for stress ranging from 100 to 500 MPa. Furthermore, although the hole mobility is often modeled as depending linearly on stress according to tabulated piezoresistance coefficients such as those of Ref. [Smi54], recent works have exhibited superlinear relationships between bulk

mobility and stress [Shi04, Wan04] suggesting even greater potential advantages in the high stress regime.

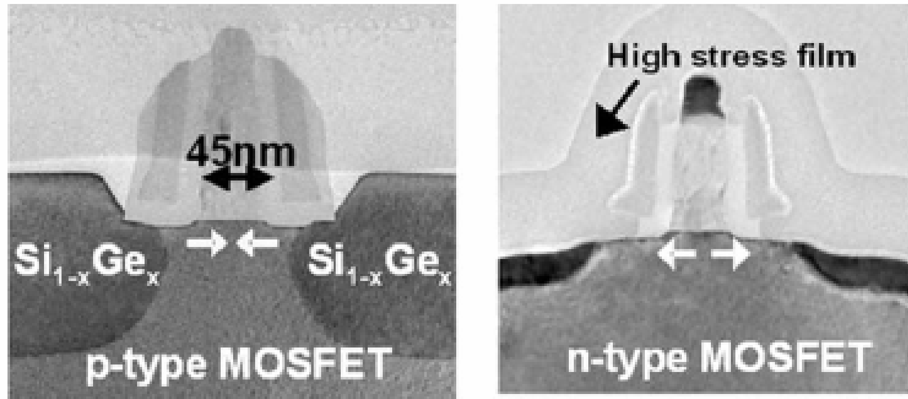

Figure 1-3: TEM micrograph of 45 nm p-type and n-type MOSFET. Uniaxial stress is introduced in the channel, by either $Si_{1-\gamma}Ge_{\gamma}$ epitaxially grown source/drain or high stress overlayer film. Image from [Tho04].

1.3 Quantum Correction in Semiclassical Monte Carlo Method

The MC methods used in this PhD research is a full-band Monte Carlo plus gate-directional quantum correction. The quantum correction is based on one-dimensional effective mass Schrödinger equation solved in the gate direction in all positions through the device. The combined Poisson's equation and Schrödinger's equation provide (real and virtual) driving electrostatic forces. The quantum confinement in inversion layers of a MOSFET is of special importance. Based on this established method, new features were added for situations encountered in non-classical CMOS device simulations. Details of these quantum correction methods in this MC research will be discussed in following sections.

1.4 Monte Carlo at the University of Texas (MCUT)

The device simulation group led by Dr. Banerjee and Dr. Register in the Microelectronics Research Center of The University of Texas at Austin has developed a MC tool named "Monte Carlo at the University of Texas" (MCUT). This code borrows the robust framework of the University of Illinois Urbana-Champaign code MoCa [Win01] [Win03], and adds the vast and varied physics necessary to treat strain and valley-dependent quantum confinement to achieve the combined effects of quantum confinement and strain within a full-band structure treatment. The code provides simulation capability for both electrons and holes devices, and for both conventional devices and "non-classical" CMOS devices such as FinFETs. The following sections will focus on MUCT simulations on biaxially stressed Si nMOS, uniaxially stressed Si pMOS, and SiGe concept vertical pMOS. The enhanced features and capabilities of MCUT are covered in each section when they are applied to the simulations.

2 Tensily Biaxially stressed Si nMOS

2.1 Motivation

The advantages of biaxially tensile-stressed Si (TSSi), as achieved by growing strained Si on relaxed SiGe, have been briefly described in Section 1.2. It is argued that the electron mobility enhancement is due to the strain-induced splitting of the six-fold degeneracy of silicon conduction band minima; while the hole mobility enhancement is from the splitting of heavy-hole and light-hole energy bands [Obe98] (*Figure 2-1*). The strain-induced degeneracy breaking in the electron band structure of the TSSi that has been well characterized [Wan01] as shown in *Figure 2-2*. In particular, the six lowest, Δ energy valleys in the conduction band separate into two groups: two valleys with the valley centers $\mathbf{k}_l$ perpendicular to the plane of the silicon/oxide interface and to the direction of transport, (in other words, $\mathbf{k}_l$ is parallel to the strained-Si growth direction) $l=\Delta_2$, and four higher energy in-plane valleys, $l=\Delta_4$. The strain-dependent energy difference $E^{(\gamma)}(\mathbf{k}_{\Delta 4}) - E^{(\gamma)}(\mathbf{k}_{\Delta 2})$ leads to segregation of low energy carriers toward the Δ_2 valleys that have the lower Si "transverse" effective mass (of ~0.19 m_e) in the direction of transport. This energy difference also reduces the densities of final states for low energy carriers, reducing the scattering rates accordingly [Wan93] [Wan01]. The mobility/drain current enhancement has been reported to be in the range of 50% to 110%, subject to different channel lengths, strain strength, and other device fabrication conditions [Wel94] [Rim98] [Rim02] [Hoy02]. Analyzing these experimental results, it should be noted that channel

length (thus short channel effect and hot carriers) and device structure played important roles, making the MC method preferable to mobility calculations based on effective mass or more accurate **k•p** sub-band structure.

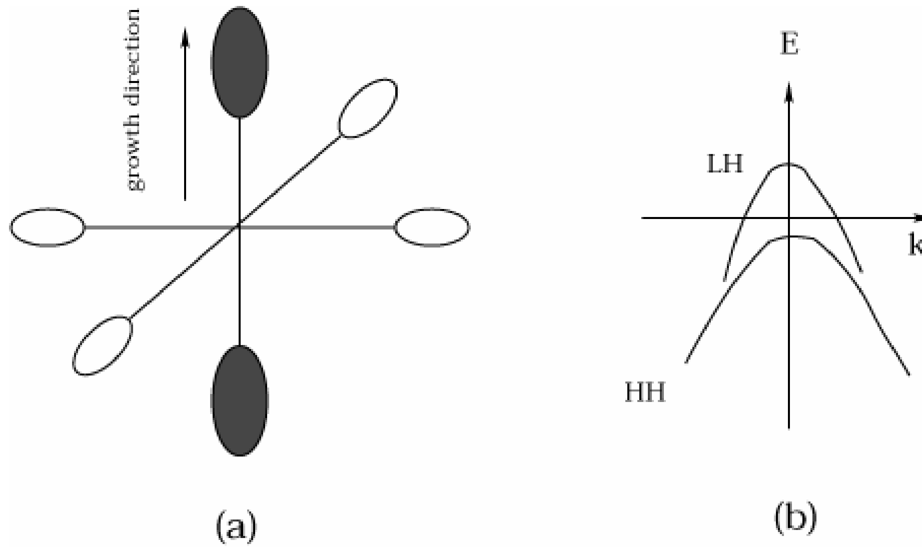

Figure 2-1: (a) Schematic equi-energy surfaces of the six split valleys in the first conduction band in TS-Si. The minimum energies of the dark valleys are lower than those of the open ellipsoids. (b) Schematic energy diagrams of heavy hole (HH) and light hole (LH) bands in TS-Si.

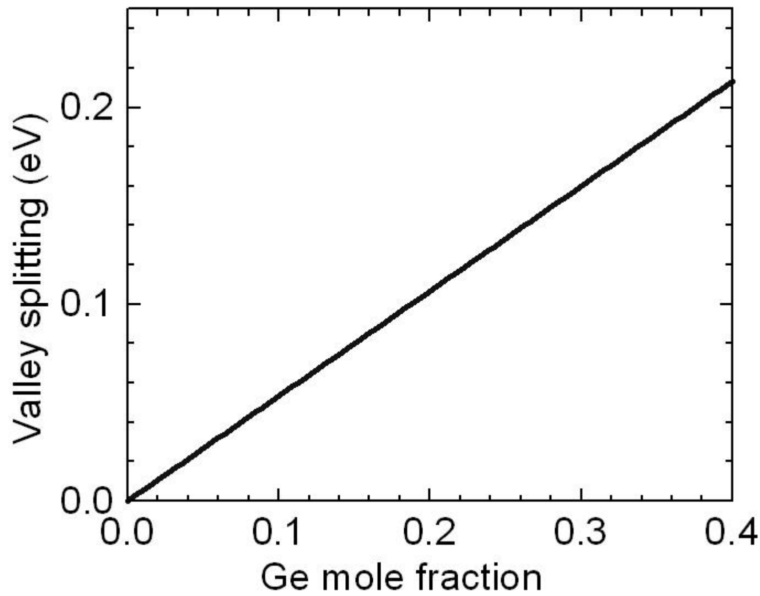

Figure 2-2: Δ_2 and Δ_4 valley splitting vs. Ge mole fraction in $Si_{1-y}Ge_y$ buffer. Note the essentially linear dependence.

2.2 Strain Effects

The electron band structure of TSSi on a relaxed $Si_{1-x}Ge_x$ substrate is calculated by the empirical pseudo-potential method (EPM) [Fri89] as a function of the substrate Ge mole fraction x. Acoustic and optical phonon scattering rate tables are calculated based on this full-band structure via Fermi's Golden Rule [Wan93]. *Figure 2-4* illustrates the agreement between our theoretical work, and the experimental and theoretical work of others for bulk unstrained Si [Jac83][Can75][Fis88]; *Figure 2-5*, the effects of strain on bulk mobility. In addition to phonon scattering in these device simulations, screened Coulomb scattering by impurities and surface roughness scattering are considered. The coupling potentials for the various phonon scattering processes were tuned by

matching the velocity and energy vs. driving electric field curve for unstrained bulk Si, and then the coupling parameters for surface roughness scattering [Rav00] were adjusted to match universal mobility data as a function of interface-normal effective field $E_{eff}(x) = \int E_y(x,y)n(x,y)dy \Big/ \int n(x,y)dy$ for unstrained Si [Tag94]. As pointed out by Fischetti and Laux [Fis96], in strained silicon, in-plane and out-of-plane valleys only differ by an energy shift approximately; the valley shape remains the same. We then follow the *approximation* that the same phonon coupling potentials can be used for strained silicon device simulation; this approximation is also used by Oberhuber et al. [Obe98] in treatment of the strained silicon valence band. The scattering rates, themselves, however change with the band structure and the availability of final states. We implement the surface roughness scattering mechanism following Goodnick's model [Goo85] with surface roughness rms value and correlation length tuned to fit universal mobility data. In strained silicon, both the case where the surface roughness parameters are held constant and, consistent with work of Ref. [Fis02], where surface roughness coupling parameters for strained Si are reduced are simulated. The band structure and the various band-structure dependent quantities are calculated with a resolution in the substrate germanium content $\Delta x = 0.05$ (5%) over the range of 0.0 to 0.45. For intermediate values of mole fraction, a quadratic interpolation scheme is used.

Captured in these calculations is the strain-induced degeneracy breaking in the electron band structure of the TSSi that has been well characterized [Wan02] (Fig. 3). In particular, the six lowest Δ energy valleys in the conduction band separate into two groups: two valleys with the valley centers $\mathbf{k}_l$ perpendicular to the plane

of the silicon/oxide interface and to the direction of transport, (in other words, $\mathbf{k}_l$ is parallel to the strained-Si growth direction) $l=\Delta_2$, and four higher energy in-plane valleys, $l=\Delta_4$. The strain-dependent energy difference $E^{(\gamma)}(\mathbf{k}_{\Delta 4}) - E^{(\gamma)}(\mathbf{k}_{\Delta 2})$ leads to segregation of low energy carriers toward the Δ_2 valleys that have the lower Si "transverse" effective mass (of ~0.19 m_e) in the direction of transport. This energy difference also reduces the densities of final states for low energy carriers, reducing the scattering rates accordingly [Wan93][Wan02].

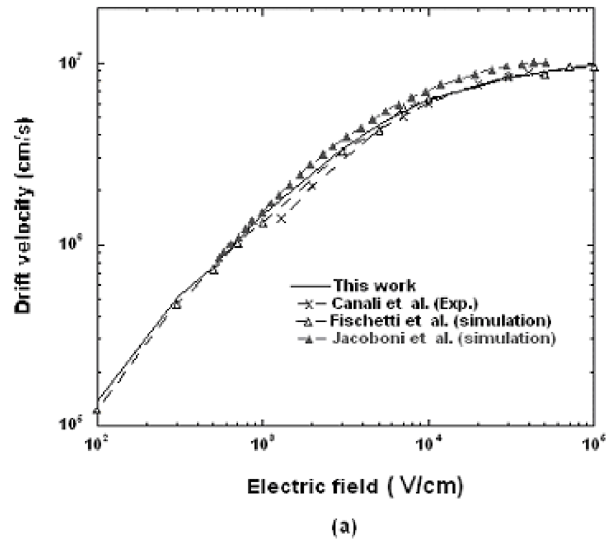

Figure 2-3: Drift velocity of electrons in bulk Si as a function of electric field. For comparison, the other measurements and simulation results are also shown. [Jac83][Can75][Fis88]

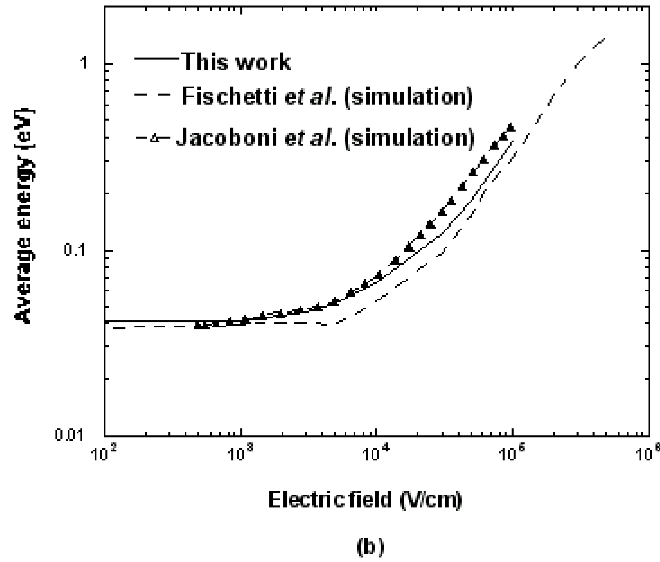

Figure 2-4: Average energy of electrons in bulk Si as a function of electric field. For comparison, the other measurements and simulation results are also shown. [Jac83][Fis88]

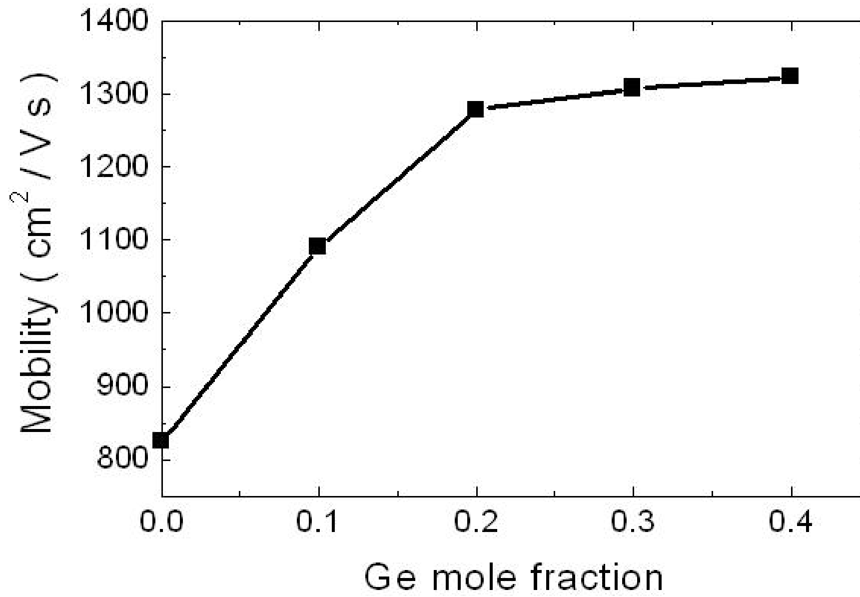

Figure 2-5: Low-field mobility vs. Ge mole fraction in $Si_{1-x}Ge_x$ buffer. Note the saturation of mobility for increasing strain as the valley-splitting-induced transfer to the Δ_2 valleys is completed for these low-energy carriers.

2.3 Valley-edge Quantum Corrections

In MOSFETs, perpendicular to the interface of silicon channel and gate oxide, size quantization in the inversion layer is important in at least two well known respects: decreased capacitive coupling of the channel to the gate and degeneracy breaking among the Δ valleys. The inversion layer exists in a narrow energy well for carriers that results in the formation of sub-bands. A resulting shift of the charge probability density of these quantum-confined states away from the Si/oxide (SiO_2 or otherwise) interface leads to the decreased capacitive coupling of the channel to the gate. Due to the differences in interface-normal effective masses, the lower interface-perpendicular mass in-plane Δ_4 valleys have larger quantum confinement/sub-band energies breaking the degeneracy between the in-plane Δ_4 valleys and the out-of-plane Δ_2 valleys even in the absence of strain. With tensile strain, lower energy electrons are partially segregated toward the Δ_2 valleys and inter-valley scattering for these low energy carriers is reduced. In MCUT a "quantum-corrected" position and strain-dependent band structure $E^{(\gamma)}{}_{qc}(\mathbf{k},x,z)$ is used to *approximate* these two effects, *and only these two effects*, of quantum confinement within the channel within this otherwise semiclassical approach.

The procedure for calculating the *valley-edge* quantum corrections in MCUT is that of [Win03], with advantages and limitations described therein, except that the valleys are considered separately. Let $E^{(\gamma)}(\mathbf{k},x,z)$ be the actual/uncorrected zero-field (flat-band neglecting material gradients) band structure as a function of position normal to the interface (z) including the band discontinuity at the

Si/oxide interface, and, potentially, in graded systems, position along the channel (x). Consider first the Δ valley centers $\mathbf{k}_l$, $l = \Delta_2, \Delta_4$. At the beginning of each time step in the simulation, for each valley and for each position (grid point) along the channel (x-direction), the one-dimensional effective-mass Schrödinger's equation normal to the interface

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{m_{z,l}(x,z)}\frac{\partial}{\partial z}\psi_{l,i}^{(\gamma)}(x,z)\right)+\left[E^{(\gamma)}(\mathbf{k}_l,x,z)+(-e)V(x,z)\right]\psi_{l,i}^{(\gamma)}(x,z)=E_{l,i}^{(\gamma)}(x)\psi_{l,i}^{(\gamma)}(x,z)$$

(2-1)

is solved to obtain the quantum-confined eigenstates i, $\psi^{(\gamma)}{}_{l,i}(x,z)$, where $V(x,z)$ is the electrostatic potential (in units of energy/charge), as per the example of *Figure 2-7*. Within a Boltzmann approximation, an equilibrium quantum mechanical electron density $n_{qc,l}^{(eq,\gamma)}(x,z)$ is then calculated *independently for each valley* from the quantum confined states,

$$\begin{aligned} n_{qc,l}^{(eq,\gamma)}(x,z) &= \sum_i \left|\psi_{l,i}^{(\gamma)}(x,z)\right|^2 \int_0^\infty g_{2D}(E_{//})\exp\left[-\frac{E_{//}+E_{l,i}^{(\gamma)}(x)-E_F(x)}{k_B T}\right]dE_{//} \\ &= \sum_i \left|\psi_{l,i}^{(\gamma)}(x,z)\right|^2 \frac{m_{l,x}^{1/2}m_{l,y}^{1/2}}{\pi\hbar^2}k_B T\exp\left[-\frac{E_{l,i}^{(\gamma)}(x)-E_F(x)}{k_B T}\right] \end{aligned}$$

(2-2)

as per *Figure 2-8*. Boltzmann statistics are used for the calculation of the quantum corrections for reasons discussed in Ref. [Win03], and to be consistent with the Boltzmann statistics approached in the Monte Carlo simulation as a result of

scattering without consideration of final state occupancies and, thus, Pauli exclusion. (If the basic Monte Carlo method were improved to better reflect Pauli exclusion effects and the limit of Fermi statistics, then including Fermi statistics within the quantum correction should not then pose a prohibitive additional problem.) The "quantum-corrected" band structure $E^{(\gamma)}{}_{qc}(\mathbf{k},x,z)$ is then *defined* to produce the same position-dependent *local* equilibrium electron densities

$$n_{qc,l}^{(eq,\gamma)}(x,z) \equiv N_l(x,z,T)\exp\left[-\frac{E_{qc}^{(\gamma)}(\mathbf{k}_l,x,z)-eV(x,z)-E_F(x)}{k_B T}\right]$$

(2-3)

where N_l is the effective density of states per valley (which is essentially valley independent except under large strain; here the *l* subscript is just a reminder that this is the per-valley effective density of states rather than that of the conduction band as a whole). Equation (2-3) can then be compared to the classical equilibrium electron concentration $n_{cl,l}(x,z)$

$$n_{cl,l}^{(eq,\gamma)}(x,z) \equiv N_l(x,z,T)\exp\left[-\frac{E^{(\gamma)}(\mathbf{k}_l,x,z)-eV(x,z)-E_F(x)}{k_B T}\right]$$

(2-4)

to obtain the valley and position-dependent quantum corrections

$$E_{qc}^{(\gamma)}(\mathbf{k}_l,x,z)-E^{(\gamma)}(\mathbf{k}_l,x,z) \equiv k_B T \ln\left[\frac{n_{cl,l}^{(eq,\gamma)}(x,z)}{n_{qc,l}^{(eq,\gamma)}(x,z)}\right] \qquad (2\text{-}5)$$

where the choice of E_F for calculation of these equilibrium carrier concentrations is irrelevant. (In practice we simply choose $E_F \equiv 0$ eV in our energy reference to obtain

$$E_{qc}^{(\gamma)}(\mathbf{k}_l, x, z) - eV(x, y) \equiv k_B T \ln\left[\frac{N_l(x,z)}{n_{qc,l}^{(eq,\gamma)}(x,z)\Big|_{E_F=0}}\right] \qquad (2\text{-}6)$$

directly from Equation (2-4), avoiding the need to actually calculate $n_{cl,l}^{(eq,\gamma)}(x,z)$; Equation (2-5) is for clarity of concept.) The *electrons within the Monte Carlo ensemble* with accompanying distribution function $n_{MC}(x,z)$ are then moved subject to the quantum-corrected full-band structure $E^{(\gamma)}{}_{qc}(\mathbf{k},x,z)$—where the quantum-corrected *full*-band structure and the movement of carriers therein will be considered momentarily. At the beginning of the next time step a new electrostatic potential $V(x,z)$ to be used in Equation (2-1) is calculated from the new ensemble Monte Carlo distribution $n_{MC}(x,z)$,

$$\nabla \cdot [\varepsilon \nabla V(x,z)] = -e[-n_{MC}(x,z) + N_D^+(x,z) - N_A^-(x,z)], \qquad (2\text{-}7)$$

so that the space and time-dependent quantum corrections are calculated self-consistently with the ensemble Monte Carlo distribution of carriers, as described in detail in the "flow chart" section to follow. The carrier temperature T in the above calculations could also be obtained approximately as a function of position from the Monte Carlo carrier ensemble as in Ref. [Win03], but this was not done for the examples in the illustrative work later. This procedure for calculating the band-edge quantum corrections is that of Ref. [Win03] except that the valleys are considered separately. This real-space and now valley-dependent quantum

correction addresses the reduced capacitive coupling of the channel to the gate as before, and now also the quantum-confinement-induced lifting of the energy valley degeneracy between the Δ_2 and the Δ_4 valleys.

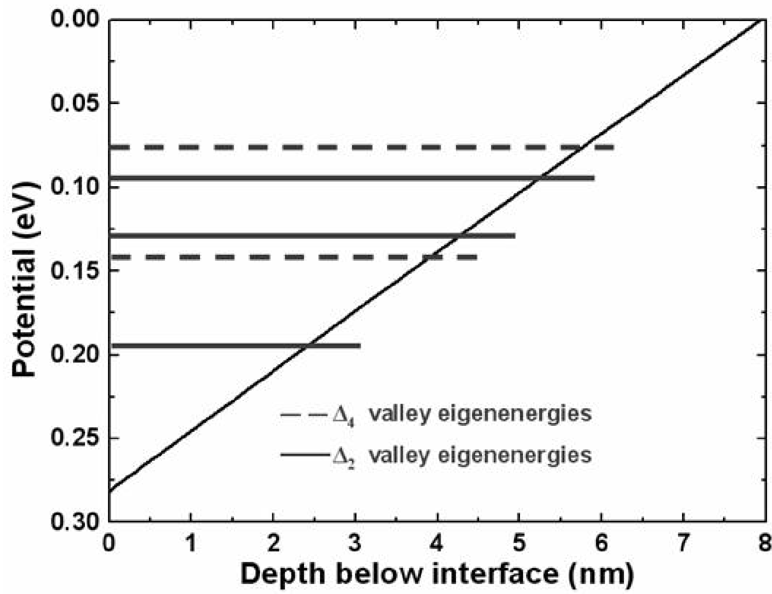

Figure 2-6: Eigenvalues of the energy well for longitudinal and transverse valleys

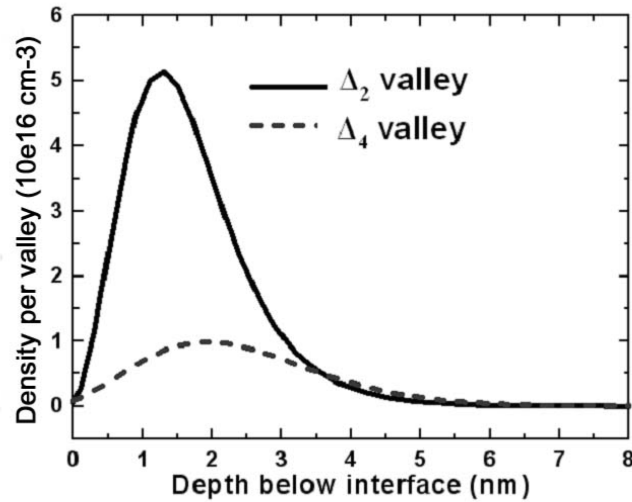

Figure 2-7: Quantum mechanical electron distribution density per valley of Δ_2 and Δ_4 valleys as a function of depth below interface.

2.4 Full-band Quantum Corrections via "Effective Strain"

The above quantum corrections are, of course, only well defined and physically reasonable for near-valley center states. Near the drain, however, a significant fraction of carriers can be expected to occupy high-energy states. What the quantum corrections for such states far from the valley centers *should be* is not clear and certainly not available through an effective mass approximation. What they are also is likely of little physical consequence; for these latter energetic states the relative degree of and effects of quantum confinement are significantly reduced. Still, there must be some interpolation/extrapolation process for the quantum-corrected band structure from the valley centers to the rest of the Brillouin zone in order to run a *full-band* Monte Carlo. Furthermore, the scattering rates will vary, particularly for low-energy carriers, as a function of the quantum-confinement-induced valley splitting. In this work we take advantage of the similarity of the effect of strain and the quantum corrections on the band structure—noting that the strain has little effect on the effective masses [Fis96]—to address both of the above quantum-correction-related issues via the definition of an "effective strain" or as it is actually calculated, a position-dependent substrate Ge "effective mole fraction" $\gamma_{eff}(x,z)$. For example, if there is no actual strain, the substrate mole fraction $\gamma(x,z)$ that would produce valley-splitting equal to $E_{qc}(\mathbf{k}_{\Delta 4},x,z) - E_{qc}(\mathbf{k}_{\Delta 2},x,z)$ through strain is simply chosen as the effective mole fraction $\gamma_{eff}(x,z)$, as per *Figure 2-9*. If the channel material is actually strained ($\gamma \neq 0$), the quantum-confinement-induced valley splitting is

simply added to the strain-induced valley splitting, and then this total splitting is translated back to the effective strain, as per *Figure 2-11*. In either case, the quantum corrected band-structure used in the Monte Carlo simulations becomes simply

$$E_{qc}^{(\gamma)}(\mathbf{k},x,z) = E_{qc}^{(\gamma)}(\mathbf{k}_{\Delta_2},x,z) + [E^{[\gamma_{eff}(x,z)]}(\mathbf{k},x,z) - E^{[\gamma_{eff}(x,z)]}(\mathbf{k}_{\Delta_2},x,z)] \tag{2-8}$$

that smoothly interpolates between the valley-centered quantum corrections calculated above. Furthermore, scattering rates and other simulation requirements, including some to follow, need only be calculated as a function of (effective) strain, not strain and degree of quantum confinement.

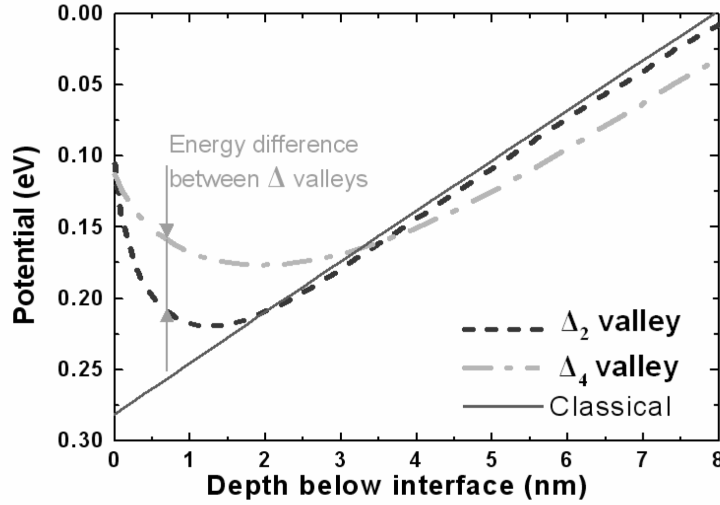

Figure 2-8: Effects of quantum corrections alone on the Δ_2 and Δ_4 valley-minimum energies for unstrained Si exhibiting position dependent additional valley splitting.

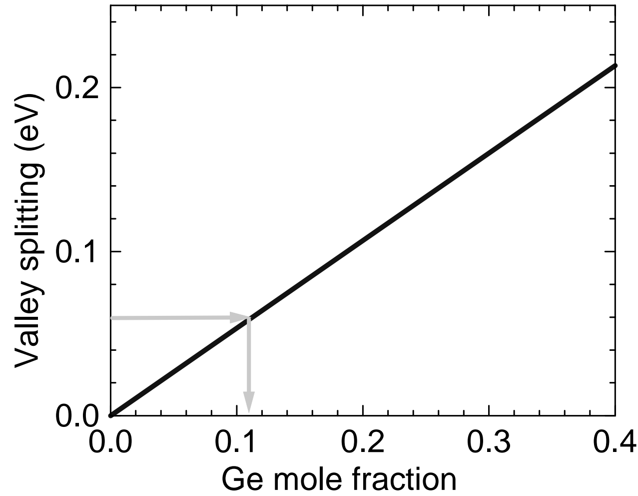

Figure 2-9: An illustration of the calculation of the effective strain/mole fraction γ_{eff} (at one position) from the quantum-corrected valley-minimum splitting, using the strain vs. valley splitting results of Figure 2-2.

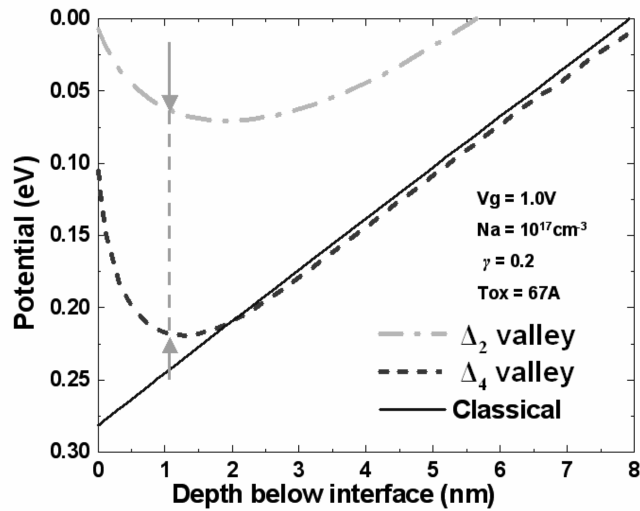

Figure 2-10: Effects of both quantum correction and tensile strain on the Δ_2 and Δ_4 valley-minimum energies for TSSi grown on $Si_{1-\gamma}Ge_{\gamma}$ buffer.

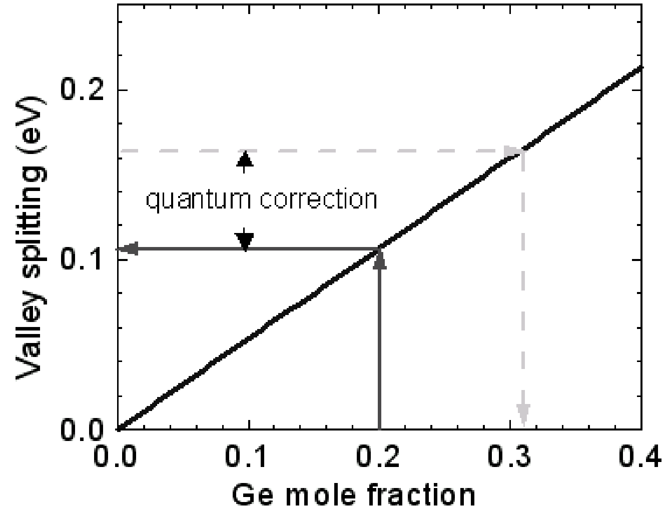

Figure 2-11: An illustration of the calculation of the effective strain/mole fraction γ_{eff} (at one position) starting with the real strain-induced valley splitting and then adding quantum-correction induced valley splitting to obtain a total valley splitting and corresponding new effective strain.

2.5 Transport within a Spatially-varying Band Structure

The particle motion in real space and **k**-space between scattering events are then calculated from, respectively,

$$\frac{d\vec{r}}{dt} = \frac{1}{\hbar}\left[\nabla_{\mathbf{k}} E^{(\gamma)}(\mathbf{k})\right]_{\gamma=\gamma_{eff}(x,z)} \tag{2-9}$$

and

$$\hbar\frac{d\mathbf{k}}{dt} = e\nabla_{\mathbf{r}} V(x,z) - \nabla_{\mathbf{r}} E_{qc}^{[\gamma_{eff}(x,z)]}(\mathbf{k},x,z) \tag{2-10}$$

or, as the latter equation is actually calculated, from equation (2-8) and the chain rule of calculus,

$$\hbar\frac{d\mathbf{k}}{dt}=e\nabla_{\mathbf{r}}V(x,z)-\nabla_{\mathbf{r}}E_{qc}^{[\gamma_{eff}(x,z)]}(\mathbf{k}_{\Delta_2},x,z)$$
$$-\nabla_{\mathbf{r}}\gamma_{eff}(x,z)\left\{\frac{\partial}{\partial\gamma}[E^{(\gamma)}(\mathbf{k})-E^{(\gamma)}(\mathbf{k}_{\Delta_2})]\right\}\Bigg|_{\gamma=\gamma_{eff}(x,z)} \cdot \quad (2\text{-}11)$$

The term in $\nabla_{\mathbf{k}}$ on the right-hand-side of equation (2-9) and the position-independent term in $\partial/\partial\gamma$ on the right-hand side of equation (2-11) are pre-tabulated as a function of γ and **k** when the strained band structures are calculated. It should be noted that anytime the band structure changes with position, due to quantum corrections as here or simply due to variation in the material and/or strain, consideration of the **k**-dependent gradient of the band structure with respect to position is necessary; it is the gradient with **r** of the band structure *at and only at the electrons location in **k**-space*, along with that of the external potential $-eV(x,z)$, that produces the electron's crystal momentum change through energy conservation requirements, as per equation (2-10).

2.6 Model Devices

For the MOSFET simulations we consider two 0.25 μm channel length nMOSFETs, one with a relaxed Si channel and the other with a TSSi channel corresponding to growth on a relaxed $Si_{0.8}Ge_{0.2}$ buffer. These relatively long channel MOSFETs allow a good estimation of the channel electron mobility, which can be compared to the experiments of Rim *et al.*'s [Rim00]. As shown in *Figure 2-13*, the MOSFET has 67 Å oxide thickness, 130 Å thick p-type strained silicon channel with 1×10^{17} cm^{-3} acceptor doping, and a 750 Å anti-punch-

through region underneath of 1×10^{18} cm^{-3} acceptor doping. The n^+ source/drain is implanted in the strained silicon, and is chosen to be a Gaussian distribution with peak density of 1×10^{20} cm^{-3}. However, the simulator is capable of handling experimental or simulated doping profiles. The polysilicon gate length is 0.25 μm and doping is 3×10^{20} cm^{-3} (*Figure 2-13*). In the strained device, because the inversion layer is within the TSSi, we neglect the $Si_{0.8}Ge_{0.2}$ buffer in transport simulation (but not the resulting strain).

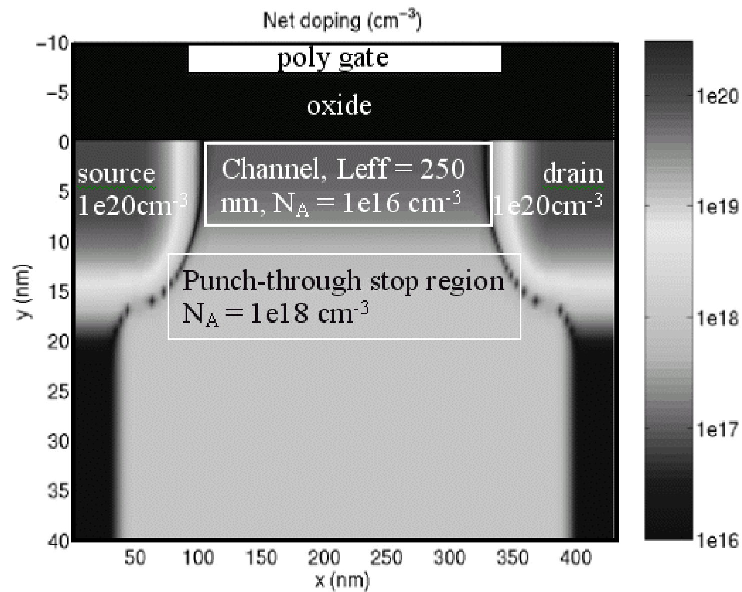

Figure 2-12: 2D device doping profile.

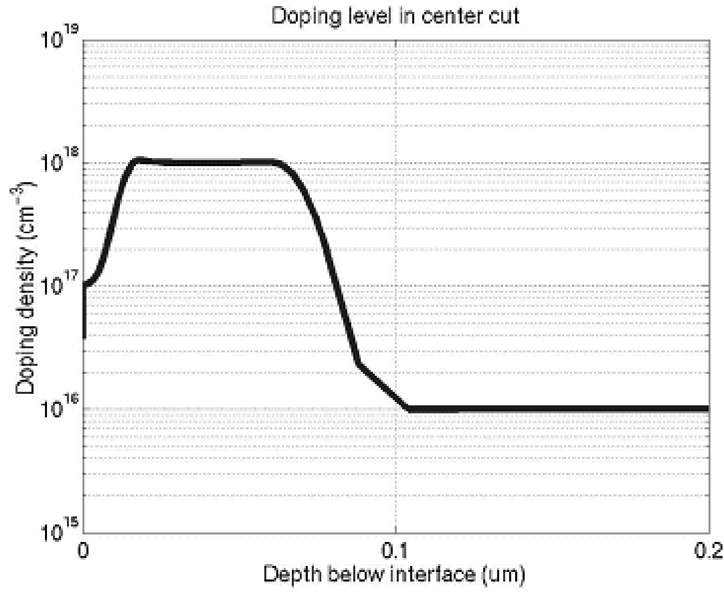

Figure 2-13: Doping density along cross-section normal to gate at mid-channel.

2.7 Mobility Calculations

To obtain low field electron mobility, a small drain-to-source voltage (0.1 V) was applied and gate voltage varied from 0.5 V to 2.4 V to vary the normal effective field. The channel mobility for the simulated devices is calculated as follows: At any point in the channel, neglecting small diffusive contributions, drive current is the product of channel mobility μ, sheet charge density Q and driving field.

$$I = \mu \cdot Q(x) \cdot \left(-\frac{dV}{dx} \right). \tag{2-12}$$

Equation (2-12) can be rewritten as

$$\frac{-I}{Q(x)} dx = \mu dV , \tag{2-13}$$

and, assuming constant channel mobility, integrated from source to drain to obtain

$$\mu = \frac{(-I)\cdot \int \frac{1}{Q(x)}dx}{V_{DS}} . \qquad (2\text{-}14)$$

Though mobility in the source/drain area is not equal to channel mobility, since carrier density $Q(x)$ is much higher in source/drain region than in the channel, the channel region dominates the calculation of mobility and provides a reasonable estimate of average channel mobility.

Figure 2-14 and *Figure 2-15* show the simulated drain current and associated calculated mobility as a function of gate voltage and corresponding effective field normal to the interface. Note that the mobility saturates as expected as a function of strain once the transfer of these, for mobility measurements, inherently low energy carriers to the Δ_2 valleys is essentially complete. *Figure 2-16* and *Figure 2-17* show calculated mobilities as a function of oxide-normal effective field for no strain ($\gamma = 0$) and $\gamma = 0.2$, as well as relevant experimental results for comparison [Tag94][Rim00][Wel94]. Note that the MC simulations for fixed surface roughness scattering parameters apparently underestimate the actual increase in the channel mobility. Fischetti et al. [Fis02] have performed more rigorous calculations of channel mobility (calculation of which is a boundary condition of MCUT, not a goal) and indicate that the strain- and quantum-induced symmetry-breaking and the resulting reduction in conductivity effective mass and scattering rates alone is not sufficient to explain the experimentally obtained increase in mobilities in TSSi channels; they suggest a fortuitous decrease in the surface roughness scattering. Accordingly, to achieve better agreement with the experimental results we have reduced that of the surface

roughness coupling potential (rms roughness value) by 44% in this work to achieve better agreement as also shown in *Figure 2-17*.

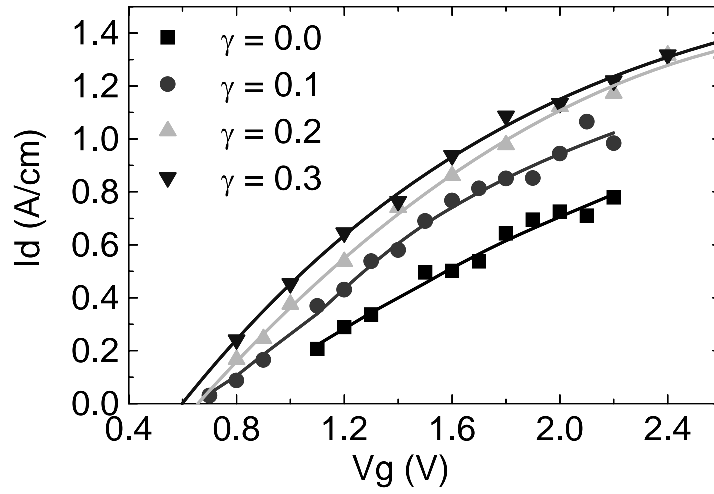

Figure 2-14: Drain current vs. gate voltage vs. normal effective field in MOSFET with TSSi channel on relaxed $Si_{1-\gamma}Ge_{\gamma}$ buffer. γ is the Ge mole fraction in the buffer layer such that $\gamma = 0$ denotes unstrained silicon.

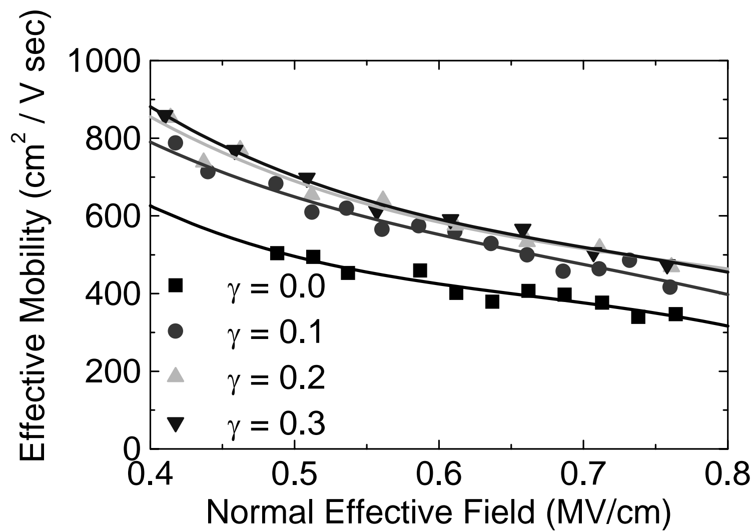

Figure 2-15: Channel electron mobility vs. normal effective field in MOSFET with TSSi channel on relaxed $Si_{1-\gamma}Ge_{\gamma}$ buffer. γ is the Ge mole fraction in the buffer layer such that $\gamma = 0$ denotes unstrained silicon. For TSSi with buffer of $Si_{0.8}Ge_{0.2}$ and $Si_{0.7}Ge_{0.3}$, the mobility enhancements are very close as the carrier transfer to the Δ_2 valleys saturates.

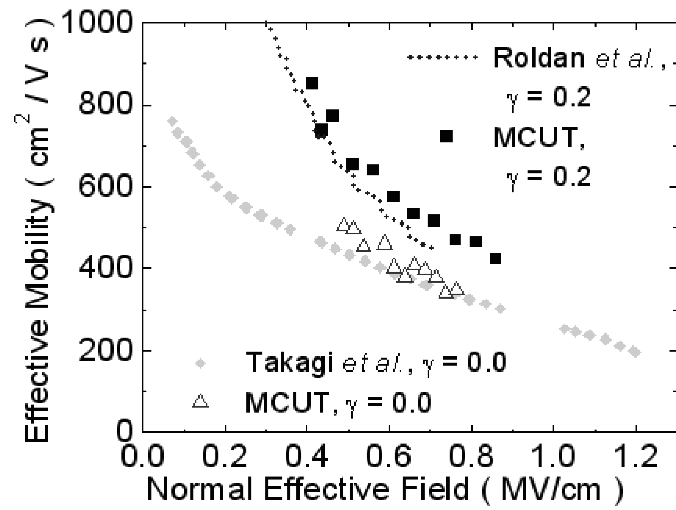

Figure 2-16: Results of MCUT device simulation are compared with other simulation results [Rol96] and the universal mobility curve [Tag94]. Open triangles: MCUT MOSFET simulation on unstrained Si; solid squares: MCUT MOSFET simulation on TSSi.

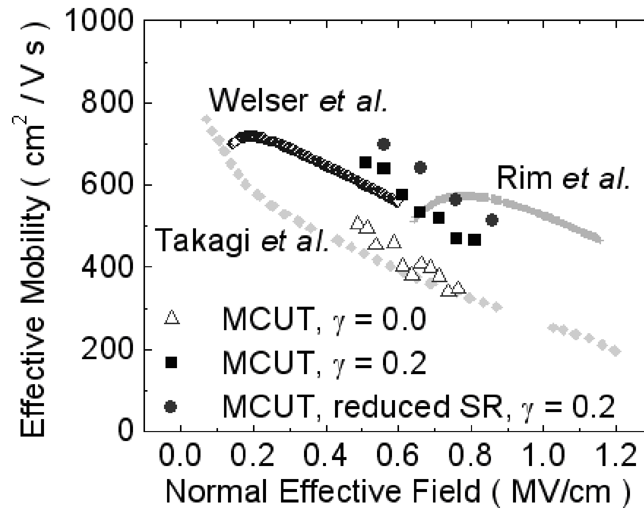

Figure 2-17: MCUT TSSi MOSFET (2D) simulations with and without a reduced rms amplitude of the Si/SiO$_2$ interface roughness are compared to experiment data [Wel94][Rim00]. Unstrained Si universal mobility data [Tag94] are plotted as reference. Solid circle: MCUT TSSi MOSFET simulation with reduced surface roughness scattering rate; solid square: MCUT TSSi MOSFET simulation with fixed surface roughness scattering rate; open triangle: MCUT unstrained Si MOSFET simulation.

2.8 Illustrative High-field and Deeply Scaled Device Results

Figure 2-18 shows I_D vs. V_D curves for $V_G - V_T = 1.0$ V and $V_G - V_T = 2.0$ V showing the enhancement in drive current due to strain and, additionally, due to the apparent reduction in surface roughness discussed above for relatively large devices. *Figure 2-19* and *Figure 2-20* shows the carrier velocity and energy distribution along the channel for $V_G - V_T = 2.0$ V and $V_D = 0.5$ V. The results of *Figure 2-18* suggest the possibility for significantly increased drive current, approximately a factor of 2 increase in the drive current for $V_G - V_T = 1.0$ V. *Figure 2-21* illustrates the simulation of devices with channel lengths of only a few

tens of nanometers (device after Ref. [Mit00]). Simulations such as this last one are the end goal of developing MCUT, and in the future such simulations will be performed using MCUT to analyze how and to what degree mobility enhancement and such drive current enhancement will translate to conventional and non-classical CMOS with channel lengths of a few tens of nanometers.

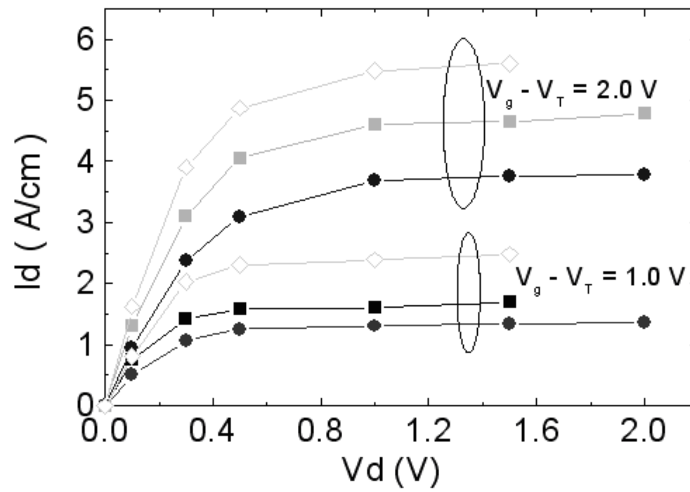

Figure 2-18: I_d-V_d curve. Drain currents as a function of drain voltage for nMOSFETs with TSSi (γ = 0.2) channel and an unstrained Si (γ = 0.0) channel. Closed circles: unstrained Si; closed squares: TSSi with fixed surface roughness parameters; open diamonds: TSSi with reduced surface roughness scattering consistent with Figure 2-17. For the same $V_G - V_T$, enhancement of the drive current is observed in the TSSi.

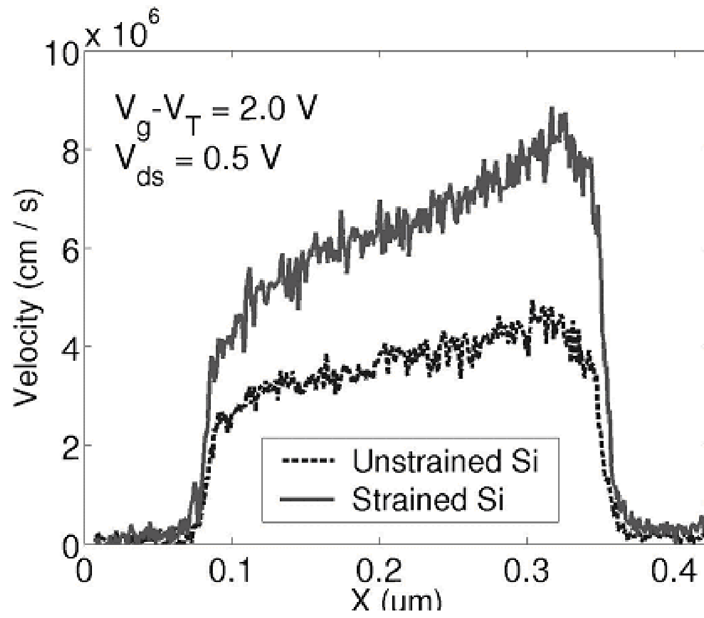

Figure 2-19: Average carrier transport velocity along the channel for fixed V_D and $V_G - V_T$.

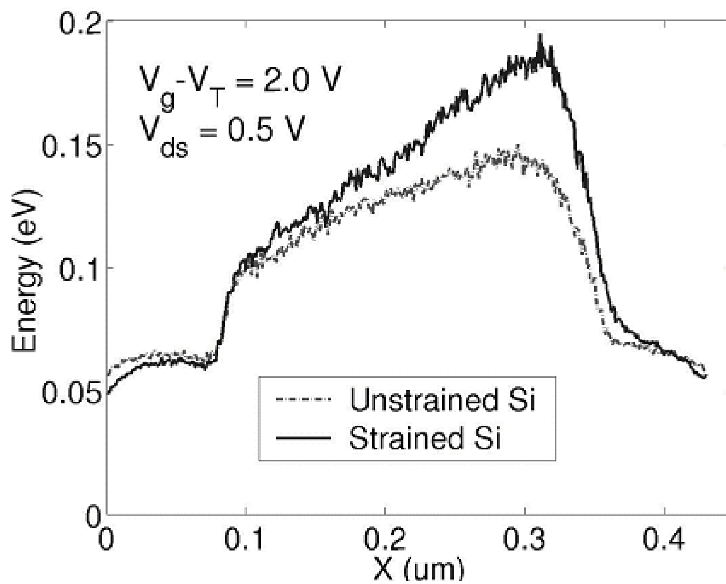

Figure 2-20: Average carrier kinetic energy distribution along the channel for fixed V_D and $V_G - V_T$. Comparison is made between a TSSi channel with reduced surface roughness scattering and a relaxed Si channel. Increases are observed in both average energy and velocity within the channel for TSSi as compared to relaxed Si.

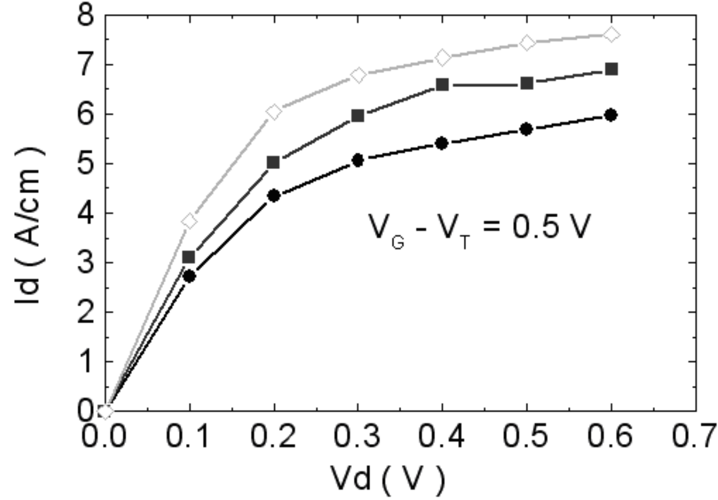

Figure 2-21: I_d-V_d curve for a 50 nm channel length well-tempered device after Ref. [Mit00]. Drain currents as a function of drain voltage for nMOSFETs with TSSi (γ = 0.2) channel and an unstrained Si (γ = 0.0) channel. Closed circles: unstrained Si; closed squares: TSSi with fixed surface roughness parameters; open diamonds: TSSi with reduced surface roughness scattering consistent with Figure 2-17. For the same $V_G - V_T$, (V_T's are different in these three cases) enhancement of the drive current is observed in the TSSi, although perhaps a bit less than the longer channel devices for this isolated example.

2.9 Scalability of TSSi nMOS on Technology Roadmap

From the above discussion and simulation results, we have realized that the underlying physical principle of TSSi is the strain-induced degeneracy breaking between the interface-normal- and interface-parallel-oriented conduction band Δ valleys which leads to a lower conductivity effective mass and reduced inter-valley scattering and, thus, enhanced mobility [Fis96][Fan04]. However, size quantization in strong inversion also breaks the degeneracy among the Δ valleys in much the same way, so that both systems have substantial degeneracy breaking;

strained Si simply has more. Therefore the actual advantage of SSi over Si may be less than otherwise expected, especially in high gate field regime [Fis02].

Because of the smaller bandgap, SSi nMOS has a smaller threshold voltage (V_{th}) than its unstrained Si counterpart, if all processing steps are kept the same. Ideally the same threshold voltage is required for SSi devices in order to keep the same I_{on}/I_{off} ratio at V_{dd}. So SSi devices need to have V_{th} increased, either by using a gate material with a higher work function, or by increasing the channel doping level. While gate work function engineering is very difficult, increasing channel doping concentration seems to be a practical solution. However, higher doping results in increased effective field and impurity scattering rate in the channel, thus reducing channel mobility. Combining these effects: more valley splitting, higher channel doping, and possibly smoother interface, can we foresee SSi nMOS scalability? For the next generation nMOS, is SSi a better choice than Si? More device simulations with decreasing channel lengths were performed with MCUT to address this question.

We choose physical channel lengths of 240 nm, 100 nm, 40 nm and 25 nm, with the gate length being 250 nm, 171 nm, 86 nm, and 50 nm, respectively. These lengths represent the scaling trend from 2002 to 2007. With channel length scaling, V_{dd} will decrease from 2.5 V to 1.1 V, while V_T changes from 0.7 V to 0.4 V. In our simulation, we took device doping profiles simplified from experiments [Rim00] and "well-tempered" bulk-Si nMOS devices [Mit00]. All the devices use anti-punchthrough implants by (super) halo doping. The surface channel p-type doping levels are plotted in *Figure 2-22*. Similar device profiles are used in SSi nMOS, except that more intensive halo implant are applied to

compensate for the effect of bandgap reduction of SSi. In our simulations, the halo doping levels in SSi nMOS were tuned to achieve the same V_T as for bulk-Si by examining $I_D - V_G$ curves. Compared with Si nMOS, the channel surface doping in SSi is 76% ~ 30% higher, with the factor decreasing with channel length. Because of the increased channel doping, the transverse gate field in short channel devices is higher, as shown in *Figure 2-23*. As pointed out in section 2, higher gate field causes more degeneracy-breaking among Δ valleys, as shown in *Figure 2-24*. The Δ_2 valleys with lower energy attract more carriers. For example, more than 75% of electrons reside in Δ_2 valleys, which happens when $E(\Delta_2) - E(\Delta_2) \approx 100$ meV [Fis02]. The mobility enhancement due to valley splitting is plotted in *Figure 2-25*, for strained Si in the absence of a gate bias. It is clearly seen that mobility enhancement in strained Si reduces at valley splitting greater than 100 meV, which happens when nMOS L_{eff} shrinks from 40 nm to 25 nm. For a 25 nm nMOS device, though SSi channels have more valley splitting energy than bulk Si does, the mobility enhancement diminishes. Since SSi nMOS devices have increased channel doping, it results in higher ionized impurity scattering rate. In ultra-short devices, both Si and SSi MOSFETs also undergo carrier heating and velocity overshoot. These are all captured by MCUT, and I_{on} enhancement as a function of channel length is shown in *Figure 2-26*, in the term of "V_T-matched".

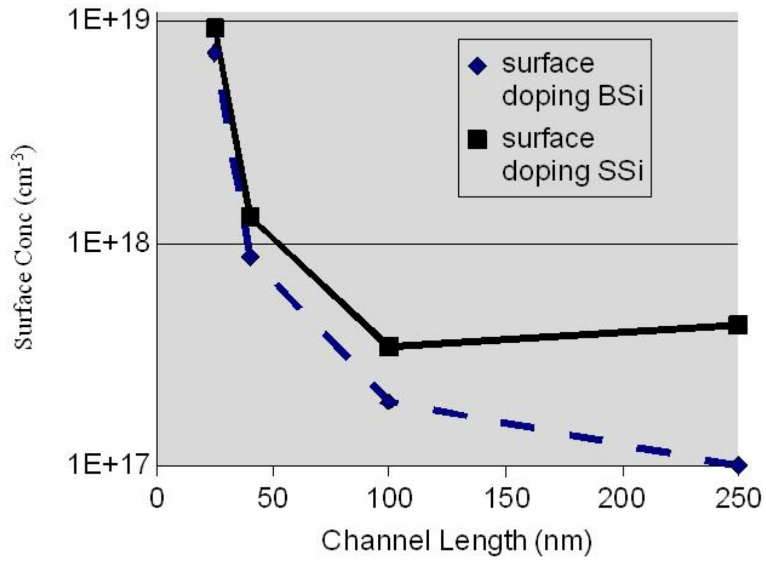

Figure 2-22: Channel surface doping in strained Si and bulk Si nMOSFETs.

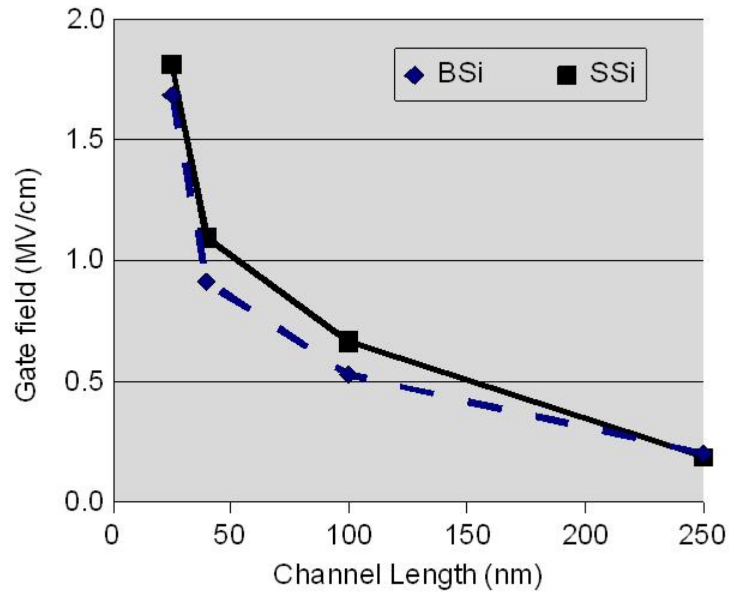

Figure 2-23: Average gate field in strained Si and bulk Si nMOSFETs.

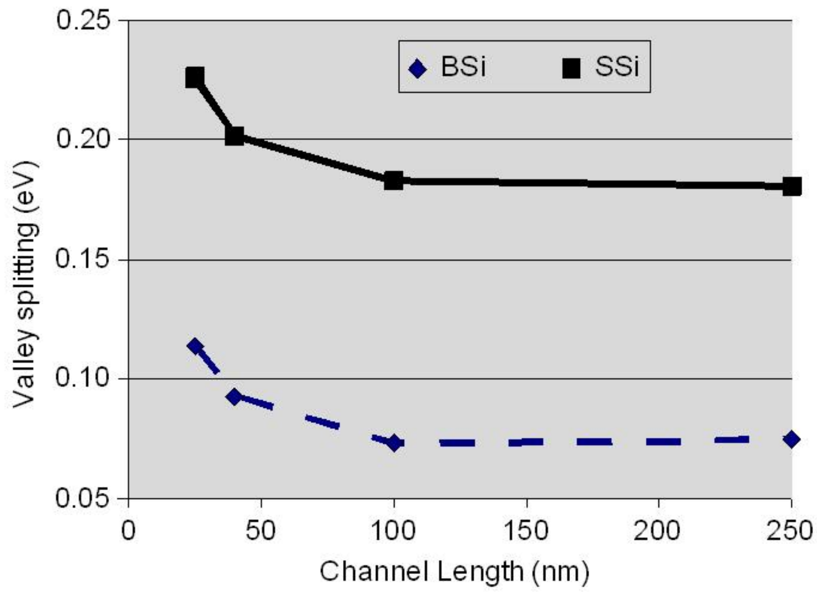

Figure 2-24: Energy splitting between Δ_2 and Δ_4 valleys in SSi and bulk Si channels.

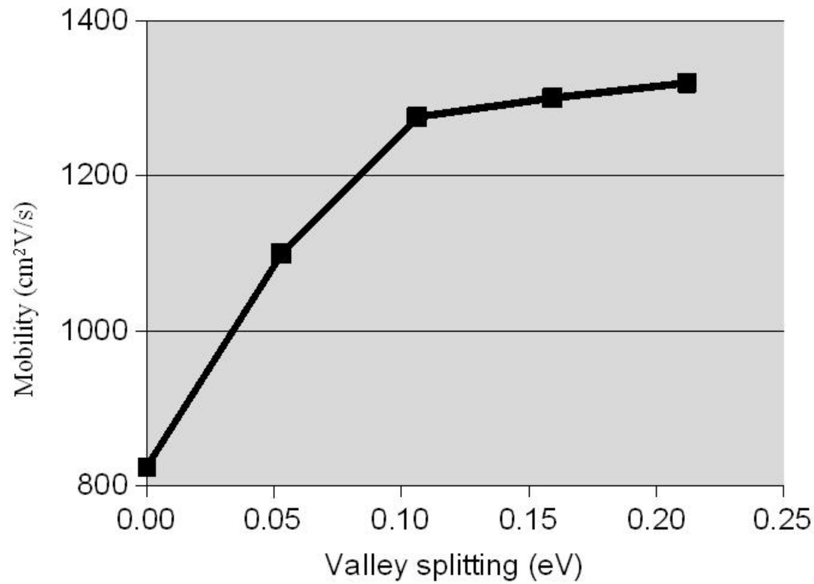

Figure 2-25: Low-field bulk mobility plotted as a function of energy splitting between Δ_2 and Δ_4 valleys.

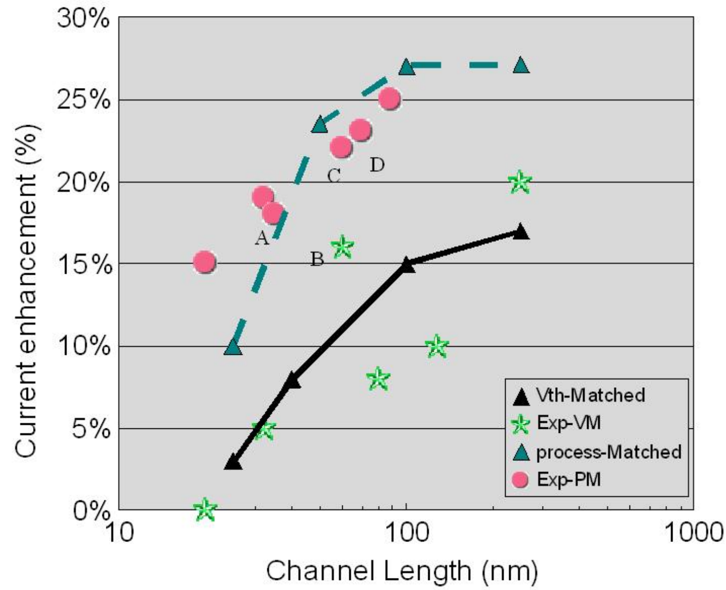

Figure 2-26: Drain current comparison between Si and SSi nMOSFETs. Lines represent MCUT simulations in "V_T-matched" and "process-matched" conditions. Experimental results come from: a. Ref [Xia03], b. Ref [Rim02a], c. Ref [Lee02], d. Ref [Hwa03], the rest from Ref [Goo03].

In the most recent development on SSi devices, gate workfunction engineering has been applied to counter the V_T change [Xia03]. Though more complex gate process steps are required, the SSi device can be rid of increased channel dopants and thus keep the drain current advantage. In our work we assume ideal gate material for SSi device to counter the V_T change to get the so-called "process-matched" I_{on} comparison between bulk Si and SSi nMOS, and the results are shown in *Figure 2-26* too. In these simulations identical doping profiles were used in bulk Si and SSi nMOS, and gate biases were applied to maintain same overdrive ($V_G - V_T$) in comparison.

Shown in *Figure 2-26*, MCUT simulation results in the "V_T-matched" and "process-matched" categories are both consistent with experimental data (VT-matched: [Goo03][Rim02a]; process-matched: [Goo03] [Xia03] [Lee02] [Hwa03]). Both

our theoretical estimations and published experiments predict that SSi nMOS loses its performance advantage over unstrained Si at around 25 nm L_{ef}, in the threshold-voltage-matched case. Compared to other Monte Carlo simulation, this work takes quantum confinement induced valley splitting in both bulk Si and SSi into consideration, thus giving a better assessment of current enhancement of SSi nMOS over bulk devices.

2.10 Conclusion

Biaxially tensile-stressed Si nMOS shows advantages over bulk Si MOSFETs in terms of low field channel mobility and electron velocity overshoot at submicron channel lengths. However, at ultra-short channel lengths, increased channel doping and gate quantization effects have brought diminishing returns to SSi over bulk Si. For nMOSFETs, our semiclassical Monte Carlo simulation comparison of bulk Si and SSi devices in the channel length range from 240 nm to 25 nm indicates that SSi will not be attractive as for nMOS channel material at the 25 nm channel lengths if V_Ts are kept constants. Though oxide interface roughness remains an unsolved issue in these devices, other high mobility channel materials which do not exhibit bandgap lowering, such as uniaxial strained Si and III-V compound semiconductors, may be better choices.

3 Uniaxially Stressed Si CMOS

3.1 Motivation

Recent experimental and theoretical works point to uniaxially-stressed Si as a possible channel material for next generation CMOS [Tho05]. As compared to tensile biaxially-stressed Si grown on relaxed SiGe substrates [Lei02] [Fis03] [Hoy02] uniaxial-stressed Si has the advantage that both theoretical estimations and experimental results indicate that the hole channel mobility improvement is maintained at high gate fields [Gil04] [Tho04]. Experimentally, uniaxial compressive stress has been applied to the channel area via source/drain hetero-epitaxial deposition of SiGe [Tho04][Gha03][Chi04], or by compressive capping layers on top of the gate [Arg04], while four-point or ring wafer bending is the main technique for experimentally measuring piezoresistance [Bea92]. Greater than 50% stress-induced hole channel mobility improvement and 10% to 25% drain current enhancement has been demonstrated in strained Si devices for stress ranging from 100 to 500 MPa [Tho04][Gha03][Chi04][Arg04]. Furthermore, although the hole mobility is often modeled as depending linearly on stress according to tabulated piezoresistance coefficients such as those of Ref. [Smi54], recent studies have exhibited super-linear relationships between bulk mobility and stress [Shi04][Wan04] suggesting even greater potential advantages in the high stress regime.

For this work, the low-field bulk mobility of uniaxially-stressed Si is studied

theoretically but over a wider range of stress than considered previously. In addition, thermal velocities that, along with the low-field mobilities, are relevant for short channel devices [Lun97] are also calculated. This study focuses on transport in the [110] and $[1\bar{1}0]$ direction for [110] compressive and tensile uniaxially-stressed Si. Considered stress, both tensile and compressive, begins at 50 MPa within the realm of bulk piezoresistance measurements [Smi54], enters the GPa stress regime as now appears possible via process induced stress [Arg06], and continues up to 4 GPa in a search for limiting behavior. The valence band structures are calculated with a **k•p** Hamiltonian method. The thermal velocities, along with thermal average effective masses and phonon scattering rates (the latter being dependent on the stress-dependent densities-of-states) are obtained directly from the band structure. MCUT is then used operating in the bulk limit to accurately calculate the bulk mobility in the presence of the highly non-parabolic, anisotropic band structures and inelastic scattering.

3.2 Band Structures

For this work, the **k•p**-based bandstructure program Nextnano3 [Bir02] was used for uniaxially-stressed Si band structures, with the deformation potentials for stressed Si obtained from [Fis03]. And as in [Fis03], a 6x6 Hamiltonian was used in Nextnano3 to calculate the “heavy-hole” (HH), “light-hole” (LH) and “split-off” (SO) hole bands. Because it is a perturbation method, the reliability of the calculated band structure at high energies is limited. However, the approach should be reliable for the thermal velocity and low-field mobility calculations of this work.

Under increasing stress along the [110] direction, the warping of the Si HH band (here, in the presence of anti-crossings, defined as the lowest lying hole band/highest lying valence band) is gradually reduced with increasing splitting between the HH and LH band, as seen in *Figure 3-1*.

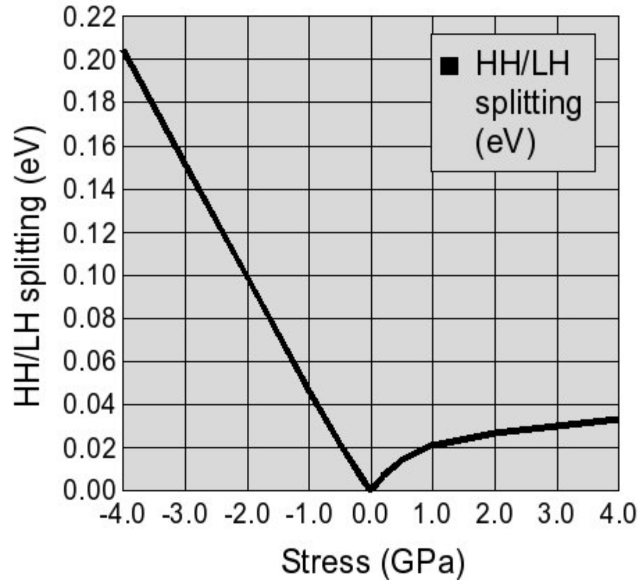

Figure 3-1: Heavy hole and light hole band splitting in uniaxially-stressed Si. Here and for subsequent figures the positive stress axis corresponds to tensile stress, and the negative stress axis to compressive stress.

Note that with band occupation ratios of HH:LH:SO ≈ 65:10:1 for unstrained Si and with the ratios shifting further in favor of the HH band with increasing stress, hole transport is governed largely by the properties of the HH band. The anti-crossing between the HH and LH bands in the [110] direction makes the effective mass of the HH band in the [110] direction smaller near the Γ point and the iso-energy surfaces near ellipsoidal in the (001) plane at low energies, as shown in *Figure 3-2* and *Figure 3-3*, for a 500 MPa stress. In *Figure 3-3*, the 10 meV energy surface of the HH band resembles an ellipsoid with low m^* in the [110] direction,

the 50 meV surface—approximately the thermal average energy—is clearly becoming warped, and the 100 meV surface is quite similar in shape to the bulk Si energy contour with almost no advantage for transport in the [110] direction. The reduced warping at low energies can be understood as degeneracy breaking among the energy "wings" in the otherwise 12 equivalent <110> directions in bulk Si HH band. Under [110] compressive stress, the energy of the wings extending in $[1\bar{1}0]$, $[\bar{1}10]$, $[011]$, $[0\bar{1}\bar{1}]$, $[101]$, and $[\bar{1}0\bar{1}]$ are lowered relative to that of the remaining six wings. These relative shifts, together with the HH/LH band splitting, also cause a redistribution of carriers within the band structure. As the band splitting increases with stress, the region of the band structure below the anti-crossing where the band structure appears ellipsoidal increases, allowing more carriers within a thermal distribution to have a small effective mass in the [110] direction.

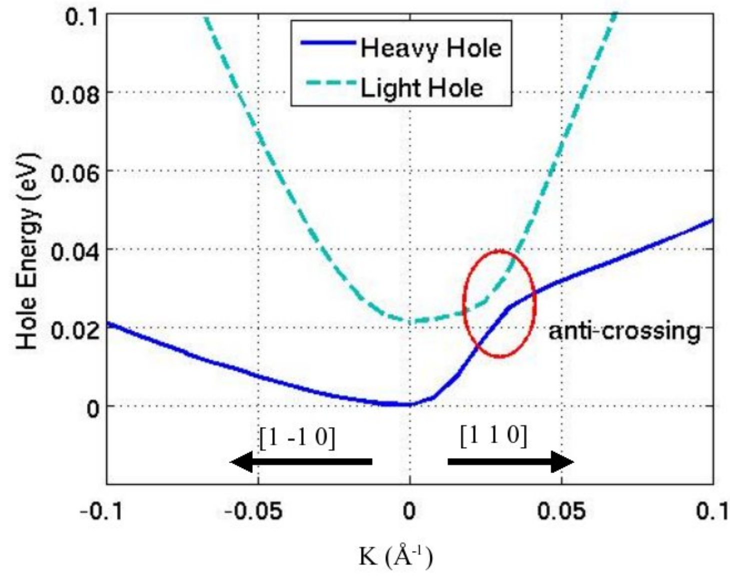

Figure 3-2: Heavy-hole and light-hole energy-K dispersion in [110] and $[1\bar{1}0]$ directions with compressive [110] uniaxial-stress of 500 MPa.

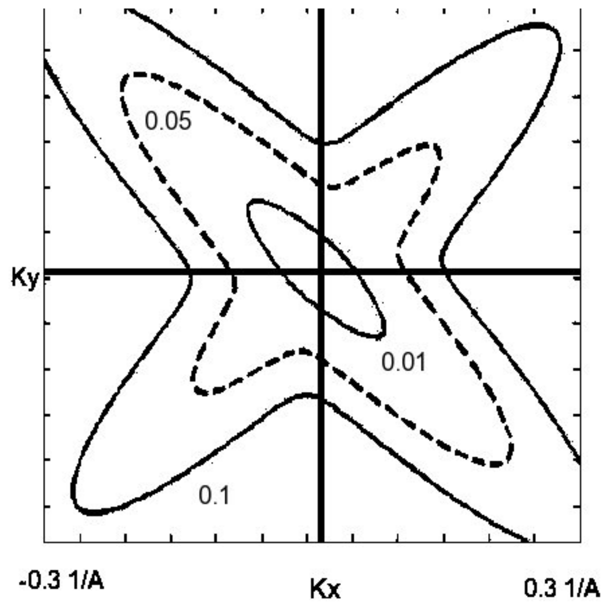

Figure 3-3: 2-D heavy-hole iso-energy contours (10 meV, 50 meV and 100 meV) for [110] uniaxially-stressed Si in the (001) plane with a compressive stress of 500 MPa in the [110]

direction.

To better quantify the level of asymmetry in uniaxially-stressed Si hole bands as a function of stress magnitude, a thermal average of the reciprocal of the effective mass $m_{\hat{\mathbf{n}}}$ in the chosen transport direction $\hat{\mathbf{n}}$ is calculated from the full-band structure by averaging it over **k**-space and all hole bands α using Boltzmann energy statistics at room temperature:

$$\frac{\hbar^2}{m_{\hat{\mathbf{n}}}} = \sum_{\alpha} \int \frac{\partial^2 E_{\alpha}(\mathbf{k})}{\partial k_{\hat{\mathbf{n}}}^2} \exp(-\frac{E_{\alpha}(\mathbf{k})}{k_B T}) d^3\mathbf{k} \Bigg/ \sum_{\alpha} \int \exp(-\frac{E_{\alpha}(\mathbf{k})}{k_B T}) d^3\mathbf{k} \, . \qquad (3\text{-}1)$$

The mass reciprocal is used because of its nominally linear relationship to mobility. Results are shown in *Figure 3-4*. Note that in unstrained Si, consistent with essentially isotropic low field mobility in bulk Si [Ott75], $m_{\hat{\mathbf{n}}}$ is essentially isotropic despite the highly anisotropic multi-band band structure (a reminder that considering the directional dependence of the band curvature only at the Γ point can be highly misleading.) However, stress and the associated symmetry reduction in the band structure breaks the symmetry in the thermal average mass reciprocal, advantageously increasing it under appropriate combinations of stress and channel orientation.

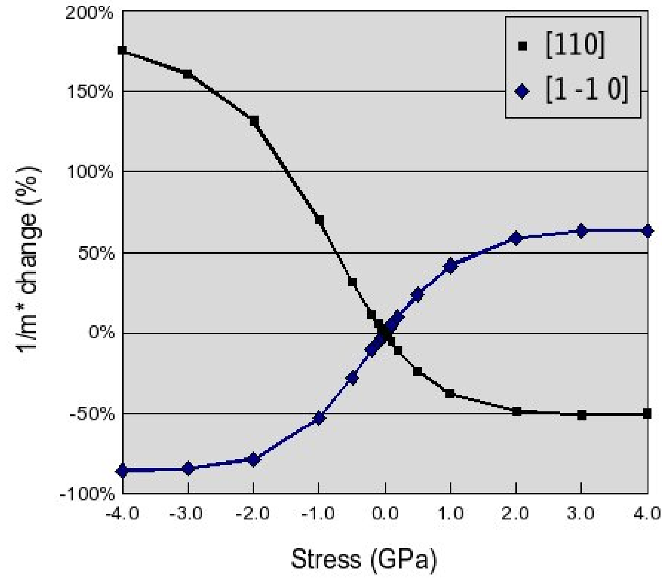

Figure 3-4: Percentile change in the thermal average of the inverse effective mass in the [110] and $[1\bar{1}0]$ *directions for [110] uniaxially-stressed Si as function of stress. The zero stress value is 2.79 times the reciprocal of the free space mass. For reference, the average inverse effective mass value in the [001] direction in unstrained Si is 2.80.*

Similarly, the Boltzmann distribution-weighted average of the thermal velocities in the transport direction, v_{th}, was obtained from

$$v_{th} = \sum_{\alpha} \int |\mathbf{v} \cdot \hat{\mathbf{n}}| \exp(-\frac{E_{\alpha}(\mathbf{k})}{k_B T}) d^3\mathbf{k} \Big/ \sum_{\alpha} \int \exp(-\frac{E_{\alpha}(\mathbf{k})}{k_B T}) d^3\mathbf{k} \,. \tag{3-2}$$

Results are shown in *Figure 3-5* where again significant and growing anisotropy is apparent under stress. (Note that for the highly isotropic HH band there can be both forward and backward moving carriers in a given half-**k**-space, and, thus, the integral over all **k**-space of the directed velocity magnitude was performed.)

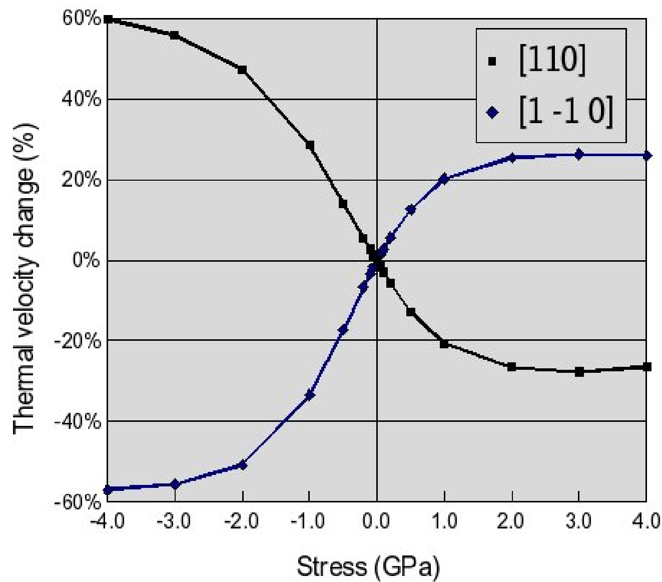

Figure 3-5: Percentile change of the thermal velocity in the [110] and $[1\bar{1}0]$ *directions for [110] uniaxially-stressed Si as function of stress. The unstrained Si thermal velocity is* 8.9×10^6 *cm/s at room temperature.*

Another important feature of the impact of uniaxial stress on the band structure is the reduction of the DOS and the corresponding reduction in the phonon scattering rates. The phonon scattering processes included in bulk simulations were longitudinal acoustic, transverse acoustic and optical phonons. The scattering rates were calculated with Fermi's golden rule based on the band structures. Calibration on the scattering process has been done [Wan02] with unstrained Si with respect to drift velocity and carrier energy [Ott75][Jac77]. The shorter low-energy contours in *Figure 3-3* than in bulk Si are indicative of this reduced DOS. The reduced DOS in the HH band results in smaller intra-band phonon scattering rates. The HH/LH energy splitting also leads to fewer inter-band scattering events. Analogous to above, the thermal average of the

reciprocal of the scattering rate, the mean time between scattering events which is again nominally proportional to the mobility, were calculated as a function of stress from

$$\tau_{th} = \frac{1}{R_{th}} = \sum_{\alpha} \int \frac{1}{R_{\alpha}(\mathbf{k})} \exp(-\frac{E_{\alpha}(\mathbf{k})}{k_B T}) d^3\mathbf{k} \Big/ \sum_{\alpha} \int \exp(-\frac{E_{\alpha}(\mathbf{k})}{k_B T}) d^3\mathbf{k} \,. \tag{3-3}$$

Results are shown in *Figure 3-6*. At low strains the change in the scattering rate is small, but at higher strain the increase in the mean time between scattering events clearly becomes significant.

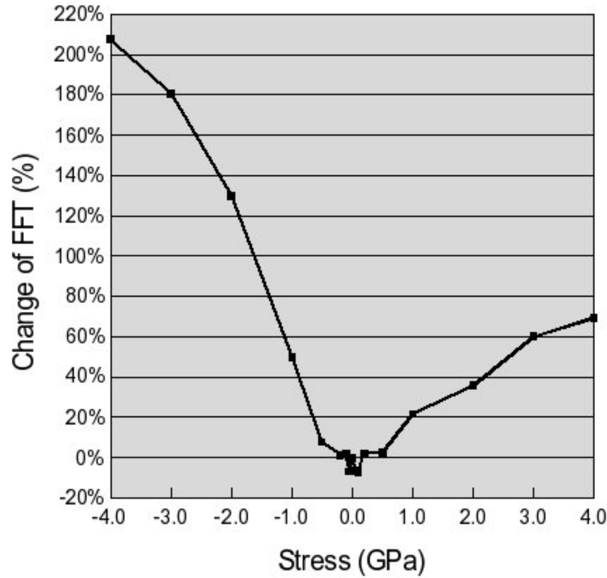

Figure 3-6: Percentile change of the thermal average of inverse scattering rates (proportional to free flight time FFT) for [110] uniaxially-stressed Si as function of stress. The zero stress FFT (phonon scattering only) is ~0.15 ps at room temperature.

3.3 Bulk Mobility

While the results obtained from Equations (3-1) and (3-3) are qualitatively

valuable, calculation of actual mobility changes is, of course, a bit more complicated than simply multiplying these two results together. And with multiple non-parabolic anisotropic bands and inelastic scattering we have turned to the use of the full-band ensemble Monte Carlo device simulator MCUT operating in the bulk limit for this purpose. Mobility calculations were performed at room temperature under constant low electrostatic fields of 1×10^3 V/cm applied in directions both parallel and perpendicular to stress. As shown in the results of *Figure 3-7*, a linear relationship between mobility and stress holds in the low stress range of –200 to 200 MPa. The slopes of these lines are close to the piezoresistance coefficient measured 50 years ago [Smi54]. However, in particular, for large compressive [110] stress, the combination of the reductions in $m_{[110]}$ (*Figure 3-4*) and in the scattering rate (*Figure 3-6*) result in a dramatic super-linear increase in $\mu_{[110]}$ that reaches 470% at 2 GPa. As for the thermal velocity under the same conditions, not until the 4 GPa limit of our simulations is approached does this mobility improvement with compressive stress finally begin to roll off toward apparent saturation. By comparison, the reduction of $\mu_{[1\bar{1}0]}$ under tensile stress quickly rolls off near saturation with increasing stress ($\mu_{[1\bar{1}0]}$ actually eventually increases again slightly at large stress) as the increase in conductivity effective mass and decrease in scattering rate work against each other. Under tensile stress the mobility advantage offered, for the perpendicular component $\mu_{[1\bar{1}0]}$ this time, is relatively much smaller that that offered under compressive stress for $\mu_{[110]}$. However, in absolute terms it is still quite

significant, exceeding 100% by approximately 1.5 GPa.

The results of these calculations, as also shown in *Figure 3-7*, are qualitatively and mostly quantitatively consistent with experimental behavior where available [Gil04] [Tho04] [Smi54] [Shi04] as well as prior theoretical calculations [Wan04]. Our bulk mobility results are most reliably compared to Smith's pioneering calculations of bulk piezoresistance coefficients [Smi54] at low stress levels up to 50 MPa. The study of piezoresistance was extended to 250 MPa for surface channel mobility with a wafer bending technique [Tho04]. Moreover, recent experiments have used SiGe heteroepitaxy in the source/drain to produce longitudinal strain in the channel where around 500 MPa stress in channel produced approximately 50% surface channel mobility enhancement [Gil04] [Shi04]. However, the comparison to these latter surface channel mobility results should be considered primarily qualitatively. While percent enhancements are shown, the absolute mobility in the surface channels is, of course, lower due to surface roughness scattering. The non-crystal-momentum-randomizing surface roughness scattering, in turn, could respond somewhat differently to band structure changes than does the essentially randomizing phonon scattering. Thus, while the near perfect agreement with measured surface channel mobility in the super-linear region near 500 MPa for $\mu_{[110]}$ under compressive stress might be considered somewhat fortuitous, the quantitative disagreement for $\mu_{[1\bar{1}0]}$ under tensile stress is not overly problematic. Indeed, in the latter case, at low stress the bulk mobility behavior calculated in this work falls closer to Smith's piezoresistance-coefficients for stressed bulk Si than does the behavior of the measured surface channel mobility. Thus, regarding mobility,

we emphasize the qualitative behavior even though the quantitative agreement elsewhere is also quite good. For thermal velocities, of course, surface roughness scattering is not an issue.

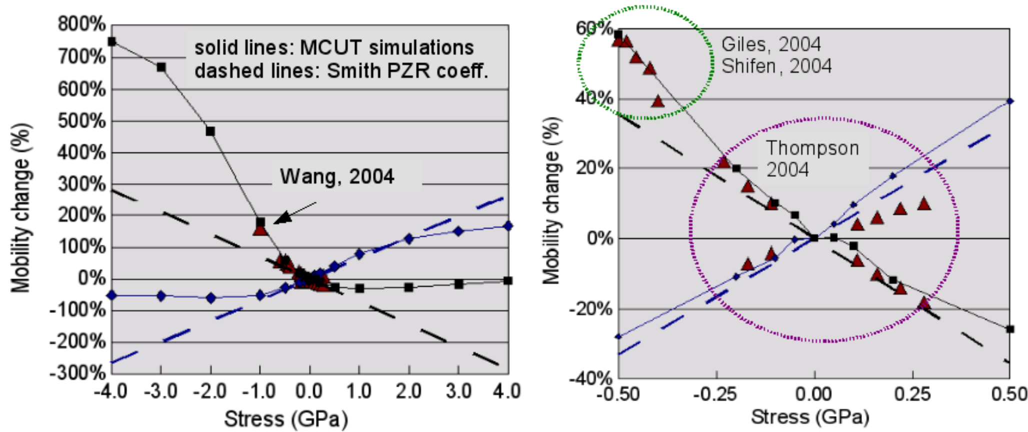

Figure 3-7: Low field mobility of uniaxially-stressed Si vs. stress strength. Dashed lines follow Smith's piezoresistance coefficient [Smi54]. Squares and diamonds plot MCUT bulk simulation results for transport parallel and perpendicular to stress directions, respectively. Experimental results from [Gil04][Tho04][Shi04] and simulation results from [Wan04] are plotted as triangles for comparison. Results are shown on two scales with the lower stress region shown in detail in right hand side.

3.4 Discussion

Figure 3-7 illustrates the potential mobility advantages of large uniaxial stress for silicon pMOS. It also illustrates that experimental work (for example, [Shi04]) is just passing from the low stress regime of near linear behavior to enter the high stress regime of strong super-linear behavior. A recent technical advance has promised a reliable stress greater than 1 GPa in pMOS channels [Arg06] placing devices into a regime where mobility enhancements could actually far exceed expectations based on linear extrapolation from Smith's piezoresistance

coefficients [Smi54].

Again, however, it has now long been known that for very short channel devices mobility is not the whole story even if it remains important, and that thermal velocity plays an increasingly important role in short channel devices [Lun97]. Therefore, calculated thermal velocities have also been presented, *Figure 3-5*. While the approximately 29% and 47% increases in thermal velocity in the [110] direction under 1 GPa and 2 GPa, respectively, of [110] compressive stress may seem small by comparison to the approximately 180% and 470% changes in mobility under the same respective conditions, in absolute terms these changes in thermal velocities are quite large. For comparison, the maximum increase in thermal electron velocity for biaxially strained Si is only approximately 17%, assuming parabolic ellipsoidal conduction band energy valleys that are shifted but not distorted by strain leading to a conductivity effective mass of approximately 0.19 m_e. For [110] tensile stressed Si, a change of approximately 32% in the electron mobility is expected at about 2 GPa, allowing for a distortion in the shape of the energy valleys themselves that results in an approximately 0.15 m_e conductivity effective mass parallel to the stress as reported in Fig. 13 of Ref [Uch05]. Finally, while the rates of increase of thermal velocity and mobility with compressive [110] stress and parallel [110] transport peak by approximately 1 GPa and 2 GPa respectively, increasing stress continues to provide significant improvements in both up to approximately the 4 GPa limit of our simulations where each finally appears near, if not quite at, saturation.

4 PMOS Wafer Crystal Orientation

4.1 Motivation

Impediments to scaling have led to the consideration of new materials and non-conventional MOSFET device structures such as Silicon on Insulator (SOI) and FinFETs. For example, CMOS fabricated on hybrid substrate are being investigated to solve the scaling limitation of devices to provide higher performance [Yan03]. In this hybrid substrate and other technologies, p-channel and n-channel devices can have different crystal orientations in surfaces. Thus, it is possible to select the best surface and channel direction for n- and p-MOSFETs separately according to the strong crystal orientation dependent of the channel mobility [Yan03][Mom02][Oni03][Mom03][Mom02a][Tag98].

The low-field mobility of Si n-MOSFETs on (001) substrate is larger than that on (111) or (110) substrate in experiments [Mom02]. For a (001) silicon substrate, quantum confinement effects in the inversion region lower the energy of the two Δ_2 valleys (valleys in the gate direction) of the six equivalent valleys in the conduction band and increase the carrier concentration in those two valleys [Fis02]. Thus, carriers have a smaller final density of states (DOS) to which to scatter, and a smaller conductivity effective mass in transport direction, both of which increase the electron mobility. For (110) or (111) substrate directions, there are 4 or 6 degenerate lowest energy valleys, respectively, and the resulting electron mobility is smaller due to higher DOS and conductivity effective mass in

each case.

However, the Si valence bands are much more complicated. The equi-energy surfaces are anisotropic and non-parabolic even for room temperature carrier distributions. In conventional Si CMOS technology, (001) wafers are usually selected for minimized surface defects that could lead to device aging, and higher n-MOSFET low field mobility. However, modern CMOS technologies have raised the possibility of separately choosing the p-MOSFET surface orientation, which stimulates both experimental and theoretical research on this topic. Furthermore, as devices are scaled toward their ultimate limit of ballistic channel transport, the role of mobility, per se, and thus the disadvantage associated with lower p-channel mobility may be reduced [Fis03].

There have been several reports about the properties of CMOS, particularly channel mobility and drive current, on wafers of various orientations [Oni03] [Mom03] [Mom02a] [Tag98]. However, the gate oxide interface quality dependence on the substrate orientation is a big concern in fabrication and in most experiments with different wafer orientations. A lot of effort has been put in Si/SiO_2 interface quality control. Still it is very difficult to isolate inherent band structure and gate quantization effects from the effects of differing surface roughness/quality, which may or may not be inherent, for different wafer orientations. This uncertainty plus the high costs and risks of novel silicon technology development provides ample motivation for device simulation work on orientation effects in p-channel devices. Among the existing device simulation methods, the quantum-corrected full-band semi-classical Monte Carlo (MC) technique is more predictive and can provide a more complete

understanding of carrier transport than drift-diffusion or hydrodynamic transport models [Jac83]. MC has the ability to accurately treat full-band structure, high energy carriers, and non-local field effects such as ballistic transport and velocity overshoot that have become increasingly important as devices are scaled down below 0.1 μm channel lengths. Via "quantum corrections" MC already has been adapted to model effects of quantum confinement in gate direction for electrons, which suggests the possibility of doing the same for holes.

4.2 Band Structure Change under Confinement

Despite the complexity of the hole band structure, conductivity is, of course, nearly isotropic in bulk. Therefore changes in transport associated with surface orientation beyond those associated with surface roughness changes, including changes between different channel directions for a give surface orientation for which surface roughness scattering should be the same, must be associated with quantum confinement effects. Here approximate treatments of quantum confinement effects for hole transport as a function of surface orientation within an otherwise semiclassical approach surface orientation as they affect transport are discussed.

The sub-band structures of Si hole bands have been re-calculated under various gate confinement field to reproduce the result of an early similar approach by Fischetti et al [Fis03]. Shown in *Figure 4-1* are the sub-band energy levels in (100) and (110) surface orientations. The terms HH, LH and SO are used here as to match the results in [Fis03]; they represent just the first, second and higher level sub-bands without the conventional " heavy " and " light " m^*

meanings. The biggest difference in the sub-band levels in (100) and (110) surface orientations lies in that the energy separation between the first and second band is much larger for the (110) surface. The sub-band shape in $k_{//}$ plane is shown in *Figure 4-2* for (100) and (110) surfaces separately. Confinement actually reduces the band warping in the first sub-band for the (110) surface, presumably by reducing the degeneracy between the hh and lh bands, and produces a reduction of the conductivity effective mass in $\left[01\bar{1}\right]$ direction in $k_{//}$ plane. The sub-band shape in (100) also changes under confinement but on a less significant scale.

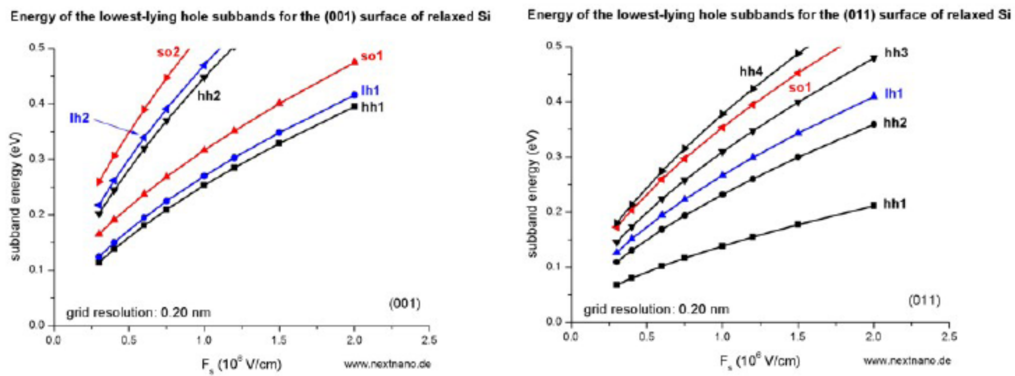

Figure 4-1: The hole subband energy levels, as measured from the surface potential, formed in a triangular well in (left): (100) and (right): (110) surface orientations. The symbols are calculated value; the connecting lines only a guide to the eye.

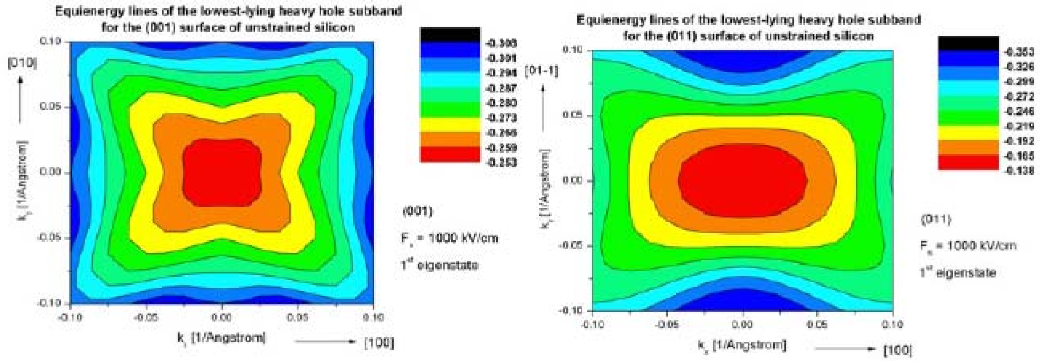

Figure 4-2: The first subband E-k dispersion in $k_{//}$ plane for (left): (100) and (right): (110) surface orientations separately.

4.3 Confinement to Stress Mapping

In the unstrained and unconfined case, the first and second energy bands (namely the heavy-hole and light-hole bands) are energy-degenerate at the Γ-point, which causes the strong warping of the lowest lying (heavy-hole) band. Differing degrees of gate-induced quantum confinement between the two bands breaks the degeneracy. The strain effect in Si, on the other hand, also breaks the energy degeneracy in hole bands. The band energy contours for [100] and [110] uniaxial tensily-stressed Si are plotted in *Figure 4-3*, which shows a very similar effect to gate confinement. The physical link of degeneracy breaking between strain and confinement and the similar results as per *Figure 4-2* and *Figure 4-3*, invites us to employ the similarity to our advantage in modeling.

In semiclassical MCUT, the band structures and associated phonon scattering rates are pre-tabulated. Though sub-band reformation occurs with applied gate bias, re-calculation of the band structure and phonon scattering rates with each bias condition and position within the channel is computationally prohibited at

this level of simulation. To maintain the computational economy of a semi-classical formalism while addressing quantum confinement effects within the complicated hole band-structure, we apply a mapping from quantum confinement effects to strain effects. The mapping methodology has been chosen to take the energy splitting between the first and second hole band (or in other terms, heavy-hole and light-hole band) as the mapping gauge. The (001) surface confinement-induced and strain-induced energy splittings are plotted in *Figure 4-4*, which could be understood as if a (001) confinement field of 0.5 MV/cm maps to 0.4 GPa tensile stress in (001) direction. The [110] mobility change as a function of (100) uniaxial tensile stress is plotted in *Figure 4-5*. Following the mapping methodology, it can be extrapolated that the mobility under confinement field of 0.5 MV/cm is about 11% larger than bulk mobility value, when other confinement effects, such as quantum-confinement enhanced phonon scattering (due to increased initial and final state overlap rather than densities of state changes that are considered) and surface roughness scattering, are not considered.

In the case of (110) surface orientation, [110] uniaxial stress is used to map to the quantum confinement. The confinement-induced band splitting is much larger than that in (001) orientation. Once again, the mapping relationship is shown in *Figure 4-6*, and the stress-induced mobility change is shown in *Figure 4-7*. For example, 0.5 MV/cm confinement causes the first and second hole band separation of 58.6 meV; this is similar in effect to 1.2 GPa [110] uniaxial stress, which is has been shown in *Figure 4-7* to induce a 230% mobility change over unstrained Si in the conduction channel. Comparing (110) and (100) surface orientations, the mobility ratio would be approximately (1+230%)/(1+11%)≈3.

This is qualitatively comparable to the theoretical calculation result (μ(011)[110] to μ(001)[110] ratio ≈ 2 at surface charge density of 8×10^{12} cm^{-2}) of the accurate six-band **k•p** Hamiltonian method, as well as the experimental results (μ(011)[110] to μ(001)[110] ratio ≈ 3 at surface charge density of 1×10^{13} cm^{-2}) shown in Fig. 11 of Ref. [Fis03]. To match experimental results [Sat69][Sat71][Fis03], full-quantum phonon scattering and surface roughness scattering need to be considered or more complex quantum correction method needs to be applied with semi-classical MCUT.

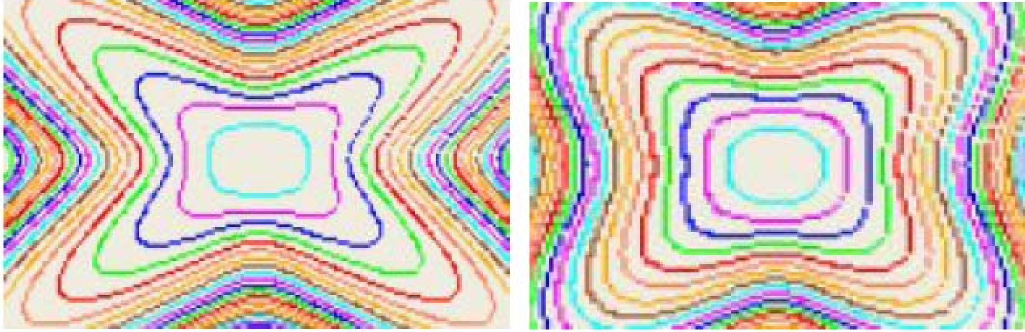

*Figure 4-3: The first band E-**k** dispersion in $k_{//}$ plane. (left) [100] uniaxial tensile stress, (100) plane; (right) [110] uniaxial tensile stress, (110) plane, spanned by [001] and $[1\bar{1}0]$ axes.*

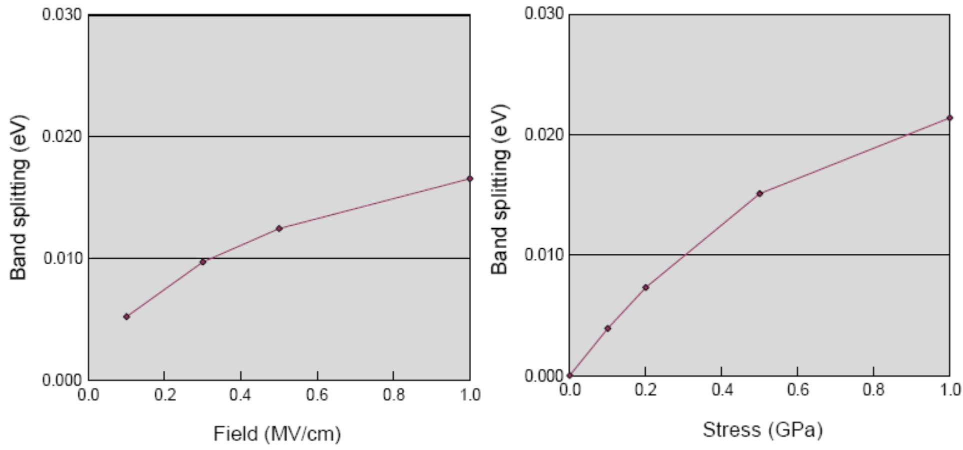

Figure 4-4: Band splitting between the first and second hole bands induced by (left) [001] gate confinement field; and (right) [001] uniaxial stress.

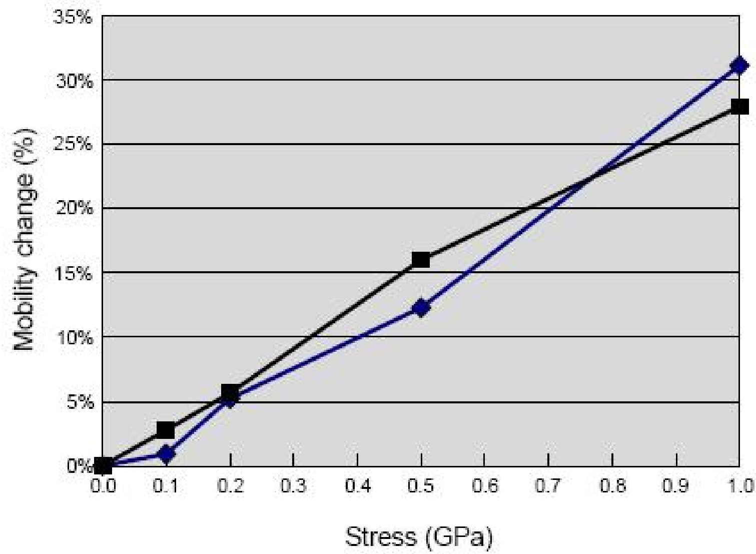

Figure 4-5: [110] Hole transport mobility in [100] and [110] directions in (001) surface orientation. Curves show percentile change with [001] uniaxial tensile stress.

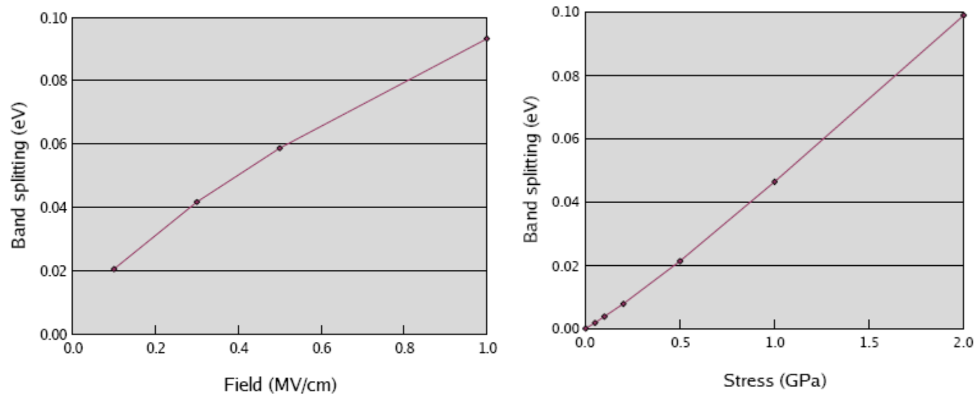

Figure 4-6: Band splitting between the first and second hole bands induced by (left) [110] gate confinement field; and (right) [110] uniaxial stress.

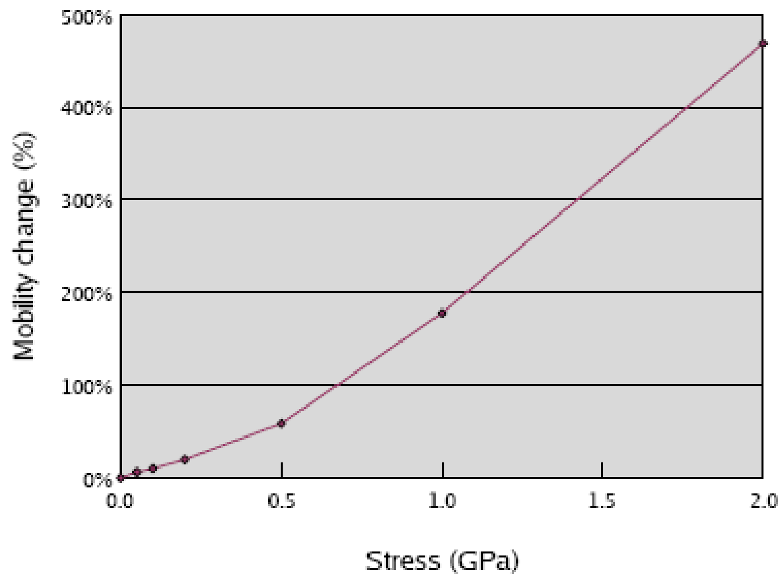

Figure 4-7: [110] transport mobility percentile change with [110] uniaxial stress.

5 Optimum Combination of Stress and Crystal Orientation

5.1 Position Dependent Stress

As described earlier in related published process techniques (for example, [Ran05]), uniaxial strain is introduced into pMOS transistors by depositing SiGe alloy in a recessed etched trench in source and drain regions on each side of channel. For a pMOS device of $Si_{0.83}Ge_{0.17}$ source/drain regions, simulation results showed approximately 500 MPa of uniaxial compression in the channel [Tho05]. This technique is only applicable for nanometer scale channel lengths; in longer devices the compressive force would not have penetrated far enough to strain the entire channel. The strain field is also function of the heat generated during processing after the source/drain epitaxial growth. Even in devices with ~100 nm channel length as shown in *Figure 5-1*, the strain field in the channel is far from uniform. For this position-dependent strain field, a new feature is added to MCUT to handle this.

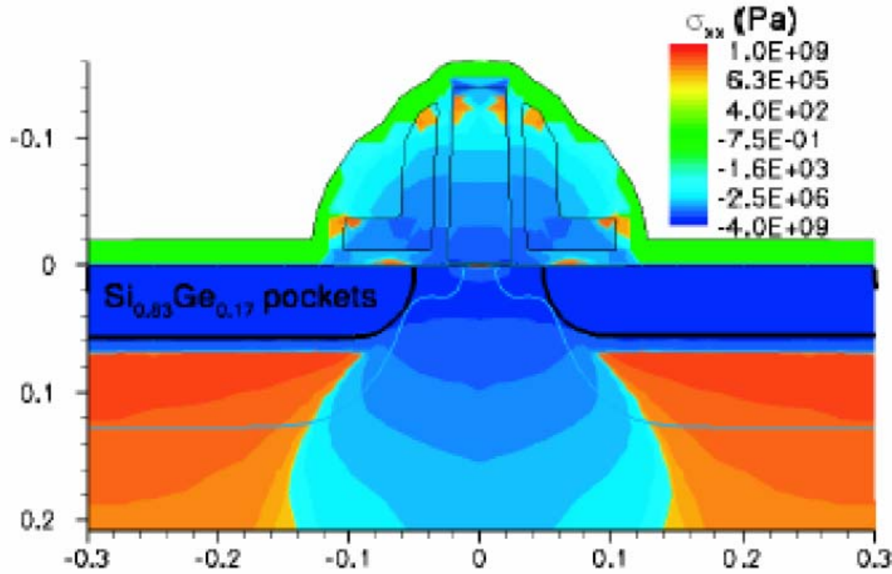

Figure 5-1: Stress along the channel in a strained Si pMOS transistor, modeled by ISE04 [Ran05]

In order to handle position-dependent materials (i.e., different stressed Si), MCUT needs to read in the band structures and phonon scattering rates of those materials. First a second order polynomial fitting is assumed for every **k** point of the entire *E*-**k** dispersion of various stressed Si materials, as

$$E_{\sigma}(\mathbf{k}) = E_0(\mathbf{k}) + a(\mathbf{k})\cdot\sigma + b(\mathbf{k})\cdot\sigma^2 . \tag{5-1}$$

K-dependent a(**k**) and b(**k**) are calculated by Matlab on each and every **k**. The polynomial form helps to reduce memory usage for storing band structures. However, this technique cannot be used in the reverse *E*-**k** lookup and phonon scattering rate tables. For a certain energy at a given stress value, the crystal momentum cannot be found by extrapolating from the **k** found with zero stress value. The scattering rates do not follow polynomial function of stress either. So these two tables require a great deal of RAM during simulation. In reverse *E*-**k** lookup process, MCUT first rounds up the stress value to one of the columns in

the table, then finds an E-$\mathbf{k}$ pair with that approximate stress. This process is only to provide a set of possible $\mathbf{k}$ vectors for the given energy in scattering events; the following energy and momentum conservation check still uses the accurate E-$\mathbf{k}$ polynomials. The phonon scattering rate calculation uses the interpolation method.

Strain in Si also affects the band gap and band edge. The effect in transport is explained in the equation below (similar to Eq. (2-11)).

$$
\begin{aligned}
\hbar \frac{d\mathbf{k}}{dt} &= e\nabla_{\mathbf{r}} V(x,z) - \nabla_{\mathbf{r}} E_{qc}^{[\gamma_{eff}(x,z)]}(\mathbf{k}_{\Gamma}, x, z) \\
&\quad - \nabla_{\mathbf{r}} \gamma_{eff}(x,z) \left\{ \frac{\partial}{\partial \gamma} [E^{(\gamma)}(\mathbf{k}) - E^{(\gamma)}(\mathbf{k}_{\Gamma})] \right\} \Bigg|_{\gamma=\gamma_{eff}(x,z)}
\end{aligned} \tag{5-2}
$$

In simulation, a stress value is allocated to each mesh point in the device. Carriers then are automatically adjusted for their energy and velocity depending on their physical location. Phonon scattering rates and mechanisms also become function of the local stress value. The library file size increase for one dimension and the code requires around 1 GB RAM to run. The simulation speed is not noticeably decreased as the dominating computation is still the Poisson equation solver.

5.2 Simulations with Position-Dependent Stress

The device simulation was carried out on 50 nm "well-tempered" pMOS structures [Mit00]. For this demonstrative purpose, the stress profile within the channel is simplified compared to that shown in *Figure 5-1*. The stress increases laterally from source to channel to reach a maximum stress level, stays constant in

the channel, then drops back to zero in the drain end. This setup is intended to model the source/drain SiGe epitaxy technology. The channel maximum stress is varied from 0.0 (unstrained) to 1.5 GPa. The drain current enhancement is plotted in *Figure 5-2*.

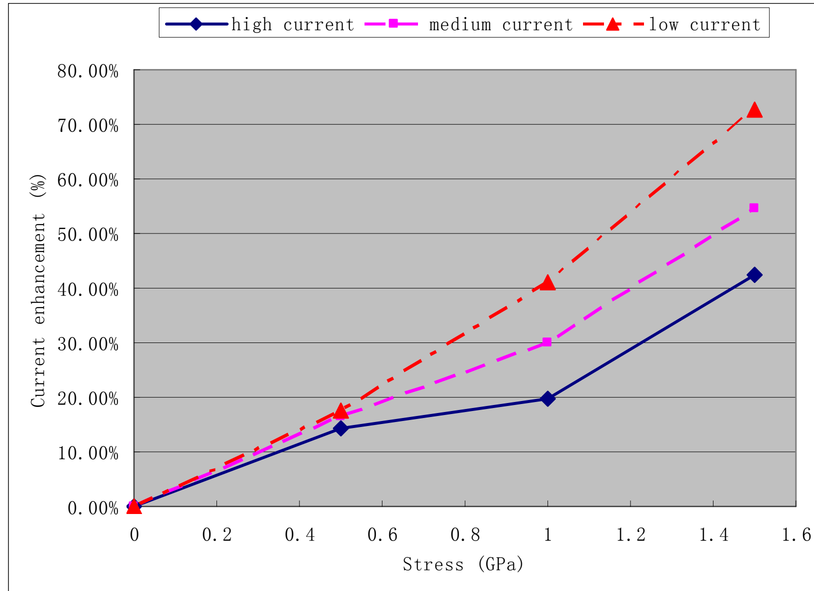

Figure 5-2: Drain current enhancement as a function of channel maximum stress. "High current" simulation was achieved with gate bias -1.5 V and drain bias -0.5 V; "medium current" simulation was with gate bias -1.2 V and drain bias -0.2 V; and "low current" simulation was with gate bias of -1.2 V and drain bias -50 mV. Source and substrate were grounded.

Comparing *Figure 5-2* with bulk properties *Figure 3-5* and *Figure 3-7*, we find that device current response with stress follows the thermal velocity change under the "high current" condition. Under the "low current" condition, the current enhancement rate is higher as mobility plays a greater role with the reduced drain voltage. However, as discussed in [Lun97], within short channel devices (50 nm in this example), the limiting factors under normal operating conditions are source to channel injection efficiency which is related to but not the same as mobility

(which shall be discussed in Section 7) and carrier injection velocity (which is the thermal average velocity shown in *Figure 3-5*.)

5.3 Combination of Stress and Quantum Confinement

The concept of confinement-to-stress mapping is introduced in Section 4.3. Since the purpose of both applying stress in channel and choosing wafer orientation is to increase performance (especially for pMOS), optimizing both simultaneously should provide the maximum performance enhancement. For a conventional (001) wafer orientation, there are still gate confinement effects on sub-bands, phonon scattering rates and confinement-induced scattering. Though MCUT runs on a semiclassical environment, various correction methods are used to accommodate the quantum effects. In the uniaxially stressed Si pMOS simulation, modeling quantum effects is important as confinement effects in valence band are both significant and quite complicated. *Figure 4-2* shows the first sub-band of unstrained Si in (001) and (110) wafer orientations. For uniaxially stressed Si, the consequences of confinement are still under investigation with (mostly) the $\mathbf{k} \cdot \mathbf{p}$ method. Our findings show that in conventional (001) wafers, the quantum confinement effects are insignificant as compared to those for uniaxial stress in the [110] direction. As shown in *Figure 5-3* and *Figure 5-4*, the HH band (i.e., the first sub-band) does not change shape much under confinement the field. Because of the heavy-hole/light-hole energy splitting, most of the carriers stay in the HH band under moderate field, so the characteristics of transport (for example, low field mobility) are almost determined by the HH band only. The change of E-$\mathbf{k}$ diagram under

confinement, as shown in *Figure 5-4*, is much less than the anisotropy induced by uniaxial stress in [110] direction. So in MCUT simulations, the quantum effects being modeled are limited to the confinement-induced phonon scattering change.

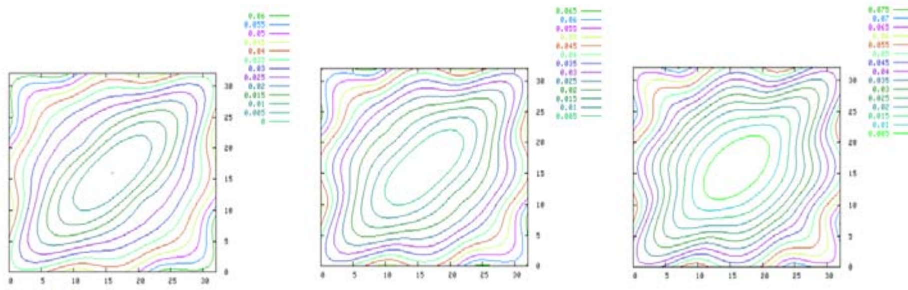

Figure 5-3: 2-D energy contours of heavy-hole band (or in other term, first subband),for a [001] confinement field of 0.1 MV/cm, 0.5 MV/cm, and 1.0 MV/cm separately, from left to right, in energy separation of 5 meV, from band edge upto 100 meV. The range of k is from -0.1 Å^{-1} to 0.1 Å^{-1}. At different confinement field (0.1, 0.5 and 1.0 MV/cm), the band shape changes but a little.

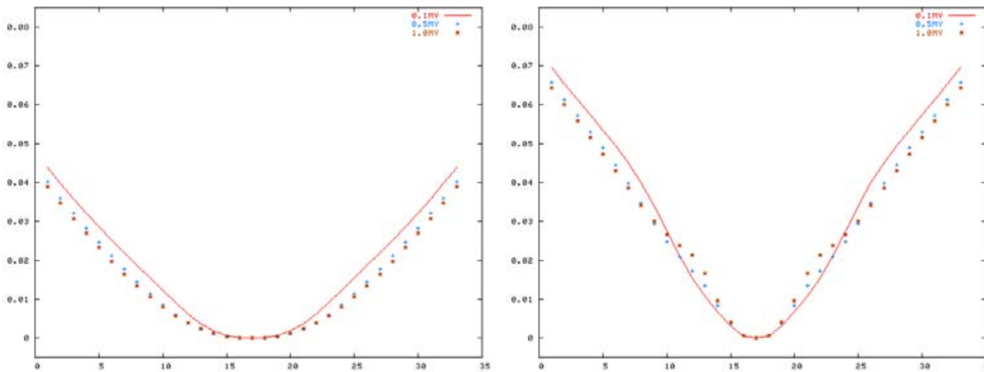

Figure 5-4: This plotting shows the above figure, the E-k dispersion, in $[110]$ *and* $[1\bar{1}0]$ *direction, with individual lines represent different* [001] *confinement fields of 0.1, 0.5 and 1.0 MV/cm. The range of k is from -0.1 Å^{-1} to 0.1 Å^{-1}, with energy unit in eV. The change is rather small, which implies that uniaxial stressed Si behaves similarly in strong confinement as in bulk.*

However, for the non-conventional (110) wafer orientation, the result is quite different. The band diagram change under confinement is dramatic as shown

previously in *Figure 4-1* and *Figure 4-2*. The ellipsoidal shape of HH band, as compared to the warped shape of bulk unstrained Si HH, is actually similar to the band diagram of [110] uniaxial stressed Si. *Figure 5-5* compares the first sub-band contours of strongly stressed Si with little confinement to that for confined but unstrained Si. The strongly confined band structure (left hand side) gives an average thermal velocity of 1.353×10^7 cm/s in vertical direction, and 8.961×10^7 cm/s in the horizontal direction. The strongly stressed band structure (right hand side) gives an average thermal velocity of 1.395×10^7 cm/s in vertical direction and 9.135×10^7 cm/s in horizontal direction. The figures are similar and velocities are close, which provides confidence in ability to map between the two scenarios, that is to map stress and confinement to just stress.

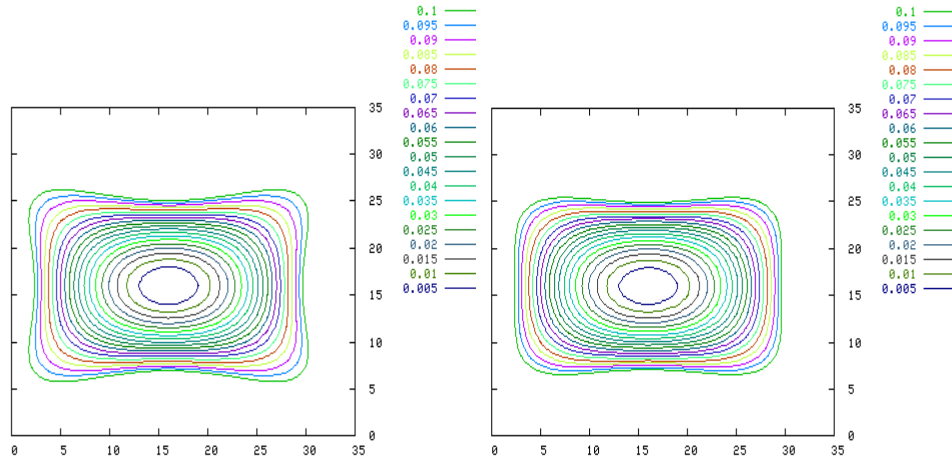

Figure 5-5: Comparison between (110) orientation confinement effect and uniaxial stress effect. In the left hand side case, there is no stress; a field of 1.0 MV/cm is applied in gate (110) direction. In the right hand side, the confinement field is small 0.1 MV/cm, but a compressive uniaxial stress of 3.0 GPa is applied on transport direction. The contours show the first sub-band (heavy-hole) from band edge to 100 meV, in steps of 5 meV.

Though this work is in its infancy, this mapping between strain and confinement

should allow consideration of both simultaneously, as discussed for electrons in Section 2. Being able to investigate wafer orientation effects in Si pMOS, which are believed to be a quantum mechanical effects, with semiclassical Monte Carlo technique brings with it all of the existing advantages of full-band semiclassical MC. For example, the Poisson-Schrödinger self-consistent infrastructure is well built with MCUT; scattering mechanisms are complete and well calibrated; non-local field effects and other short channel effects are treated naturally, and device structures can be easily read in.

6 Dual Gate NMOS Simulation

6.1 Alternative quantum correction scheme

Theoretically, dual gate Si nMOSFETs (the technology is also called as "double gate nMOSFETs", "FINFETs", "SOI", etc) simulation is quite similar to conventional bulk Si nMOS simulation, as long as the simulator takes care of the structural difference in its self-consistent Poisson's and Schrödinger's equation solvers. MCUT deals with the quantum correction via Schrödinger's equation solver and the "effective strain" mapping scheme (Section 2.4). Under gate confinement, the band edge of energy valleys Δ_2 and the Δ_4 becomes non-degenerate, due to different effective masses in gate direction. The induced valley shift and inter-valley phonon scattering rate change are modeled by mapping to strain effects.

In dual gate nMOS, with Si body thickness around 100 nm or lower, the confinement effects get stronger than in bulk devices. Full quantum study shows the phonon scattering rate as a function of confinement. Scattering rate increases as gate bias increases. With the previous quantum correction method, this effect is not modeled. A new quantum correction method for ultra-strong confinement becomes necessary, especially for the dual gate nMOS simulations.

In the full quantum 2-D MC simulator, SEMC2D [Che03], the subband levels in the device are calculated, then the phonon scattering rates (i.e., imaginary self-energies) are accordingly calculated. The scattering rates as a function of carrier

kinetic energy and confinement are drawn in *Figure 6-1*. The effects shown there are quite significant especially in ultra-small scale CMOS while conventional semiclassical Monte Carlo simulators usually neglect them and might have used empirical surface roughness scattering models to match experimental results. In *Figure 6-2* the scattering rates are re-plotted with energy reference fixed at band edge at no confinement. At this fixed energy reference, the phonon scattering rates at various confinement situations "converge" well with the bulk scattering rates value.

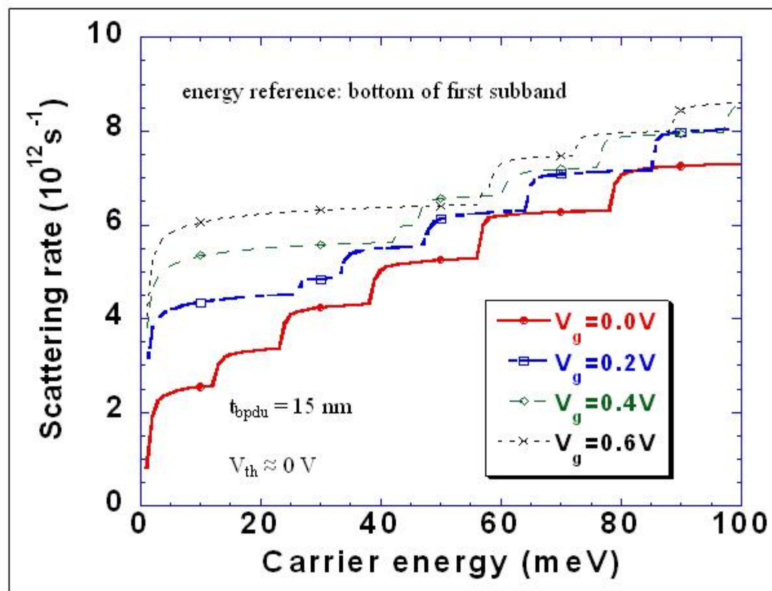

Figure 6-1: Phonon scattering rates as function of carrier energy in confined channels. Scattering rates form steps with sub-band levels, while they increase with confinement.

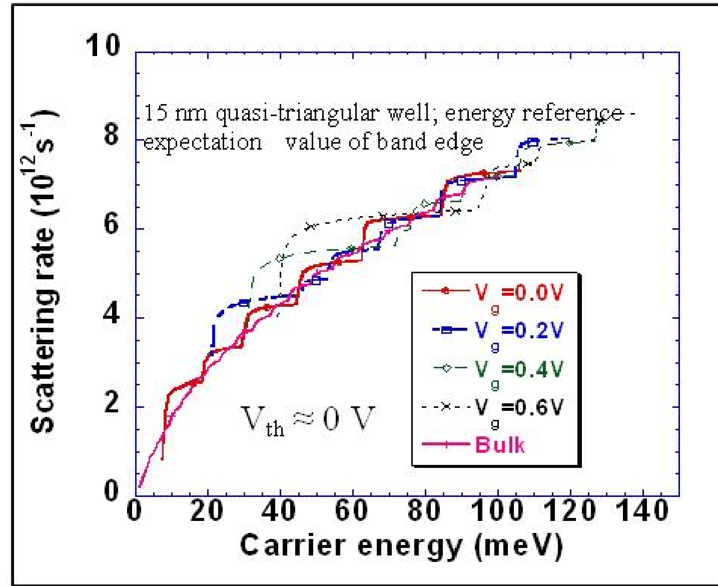

Figure 6-2: Replot scattering rates with energy reference fixed to bulk band edge at no confinement.

This interesting phenomenon can be used in MCUT as a new method for the quantum correction on phonon scattering rates. We consider the summation of carrier kinetic energy and the quantum shift of band edge to be "carrier quantum energy", as shown in *Figure 6-3*. In MCUT, the quantum shift of band edge is the first sub-band level, position and valley dependent, solved by the one-dimensional effective-mass Schrödinger's equation normal to the interface. In this version of MCUT, the phonon scattering rates are function of the "total energy", summation of the first sub-band energy ΔE and classical kinetic energy. As shown in *Figure 6-3*, due to the difference in m* in confinement direction, ΔE are different for different valleys/bands. Besides the effect that intra-valley scattering rates are increased in inversion layers, some inter-valley/band scattering between valleys/ bands become energy prohibited, as described by equation set:

$$
\begin{aligned}
&\Gamma^{(QC)}_{(1\to1)}(E_{CL})=\gamma_{(1\to1)}(E_{CL}+\Delta E_1);\\
&\Gamma^{(QC)}_{(1\to2)}(E_{CL})=0;\quad (if\ E_{CL}+\Delta E_1 \pm E_{ph} < \Delta E_2)\\
&\Gamma^{(QC)}_{(1\to2)}(E_{CL})=\gamma_{(1\to2)}(E_{CL}+\Delta E_1);\quad (otherwise)
\end{aligned}
\tag{6-1}
$$

where $\Gamma^{(QC)}$ is the quantum corrected scattering rate and γ is the pre-tabulated bulk scattering rate, E_{CL} is the carrier kinetic energy, E_{ph} is the maximum phonon energy and subscript 1, 2 denotes different bands/valleys. This treatment effectively changes the density-of-states in each band/valley to emulate the sub-band reformation and phonon scattering rate increase in inversion layers.

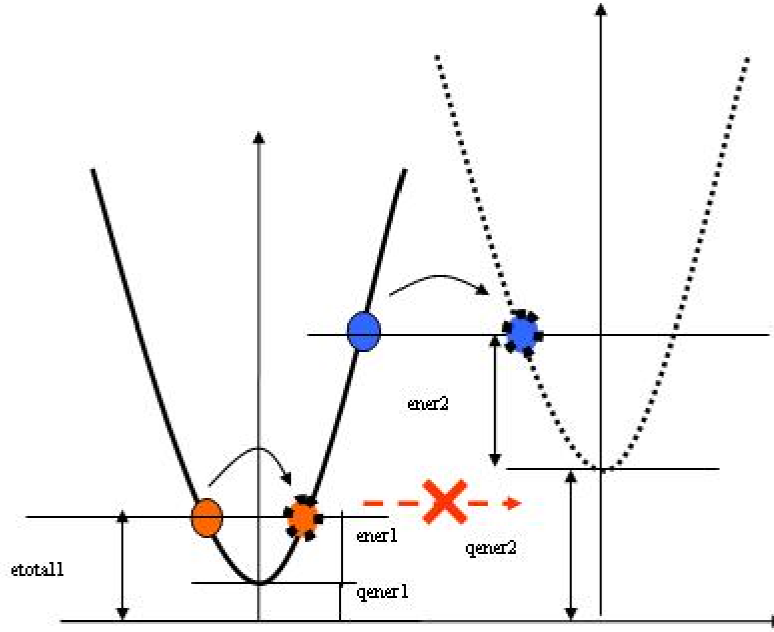

Figure 6-3: With confinement, when valleys/bands become non-degenerate, intra- and inter-valley phonon scattering rates are function of "total energy", while some inter-valley scattering becomes energy-prohibited. This effectively alternates valley/band concentration in simulations.

6.2 MCUT simulations with dual gate Si nMOS

The effect of quantum confinement is more evident in dual gate FINFETs. When FINFET body thickness decreases, the confinement in body increases, which allocates carriers to low energy valleys (*Figure 6-4*) and also increases scattering rates as confinement shifts subbands (*Figure 6-5*). In Si nMOS case, the low energy Δ_2 valleys have lower conductivity m*, so the two effects compete with each other. At high confinement, the phonon-limited mobility decrease with body thickness [Sho99] (*Figure 6-6*).

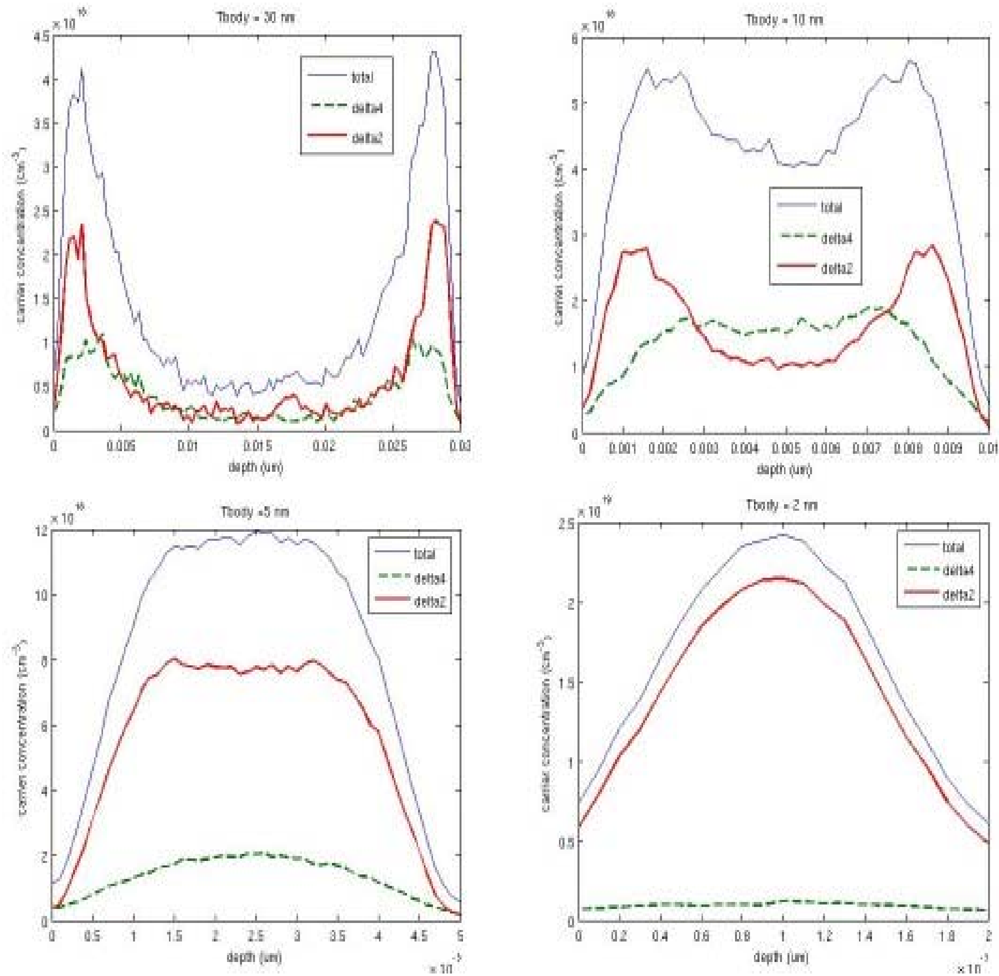

Figure 6-4: Valley population in Δ_2 and Δ_4 valleys in Si dual gate MOSFET body. t_{body} shrinks from 30 nm down to 2 nm. The solid red lines represent Δ_2 valley concentration and the green dashed lines for Δ_4 valleys.

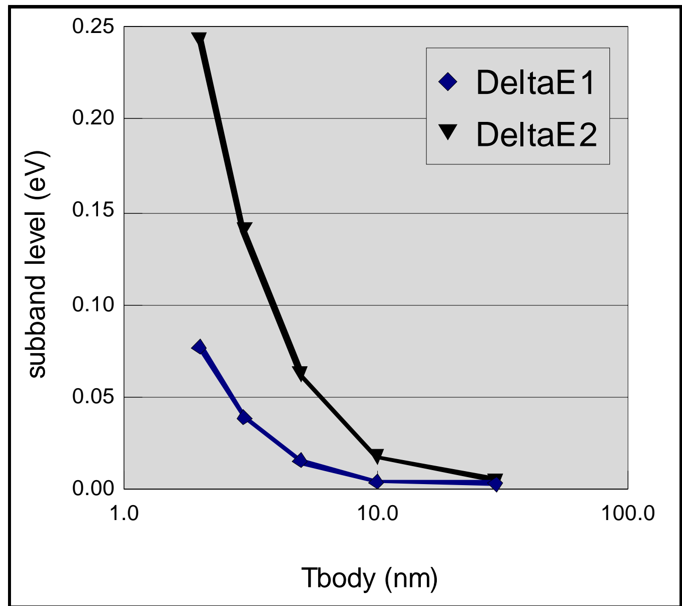

Figure 6-5: Δ_2 and Δ_4 valley subband level in Si dual gate MOSFETs, with the energy denoted by $\Delta E1$ and $\Delta E2$.

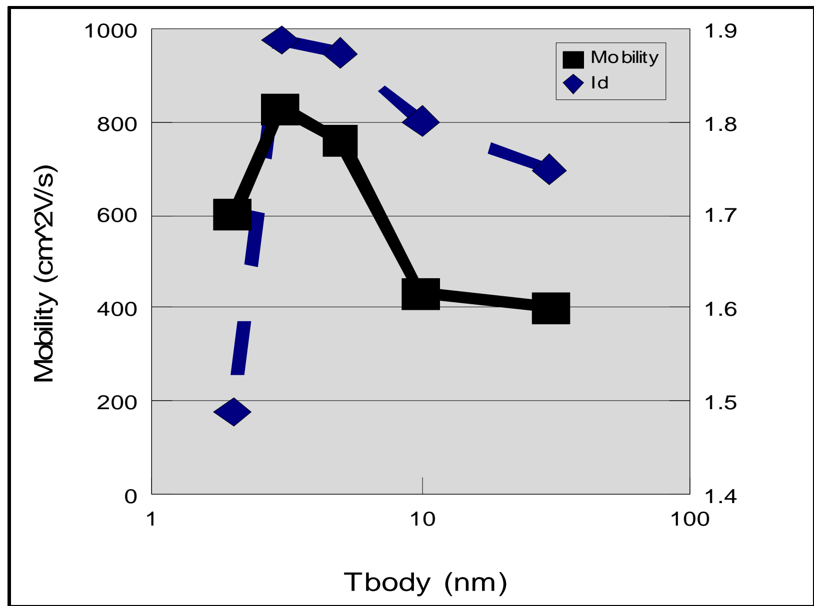

Figure 6-6: Phonon-limited electron mobility in Si dual gate MOSFETs. The confinement enhanced surface roughness scattering is not included here.

7 Injection Efficiency with Bandgap Engineering

7.1 Motivation

Strained-SiGe-on-Si has been tried in vertical pMOS structures. Holes mobility in strained-SiGe is subject to the competing effects of enhancement by band splitting and degradation by alloy scattering as compared to Si. Unfortunately, the highly strained SiGe on top of Si substrate cannot be grown thick before it relaxes, so that the mobility-enhancing effect of band splitting is not generally attractive. A potential new variation on this structure is the use of bandgap engineering to improved source-to-channel injection efficiency. When Ge mole fraction varies (or forms hetero-junction steps) in the channel, the transporting carriers gain (or loose) kinetic energy at these hetero-junction steps (*Figure 7-1*). When carriers gain enough energy by passing over one of these steps it should trigger optical phonon emission, and although they scatter, they should loose enough energy to prevent them from being scattering back into the source. This is not the case for the competing processes of ionized impurity, surface roughness and low-energy acoustic phonon scattering which dominate for low energy carriers. Thus band-gap engineering might be able to increase device injection efficiency.

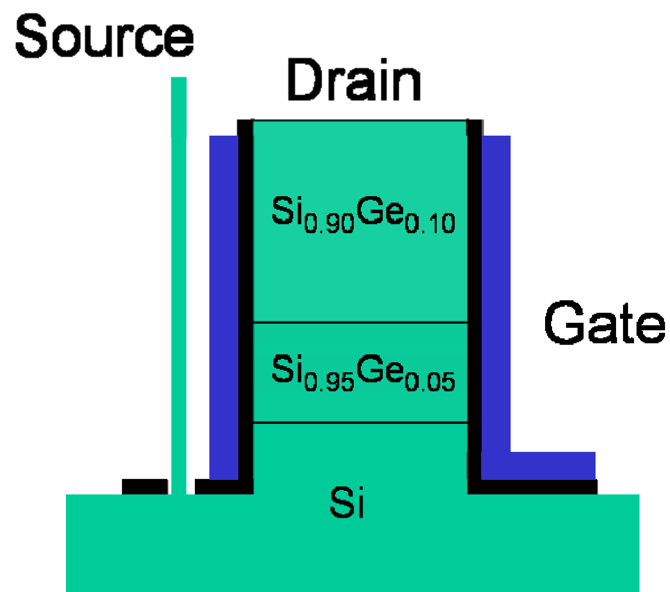

Figure 7-1: Vertical pMOS structure with strained-SiGe grown on top of Si. Note the Ge mole fraction in SiGe forms steps from source to drain.

7.2 Simulation Results

The vertical pMOS structure with Ge mole fraction steps will be carefully designed to maximize channel injection efficiency and thus drive current. Injection efficiency enhancements are expected to resemble the test result shown in *Figure 7-2*.

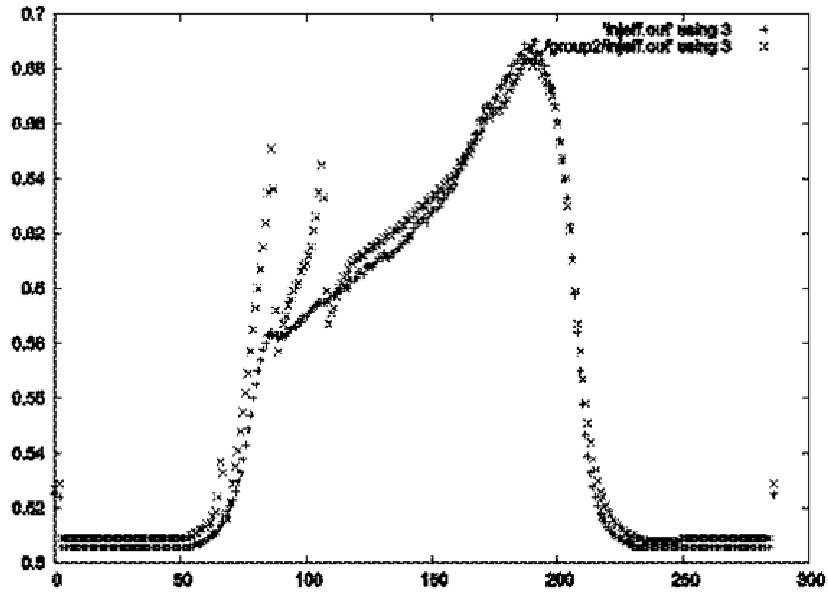

Figure 7-2: Injection efficiency comparison between a device with Ge mole fraction steps and a control device. Note the peaks represent the Ge mole fraction steps in channel.

8 Conclusions and recommendations

8.1 Conclusions

This work focused on semiclassical Monte Carlo simulation with Si CMOS devices. Though the basic infrastructure of semiclassical MC device simulation has been established for a long time, work to improve the method and to extend the method to new applications has never stopped.

As the feature size of CMOS shrinks into nanometer regime, quantum mechanicals corrections in semiclassical MC have evolved from a novel to a near-mandatory feature. Since Schrödinger's equation solver was introduced into MC as the base for quantum correction, the algorithm has been continually improved. In MCUT, the valley-dependent quantum correction based on valley effective mass Schrödinger's solver was the first improvement over older MoCa [Win03]; the concept of "effective stress" based on the combination of stress-induced and confinement-induced valley energy splitting was the direct result of that.

The valley-dependent quantum correction worked well for nMOS, in which electrons are the majority carriers. The pMOS, or hole carrier, acts differently with strong confinement. The full quantum analysis showed the valence band structures were altered in shape by the confinement not merely shifted in energy, thus changing the hole transporting characteristics. The conventional quantum correction algorithms do not apply for this category; in fact, there had been less hole transport study than electron's with MC technique due to the complexity of valence band structure. In this Ph.D. work, the hole transport was addressed

within a full-band semiclassical MC technique, and a new quantum correction mechanism on the full band structure was applied to strong confinement situations. The new quantum correction method analyzes the similarity of band structures of Si under gate confinement and Si under uniaxial stress, and then utilizes it to do the correction directly on band structures.

While not perfect, these quantum corrections for otherwise semi-classical MC aim to achieve a balance between accurate modeling and acceptable computation complexity. Compared with a first-principles treatment of transport within individual sub-bands, the correction technique saves the re-calculation of band structure over each bias and position in the device; rather, a correction based on the confinement strength is added to the bulk band structure for a very low cost of computation power. The accuracy requirement is also guarded by the corrected band structures mimicking the real band structure under confinement.

Modeling of Stress Si has been extensively discussed in this thesis. Modeling stressed Si CMOS the major application of the tool MCUT over the last five years. Since the semiconductor industry entered the sub-micron regime in 1990's, the high mobility in stressed Si has raised widespread interest. Compared with other contemporary non-conventional CMOS technologies, such as SOI, FINFET, nanowire, etc, use of stressed Si requires the least amount of change to the current Si fabrication process, and it has also been proven effective with various experiments, and recent generations of Si products [Tho02]. Moreover, there are many variations on the use of stress. Biaxial stress with Si over SiGe substrate and uniaxial stress with SiGe epitaxy are the two main topics in this study. At the moment stress is limited to the major surface orientations such as <100> and

<110> due to compatibility with existing fabrication.

For MCUT simulation of non-conventional channel materials, calculation of the bulk strained band structure is the first step. The full Brillouin Zone energy (E) vs. wavevector **k** dispersion is calculated and recorded in the library; then phonon scattering rates are then calculated with this full band structure and also recorded. A full quantum analysis on band structure under gate confinement is carried out and an empirical mapping equation between confinement strength and stress is established. In MC simulation, full band structure and phonon scattering rates (together with other empirical scattering rates) are read in for carrier transport and scattering events. The Poisson's equation solver determines the potential profile which directs carrier transport, and the Schrödinger's equation solver decides the correction to put on the band structures and phonon scattering rates. These two solvers run iteratively during simulation to update the potential and corrections depending on carrier concentration. Thus MCUT self-consistently runs the simulation till stable result is achieved.

With all the advantages of the MC technique in modern simulation field, MCUT is still not a perfect tool. Perhaps quantum device simulation will eventually replace semiclassical simulation as the reference standard. However, while full quantum transport simulation is still in the development phase and struggling to accurately model scattering, semiclassical MC takes advantage of the well-established infrastructure and empirical scattering models to satisfy the simulation needs of contemporary and at least near future generations of semiconductor technology.

8.2 Recommendations for future research

Though semiclassical MC may be doomed to be replaced by full quantum transport simulation as a reference standard in the future, for the current and near future device generations there is still both value and room for improvement. Even with fully quantum transport simulators available in the future, semiclassical Monte Carlo simulation will likely remain a more practical alternative for many applications. (Certainly, drift-diffusion and hydrodynamic simulators are still being used and even improved upon long after semiclassical Monte Carlo became the reference standard.) Over the past few years MCUT has seen many modifications and improvements, and its simulation capabilities have been increased greatly over its predecessor semi-classical MC simulators. So the first recommendation is now simply the application of MCUT to various topics in the current generation of CMOS.

Most of the uniaxial stressed Si CMOS are built on conventional (100) wafer surface with <110> channel orientation, to allow use of existing fabrication technology. As discussed in Section 4, PMOS devices built on (110) wafer orientation have higher mobility than conventional ones, which was attributed to different gate confinement effects in (110) wafers. The work to map this quantum confinement effect to uniaxial stress effect has been (partially) completed (Section 4.3), and a proposal to combine the effect of quantum confinement and channel stress is made in Section 5. This may be an interesting topic to continue research on. If simulation indicates that the benefit of high mobility devices exceeds the cost, it may provide a model for the next generation

of CMOS.

The conceptual injection enhancement SiGe PMOS discussed in Section 7 also is worthy of continual research. Despite the difficulties in real fabrication of these devices, the theoretically increased injection efficiency is a unique feature. In ultra short channel MOSFETs, the source-to-channel injection efficiency is more important than channel carrier mobility, and the ways of changing the injection efficiency are very limited. Simulation is, of course, much easier than actually building SiGe PMOS, but it is still a challenge. Continual research on this subject could be rewarding.

The MCUT program could see continuous improvements too. One on-going project is the addition of Pauli exclusion in the calculation of phonon scattering. Impurity scattering, alloy scattering and surface roughness scattering are largely based on empirical formula. For new materials and structures, it is likely that the formulae will need to be updated, or, better replaced with more first-principles full band scattering rates (with quantum correction).

References

[Arg04] R. Arghavani, Z. Yuan, N. Ingle, K-B Jung, M. Seamons, S. Venkataraman, V. Banthia, K. Lilja, P. Leon, G. Karunasiri, S. Yoon, and A. Mascarenhas, "Stress Management in Sub-90-nm Transistor Architecture", IEEE Transactions on Electron Devices, 51 (10), pp. 1740 – 1743, 2004.

[Arg06] R. Arghavani, L. Xia, H. M'Saad, M. Balseanu, G. Karunasiri, A. Mascarenhas, and S. E. Thompson, "A reliable and manufacturable method to induce a stress of >1 GPa on a p-channel MOSFET in high volume manufacturing", IEEE Electron Device Letters, 27 (2), p. 114, 2006.

[Bea92] R. E. Beaty, R. C. Jaeger, J. C. Suhling, R. W. Johnson, and R. D. Butler, "Evaluation of piezoresistive coefficient variation in silicon stress sensors using a four-point bending test fixture", IEEE Transactions on Components, Hybrids, and Manufacturing Technology, 15 (5) , pp. 904 – 914, 1992.

[Bir02] S. Birner, S. Hackenbuchner, J.A. Majewski, D. Mamaluy, M. Sabathil, G. Zandler, "nextnano3 – a state-of-the-art simulation tool for 3D quantum nanodevices", Annual Report 2001, Walter Schottky Institute, TU Munich, 2002.

[Can75] C. Canali, C. Jacoboni, F. Nava, G. Ottaviani and A. Alberigi-Quaranta, "Electron drift velocity in silicon", Physical Review B, 12(4): 2265 – 2284, 1975.

[Che76] J. R. Chelikowsky and M. L. Cohen, "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blend semiconductors," *Phys. Rev.* **B** 14, p. 556, 1976.

[Che80] A. B. Chen and A. Sher, "Electronic structure of III-V semiconductors and alloys using simple orbitals," *Phys. Rev*. **B** 22, p. 3886, 1980.

[Che03] W. Chen et al, Physica E, 19, p. 28, 2003.

[Chi04] P. R. Chidambaram, "35% drive current improvement from recessed-SiGe drain extensions on 37 nm gate length PMOS," in Tech. Dig. Symp. VLSI, ses. 6.1, 2004.

[Fan04] X.-F. Fan, X. Wang, B. Winstead, L. F. Register, U. Ravaioli, and S. K. Banerjee, "MC simulation of strained-Si MOSFET with full-band structure and quantum correction", IEEE Transactions on Electron Devices, Vol. 51, No. 6, p. 962, June 2004.

[Fis88] M. V. Fischetti and S. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," *Phys. Rev*. **B** 38, p. 9721, 1988.

[Fis96] M. V. Fischetti and S. E. Laux, "Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys", Journal of Applied Physics, 80 (4): 2234 – 2252, 1996.

[Fis02] M. V. Fischetti, F. Gamiz, and W. Hansch, "On the enhanced electron mobility in strained-silicon inversion layers," Journal of Applied Physics, 92 (12): 7320 – 7324, 15 December 2002.

[Fis03] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, "Six-band **k•p** calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness", Journal of Applied Physics, 94 (2), p. 1079, 2003.

[Fri89] P. Friedel, M. S. Hybertsen, and M. Schluter, "Local empirical pseudopotential approach to the optical properties of Si/Ge superlattices," *Phys. Rev*. **B** 39, p. 7974, 1989.

[Gha03] T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A. Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, M. Bohr, "A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors", IEDM Technical Digest, pp. 11.6.1 - 11.6.3, 2003.

[Gil04] M. D. Giles, M. Armstrong, C. Auth, S. M. Cea, T. Ghani, T. Hoffmann, R. Kotlyar, P. Matagne, K. Mistry, R. Nagisetty, B. Obradovic, R. Shaheed, L. Shifren, M. Stettler, S. Tyagi, X. Wang, C. Weher, K. Zawadzki, "Understanding stress enhanced performance in Intel 90nm CMOS technology", Tech. Dig. Symp. VLSI, ses. 12.2, 2004.

[Goo85] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy and O. L. Krivanek, "Surface roughness at the Si(100)-SiO_2 interface", Physical Review B, 32, p. 8171, 1985

[Har76] J. W. Harrison and J. R. Hauser, "Alloy scattering in ternary III-V compounds," Phys. Rev. B 13, p. 5347, 1976.

[Hoy02] J. L. Hoyt, H. M. Nayfeh, S. Eguchi, I. Aberg, G. Xia, T. Drake, E. A. Fitzgerald, D. A. Antoniadis, "Strained silicon MOSFET technology", IEDM 2002.

[Hwa03] J. R. Hwang, J. H. Ho, S. M. Ting, T. P. Chen, Y. S. Hsieh, C. C. Huang, Y. Y. Chiang, H. K. Lee, A. Liu, T. M. Shen, G. Braithwaite, M. Curie, N. Gerrish, R. Hammond, A. Lochtefeld, F. Singaporewala, M. Bulsara, Q. Xiang, M. R. Lin, W. T. Shiau, Y. T. Loh, J. K. Chen, S. C. Chien, and F. Wen, "Performance of 70nm strained-silicon CMOS devices", VLSI Technology 2003, pp. 103–104.

[Itr04] Online accessible at http://public.itrs.net/.

[Jac77] C. Jacoboni, C. Canali, G. Ottaviani, and A. Alberigi-Quaranta, "Review of some charge transport properties of silicon", Solid State Electronics, Vol. 20, p. 77, 1977.

[Jac83] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Rev. of Mod. Phys*. Vol. 55, p. 645, 1983.

[Jan98] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, "Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters", Physical Review B, Vol. 57, No. 11, pp. 6493-6507, 1998.

[Jun04] C. Jungemann, N. Subba, J.-S. Goo, C. Riccobene, Q. Xiang, and B. Meinerzhagen, "Investigation of strained Si/SiGe devices by MC simulation", Solid-State Electronics, Vol. 48, p. 1417, 2004.

[Kri06] S. Krishnan, M. Fischetti, and D. Vasileska, "Self-consistent full band two-dimensional Monte Carlo two-dimensional Poisson device solver for modeling SiGe p-channel devices", Journal of Vacuum Science and Technology B, 24 (4), pp. 1997- 2003, 2006.

[Lee02] B. H. Lee, A. Mocuta, S. Bedell, H. Chen, D. Sadana, K. Rim, P. O'Neil, R. Mo, K. Chan, C. Cabral, C. Lavoie, D. Mocuta, A. Chakravarti, R. M. Mitchell, J. Mezzapelle, F. Jamin, M. Sendelbach, H. Kermel, M. Gribelyuk, A. Domenicucci, K. A. Jenkins, S. Narasimha, S. H. Ku, M. Ieong, I. Y. Yang, E. Leobandung, P. Agnello, W. Haensch, and J. Welser, "Performance enhancement on sub-70 nm strained silicon SOI MOSFETs on ultra-thin thermally

mixed strained silicon/SiGe on insulator (TM-SGOI) substrate with raised S/D", IEDM 2002, pp. 946–948.

[Lei02] C. W. Leitz, M. T. Currie, M. L. Lee, Z.-Y. Cheng, D. A. Antoniadis, and E. A. Fitzgerald, "Hole mobility enhancements and alloy scattering-limited mobility in tensile strained Si/SiGe surface channel metal-oxide-semiconductor field-effect transistors", Journal of Applied Physics, Vol. 92, No. 7, pp. 3745-51, October 2002.

[Lun97] M. Lundstrom, "Elementary scattering theory of the Si MOSFET", IEEE Electron Device Letters 18 (7), 361 (1997)

[Maj87] J. A. Majewski and P. Vogl, "Simple model for structural properties and crystal stability of *sp*-bonded solids," *Phys. Rev.* **B** 35, p. 9666, 1987.

[Mii91] Y. J. Mii, Y. H. Xie, E. A. Fitzgerald, D. Monroe, F. A. Thiel, B. E. Weir, and L. C. Feldman, "Extremely high electron-mobility in Si/Ge_xSi_{1-x} structures grown by molecular-beam epitaxy", Applied Physics Letters, 59(13): 1611 – 1613, 23 Sep 1991.

[Mit00] MIT well-tempered device structure, http://www-mtl.mit.edu/

[Mom02] H. S. Momose, T. Ohguro, S. Nakamura, Y. Toyoshima, H. Ishiuchi, and H. Iwai, "Ultrathin gate oxide CMOS on (111) surface-oriented Si substrate", IEEE Trans. on Electron Devices, 49 (9), p.1597, 2002.

[Mom03] H. S. Momose, T. Ohguro, K. Kojima, S. Nakamura, and Y. Toyoshima, "1.5-nm gate oxide CMOS on (110) surface-oriented Si substrate", IEEE Trans. on Electron Devices, 50 (4), p.1001, 2003.

[Nel93] S. F. Nelson, K. Ismail, J. O. Chu, and B. S. Meyerson, "Room-temperature electron-mobility in strained Si/SiGe heterostructures", Applied Physics Letters, 63(3): 367 – 369, 19 Jul 1993.

[Obe98] R. Oberhuber, G. Zandler, and P. Vogl, "Subband structure and mobility of two-dimensional holes in strained Si/SiGe MOSFET's", Physical Review B, 58 (15): 9941 – 9948, 1998.

[Oni03] K. Onishi, C. S. Kang, R. Choi, H. Cho, Y. H. Kim, S. Krishnan, M. S. Akbar, and J. C. Lee, "Performance of polysilicon gate HfO_2 MOSFETs on (100) and (111) silicon substrates", IEEE Electron Device Letters, 24 (4), p.254, 2003.

[Ott75] G. Ottaviani, L. Reggiani, C. Canali, F. Nava, and A. Alberigi-Quaranta, "Hole drift velocity in silicon", Physical Review B, 12 (8), p.3318, 1975.

[Ran05] H. E. Randell, "Applications of stress from boron doping and other challenges in silicon technology", PhD dissertation, University of Florida, 2005.

[Rav00] U. Ravaioli, B. Winstead, C. Wordelman, and A. Kepkep, "Monte Carlo simulation for ultra-small MOS devices", Superlattices and Microstructures, 27 (2/3), p.137, 2000.

[Ren82] S. Y. Ren, J. D. Dow, and D. J. Wolford, "Pressure dependence of deep levels in GaAs," *Phys. Rev*. **B** 25, p. 7661, 1982.

[Ric87] S. L. Richardson, M. L. Cohen, S. G. Louie, and J. R. Chelikowsky, "Conduction-band-edge charge densities in elemental and compound semiconductors," *Phys. Rev*. B 35, p. 1388, 1987.

[Rid77] B. K. Ridley, "Reconciliation of the Conwell-Weisskopf and Brooks-Herring formulae for charged-impurity scattering in semiconductors," *J. Phys. C.* Vol. 10, p. 1589, 1977.

[Rim98] K. Rim, J. L. Hoyt, and J. F. Gibbons, "Transconductance enhancement in deep submicron strained-Si n-MOSFETs", IEDM 1998.

[Rim00] K. K. Rim, J. L. Hoyt, and J. F. Gibbons, "Fabrication and Analysis of Deep Submicron Strained-Si N-MOSFET's", IEEE Trans. on Electron Devices, 47 (7): 1406 – 1415, July 2000.

[Rim02] K. Rim, S.Narasimha, M. Longstreet, A. Mocuta, and J. Cai, "Low field mobility characteristics of sub-100 nm unstrained and strained Si MOSFETs", IEDM 2002.

[Rim02a] K. Rim, J. Chu, H. Chen, K. A. Jenkins, T. Kanarsky, K. Lee, A. Mocuta, H. Zhu, R. Roy, J. Newbury, J. Ott, K. Petrarca, P. Mooney, D. Lacey, S. Koester, K. Chan, D. Boyd, M. Ieong, and H.-S.Wong, "Characteristics and device design of sub-100 nm strained Si N- and pMOSFET's," in Proc. Symp. VLSI Technology, Honolulu, HI, June 2002, pp. 98–99.

[Rol96] J. B. Roldan, F. Gamiz, J. A. LopezVillanueva and J. E. Carceller, "A Monte Carlo study on the electron-transport properties of high-performance strained-Si on relaxed $Si_{1-x}Ge_x$ channel MOSFETs", Journal of Applied Physics, 80 (9): 5121 – 5128, 1 November 1996.

[Sat69] T. Satô, Y. Takeishi, and H. Hara, Jpn. J. Appl. Phys., 8, 588 (1969).

[Sat71] T. Satô, Y. Takeishi, H. Hara, and Y. Okamoto, Phys. Rev. B, 4, 1950 (1971).

[Shi04] L. Shifren, X. Wang, P. Matagne, B. Obradovic, C. Auth, S. Cea, T. Ghani, J. He, T. Hoffman, R. Kotlyar, Z. Ma, K. Mistry, R. Nagisetty, R. Shaheed, M. Stettler, C. Weber, and M. D. Giles, “Drive current enhancement in p-type metal-oxide-semiconductor field-effect transistors under shear uniaxial stress”, Applied Physics Letters, 85 (2), p. 6188, 2004.

[Sho99] M. Shoji et al, JAP, 85 (5), p. 2722, 1999.

[Smi54] C. S. Smith, “Piezoresistance effect in germanium and silicon”, Physical Review, Vol. 94, No. 1, pp. 42-49, April, 1954.

[Tag94] S. Tagaki, 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. on Electron Devices, 41(12): 2357 – 2362, December 1994.

[Tag98] Y. Tagawa, and Y. Awano, “Enhanced hole drift velocity in sub-0.1 μm Si devices caused by anisotropic velocity overshoot”, IEDM 1998, p.206.

[Tho02] S. Thompson, N. Anand, M. Armstrong, C. Auth, B. Arcot, M. Alavi, P. Bai, J. Bielefeld, R. Bigwood, J. Brandenburg, M. Buehler, S. Cea, V. Chikarmane, C. Choi, r. Frankovic, T. Ghani, G. Glass, W. Han, T. Hoffmann, M Hussein, P. Jacob, A. Jan, S. Joshi, C. Kenyon, J. Klaus, S. Klopcic, J. Luce, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, P. Nguyen, H. Pearson, T. Sandford, R. Schweinfurth, R. Shaheed, S. Sivakumar, M. Taylor, B. Tufts, C. Wallace, P. Wang, C. Weber, and M. Bohr, IEEE Technical Digest 3.2, 2002.

[Tho04] 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”, IEDM Technical Digest, pp. 221 – 224, 2004.

[Tho04a] S. E. Thompson, M. Armstrong, C. Auth, S. Cea, R. Chau, G. Glass, T. Hoffman, J. Klaus, Z. Ma, B. Mcintyre, A. Murthy, B. Obradovic, L. Shifren, S. Sivakumar, S. Tyagi, T. Ghani, K. Mistry, M. Bohr, and Y. El-Mansy, “A logic nanotechnology featuring strained-silicon”, IEEE Elec. Device Lett., pp. 191 – 193, 2004.

[Tho05] S. Thompson, R. S. Chau, T. Ghani, K. Mistry, S. Tyagi, and M. T. Bohr, “In search of ‘forever’, continued transistor scaling one new material at a time”, IEEE Transactions on Semiconductor

Manufracturing, Vol. 18, No. 1, pp. 26-36, February 2005.

[Vog92] T. Vogelsang and W. Hänsch, "A novel approach for including band-structure effects in a Monte Carlo simulation of electron transport in silicon," *J. Appl. Phys*. Vol. 70, p. 1492, 1992.

[Xia03] Q. Xiang, J. Goo, J. Pan, B. Yu, S. Ahmed, J. Zhang, and M.-R. Lin, "Strained silicon NMOS with nickel-silicide metal gate", VLSI Technology, 2003, pp. 101–102.

[Uch05] 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", IEDM Technical Digest, 6.1, 2005.

[Wan93] X. Wang, V. Chandramouli, C. M. Maziar, and Al F. Tasch Jr., "Simulation program suitable for hot carrier studies: an efficient multiband Monte Carlo model using both full and analytic band structure description for silicon," *J. Appl. Phys*. Vol. 73, p. 3339, 1993.

[Wan01] Xin Wang, Ph.D. dissertation.

[Wan04] E. Wang, P. Matagne, L. Shifren, B. Obradovic, R. Kotlyar, S. Cea, J. He, Z. Ma, R. Nagisetty, S. Tyagi, M. Stettler, and M. D. Giles, "Quantum mechanical calculation of hole mobility in silicon inversion layers under arbitrary stress", IEDM Tech. Dig., pp. 147 - 150, December 2004.

[Wel94] J. Welser, J. L. Hoyt, S. Takagi, and J. F. Gibbons, "Strain dependence of the performance enhancement in strained-Si n-MOSFETs", in Technical Digest., International, page(s): 373 – 376, 11 – 14 December 1994.

[Win01] B. Winstead, and U. Ravaioli, "A coupled Schrödinger/Monte Carlo technique for quantum-corrected device simulation", Device Research Conference, 2001, Page(s): 169 – 170.

[Win03] B. Winstead, and U. Ravaioli, "A Quantum Correction Based on Schrödinger Equation Applied to Monte Carlo Device Simulation", IEEE Trans. on Electron Devices, Volume 50, Issue 2, Page(s): 440 – 446, February, 2003

Vita

Xiaofeng Fan was born in Shaoyang, China on November 23rd, 1978. He entered the "Special class for gifted young" in the University of Science and Technology of China at the age of 16 to begin the college education, focused on mathematics and physics in curriculum. He obtained the B.S. degree in July 1998 and continued onto the graduate program in the same university. His coursework and research was focused on quantum mechanics and quantum optics, specialized in quantum communication and quantum computation. Supervised by Prof. Guo, Guangcan, Xiaofeng Fan obtained his Master of Science degree in July 2001. In the same year, he joined the University of Texas at Austin to enroll in the Ph.D. program in the Electrical and Computer Engineering department. He started working with Dr. Sanjay K. Banerjee and Dr. Leonard F. Register toward the doctoral degree in solid state engineering in 2002. The research area during the Ph.D. program is semiclassical Monte Carlo device simulation with non-classical Si CMOS.

Mail address: 3719 S. Gekeler Lane, #42
Boies, ID 83706

The dissertation was typed by the author.