Technology for SiGe Heterostructure-Based CMOS Devices

by

Mark Albert Armstrong

Bachelor of Science in Electrical Science and Engineering
Massachusetts Institute of Technology, June 1992

Master of Science in Electrical Engineering and Computer Science
Massachusetts Institute of Technology, September 1993

Submitted to the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Electrical Engineering and Computer Science
at the
Massachusetts Institute of Technology

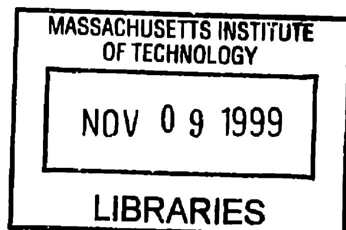
September 1999

© Massachusetts Institute of Technology, 1999. All Rights Reserved.

Author ....
Department of Electrical Engineering and Computer Science
June 30, 1999

Certified by ....
Dimitri A. Antoniadis
Professor of Electrical Engineering
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Professor of Electrical Engineering
Graduate Officer


MASSACHUSETTS INSTITUTE OF TECHNOLOGY
NOV 0 9 1999
LIBRARIES
ARCHIVES

Technology for SiGe Heterostructure-Based CMOS Devices

by

Mark Albert Armstrong

Submitted to the Department of Electrical Engineering and Computer Science on June 30, 1999, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Science and Engineering

Abstract

Bulk silicon is currently the substrate material of choice for the manufacture of high-peformance digital circuits due to its highly-developed processing technology and the relatively low cost for high-quality substrates. Silicon-based MOSFETs have reached remarkable levels of performance through device scaling. However, with each technology generation, it is becoming harder and harder to improve device performance at the same pace through traditional scaling methods alone. Short-channel effects such as velocity saturation and drain-induced barrier lowering have placed an fundamental limit on the ultimate performance of bulk Si MOSFETS.

One way to raise this limit is to increase the carrier mobilities in the channel. This can be done using high-mobility Si and SiGe strained-layers. Unlike III-V-based high-mobility materials, Si/SiGe strained-layers have the advantage of being largely compatible with mainstream Si processing, which is important from a financial feasibility standpoint. This thesis examines several issues related to Si/SiGe strained-layer devices and their integration into mainstream CMOS.

The first part of this work strives to predict the performance leverage of high-mobility Si/SiGe over bulk Si devices and circuits in a realistic manner. Two-dimensional hydrodynamic simulations are used to predict static device characteristics including effects of series resistance, velocity saturation and velocity overshoot. The simulations show enhanced current drive over bulk Si devices at 0.2 µm effective channel length and highlight the importance of velocity overshoot in high-mobility submicron devices. The circuit performance of Si/SiGe devices is determined from transient simulations of CMOS ring oscillators including the effects of parasitic capacitance and drain-to-source voltage at the onset of saturation $V_{DS,sat}$. The simulations show a 4 to 6-fold reduction in power-delay product as compared to bulk CMOS oscillators operated at 2.5 V with the same design rules.

The remainder of the thesis focuses on the fabrication and characterization of strained-Si NMOS devices. The vehicle for this work is a novel short-flow, single-mask MOSFET which can be fabbed in as little as a week. This device is superior to simple Hall mobility structures which suffer from leakage through the substrate, an inability to control the carrier concentration and the uncertainty associated with the Hall scattering factor.

I investigate a novel buried-channel strained-Si NMOS structure incorporating an n-type donor layer beneath the strained-Si channel to encourage occupation of the buried channel and increase the overall mobility. Peak mobility in a structure without a donor layer reproduces the best results in the literature for buried-channel strained-Si NMOS devices. For structures with donor layers, Coulomb scattering from charges in the donor layer eradicates any benefit from increased buried-channel occupation.

I also investigate the effect of well implants on the mobility of surface-channel strained-Si NMOS devices. Similar to the universal mobility curve in bulk Si, mobility at low perpendicular electric field degrades with increasing implant dose while high field mobility is unaffected. The mobility is largely unaffected by a neutral implant species at the same dose. This leads to the conclusion that the material quality of the strained-layer is not affected by the implant, and that the mobility degradation is due solely to increased ionized impurity scattering.

Thesis Supervisor: Dimitri A. Antoniadis
Title: Professor of Electrical Engineering

Acknowledgments

I would like to thank my advisor, Professor Dimitri Antoniadis, for supporting this research through all its twists and turns. Professor Gene Fitzgerald and Dr. Khalid Ismail of IBM also provided guidance during this work. Professor Gene Fitzgerald and Professor Jesus del Alamo acted as readers for this thesis.

I would like to thank Dr. Jack Chu of IBM for growing the wafers used in this work and for providing encouraging words when I needed them. In addition I acknowledge IBM for supporting part of this research.

I have benefited from interaction with so many students in and around MTL, as is the multifaceted and collaborative nature of this field. The contributions of past and present members of my research group have been invaluable, including Dr. Jarvis Jacobs, Dr. Isabel Yang, Dr. Melanie Sherony, Dr. Zachary Lee, Andy Wei, Keith Jackson, Tony Lochtefeld, Andy Ritenour, Hasan Nayfeh, Jim Fiorenza and Dr. Jong-Ho Lee. Several students from Professor Gene Fitzgerald's group have also helped, including Tom Langdo and Matt Currie.

This work would not have been possible without the skilled staff of MTL, past and present, including Dr. Vicky Diadiuk, Pat Burkhart, Paul Tierney, Bernard Alamariu, Barry Farnsworth, Paul McGrath, Joe Walsh, Kurt Broderick, Octavio Hurtado, Ron Stoute, Wayne Price, Jim Bishop, Dan Adams, Joe DiMaria, Myron Freeman, Tom Wingard and Tom Lohman.

Most of all, I thank my friends here for all the fun times which will make me miss Boston greatly, and my family for never-ending support and encouragement.

Contents

1 **Introduction** 17
1.1 Motivation .17
1.2 Goals .20
1.3 Outline .22

2 **Si/SiGe Strained-Layers** 25
2.1 Introduction .25
2.2 Lattice Structure .25
2.3 Band Structure .29
Relaxed-$Si_{1-x}Ge_x$ .29
Strained-$Si_{1-x}Ge_x$ on Relaxed-Si .30
Strained-$Si_{1-x}Ge_x$ on Relaxed-$Si_{0.7}Ge_{0.3}$ .32
Strained-Si on Relaxed-$Si_{1-x}Ge_x$ .35
2.4 Transport .37
Relaxed-$Si_{1-x}Ge_x$ .37
Strained-$Si_{1-x}Ge_x$ ($x < 0.6$) .39
Strained-$Si_{1-x}Ge_x$ ($x > 0.6$) .42
Strained-Si .43

3 **Si/SiGe Device Structures** 49
3.1 Introduction .49
3.2 PMOS Structures .49
3.3 NMOS Structures .53
3.4 CMOS Structures .55
(a) Buried-Channel CMOS .57
(b) Surface-Channel CMOS .59
(c) Surface/Buried-Channel CMOS .59

4 **Device and Circuit Performance** 63
4.1 Introduction .63
4.2 Device Performance Analysis .63
Velocity Saturation .63
Long-Channel Characteristics .65
Short-Channel Characteristics .66
4.3 Device Simulation .66

Simulation Inputs 67
Simulation Results 70
4.4 Circuit Performance Analysis 75
CMOS Inverter 76
Interconnect Capacitance Limit 78
Gate Capacitance Limit 79
Mobility Dependence 79
Saturation Voltage 81
4.5 Circuit Simulation 83

5 Short-Flow MOSFET **87**
5.1 Introduction 87
Drift Mobility vs. Hall Mobility 88
5.2 Process Flow 89
5.3 Layout and Operation 95
5.4 Measurement 99

6 Buried-Channel Strained-Si NMOS **103**
6.1 Introduction 103
6.2 Comparison to Previous Work 103
6.3 Effect of Donor Layer 111
6.4 Effect of In-Situ P-Well 117

7 Surface-Channel Strained-Si NMOS **123**
7.1 Introduction 123
7.2 Device Fabrication 124
7.3 Mobility Extraction 125
7.4 Results 125

8 Conclusion **129**
8.1 Summary 129
8.2 Future Work 131

References **133**

A Low Temperature Process Technology **141**
A.1 Introduction 141
A.2 Source/Drain Junctions 142
Reverse Leakage 143
Resistivity 147
Junction Depth 148
A.3 Gate Oxide 150
Thin Oxide Deposition 150
Electrical Characterization 151

List of Figures

Fig. 1.1. Effective velocity veff of electrons in the channel vs. SIA Technology Roadmap generation for NMOS. Diamond points are calculated using (1.1) with I_{on} = 600 μA/μm and V_T = 0.3 V and V_{DD} and t_{ox} as listed on the lower two axes. Squares are the same but with a minimum allowed t_{ox} of 2 nm. 19

Fig. 1.2. Effective velocity in the channel vs. DIBL for bulk-Si NMOS and Si/SiGe NMOS (labelled HNMOS for heterostructure-based NMOS). The HNMOS curve is simulated in Chapter 4. The bulk curve is measured but is the same as simulation. 20

Fig. 1.3. Key issues for integration of SiGe-based strained-layer technology into mainstream CMOS applications. 21

Fig. 2.1. Simplified lattice-structure for growth of (a) SiGe on relaxed-Si and (b) Si on relaxed-SiGe. The epitaxial film can be pseudomorphic or dislocated, corresponding to a strained or relaxed condition [25]. 27

Fig. 2.2. Critical thickness as a function of Ge content x for growth of strained-$Si_{1-x}Ge_x$ on a relaxed-Si substrate. Solid line shows the result of methods sensitive to the structural properties of the film [55]. Dashed line shows average results of methods sensitive to the electrical properties [25]. 28

Fig. 2.3. Bandgap vs. Ge content for relaxed-$Si_{1-x}Ge_x$ at 296K [50]. 30

Fig. 2.4. Valence band offset ΔE_V vs. Ge content x for compressively-strained-$Si_{1-x}Ge_x$ on relaxed-Si at room temperature. The valence band is higher in strained-SiGe. 31

Fig. 2.5. Room temperature band alignments of three technologically significant layer combinations using consensus of data from the literature: (a) strained-$Si_{0.7}Ge_{0.3}$ on relaxed-Si, (b) strained-$Si_{0.2}Ge_{0.8}$ on relaxed-$Si_{0.7}Ge_{0.3}$ and (c) strained-Si on relaxed-$Si_{0.7}Ge_{0.3}$. 33

Fig. 2.6. (a) Conduction and (b) valence band offsets for strained-$Si_{1-x}Ge_x$ on relaxed-$Si_{0.7}Ge_{0.3}$ vs. strained-layer Ge content x [52][13][53]. Conduction and valence bands are higher in the strained-$Si_{1-x}Ge_x$ for positive values of ΔE_C and ΔE_V. 34

Fig. 2.7. (a) Conduction and (b) valence band offsets for strained-Si on relaxed-$Si_{1-x}Ge_x$ vs. substrate Ge fraction x comparing several theoretical predictions [52][13][53] to the MOS C-V data of Welser [45]. Both the conduction and valence bands are lower in strained-Si. 36

Fig. 2.8. Low-field mobility for (a) electrons and (b) holes in undoped, relaxed-$Si_{1-x}Ge_x$ as a function of Ge content x. Data is a calculation [58] based on the measured data of

[60]......38

Fig. 2.9. Room-temperature in-plane hole mobility in undoped strained-$Si_{1-x}Ge_x$ on relaxed-Si vs. Ge content. Several theoretical predictions (lines) [58][61][62][66][74] are compared to various experiments (x marks) [4][5][73][72][71][70][67][68][69]......40

Fig. 2.10. (a) Simplified valence band E vs. k diagram around the Γ-point of Si showing the splitting of the light-hole (LH) and heavy-hole (HH) bands under strain [74]. (b) Conduction band surfaces of constant energy in k-space of Si showing the splitting of the 6-fold degenerate Δ-minimum into Δ_2 and Δ_4-minima under strain [75]. Filled ovals represent those lower in energy and thus relevant to electron transport. This diagram is schematic only and does not reflect changes in effective mass. These diagrams are relevant for both Si and $Si_{1-x}Ge_x$ for $x < 0.8$. For $x > 0.8$, the conduction band in relaxed material changes from Si-like with a 6-fold degenerate Δ-minimum to Ge-like with an 8-fold degenerate L-minimum [50]......41

Fig. 2.11. Electron mobility in buried-channel strained-Si structures on relaxed-$Si_{1-x}Ge_x$ vs. substrate Ge content comparing several theoretical calculations [80][15][33][58][81] to various experimental results [8][83][1][22][84] including this work. Lines and +, x and * marks are theory while filled symbols represent measured MOS-gated results and open symbols represent measured Schottky-gated results......44

Fig. 2.12. Peak electron and hole mobility in surface-channel strained-Si NMOS and PMOS devices on relaxed-$Si_{1-x}Ge_x$ vs. substrate Ge content comparing various experimental results [82][8][85][27] and this work......46

Fig. 3.1. Layer structure of SiGe-based PMOS devices. – marks indicate acceptors. The gate oxide would be placed immediately on top of these structures. PMOS devices have been fabricated by various groups based on structures (a) [4][5][6][68], (b) [9] and (d) [27][85]. Structure (c) has not been demonstrated in this form, though a similar structure with a relaxed-$Si_{0.7}Ge_{0.3}$ cap in place of the strained-Si cap has been fabricated with a Schottky gate [2]......50

Fig. 3.2. Band structures of the four PMOS layer structures of Fig. 3.1: (a) solid line, (b) small dash, (c) medium dash and (d) long dash. Fermi level is at zero potential. Diagram is for following conditions: t_{ox} = 25 Å, $N_D = 5 \times 10^{17}$ cm^{-3}, V_{GS} = 1.5 V and p^+ poly gate workfunction. For (a), (b) and (c): t_{cap} = 30 Å and $t_{channel}$ = 50 Å. Additionally, for (b) and (c): t_{spacer} = 40 Å and $N_{A,acceptor\text{-}layer} = 2 \times 10^{12}$ cm^{-2}. For (d) there is no cap and $t_{channel}$ = 80 Å. Note how most carriers end up in the cap no matter what layer structure is used......52

Fig. 3.3. Layer structure of SiGe-based NMOS devices. + marks indicate donors. The gate oxide would be placed immediately on top of these structures. NMOS devices have been fabricated by various groups based on structure (a) [8], (b) [22][29] and (c) [46]. Structure (b) is also favored for Schottky-gate devices [84][83]. MOS devices based on structures (b) and (c) are investigated in this work......54

Fig. 3.4. Band structures of the three NMOS layer structures of Fig. 3.3: (a) solid line, (b) small dash and (c) long dash. Fermi level is at zero potential. Diagram is for following conditions: t_{ox} = 25 Å, $N_A = 5 \times 10^{17}$ cm^{-3}, V_{GS} = 1.5 V and n^+ poly gate workfunction. For (a) and (b) the thin Si cap is ignored and t_{cap} = 30 Å and $t_{channel}$ = 50 Å. Additionally, for (b): t_{spacer} = 40 Å and $N_{D,donor\text{-}layer} = 2 \times 10^{12}$ cm^{-2}. For (c) there is no cap and $t_{channel}$

= 80 Å. Note how most carriers end up in the cap no matter what layer structure is used. 55

Fig. 3.5. Layer structure of SiGe-based CMOS devices. + marks indicate donors. The gate oxide would be placed immediately on top of these structures. 56

Fig. 3.6. Electron, hole and doping concentration and band diagram in inversion for (a) NMOS and (b) PMOS device based Fig. 3.5(a). Dashed line is net doping or the fermi level. Layer thicknesses are in Å: $t_{ox} = 50$, $t_{cap} = 10$, $t_{p\text{-channel}} = 40$, $t_{n\text{-channel}} = 40$. The thin Si cap is modeled as an extension of the p-channel due to quantum-effect raising of the minimum energy level. The donor layer dose is 2.6×10^{12} cm^{-2} located 20 Å below the n-channel. Gate is p$^+$ poly and well doping is 3×10^{17} cm^{-3} starting 300 Å below the n-channel for both NMOS and PMOS. 58

Fig. 3.7. Electron and hole concentration vs. gate bias in HCMOS structure showing symmetric threshold voltages for NMOS and PMOS. Parameters are as in Fig. 3.6. Dashed line indicated electrons confined in buried strained-Si layer. 59

Fig. 3.8. Carrier profiles and band structure in inversion of (a) NMOS and (b) PMOS devices based on Fig. 3.5(c). Dashed line is net doping or the fermi level. The strained-Si and strained-$Si_{0.2}Ge_{0.8}$ layers are both 50 Å thick. Oxide thickness is 25 Å. Gate is p$^+$ poly for PMOS and n$^+$ poly for NMOS. Well doping is 5×10^{18} cm^{-3} starting 100 Å below the surface. 61

Fig. 3.9. Electron and hole concentration vs. gate bias in NMOS and PMOS devices based on Fig. 3.5(c) comparing (a) p$^+$ poly on PMOS and n$^+$ poly on NMOS and (b) midgap gate workfunction. Well doping for (a) is 5×10^{18} cm^{-3} starting 100 Å below the surface and for (b) is 5×10^{17} cm^{-3} starting 300 Å below the surface. Other parameters are as in Fig. 3.8. Dashed line indicates holes confined in buried strained-$Si_{0.2}Ge_{0.8}$ layer. 62

Fig. 4.1. Velocity-field characteristics in silicon comparing two different mobilities plotting using (4.2) with $\beta = 1$. 64

Fig. 4.2. Long-channel device characteristics comparing two mobilities. (a) I_D-V_{DS} characteristics for $V_{GS} = 0$ to 5 step 1 V. (b) I_D-V_{GS} characteristics for $V_{DS} = 0$ to 5 step 1 V. Device parameters: L = 5 μm, $t_{ox} = 1000$ Å, $V_T = 1$ V. 65

Fig. 4.3. Short-channel device characteristics incorporating velocity saturation comparing two mobilities. (a) I_D-V_{DS} characteristics for $V_{GS} = 0$ to 1.2 step 0.3 V. (b) I_D-V_{GS} characteristics for $V_{DS} = 0$ to 1.2 step 0.3 V. Device parameters: L = 0.1 μm, $t_{ox} = 20$ Å, $V_T = 0.3$ V, $v_{sat} = 6 \times 10^6$ cm/s..67

Fig. 4.4. Velocity-field characteristics in Si and Si/SiGe material. 69

Fig. 4.5. I_D-V_{DS} characteristics for different V_{GS}. (a) HNMOS and (b) HPMOS compared to bulk counterparts. Also shown is the comparison between DD and hydro simulations. DD and hydro give the same results for bulk Si devices. 72

Fig. 4.6. Simulated I_D-V_{GS} characteristics for HNMOS and HPMOS at several different V_{DS} values. 74

Fig. 4.7. Effective carrier velocity vs. DIBL comparing HNMOS, HPMOS and bulk NMOS. The data points are labelled with the effective channel lengths. The HPMOS points have effective lengths corresponding to the HNMOS points.............................76

Fig. 4.8. CMOS inverter with power supply V_{DD} and load capacitance C_L. The NMOS pull-down transistor is on bottom and the PMOS pull-up, indicated by the bubble on the gate, is on top. 77

Fig. 4.9. Piecewise linear model for short-channel MOSFET comparing two different mobilities. 82

Fig. 4.10. Simulated power-delay-product vs. stage delay for Si/SiGe HCMOS and bulk Si CMOS with V_{DD} as an implicit variable, comparing unloaded and loaded ring oscillators. The corresponding V_{DD} values are indicated on the curves. 84

Fig. 4.11. I_D-V_{DS} locus traced by NMOS and PMOS transistors in a ring oscillator (W_n = 1 μm, W_p = 2.5 μm, C_L = 10 fF, V_{DD} = 1 V) superimposed on the I_D-V_{DS} characteristics of those transistors at V_{GS} = 1 V comparing bulk Si CMOS with Si/SiGe HCMOS. NMOS is plotted with positive current while PMOS is plotted with negative current...85

Fig. 5.1. (1) Schematic and (b) SEM cross section of short-flow MOSFET. Rough material is Ti/Al. The SEM shows a 0.1 μm line, which is much smaller than the 100 μm channel lengths of actual devices. 90

Fig. 5.2. Process flow for short-flow MOSFET 91

Fig. 5.3. Short-flow MOSFET after gate stack patterning and 20 seconds BOE etch. The structure shown is amorphous Si undercut by LTO 93

Fig. 5.4. Example implant profile (Phosphorous 25keV $1\times10^{15}cm^{-2}$). 94

Fig. 5.5. Ti/Al contact diffusion barrier lifetime vs. temperature. 96

Fig. 5.6. Ring transistor mask layout. Large arrow indicates view shown in Fig. 5.7....96

Fig. 5.7. SEM of inside corner of ring transistor gate. Gate region is the raised area to the upper right and the drain is to the lower left. 97

Fig. 5.8. Isopotential contours for square ring of uniform sheet resistance bounded by ideal contacts on inside and outside edges 98

Fig. 5.9. Inside corner of gate of NMOS device after arcing (a) without and (b) with 10 MΩ series resistor. Gate is to upper right and is darker due to charging. 100

Fig. 6.1. Buried-channel NMOS structure (a) without and (b) with donor layer. Strained-Si cap is not indicated in the conduction band because ideally it is of negligible thickness 104

Fig. 6.2. Layer structure of Welser [8] compared to HNMOS-15. The gate oxide is immediately above this structure. 105

Fig. 6.3. Effective mobility of HNMOS-15 compared to Welser. Also shown are bulk monitors from both works 106

Fig. 6.4. Welser's mobility curve compared to a simple two-channel model. Also shown is the ratio of buried to total carriers. 109

Fig. 6.5. HNMOS-15 mobility curve compared to the two-channel model. Also shown is the ratio of buried to total carriers 110

Fig. 6.6. Layer structure of runs HNMOS-10, 13, and 15. 112

Fig. 6.7. Effective mobility vs. effective field for HNMOS-10,13 and 15. The three wafers have different phosphorous doses in the donor layer. Arrows point to the co-pro-

cessed bulk mobility data. Inset shows peak effective mobility vs. donor layer dose..114

Fig. 6.8. Schematic showing comparison between (a) oxide fixed charge scattering in bulk Si inversion layers and (b) donor layer charge scattering in strained-Si quantum wells. 115

Fig. 6.9. Effective mobility of co-processed bulk monitors compared to Sun and Plummer thermal oxide data. The labels point to the Sun and Plummer data with $N_A = 9x10^{14}$ cm^{-2}. The bulk data corresponds to Table 6.3. Inset shows peak effective mobility vs. fixed charge taken from the Sun and Plummer paper. 116

Fig. 6.10. Normalized reduction in peak effective mobility of HNMOS-15 buried strained-Si channel with donor layer charge compared to normalized reduction in peak effective mobility of inversion layer in silicon with thermal oxide including oxide fixed charge. 117

Fig. 6.11. Layer structure of runs HNMOS-8, 9 and 10. 118

Fig. 6.12. Doping profiles of runs HNMOS-8, 9 and 10. The large hump is boron doping from SIMS. The small spike is phosphorous doping which is a guess based on the intended profile with total dose determined form C-V data. 119

Fig. 6.13. (a) Effective mobility vs. effective field for HNMOS-8, 9 and 10. Arrows point to co-processed bulk wafers. (b) Simulated ratio of buried to total carriers vs. effective field. In this simulation a 10 Å Si cap was included with $\Delta E_C = 0.2$ eV. The difference in observed mobility between the three samples cannot be explained by differences in carrier location. 121

Fig. 7.1. Layer structure and band diagram of surface-channel strained-Si NMOS.....123

Fig. 7.2. Well implant profiles as calculated by SUPREM. Solid lines are boron and dashed lines are silicon (approximated in SUPREM by a phosphorous implant at same conditions). Tilt is 7 degrees. 124

Fig. 7.3. μ_{eff} vs. E_{eff} for (a) bulk Si and (b) strained-Si NMOS comparing B and Si well implants at several doses 126

Fig. A.1. Problem areas for process integration of SiGe materials 142

Fig. A.2. (a) Characteristics of As and BF_2-implanted (25 keV, 1 x 10^{15} cm^{-2}) junctions in bulk Si after 600 °C anneal. (b) Leakage characteristics of BF_2-implanted (25 keV, 1 x 10^{15} cm^{-2}) junctions in relaxed-$Si_{0.7}Ge_{0.3}$ after 600 °C anneal. Diodes with intersecting growth defects have more leakage. 144

Fig. A.3. (a) Histogram of leakage current at 5 V reverse bias of BF_2-implanted (7 keV, 9 x $10^{14}cm^{-2}$) diodes comparing non-preamorphized and Ge-preamorphized (20 keV, 1 x 10^{15} cm^{-2}) cases. (b) Dopant concentration vs. distance from the surface as calculated by TSUPREM4 comparing Ge, BF_2 and BF_2 without channeling. Also shown is the approximate position of end-of-range (EOR) damage. 146

Fig. A.4. (a) Sheet resistance and (b) ideality factor vs. annealing time at 600°C of unsintered diodes. 148

Fig. A.5. SIMS profiles of (a) As and (b) BF_2 implants before and after annealing at 600 °C for 60 minutes. No implant diffusion is detectable. 149

Fig. A.6. Oxide thickness vs. deposition time for slow LTO deposition recipe. Flow ratio

was 1:3 silane:oxygen Wafers were uncleaned new wafers except for the first point, which received a piranha clean and HF dip. 151

Fig. A.7. C-V plots for Al-gate capacitors: (a) thermal oxide before sinter, (b) LTO before sinter, (c) thermal oxide after sinter and (d) LTO after sinter. Solid line is high-frequency and dashed is quasistatic. There is some surface leakage visible in the quasi-static C-V plot for the LTO sample. 152

Fig. A.8. Interface state density extraction using the conductance method for (a) 2869 Å LTO on bulk Si, (b) 2790 Å LTO on 10 Å strained-Si on relaxed-$Si_{0.7}Ge_{0.3}$ and (c) 3018 Å LTO on relaxed-$Si_{0.7}Ge_{0.3}$. Width of the conductance peak was not considered. ...154

List of Tables

TABLE 2.1
TECHNOLOGICALLY SIGNIFICANT SI/SIGE STRAINED-LAYERS 29
TABLE 2.2
PROPERTIES OF RELAXED MATERIALS 39
TABLE 2.3
PROPERTIES OF STRAINED MATERIALS....................................47
TABLE 4.1
CALCULATED AND EXPERIMENTAL DC PERFORMANCE OF SI AND SI/SIGE MOSFETS75
TABLE 4.2
RESULTS OF ANALYSIS OF CMOS INVERTER................................80
TABLE 5.1
SHORT-FLOW MOSFET PROCESS...92
TABLE 5.2
GEOMETRY FACTORS FOR VARIOUS DEVICE SIZES 99
TABLE 6.1
MATERIAL PARAMETERS FOR MEDICI SIMULATIONS.........................108
TABLE 6.2
MOSFET ELECTRICAL PARAMETERS FOR HNMOS-10, 13 AND 15...............113
TABLE 6.3
PARAMETERS OF CO-PROCESSED BULK WAFERS.............................116
TABLE 6.4
MOSFET ELECTRICAL PARAMETERS FOR HNMOS-8, 9 AND 10.................119

Chapter 1

Introduction

1.1 Motivation

Bulk silicon is currently the substrate material of choice for the manufacture of high-performance digital circuits due to its highly-developed processing technology and the relatively low cost for high-quality substrates. Silicon-based MOSFETs have reached remarkable levels of performance through device scaling. However, with each technology generation, it is becoming increasingly harder to improve device performance through traditional scaling methods alone. New technologies will be necessary to maintain the expected level of generational performance improvement. New low-power applications, such as portable computing, impose additional demands of high-speed operation at low supply voltages. It is mandatory from a financial feasibility standpoint that new technologies be compatible with existing Si processing techniques, so as not to require extensive retooling.

There have been previous suggestions to use complementary heterostructure field-effect transistors in GaAs/AlGaAs in order to make use of the high electron mobility in this material system [3]. However, the problem of the low hole mobility which plagues GaAs

was not solved. In addition, that technology relied on making Schottky gates, which result in several orders of magnitude higher gate leakage current than in MOS devices.

One promising new technology is SiGe strained-layer epitaxy, which aims to improve silicon MOSFET performance by offering increased electron and hole mobilities. Unlike GaAs-based high-mobility technologies, SiGe epitaxy is largely Si-compatible and thus readily integrable into mainstream CMOS fabrication lines. By allowing an increase in hole mobility, balanced operation of NMOS and PMOS devices can be achieved using SiGe technology, which leads to reduced circuit area and faster performance. The high current drive at low supply voltage afforded by high carrier mobilities makes low-power digital applications favorable for SiGe. The combination of high-mobility buried-channel operation with an oxide gate insulator, impossible in any other material system, points to low power analog applications.

Although it is inconceivable that SiGe-based CMOS will replace Si CMOS in ULSI applications in the near future, there is increasing interest in application-specific designs which require low power consumption and high speed (e.g. cellular phones, other portable electronics, optoelectronic receivers, etc.). This is where we believe that there is ample room for implementing SiGe-based CMOS, making use of its potential performance leverage over Si CMOS.

A simple analysis of the 1997 SIA National Technology Roadmap for Semiconductors requirements for NMOS using the equation,

$$v_{eff} = \frac{I_{on}}{Q_{inv}} = \frac{I_{on}}{C_{ox}(V_{DD} - V_T)}, \qquad (1.1)$$

illustrates the need for higher effective velocities, v_{eff}, for electrons in the channel at shorter length generations in order to make up for diminishing inversion layer charge, Q_{inv}, as shown in Fig. 1.1. The situation becomes worse if a method is not found for reducing the gate insulator thickness to below 20 Å of equivalent oxide thickness. The question is, can Si NMOS achieve these velocities?

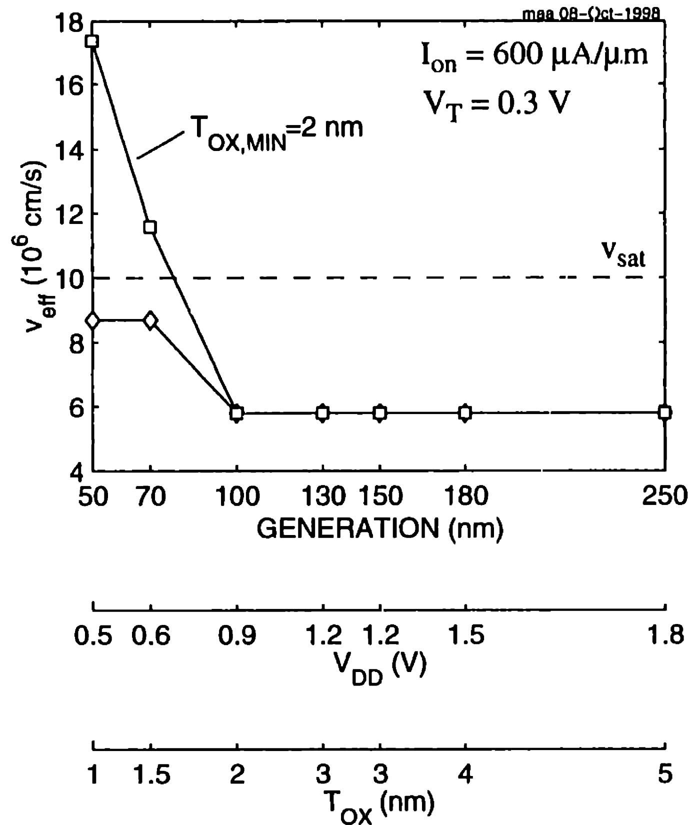


Fig. 1.1. Effective velocity v_{eff} of electrons in the channel vs. SIA Technology Roadmap generation for NMOS. Diamond points are calculated using (1.1) with I_{on} = 600 μA/μm and V_T = 0.3 V and V_{DD} and t_{ox} as listed on the lower two axes. Squares are the same but with a minimum allowed t_{ox} of 2 nm.

We can answer this question by looking at the quantity $\frac{g_m}{WC_{ox}}$, a measure of carrier velocity near the source approximately equal to v_{eff}, versus DIBL, a measure of short-channel effects. For a given source/drain structure, this trade-off is insensitive to most device parameters, such as N_A, t_{ox} or V_T, but does depend greatly on low-field mobility μ_0. Fig. 1.2 compares simulated results for a high mobility strained-Si NMOS device (μ = 2500 cm^2/Vs, see Chapter 4) to measured results for bulk Si NMOS. For reasonable values of DIBL, around 100 mV/V, bulk Si NMOS does not achieve the v_{eff} required for generations beyond 70 nm. However, the requirement can be met in theory by the strained-Si NMOS (labeled HNMOS for heterostructure-based NMOS), due to the high electron mobility in strained-Si material.

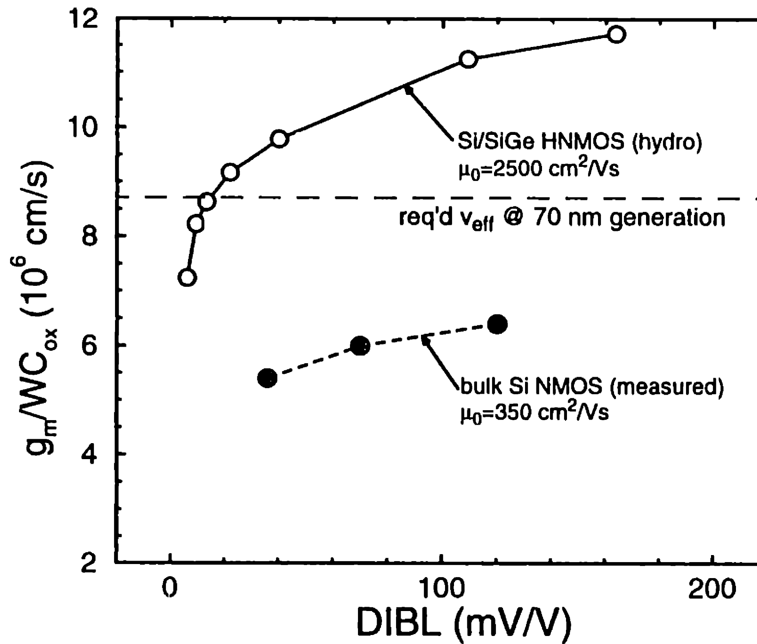


Fig. 1.2. Effective velocity in the channel vs. DIBL for bulk-Si NMOS and Si/SiGe NMOS (labelled HNMOS for heterostructure-based NMOS). The HNMOS curve is simulated in Chapter 4. The bulk curve is measured but is the same as simulation.

1.2 Goals

The key issues for integration of SiGe into mainstream CMOS fall into three main areas as shown in Fig. 1.3: materials, process and operation. These three areas are interrelated. For instance, the number of dislocations in the substrate, a materials issue, is related to the Ge-content in the substrate. The Ge-content in the substrate may also affect the thermal budget, a process issue, and the amount of self heating, an operation issue.

This thesis examines issues in all of these areas. The first part of this work strives to predict the performance leverage of high-mobility Si/SiGe over bulk Si devices and circuits in a realistic manner. This is an operation-related issue. Two-dimensional hydrodynamic simulations are used to predict static device characteristics including effects of series resistance, velocity saturation and velocity overshoot. The simulations show enhanced current drive over bulk Si devices at 0.2 μm effective channel length and highlight the importance of velocity overshoot in high-mobility submicron devices. The circuit performance of Si/SiGe devices is determined from transient simulations of CMOS ring oscilla-

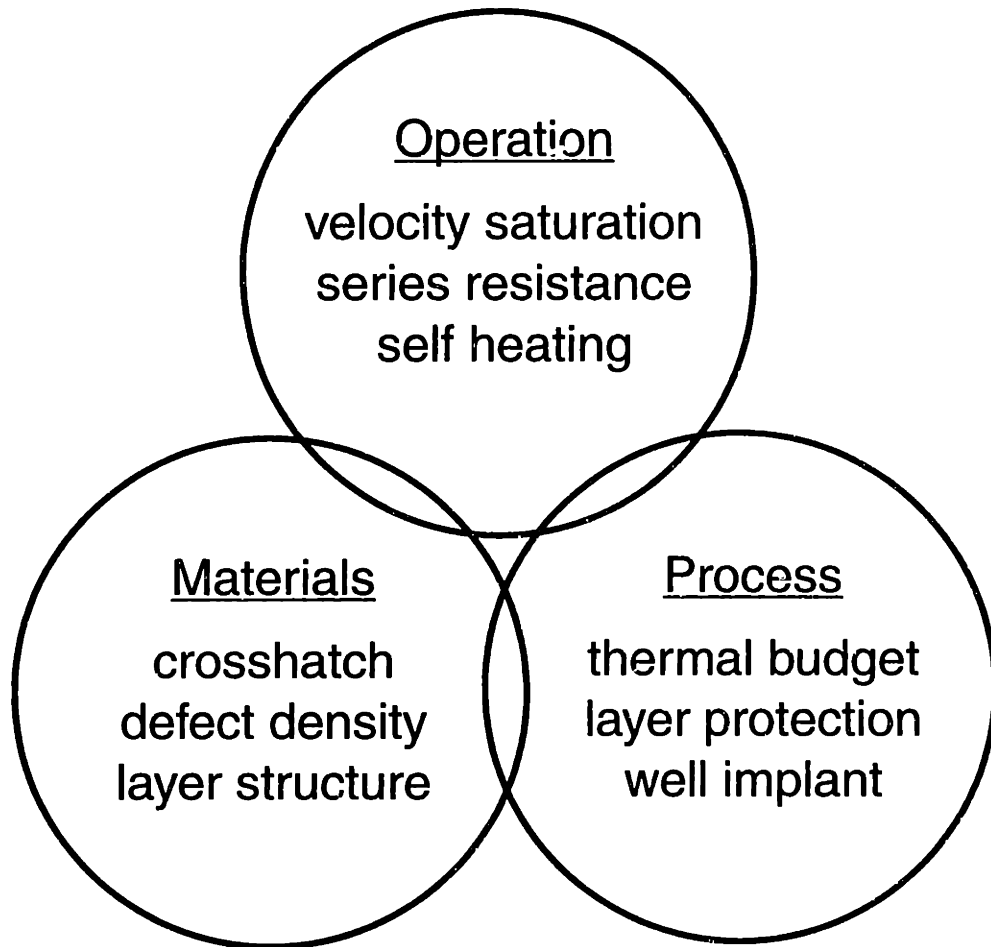


Fig. 1.3. Key issues for integration of SiGe-based strained-layer technology into mainstream CMOS applications.

tors including the effects of parasitic capacitance and drain-to-source voltage at the onset of saturation $V_{DS,sat}$. The simulations show a 4 to 6-fold reduction in power-delay product as compared to bulk CMOS oscillators operated at 2.5 V with the same design rules.

The remainder of the thesis focuses on the fabrication and characterization of strained-Si NMOS devices. The vehicle for this work is a novel short-flow, single-mask MOSFET which can be fabricated in as little as a week. This device is superior to simple Hall mobility structures which suffer from leakage through the substrate, an inability to control the carrier concentration and the uncertainty associated with the Hall scattering factor.

I investigate a novel buried-channel strained-Si NMOS structure incorporating an n-type donor layer beneath the strained-Si channel to encourage occupation of the buried channel

and increase the overall mobility. This is primarily a materials-related issue of optimum layer structure. Peak mobility in a structure without a donor layer reproduces the best results in the literature for buried-channel strained-Si NMOS devices. For structures with donor layers, Coulomb scattering from charges in the donor layer eradicates any benefit from increased buried-channel occupation.

I also investigate the effect of well implants on the mobility of surface-channel strained-Si NMOS devices. This is a process-related issue. Similar to the universal mobility curve in bulk Si, mobility at low perpendicular electric field degrades with increasing implant dose while high field mobility is unaffected. The mobility is largely unaffected by a neutral implant species at the same dose. This leads to the conclusion that the material quality of the strained-layer is not affected by the implant, and that the mobility degradation is due solely to increased ionized impurity scattering.

1.3 Outline

Chapter 2 describes the general properties of Si and SiGe strained-layers including lattice and band structure and electron and hole mobility.

Chapter 3 introduces the layer structure and operation of NMOS and PMOS devices incorporating high mobility Si and SiGe strained-layers, including designs suitable for CMOS applications.

Chapter 4 presents both a theoretical and numerical analysis of the impact of mobility on submicron CMOS device and circuit performance.

Chapter 5 describes a short-flow MOSFET process for rapid evaluation of device performance of candidate Si/SiGe heterostructures.

Chapter 6 is an experimental investigation of the effect of donor layers on buried-channel strained-Si NMOS devices.

Chapter 7 is an experimental investigation of the effect of well implants on surface-channel strained-Si NMOS devices.

Chapter 8 summarizes the key contributions of this thesis and provides suggestions for future work.

Appendix A presents various issues and experiments related to low temperature processing technology.

Chapter 2

Si/SiGe Strained-Layers

2.1 Introduction

This section covers the material and electrical properties of Si and SiGe alloys and strained-layers relevant to high-mobility MOSFET applications. I discuss the lattice structure, band structure and mobility of the technologically significant strained-layer combinations.

2.2 Lattice Structure

Both Si and Ge exist in crystal form in a diamond lattice structure [23]. A SiGe alloy is created by mixing Ge and Si atoms within the same lattice. The macroscopic lattice constant of a uniformly mixed alloy in its relaxed state will be approximately a linear interpolation between the lattice constants of pure Si and Ge,

$$a_{SiGe} = a_{Si} + (a_{Ge} - a_{Si})x\,, \qquad (2.1)$$

where x is the fraction of Ge in the mixture, $a_{Ge} = 5.658$ Å and $a_{Si} = 5.428$ Å [24].

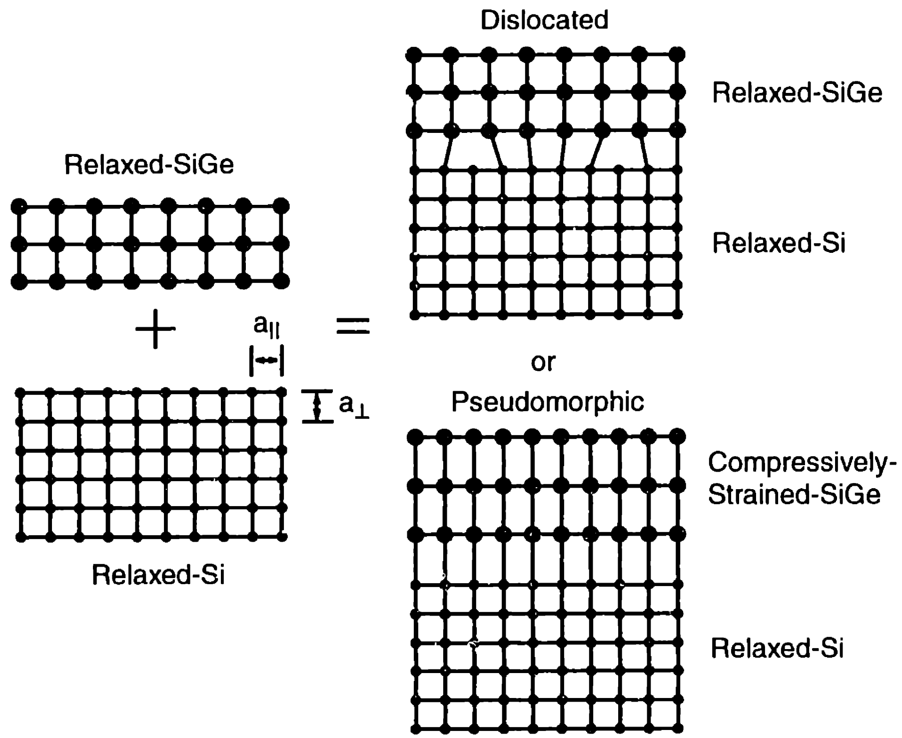
Of greater interest are thin SiGe layers or layer stacks which exhibit unique properties. One way to form SiGe layers in practice is by ultra-high vacuum chemical vapor deposi-

tion (UHVCVD), which involves flowing the proper amounts of silane and germane into a reaction chamber, where they react and SiGe films are deposited on a Si substrate [26]. By modulating the gas flows, stratified layers of varying Ge content, or heterostructures, can be grown. The layers can also be doped n- or p-type by flowing diborane and phosphine into the chamber, resulting in a modulation-doped heterostructure.

Because the lattice constant varies as function of Ge content, there will be a lattice mismatch between the epitaxial film and the substrate. For instance, the relaxed lattice constants of Si and $Si_{0.7}Ge_{0.3}$ differ by about 1.3%. When a thin $Si_{0.7}Ge_{0.3}$ layer is grown on top of relaxed-Si, as shown in Fig. 2.1(a), it is energetically favorable for the film to deform so that the in-plane lattice constant $a_{||}$ is reduced to match that of the substrate, while the out-of-plane lattice constant $a_{\perp}$ will increase according to the elastic properties of the lattice in much the same way as a block eraser will react to being pinched between the fingers. This is called pseudomorphic growth and the SiGe film is said to be under compressive strain or stress.

When the film exceeds a certain critical thickness dislocations form to relieve the strain and the film relaxes. Fig. 2.2 shows how critical thickness is a decreasing function of the Ge content. The critical thickness is fundamentally a function of the lattice mismatch, rather than Ge content itself, so this figure can be used to estimate the critical thickness of $Si_{1-x}Ge_x$ grown on $Si_{1-y}Ge_y$. The upper curve was determined by People from channeling and TEM [55], methods which are sensitive the structural properties of the film such as the macroscopic lattice constant. Other researchers have found a critical thickness of about half this value through methods sensitive to the electrical properties of the film [25] such as carrier lifetime, which is strongly dependent on the number of dislocations. The relaxation is not abrupt upon exceeding the critical thickness, but rather is a function of the degree to which the critical thickness is exceeded. Pseudomorphic films of thickness less than the critical thickness are stable at all temperatures and will not relax during processing. However, the heterostructure can be destroyed by Ge diffusion [29].

Strained-Si is the result of depositing Si on top of a relaxed-SiGe substrate as shown in Fig. 2.1(b). This material will be under tensile strain and the in-plane lattice constant will



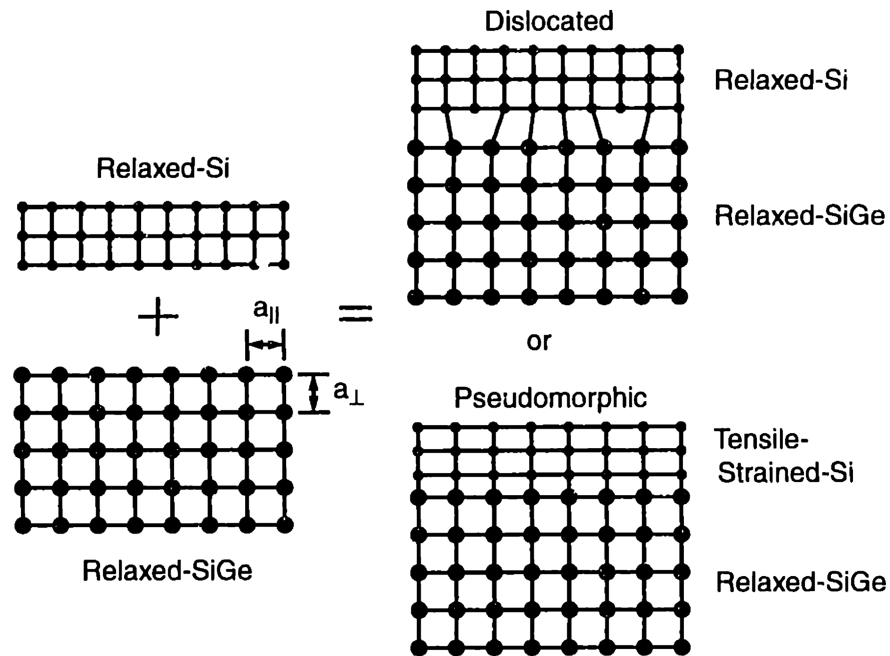


Fig. 2.1. Simplified lattice-structure for growth of (a) SiGe on relaxed-Si and (b) Si on relaxed-SiGe. The epitaxial film can be pseudomorphic or dislocated, corresponding to a strained or relaxed condition [25].

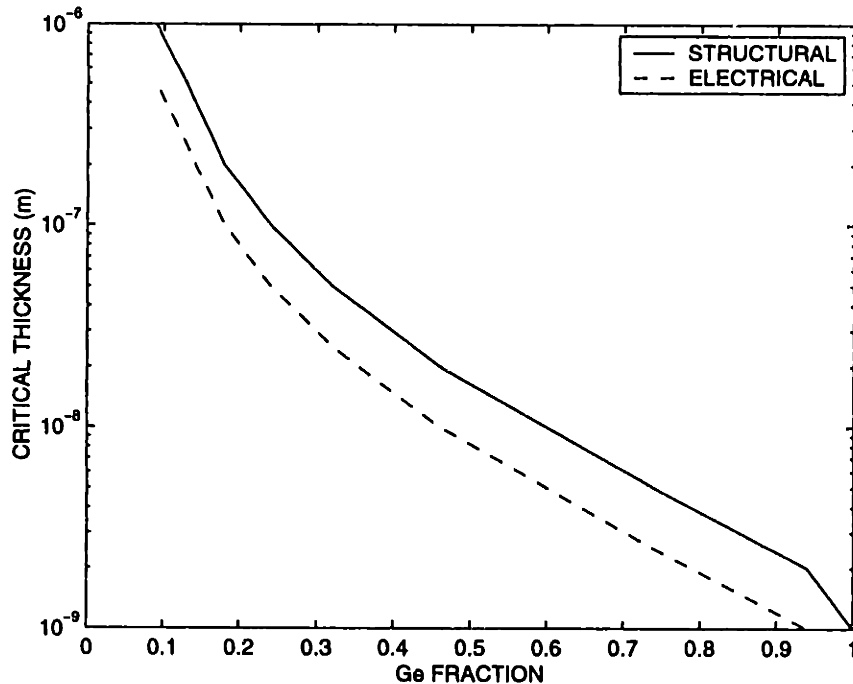


Fig. 2.2. Critical thickness as a function of Ge content x for growth of strained-$Si_{1-x}Ge_x$ on a relaxed-Si substrate. Solid line shows the result of methods sensitive to the structural properties of the film [55]. Dashed line shows average results of methods sensitive to the electrical properties [25].

be increased over equilibrium. A relaxed-SiGe substrate can be created using a graded buffer technique [56] where the Ge content is slowly graded up over several microns starting with a Si substrate. This causes threading dislocations to bend so that they travel more parallel to the wafer surface than perpendicular. These dislocations will tend to terminate on each other forming loops or at the edges of the wafer and the resulting SiGe material on top has a threading dislocation density of about 10^5 cm^{-2}, greatly reduced from SiGe of the same mole content grown directly on Si. Dislocation density is still much worse than bulk silicon, which has less than 10 cm^{-2}.

In general it is possible to grow $Si_{1-x}Ge_x$ on $Si_{1-y}Ge_y$, where the grown film is tensile strained if $x < y$ and compressively strained if $x > y$, although only certain combinations are technologically interesting for MOSFET applications. For instance, the critical thickness puts an upper limit on the difference between x and y in order to have films of usable thickness. The band structure and mobility also favor certain layer combinations as will be

covered in the next sections. The most interesting layer combinations are summarized in Table 2.1.

TABLE 2.1
TECHNOLOGICALLY SIGNIFICANT SI/SIGE STRAINED-LAYERS

Strained-Layer	on	Substrate	Features
$Si_{0.7}Ge_{0.3}$	on	Si	compressive strain hole confinement Si substrate - no graded buffer
$Si_{0.2}Ge_{0.8}$	on	$Si_{0.7}Ge_{0.3}$	compressive strain hole confinement high hole mobility
Si	on	$Si_{0.7}Ge_{0.3}$	tensile strain electron confinement high electron and hole mobility

2.3 Band Structure

In order to design devices with strained-layers, we need to know the bandgaps of the layer and the substrate and the band offsets between the layer and the substrate. These parameters are functions of both the Ge content and the amount of strain in the materials, and the temperature. The emphasis in this section is obtaining accurate room temperature data for the material combinations of Table 2.1.

Relaxed-$Si_{1-x}Ge_x$

Fig. 2.3 shows the bandgap of relaxed-$Si_{1-x}Ge_x$ as a function of x from optical absorption measurements on substrates pulled from melt by Braunstein [50]. The energy bandgap of a SiGe alloy varies as a non-linear function of Ge content, dropping off rapidly at Ge contents above 80%. This corresponds to where the conduction band changes from Si-like with a Δ-minimum to Ge-like with an L-minimum.

Notice that in this data the values of E_g at the extremes of pure Si and Ge are slightly less than the generally accepted values of $E_{g,Si}$ = 0.66 eV and $E_{g,Ge}$ = 1.12 eV at room temper-

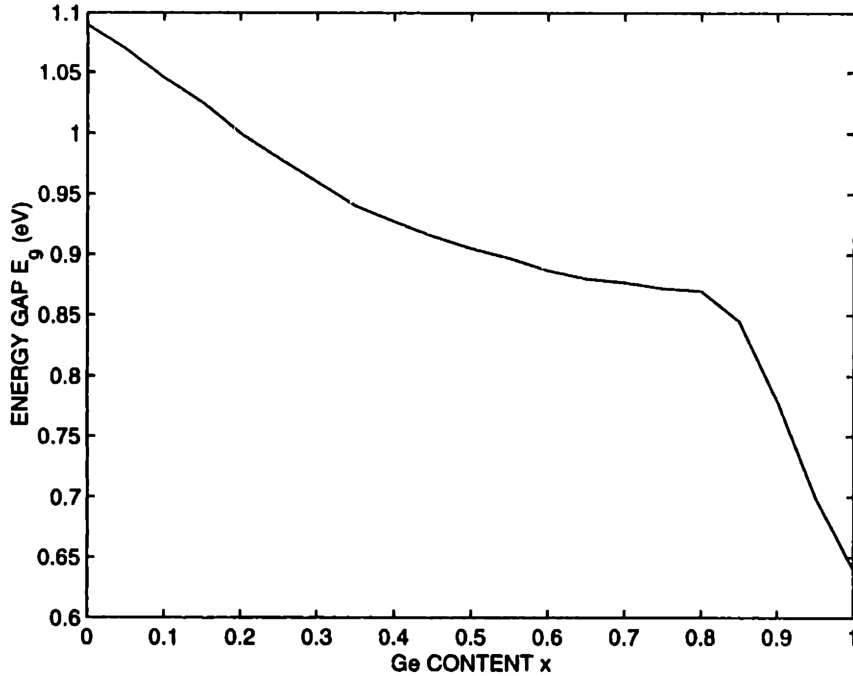


Fig. 2.3. Bandgap vs. Ge content for relaxed-$Si_{1-x}Ge_x$ at 296K [50].

ature. In the literature this curve is often used to calculate the bandgap difference between different alloys rather than the absolute values.

The electron affinity χ of relaxed-$Si_{1-x}Ge_x$ as a function of Ge content x is not available in the literature. Instead, it is generally assumed that the electron affinity χ is the same in Si and relaxed-$Si_{1-x}Ge_x$ for low values of x [51]. One can probably assume it lies between the extremes of $\chi_{Si} = 4.05$ and $\chi_{Ge} = 4.00$ for all values of x. Welser estimated an upper bound on $\Delta\chi = 180$ meV between relaxed-Si and relaxed-$Si_{0.62}Ge_{0.38}$ from flatband shift of MOS C-V measurements [45]. However, Welser noted that the flatband shift due to $\Delta\chi$ is hard to separate from the shift due to oxide fixed charge Q_F which was also unknown.

Strained-$Si_{1-x}Ge_x$ on Relaxed-Si

Fig. 2.4 shows the valence band offset ΔE_V between strained-layer and substrate for strained-$Si_{1-x}Ge_x$ grown on relaxed-Si [51][52][53] at room temperature. The plotted line is simply:

$$\Delta E_V = 0.74x, \tag{2.2}$$

in eV where the valence band is higher in energy in the strained-SiGe layer. This particular curve has received special attention because it is important for the design of npn heterojunction bipolar transistors (HBT) which have been the main application of SiGe materials to date. It has been confirmed experimentally by comparing the current gain of Si and Si/SiGe HBTs through $x < 0.4$ [51] and through photocurrent spectroscopy on Si/SiGe superlattices through $x < 0.7$ [54]. The conduction band offset ΔE_C is less than about 20 meV for all x.

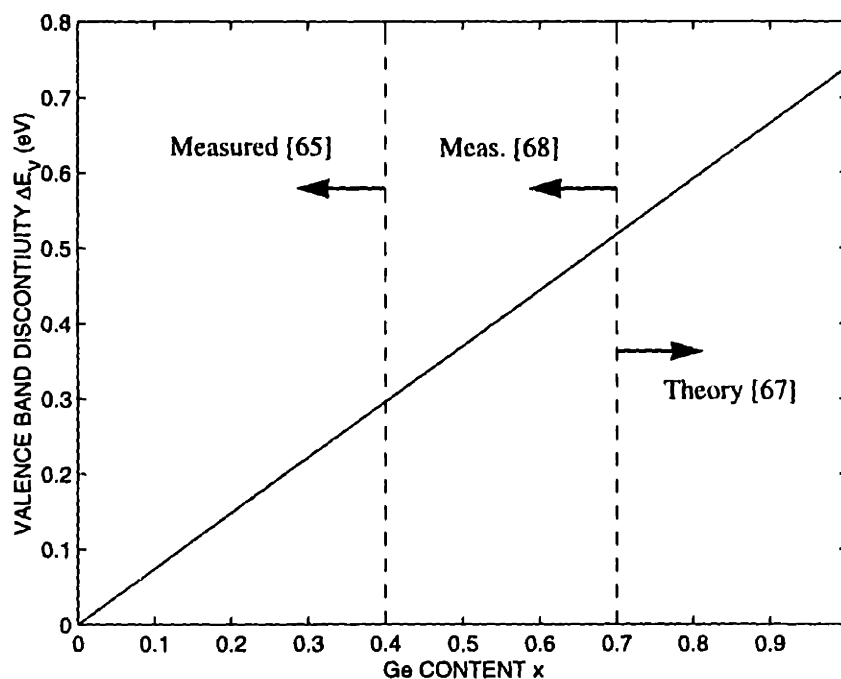


Fig. 2.4. Valence band offset ΔE_V vs. Ge content x for compressively-strained-$Si_{1-x}Ge_x$ on relaxed-Si at room temperature. The valence band is higher in strained-SiGe.

Often the applicable temperature for the data in the literature is unclear. King noted that this data is fairly independent of temperature in the range $50K < T < 300K$ and $0 < x < 0.5$ [51] which may explain why his data matches well with earlier experiments carried out at $T = 90K$ [54].

Based on this data and the bandgap of relaxed-Si we can draw the band alignments for strained-$Si_{0.7}Ge_{0.3}$ on relaxed-Si, shown in Fig. 2.5(a). In general, compressively-

strained-SiGe will have a higher valence band than the surrounding material, and will serve as a quantum well for holes.

Strained-$Si_{1-x}Ge_x$ on Relaxed-$Si_{0.7}Ge_{0.3}$

Fig. 2.6 shows the results of several theoretical predictions [52][13][53] for the band offsets of strained-$Si_{1-x}Ge_x$ on $Si_{0.7}Ge_{0.3}$ as a function of the Ge content x. This case has less experimental confirmation than the case of strained-SiGe on Si. However, for $x > 0.3$ the band offsets should behave similarly to the case of strained-SiGe on relaxed-Si, which has been well characterized. In addition, for $x < 0.3$ the band offsets should approach those of strained-Si on $Si_{0.7}Ge_{0.3}$ which has also been experimentally explored.

For the valence band offset there is fair agreement between the various theories for $x > 0.3$. People's theory for valence band offset based on the work of Van de Walle [57] follows the general function:

$$\Delta E_V = (0.74 - 0.53y)(x - y) \tag{2.3}$$

in eV where x is the Ge content of the strained-layer and y is the Ge content of the relaxed substrate. In the present case of a relaxed-$Si_{0.7}Ge_{0.3}$ substrate we have $y = 3$ and:

$$\Delta E_V = 0.58(x - 0.3) \tag{2.4}$$

in eV. Rieger and Schaffler, on the other hand, both agree with:

$$\Delta E_V = 0.71(x - 0.3) \tag{2.5}$$

in eV.

As for the conduction band offset, both theories predict a fairly small value with Schaffler's being near zero. This coincides with the results for strained-SiGe on relaxed-Si where negligible conduction band offsets were found.

Using Rieger's and Schaffler's theories for ΔE_V and assuming a negligible ΔE_C we can draw the band alignments for strained-$Si_{0.2}Ge_{0.8}$ on relaxed-$Si_{0.7}Ge_{0.3}$ as shown in Fig. 2.5(b). The valence band in strained-$Si_{0.2}Ge_{0.8}$ is quite high compared to the substrate and there will be many more holes in the strained-layer than the surrounding material.

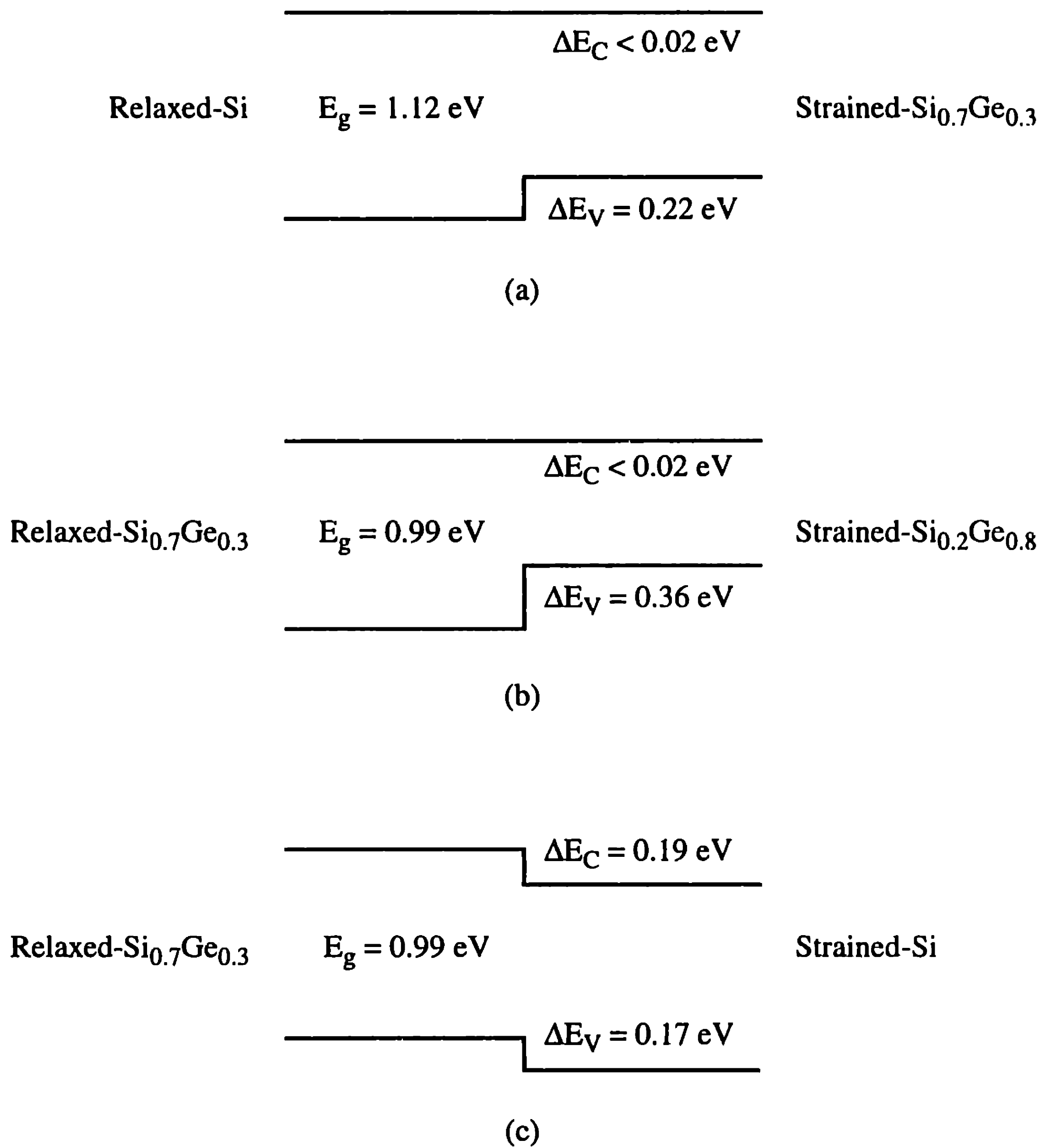
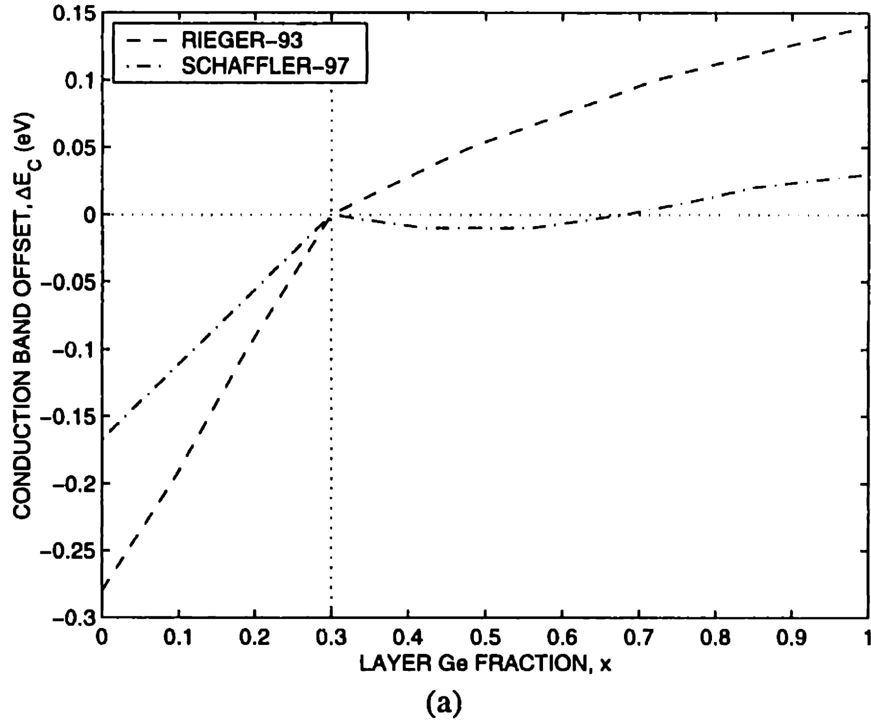
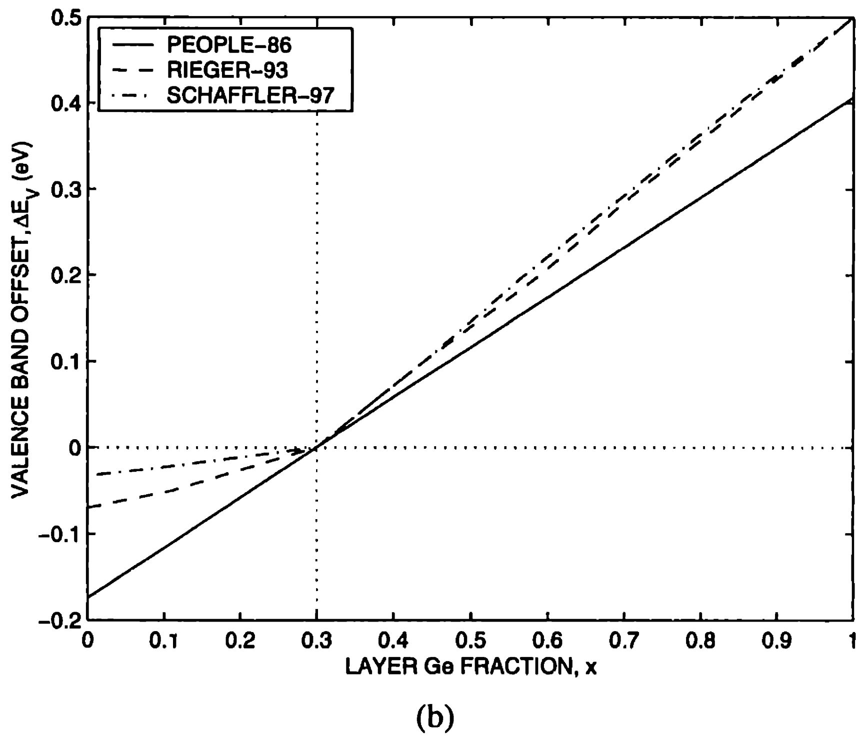


Fig. 2.5. Room temperature band alignments of three technologically significant layer combinations using consensus of data from the literature: (a) strained-$Si_{0.7}Ge_{0.3}$ on relaxed-Si, (b) strained-$Si_{0.2}Ge_{0.8}$ on relaxed-$Si_{0.7}Ge_{0.3}$ and (c) strained-Si on relaxed-$Si_{0.7}Ge_{0.3}$.



(a)



(b)

Fig. 2.6. (a) Conduction and (b) valence band offsets for strained-$Si_{1-x}Ge_x$ on relaxed-$Si_{0.7}Ge_{0.3}$ vs. strained-layer Ge content x [52][13][53]. Conduction and valence bands are higher in the strained-$Si_{1-x}Ge_x$ for positive values of ΔE_C and ΔE_V.

Strained-Si on Relaxed-$Si_{1-x}Ge_x$

Fig. 2.7 shows the conduction band and valence band offsets for strained-Si grown on a relaxed-$Si_{1-x}Ge_x$ substrate as a function of the substrate Ge content x comparing several theoretical predictions [52][13][53] to the MOS C-V data of Welser [45]. Both the conduction and valence bands are lower in the strained-Si layer.

I produced People's curve in the ΔE_C plot following Welser's formula:

$$\Delta E_C = \Delta E_V + E_{g,\,\text{strained-Si}} - E_{g,\,\text{relaxed-SiGe}}, \tag{2.6}$$

where ΔE_V and $E_{g,\text{strained-Si}}$ are from People [52] and $E_{g,\text{relaxed-SiGe}}$ is from Braunstein [50]. Welser's ΔE_C agrees with both People's and Schaffler's theories for $x < 0.4$ and follows the functional form:

$$\Delta E_C = 0.63x \tag{2.7}$$

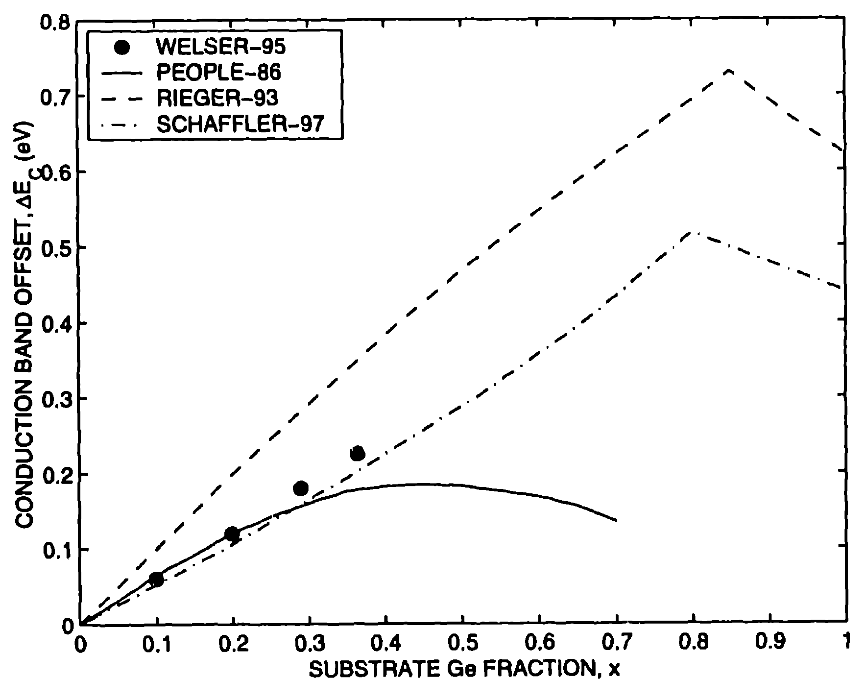
in eV. For $x > 0.4$, People's theory diverges significantly from Schaffler's and Rieger's, predicting a peak in ΔE_C at $x = 0.45$ while the latter two predict a peak at around $x = 0.8$.

Welser's data for ΔE_V agrees with only People's theory for $x < 0.4$ and follows the functional form:

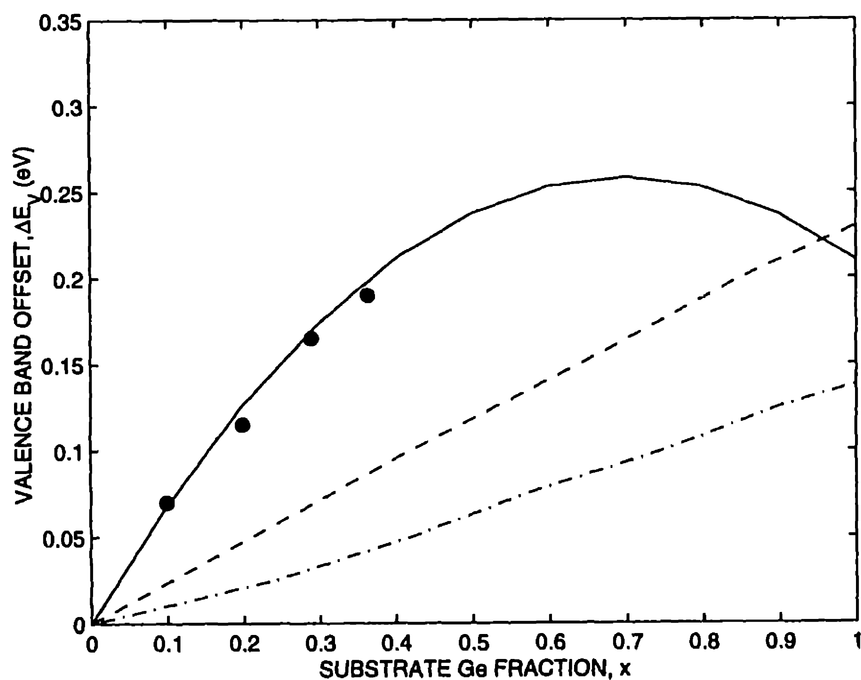
$$\Delta E_V = (0.74 - 0.53x)x \tag{2.8}$$

in eV [52]. For $x > 0.4$, People predicts a peak in ΔE_V at $x = 0.7$ while Schaffler and Rieger both predict a monotonic rise.

Strained-Si has only recently attracted attention following the introduction of the graded-buffer technique for producing low-dislocation-density, relaxed-SiGe substrates in 1991 [56] and hence there is much less consensus in the literature concerning the band offsets than for strained-SiGe grown on relaxed-Si. We will use Welser's data since it has some experimental confirmation behind it and agrees with at least one of the theories. Based on Welser's strained-Si data and Braunstein's relaxed-SiGe data (Fig. 2.3) we can draw the band alignments for strained-Si grown on relaxed-$Si_{0.7}Ge_{0.3}$ as shown in Fig. 2.5(c). The strained-layer has a lower conduction and valence band than the relaxed substrate. Electrons will tend to accumulate in the strained-Si, while holes will tend to accumulate in the



(a)



(b)

Fig. 2.7. (a) Conduction and (b) valence band offsets for strained-Si on relaxed-$Si_{1-x}Ge_x$ vs. substrate Ge fraction x comparing several theoretical predictions [52][13][53] to the MOS C-V data of Welser [45]. Both the conduction and valence bands are lower in strained-Si.

relaxed-$Si_{0.7}Ge_{0.3}$. In determining E_g of relaxed-$Si_{0.7}Ge_{0.3}$ from Fig. 2.3, an offset was added to make $E_g = 1.12$ eV for relaxed-Si.

2.4 Transport

Low-field mobility for electrons and holes is given by:

$$\mu = \frac{e\tau}{m^*}, \tag{2.9}$$

where e is the electronic charge, τ is the momentum relaxation time and m^* is the conductivity effective mass.

Mobility is strongly dependent on temperature and, in non-isotropic material, direction of transport. We are interested in the room temperature, in-plane mobilities of the materials listed in Table 2.1.

Relaxed-$Si_{1-x}Ge_x$

Fig. 2.8 shows the mobility in undoped relaxed-$Si_{1-x}Ge_x$ as a function of Ge content x [58]. The mobility in such bulk alloys is generally lower than in bulk Si until the alloy approaches pure Ge. This is due to alloy scattering.

The momentum relaxation time in Si is made up of several contributions:

$$\frac{1}{\tau_{Si}} = \frac{1}{\tau_{ii}} + \frac{1}{\tau_{ac}} + \frac{1}{\tau_{op}}, \tag{2.10}$$

such as ionized-impurity scattering τ_{ii}, acoustic phonon scattering τ_{ac} and optical phonon scattering τ_{op}. The random distribution of Si and Ge atoms in the lattice causes a variation in the potential which tends strongly to scatter the carriers and which has a scattering time τ_{alloy} associated with it. This scattering time is combined with the rest:

$$\frac{1}{\tau_{SiGe}} = \frac{1}{\tau_{Si}} + \frac{1}{\tau_{alloy}}, \tag{2.11}$$

reducing the overall mobility in SiGe. τ_{alloy} follows a dependence:

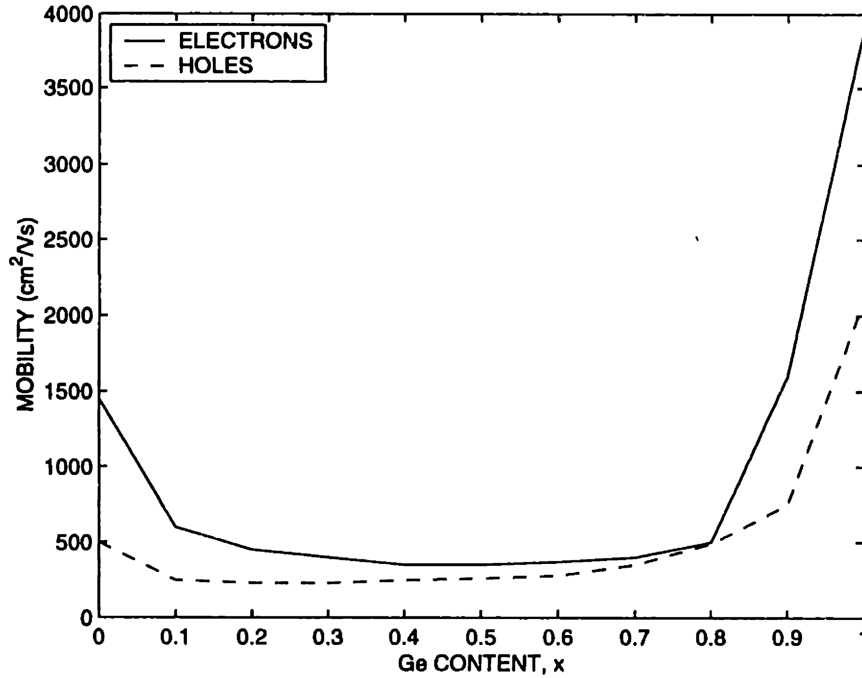


Fig. 2.8. Low-field mobility for (a) electrons and (b) holes in undoped, relaxed-$Si_{1-x}Ge_x$ as a function of Ge content x. Data is a calculation [58] based on the measured data of [60].

$$\frac{1}{\tau_{alloy}} \propto U^2 x(1-x), \tag{2.12}$$

where U is the alloy scattering potential, a constant with different values for electrons and holes, and x is the Ge content [62]. This leads to the bowl shaped curves in Fig. 2.8.

There is some debate on the validity of Fig. 2.8 because it is based on the data of Busch [60] which is Hall mobility data for polycrystalline material. For instance, Manku [62] and Levitas [63] found a higher hole mobility in SiGe alloys, while Glicksman [59] found a higher electron mobility. However, Schaffler's analysis [53] and other recent data [64][65][79] seems to confirm the strong alloy scattering potential at least for holes.

We are primarily interested in the value at x = 0.3 which is listed in Table 2.2 along with values for pure Si and Ge.

TABLE 2.2
PROPERTIES OF RELAXED MATERIALS

	Si	$Si_{0.7}Ge_{0.3}$	Ge
Electron Mobility (cm^2/Vs)	1450	400	3900
Hole Mobility (cm^2/Vs)	450	230	1800

Strained-$Si_{1-x}Ge_x$ ($x < 0.6$)

Fig. 2.9 shows the room-temperature in-plane hole mobility in strained-$Si_{1-x}Ge_x$ on relaxed-Si as a function of Ge content x comparing several theoretical predictions to various experiments. The theoretical data is for undoped or lightly-doped material. The experimental data is for both undoped and modulation-doped heterostructures. Mobility in these structures is often a function of the applied vertical field. The figure shows peak mobilities which typically occur at low vertical fields.

The shape of the theoretical curves and the variation between them can be explained in the following way. Strain leads to a modification of the valence band structure of SiGe as shown in Fig. 2.10(a) [74]. In both the compressive and tensile strain case the light-hole and heavy-hole bands are split. In the case of strained-$Si_{1-x}Ge_x$ on relaxed-Si we have compressive strain, so the heavy-hole band is higher in energy and will dominate the transport properties of the material as it will contain most of the holes.

The splitting of the LH and HH bands leads to a reduced intervalley scattering rate because of a reduced density of states available for scattering. This in turn increases the phonon-related momentum relaxation time and increases the mobility. The higher the strain, the more splitting occurs and thus the trend towards higher mobility with increasing strained-layer Ge content in Fig. 2.9.

The effective mass also changes due to the strain. In general, the effective mass decreases with compressive strain and increases with tensile strain [66][76]. In fact, the so-called heavy-hole band may have a lighter effective mass than the light-hole band under some

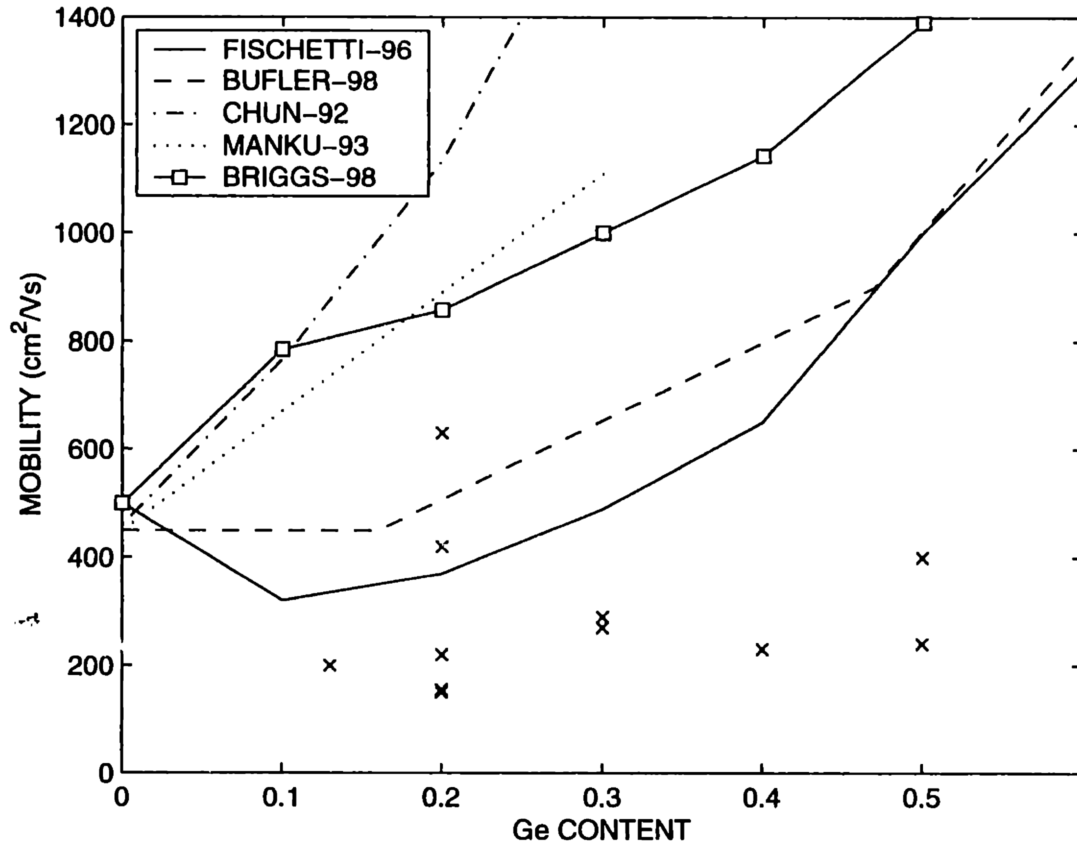


Fig. 2.9. Room-temperature in-plane hole mobility in undoped strained-$Si_{1-x}Ge_x$ on relaxed-Si vs. Ge content. Several theoretical predictions (lines) [58][61][62][66][74] are compared to various experiments (x marks) [4][5][73][72][71][70][67][68][69].

conditions. In the case of strained-$Si_{1-x}Ge_x$ on relaxed-Si we have compressive strain and thus the mobility should increase due to a decreased m* as well.

Some curves in Fig. 2.9 show a dip in mobility, though, and this is caused by alloy scattering due to the presence of both Si and Ge in the lattice. As shown in (2.12), alloy scattering follows an $x(1-x)$ dependence, which leads to the depression in mobility at medium x. The lack of a dip for some of the curves can be explained by the different values used for alloy scattering potential U. The amount of alloy scattering is determined by the alloy scattering potential U as shown in (2.12). The different theoretical predictions in Fig. 2.9 use different values for U: Fischetti and Bufler have higher values ($U > 0.7$ eV) than the other curves ($U < 0.4$ eV). The higher the value for U, the more pronounced the dip in mobility. The curve of Briggs, however, is an exception, for he uses a high value for U (U

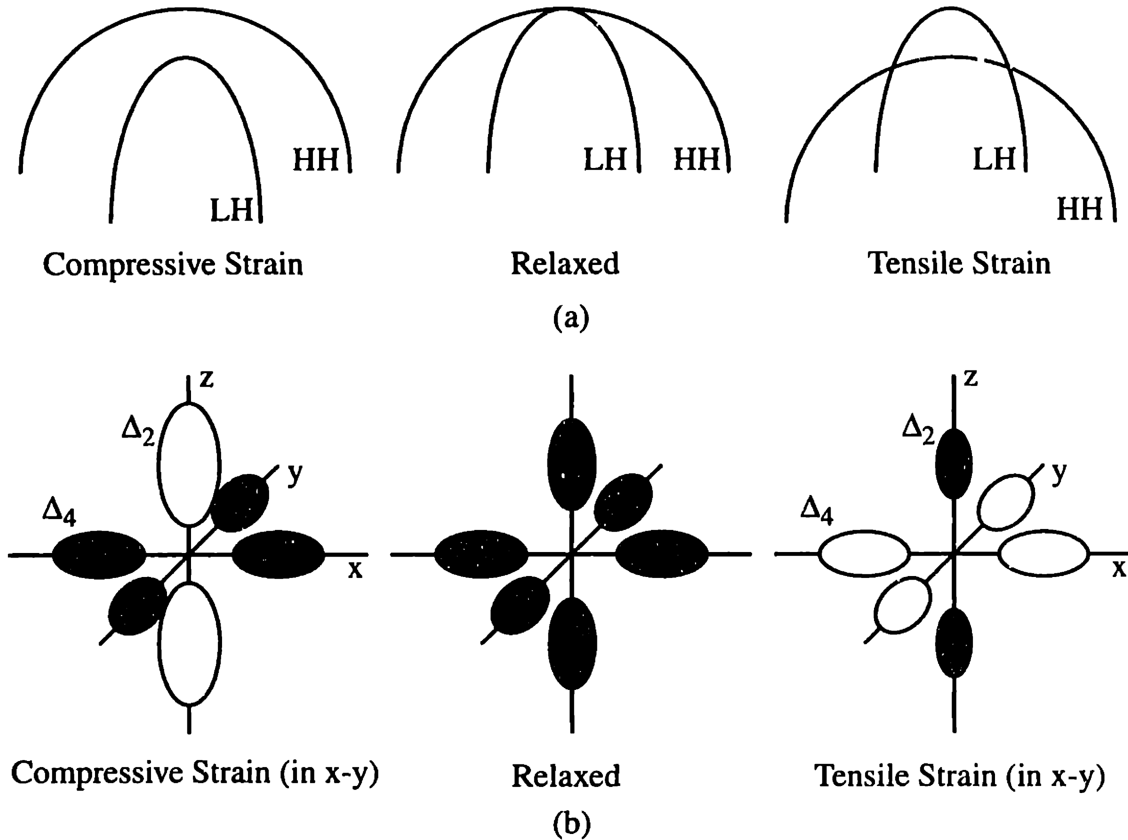


Fig. 2.10. (a) Simplified valence band E vs. k diagram around the Γ-point of Si showing the splitting of the light-hole (LH) and heavy-hole (HH) bands under strain [74]. (b) Conduction band surfaces of constant energy in k-space of Si showing the splitting of the 6-fold degenerate Δ-minimum into Δ_2 and Δ_4-minima under strain [75]. Filled ovals represent those lower in energy and thus relevant to electron transport. This diagram is schematic only and does not reflect changes in effective mass. These diagrams are relevant for both Si and $Si_{1-x}Ge_x$ for $x < 0.8$. For $x > 0.8$, the conduction band in relaxed material changes from Si-like with a 6-fold degenerate Δ-minimum to Ge-like with an 8-fold degenerate L-minimum [50].

= 1.4 eV), but does not show a significant dip. Thus, the importance of other scattering mechanisms, such as phonon scattering, must be enhanced in his simulation. Unfortunately, U and other parameters are not very well known and this leads to the large differences between various theoretical models. Theoretically, U for holes should be approximately the change of the valence band offset with composition [53]. In the case of strained-$Si_{1-x}Ge_x$ on relaxed-Si we have U = 0.74 eV from (2.2).

Looking at the experimental points in Fig. 2.9, we see that actual room-temperature mobilities measured to date have been very disappointing. Values are generally less than 300 cm^2/Vs. It is not known why this is the case. Most encouraging is the value of 630 cm^2/Vs recorded by Lander [68] for $Si_{0.8}Ge_{0.2}$ on Si. This was a non-modulation-doped structure, indicating that ionized-impurity scattering may be a limiting factor in some of the other results.

In-plane electron mobility in undoped strained-$Si_{1-x}Ge_x$ on relaxed-Si is expected to be lower than in bulk Si for all Ge contents, and less than 500 cm^2/Vs for $0.1 < x < 0.9$ [58][75].

At high parallel fields, the carrier velocity will saturate to a value v_{sat}. At this point, transport is dominated by the optical phonon scattering mechanism. v_{sat} is related to the optical phonon energy, which is in turn related to the stiffness of the lattice, which is a function of the bond strength between the atoms in the lattice. For instance, v_{sat} increases going from Ge ($6x10^6$ cm/s) to Si ($1x10^7$ cm/s) to diamond ($2.7x10^7$ cm/s) due to increasing bond strength. Thus, one might expect a slight decrease in SiGe alloy. The saturated velocity of holes in strained-$Si_{1-x}Ge_x$ is the subject of some debate. For instance, Bhaumik [28] calculated a 14% increase in v_{sat} in strained-$Si_{0.7}Ge_{0.3}$ on relaxed-Si compared to bulk Si, while Bufler calculated a 24% decrease [61].

Strained-$Si_{1-x}Ge_x$ ($x > 0.6$)

Due to the poor performance of low Ge content strained-SiGe on relaxed-Si in terms of hole mobility, there has been recent interest in higher Ge content layers. The reasoning behind this is that the alloy scattering term, which is proportional to $x(1-x)$, should become less of a factor for $x > 0.6$. However, it is difficult to grow strained-$Si_{1-x}Ge_x$ with $x > 0.6$ on relaxed-Si because the critical thickness is so small. For instance, Goto found a lower mobility for $x = 0.7$ than for $x = 0.5$ [71].

In order to avoid this problem, a relaxed-SiGe substrate can be used. A relaxed-SiGe substrate can be achieved through a linear graded Ge content buffer on a Si substrate as described previously. First tried was strained-Ge on relaxed-$Si_{0.3}Ge_{0.7}$, for which Konig

found a hole mobility of 1300 cm^2/Vs [77]. One issue with growth of such high Ge content films is three-dimensional nucleation, or islanding. In order to avoid this, the growth temperature must be lowered with increasing Ge content during growth of the graded buffer.

In order to reduce difficulties associated with the relaxed $Si_{0.3}Ge_{0.7}$ buffer, strained-SiGe on a relaxed-$Si_{0.7}Ge_{0.3}$ buffer have been explored. Ismail reported a peak mobility of 1050 cm^2/Vs [2] for strained-$Si_{0.2}Ge_{0.8}$ and Arafa found a mobility of 700 cm^2/Vs for strained-$Si_{0.4}Ge_{0.6}$ [78]. For MOSFETs we wish to end up with a pure Si cap since oxidation of Ge results in high interface-state density. The relaxed-$Si_{0.7}Ge_{0.3}$ substrate has the advantage that is compatible with a Si cap. In fact, such a cap would be under tensile strain and exhibit high electron mobility as described below. A device based on this principle will be analyzed later.

As for high-field transport, Bufler calculated a 50% decrease in v_{sat} for strained-$Si_{0.3}Ge_{0.7}$ on relaxed-$Si_{0.7}Ge_{0.3}$ compared to bulk Si [61].

Strained-Si

Strained-Si grown on relaxed $Si_{1-x}Ge_x$ is very interesting from a device point of view because it exhibits enhanced electron mobility while also retaining a high-quality Si/SiO_2 interface. It serves as a potential well for electrons and thus can be used for both buried and surface channel NMOS devices. However, the hole surface mobility is not enhanced all that much above bulk, and it cannot be used to confine holes in a buried channel.

First, let us examine buried-channel electron transport. Fig. 2.11 shows the electron mobility in buried-channel strained-Si structures on relaxed-$Si_{1-x}Ge_x$ as a function of substrate Ge content x comparing several theoretical calculations to various experimental results. The theoretical results are only compared to buried channel because they do not include surface scattering.

The enhancement of mobility in strained-Si relative to unstrained-Si is due to the reduction of the carrier conductivity effective mass and the intervalley scattering rates as a result

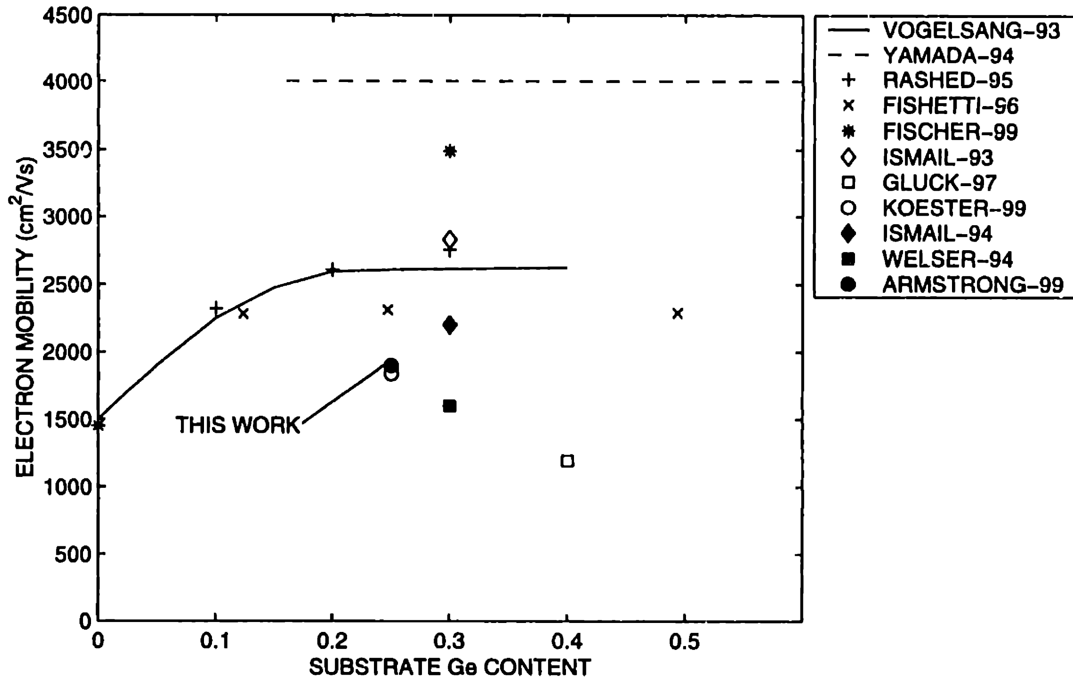


Fig. 2.11. Electron mobility in buried-channel strained-Si structures on relaxed-$Si_{1-x}Ge_x$ vs. substrate Ge content comparing several theoretical calculations [80][15][33][58][81] to various experimental results [8][83][1][22][84] including this work. Lines and +, x and * marks are theory while filled symbols represent measured MOS-gated results and open symbols represent measured Schottky-gated results.

of a split of the 6-fold degenerate Si conduction band minimum into 2-fold and 4-fold degenerate valleys as shown in Fig. 2.10(b). In the case of strained-Si on $Si_{1-x}Ge_x$ we have tensile strain, so the Δ_2-minimum is lowered with respect to the Δ_4-minimum and is preferentially occupied.

The conductivity effective mass is lowered in the following way. The conductivity effective mass m* for transport parallel to the interface in bulk Si is determined from the contributions of all six valleys, and is given by:

$$\frac{1}{m^*} = \frac{1}{3}\left(\frac{1}{m_l} + \frac{2}{m_t}\right), \tag{2.13}$$

where m_l is the longitudinal mass and m_t is the transverse mass of any one valley. In tensile-strained-Si, only the transverse masses of the Δ_2-minima contribute for transport perpendicular to the growth direction. This is called in-plane transport and is parallel to the plane of the four Δ_4-minima in Fig. 2.10 (which do not contribute because they are too

high in energy to be occupied). Thus, the effective mass is given in this case by:

$$\frac{1}{m^*} = \frac{1}{m_t}. \tag{2.14}$$

Note that for $x < 0.5$, the values of m_l and m_t do not change significantly from their bulk Si values. Thus, from m^* considerations alone we expect an increase in mobility by a factor of 1.36 in strained-Si over bulk Si assuming that the Δ_4 valleys are not occupied and the momentum relaxation time is the same between the two materials. This is actually the case as the temperature is dropped to 77K and below, where phonon scattering is reduced [82].

At room temperature, phonon scattering is important and the mobility in strained-Si is increased even further over bulk Si due to a reduction in intervalley scattering between the Δ_2 and Δ_4-minima [81]. This increases the mobility through an increase in momentum relaxation time τ. The greater the Ge content in the substrate, the more separation there is between the Δ_2 and Δ_4-minima and the greater the increase in mobility as shown in Fig. 2.11. However, the mobility enhancement in strained-Si saturates above $x = 0.2$. In general, experimental results are fairly close to the theory. Electron mobility at room temperature in strained Si has been measured to be as high as 2830 cm^2/Vs [1].

Now we turn to surface-channel electron and hole transport. Fig. 2.12 shows the peak surface electron and hole mobility in surface-channel strained-Si NMOS and PMOS devices as a function of substrate Ge content. There are not as many experimental data available for such structures, and very few theoretical analyses have been done which include surface scattering [86]. NMOS results show that the enhancement over bulk still exists even with surface scattering effects.

PMOS results have been very low considering the high values calculated by Fischetti [58] and also by Fischer [81] for holes in strained-Si on relaxed-$Si_{1-x}Ge_x$ without surface scattering included. They calculated 1650 and 1760 cm^2/Vs, respectively.

High-field transport in strained-Si has been explored by several groups in theory [87][15] and experimentally [46]. In general, the value for v_{sat} is found to be the same as for bulk Si, though Rim [46] inferred an increased energy relaxation time by comparing device

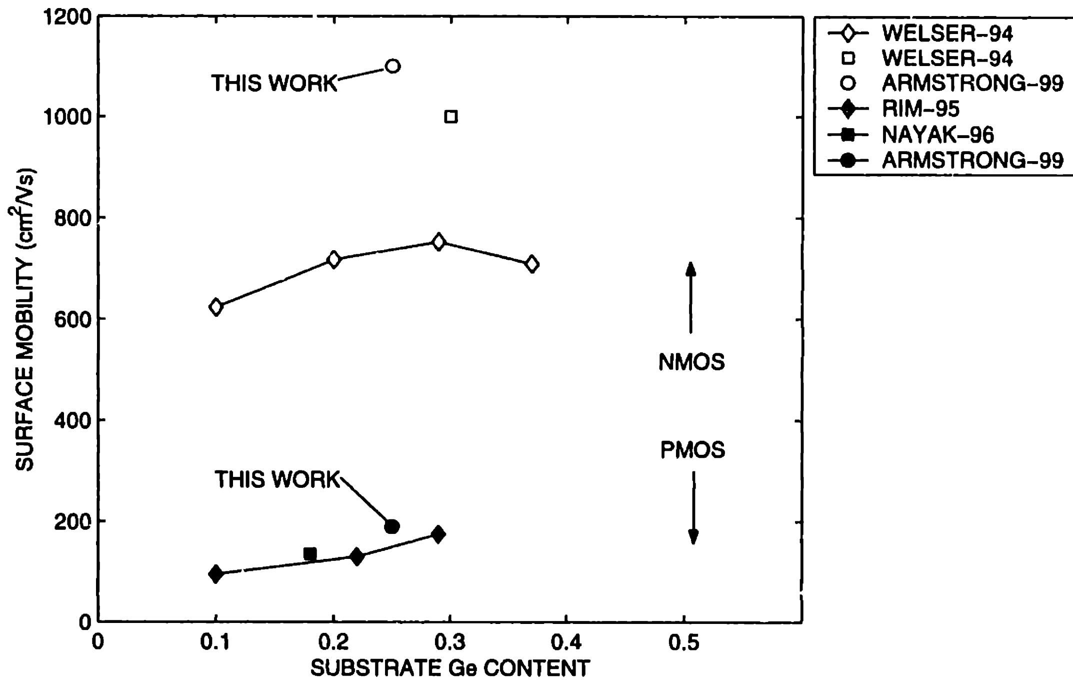


Fig. 2.12. Peak electron and hole mobility in surface-channel strained-Si NMOS and PMOS devices on relaxed-$Si_{1-x}Ge_x$ vs. substrate Ge content comparing various experimental results [82][8][85][27] and this work.

results to simulation. This could lead to increased velocity overshoot effects and increased current drive in short-channel devices.

The room-temperature, undoped, in-plane mobilities of various strained materials are summarized in Table 2.3. These represent the highest experimental values achieved to date in these materials when experimental data is available. Otherwise, experimental data from similar structures are quoted, or, when no data is available, theoretical values are quoted.

TABLE 2.3
PROPERTIES OF STRAINED MATERIALS

	$Si_{0.3}Ge_{0.7}$ on Si	$Si_{0.2}Ge_{0.8}$ on $Si_{0.7}Ge_{0.3}$	Si on $Si_{0.7}Ge_{0.3}$
Electron Mobility (cm^2/Vs)	350[a]	350[b]	2830[c]/1100[d]
Hole Mobility (cm^2/Vs)	630[e]	1050[f]	1760[g]/190[h]

a. Theory [58].
b. Theory [58].
c. Measured, buried-channel [1].
d. Measured, surface-channel, this work.
e. Measured, buried-channel [68].
f. Measured, buried-channel [2].
g. Theory, buried-channel [81].
h. Measured, surface-channel, this work.

Chapter 3

Si/SiGe Device Structures

3.1 Introduction

Based on the band offset and mobility information of the previous chapter we are ready to explore devices which make use of Si and SiGe strained-layers for their high mobility properties. In our examples, we will use the strained-layer combinations of Table 2.1 for which we have band offset and mobility information. All the basic types of devices use layers similar to those.

We are primarily interested in the VLSI application of strained-layers, so we will investigate MOSFETs (metal-oxide-semiconductor field-effect transistor) due to the low gate leakage. First, I will look at PMOS structures, second, NMOS structures and last, some structures suitable CMOS applications.

3.2 PMOS Structures

Fig. 3.1 shows the layer structure of several types of PMOS devices based on SiGe technology. Early work focused on (a) due to the ease of fabrication, as it uses a plain Si substrate [4][5][6]. The device is similar to a bulk silicon MOSFET with the addition of a

strained-$Si_{0.7}Ge_{0.3}$ buried hole channel beneath a Si cap layer. The hole channel is typically thicker than 50 Å to completely contain the hole wave function and less than 100 Å to be less than the critical thickness. Not all work has used x = 0.3, but this is a typical value. However, the upper limit on x is about 0.5 due to critical thickness limitations [71]. The Si cap is necessary to avoid oxidation of SiGe which results in poor interface quality. The Si cap is typically between 50 and 70 Å. The impact of the cap thickness is discussed below.

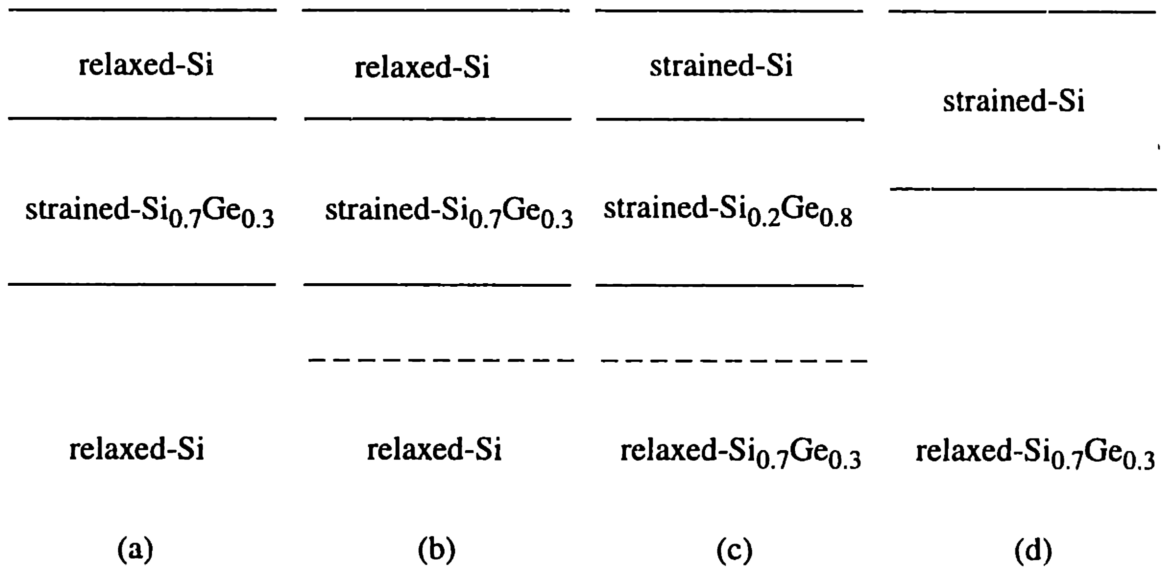


Fig. 3.1. Layer structure of SiGe-based PMOS devices. – marks indicate acceptors. The gate oxide would be placed immediately on top of these structures. PMOS devices have been fabricated by various groups based on structures (a) [4][5][6][68], (b) [9] and (d) [27][85]. Structure (c) has not been demonstrated in this form, though a similar structure with a relaxed-$Si_{0.7}Ge_{0.3}$ cap in place of the strained-Si cap has been fabricated with a Schottky gate [2].

At low gate voltage, and hence low vertical field and carrier concentration, holes are confined in the strained-$Si_{0.7}Ge_{0.3}$ due to the raised valence band and should enjoy a high mobility for several reasons. First, the compressive strain in the pseudomorphic SiGe layer changes the band structure in the layer in a favorable way so as to increase the mobility. Second, the carriers in the layer are separated from the oxide-Si interface and thus the effect of surface scattering is reduced. Third, modulation doping can be employed during

growth of the layer, leaving the hole channel undoped and eliminating impurity scattering. That is, the well doping, necessary for control of short-channel effects, can be ended at a point just below the hole channel. However, as shown in Fig. 2.9, mobilities for (a) have been disappointing, less than 300 cm^2/Vs, though recent results have indicated mobilities as high as 630 cm^2/Vs are possible [68].

At high gate voltage and vertical field, the hole concentration in the buried channel becomes constant and holes populate the Si cap layer, as shown in Fig. 3.2. Holes in the Si cap have about the same mobility as in a bulk inversion layer. Thus, the overall mobility of the device is reduced as more carriers travel in the cap. The population of the Si cap can be reduced by using a thinner Si cap layer. However, with a thin Si cap layer surface roughness scattering will reduce the buried channel mobility placing a lower limit on the Si cap thickness.

There are other ways to reduce occupation of the Si cap. Some groups have proposed grading the Ge content in the buried channel to counteract the vertical field and encourage occupation of the bottom of the buried channel [67]. Other groups have used a boron or gallium-doped acceptor-layer to reduce the vertical field near the surface [9], as shown in Fig. 3.1(b). There is a drawback to these methods, however: the further away the holes travel from the oxide interface, the lower the effective gate capacitance is, and the lower the transconductance. For buried channel devices it is convenient to define an effective oxide thickness $t_{ox,eff}$ which takes into account the intervening cap layer thickness t_{cap}:

$$t_{ox,eff} = t_{ox} + t_{cap}\frac{\varepsilon_{ox}}{\varepsilon_{cap}}, \qquad (3.1)$$

where ε_{ox} and ε_{cap} are the permittivities of oxide and the cap layer, respectively.

Another issue with acceptor-layers in logic applications is the inability to turn off the channel adequately. If the acceptor-layer has too much dose or is located too far away from the channel it will not deplete fully and there will be a parallel path which will be a source of static power dissipation. In other words, the off-current of the device will be unacceptably high. The acceptor-layer can be located either above or below the channel,

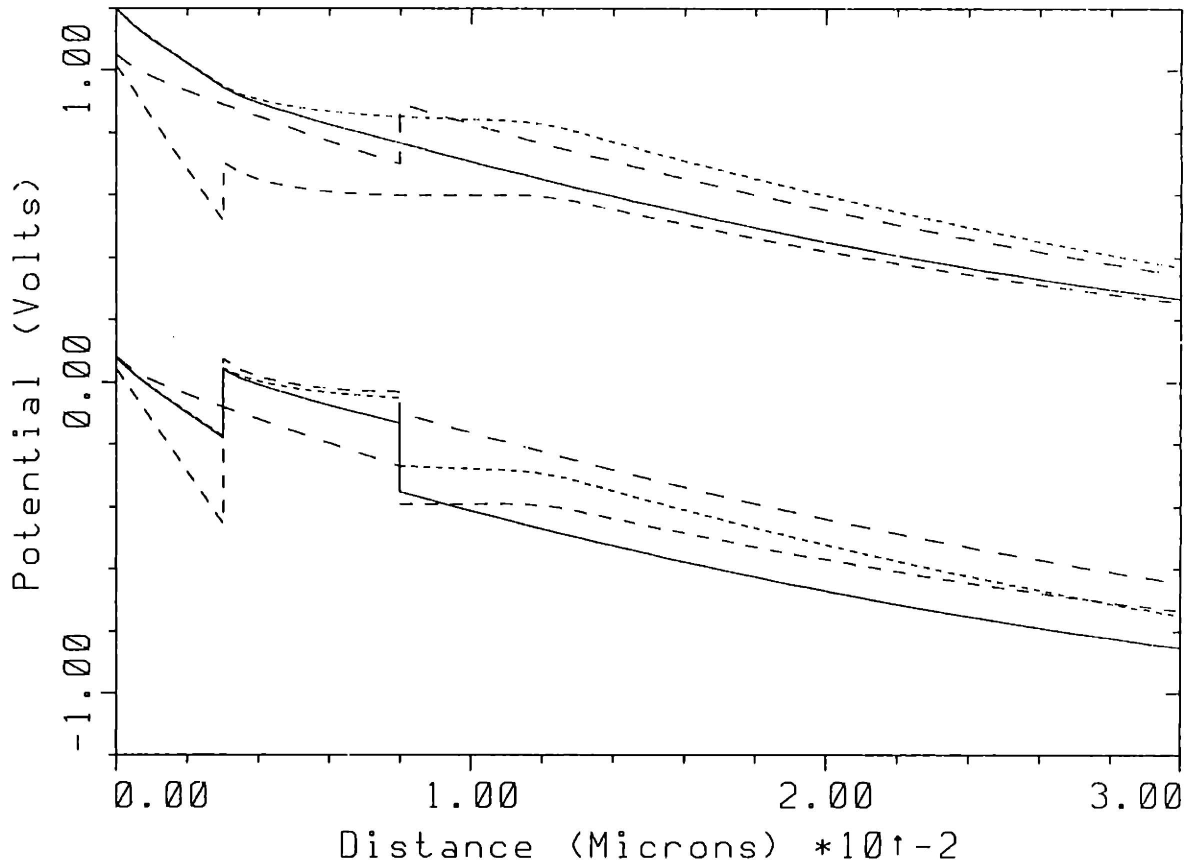


Fig. 3.2. Band structures of the four PMOS layer structures of Fig. 3.1: (a) solid line, (b) small dash, (c) medium dash and (d) long dash. Fermi level is at zero potential. Diagram is for following conditions: t_{ox} = 25 Å, $N_D = 5 \times 10^{17}$ cm^{-3}, V_{GS} = 1.5 V and p^+ poly gate workfunction. For (a), (b) and (c): t_{cap} = 30 Å and $t_{channel}$ = 50 Å. Additionally, for (b) and (c): t_{spacer} = 40 Å and $N_{A,acceptor\text{-}layer} = 2 \times 10^{12}$ cm^{-2}. For (d) there is no cap and $t_{channel}$ = 80 Å. Note how most carriers end up in the cap no matter what layer structure is used.

or both, but location above the channel leads to increased effective oxide thickness so we prefer below.

The low hole mobilities observed in low Ge content layers has led some groups to use a higher Ge content layer on a relaxed-SiGe substrate, typically in combination with an acceptor-layer as shown in Fig. 3.1(c). So far, only Schottky-gated structures have been fabricated with relaxed-SiGe caps on top of the buried channel. However, this technology could potentially be combined with a Si cap to yield a MOSFET device. A record mobility of 2830 cm^2/Vs was obtained in strained-$Si_{0.2}Ge_{0.8}$ on relaxed-$Si_{0.7}Ge_{0.3}$ [2]. Pure Ge on

relaxed-$Si_{0.5}Ge_{0.5}$ and relaxed-$Si_{0.3}Ge_{0.7}$ has also been tried with the aim of reducing the amount of alloy scattering, with modest results [77][95].

More recently, several groups have investigated surface-channel strained-Si PMOS devices [27][85], shown in Fig. 3.1(d), hoping to take advantage of predictions of high hole mobility [58][81] in combination with the benefits of surface-channel transport, namely that the effective oxide thickness is equal to the actual oxide thickness. Strained-Si thickness is typically greater than 100 Å. Results have not lived up to expectations, as shown in Fig. 2.12, with mobilities similar to bulk Si inversion layers.

3.3 NMOS Structures

Fig. 3.3 shows the layer structure of several types of NMOS devices based on strained-Si. Operation of these structures is analogous to the PMOS structures of the previous section. Structure (a), for example, is for the most part the inverse of Fig. 3.1(a), featuring a strained-Si electron channel on a relaxed-$Si_{0.7}Ge_{0.3}$ buffer beneath a relaxed-$Si_{0.7}Ge_{0.3}$ cap. The strained-Si channel and relaxed-SiGe cap are typically 50 Å to 100 Å thick. However, in this case, an additional thin strained-Si cap is necessary to avoid oxidation of SiGe. Only about 10 Å is needed, but this cap can affect device operation greatly [45].

Electrons preferentially populate the strained-Si due to its lower conduction band energy. Similar to PMOS buried-channel structures, the electrons have high mobility in the buried channel layer due to reduced surface roughness scattering, in addition to the enhancement due to strain. As shown in Fig. 2.11, devices based on structure (a) have peak mobilities around 2000 cm^2/Vs at low perpendicular fields, though mobility as high as 2830 cm^2/Vs has been observed in strained-Si.

At higher fields, mobility is degraded as the Si/SiGe cap fills with electrons which has low mobility due to alloy scattering, as shown in Fig. 3.4. The thin Si cap is too thin to completely contain the electron wave function and does not contribute to enhanced electron mobility. One way to reduce the cap occupation, aside from using a thinner cap, is with a donor layer, as shown in Fig. 3.3(b). As mentioned in the previous section, it is advanta-

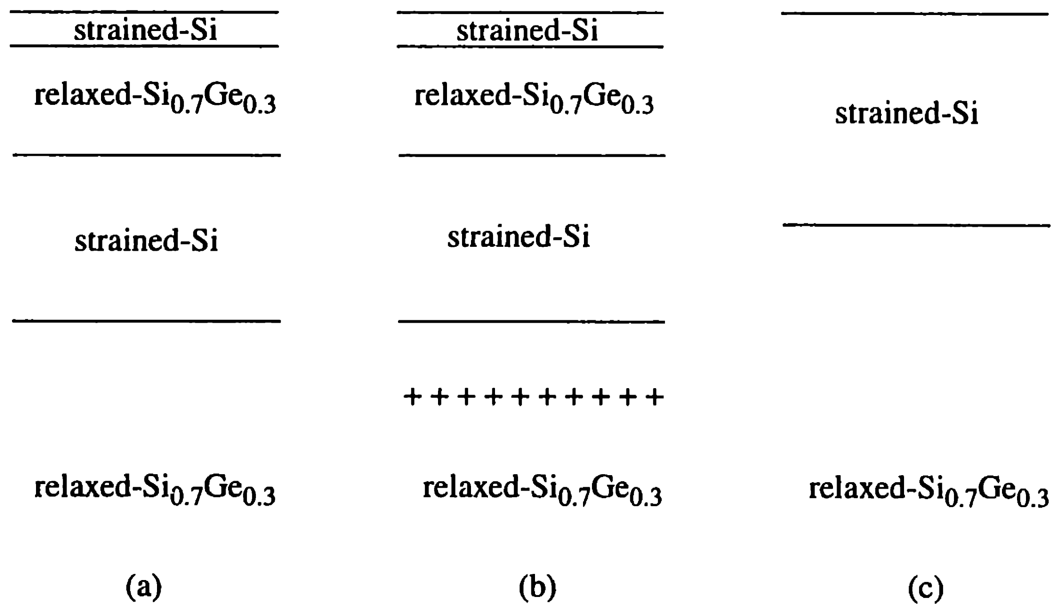


Fig. 3.3. Layer structure of SiGe-based NMOS devices. + marks indicate donors. The gate oxide would be placed immediately on top of these structures. NMOS devices have been fabricated by various groups based on structure (a) [8], (b) [22][29] and (c) [46]. Structure (b) is also favored for Schottky-gate devices [84][83]. MOS devices based on structures (b) and (c) are investigated in this work.

geous to place the donor layer below the channel. This leads to technical complications, though, due to the long tail-off of phosphorous doping during CVD growth. This can lead to mobility degradation due to unwanted dopants in the buried channel. To combat this problem, a growth interrupt technique may be used, in which the wafer is removed from the chamber after donor layer growth and the chamber is purged, and then growth is continued [22]. Another issue involved with donor layers is mobility degradation due to Coulomb scattering from the donor layer charges themselves. This thesis further examines this issue in buried-channel strained-Si NMOS devices.

More recently there has been interest in the surface-channel strained-Si NMOS structure of Fig. 3.3(c). Peak mobilities in this structure are typically around 1000 cm^2/Vs as shown in Fig. 2.12. This structure eliminates the need for a thin Si cap layer which is difficult to maintain during processing. Though the peak mobility at low perpendicular fields may not be as high as in the buried channel case due to surface roughness scattering, the high field mobility is higher, which is more important for VLSI applications which operate near

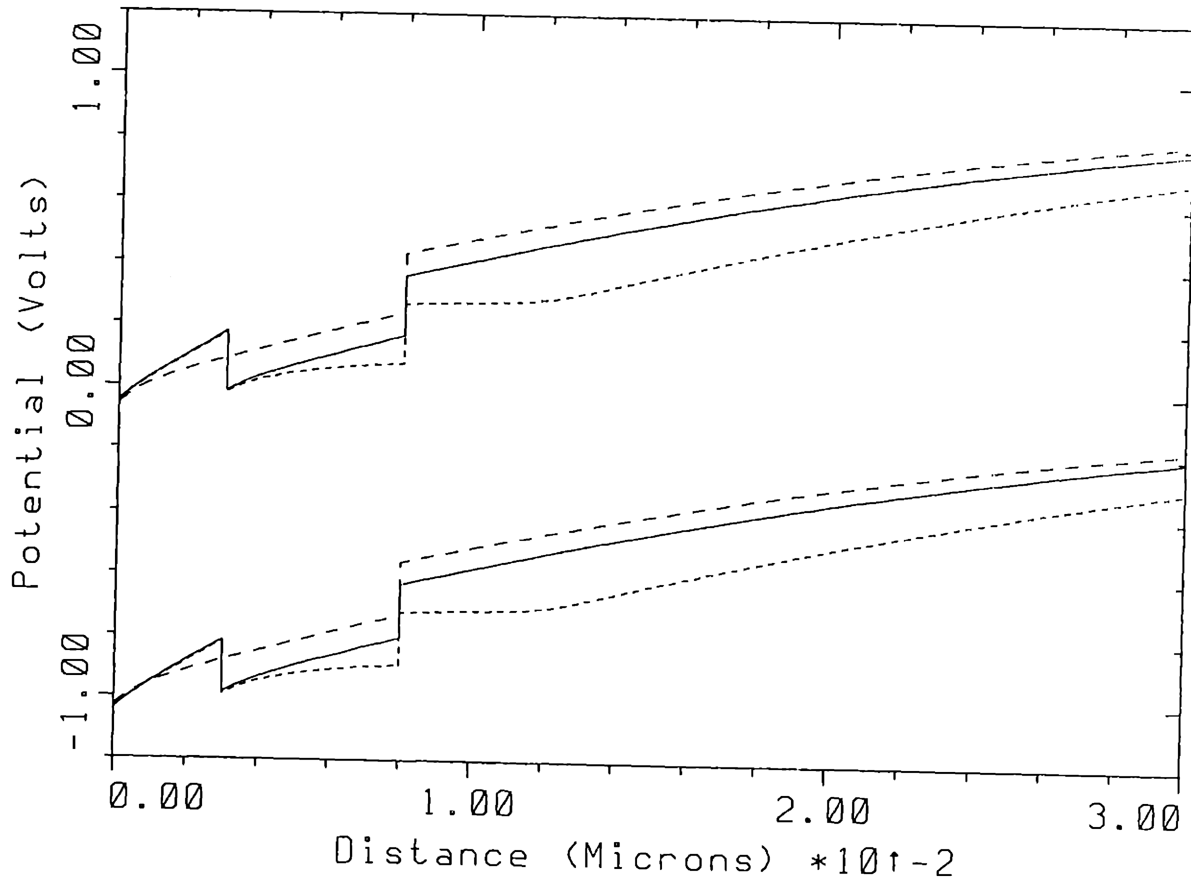


Fig. 3.4. Band structures of the three NMOS layer structures of Fig. 3.3: (a) solid line, (b) small dash and (c) long dash. Fermi level is at zero potential. Diagram is for following conditions: $t_{ox} = 25$ Å, $N_A = 5 \times 10^{17}$ cm^{-3}, $V_{GS} = 1.5$ V and n^+ poly gate workfunction. For (a) and (b) the thin Si cap is ignored and $t_{cap} = 30$ Å and $t_{channel} = 50$ Å. Additionally, for (b): $t_{spacer} = 40$ Å and $N_{D,donor\text{-}layer} = 2 \times 10^{12}$ cm^{-2}. For (c) there is no cap and $t_{channel} = 80$ Å. Note how most carriers end up in the cap no matter what layer structure is used.

breakdown fields [8]. The effect of well implants on surface-channel strained-Si NMOS is further investigated in this thesis.

3.4 CMOS Structures

We would like to use the high mobility devices presented in the previous sections to enhance the performance of CMOS. We want to achieve high hole and electron mobility in the same layer structure, so as to avoid the need to etch down to expose the appropriate layers resulting in a non-planar surface. However, the ideal layer structures for NMOS and PMOS are not necessarily the same. Welser has discussed the possibility of combining

surface channel NMOS with buried channel PMOS or the inverse in the same layer structure [7]. However, in his design, the electron channel is strained-Si while the hole channel is relaxed-$Si_{0.7}Ge_{0.3}$. This will result in low hole mobility due to alloy scattering.

Fig. 3.5 compares several layer structures suitable for SiGe heterostructure-based CMOS devices. The most ambitious is (a), which combines buried electron and hole channels in the same structure with a donor layer to enhance occupation of the bottom channel. This leads to high peak mobility but low mobility as high perpendicular field. Another drawback of structure (a) is its sheer complexity. In contrast, (b) uses a simple strained-Si surface-channel for both NMOS and PMOS. This design increases electron mobility at all perpendicular fields, but the hole mobility is not significantly enhanced over bulk. Structure (c) is proposed in this thesis as a way to combine the high hole mobility available in strained-$Si_{0.2}Ge_{0.8}$ with surface-channel strained-Si NMOS without the complexity of structure (a).

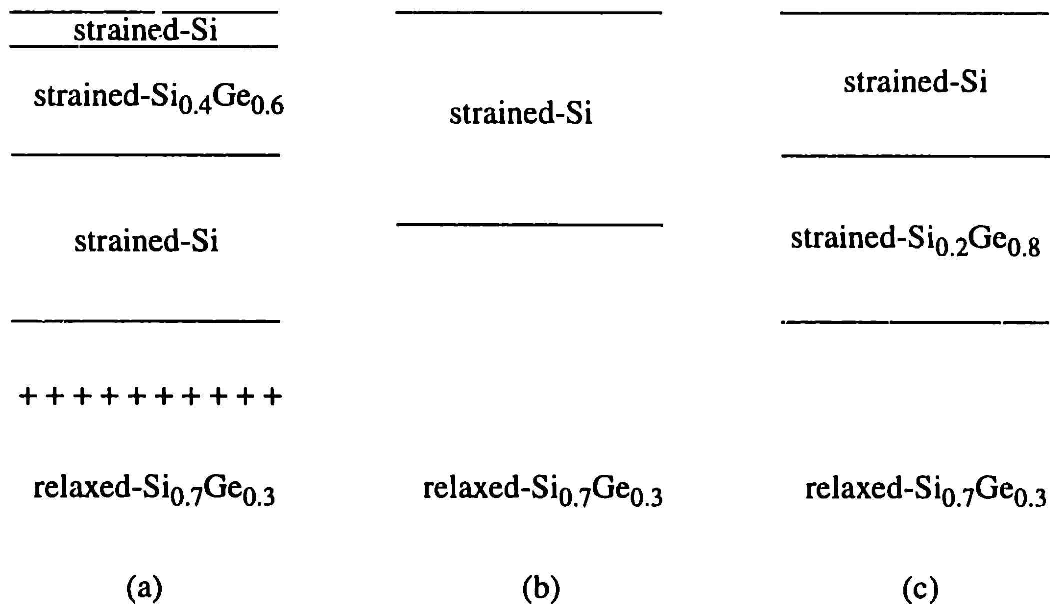


Fig. 3.5. Layer structure of SiGe-based CMOS devices. + marks indicate donors. The gate oxide would be placed immediately on top of these structures.

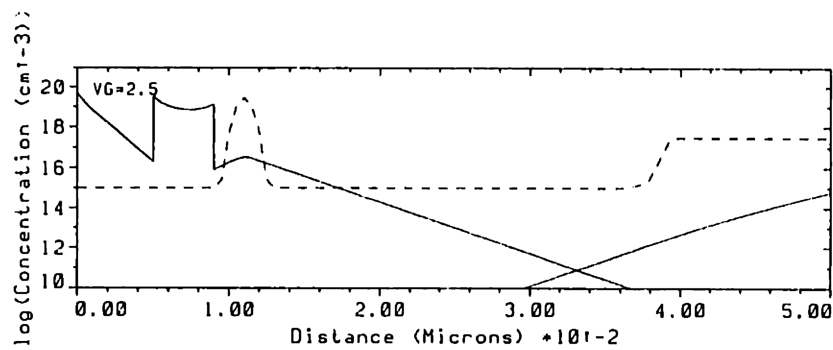
(a) Buried-Channel CMOS

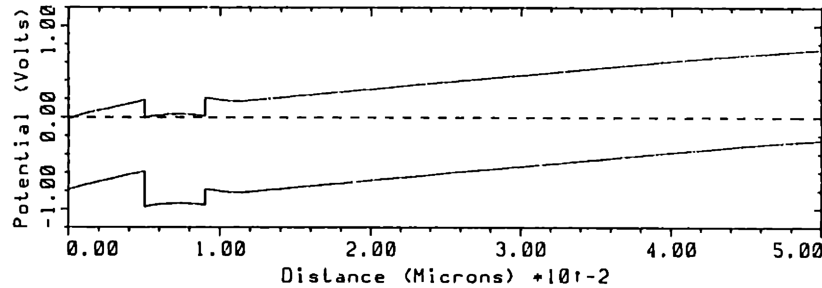
Fig. 3.5(a) was first proposed in [30] where it was called HCMOS for heterostructure-based CMOS. The individual devices polarities are referred to as HNMOS and HPMOS. The performance leverage of the structure over bulk Si CMOS for submicron devices and circuits is investigated in the following chapter.

The proposed design provides for both a compressively-strained SiGe hole channel and a tensely-strained Si electron channel in a single structure. The layers are grown upon a low-defect-density, relaxed-SiGe buffer achieved using a graded-buffer technique. The p-well is *in-situ* doped during growth of the relaxed buffer, while the n-well is created by ion implantation prior to growth of the channel layers. An undoped spacer is grown above the well doping in order to reduce the body effect. An n-type donor-layer is used below the n-channel to bend the energy bands so as to avoid inversion of the low-mobility silicon surface channel. The strained Si electron channel is separated from the donor-layer by an undoped setback to minimize ionized impurity scattering. The Si cap layer allows for growth of a high quality gate oxide. An *in-situ* doped p^+ poly-Si gate is used for both device polarities. The devices can be isolated by shallow trench isolation.

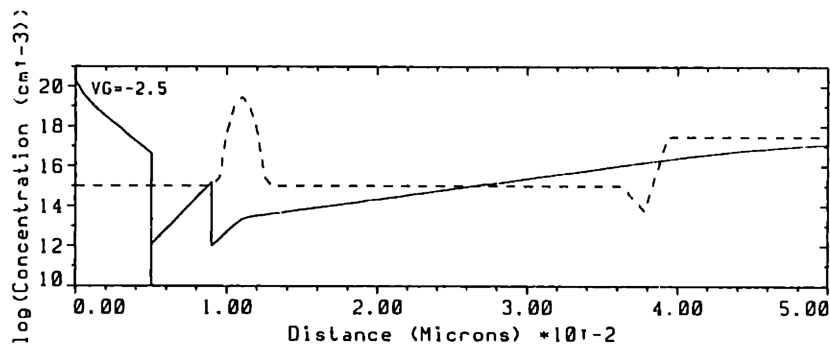
Fig. 3.6 shows the electron, hole and doping concentration and the band structure in inversion for NMOS and PMOS devices based on HCMOS. Note how the NMOS device is degraded by occupation of the Si/SiGe cap, even though a donor layer is used. The PMOS device also may suffer from mobility degradation at high fields due to surface roughness scattering because the Si cap is so thin.

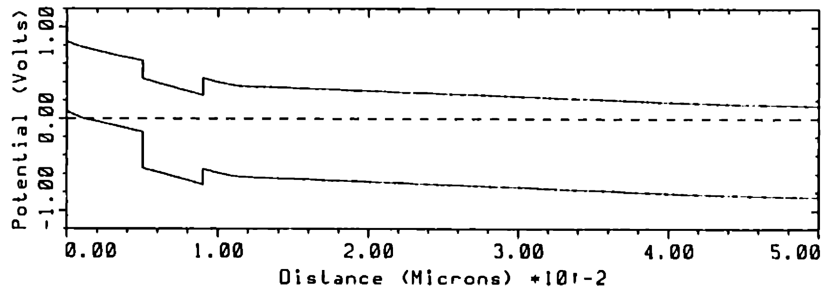
Fig. 3.7 shows the electron and hole density in the NMOS and PMOS device, respectively, vs. gate bias. The hole density is higher than the electron density for a given gate bias due to the proximity of the p-channel to the oxide interface. This partially compensates for the lower hole mobility, and, consequently, leads to more symmetric current drive.





(a)





(b)

Fig. 3.6. Electron, hole and doping concentration and band diagram in inversion for (a) NMOS and (b) PMOS device based Fig. 3.5(a). Dashed line is net doping or the fermi level. Layer thicknesses are in Å: $t_{ox} = 50$, $t_{cap} = 10$, $t_{p\text{-channel}} = 40$, $t_{n\text{-channel}} = 40$. The thin Si cap is modeled as an extension of the p-channel due to quantum-effect raising of the minimum energy level. The donor layer dose is 2.6×10^{12} cm^{-2} located 20 Å below the n-channel. Gate is p^+ poly and well doping is 3×10^{17} cm^{-3} starting 300 Å below the n-channel for both NMOS and PMOS.

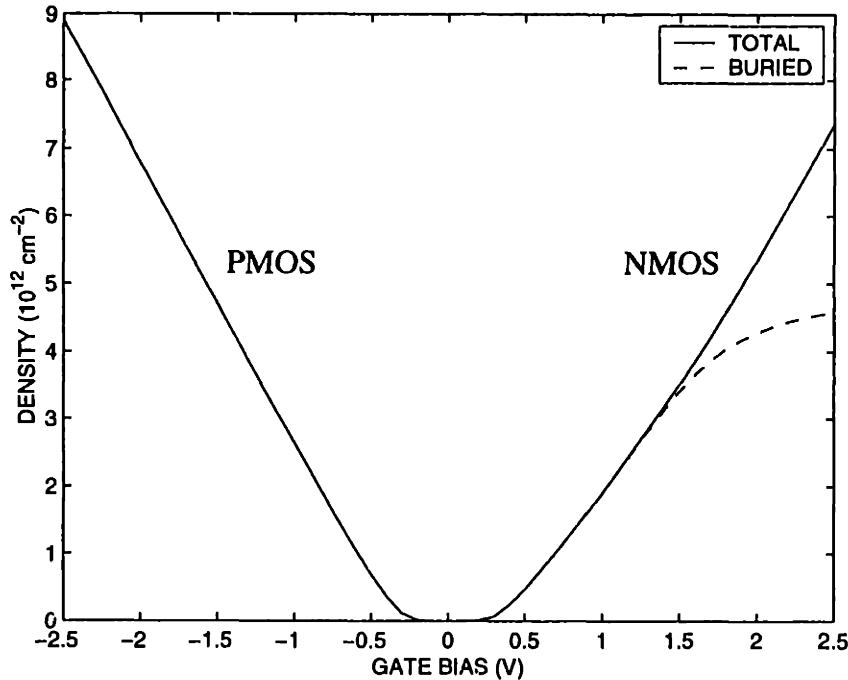


Fig. 3.7. Electron and hole concentration vs. gate bias in HCMOS structure showing symmetric threshold voltages for NMOS and PMOS. Parameters are as in Fig. 3.6. Dashed line indicated electrons confined in buried strained-Si layer.

(b) Surface-Channel CMOS

The surface-channel strained-Si CMOS of Fig. 3.5(b) is a much simpler alternative for SiGe-based CMOS. The band structure for PMOS and NMOS devices of this type were shown in Fig. 3.2 and Fig. 3.4, respectively. The only drawbacks of this structure are the low hole mobility, as shown in Fig. 2.12, and the need for a relaxed-SiGe substrate.

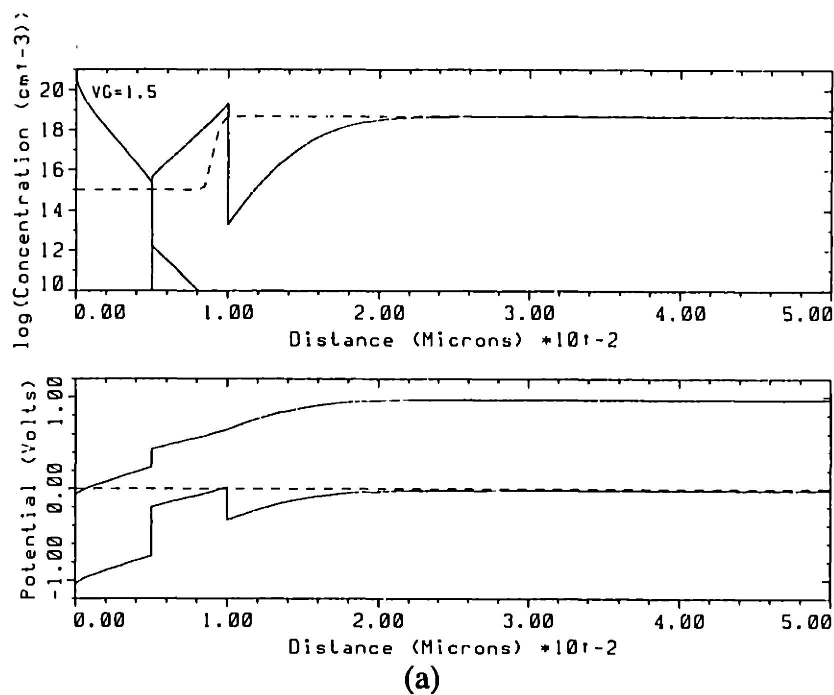
(c) Surface/Buried-Channel CMOS

Fig. 3.5(c) is a way of achieving high hole mobility in combination with the high electron mobility in surface strained-Si without unneeded complexity. The structure features a buried strained-$Si_{0.2}Ge_{0.8}$ layer to enhance hole mobility at low perpendicular fields.

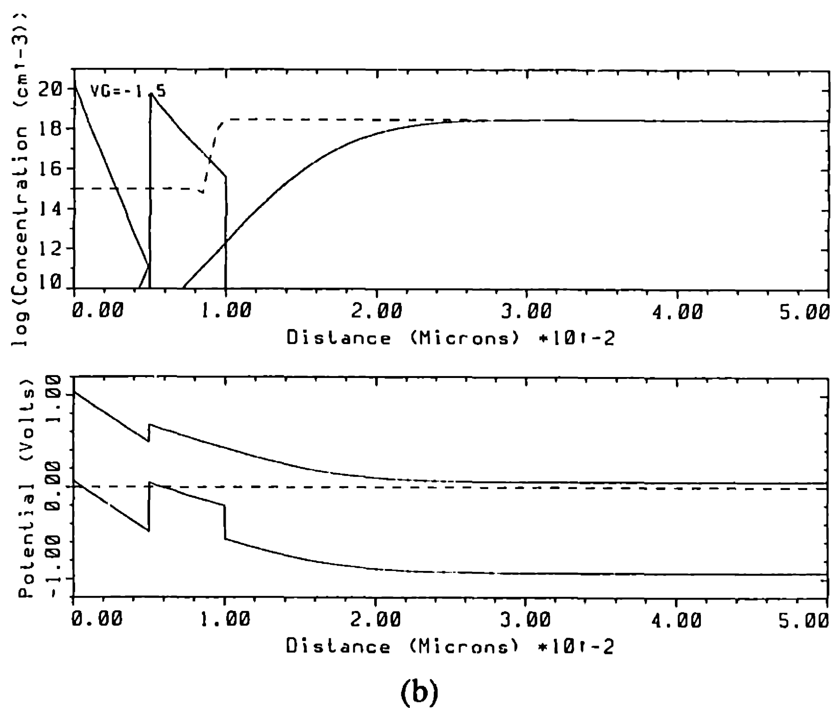
Fig. 3.8 shows the carrier profiles and band structure of NMOS and PMOS devices based on Fig. 3.5(c) in inversion and Fig. 3.9 shows the carrier density vs. gate bias. The NMOS device acts as a surface-channel strained-Si device, unaffected by the strained-$Si_{0.2}Ge_{0.8}$ layer. The PMOS device acts as a buried-channel device until high fields invert the

strained-Si surface layer. However, this is not extremely detrimental since the hole mobility here is the same as in bulk Si. Note that even though the buried hole channel will have an increased effective oxide thickness (by 16 Å for a 50 Å strained-Si cap), the gain in mobility should outweigh this drawback.

This layer structure also shows promise for use with midgap gate materials. Fig. 3.9(b) shows the carrier density vs. gate bias for NMOS and PMOS devices with a midgap gate. The smaller effective bandgap, by which I mean the separation between the lowest conduction band point and highest valence band point, leads to a reduction in the spread between NMOS and PMOS threshold voltages from around 1.1 V to 0.6 V. This solves the problem of one or both threshold voltages being too high when using midgap gate workfunction material on bulk Si.

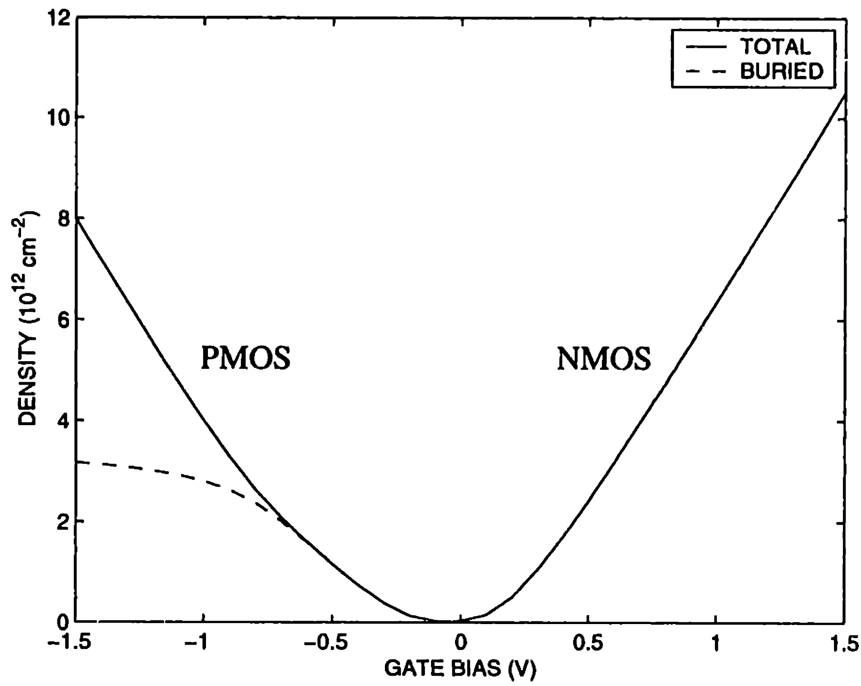


(a)

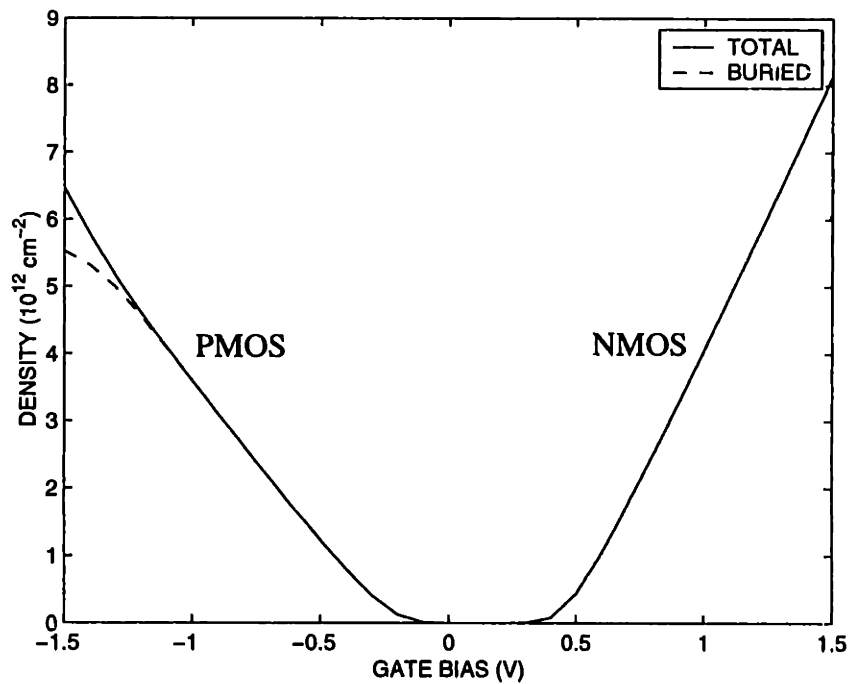


(b)

Fig. 3.8. Carrier profiles and band structure in inversion of (a) NMOS and (b) PMOS devices based on Fig. 3.5(c). Dashed line is net doping or the fermi level. The strained-Si and strained-$Si_{0.2}Ge_{0.8}$ layers are both 50 Å thick. Oxide thickness is 25 Å. Gate is p^+ poly for PMOS and n^+ poly for NMOS. Well doping is 5×10^{18} cm^{-3} starting 100 Å below the surface.



(a)



(b)

Fig. 3.9. Electron and hole concentration vs. gate bias in NMOS and PMOS devices based on Fig. 3.5(c) comparing (a) p^+ poly on PMOS and n^+ poly on NMOS and (b) midgap gate workfunction. Well doping for (a) is 5×10^{18} cm^{-3} starting 100 Å below the surface and for (b) is 5×10^{17} cm^{-3} starting 300 Å below the surface. Other parameters are as in Fig. 3.8. Dashed line indicates holes confined in buried strained-$Si_{0.2}Ge_{0.8}$ layer.

Chapter 4

Device and Circuit Performance

4.1 Introduction

This chapter explores the potential performance of high-mobility Si/SiGe MOSFETs in comparison to bulk silicon MOSFETs for sub-micron CMOS applications. Device and circuit performance is explored both through hand analysis and detailed numerical simulation. The numerical simulations focus on the HCMOS structure presented in the previous chapter.

4.2 Device Performance Analysis

This section uses analytical models to explore the dependence of MOSFET device performance on mobility in both the long-channel and short-channel limits.

Velocity Saturation

Fig. 4.1 shows the velocity-field characteristics of carriers in silicon. At low electric fields, the drift velocity v of carriers in silicon is related to the parallel electric field E through the low-field mobility μ_0:

$$v = \mu_0 E. \tag{4.1}$$

At higher fields, the velocity deviates from this relationship. A first order model for this behavior is:

$$v = v_{sat} \frac{E/E_C}{[1 + (E/E_C)^{\beta}]^{1/\beta}}, \tag{4.2}$$

where v_{sat} is the saturated velocity and β is a fitting parameter. Commonly accepted values for these parameters are $v_{sat} = 10^7$ cm/s for both holes and electrons and $\beta = 1.0$ for holes and 1.3 for electrons. E_C is the critical field, defined by the relationship:

$$E_C = \frac{v_{sat}}{\mu_0}. \tag{4.3}$$

At very high fields, the carrier velocity approaches the saturated velocity v_{sat}.

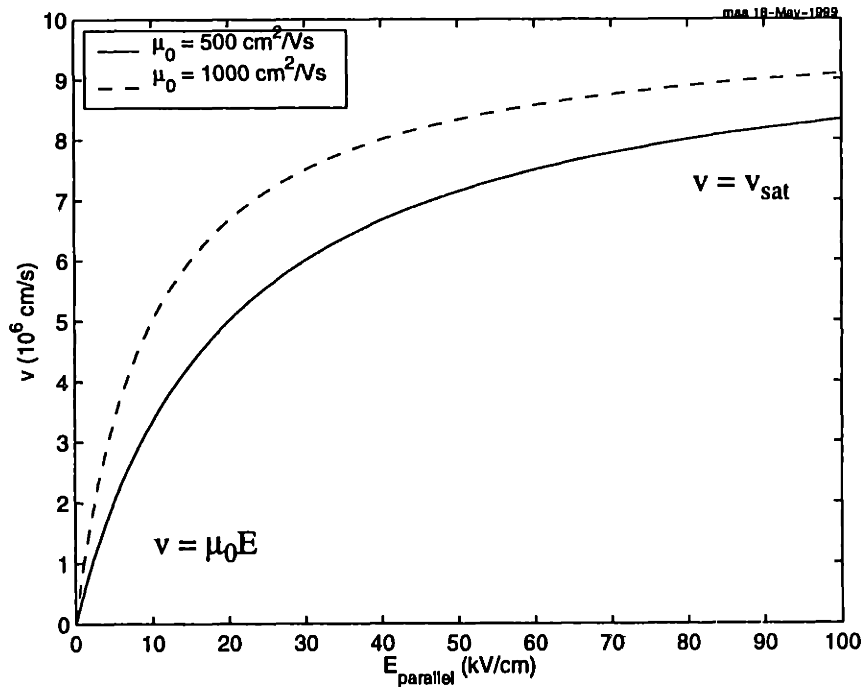


Fig. 4.1. Velocity-field characteristics in silicon comparing two different mobilities plotting using (4.2) with $\beta = 1$.

Physically, at low fields the carrier transport is dominated by Coulomb scattering and acoustical phonon scattering. At higher fields, optical phonon scattering becomes important and results in velocity saturation.

Long-Channel Characteristics

In a long-channel MOSFET the lateral electric fields are low and we are in the regime of (4.1). Assuming a constant mobility μ it is possible to derive a simple model for device operation:

$$I_D = \frac{W}{L}\mu C_{ox}\left[(V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2\right], \qquad V_{DS} \leq V_{DS,sat} \tag{4.4}$$

$$I_{D,sat} = \frac{W}{L}\mu C_{ox}\frac{(V_{GS} - V_T)^2}{2}, \qquad V_{DS} > V_{DS,sat} \tag{4.5}$$

where

$$V_{DS,sat} = V_{GS} - V_T, \tag{4.6}$$

neglecting the body effect. This dependence is shown in Fig. 4.2 for an L = 5 μm MOSFET. Here we see that current is simply proportional to mobility.

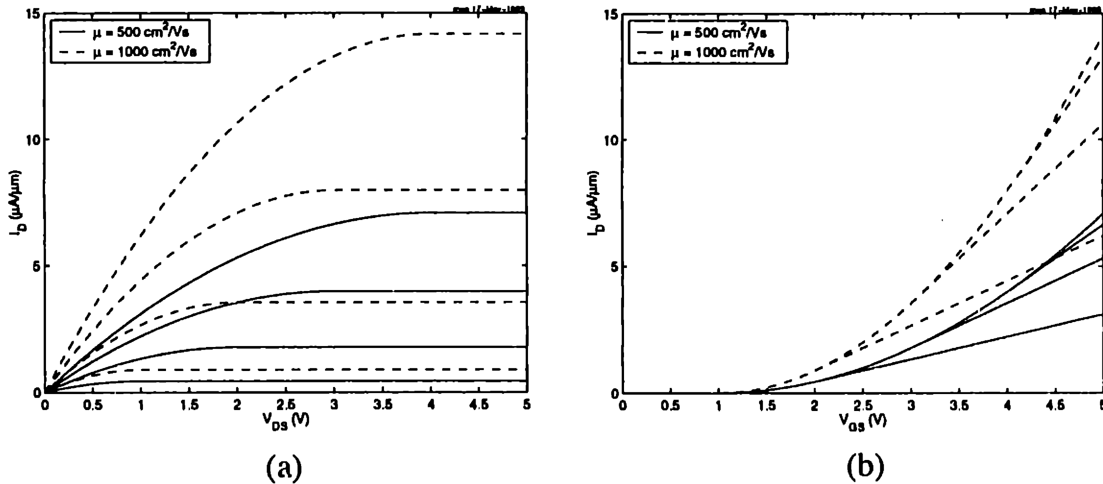


Fig. 4.2. Long-channel device characteristics comparing two mobilities. (a) I_D-V_{DS} characteristics for V_{GS} = 0 to 5 step 1 V. (b) I_D-V_{GS} characteristics for V_{DS} = 0 to 5 step 1 V. Device parameters: L = 5 μm, t_{ox} = 1000 Å, V_T = 1 V.

Short-Channel Characteristics

In contrast, for sub-micron devices several non-linear effects come into play to modify the dependence of current drive on mobility such as velocity saturation, series resistance and charge-sharing. These phenomena can eradicate the benefits of increased mobility seen at long-channel lengths. Let us now examine velocity saturation by itself.

To include the effects of velocity saturation, the basic MOSFET equations can be modified in the following way [49]:

$$I_D = \frac{I_{D,\text{ no velocity saturation}}}{1 + V_{DS}/(LE_C)}, \qquad V_{DS} \le V_{DS,sat} \tag{4.7}$$

$$I_{D,sat} = \left.\frac{I_{D,\text{ no velocity saturation}}}{1 + V_{DS}/(LE_C)}\right|_{V_{DS} = V_{DS,sat}}, \qquad V_{DS} > V_{DS,sat} \tag{4.8}$$

where

$$V_{DS,sat} = LE_C\left[\sqrt{1 + \frac{2(V_{GS} - V_T)}{LE_C}} - 1\right]. \tag{4.9}$$

This assumes $\beta = 1$ in (4.2), which is not accurate for electrons, but should give a rough idea of the behavior.

This situation is shown in Fig. 4.3 for a 0.1 µm MOSFET. A value for $v_{sat} = 6\text{x}10^6$ cm/s is chosen to better reproduce experimental device data with this simple model. Notice that the increase in current drive is reduced in this case compared to the long-channel case. In fact, in the limit of complete velocity saturation, or $LE_C \to 0$, we have in saturation:

$$I_{D,sat} = WC_{ox}(V_{GS} - V_T)v_{sat}, \tag{4.10}$$

where v_{sat} is the saturated carrier velocity. Thus, in the limit of velocity saturation, the device current in saturation becomes independent of mobility (and channel length).

4.3 Device Simulation

The preceding analysis neglects important two-dimensional short-channel effects such as

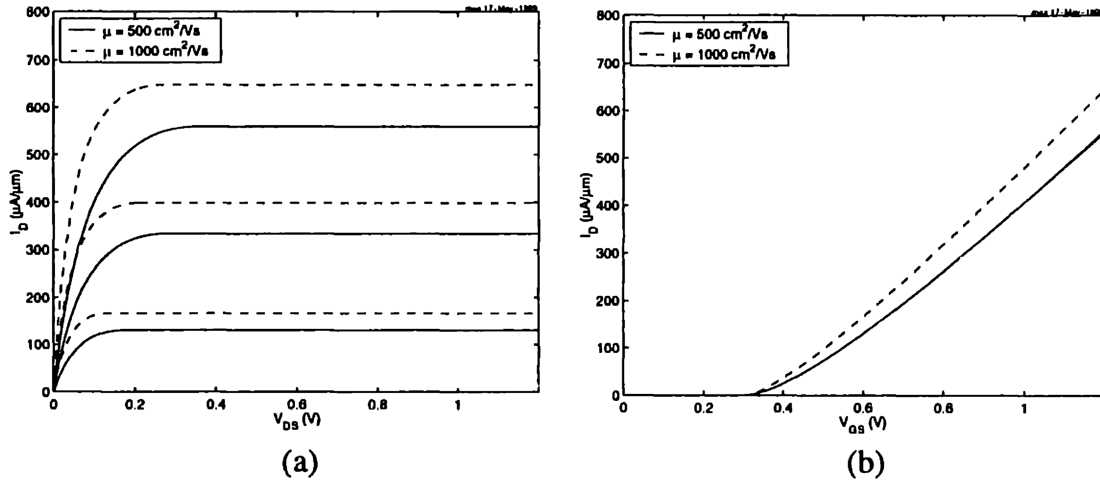


Fig. 4.3. Short-channel device characteristics incorporating velocity saturation comparing two mobilities. (a) I_D-V_{DS} characteristics for $V_{GS} = 0$ to 1.2 step 0.3 V. (b) I_D-V_{GS} characteristics for $V_{DS} = 0$ to 1.2 step 0.3 V. Device parameters: L = 0.1 μm, t_{ox} = 20 Å, V_T = 0.3 V, $v_{sat} = 6x10^6$ cm/s.

charge-sharing. Series resistance also becomes important as the channel resistance becomes a smaller fraction of the overall resistance. At extreme sub-micron channel lengths and high mobilities, carrier energy relaxation becomes important, leading to velocity overshoot which serves to mitigate the effects of velocity saturation.

These issues were addressed through 2-D numerical simulation with Avant! MEDICI. Both Si and Si/SiGe devices were simulated in NMOS and PMOS polarity, using drift-diffusion and hydrodynamic equations. In order to take advantage of the highest electron and hole mobilities available in Si/SiGe, I simulated the HCMOS structure of Fig. 3.5(a).

Simulation Inputs

Inputs to the program include both structural parameters, such as the layer structure and doping, and transport parameters, such as the low-field mobility and the transverse and parallel field dependence of the mobility.

Structural Parameters

Oxide thickness t_{ox} was 50 Å. Junction profiles were Guassian with junction depth x_j of 380 Å for NMOS and 630 Å for PMOS. For bulk devices, substrate doping was a constant 3×10^{17} cm^{-3} for both NMOS and PMOS. For the Si/SiGe devices, the HCMOS structure of Fig. 3.5(a) was used. Layer thicknesses were: t_{cap} = 10 Å, $t_{n\text{-}channel}$ = 40 Å and $t_{p\text{-}channel}$ = 40 Å. The donor layer dose was 2×10^{12} cm^{-2} with an undoped spacer of 20 Å beneath the strained-Si electron-channel. A 300 Å undoped spacer was used above an in-situ doped well of concentration 3×10^{17} cm^{-3} for both HNMOS and HPMOS.

No charge accumulation in the Si cap layer was allowed by specifying that layer to be insulating, rather than semiconducting. This is justified on the basis of solutions to the Schroedinger equation for this structure which show that for a 10 Å cap the lowest energy level is near the top of the well [45]. A more accurate way to treat this would be to artificially raise the conduction band in the cap, but this is a small correction. One could also assume the cap was completely oxidized away right up to the SiGe cap.

The bandgap of relaxed-$Si_{0.3}Ge_{0.7}$ was 1.03 eV. In the strained-Si electron channel E_C and E_V were offset by -0.22 and -0.15 eV, respectively, with respect to the relaxed-$Si_{0.3}Ge_{0.7}$. In the strained-$Si_{0.4}Ge_{0.6}$ hole channel E_C and E_V were offset by 0.03 and 0.18 eV, respectively, from the relaxed-$Si_{0.3}Ge_{0.7}$. Permittivity in $Si_{1-x}Ge_x$ was calculated from a linear interpolation between the values in pure Si and pure Ge, where $\varepsilon_{Si} = 11.8\varepsilon_0$ and $\varepsilon_{Ge} = 16\varepsilon_0$.

Transport Parameters

The simulation assumed low-field mobilities of 2500 and 800 cm^2/Vs for SiGe HNMOS (for heterostructure-base NMOS) and HPMOS, respectively, slightly less than the measured results in similar structures [1][2]. The surface mobility of the bulk devices followed a universal-curve dependence on perpendicular field, which was previously calibrated to experimental data from bulk MOSFETs [19][20]. In the case of the SiGe devices, no perpendicular field dependence of mobility was assumed, with the justification that the carri-

ers are confined in a quantum well away from the oxide/silicon interface and the attendant surface roughness scattering [22].

In the drift-diffusion solution, the same model for mobility dependence on parallel field was used for the SiGe devices as for the bulk devices, which was previously calibrated to experimental data from bulk MOSFETs [20]. The same value for the saturated velocity v_{sat} was used for Si and SiGe materials (10^7 cm/s for both electrons and holes). Fig. 4.4 shows the velocity-field characteristics for Si and Si/SiGe NMOS and PMOS. The expression (4.2) is used with β = 1.0 for holes and 1.3 for electrons. These values were found to give a good match between simulated and experimental results for bulk silicon devices.

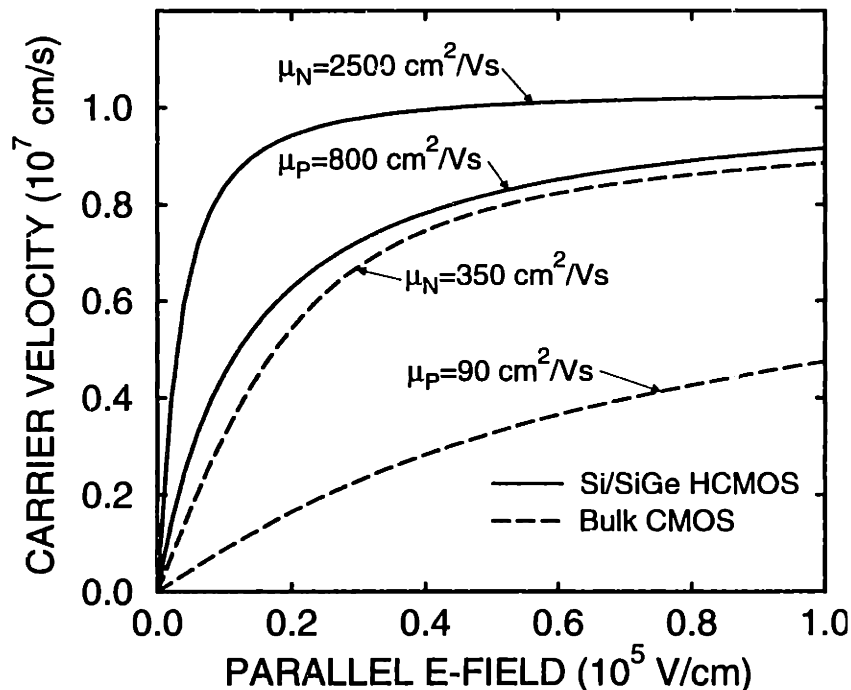


Fig. 4.4. Velocity-field characteristics in Si and Si/SiGe material.

Several things are apparent from this figure. At high fields, electron velocities in both Si (labelled Bulk CMOS) and strained-Si (labelled Si/SiGe HCMOS) approach v_{sat}. However, higher values of low-field electron mobility lead to the earlier onset of velocity saturation with field. For holes, there is much greater improvement in velocity over the entire curve due to the low hole mobility in bulk silicon. Notice that the bulk electron curve

almost matches that of holes in strained-SiGe (labelled Si/SiGe HCMOS). This could lead to balanced operation of CMOS.

For hydrodynamic simulations, an energy-dependent mobility model was used. The results of the hydrodynamic and D-D mobility models were exactly equivalent in the limiting case of uniform parallel electric field. The energy dependence of mobility was determined from the relationship:

$$\frac{3}{2}\frac{kT}{q} = \mu(E)E^2\tau_w, \tag{4.11}$$

where τ_w is the energy relaxation time. The value to use for τ_w is not well established, primarily because there is not much experimental data to go on. It is energy dependent itself, and thus must be chosen to be valid in the range of energies of interest in the simulation. I used the value $\tau_w = 0.1$ ps, although others have used values from 0.1 to 0.4 ps. The longer the energy relaxation time, the more velocity overshoot will occur and the higher the simulated current drive will be. Recent work has shown evidence of an increased τ_w for electrons in strained-Si due to the reduced intervalley scattering rate [46].

Total series resistances $R_T = 2R_S$ of 600 $\Omega\mu m$ for NMOS and 1800 $\Omega\mu m$ for PMOS were assumed. These values are representative of a 0.25 μm bulk technology [19]. For comparison purposes, the same values were used for SiGe and bulk simulations. Effective channel length and series resistance were determined by comparing devices of various lengths. We realize that, for bulk Si devices, series resistances of roughly half these values are achievable, while for SiGe devices, it may be difficult to achieve the above values. However, a reduction in series resistance of the bulk Si devices will not eradicate the performance advantage of SiGe devices at this channel length.

Simulation Results

The I-V characteristics of $L_{eff} = 0.2$ μm HNMOS and HPMOS are shown in Fig. 4.5. They were calculated using a drift-diffusion (DD) model and a hydrodynamic model where carrier mobility is a function of carrier temperature [10]. For comparison, results of bulk Si

devices with the same channel length, oxide thickness and threshold voltage as their Si/SiGe counterparts were also calculated.

The increased mobility in the SiGe devices provides higher current drives than the bulk devices: 23% for the n-FET and 125% for the p-FET at $V_{GS} = 1.5$ V. The increased mobility, and the resulting higher current drive, of the SiGe devices is due to a combination of factors, including: 1) reduced effective masses and intervalley scattering rates resulting from the strained-layer band structure, 2) reduced ionized impurity scattering due to undoped channel layers, and 3) reduced surface scattering (perpendicular field dependence) as mentioned above.

In the Si/SiGe devices, the hydrodynamic calculation indicates a current level that is higher than that calculated using DD. The difference amounts to 25% in the HNMOS and 5% in the HPMOS. This reflects an increase in the average electron velocity in the channel. In contrast, hydrodynamic solutions for bulk Si devices are coincident with DD solutions at this channel length. Velocity overshoot has been deduced from transconductance measurements in Si n-MOSFETs, and was only possible when the channel length was reduced to below 0.1 μm at room temperature [14]. In view of the higher mobility in our suggested structure it is not surprising that velocity overshoot is observed at 0.2 μm. Consistent with our observation is the prediction of Yamada et. al using Monte Carlo calculation [15], that a significant velocity overshoot in a 0.18 μm gate Si/SiGe n-type modulation-doped transistor would take place.

Pinto et. al [16] have investigated the effect of increasing the carrier mobility on velocity overshoot and, consequently, on current and transconductance, and predicted a significant improvement in device performance as the charge carrier mobility is increased. However, an interesting conclusion of that paper is that the device performance, as predicted by the hydrodynamic model, for $\mu = 2500$ will not be much different than for $\mu = 25000$. This is because the $\mu(E)$ relation used in the hydro model become insensitive to mobility as the μ approaches infinity.

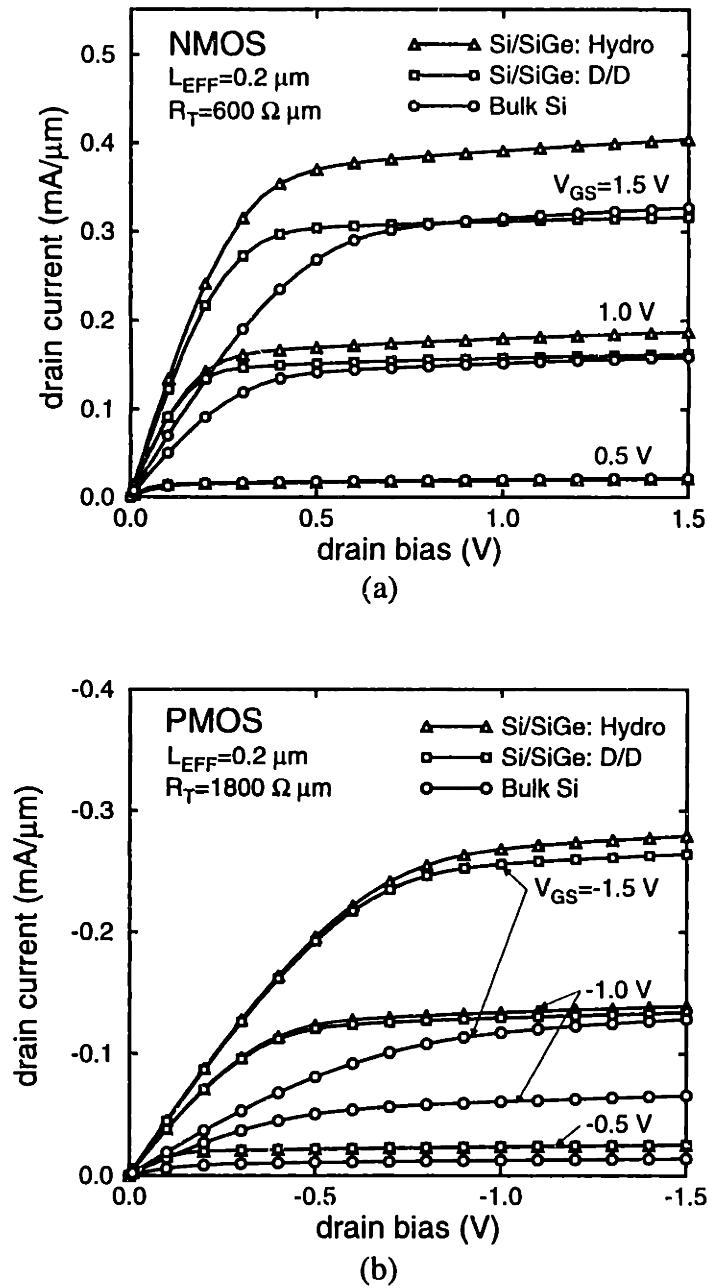


Fig. 4.5. I_D-V_{DS} characteristics for different V_{GS}. (a) HNMOS and (b) HPMOS compared to bulk counterparts. Also shown is the comparison between DD and hydro simulations. DD and hydro give the same results for bulk Si devices.

It is important to note that the current saturates at V_{DS} = 0.4 V (0.8 V) in the HNMOS (HPMOS) at V_{GS} = 1.5 V, despite the inclusion of a series resistance (the corresponding numbers with no series resistance are 0.25 V and 0.5 V, respectively). Also, the saturated current level of the p- and n-FETs in our case differ only by about a factor of 1.4, whereas the difference amounts to about a factor of 2.5 in Si technology. This is attributed mostly to the higher hole mobility in the strained SiGe channel. The proximity of the p-channel to the surface compared to the n-channel and the slightly lower threshold voltage also result in a higher carrier density at a given bias.

The subthreshold behavior is shown in Fig. 4.6. The subthreshold slopes at V_{DS} = 50 mV of the n- and p-FET are 88 and 87 mV/decade, respectively. The devices exhibit reasonable resistance to short channel effects, with DIBL values of 21 mV/V for n-FET and 15 mV/V for the p-FET. DIBL was calculated as the parallel shift in log I_D-V_{GS} curves between V_{DS} = 0.05 and 1.5 V. Junction depth in all devices was 0.55 μm. The off-state drain current in both devices is about 5×10^{-10} A/μm at zero gate bias, which meets the requirement for low standby power consumption. The extrinsic transconductances, g_m, of the n and p-FET are 441 and 288 mS/mm, respectively, at $V_{DS} = V_{GS}$ = 1.5 V. In comparison, the simulated numbers for bulk Si n- and p-FETs of the same design rule and including the same series resistance are 347 and 130 mS/mm, respectively. After correction for series resistance, the intrinsic transconductances, g_{mi}, of the SiGe n- and p-FET are 516 and 404 mS/mm, respectively, as compared to 392 and 152 mS/mm for bulk devices.

A useful figure for comparing between technologies is g_m/WC_{ox}, which has the units of velocity, and represents the average carrier velocity in the channel. In this equation, C_{ox} represents the effective oxide capacitance, which is slightly less than the physical oxide capacitance in the SiGe devices due to the confinement of carriers in subsurface quantum wells. Assuming a metallic gate, the effective oxide thickness is 65 (53) Å for SiGe HNMOS (HPMOS), as compared to 50 Å for the bulk devices, taking into account the relative permittivities of oxide and silicon. At V_{DS} = 1.5 V, the values of g_m/WC_{ox} for the SiGe HNMOS and HPMOS are 8.2×10^6 and 4.4×10^6 cm/s, respectively, as compared to 5.0×10^6 and 1.9×10^6 cm/s for the bulk Si devices. After correction for series resistance,

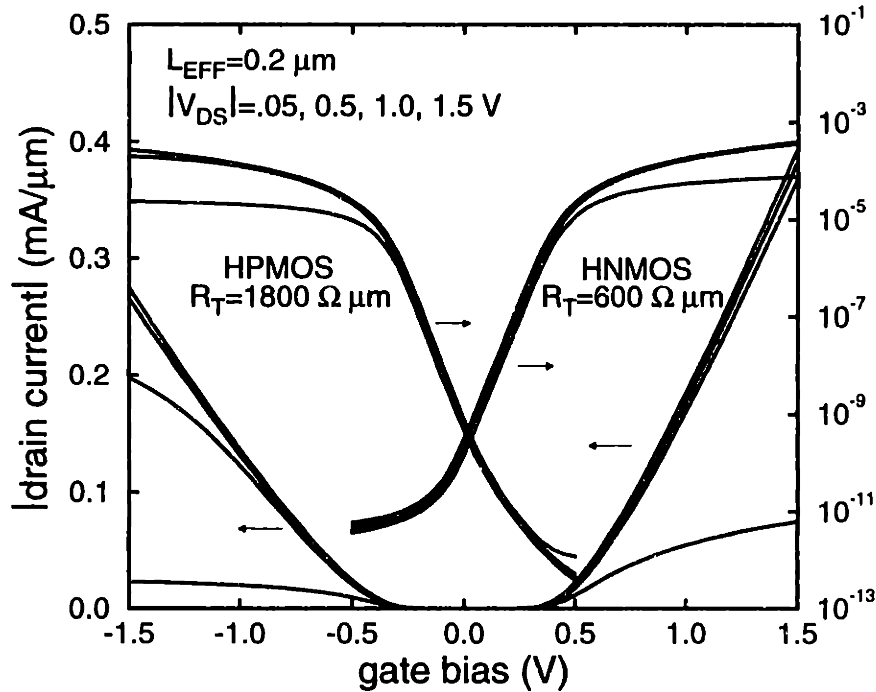


Fig. 4.6. Simulated I_D-V_{GS} characteristics for HNMOS and HPMOS at several different V_{DS} values.

these values increase to 9.6×10^6 and 6.2×10^6 cm/s for the SiGe n- and p-FET, respectively, as compared to 5.7×10^6 and 2.2×10^6 cm/s for the bulk Si devices. The above calculated DC characteristics of heterojunction and bulk MOSFETs are summarized in Table 4.1. The bulk PMOS device has worse DIBL than the HPMOS because the well doping was 1.4×10^{17} cm^{-3} for the bulk PMOS (2.4×10^{17} cm^{-3} for the NMOS) as compared to 3.0×10^{17} cm^{-3} for the HPMOS and HNMOS. Experimental values for other 0.1 to 0.25 μm effective channel length devices [17][18][19] are given for comparison.

It has been suggested by Hu, et. al [20] that there exists a fundamental trade-off between g_m/WC_{ox} and V_T/V_{DS}, or DIBL, for a given source and drain doping profile. This essentially is a trade-off between device performance the amount of short-channel effects. Only by increasing the surface mobility can one jump off this curve to attain higher performance at a given DIBL.

TABLE 4.1
CALCULATED AND EXPERIMENTAL DC PERFORMANCE OF SI AND SI/SIGE MOSFETS

		$t_{ox,eff}$ (nm)	V_T (V)	R_S ($\Omega\,\mu m$)	g_{me} (mS/mm)	g_{mi} (mS/mm)	g_{mi}/WC (10^7 cm/s)	SS^{-1} (mV/dec)	DIBL (mV/V)
Calculated	0.2 μm HNMOS	6.5	0.38	600	441	516	0.96	86	23
	0.2 μm HPMOS	5.3	-0.31	1800	288	404	0.62	87	18
	0.2 μm NMOS	5.0	0.34	600	347	392	0.57	78	24
	0.2 μm PMOS	5.0	-0.24	1800	130	152	0.22	80	21
Experimental	0.1 μm NMOS	4.0	0.25	125	500	550	0.64	125	120
	0.1 μm PMOS	4.0	-0.15	250	300	350	0.38	140	180
	0.15 μm NMOS	5.5	0.50	300	360	380	0.60	86	70
	0.15 μm PMOS	5.5	-0.50	700	200	215	0.28	88	67
	0.25 μm NMOS	7.0	0.40	600	240	280	0.49	89	40
	0.25 μm PMOS	7.0	-0.40	1800	130	150	0.24	88	45

This trade-off is illustrated in Fig. 4.7, which shows g_m/WC_{ox} as a function of V_T/V_{DS} for several state-of-the-art bulk NMOS technologies. The simulated data points for SiGe HNMOS and bulk NMOS devices are plotted on this figure. Lower series resistances of 300 and 900 Ω μm for HNMOS and HPMOS, respectively, were assumed in generating this figure only. The higher electron mobility in the HNMOS results in higher effective carrier velocity than achievable in bulk NMOS. Velocity overshoot is predicted for channel lengths less than 0.15 μm. Note that the HPMOS curve matches the performance of bulk NMOS.

These results show the performance leverage of this particular design and also address the more general issue of the importance of mobility in sub-micron device performance [16], showing how the benefits of increased mobility persist at $L_{eff} = 0.2\ \mu m$ despite the onset of velocity saturation.

4.4 Circuit Performance Analysis

The previous section showed the benefits of high mobility strained layers over bulk in terms of static device characteristics. Circuit performance of Si/SiGe devices has been less

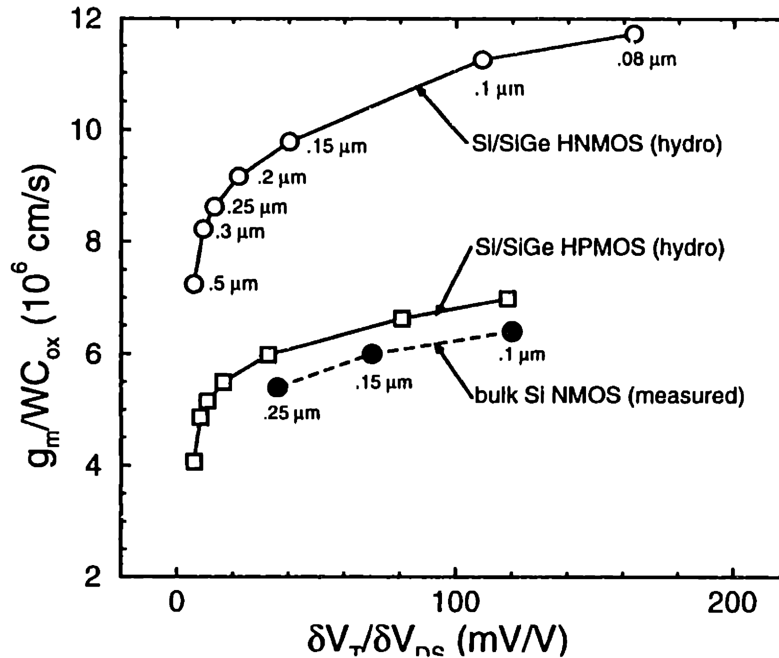


Fig. 4.7. Effective carrier velocity vs. DIBL comparing HNMOS, HPMOS and bulk NMOS. The data points are labelled with the effective channel lengths. The HPMOS points have effective lengths corresponding to the HNMOS points.

thoroughly investigated [30][47][48]. In this section I use analytical equations to explore the potential performance of Si/SiGe-based CMOS circuits.

CMOS Inverter

Let us focus our attention on the CMOS inverter, shown in Fig. 4.8. This could represent one stage in a ring oscillator or a single gate in a chain of logic gates. In this circuit, when the input voltage V_{in} goes low, the PMOS transistor turns on and charges up the load capacitance, and the output node goes to V_{DD}. On the other hand, when V_{in} goes high, the NMOS transistor turns on and discharges the load capacitance, and the output node goes low.

Let us assume for now all the capacitance is represented by the load capacitance C_L and furthermore, let us model the transistors as switched current sources with value $I_{D,sat}$. A single switching event will then be the charging or discharging of the load capacitance by the transistor current source. The time associated with this switch, called the propagation delay t_{pd}, can be written as:

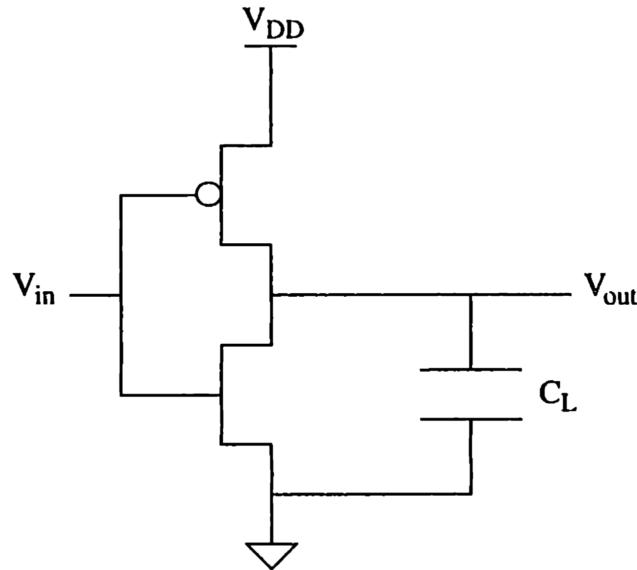


Fig. 4.8. CMOS inverter with power supply V_{DD} and load capacitance C_L. The NMOS pull-down transistor is on bottom and the PMOS pull-up, indicated by the bubble on the gate, is on top.

$$t_{pd} = \frac{C_L V_{DD}}{I_{D,sat}}. \tag{4.12}$$

This is based on the assumption that the device spends the majority of its time in saturation during the switching cycle. There will be a propagation delay associated with the NMOS, t_{pdn}, and also with the PMOS, t_{pdp}. The average stage delay, which I will refer to as simply t_{pd}, will be the average of these two:

$$t_{pd} = \frac{t_{pdn} + t_{pdp}}{2}. \tag{4.13}$$

We are also interested in the power P, which, assuming static and short-circuit power are negligible, can be written as:

$$P = \frac{C_L V_{DD}^2}{t_{pd}}, \tag{4.14}$$

and the power-delay product:

$$P t_{pd} = C_L V_{DD}^2, \tag{4.15}$$

which is a measure of the energy required per switch.

In general, this inverter will be driving subsequent logic stages. In this case, some of the load capacitance will be made up of the gate capacitances of the subsequent stages, some of it will come from the drain junctions of the driving stage and the rest will be parasitic interconnect capacitance. Typically, in dense logic the effective load capacitance will be 1/3 gate, 1/3 junction and 1/3 interconnect capacitance. In order to get a feeling for how this will affect the performance of our inverter, I will examine two limits: (1) pure interconnect capacitance and (2) pure gate capacitance. Within these limits, I analyze the inverter performance for both long-channel and short-channel (velocity saturated) devices.

Interconnect Capacitance Limit

Let us assume that the load capacitance is all interconnect, meaning that it does not depend on device geometry in any way, and is represented by C_L. Let us now examine the performance in the limit of long-channel devices. This means our saturated current is given by (4.5), which I will simplify by assuming $V_T << V_{DD}$:

$$I_{D,sat} = \frac{W}{L}\mu C_{ox}\frac{V_{DD}^2}{2}. \tag{4.16}$$

Plugging this into (4.12) and (4.14) yields for the propagation delay:

$$t_{pd} = \frac{2LC_L}{W\mu C_{ox}V_{DD}}, \tag{4.17}$$

and for the power:

$$P = \frac{W}{L}\mu C_{ox}V_{DD}^3. \tag{4.18}$$

On the other hand, in the limit of extremely short channel devices we have from (4.10):

$$I_{D,sat} = WC_{ox}V_{DD}v_{sat}, \tag{4.19}$$

which results in a propagation delay:

$$t_{pd} = \frac{C_L}{WC_{ox}v_{sat}}, \tag{4.20}$$

and a power:

$$P = WC_{ox}v_{sat}V_{DD}^2. \tag{4.21}$$

The power-delay product in both cases is given by (4.15).

Gate Capacitance Limit

Let us now assume C_L from Fig. 4.8 is due to the gate capacitances of the next stage, not shown in the figure, and that the next stage has the same geometries (L and W) as the driving stage. Then we essentially have:

$$C_L = (W_nL_n + W_pL_p)C_{ox}, \tag{4.22}$$

which I will approximate as $2WLC_{ox}$.

For long-channel devices we have from (4.17):

$$t_{pd} = \frac{4L^2}{\mu V_{DD}} \tag{4.23}$$

and the expression for power (4.18) is unchanged, as C_L does not appear in it.

For short-channel devices we have from (4.20):

$$t_{pd} = \frac{2L}{v_{sat}} \tag{4.24}$$

and again the expression for power (4.21) is unchanged.

The power-delay product in the gate capacitance limit is from (4.15):

$$Pt_{pd} = 2WLC_{ox}V_{DD}^2. \tag{4.25}$$

Mobility Dependence

The results of the above analysis of the CMOS inverter are summarized in Table 4.2. From this table we can come to several conclusions. First, for long-channel, mobility-limited devices, t_{pd} is inversely proportional to μV_{DD} and P is proportional to μV_{DD}^3. Thus, one can use a higher mobility to increase performance in one of two ways:

1. One can reduce power while keeping speed constant. For instance, with twice the

mobility, one can operate at the same speed (t_{pd}) with half the gate drive (V_{DD}) and thus one quarter the power (or power-delay product). This is useful for high-performance, low-power applications such as portable terminals.

2. One can increase speed at the same power. For instance, with twice the mobility, one can operate at $2^{-1/3}$ or 0.8 times V_{DD} for the same power dissipation and $2^{-2/3}$ or 0.6 times t_{pd}. This is useful for high-performance systems with relaxed power requirements such as desktop computers.

In general, in the long-channel limit, power at constant speed goes as μ^{-2}, and speed at constant power goes as $\mu^{-2/3}$. Speed at constant power-delay product goes as μ^{-1} since in this case V_{DD} is constant. This is true in both the gate capacitance limited and interconnect capacitance limited cases.

TABLE 4.2
RESULTS OF ANALYSIS OF CMOS INVERTER

	Gate Capacitance Limit	Interconnect Capacitance Limit
Mobility Limit	$t_{pd} = \frac{4L^2}{\mu V_{DD}}$	$t_{pd} = \frac{2LC_L}{W\mu C_{ox}V_{DD}}$
	$P = \frac{W}{L}\mu C_{ox}V_{DD}^3$	$P = \frac{W}{L}\mu C_{ox}V_{DD}^3$
	$Pt_{pd} = 2WLC_{ox}V_{DD}^2$	$Pt_{pd} = C_L V_{DD}^2$
Velocity Saturation Limit	$t_{pd} = \frac{2L}{v_{sat}}$	$t_{pd} = \frac{C_L}{WC_{ox}v_{sat}}$
	$P = WC_{ox}v_{sat}V_{DD}^2$	$P = WC_{ox}v_{sat}V_{DD}^2$
	$Pt_{pd} = 2WLC_{ox}V_{DD}^2$	$Pt_{pd} = C_L V_{DD}^2$

Turning to the short-channel, velocity saturation limited case, we see that t_{pd} is inversely proportional to v_{sat} and P is proportional to $v_{sat}V_{DD}^2$. Obviously, in this case mobility does not enter into the equations and does not affect circuit performance. Note that, like the long-channel case, this is true regardless of the capacitive loading situation. Realistic CMOS circuits reside somewhere in between the mobility and velocity limited cases, and the true effect of mobility can only be determined through numerical simulation.

Saturation Voltage

One effect mobility can have on velocity-saturated MOSFETs is a lowering of the drain-to-source voltage at the onset of saturation, $V_{DS,sat}$. In the limit of $LE_C \rightarrow 0$, (4.9) reduces to:

$$V_{DS,sat} = \sqrt{2LE_C(V_{GS} - V_T)}. \quad (4.26)$$

Subbing in for $E_C = v_{sat}/\mu$ we find:

$$V_{DS,sat} = \sqrt{\frac{2Lv_{sat}(V_{GS} - V_T)}{\mu}} \quad (4.27)$$

for very short-channel devices.

This expression for $V_{DS,sat}$ tends to give too high values because it was derived from the condition $\partial I_D / \partial V_{DS} = 0$. What we really want to know is when the current has reached about 90% of $I_{D,sat}$. This number will be closer to the rough approximation:

$$V_{DS,sat} = LE_C = \frac{Lv_{sat}}{\mu}, \quad (4.28)$$

though this will tend to give low values as it is derived from extrapolating the initial slope of the I_D-V_{DS} characteristics to $I_{D,sat}$. Thus, we see that $V_{DS,sat}$ of a velocity-saturated device will have a dependence on mobility somewhere between μ^{-1} and $\mu^{-1/2}$. Note also that $V_{DS,sat}$ does not depend on μ in pinchoff-dominated devices.

I will now estimate the effect $V_{DS,sat}$ has on propagation time. Let us model the transistor as a piecewise linear source as shown in Fig. 4.9. We will model the linear regime as having a resistance:

$$R = \frac{V_{DS,sat}}{I_{D,sat}}. \quad (4.29)$$

The switching time will then be the sum of the time when the transistor is in saturation, $t_{pd,sat}$, and the time when the transistor is in the linear regime, $t_{pd,lin}$:

$$t_{pd} = t_{pd,lin} + t_{pd,sat}. \quad (4.30)$$

From (4.12) we find:

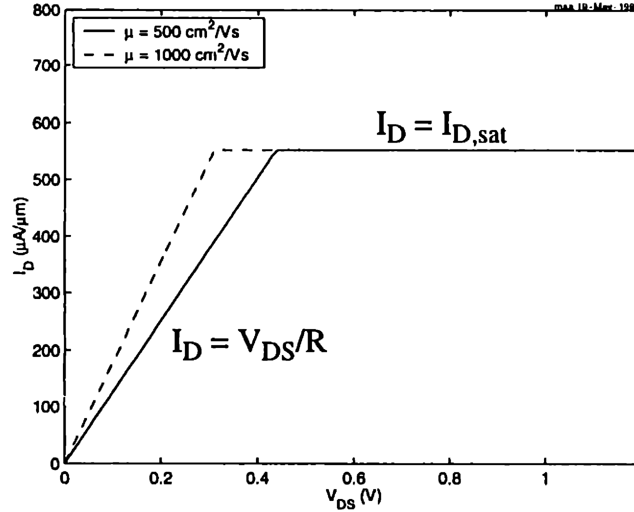


Fig. 4.9. Piecewise linear model for short-channel MOSFET comparing two different mobilities.

$$t_{pd,sat} = \frac{C_L(V_{DD} - V_{DS,sat})}{I_{D,sat}}. \tag{4.31}$$

We can estimate $t_{pd,lin}$ from 3 times the RC time constant:

$$t_{pd,lin} = 3RC_L = 3\frac{V_{DS,sat}}{I_{D,sat}}C_L. \tag{4.32}$$

Thus we have for the total propagation delay:

$$t_{pd} = \frac{C_L}{I_{D,sat}}(V_{DD} + 2V_{DS,sat}). \tag{4.33}$$

Well-designed short-channel MOSFETs operate at around V_{DD}/L = 150 kV/cm and are well described by using $v_{sat} = 6x10^6$ cm/s. Using these numbers with (4.28) and (4.33) I find a 7% decrease in t_{pd} going from $\mu = 500\ cm^2/Vs$ to $1000\ cm^2/Vs$.

The preceding analysis assumes that the node voltage at the output of the transistor must go completely from rail to rail. Since the effects of $V_{D,sat}$ are important only as the node voltage approaches the rail, the actual effects of $V_{D,sat}$ and hence μ on the performance could be reduced. This points out the need for numerical simulation in evaluating the circuit performance of high-mobility devices.

4.5 Circuit Simulation

In order to determine the circuit performance leverage of Si/SiGe HCMOS over bulk Si CMOS, an 11-stage CMOS inverter ring oscillator was simulated using HSPICE (Meta-Software [21]) with both SiGe and bulk device models. The level 28 (modified BSIM) MOSFET models were fitted to simulated (MEDICI) IV data for $L_{eff} = 0.2$ µm, $t_{ox} = 50$ Å devices. The SiGe and bulk devices have the same parameters, including channel length, oxide thickness and threshold voltage. A conservative series resistance of 600 (1800) Ω µm for NMOS (PMOS) was included.

To illustrate the effects of capacitive loading, such as wiring capacitance, both unloaded and moderately loaded ($C_L = 10$ fF) ring oscillators were simulated. In the unloaded case, minimum size devices ($W_p = W_n = 1$ µm) were used, while in the loaded case, a ratio of 2.5:1 W_p:W_n was used. It is important to note that those values have not been fully optimized for the highest speed performance; to take full advantage of high-mobility SiGe CMOS, the circuit must be redesigned. The fan-in and fan-out were both one. Including gate and junction capacitances the total effective load per stage C_{eff} was roughly 5 fF in the unloaded case and 19 fF in the loaded case.

Fig. 4.10 shows the power-delay product as a function of average stage delay comparing Si/SiGe HCMOS to bulk Si CMOS for the unloaded and loaded cases. For our threshold voltage values, dynamic power dissipation dominates over static and short-circuit dissipation. Thus, the power-delay product is approximately $\frac{1}{2}C_{eff}V_{DD}^2$, where C_{eff} is the total effective load capacitance (gate, diffusion and interconnect) and V_{DD} is the supply voltage.

It is well known that there is a fundamental trade-off between power and delay: as V_{DD} drops, the energy per switch decreases as V_{DD}^2, but only at the expense of increased delay and thus poorer performance. Assuming that the device length, width, oxide thickness, series resistance, threshold voltage and interconnect and junction capacitance are all constrained to minimum/optimum values, the only way to increase performance at a fixed

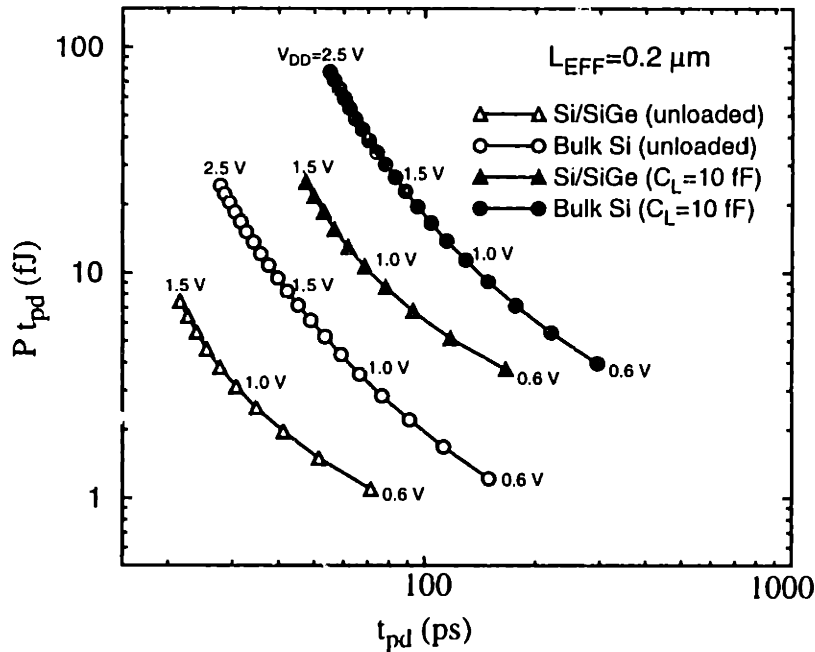


Fig. 4.10. Simulated power-delay-product vs. stage delay for Si/SiGe HCMOS and bulk Si CMOS with V_{DD} as an implicit variable, comparing unloaded and loaded ring oscillators. The corresponding V_{DD} values are indicated on the curves.

supply voltage is to increase the mobility in the n-FET and/or p-FET as in our proposed structure.

Our simulations show that one can run the loaded Si/SiGe circuit at 1.2 V while retaining that same stage delay as the bulk Si circuit running at 2.5 V. This results in a 4.6 times improvement in power-delay product at a stage delay of 55 ps. Similarly, the unloaded Si/SiGe circuit can be run at 1.1 V for the same performance of the bulk Si circuit at 2.5 V, resulting in a 6.4 times improvement in power-delay product at a stage delay of 28 ps. For devices of the same design rules, the stage delay attained by the Si/SiGe circuit at V_{DD} approaching 1.5 V cannot be achieved by bulk Si CMOS within its breakdown voltage limits. It is important to note that the improved performance of the 0.2 μm Si/SiGe HCMOS is not achieved at the expense of increased short channel effects.

Fig. 4.11 illustrates the I_D-V_{DS} locus traced by NMOS and PMOS transistors in a ring oscillator superimposed on the I_D-V_{DS} characteristics of the transistors comparing bulk Si

CMOS and Si/SiGe HCMOS ring oscillators. The performance advantage of Si/SiGe results largely from the increased current drive of the Si/SiGe devices, especially the p-FET. However, since part of the switching cycle occurs in the linear regime, the rapid rise of the drain current with drain voltage, or the low $V_{DS,sat}$, of the high-mobility devices also contributes to the faster switching speed.

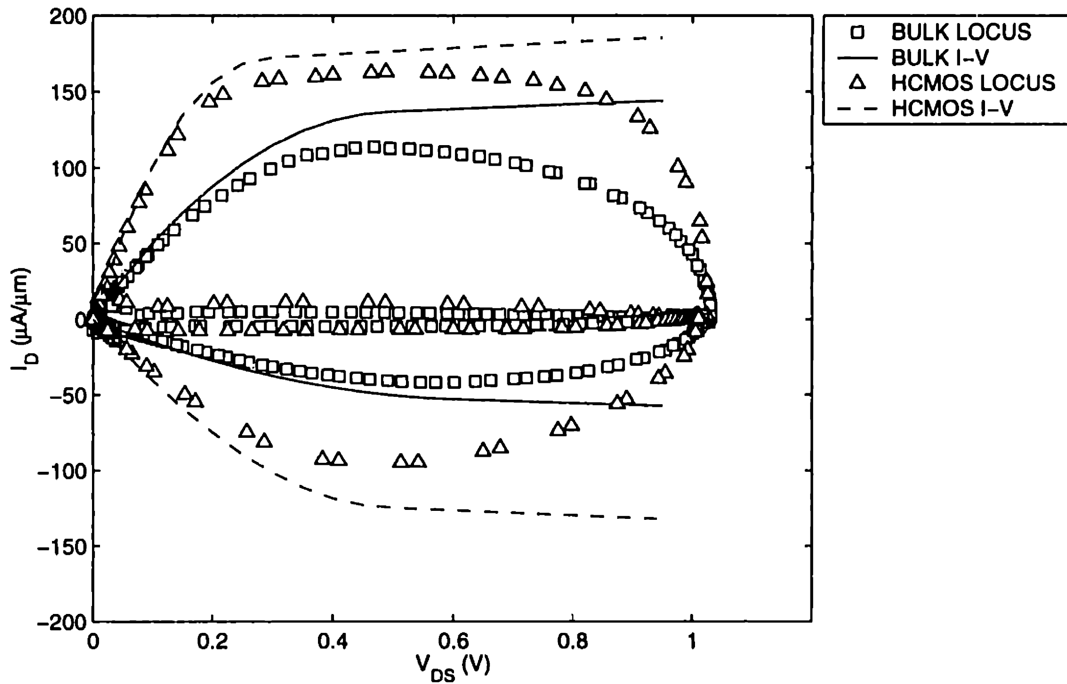


Fig. 4.11. I_D-V_{DS} locus traced by NMOS and PMOS transistors in a ring oscillator (W_n = 1 µm, W_p = 2.5 µm, C_L = 10 fF, V_{DD} = 1 V) superimposed on the I_D-V_{DS} characteristics of those transistors at V_{GS} = 1 V comparing bulk Si CMOS with Si/SiGe HCMOS. NMOS is plotted with positive current while PMOS is plotted with negative current.

Chapter 5

Short-Flow MOSFET

5.1 Introduction

This chapter describes the short-flow MOSFET process used to evaluate strained-Si substrates in my experiments, including the process flow, device layout, theory of operation and measurement tips.

It is not necessary to make short-channel MOSFETs to extract most types of information of interest about strained-Si substrates. For instance, from long-channel MOSFETs one can extract:

- effective mobility vs. effective field,
- capacitance vs. gate voltage, and
- interface state density vs. surface potential.

A thin gate oxide is also not necessary to extract this information, though a thick gate oxide will reduce the usefulness of C-V as a probe to measure the thickness and band offsets of thin Si/SiGe layers near the surface.

On the other hand, effects which only occur in short-channel devices cannot be explored with long-channel MOSFETs, such as:

- velocity saturation and
- degradation due to source/drain series resistance.

A traditional MOSFET fabrication process can take several months to complete. Even when stripped to the essentials, there are four photolithography steps: active area, polysilicon, contact cut and metal, three of which require alignment.

In order to speed up fabrication, I have developed a short-flow MOSFET process [34] which has a single photo step and requires as little as one week to complete in a university lab. In addition, I have made the process compatible with chemical and thermal exposure limitations of SiGe substrates. This process yields long-channel, thick-oxide MOSFETs and thus has the limitations mentioned earlier.

Drift Mobility vs. Hall Mobility

I prefer to make MOSFETs rather than Hall mobility structures to evaluate the mobility of strained-Si substrates even though Hall mobility is relatively quick and easy to measure. This is because the drift mobility differs from the Hall mobility, and it is the drift mobility which really matters for MOSFET performance.

For a sample in which there are many more electrons than holes, the drift mobility can be expressed as

$$\mu = \frac{\sigma}{qn}, \tag{5.1}$$

where σ is the conductivity, n is the electron concentration and q is the electronic charge.

This differs from the Hall mobility,

$$\mu_H = r\frac{\sigma}{qn}, \tag{5.2}$$

which contains the factor r, given by

$$r = \frac{\langle\tau^2\rangle}{\langle\tau\rangle}, \tag{5.3}$$

where the parameter τ is the mean free time between carrier collisions, which depends on

the scattering mechanism and carrier energy.

For nondegenerate semiconductors with spherical constant-energy surfaces, r = 1.18 for phonon scattering and r = 1.93 for ionized impurity scattering [23].

5.2 Process Flow

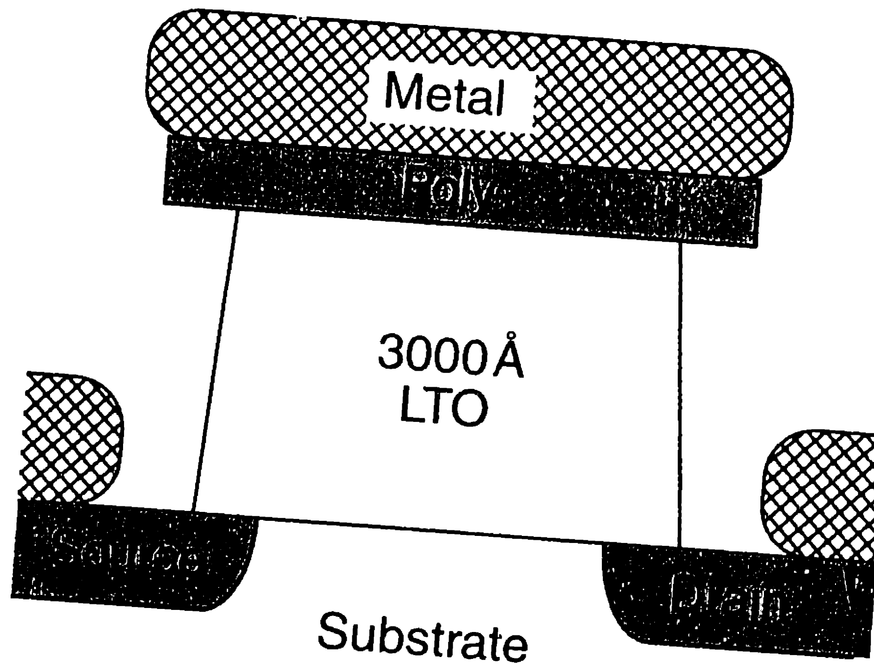
The structure, shown in Fig. 5.1, features a 3000Å gate oxide capped by 500Å polysilicon. The fabrication process is depicted in Fig. 5.2. Early in the fabrication process, this stack is patterned and an isotropic wet etch is used to undercut the poly by 300Å. After source/drain implant and anneal, Ti/Al is evaporated at zero tilt so that the angle of incidence is perpendicular to the wafer, similar to a "lift-off" process. The undercut keeps metal off the oxide sidewalls, ensuring discontinuity between S/D and gate metal as long as the metal height does not exceed the oxide thickness. This trick eliminates the need for contact cut and metal photolithography steps. The use of a ring-transistor layout eliminates the need for an isolation photo step as discussed later in this chapter.

Though easy to make, the structure is not without limitations. Because the gate oxide must be relatively thick to accommodate a reasonable metal height, the process is useful for fabricating long-channel devices only. With t_{ox} = 3000Å, the rule-of-thumb $L_{eff}/t_{ox} \approx 40$ suggests a minimum device length of about 10μm.

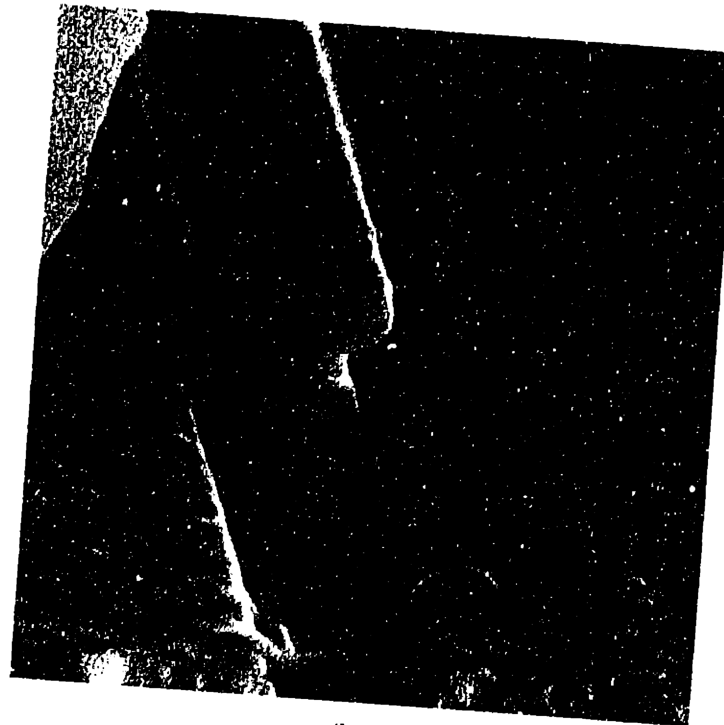
Table 5.1 lists the process steps in more detail. Machine names are nicknames for equipment in the Integrated Circuits Laboratory. Many steps are modified from standard practice to integrate SiGe substrates. This process flow can be modified to explore the effects of various processing issues on mobility, such as adding a well implant before the first step, or adding various thermal treatments after source/drain implant.

Starting Wafers

Starting wafers have a SiGe heterostructure already grown by UHVCVD as described elsewhere in this thesis. At this point, thin SiGe layers are vulnerable near the surface, protected only by a thin Si cap.



(a)



(b)

Fig. 5.1. (1) Schematic and (b) SEM cross section of short-flow MOSFET. Rough material is Ti/Al. The SEM shows a 0.1 μm line, which is much smaller than the 100 μm channel lengths of actual devices.

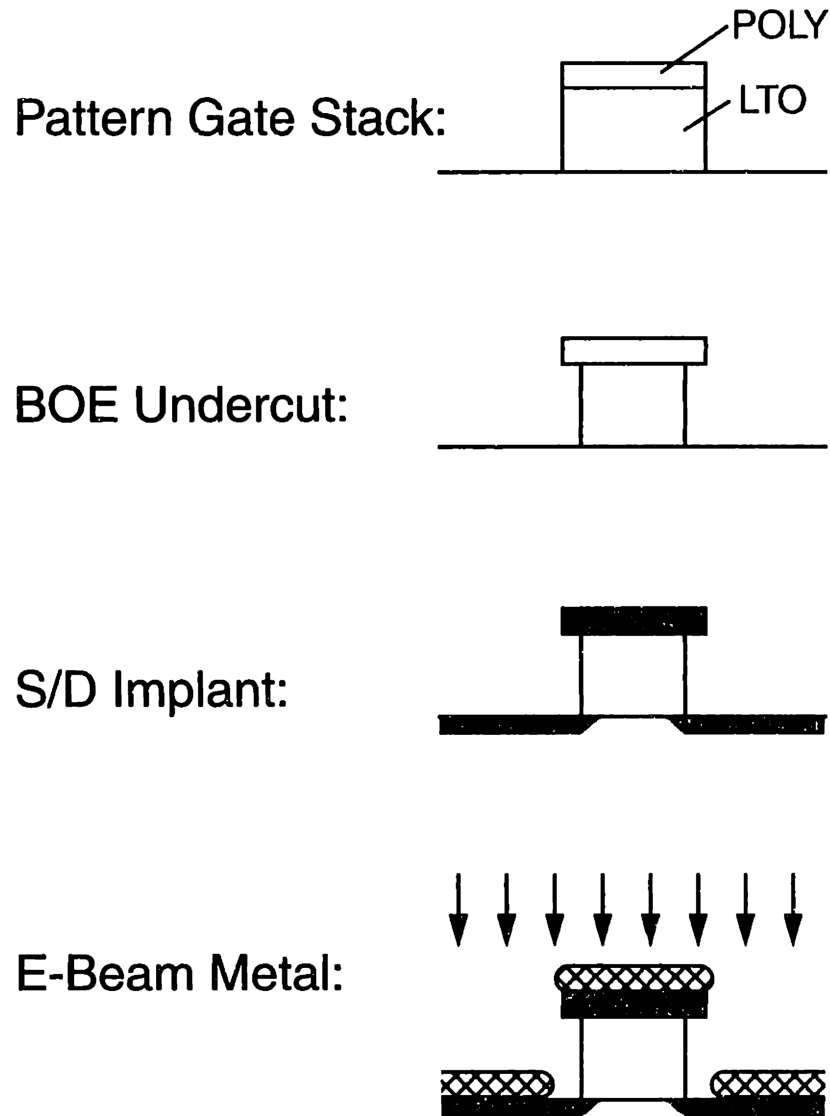


Fig. 5.2. Process flow for short-flow MOSFET

Gate Stack Deposition

Before depositing the gate oxide, the wafers are RCA cleaned. The RCA clean consists of three parts:

1. SC-1 organic clean (5:1:1 $H_2O:H_2O_2:NH_4OH$ at 80°C),
2. native oxide removal (50:1 $HF:H_2O$) and
3. SC-2 ionic clean (6:1:1 $H_2O:H_2O_2:HCl$ at 80°C).

The organic clean solution etches SiGe rapidly and even attacks pure Si to some extent due to the NH_4OH component [35]. Thus, in order to preserve thin surface layers, the organic clean time is reduced to 30 seconds from the standard 10 minutes. The ionic clean

TABLE 5.1
SHORT-FLOW MOSFET PROCESS

Step	Description
Gate Stack Deposition	
rca	Short RCA clean (30 second dip in each bath)
tubeA7	LPCVD 3000Å LTO at 400°C
tubeA6	LPCVD 500Å a-Si at 560°C
Backside Clear	
HMDS	
coater	Coat frontside
developer	Hardbake
AME5000	RIE a-Si from backside
oxide	BOE LTO from backside
asher	
Gate Stack Patterning	
HMDS	
coater	
stepper2	Expose ring-transistor mask
developer	
AME5000	RIE a-Si using CF_4/O_2
AME5000	RIE LTO using CHF_3/O_2, leaving 100Å
asher	
oxide	BOE 4 seconds for 300Å undercut
Source/Drain/Gate Implantation	
vendor	Implant source/drain and gate
pre-metal	Short piranha clean (60 seconds)
tubeB4	Anneal in N_2 at 600°C for 30 minutes
Metallization	
pre-metal	Short piranha (60 sec) + 50:1 HF dip (45 sec)
e-beam	Evaporate 500Å Ti + 1000Å Al, zero tilt
e-beam	Evaporate 5000Å Al onto backside
tubeB8	Sinter in N_2/H_2 at 400°C for 15 minutes

time is also reduced to 30 seconds from the usual 15 minutes. The abbreviated cleaning schedule does not affect the quality of MOS capacitors fabricated on bulk Si wafers.

Next, low-temperature oxide is deposited via LPCVD, followed by amorphous silicon. LTO is used to minimize the thermal budget and results in excellent MOS capacitors as described elsewhere in this thesis. Amorphous silicon is used to achieve efficient dopant activation via solid phase recrystallization.

Backside Clear

The gate stack layers are etched from the backside of the wafers to facilitate good electrical contact to the device substrate.

Gate Stack Patterning

First, resist is patterned with ring-transistor structures. Note this is the first and only photo step. Second, the a-Si/LTO stack is etched anisotropically down to the substrate using a sequence of reactive ion etches to produce a straight sidewall. Third, after resist removal, a short dip in buffered oxide etch introduces an undercut beneath the a-Si while preserving the straight LTO sidewall. BOE etches oxide isotropically but does not attack silicon. The device after this step is shown in Fig. 5.3.

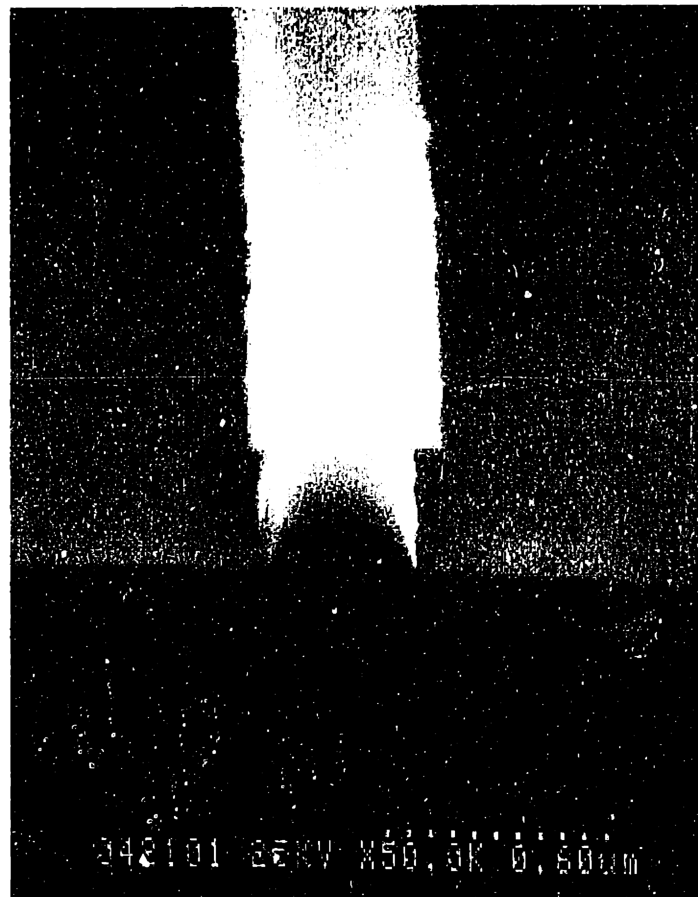


Fig. 5.3. Short-flow MOSFET after gate stack patterning and 20 seconds BOE etch. The structure shown is amorphous Si undercut by LTO.

Source/Drain/Gate Implantation

The source/drain and gate are doped simultaneously via ion implantation. Phosphorous or arsenic is implanted for NMOS devices while boron is used for PMOS. The energy of the implant is chosen to fully dope the gate material without punching through the LTO into the substrate as shown in Fig. 5.4.

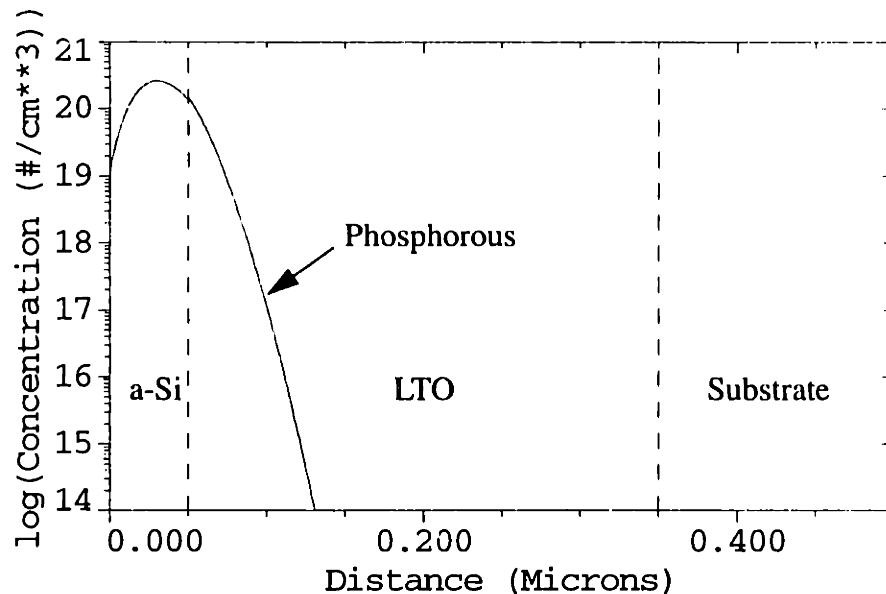


Fig. 5.4. Example implant profile (Phosphorous 25keV $1\times10^{15}cm^{-2}$).

The overhang of the gate stack shields the substrate from the ion implant. However, if dopants do not reach the corner of the gate oxide, a high source/drain series resistance results. Thus, four implants are performed at 90° rotations and at a tilt of 7° to eliminate the overhang's shadow.

After implant, a short piranha clean (3:1 H_2SO_4:H_2O_2) is performed. The time is shortened to 30 seconds from the standard 10 minutes to avoid etching of the heavily doped SiGe. A $600^{o}C$ anneal then activates the dopants. Efficient activation is achieved at this temperature through solid phase recrystallization.

Metallization

Prior to metal deposition, the wafers are piranha cleaned and the native oxide is removed with an HF dip (50:1 H_2O:HF). The length of the HF dip must be controlled as it adds to the amount of undercut. Particulates introduced at this point are the main yield limiter in this process, leading to bridging shorts between gate and source/drain metal. Thus it is important to thoroughly rinse and dry the wafers after this step and take care not to generate particulates during transfer to the e-beam.

Next, metal is electron-beam evaporated onto the wafers at perpendicular incidence, similar to a "lift-off" process, using custom-made wafer holders. A Ti/Al stack is deposited, Ti serving as a diffusion barrier to Al against spiking during sinter. If the metal thickness is less than the oxide thickness, there will be a discontinuity between source/drain and gate metal.

Since the thinnest possible oxide is desired, the minimum amount of metal is deposited. About 500Å of Ti is necessary to protect against spiking, as shown in Fig. 5.5 [36]. Good electrical contact to needle probes or bonding wires requires at least 1000Å of Al. Thus, the total metal height is about 1500Å.

Finally, Al is evaporated onto the backside to facilitate good electrical contact to bulk and the devices are sintered.

5.3 Layout and Operation

The short-flow MOSFET layout is a "ring transistor" or "edgeless FET" geometry as shown in Fig. 5.6. The dark areas in the figure are the gated regions while the light areas are the source/drain regions. The inside of the corner marked in the figure is shown at an angle in Fig. 5.7.

The isolation ring is used to allow a substrate bias to be applied. In this case, the area beneath the ring is brought into accumulation, electrically disconnecting the source area

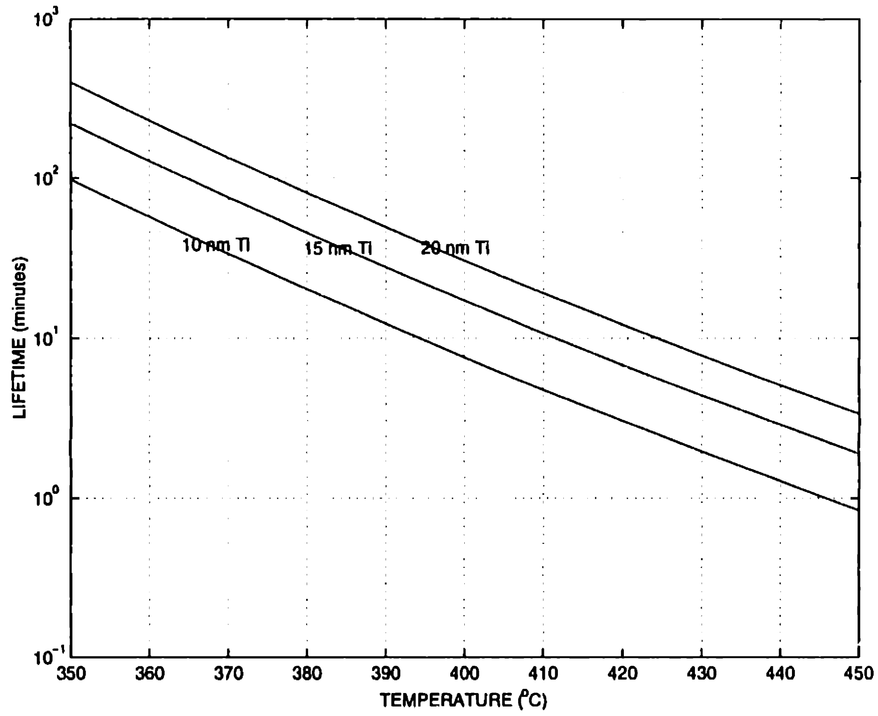


Fig. 5.5. Ti/Al contact diffusion barrier lifetime vs. temperature.

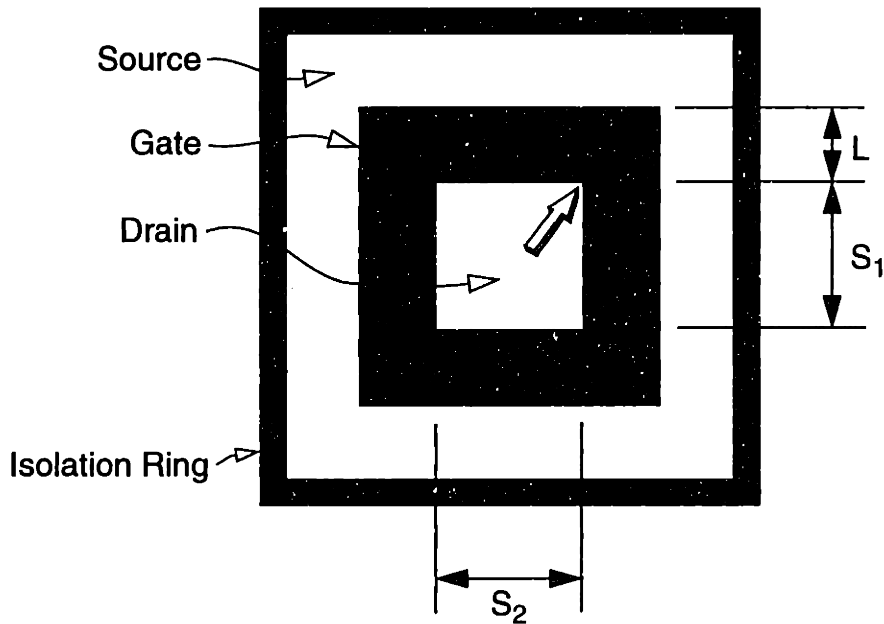


Fig. 5.6. Ring transistor mask layout. Large arrow indicates view shown in Fig. 5.7.

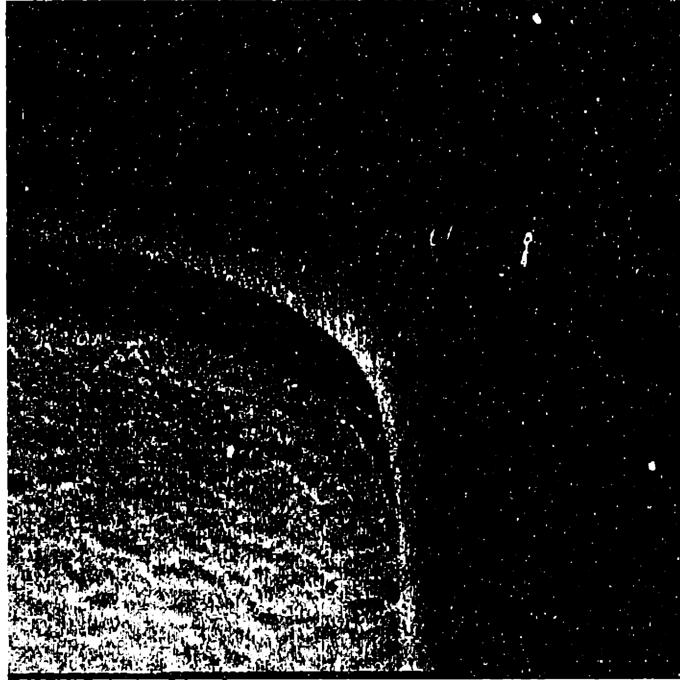


Fig. 5.7. SEM of inside corner of ring transistor gate. Gate region is the raised area to the upper right and the drain is to the lower left.

from the rest of the wafer. Without the isolation ring, the source-to-substrate diode area would be too large to allow a bias to be applied.

Device operation is similar to a conventional, rectangular FET with the exception that it is impossible to define a specific width or length. Instead, the channel is longer in the corners than on the sides.

In the linear regime, the inversion layer can be approximated as region of uniform sheet resistance bounded by ideal contacts at the source and drain. The isopotential contours for this case are shown in Fig. 5.8.

The resistance between the contacts is given by

$$R = R_{\square}G, \tag{5.4}$$

where $R_{\square}$ is the sheet resistance and G is a geometry-dependent value or "geometry factor". A simple expression for G as a function of L, S_1 and S_2 as defined in Fig. 5.6 does not exist except in the limit $L \ll S_1, S_2$, where $G = \frac{L}{2S_1 + 2S_2}$.

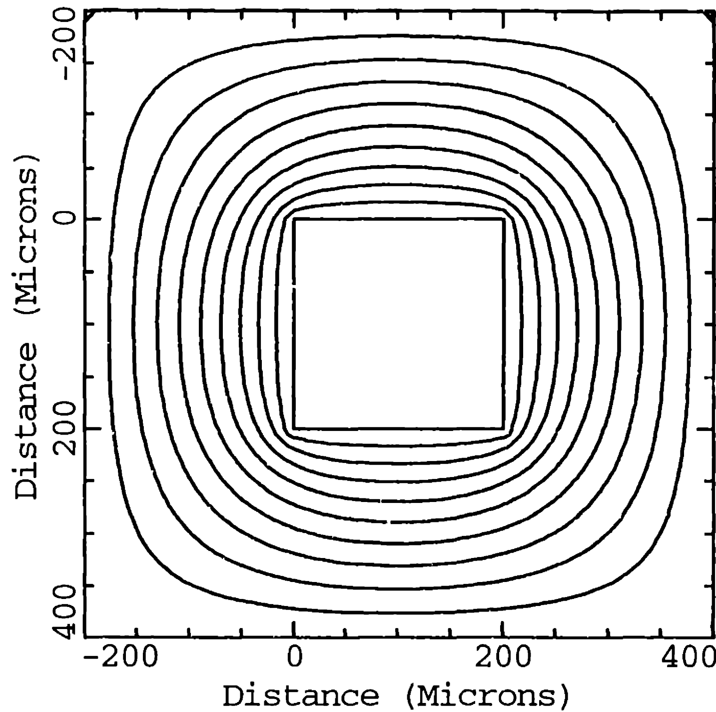


Fig. 5.8. Isopotential contours for square ring of uniform sheet resistance bounded by ideal contacts on inside and outside edges.

Accurate values for G were obtained through numerical solution of the Laplace equation using MEDICI. Calculated values of G for various device sizes are given in Table 5.2. Device names correspond to those on the MOBIL and TMASK-CP masks used in this work.

Following from (5.4), the linear regime device current can be modeled by

$$I_D = \frac{1}{G}\mu_{eff}C_{ox}(V_{GS} - V_T)V_{DS}, \tag{5.5}$$

where the familiar $\frac{W}{L}$ ratio is replaced by $\frac{1}{G}$. This equation is used to extract the effective mobility, μ_{eff}.

Series resistance can be extracted from devices of several different geometries using a method such as Terada-Muta so long as the resistance does not have geometry-dependent components. Series resistance is expected to be low in this process because the source and drain are fully metallized. However, the fact that the metal is so thin can cause problems

TABLE 5.2
GEOMETRY FACTORS FOR VARIOUS DEVICE SIZES

Name	L (μm)	S_1 (μm)	S_2 (μm)	G
500M250	500	250	250	.240
200M250	200	250	250	.138
100M250	100	250	250	.081
50M250	50	250	250	.045
1H200	200	200	200	.160
1H150	233	150	200	.191
0.75H200	150	200	200	.131
0.75H150	175	150	200	.160
0.75H100	225	100	200	.206
0.5H200	100	200	200	.097
0.5H150	117	150	200	.121
0.5H100	150	100	200	.160

with needle probes, which have the tendency to push the metal away, resulting in a spurious high series resistance.

The area of the gate is given by

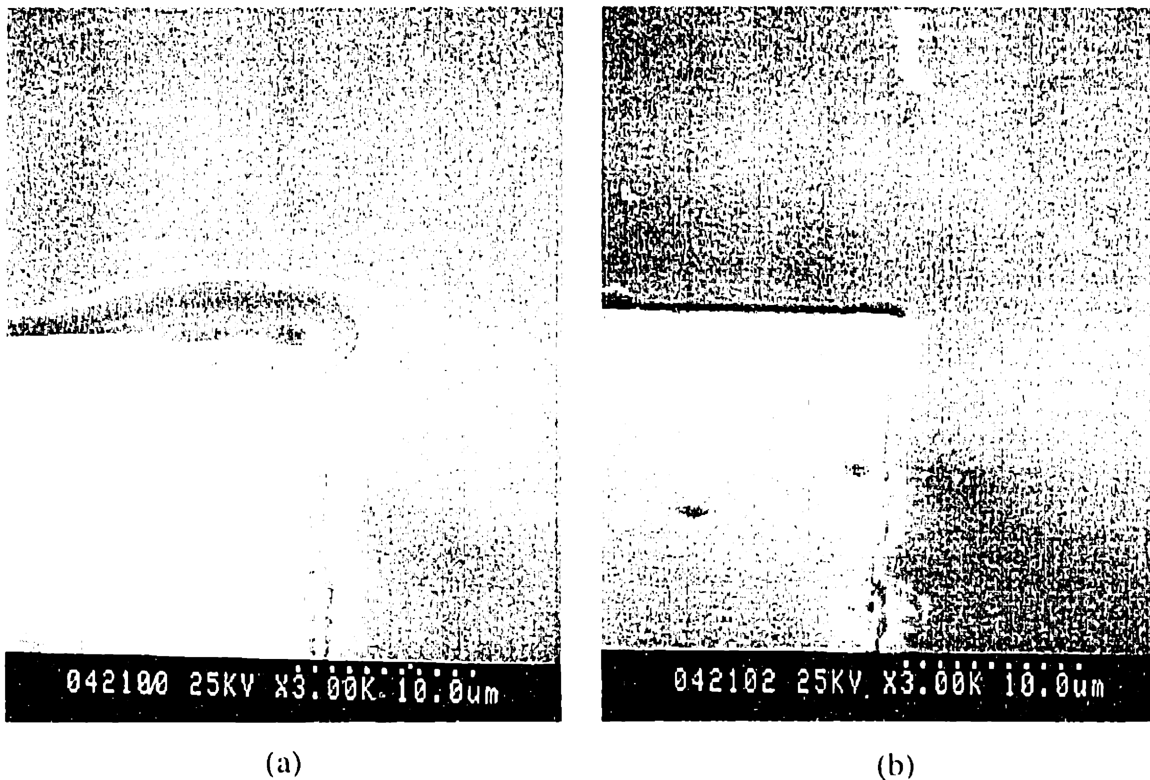
$$A = 2(L - 2\Delta L)(2L + S_1 + S_2), \qquad (5.6)$$

where $2\Delta L$ is the linewidth shrink due to overetching and undercutting during fabrication. The area is needed to calculate the gate oxide thickness from capacitance measurements.

5.4 Measurement

Due to the thick gate oxide of the short-flow MOSFET, large gate voltages must be applied to obtain effective electric fields of sufficient magnitude in the inversion layer. For instance, E_{eff} = 1 MV/cm requires V_{GS} = 170 V for a moderately doped substrate. Unfortunately, the devices typically breakdown through the air between the gate and source/drain at voltages around 100 V, which is a field of around 3 MV/cm. This results in signif-

icant damage to the device as shown in Fig. 5.9(a).



(a) (b)

Fig. 5.9. Inside corner of gate of NMOS device after arcing (a) without and (b) with 10 MΩ series resistor. Gate is to upper right and is darker due to charging.

That breakdown fields of this magnitude can be achieved without dielectric isolation is actually quite remarkable. It derives from the fact that breakdown in air is an avalanche mechanism. For small gaps, as is the case here, any ions generated by the field will most likely travel all the way to the electrode without a collision, resulting in a leakage current but no breakdown [37]. Since the gaps are on the order of the average distance between air molecules, 1 μm at atmospheric pressure, I believe that breakdown occurs along an indirect electric field line extending several microns out into the air.

Any attempt to increase the breakdown voltage by replacing air with a high-breakdown material such as transformer oil fails because it simply places more molecules in the gap. Measurement under vacuum may be beneficial, but was not tried. Instead, it was found

that minimizing the arc damage as described below was sufficient to allow repeatable measurements to high voltages.

The damage caused by arcing can be minimized by inserting a 10 MΩ series resistance immediately above the gate electrode. This limits the instantaneous arcing current to below 10 μA at an applied bias of 100 V. This is typically enough to prevent large amounts of metal from being vaporized at high current density areas along the edges of the gate and at the contact point between probe tip and metal, as shown in Fig. 5.9(b). However, one must make sure that the steady-state gate leakage current is below 1 μA to limit the voltage drop across the resistor to less than 1 V.

The HP4145 semiconductor parameter analyzer from Hewlett-Packard is used for I-V measurements. This instrument only goes to 100 V In order to obtain the gate bias range of interest, one can bias the source at -70 V and reference everything to that voltage. This means biasing the chuck at -70 V and the drain at -69.9 V in order to obtain $V_{BS} = 0$ V and $V_{DS} = 0.1$ V. Then sweeping the gate from -100 to 100 V produces a V_{GS} swing from -30 to 170 V. Of course, one must be careful to avoid electric shock when voltages above 42 V are present.

Chapter 6

Buried-Channel Strained-Si NMOS

6.1 Introduction

Strained-Si can be used to create high-electron-mobility devices of buried or surface-channel configuration. This chapter focuses on buried channel, while the next chapter focuses on surface-channel.

In this chapter I discuss the results of buried-channel Si/SiGe NMOS devices fabricated with the short-flow MOSFET process. Wafers were grown by UHVCVD at IBM and processing was carried out at MIT. First I compare my results to previous work. Then, I present the results of an effort to increase mobility through the addition of a buried donor layer. Last, I discuss the effect of an in-situ-doped p-well on mobility.

6.2 Comparison to Previous Work

A buried-channel strained-Si NMOS device has a layer and band structure like that shown in Fig. 6.1(a). The whole structure is fabricated on a relaxed SiGe buffer achieved using a graded-buffer technique [26]. The strained-Si buried channel is grown on top of that relaxed buffer. Electrons preferentially populate the strained-Si due to its lowered conduc-

tion band energy. A relaxed SiGe cap layer is grown on top of the strained-Si in order to confine the carriers away from the oxide interface. A strained-Si cap layer is grown on top of the whole structure in order to avoid oxidation of SiGe which results in poor oxide-interface quality. However, electron population of this layer is undesirable as it has lower mobility than the buried channel. Ideally, the Si cap is almost completely removed during processing.

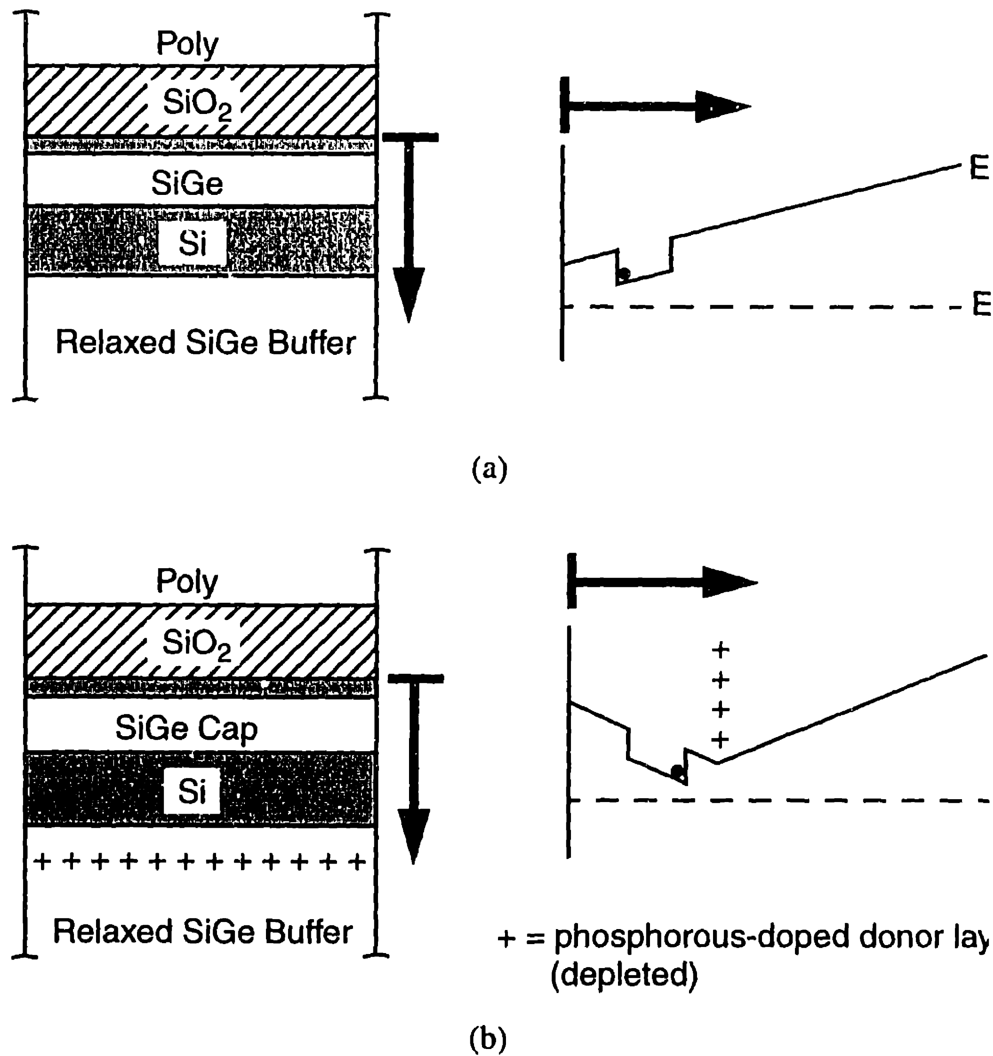


Fig. 6.1. Buried-channel NMOS structure (a) without and (b) with donor layer. Strained-Si cap is not indicated in the conduction band because ideally it is of negligible thickness.

Buried-channel strained-Si NMOS devices of this type have been reported by several groups [29][22][8]. My comparison will be to the work of Welser [8].

Fig. 6.2 shows the layer structure of Welser's buried-channel strained-Si device compared to my device labeled HNMOS-15. Thicknesses were determined by TEM in Welser's case and are inferred from a TEM of a similar structure (HNMOS-10) in my case. Welser has no cap layer as it has been completely oxidized away during the 750 C oxidation process used to grow a 12 nm thermal gate oxide. This was confirmed by micro-Auger spectroscopy directly on measured samples.

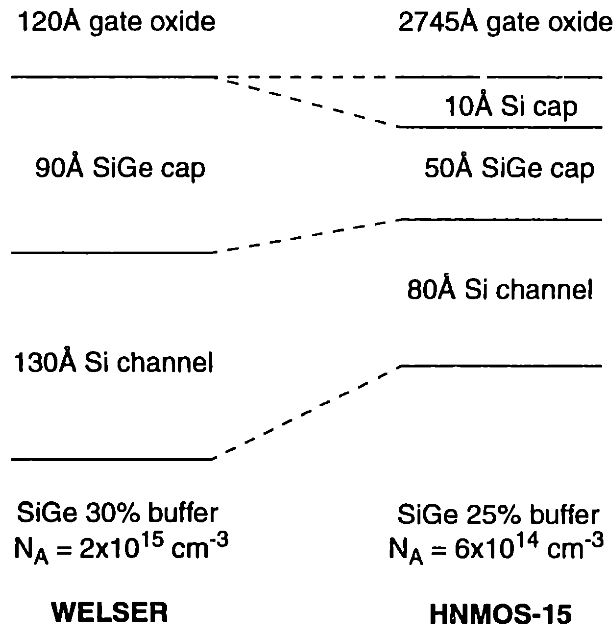


Fig. 6.2. Layer structure of Welser [8] compared to HNMOS-15. The gate oxide is immediately above this structure.

In contrast, my gate oxide is LTO deposited on a Si cap of 10Å estimated thickness. The structure is grown with a 26Å Si cap from which some thickness is removed by formation of an original native oxide, the removal of that oxide by an HF dip between the SC-1 and SC-2 cleans during the pre-gate-oxide RCA clean, and the reformation of a native oxide during SC-2 clean. The formation of two native oxides should consume between 10 and 15 Å of Si leaving between 11 and 16 Å of the Si cap.

The SC-1 solution is known to etch Si at a rate of around 5 Å/min [38] and thus should only remove 2.5 Å during the 30 s SC-1 clean time of this process. SC-2 does not etch Si

at an appreciable rate. Thus, I assume I have approximately 10 Å of Si cap remaining beneath the gate oxide.

Fig. 6.3 shows the effective mobility μ_{eff} vs. effective field E_{eff} comparing Welser to HNMOS-15. μ_{eff} was determined from the linear regime drain current using

$$I_D = \frac{1}{G}\mu_{eff}Q_{inv}V_{DS}, \qquad (6.1)$$

where for Welser, $G = L/W$ and $Q_{inv} = C_{ox}(V_{GS} - V_T)$, while for HNMOS-15, G is the geometry factor taken from Table 5.2 and Q_{inv} is extracted from simulation after careful simultaneous fitting of the C-V and log-I-V characteristics. This yields a value close to $C_{ox}(V_{GS} - V_T)$ except very close to V_T, where it is higher and tends to reduce the extracted mobility compared to using $C_{ox}(V_{GS} - V_T)$.

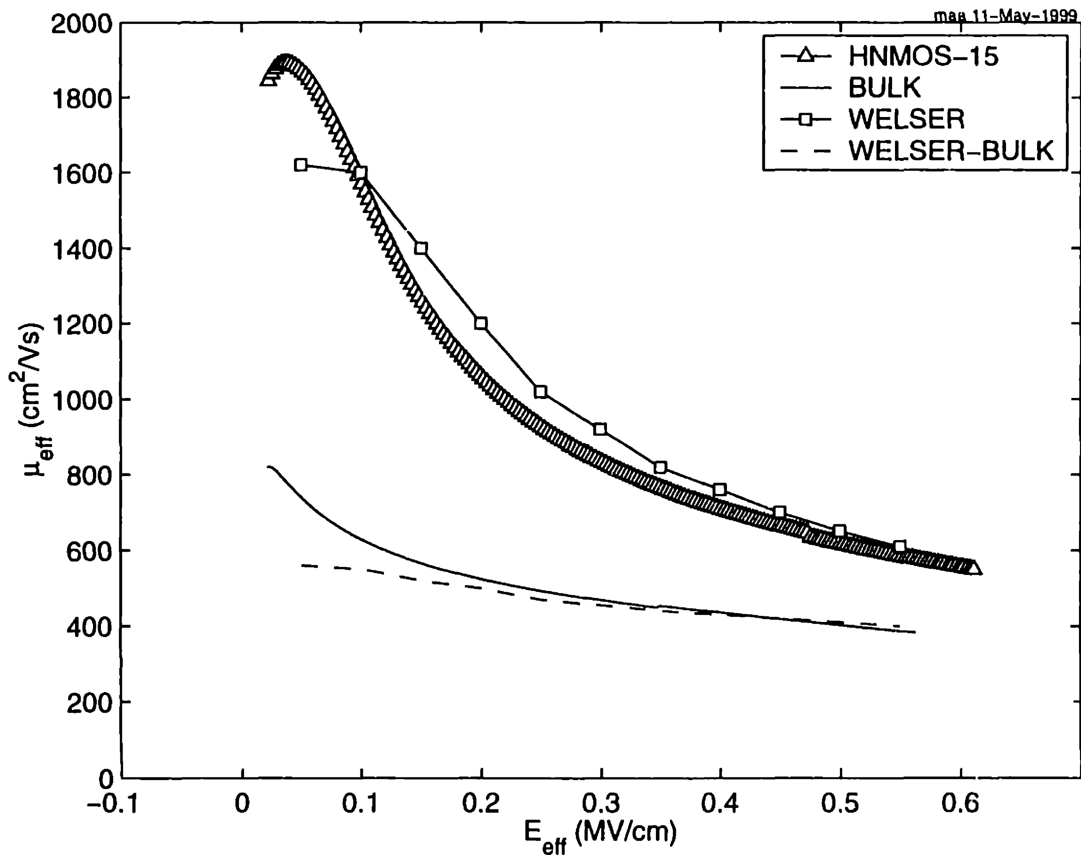


Fig. 6.3. Effective mobility of HNMOS-15 compared to Welser. Also shown are bulk monitors from both works.

E_{eff} was calculated using

$$E_{eff} = \frac{0.5Q_{inv} + Q_B}{\varepsilon_{Si}}, \quad (6.2)$$

where Q_B was approximated by $N_A x_{d,max}$ in the case of Welser, where $N_A = 2 \times 10^{15}$ cm^{-3} and $x_{d,max}$ is the maximum depletion depth in inversion, and from simulation in the case of HNMOS-15.

Agreement is quite good between Welser's results and mine for both Si/SiGe and bulk samples. Welser's low-field mobility is most likely even higher since $C_{ox}(V_{GS} - V_T)$ tends to overestimate the charge in the buried channel due to the additional effective dielectric thickness of the SiGe cap. His cap thickness of 90 Å converts to an effective oxide thickness of 30 Å due to the ratio of permittivities of oxide to SiGe. This means his effective t_{ox} at low fields in closer to 150 Å. Applying this correction factor raises his peak mobility to 2000 cm^2/Vs. This makes more sense, since his buried channel is thicker and further set back from the Si/SiO_2 interface than mine, which should lead to higher mobility.

A yet higher peak Hall mobility of 2100 cm^2/Vs was recorded by Ismail [22] in an MOS device using a layer structure similar to HNMOS-15 but with an n-type modulation-doped region below the channel to supply carriers. However, I believe this number cannot be compared directly to the drift mobility values given here due to the Hall scattering factor mentioned previously. That structure, however, forms the motivation for the donor-layer enhanced structure of the next section.

In both cases, the mobility drops rapidly with increasing effective field. Welser's mobility is slightly higher than mine at medium fields, but converges with mine at fields above 0.5 MV/cm. Welser blames the rapid fall in mobility with increasing field on occupation of the SiGe cap. I now investigate this through 1-D simulations.

Simulations were carried out using Avant! MEDICI using the structural parameters shown in Fig. 6.2 and the material parameters from Table 6.1. This results in a conduction band offset ΔE_C of 20 meV between the strained-Si and relaxed $Si_{0.7}Ge_{0.3}$. Simulations also

included an interface state density D_{it} of 10^{11} cm^{-2} eV^{-1} for both Welser's device and HNMOS-15.

TABLE 6.1
MATERIAL PARAMETERS FOR MEDICI SIMULATIONS

Material	$\varepsilon/\varepsilon_0$	χ (eV)	E_G (eV)
Relaxed $Si_{0.7}Ge_{0.3}$	13.06	4.12	1.03
Strained Si on $Si_{0.7}Ge_{0.3}$	11.80	4.32	1.00

Fig. 6.4 shows Welser's mobility curve compared to a simple two-channel model. In this model, the effective mobility is described by the weighted sum of the buried and surface channel mobilities,

$$\mu_{eff} = \mu_{buried} r + \mu_{surface}(1 - r), \tag{6.3}$$

where r is ratio of buried to total carriers n_{buried}/n_{total}. Two fits are shown in the figure. Both assume a constant value for μ_{buried} with the justification that the carriers are confined away from the surface and will be less affected by surface roughness scattering. Also, the carriers in the buried channel travel under less vertical field than those in the surface channel, as the vertical field increases through the inversion layer from Gauss's Law. μ_{buried} is determined directly from the peak mobility since all carriers are moving in the buried channel at low fields and is assigned a value of 1700 cm^2/Vs.

In the first fit, the surface channel mobility $\mu_{surface}$ is assumed to follow the same dependence on effective field as the co-processed bulk device. This fit results in a slightly higher mobility than experiment at higher fields. This may be due to the fact that the relevant effective field in the surface channel is higher than the effective field calculated from the total inversion charge, since the average field in the surface channel includes contributions from the whole of the buried channel charge plus half of the surface channel charge, resulting in lower mobility in the buried channel for a given total effective field. Another reason could be poor mobility at the $SiGe/SiO_2$ interface due to alloy scattering and increased surface scattering.

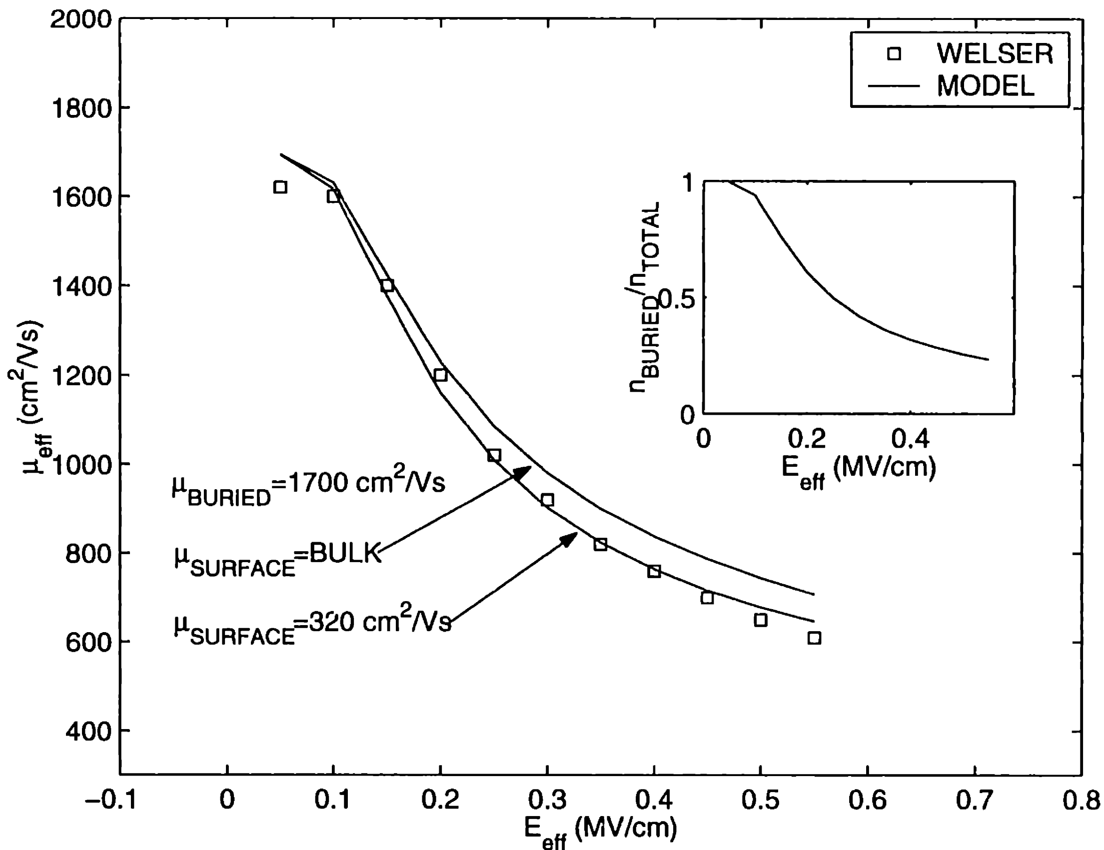


Fig. 6.4. Welser's mobility curve compared to a simple two-channel model. Also shown is the ratio of buried to total carriers.

For these reasons a second fit is shown with a constant $\mu_{surface}$ of 320 cm^2/Vs, reduced from about 400 cm^2/Vs in the case of the bulk-like surface mobility. This gives a better fit to the measured data.

The simple two-channel model of (6.3) is successful in describing the mobility curve of Welser's device. Now I will apply it to HNMOS-15. Fig. 6.5 shows the HNMOS-15 mobility data compared to the same simple two-channel model. Three model fits are shown, corresponding to different treatments of the silicon cap. The same values for $\mu_{surface}$ and μ_{buried}, 1900 and 320 cm^2/Vs, are used in all three fits.

One fit corresponds to no Si cap, meaning that the SiGe cap meets the oxide with no Si in between, similar to Welser's device. Since my SiGe cap is thinner than Welser's, I have

more carriers in the buried channel at a given effective field than does Welser, and the resultant mobility from the two-channel model is too high compared to experiment.

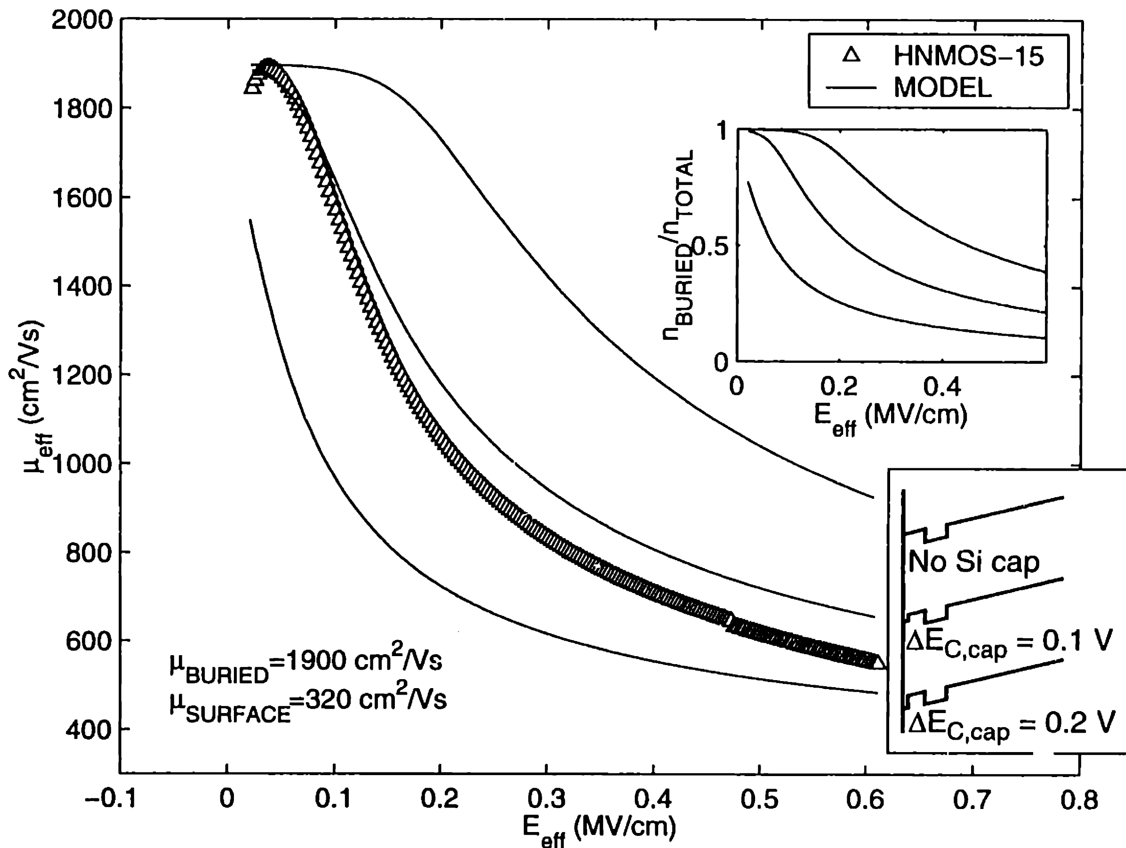


Fig. 6.5. HNMOS-15 mobility curve compared to the two-channel model. Also shown is the ratio of buried to total carriers.

Another fit corresponds to a 10Å strained-Si cap with a conduction band discontinuity ΔE_C from relaxed $Si_{0.7}Ge_{0.3}$ of 200 meV, which is the generally accepted value for this material interface. This substrate actually has closer to 25% Ge in the relaxed buffer, but this should not affect the ΔE_C too much. In this case, many carriers accumulate in the cap layer at relatively low fields. Since these carriers have the low mobility, this drags the total mobility down and the model results are too low for the experimental data.

The third fit corresponds to a 10 Å Si cap with a conduction band offset of 100 meV. The justification for raising the conduction band in the cap comes from the fact that a narrow

quantum well, such as exists in the Si cap, will have a minimum energy level which is raised from the bottom of the well. Also, since it is a shallow well, the carrier wave functions will not be completely confined and will penetrate into the adjacent SiGe material. Thirdly, there is a finite transition between the SiGe and Si which could reduce the offset in the cap. Since MEDICI does not solve the Schroedinger equation, these phenomena must be modeled in another way, such as by artificially raising the conduction band in the cap as I have done.

According to [45], the expected offset for a 10 Å wide strained-Si quantum well in relaxed-$Si_{0.3}Ge_{0.7}$ is only 10 meV, which is much less than the 100 meV value used here. A 20 Å well is expected to have an 87 meV offset. By increasing the thickness of the cap in the simulation and reducing the offset to match these values, I expect that the same results could be achieved.

The resulting fit to the experimental data for the 100 meV offset case is good, except for at high fields. This could be explained by a vertical field dependence of the buried channel mobility. Since the SiGe cap is only 50 Å in this sample, it is possible that the buried channel carriers experience some roughness scattering from the surface and Coulomb scattering from the charges in the surface channel. A slight degradation of μ_{buried} with vertical field would result in a better fit.

6.3 Effect of Donor Layer

In order to reduce occupation of the low-mobility surface channel provided by the strained-Si cap layer one can introduce a modulation-doped donor layer which bends the bands to favor occupation of the high-mobility buried channel as shown in Fig. 6.1(b). The phosphorous-doped donor layer is designed to be fully depleted at all times. All of the donor electrons are transferred to the strained Si due to the lower conduction band there. The occupation of the surface channel cannot be avoided completely, but this allows a large gate bias range over which the buried channel dominates.

This work represents an attempt to optimize the buried-channel NMOS structure with a phosphorous donor layer in order to achieve high-mobility over an increased voltage range as compared to devices without the donor layer.

Fig. 6.6 shows the layer structure of HNMOS-10, 13, and 15, which have different donor layer doses. HNMOS-13 has twice the donor layer phosphorous dose as HNMOS-10. HNMOS-15 has no donor layer at all, and also increased Si channel and SiGe cap thicknesses. HNMOS-15 was shown also in the previous section. As mentioned before, the 26Å cap is reduced to approximately 10Å after processing.

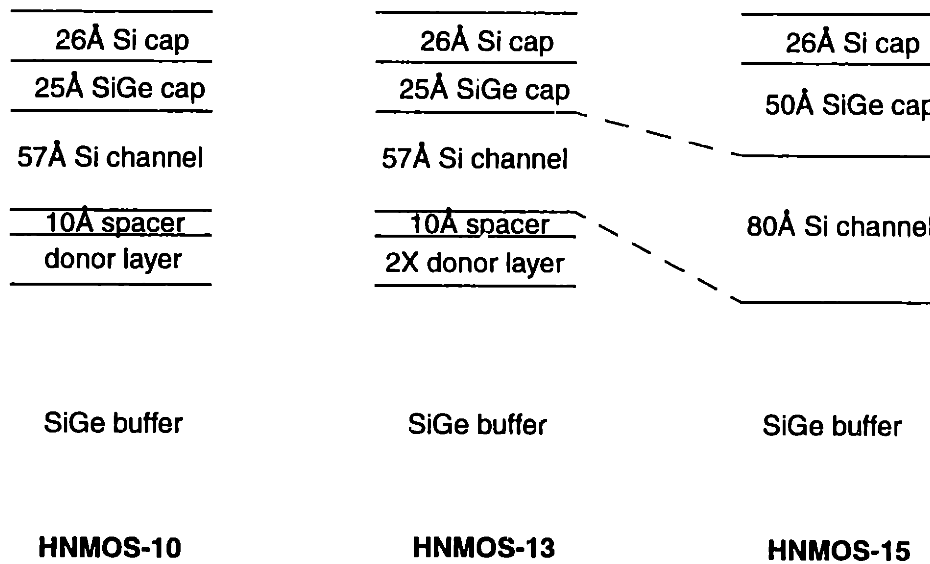


Fig. 6.6. Layer structure of runs HNMOS-10, 13, and 15.

Table 6.2 lists electrical parameters for short-flow MOSFETs fabricated on substrates from runs HNMOS-10, 13 and 15. The various HNMOS wafers were processed in separate lots with bulk monitor wafers in each lot.

t_{ox} is the effective oxide thickness determined from the maximum capacitance. D_{it} is the interface state density determined from the voltage spreading from a simulated without interface states C-V. Q_F, the fixed oxide charge, and Q_{DL}, the donor layer charge, were determined by the shift in flatband voltage from ideal. Q_F was assumed be the same as co-

TABLE 6.2
MOSFET ELECTRICAL PARAMETERS FOR HNMOS-10, 13 AND 15.

Parameter	HNMOS-10	HNMOS-13	HNMOS-15
t_{ox} (Å)	3090	3220	2745
D_{it} (10^{11} cm^{-2} eV^{-1})	4.0	4.0	1.0
Q_F (10^{11} cm^{-2})	5.6	4.4	2.0
Q_{DL} (10^{11} cm^{-2})	4.1	10.9	0.0
V_T (V)	-3.0	-7.4	0.2

processed bulk wafers, as the fixed charge is mainly related to the oxide. V_T was extracted from linear regime I-V by linear extrapolation from the point of maximum slope.

Fig. 6.7 compares the mobilities of HNMOS-10, 13 and 15 and co-processed bulk wafers from each run. Effective mobility and effective field were extracted from the linear regime drain current as described previously with inversion and depletion charges extracted from simulation after fitting to measured C-V and log-I-V data. As shown in the inset, peak mobility drops with increasing phosphorous dose in the donor layer. This is opposite to what is expected, since the donor layer should encourage occupation of the buried channel and increase the overall mobility. At least, I expect from the two-channel model that the peak mobility should remain constant, but mobility should be increased at higher fields.

From this I conclude that the donor layer reduces the peak mobility through increased Coulomb scattering from the charges in the donor layer from beneath the buried channel. I now attempt to draw a parallel between this and the degradation of bulk Si inversion layer mobility with fixed oxide charge. As shown in Fig. 6.8, the separation between oxide fixed charge and inversion charge at the Si/SiO_2 interface is on the same length scale as the separation between donor layer charge and inversion charge at the strained-Si/relaxed-$Si_{0.7}Ge_{0.3}$ interface. Fixed oxide charges are generally considered to reside in the first 50 Å of oxide, while the spacer distance in the HNMOS structures is about 10 to 20 Å.

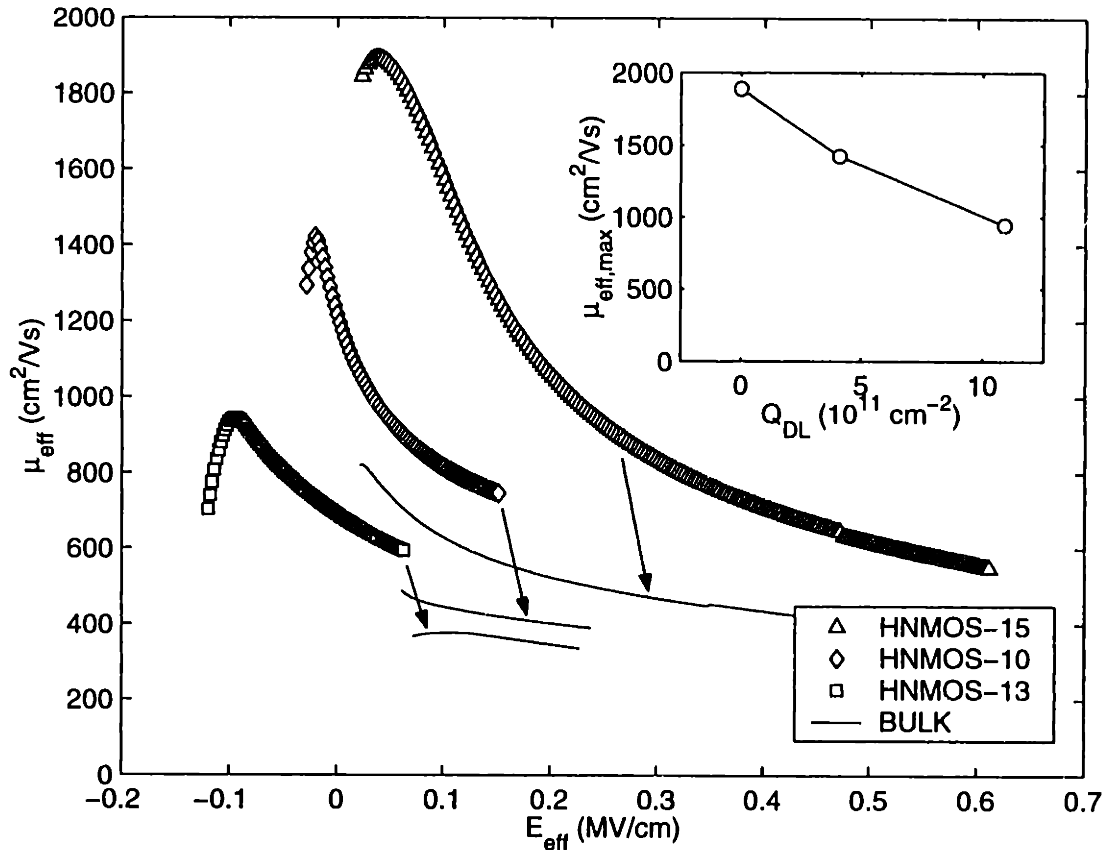


Fig. 6.7. Effective mobility vs. effective field for HNMOS-10,13 and 15. The three wafers have different phosphorous doses in the donor layer. Arrows point to the co-processed bulk mobility data. Inset shows peak effective mobility vs. donor layer dose.

Fig. 6.9 shows the co-processed bulk mobility vs. effective field compared to Sun and Plummer thermal oxide data [39]. The parameters for the bulk wafers are listed in Table 6.3. The Sun and Plummer data is for thermal oxides while my oxide is LTO. Note that CVD oxides contain a large amount of fixed charge which is distributed throughout the oxide and does not affect mobility. Only the charge near the interface affects mobility.

The BULK5 monitor corresponds well to the Sun and Plummer data. The other to bulk wafers do not have the expected dependence on Q_F. BULK4 has more fixed charge, but higher mobility than BULK3. However, the D_{it} of BULK3 is higher, which may contribute to the total interface charge. However, it is clear both have higher interface charge and thus lower overall mobility.

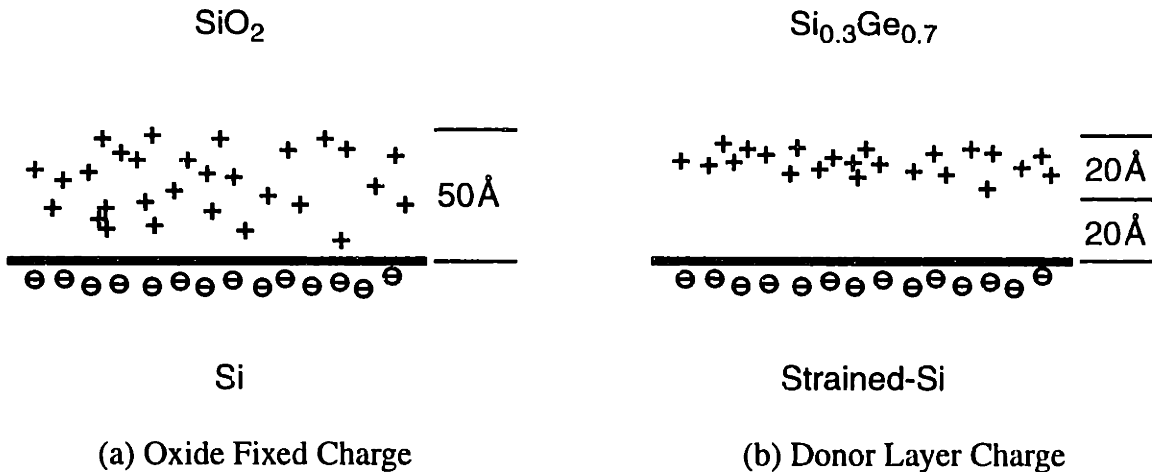


Fig. 6.8. Schematic showing comparison between (a) oxide fixed charge scattering in bulk Si inversion layers and (b) donor layer charge scattering in strained-Si quantum wells.

The thing to notice from this graph is the amount of fixed charge necessary to affect a large change in inversion layer mobility. Already at 2 x 10^{11} cm^{-2} the mobility is down 20%. The inset shows the dependence of peak mobility on fixed charge taken from the paper. Fig. 6.10 shows the normalized mobility reduction with donor layer fixed charge comparing the Sun and Plummer data for inversion layers in silicon with thermal oxide to the HNMOS donor-layer data. The Sun and Plummer data is normalized to the value at zero Q_F determined by extrapolation to zero. The dependence on charge density is similar in both cases, supporting the theory that the mobility reduction in HNMOS-15 is caused by increased Coulomb scattering from the donor layer charge.

Thus, I conclude that the donor layer technique cannot be applied to enhance the mobility of buried-channel strained-Si MOSFETs due to Coulomb scattering from the charges in the donor layer. This is because in order to donate a significant amount of charge into the buried channel, say 1 x 10^{12} cm^{-2}, you need approximately 1 x 10^{12} cm^{-2} in the donor layer, which is a huge fixed charge in terms of Coulomb scattering in inversion layers. One could try to increase the spacer thickness between the donor layer and strained-Si channel, but there is a limit to this thickness since the donor layer must be fully depleted at all times in VLSI applications. That is, there must not be a parallel conduction path through the donor layer which prevent the device from turning off completely. The amount of charge

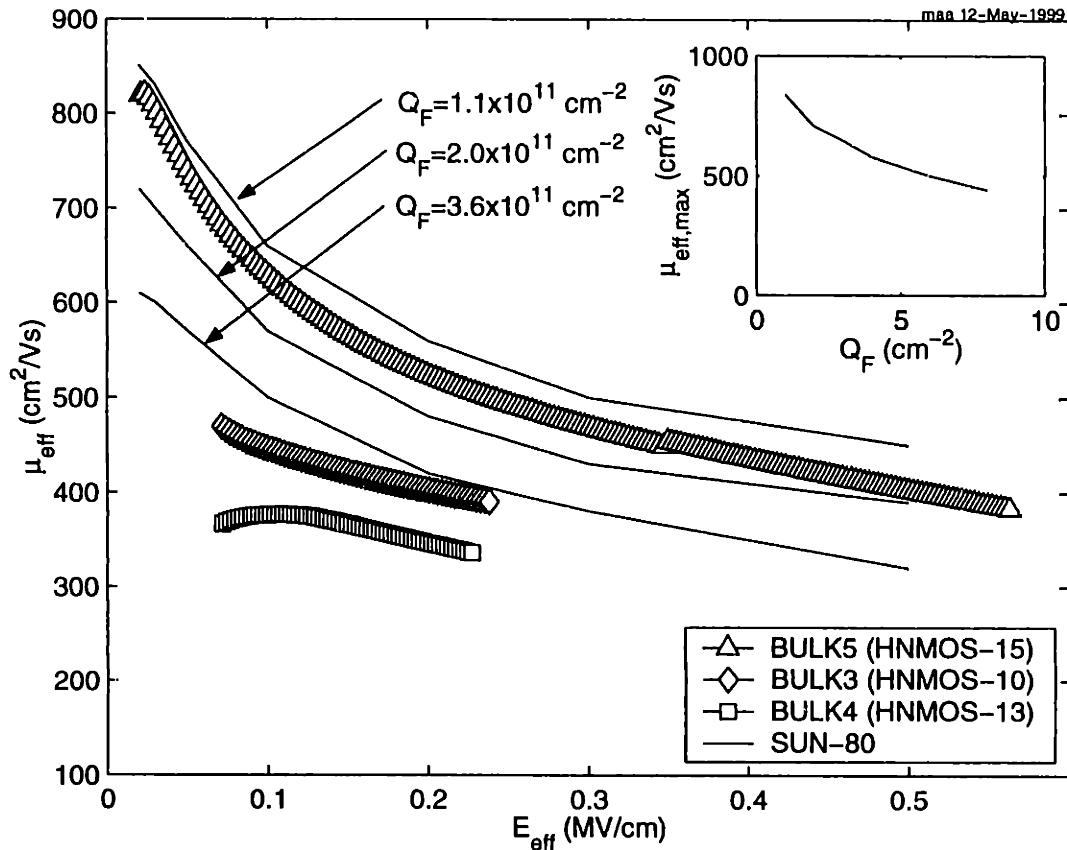


Fig. 6.9. Effective mobility of co-processed bulk monitors compared to Sun and Plummer thermal oxide data. The labels point to the Sun and Plummer data with $N_A = 9\text{x}10^{14}$ cm^{-2}. The bulk data corresponds to Table 6.3. Inset shows peak effective mobility vs. fixed charge taken from the Sun and Plummer paper.

TABLE 6.3
PARAMETERS OF CO-PROCESSED BULK WAFERS

Parameter	BULK3 (HNMOS-10)	BULK4 (HNMOS-13)	BULK5 (HNMOS-15)
t_{ox} (Å)	3090	3270	3020
N_A (10^{15} cm^{-3})	9.5	10.0	0.8
Q_F (10^{11} cm^{-2})	5.6	4.4	2.0
D_{it} (10^{11} cm^{-2} eV^{-1})	4.0	2.0	0.2

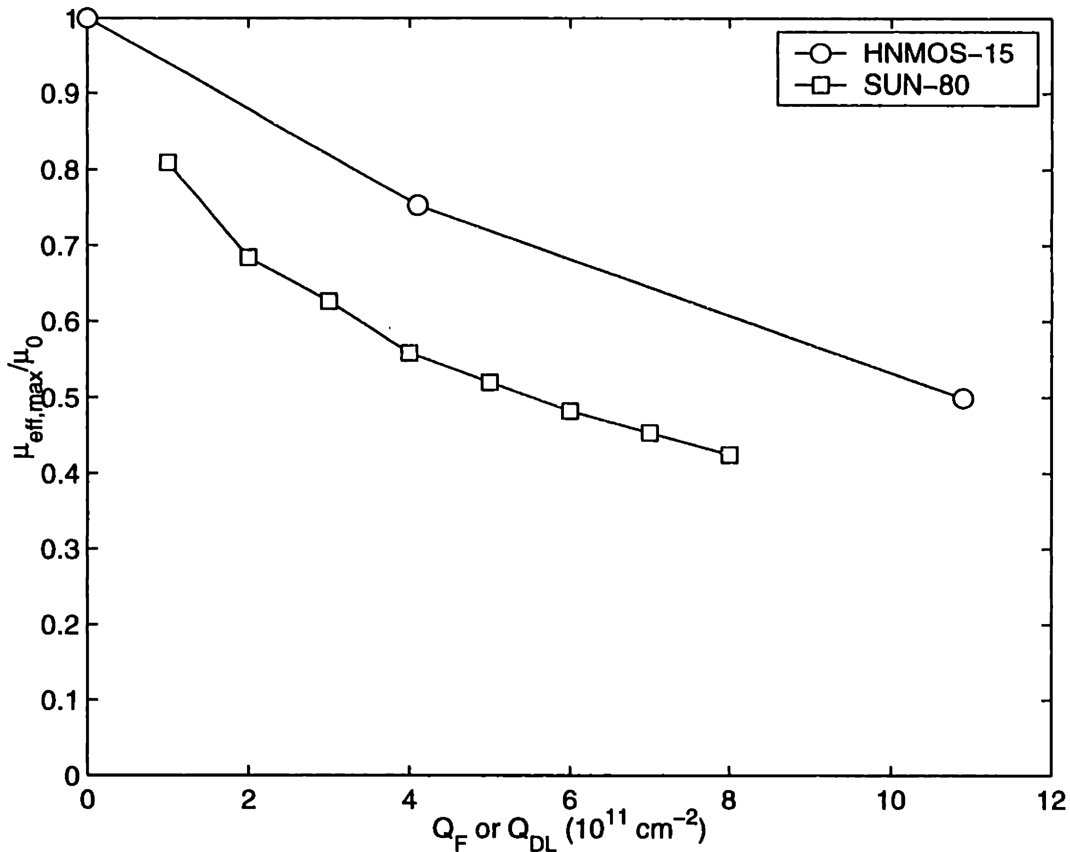


Fig. 6.10. Normalized reduction in peak effective mobility of HNMOS-15 buried strained-Si channel with donor layer charge compared to normalized reduction in peak effective mobility of inversion layer in silicon with thermal oxide including oxide fixed charge.

which can be put in the donor layer while still having it fully depleted decreases with increasing spacer thickness.

6.4 Effect of In-Situ P-Well

The structures discussed in the previous section had no intentional well doping. However, a heavily-doped p-well is necessary to maintain electrostatic integrity in sub-micron n-channel MOSFETs. In this section I explore the effect of placing a p-well beneath the channel layers of the HNMOS-10 structure.

Fig. 6.11 shows the layer structure of runs HNMOS-8, 9 and 10 as measured by TEM. These runs have similar layer thicknesses but different in-situ well doping. The germanium content in the buffer is 25% as measured by SIMS.

HNMOS-8	HNMOS-9	HNMOS-10
26Å Si cap	26Å Si cap	26Å Si cap
25Å SiGe cap	25Å SiGe cap	25Å SiGe cap
57Å Si channel	57Å Si channel	57Å Si channel
10Å spacer	10Å spacer	10Å spacer
20Å donor layer	20Å donor layer	20Å donor layer
300Å spacer	300Å spacer	
500Å p-well $N_A = 10^{18}$ cm^{-3}	500Å p-well $N_A = 5\text{x}10^{17}$ cm^{-3}	SiGe buffer
SiGe buffer	SiGe buffer	

Fig. 6.11. Layer structure of runs HNMOS-8, 9 and 10.

Fig. 6.12 shows the doping profiles of runs HNMOS-8, 9 and 10. The data represents boron doping except for the spike to the left which is the phosphorous-doped donor layer. The boron distribution was measured by SIMS while the phosphorous distribution, difficult to detect with SIMS, is a guess based on the intended profile with total dose inferred from C-V data. HNMOS-10 has no intentional well-doping but is slightly p-type due to residual contamination in the growth chamber.

Table 6.4 lists electrical parameters for short-flow MOSFETs fabricated on a bulk substrate and substrates from runs HNMOS-8, 9 and 10. Symbols are as defined in the previous section.

Fig. 6.13(a) shows the effective mobility vs. effective field extracted from short-flow MOSFETs for the HNMOS-8, 9 and 10 structures compared to bulk silicon. Effective mobility and field were extracted from the linear regime device current as described earlier. The in-situ well doping has a large effect on the mobility. The two boron-doped sam-

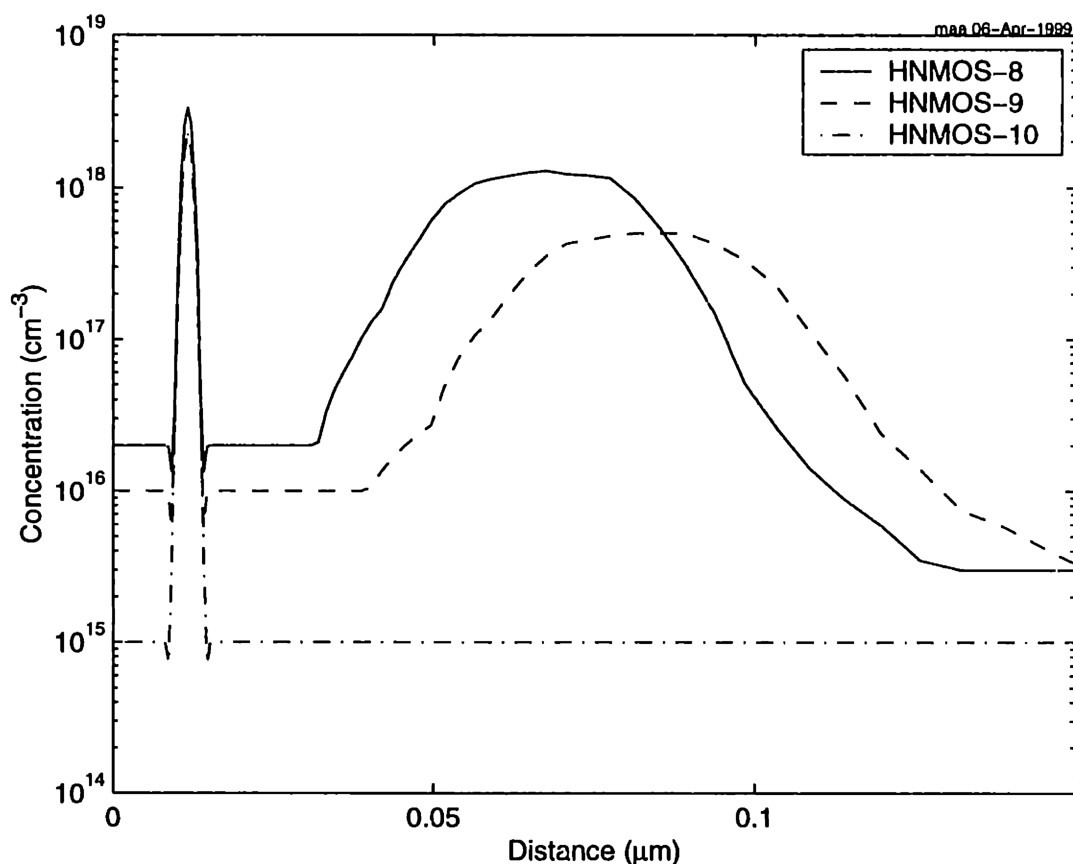


Fig. 6.12. Doping profiles of runs HNMOS-8, 9 and 10. The large hump is boron doping from SIMS. The small spike is phosphorous doping which is a guess based on the intended profile with total dose determined form C-V data.

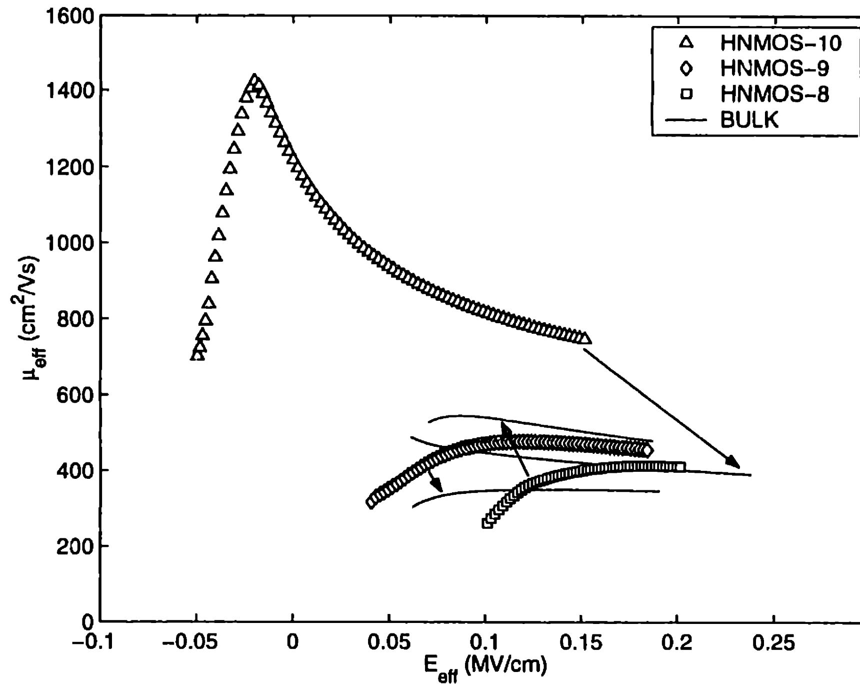
TABLE 6.4
MOSFET ELECTRICAL PARAMETERS FOR HNMOS-8, 9 AND 10.

Parameter	HNMOS-8	HNMOS-9	HNMOS-10
t_{ox} (Å)	4460	3820	3090
D_{it} (10^{11} cm^{-2} eV^{-1})	1.0	1.0	3.0
Q_F (10^{11} cm^{-2})	4.2	3.3	6.0
Q_{DL} (10^{11} cm^{-2})	6.0	4.0	4.0
V_T (V)	5.3	4.3	-5.0

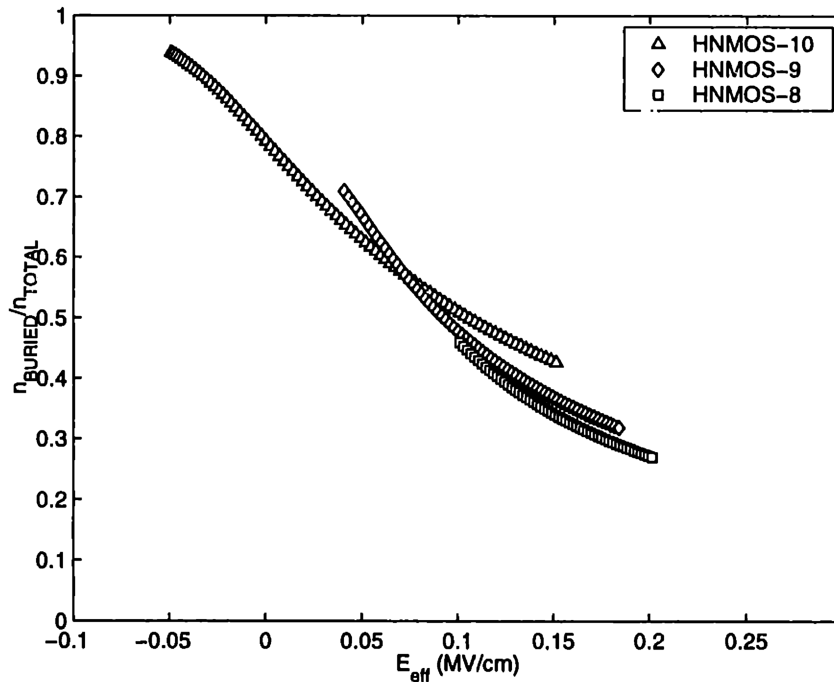
ples, HNMOS-8 and 9, have much lower mobility than HNMOS-10. This trend continues inside the first pair, with HNMOS-9 having slightly higher mobility than the heavier-doped HNMOS-8.

Fig. 6.13(b) shows the simulated ratio of electrons in the buried channel to total electrons vs. effective field for HNMOS-8, 9 and 10. In theory, the p-well will reduce this ratio due to the increased electric field at the surface due to the bulk charge, thus reducing the mobility. However, in this case the p-well does not change this ratio appreciably. Thus, the observed difference in mobility between HNMOS-8, 9 and 10 cannot be explained by a shift in carriers from the buried channel to the parasitic surface channel due to the p-well doping.

From this I conclude that the in-situ p-well must affect the subsequent growth in some way so as to ruin the mobility. The cause is unknown. Fig. 6.11 shows that the boron has trailed off completely by the time the channel layers are grown, so ionized impurity scattering is not expected. The effect of an implanted p-well on surface-channel strained-Si NMOS is investigated in the next chapter.



(a)



(b)

Fig. 6.13. (a) Effective mobility vs. effective field for HNMOS-8, 9 and 10. Arrows point to co-processed bulk wafers. (b) Simulated ratio of buried to total carriers vs. effective field. In this simulation a 10 Å Si cap was included with $\Delta E_C = 0.2$ eV. The difference in observed mobility between the three samples cannot be explained by differences in carrier location.

Chapter 7

Surface-Channel Strained-Si NMOS

7.1 Introduction

I investigate the effect of well implants on the mobility of surface-channel strained-Si NMOS devices. Fig. 7.1 shows the layer structure and band diagram for surface-channel strained-Si NMOS. Well implants are necessary to integrate NMOS and PMOS devices on the same wafer for CMOS. However, there is little data in the literature on integrated or implanted devices. Will the implant relax the strain? Will strained-Si follow a universal curve like bulk Si?

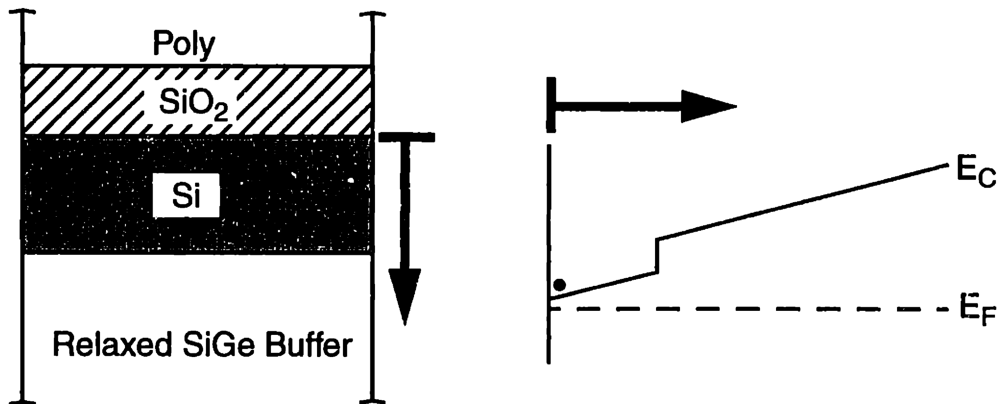


Fig. 7.1. Layer structure and band diagram of surface-channel strained-Si NMOS.

7.2 Device Fabrication

Strained-Si and bulk Si NMOS devices were fabricated with various well implants using a short-flow MOSFET process. The strained-Si substrates were grown by UHVCVD and consisted of 70 Å strained-Si on a relaxed-$Si_{0.75}Ge_{0.25}$ buffer created with a graded buffer technique. The grown layers were not intentionally doped and were grown on p-prime Si wafers. Bulk Si NMOS devices were fabricated on p-prime (10-20 Ωcm) substrates.

The wafers received a 13 keV boron p-well implant at doses of 2×10^{11}, 1×10^{12} or 5×10^{12} cm^{-2}. In addition, some dice received an electrically neutral 45 keV silicon implant while other dice were left unimplanted. Fig. 7.2 shows the well implant profiles as calculated by SUPREM. The well implants were chosen to mirror realistic well implants for modern submicron CMOS devices in terms of concentration and proximity to surface.

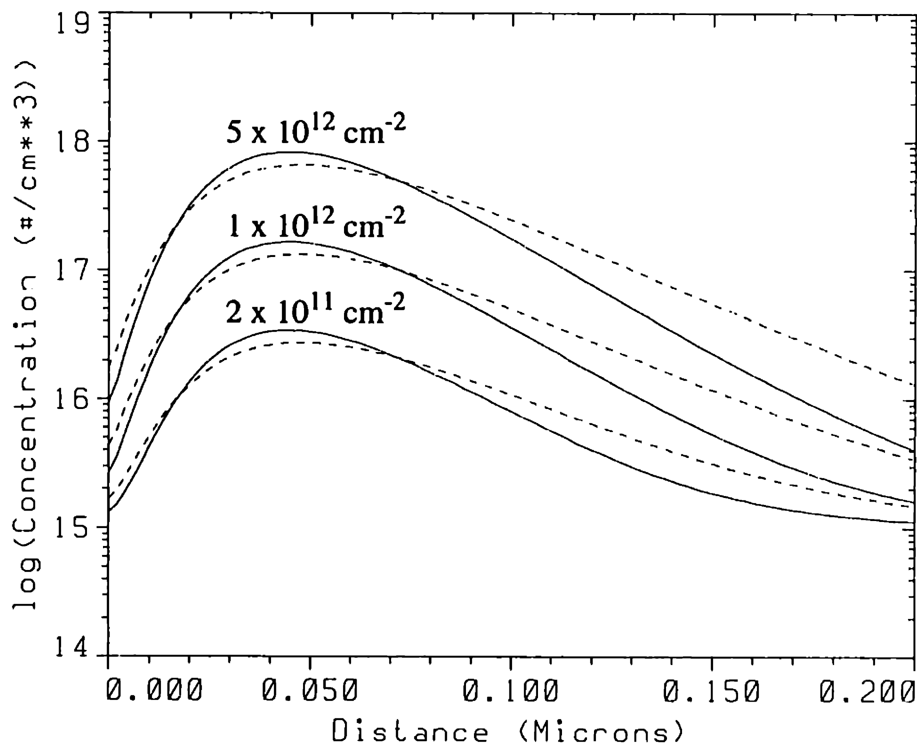


Fig. 7.2. Well implant profiles as calculated by SUPREM. Solid lines are boron and dashed lines are silicon (approximated in SUPREM by a phosphorous implant at same conditions). Tilt is 7 degrees.

Oxide thickness was around 3300 Å. The source/drain implant was P at 25 keV and 2 x 10^{15} cm^{-2}. The only anneal was a 30 minute 600 °C anneal in N_2 after source/drain implant.

7.3 Mobility Extraction

Effective mobility μ was extracted from large devices from drain current I_D at V_{DS} = 0.1 V using:

$$I_D = \frac{1}{G}\mu C_{ox}(V_{GS} - V_T)V_{DS}, \qquad (7.1)$$

where V_T is the linear extrapolated threshold voltage and G is the geometry factor described in Chapter 5. Effective field E_{eff} was found from:

$$E_{eff} = \frac{0.5Q_I + Q_B}{\varepsilon_{Si}}, \qquad (7.2)$$

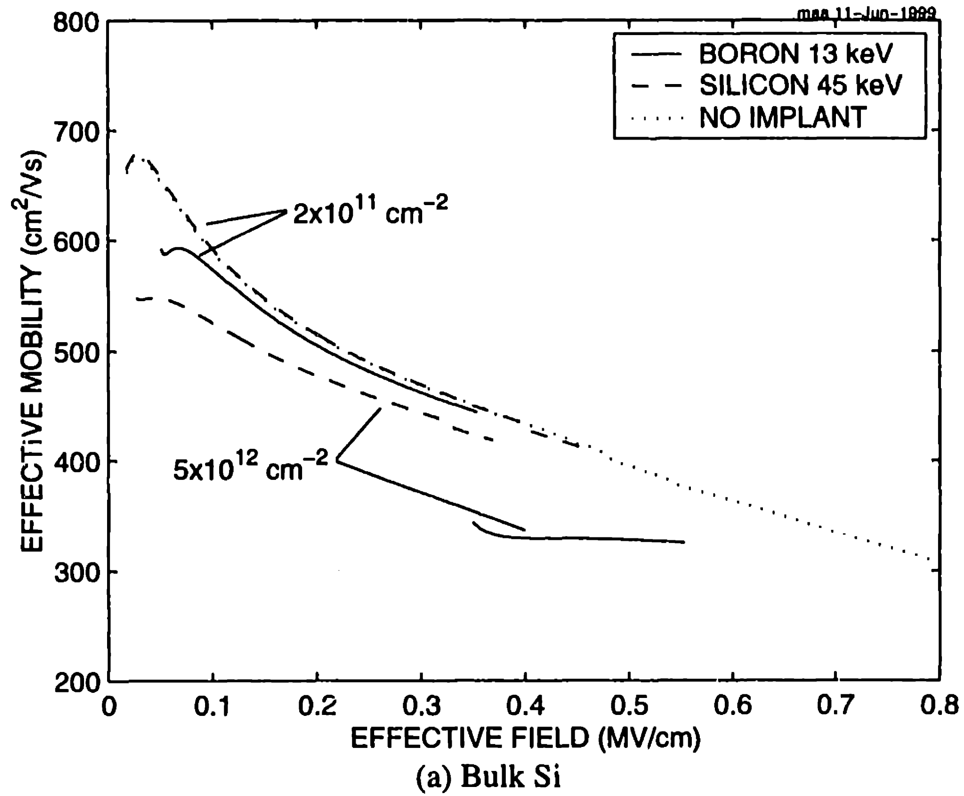
where $Q_I = C_{ox}(V_{GS} - V_T)$ and Q_B was calculated numerically in strong inversion from the well implant profiles of Fig. 7.2 assuming complete activation.

7.4 Results

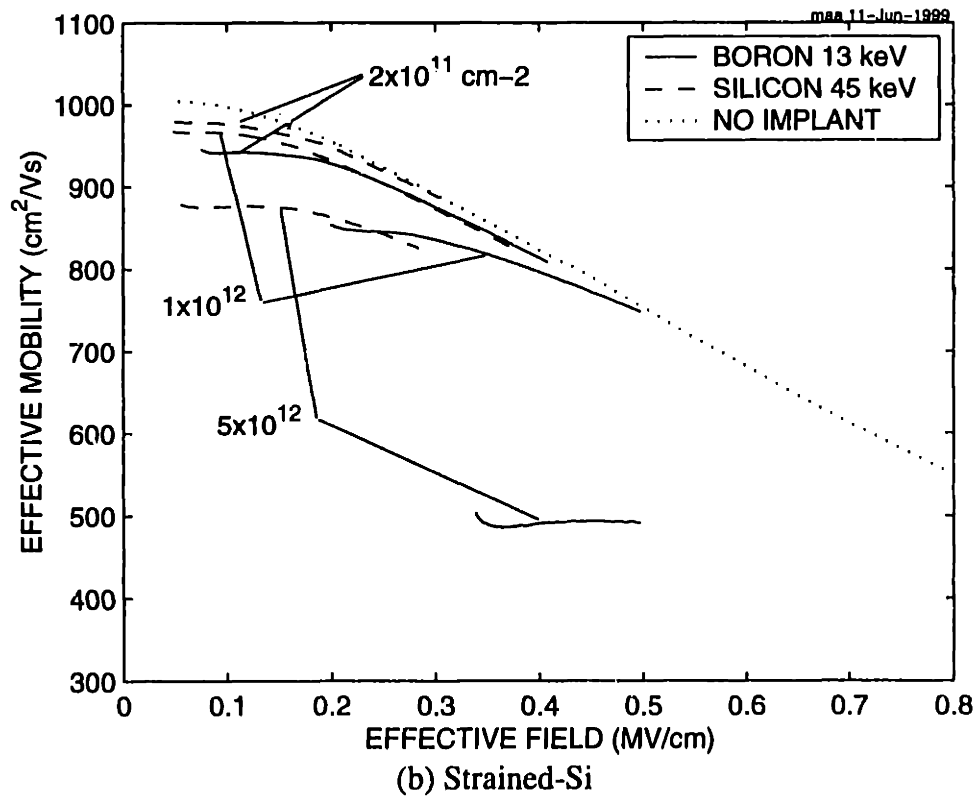
Fig. 7.3 shows μ_{eff} vs. E_{eff} for (a) bulk Si and (b) strained-Si NMOS. The medium dose bulk Si wafer was lost during processing. Due to metallization problems with this lot the maximum effective field measurable before breakdown was around 0.5 MV/cm. Peak mobility for the strained-Si device is around 1000 cm^2/Vs. A peak mobility of 1100 cm^2/Vs was recorded on a different sample.

Surface-channel strained-Si NMOS follows a universal curve behavior similar to bulk Si in that mobility is independent of doping at high E_{eff} [94]. At low fields, strained-Si retains a 1.6 times greater mobility than bulk Si regardless of implant dose. At high fields, the mobility advantage increases to 1.8 times regardless of implant dose.

The electrically neutral silicon implant reduces mobility only slightly and to the same degree in both strained-Si and bulk Si indicating that the act of implanting itself does not



(a) Bulk Si



(b) Strained-Si

Fig. 7.3. μ_{eff} vs. E_{eff} for (a) bulk Si and (b) strained-Si NMOS comparing B and Si well implants at several doses.

affect the strain of mobility in the layer but rather the mobility reduction is due solely to ionized impurity scattering.

Chapter 8

Conclusion

8.1 Summary

This thesis has examined several issues related to the integration of SiGe strained-layers into mainstream CMOS. The key contributions are summarized below, followed by recommendations for future work.

Device and Circuit Performance

The earlier part of this work strived to predict the performance leverage of high-mobility Si/SiGe devices over bulk Si devices in a realistic manner. I sought to include all the issues associated with a realistic implementation of Si/SiGe CMOS devices in the analysis. A novel layer design for Si/SiGe heterojunction CMOS (HCMOS) featuring buried electron and hole channels was optimized for minimum parasitic surface channel occupation and appropriate threshold voltage and off-current. By using a realistic layer design the issues associated with carrier motion in the buried channel were included, such as increased effective oxide thickness.

Two-dimensional hydrodynamic simulations were used to predict the static device characteristics of high-mobility Si/SiGe devices including the effects of series resistance, veloc-

ity saturation and velocity overshoot. The simulations showed enhanced current drive and transconductance over bulk Si devices at 0.2 μm effective channel length and highlighted the importance of velocity overshoot for high-mobility submicron devices. The circuit performance of Si/SiGe devices was determined from transient simulations of CMOS ring oscillators including the effects of parasitic capacitance and drain-to-source voltage at the onset of saturation $V_{DS,sat}$. The simulations showed a 4 to 6-fold reduction in power-delay product as compared to bulk CMOS oscillators operated at 2.5 V with the same design rules. This work is the most realistic and comprehensive analysis of the device and circuit performance of Si/SiGe CMOS devices to date.

Short-Flow MOSFET

The remainder of the thesis focused on the fabrication and characterization of Si/SiGe devices. Due to the complex nature of the layer structure, much optimization was needed in order to obtain high-mobility devices. Unfortunately, full MOSFET runs can take more than a month to turn around in a university lab. For this reason, I developed a short-flow, single-mask MOSFET which can be turned around in as little as a week in our lab. This device is superior to simple Hall mobility structures which suffer from leakage through the substrate, an inability to gate the electron and hole channels and the uncertainty associated with the Hall scattering factor.

Buried-Channel NMOS

I investigated a novel buried-channel strained-Si NMOS structure incorporating an n-type donor layer beneath the strained-Si channel to encourage occupation of the buried channel and increase the overall mobility. Peak mobility in a structure without a donor layer reproduced the best results in the literature for buried-channel strained-Si NMOS devices. For structures with donor layers, I found that the Coulomb scattering from charges in the donor layer eradicated any benefit from increased buried-channel occupation.

Surface-Channel NMOS

I also investigated the effect of well implants on the mobility of surface-channel strained-Si NMOS devices. Mobilities in unimplanted devices reproduced the highest reported in

the literature to date. I found that low field mobility degraded with increasing implant dose while high field mobility was not impaired by the well implant, similar to the universal mobility curve in bulk Si. I also found that the mobility was not degraded by a neutral implant species at the same dose. This leads to the conclusion that the material quality of the strained-layer is not affected by the implant, and that the mobility degradation is due solely to increased ionized impurity scattering.

8.2 Future Work

There are many issues opened up by this thesis that deserve further exploration.

The simulation work of Chapter 4 was carried out at 0.2 μm channel length, which has become outdated. This work should be extended to channel lengths of 0.1 and 0.05 μm. The more accurate band offset and mobility information currently available should be incorporated into the simulations. MEDICI has some convergence problems using hydrodynamic solutions with high-mobility structures, so a more robust simulator should be used. The ability to implement arbitrary mobility models in the simulator is also required. More accurate calibration of the energy relaxation time is needed to properly predict the amount of velocity overshoot.

As for the optimum layer structure, based on this thesis, it has become clear that donor layers are difficult to work with and provide no significant advantage in Si/SiGe NMOS devices. This statement applies only to enhancement-mode devices intended for VLSI applications, where the donor layer must be fully depleted at all times and thus placed very close to and underneath the strained-Si channel. Such modulation doping may still find use in analog applications. Buried channel structures are also difficult to work with due to the strong dependence of device performance on the thickness of the thin Si cap layer. Instead, future work should focus on the surface-channel strained-Si configuration, which exhibits quite high electron mobilities, or the buried/surface-channel CMOS suggested in Fig. 3.5(c).

I would suggest that the following issues be pursued using surface-channel strained-Si wafers or wafers of the proposed structure in Fig. 3.5(c) in conjunction with the short-flow MOSFET process developed in this thesis:

- substrate Ge content dependence of mobility
- thermal budget
- junction formation in relaxed-SiGe
- silicidation of strained-Si
- integration with silicon-on-insulator substrates
- layer thickness optimization

Conduction and valence band offsets can be explored with a thin-oxide MOSFET or MOS capacitor in conjunction with C-V on NMOS and PMOS devices. Short-channel MOSFETs can be used to explore velocity saturation and velocity overshoot.

Since the relaxed-SiGe buffer is required for a realistic implementation of SiGe-based CMOS, the quality of the buffer must be improved before mainstream integration of SiGe materials is possible. This includes the following growth-related issues:

- reduction of surface cross-hatch
- reduction of defect density

References

[1] K. Ismail, S. F. Nelson, J. O. Chu and B. S. Meyerson, "Electron transport properties in Si/SiGe heterostructures: measurements and device implications," Appl. Phys. Lett. **63**, 660 (1993).

[2] K. Ismail, J. O. Chu and B. S. Meyerson, "High hole mobility in SiGe alloys for device applications," Appl. Phys. Lett. **64**, 3124 (1994).

[3] R. A. Kiehl, P. M. Solomon and D. J. Frank, "Heterojunction FETs in III-V compounds," IBM J. Res. Develop. **34**, 506 (1990).

[4] D. K. Nayak, J. C. S. Woo, J. S. Park, K.-L. Wang and K. P. McWilliams, "Enhancement-mode quantum-well GeSi PMOS," IEEE Electron Device Lett. **12**, 154 (1991).

[5] V. P. Kesan, S. Subbanna, P. J. Restle, M. J. Tejwani, J. M. Altken, S. S. Iyer and J. A. Ott, "High performance 0.25 μm p-MOSFETs with silicon-germanium channels for 300K and 77K operation," International Electron Devices Meeting Technical Digest, 25 (1991).

[6] K. Bhaumik, Y. Shacham-Diamand, J. P. Noel, J. Bevk and L. C. Feldman, "23 GHz f_T room temperature SiGe quantum well p-MOSFETs," International Semiconductor Device Research Symposium, 349 (1993).

[7] J. Welser, J. L. Hoyt and J. F. Gibbons, "NMOS and PMOS transistors fabricated in strained silicon/relaxed silicon-germanium structures," International Electron Devices Meeting Technical Digest, 1000 (1992).

[8] J. Welser, J. L. Hoyt and J. F. Gibbons, "Electron mobility enhancement in strained Si *n*-type metal-oxide-semiconductor field-effect transistors", IEEE Electron Device Lett. **15**, 100 (1994).

[9] S. Verdonckt-Vanderbroek, E. F. Crabbe, B. S. Meyerson, D. L. Harame, P. J. Restle, J. M. C. Stork and J. B. Johnson, "SiGe channel heterojunction *p*-MOSFETs", IEEE Trans. Electron Devices **41**, 90 (1994).

[10] MEDICI User's Manual, TMA Associates (1992).

[11] B. S. Meyerson, K. Uram and F. K. LeGoues, "Cooperative growth phenomena in silicon/germanium low-temperature epitaxy," Appl. Phys. Lett. **53**, 2555 (1988).

[12] J. W. Matthews, *Epitaxial Growth Part B*, Academic Press (1975).

[13] M. M. Rieger and P. Vogl, "Electronic band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates," Phys. Rev. B **48**, 14276 (1993).

[14] G. G. Shahidi, D. A. Antoniadis and H. I. Smith, "Electron velocity overshoot at room and liquid nitrogen temperatures in Si inversion layers," IEEE Electron Device Lett. **9**, 94 (1988).

[15] T. Yamada, J. R. Zhou, H. Miyata and D. K. Ferry, "In-plane transport properties of Si/SiGe structure and its FET performance by computer simulation," IEEE Trans. Electron Devices **41**, 1513 (1994).

[16] M. R. Pinto, E. Sangiorgi and J. Bude, "Silicon MOS transconductance scaling in the overshoot regime," IEEE Electron Device Lett. **14**, 375 (1993).

[17] M. Iwase, T. Mizuno, M. Takahashi, H. Niiyama, M. Fukumoto, K. Ishida, S. Inaba, Y. Takigami, A. Sanda, A. Toriumi and M. Yoshimi, "High-performance 0.10 μm CMOS devices operating at room-temperature" IEEE Electron Device Lett. **14**, 51 (1993).

[18] H. Hanafi, P. Coane, A. Dally, T. Lii and D. Mastin, "The design, fabrication and characterization of 0.15 μm MOS devices," European Solid State Device Research Conference, (1993).

[19] W.-H. Chang, B. Davari, M. R. Wordeman, Y. Taur, C. C.-H. Hsu and M. D. Rodriguez, "A high-performance 0.25 μm CMOS technology: I - Design and characterization," IEEE Trans. Electron Devices **39**, 959 (1992).

[20] H. Hu, J. Jacobs, L. T. Su and D. A. Antoniadis, "A study of deep-submicron MOSFET scaling based on experiment and simulation," IEEE Trans. Electron Devices **42**, 669 (1995).

[21] HSPICE User's Manual, Meta-Software (1992).

[22] K. Ismail, J. O. Chu, K. L. Saenger, B. S. Meyerson and W. Rausch, "Modulation-doped *n*-type Si/SiGe with inverted interface," Appl. Phys. Lett. **65**, 1248 (1994).

[23] S. M. Sze, *Physics of Semiconductor Devices*, Wiley, 8 (1981).

[24] E. A. Fitzgerald, "GeSi/Si nanostructures," Annu. Rev. Mater. Sci. **25,** 417 (1995).

[25] S. S. Iyer, G. L. Patton, J. C. M. Stork, B. S. Meyerson, and D. L. Harame, "Heterojunction bipolar transistors using Si-Ge alloys," IEEE Trans. Electron Devices **36,** 2043 (1989).

[26] B. S. Meyerson, "UHV/CVD growth of Si and Si:Ge alloys: chemistry, physics, and device applications," Proc. IEEE **80,** 1592 (1992).

[27] D. K. Nayak, K. Goto, A. Yutani, J. Murota and Y. Shiraki, "High-mobility strained-Si PMOSFET's," IEEE Trans. Electron Devices **43,** 1709 (1996).

[28] K. Bhaumik, Y. Shacham-Diamand, J. P. Noel, J. Bevk and L. C. Feldman, "Theory and observation of enhanced, high field hole transport in $Si_{1-x}Ge_x$ quantum well p-MOSFET's," IEEE Trans. Electron Devices **43,** 1965 (1996).

[29] V. I. Kuznetsov, R. v. Veen, E. van der Drift, K. Werner, A. H. Verbruggen, and S. Radelaar, "Technology for high-performance *n*-channel SiGe modulation-doped field-effect transistors," J. Vac. Sci. Technol. B **13,** 2892 (1995).

[30] M. A. Armstrong, D. A. Antoniadis, A. Sadek, K. Ismail, and F. Stern, "Design of Si/SiGe heterojunction complementary metal-oxide-semiconductor transistors," IEEE International Electron Devices Meeting Technical Digest, 761 (1995).

[31] M. J. Hargrove, A. K. Henning, J. A. Slinkman, and J. C. Bean, "Quantum mechanical modeling of the charge distribution in a $Si/Si_{1-x}Ge_x/Si$ p-channel MOSFET," IEEE International Electron Devices Meeting Technical Digest, 735 (1994).

[32] S. P. Voinigescu, P. B. Rabkin, C. A. T. Salama, and P. A. Blakey, "2D numerical investigation of the impact of compositional grading on the performance of submicrometer Si-SiGe MOSFET's," IEEE Trans. Electron Devices **42,** 1039 (1995).

[33] M. Rashed, W.-K. Shih, S. Jallepalli, T. J. T. Kwan, and C. M. Maziar, "Monte Carlo simulation of electron transport in strained $Si/Si_{1-x}Ge_x$ n-MOSFETs," IEEE International Electron Devices Meeting Technical Digest, 765 (1995).

[34] B. Alamariu, private communication.

[35] S. Wolf and R. N. Tauber, *Silicon Processing for the VLSI Era: Volume 1 – Process Technology*, Lattice Press, 517 (1986).

[36] J. W. Mayer and S. S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs*, Macmillan, 330 (1990).

[37] M. S. Naidu and V. Kamaraju, *High Voltage Engineering*, McGraw-Hill, 25 (1982).

[38] T. Hattori, ed., *Ultraclean Surface Processing of Silicon Wafers: Secrets of VLSI Manufacturing*, 451 and 482 (1998).

[39] S. C. Sun and J. D. Plummer, "Electron mobility in inversion and accumulation layers on thermally oxidized silicon surfaces," IEEE Trans. Electron Devices **27**, 1497 (1980).

[40] M.-S. Liang, J. Y. Choi, P.-K. Ko and C. Hu, "Inversion-layer capacitance and mobility of thin gate-oxide MOSFET's," IEEE Trans. Electron Devices **33**, 409 (1986).

[41] S. M. Prokes, O. J. Glembocki, M. E. Twigg, and K. L. Wang, "The study of relaxation in asymmetrically strained $Si_{1-x}Ge_x$/Si superlattices," J. Electronic Materials **20**, 389 (1991).

[42] D. C. Houghton, "Strain relaxation kinetics in $Si_{1-x}Ge_x$/Si heterostructures," J. Appl. Phys. **70**, 2136 (1991).

[43] J. Lee, C. Hegarty, and C. Hu, "Electrical characteristics of MOSFET's using low-pressure chemical vapor deposited oxide," IEEE Electron Device Lett. **9**, 324 (1988).

[44] F. K. LeGoues, R. Rosenberg, T. Nguyen, F. Himpsel, and B S. Meyerson, "Oxidation studies of SiGe," J. Appl. Phys. **65**, 1724 (1989).

[45] J. Welser, *The Application of Strained-Silicon/Relaxed-Silicon Germanium Heterostructures to Metal-Oxide-Semiconductor Field-Effect Transistors*, Ph.D. thesis Stanford University (1994).

[46] K. Rim, J. L. Hoyt and J. F. Gibbons, "Transconductance enhancement in deep submicron strained-Si *n*-MOSFETs," International Electron Devices Meeting Technical Digest, (1998).

[47] R. Hagelauer, T. Ostermann, U. Konig, M. Gluck and G. Hock, "Performance estimation of Si/SiGe hetero-CMOS circuit," Electronics Lett. **33**, 208 (1997).

[48] U. Konig, M. Gluck, A. Gruhle, G. Hock, E. Kohn, B. Bozon, D. Nuernbergk, T. Ostermann and R. Hagelauer, "Design rules for n-type SiGe hetero FETs," Solid-State Electronics **41**, 1541 (1997).

[49] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, (1987).

[50] R. Braunstein, A. R. Moore and F. Herman, "Intrinsic optical absorption in germanium-silicon alloys," Phys. Rev. **109**, 695 (1958).

[51] C. A. King, J. L. Hoyt and J. F. Gibbons, "Bandgap and transport properties of $Si_{1-x}Ge_x$ by analysis of nearly ideal $Si/Si_{1-x}Ge_x/Si$ heterojunction bipolar transistors," IEEE Trans. Electron Devices **36**, 2093 (1989).

[52] R. People and J. C. Bean, "Band alignments of coherently strained Ge_xSi_{1-x}/Si heterostructures on <001> Ge_ySi_{1-y} substrates," Appl. Phys. Lett. **48**, 538 (1986).

[53] F. Schaffler, "High-mobility Si and Ge structures," Semicond. Sci. Technol **12**, 1515 (1997).

[54] D. V. Lang, R. People, J. C. Bean and A. M Sergent, "Measurement of the band gap of Ge_xSi_{1-x}/Si strained-layer heterostructures," Appl. Phys. Lett. **47**, 1333 (1985).

[55] R. People and J. C. Bean, "Calculation of critical layer thickness versus lattice mismatch for Ge_xSi_{1-x}/Si strained layer heterostructures," Appl. Phys. Lett. **47**, 322 (1985).

[56] E. A. Fitzgerald, Y. H. Xie, M. L. Green, D. Brasen, A. R. Kortan, J. Michel, Y. J. Mii, B. E. Weir, "Totally relaxed Ge_xSi_{1-x} layers with low threading dislocation densities grown on Si substrates," Appl. Phys. Lett. **59**, 811 (1991).

[57] C. G. Van de Walle, R. M. Martin, "Theoretical study of Si/Ge interfaces" J. Vac. Sci. Technol. B **3**, 1256 (1985).

[58] M. V. Fischetti and S. E. Laux, "Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys," J. Appl. Phys. **80**, 2234 (1996).

[59] M. Glicksman, "Mobility of electrons in Ge-Si alloys," Phys. Rev. **111**, 125 (1958).

[60] G. Busch and O. Vogt, "Elektrische Leitfähigkeit and Halleffekt von GeSi-Legierungen," Helv. Phys. Acta **33**, 437 (1960).

[61] F. M. Bufler and B. Meinerzhagen, "Hole transport in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates," J. Appl. Phys. **84**, 5597 (1998).

[62] T. Manku, J. M. McGregor, A. Nathan, D. J. Roulston, J.-P. Noel and D. C. Houghton, "Drift hole mobility in strained and unstrained doped $Si_{1-x}Ge_x$ alloys," IEEE Trans. Electron Devices **40**, 1990 (1993).

[63] A. Levitas, "Electrical properties of germanium-silicon alloys," Phys. Rev. **99**, 1810 (1955).

[64] P. Gaworzewski, K. Tittelbach-Helmrich and U. Penner, "Electrical properties of lightly doped *p*-type silicon-germanium single crystals," J. Appl. Phys. **83**, 5258 (1998).

[65] T. K. Carns, S. K. Chun, M. O. Tanner, K. L. Wang, T. I. Kamins, J. E. Turner, D. Y. C. Lie, M.-A. Nicolet and R. G> Wilson, "Hole mobility measurements in heavily doped $Si_{1-x}Ge_x$ strained layers," IEEE Trans. Electron Devices **41**, 1273 (1994).

[66] S. K. Chun and K. L. Wang, "Effective mass and mobility of holes in strained $Si_{1-x}Ge_x$ layers on (001) $Si_{1-y}Ge_y$ substrate," IEEE Trans. Electron Devices **39**, 2153 (1992).

[67] S. P. Voinigescu, C. A. T. Salama, J.-P. Noel and T. I. Kamins, "Optimized Ge channel profiles for VLSI compatible Si/SiGe p-MOSFETs," International Electron Devices Meeting Technical Digest, 369 (1994).

[68] R. J. P. Lander, G. J. Emeleus, B. M. McGregor, E. H. C. Parker, T. E. Whall, A. G. R. Evans and G. P. Kennedy, "Study of Hall and effective mobilities in pseudomorphic $Si_{1-x}Ge_x$ *p*-channel metal-oxide-semiconductor field-effect transistors at room temperature and 4.2 K," J. Appl. Phys. **82**, 5210 (1997).

[69] B. M. M. McGregor, R. J. P. Lander, P. J. Philips, E. H. C. Parker and T. E. Whall, "Temperature-dependent Hall scattering factor and drift mobility in remotely doped Si:B/SiGe/Si heterostructures," Appl. Phys. Lett. **74**, 1245 (1999).

[70] D. J. Paul, N. Griffin, D. D. Arnone, M. Pepper, C. J. Emeleus, P. J. Philips and T. E. Whall, "Low temperature characterization of modulation doped SiGe grown on bonded silicon-on-insulator," Appl. Phys. Lett. **69**, 2704 (1996).

[71] K. Goto, J. Murota, T. Maeda, R. Schutz, K. Aizawa, R. Kircher, K. Yokoo and S. Ono, "Fabrication of a $Si_{1-x}Ge_x$ channel metal-oxide-semiconductor field-effect transistor (MOSFET) containing high Ge fraction layer by low-pressure chemical vapor deposition," Jpn. J. Appl. Phys. **32**, 438 (1993).

[72] S. Verdonckt-Vandebroek, E. F. Crabbe, B. S. Meyerson, D. L. Harame, P. J. Restle, J. M. C. Stork, A. C. Megdanis, C. L. Stanis, A. A. Bright, G. M. W. Kroesen and A. C. Warren, "High-mobility modulation-doped graded SiGe-channel p-MOSFETs," IEEE Electron Device Lett. **12**, 447 (1991).

[73] P. M. Garone, V. Venkataraman and J. C. Sturm, "Hole mobility enhancement in MOS-gated Ge_xSi_{1-x}/Si heterostructure inversion layers," IEEE Electron Device Lett. **13**, 56 (1992).

[74] P. J. Briggs, A. B. Walker and D. C. Herbert, "Calculation of hole mobilities in relaxed and strained SiGe by Monte Carlo simulation," Semicond. Sci. Technol. **13**, 680 (1998).

[75] T. Manku and A. Nathan, "Electron drift mobility model for devices based on unstrained and coherently strained $Si_{1-x}Ge_x$ grown on <001> silicon substrate," IEEE Trans. Electron Devices **39**, 2082 (1992).

[76] T. Manku and A. Nathan, "Effective mass for strained p-type $Si_{1-x}Ge_x$," J. Appl. Phys. **69**, 8414 (1991).

[77] U. Konig and F. Schaffler, "p-type Ge-channel MODFETs with high transconductance grown on Si substrates," IEEE Electron Device Lett. **14**, 205 (1993).

[78] M. Arafa, P. Fay, K. Ismail, J. O. Chu, B. S. Meyerson and I. Adesida, "DC and RF performance of 0.25 μm p-type SiGe MODFET," IEEE Electron Device Lett. **17**, 449 (1996).

[79] Y. C. Chen, S. H. Li, P. K. Bhattacharya, J. Singh and J. M. Hinckley, "Direct measurement of the Hall factor for holes in relaxed $Si_{1-x}Ge_x$ ($0 < x < 1$)," Appl. Phys. Lett. **64**, 3110 (1994).

[80] T. Vogelsang and K. R. Hofmann, "Electron transport in strained Si layers on $Si_{1-x}Ge_x$ substrates," Appl. Phys. Lett. **63**, 186 (1993).

[81] B. Fischer and K. R. Hoffman, "Full-band Monte Carlo model of electron and hole transport in strained Si including inelastic acoustic phonon scattering," Appl. Phys. Lett. **74**, 2185 (1999).

[82] J. Welser J. L. Hoyt, S. Takagi and J. F. Gibbons, "Strain dependence of the performance enhancement in strained-Si *n*-MOSFETs," International Electron Devices Meeting Technical Digest, 373 (1994).

[83] M. Gluck, T. Hackbarth, U. Konig, A. Haas, G. Hock and E. Kohn, "High f_{max} *n*-type Si/SiGe MODFETs," Electronics Lett. **33**, 335 (1997).

[84] S. J. Koester, J. O. Chu and R. A. Groves, "High-f_T *n*-MODFETs fabricated on Si/SiGe heterostructures grown by UHV-CVD," Electronics Lett. **35**, 86 (1999).

[85] K. Rim, J. Welser, J. L. Hoyt and J. F. Gibbons, "Enhanced hole mobilities in surface-channel strained-Si *p*-MOSFETs," International Electron Devices Meeting Technical Digest, 517 (1995).

[86] F. Gamiz, J. B. Roldan, J. A. Lopez-Villanueva and P. Cartujo, "Coulomb scattering in strained-silicon inversion layers on $Si_{1-x}Ge_x$ substrates," Appl. Phys. Lett. **69**, 797 (1996).

[87] H. Miyata, T. Yamada and D. K. Ferry, "Electron transport properties of a strained Si layer on a relaxed $Si_{1-x}Ge_x$ substrate by Monte Carlo simulation," Appl. Phys. Lett. **62**, 2661 (1993).

[88] D. K. Nayak, K. Kamjoo, J. S. Park, J. C. S. Woo and K. L. Wang, "Rapid isothermal processing of strained GeSi layers," IEEE Trans. Electron Devices **39**, 56 (1992).

[89] M. C. Ozturk, J. J. Wortman, C. M. Osburn, A. Ajmera, G. A. Rozgonyi, E. Frey, W.-K. Chu and C. Lee, "Optimization of the germanium preamorphization conditions for shallow-junction formation," IEEE Trans. Electron Devices **35**, 659 (1988).

[90] T. Nitta, T. Ohmmi, Y. Ishihara, A. Okita, T. Shibata, J. Sugiura and N. Ohwada, "Reverse-bias current reduction in low-temperature-annealed silicon *pn* junctions by ultraclean ion-implantation technology," J. Appl. Phys. **67**, 7404 (1990).

[91] A. J. Walker, P. H. Woerlee, H. G. Pomp and N. E. B. Cowern, "Shallow boron junctions and preamorphization for deep submicron silicon technology," J. Appl. Phys. **73**, 4048 (1993).

[92] J. Kato, "The annealing time and temperature dependence of electrical dopant activation in high-dose BF_2 ion implanted silicon," J. Electrochem. Soc. **141**, 3158 (1994).

[93] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*, Wiley, (1982).

[94] S. Takagi, A. Toriumi, M. Iwase and H. Tango, "On the universality of inversion layer mobility in Si/MOSFET's: part I - effects of substrate impurity concentration," IEEE Trans. Electron Devices **41**, 2357 (1994).

[95] E. Murakami, N. Nakagawa, A. Nishida and M. Miyao, "Strain-controlled Si-Ge modulation-doped FET with ultrahigh hole mobility," IEEE Electron Device Lett. **12**, 71 (1991).

Appendix A

Low Temperature Process Technology

A.1 Introduction

One of the main advantages of SiGe epitaxial technology is compatibility with well-developed, mainstream Si processing technology. However, there are some SiGe-specific issues which must be addressed stemming from the thermal instability of epitaxially-grown SiGe heterostructures. The high mobilities observed in SiGe layers depend on the layers being strained and dislocation free. Exposure to high temperatures can degrade mobilities both through strain-relaxation and Ge diffusion. At temperatures below 950 ^{o}C Ge diffusion dominates [41] resulting in a spreading of the layers and a reduction of both strain and interface quality. At higher temperatures, relaxation through dislocation nucleation and propagation occurs [42], resulting in a reduction in strain and increase in dislocation density.

Layers which are metastable, meaning pseudomorphic but grown at low temperatures to a thickness beyond the critical thickness, are particularly sensitive to thermal strain relaxation. For instance, Kuznetsov [29] observed a degradation in n-MOSFET mobility at temperatures above 700 ^{o}C commensurate with a relaxation of his strained-layer, but his layer thickness of 200 Å was above the practical critical thickness of strained-Si on

relaxed-$Si_{0.7}Ge_{0.3}$. On the other hand, Nayak [88] observed no degradation in strain of a 500 Å strained-$Si_{0.8}Ge_{0.2}$ layer on relaxed-Si for an RTA of 1000 °C for 10 s, indicating the strain-stability of pseudomorphic layers with thickness below the critical thickness. However, in heterostructures with layer thicknesses on the order of 50 Å, Ge diffusion is still a problem.

Thus, the total thermal budget for a fabrication process for SiGe MOSFETs is limited. This has implications for several fabrication steps such as implant activation and gate oxidation which traditionally require high temperatures. These problem areas are illustrated in Fig. A.1. New alternatives will have to be developed for the modules. Low temperature alternatives for device isolation, such as shallow-trench isolation (STI), are already being developed in industry. In the following we will focus on source/drain junctions and gate oxide formation.

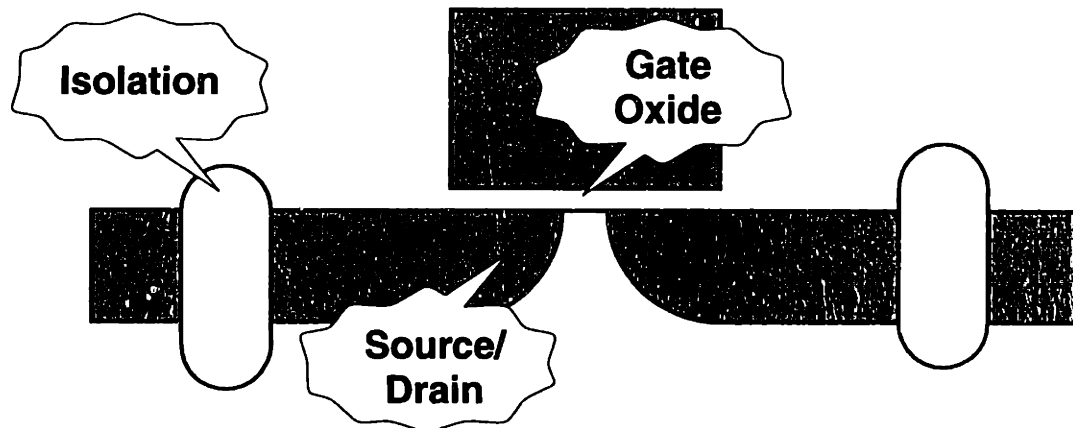


Fig. A.1. Problem areas for process integration of SiGe materials.

A.2 Source/Drain Junctions

Key issues in MOSFET source/drain formation are resistivity, reverse-leakage and junction depth. Low resistivity is necessary to ensure low resistance in series with the MOSFET channel, which harms device performance. Low reverse leakage is necessary to avoid high off-state CMOS device leakage and the resulting high static power which is undesirable for low-power applications. A shallow junction is necessary to minimize charge-shar-

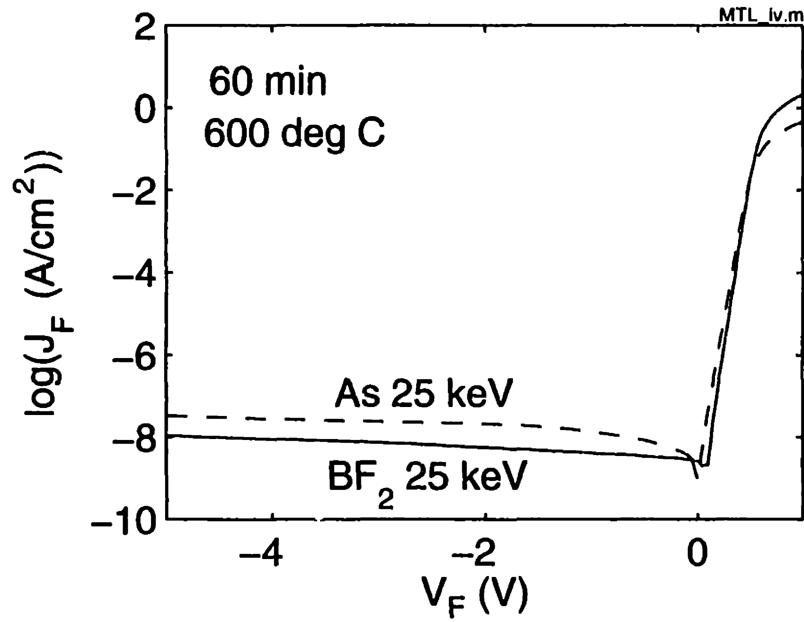
ing effects in sub-micron devices, which results in a loss of device control due to excessive drain-bias modulation of device current.

Reverse Leakage

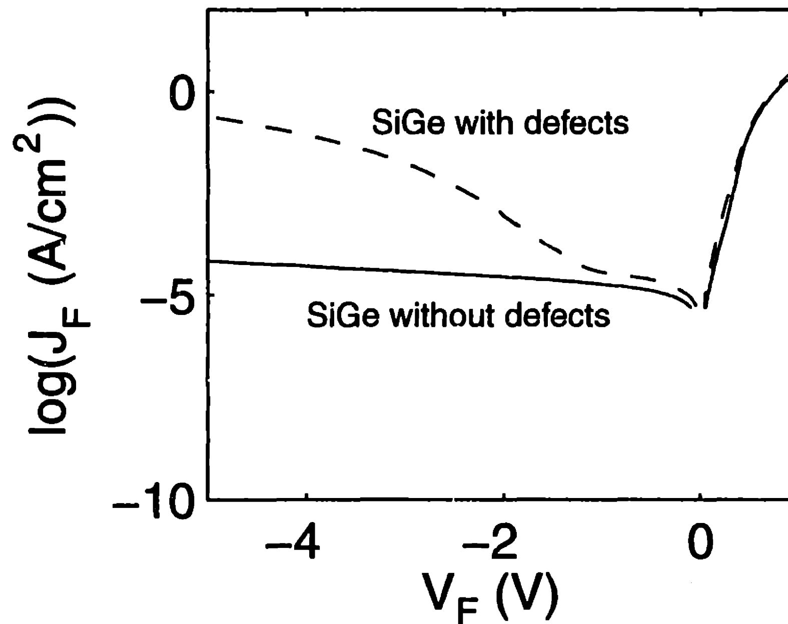
Source/drain junctions are formed by ion implantation of impurities. After implantation, the impurities occupy interstitial sites and are electrically inactive. In order to activate the dopants, a thermal anneal is needed. In modern devices, a rapid thermal anneal at around 1000 ^{o}C is used to achieve maximum activation with little diffusion. Thus, low resistivity is achieved with shallow junction depth. This anneal also serves to anneal out any residual implantation damage (displaced lattice atoms). Lattice damage attracts heavy metal impurities, which in turn serve as generation centers which can result in excessive reverse leakage if they are located in or near the depletion region of the junction.

In this work, we have sought a low-temperature technique for the formation of low-leakage, low-resistance source/drain implants. We make use of the phenomenon of solid-state recrystallization (SSR): the silicon lattice, amorphized by a high-dose implant, will recrystallize at a temperature of only around 600 ^{o}C, during which time any implanted species are incorporated substitutionally into the lattice and rendered active. The recrystallized area is of high quality except at the bottom of the implant, where it failed to completely amorphize the lattice. This transition region between amorphous and crystalline material is known as end-of-range damage, and must be kept outside the depletion region of the junction so that it does not produce leakage as a source of recombination centers.

Taking advantage of SSR, it is possible to find implant and anneal conditions which high quality, low-temperature-annealed junctions in bulk silicon. Fig. A.2(a) shows the reverse-leakage characteristics of As and BF_2-implanted junctions, suitable for use in NMOS and PMOS devices, respectively, which received only a 600 ^{o}C, 60 minute furnace anneal. Both implants were 25 keV and 1 x 10^{15} cm^{-2} and the substrates were lightly-doped (10 - 20 Ωcm). The As implant was preceded by a BF_2 implant of 50 keV and 1 x 10^{12} cm^{-2} to simulate a p-well implant and to suppress edge leakage due to the inversion of the field by



(a)



(b)

Fig. A.2. (a) Characteristics of As and BF_2-implanted (25 keV, 1 x 10^{15} cm^{-2}) junctions in bulk Si after 600 °C anneal. (b) Leakage characteristics of BF_2-implanted (25 keV, 1 x 10^{15} cm^{-2}) junctions in relaxed-$Si_{0.7}Ge_{0.3}$ after 600 °C anneal. Diodes with intersecting growth defects have more leakage.

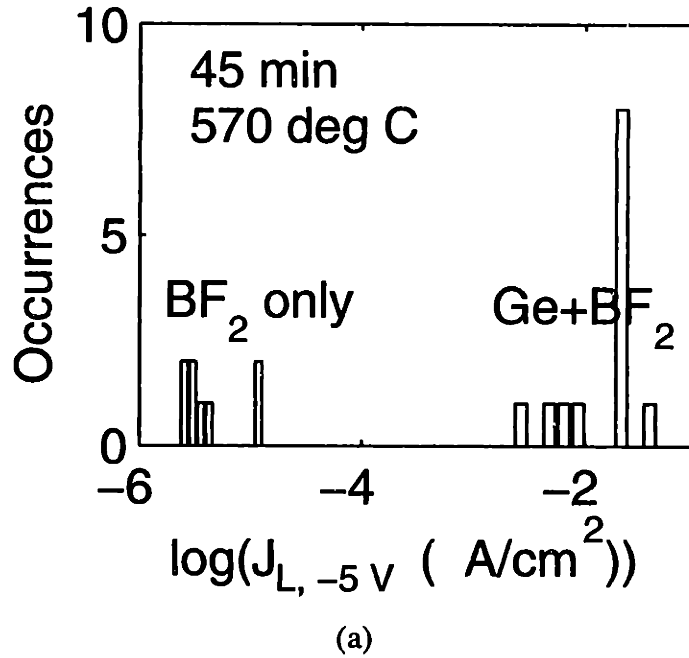
the field oxide fixed charge. The devices were sintered at 400 °C for 15 minutes in forming gas.

The reverse leakages are remarkable for this low thermal budget and comparable to the results seen for traditionally processed diodes of about 1 to 10 nA/cm^2. These results confirm previous experiments in the literature for BF_2 [91] and As [90], but without the need for ultraclean ion implantation.

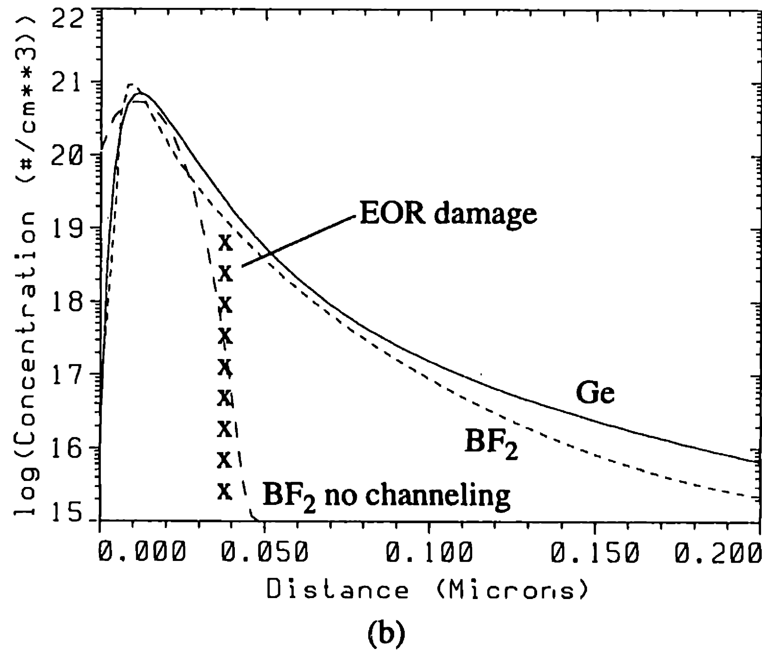
While the above results are for silicon substrates, Fig. A.2(b) shows the results of experimental BF_2 implants into relaxed-$Si_{0.7}Ge_{0.3}$ substrates. The lower-leakage curve in the figure is typical of the large area (200 μm^2) diodes on the wafer, while the higher-leakage curve is for diodes with large defects visible as pits. Given a threading dislocation density of greater than 10^5 cm^{-2} for the relaxed-$Si_{0.7}Ge_{0.3}$ virtual substrate, the diodes should on average contain several threading dislocations. Thus, it is not threading dislocations which cause catastrophic leakage in the junctions, but rather the less populous pits. Such pits attract metallic impurities which serve as recombination centers and increase reverse leakage. The pits are caused by particulate contamination before and during growth.

One way to decrease junction leakage is to use a preamorphizing implant of a heavy species which is either neutral or of the opposite type to that used for the junction implant. The purpose of the preamorphizing implant is move the end-of-range damage below the depth of the p-n junction by pushing the amorphized region to a greater depth. Fig. A.3(a) shows the results of one such experiment in the form of a histogram of leakage currents measured at 5 V reverse bias, comparing BF_2-implanted (7 keV, 9 x 10^{14}cm^{-2}) samples which did and did not receive a preamorphizing Ge implant (20 keV, 1 x 10^{15} cm^{-2}). In this case, the Ge implant apparently is not deep enough, as it only increases leakage due to unannealled end-of-range damage near the depletion region of the junction.

The situation is illustrated in Fig. A.3(b). The end-of-range (EOR) damage of a heavy implant occurs where the implant falls below roughly 10^{19} cm^{-3}, as indicated by the x marks in the figure. In the case of the BF_2 implant by itself, the EOR damage falls well above the metallurgical junction and the resultant depletion region. If the EOR damage is



(a)



(b)

Fig. A.3. (a) Histogram of leakage current at 5 V reverse bias of BF_2-implanted (7 keV, 9 x $10^{14} cm^{-2}$) diodes comparing non-preamorphized and Ge-preamorphized (20 keV, 1 x 10^{15} cm^{-2}) cases. (b) Dopant concentration vs. distance from the surface as calculated by TSUPREM4 comparing Ge, BF_2 and BF_2 without channeling. Also shown is the approximate position of end-of-range (EOR) damage.

outside the depletion region, any recombination centers it engenders will not affect the reverse leakage of the diode. However, when the BF_2 implant follows the preamorphizing Ge implant, channeling is suppressed and the implant is shallower. In addition, the EOR damage of the Ge implant is slightly deeper as it is a heavier ion. For these reasons, the EOR damage ends up inside the depletion region of the diode where recombination centers contribute to diode current very efficiently. This problem is typical for preamorphization schemes and cannot be solved by implanting the Ge at higher energy, as it is difficult to implant heavy species to large depths without producing a large amount of unannealable damage [89].

Resistivity

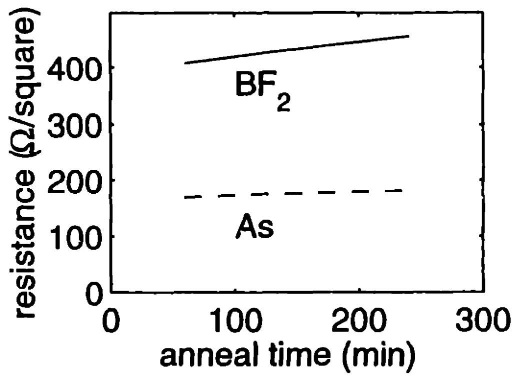
In addition to having low reverse-leakage, we require that the implanted layer present a low resistance to current flow parallel to the surface. A low S/D resistance is mandatory for HCMOS in order to observe the benefits of the high-mobility channels, as a high series resistance will degrade transconductance and maximum current drive of the devices. Fig. A.4(a) shows the sheet resistance of the implants of Fig. A.2 as a function of annealing time. Sheet resistance, which is directly proportional to impurity activation, is good for this annealing temperature due to incorporation during solid-state regrowth.

Sheet resistance, $R_{\square}$, is given by:

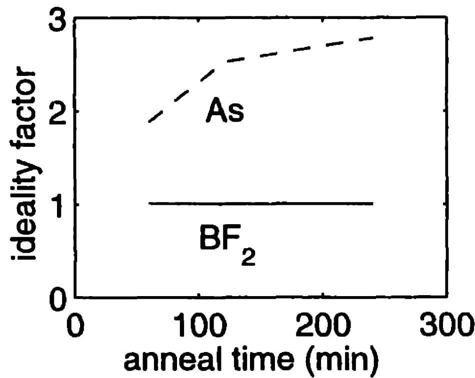
$$R_{\square} = \frac{1}{qn\mu}, \tag{1.1}$$

where q is the electronic charge, n is the carrier areal density and μ is the mobility. Assuming 100% activation ($n = 1 \times 10^{15}$ cm^{-2}) and constant mobility, though it is in fact a function of concentration and hence depth, I find $\mu = 30$ cm^2/Vs for the As implant and $\mu = 15$ cm^2/Vs for the BF_2 implant. These are of the right order for the peak doping densities of around 10^{21} cm^{-3} for these implants and indicate at least 50% activation.

Fig. A.4(a) shows how the sheet resistance increases with annealing time at 600 °C due to deactivation of carriers [92] making a shorter annealing time preferable. Ideality factor, shown in Fig. A.4(b), also degrades with increasing annealing time in the arsenic implant



(a)



(b)

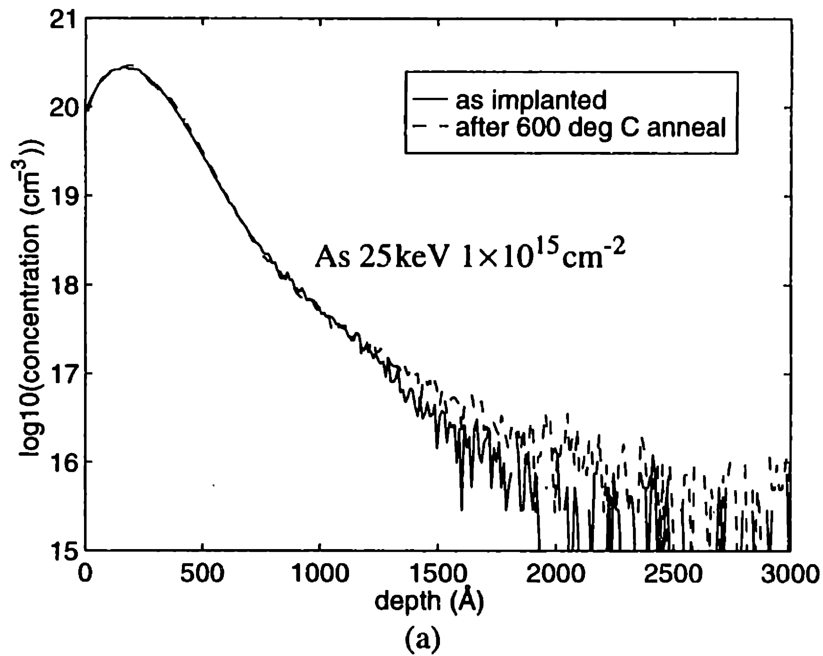
Fig. A.4. (a) Sheet resistance and (b) ideality factor vs. annealing time at 600°C of unsintered diodes.

case. This could reflect the depletion region extending into the EOR damage as the carriers deactivate.

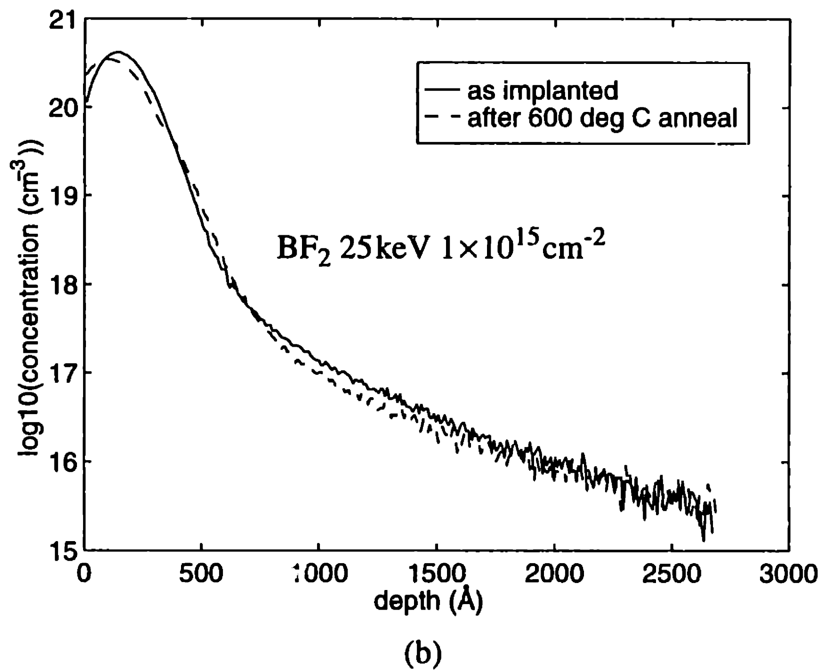
Junction Depth

Shallow junctions are required for sub-micron devices to reduce short-channel effects. A beneficial side effect of low-temperature processing is reduced dopant diffusion during the source/drain anneal, which is the main contributor to junction depth in modern devices.

Fig. A.5 shows SIMS analysis of the As and BF_2 implants of Fig. A.2, comparing the as-implanted and post-anneal profiles. The implant hardly moves during the low-temperature anneal. This is noteworthy, as it could lead to a method of forming extremely shallow junctions. Such a method could be used for improving the performance of Si- and SiGe-



(a)



(b)

Fig. A.5. SIMS profiles of (a) As and (b) BF_2 implants before and after annealing at 600 ^{o}C for 60 minutes. No implant diffusion is detectable.

based MOSFETs alike. Indeed, low-temperature processing leads the way to many possibilities, such as metal gates [90].

A.3 Gate Oxide

The gate stack is another area of device fabrication which traditionally requires high processing temperatures. Key issues for gate oxide formation are thickness, breakdown voltage, interface states and fixed charge.

Submicron devices require oxides less than 20 Å to maintain gate control of the channel. The drive to thinner oxides has led to lower and lower thermal oxidation temperatures, increasing compatibility with the limited thermal budget of SiGe heterostructures. Traditional processing uses thermal oxidation at 750 or 800 °C to form such thin oxides. However, oxide quality is reduced as the oxidation temperature is lowered. For this reason, deposited oxides have been explored as an alternative [43] taking advantage of the lower pinhole density in such oxides and the resultant higher device yield.

Deposited oxides have the beneficial side effect of not consuming silicon to produce the oxide. Recall that the poor characteristics of thermally oxidized SiGe [44] mandates a Si cap layer for SiGe heterostructure-based MOS devices. Using a deposited oxide eliminates the need to grow an extra thickness of Si cap to compensate for loss during oxidation.

Thin Oxide Deposition

This section reports on the development of a recipe optimized for slow deposition of undoped LTO in the Integrated Circuits Laboratory at MIT. The deposition rate was slowed to enable formation of thin gate oxides by reducing the gas flow rates and the ratio of silane to oxygen.

Fig. A.6 shows oxide thickness vs. deposition time for flow rates of 10:30 sccm silane:oxygen and deposition temperature of 450 °C. Thickness was unchanged by an 18

hour anneal at 600 °C in N_2. Deposition rate is about 19 Å/minute for thicker films. A 56 Å film was deposited in 1.5 minute with excellent uniformity and defect density.

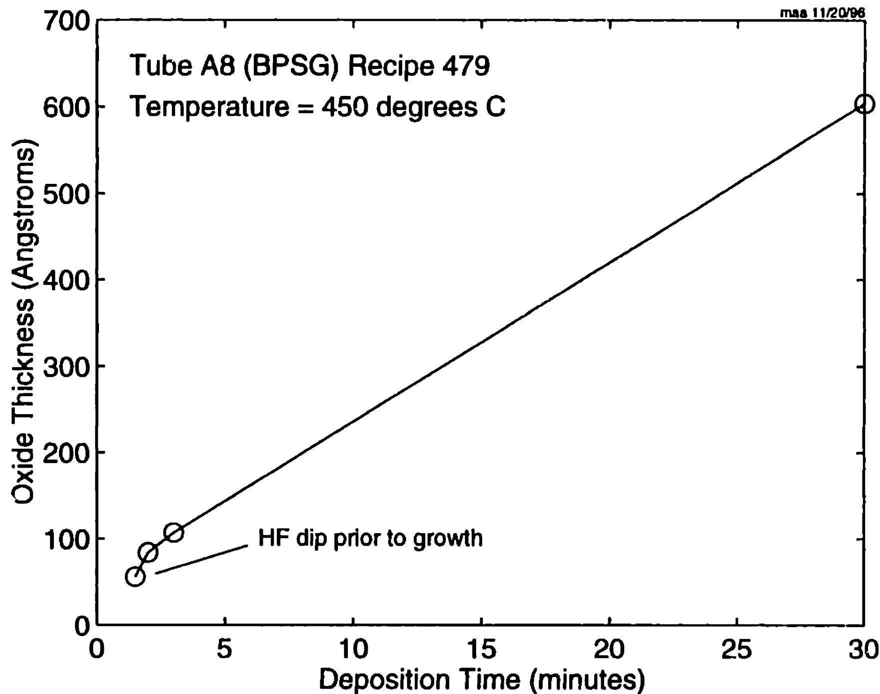


Fig. A.6. Oxide thickness vs. deposition time for slow LTO deposition recipe. Flow ratio was 1:3 silane:oxygen Wafers were uncleaned new wafers except for the first point, which received a piranha clean and HF dip.

Electrical Characterization

Aluminum gate MOS capacitors were fabricated to evaluate the electrical quality of our deposited oxides in comparison to thermal oxide. P-type wafers were piranha cleaned (3:1 H_2SO_4:H_2O_2) and HF dipped before growth. The thermal oxide was grown at 900 °C in dry O_2 for 18 minutes resulting in 88 Å thickness. The LTO oxide was deposited at 450 °C for 1.5 minutes resulting in 56 Å thickness. The LTO oxide was then annealed in N_2 at 600 °C for 48 hours. Shorter anneal times (1 hour) were found to produce similar results. After metal sputter deposition and patterning, the wafers were sintered at 400 °C for 15 minutes.

Fig. A.7 shows the high-frequency and quasistatic C-V characteristics of the aluminum gate MOS capacitors both before and after sintering. Notice the large amount of interface states present before sintering, visible as the upward shift and spreading of the quasistatic

C-V. The LTO oxide has a higher interface state density before sintering than the thermal oxide. However, after sintering the interface state density in both cases is reduced below detectable levels. The C-V characteristics of the 56 Å LTO capacitor are acceptable for MOS applications.

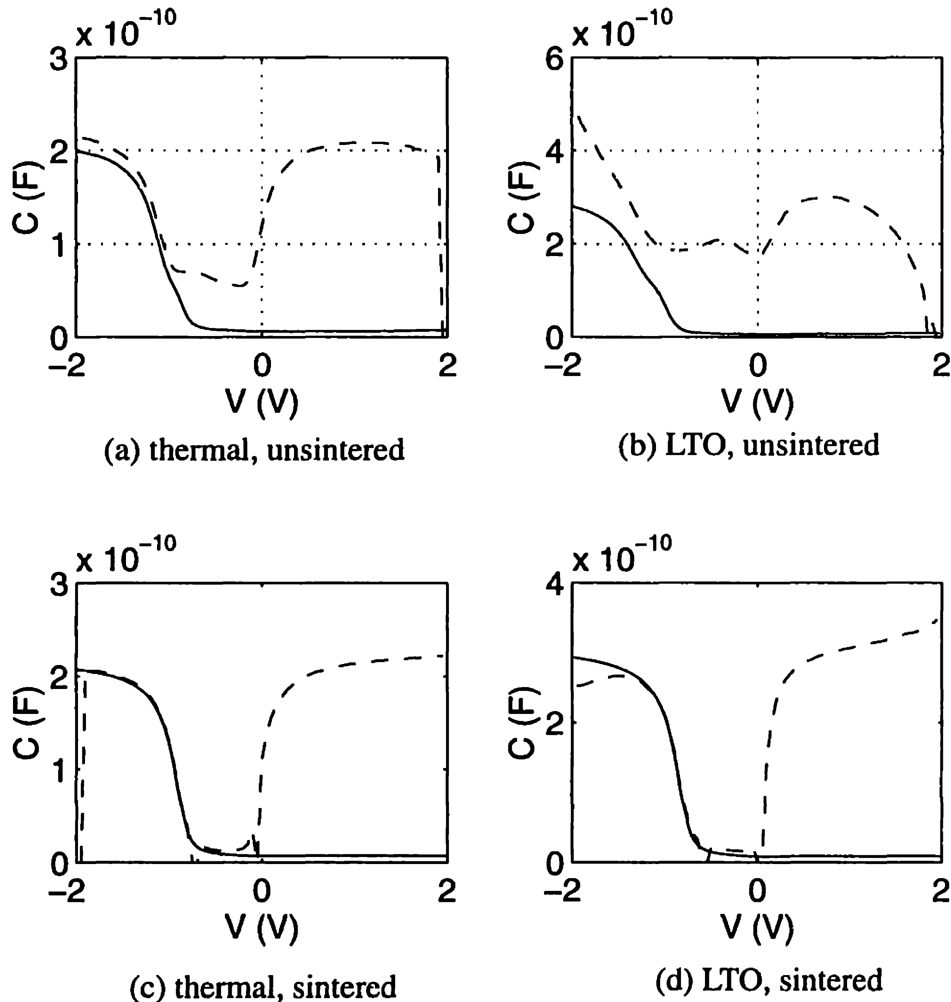


Fig. A.7. C-V plots for Al-gate capacitors: (a) thermal oxide before sinter, (b) LTO before sinter, (c) thermal oxide after sinter and (d) LTO after sinter. Solid line is high-frequency and dashed is quasistatic. There is some surface leakage visible in the quasi-static C-V plot for the LTO sample.

In addition, time-zero dielectric breakdown (TZDB) measurements show a breakdown field of above 10 MV/cm for the 56 Å LTO before sintering, which is comparable to thermal oxide.

Fig. A.8 shows the results of interface state density extraction using the conductance method [93] on short-flow MOSFETs with thick LTO oxides. This LTO was deposited using the standard recipe at 400 °C and 100 Å/s. Three cases are compared: (a) LTO on bulk Si, (b) LTO on 10 Å strained-Si on relaxed-$Si_{0.7}Ge_{0.3}$ (more precisely the HNMOS-15 structure of Fig. 6.6) and (c) LTO on relaxed-$Si_{0.7}Ge_{0.3}$.

Interface state density D_{it} is extracted at a given frequency from the maximum in the parallel conductance G_p using:

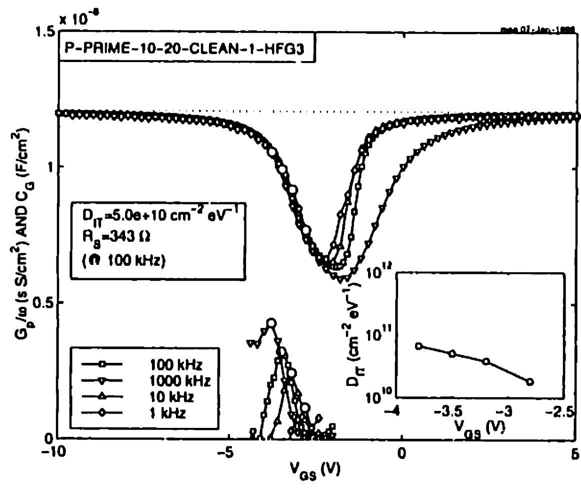
$$D_{it} = \frac{G_p}{q\omega f_D} \tag{1.2}$$

where q is the electronic charge, ω is the angular frequency and f_D is a function of the width of the conductance peak. f_D varies between 0.4 for a standard deviation $\sigma = 0$ to 0.1 for $\sigma > 5$. In other words, for a given peak height, the wider the peak, the larger the D_{it}. The data in the figure was derived assuming the ideal case with $f_D = 0.4$.

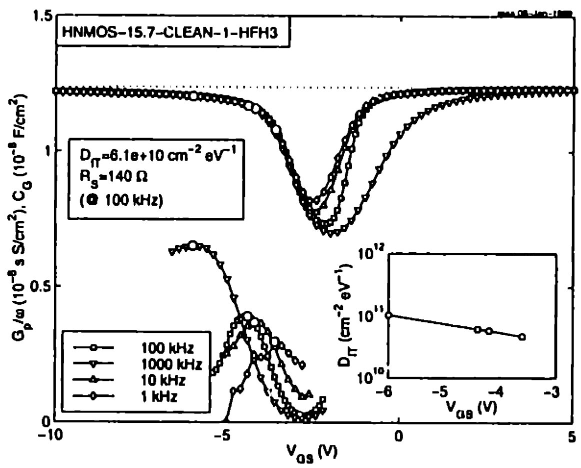
For LTO on bulk Si the minimum extracted D_{it} is 2×10^{10} cm^{-2} eV^{-1}. This number is comparable to thermal oxides and indicates the high quality of the interface. In this case the conductance peaks are very narrow and the assumption of $f_D = 0.4$ is valid.

For LTO on strained-Si the minimum extracted D_{it} is 5×10^{10} cm^{-2} eV^{-1}. This is comparable to the bulk-Si case, but some caution must be used in comparing the numbers, as the conductance peaks in this case are much wider. In fact, in order to fit the high-frequency C-V curve to simulation I have to use $D_{it} = 1 \times 10^{11}$ cm^{-2} eV^{-1} which indicates $f_D = 0.2$.

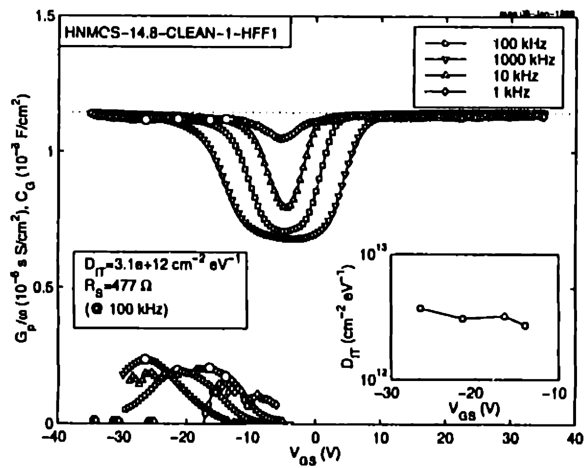
For LTO on SiGe the extracted D_{it} values are not reliable due to the large width of the peaks. However, it is clear that D_{it} is very large and close to 10^{13} cm^{-2} eV^{-1}. This sample was the result of a full 10 minute RCA clean while the previous case was the result of a shortened 30 second RCA clean. Comparing this case and the previous case we see how such a minute thickness of Si cap can drastically alter the oxide interface properties. This underscores the importance of proper cleaning procedures for SiGe heterostructure-based MOS devices.



(a) LTO on bulk Si



(b) LTO on strained-Si



(c) LTO on SiGe

Fig. A.8. Interface state density extraction using the conductance method for (a) 2869 Å LTO on bulk Si, (b) 2790 Å LTO on 10 Å strained-Si on relaxed-$Si_{0.7}Ge_{0.3}$ and (c) 3018 Å LTO on relaxed-$Si_{0.7}Ge_{0.3}$. Width of the conductance peak was not considered.