Design and Characterization of Si/SiGe heterostructure sub-100 nm bulk p-MOSFET

by

Jae-kyu Lee

B.S., Kyungpook National University (1995)
M.S., Pohang University of Science and Technology (1997)

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

June 2009

© Massachusetts Institute of Technology 2009. All right reserved.

Author

Department of Electrical Engineering and Computer Science

March 13, 2009

Certified by

Dimitri A. Antoniadis
Professor of Electrical Engineering
Thesis Supervisor

Accepted by

Terry P. Orlando
Chairman, Department committee on Graduate Students
Design and Characterization of Si/SiGe Heterostructure sub-100 nm bulk p-MOSFET

by

Jae-kyu Lee

Submitted to the Department of Electrical Engineering and Computer Science on March 13, 2009, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

As the gate length of CMOS device is scaled down to the sub-100 nanometer node, the development of devices faces many technological challenges, which are related to material and process integration. As a new channel material, compressively strained SiGe layer grown directly on the bulk Si is attractive for the p-MOSFET because of its integration compatibility with the Si-based process. The goal of this thesis is to design and fabricate bulk Si/SiGe heterostructure nano scale p-MOSFETs and characterize their performance. In designing the sub-100 nm Si/SiGe heterostructure devices, low temperature process is necessary because of the high diffusivity of Ge in the strained SiGe layer and shallow source/drain structure. In this work, nickel silicidation is used as a low temperature process for low resistance source/drain and fully silicided (FUSI) gate. E-beam lithography is used for patterning nano scale gate with hydrogen silsequioxane (HSQ) e-beam resist and proper cleaning process for CMOS process compatibility. Extraction of carrier transport parameters for deep submicron devices will be also discussed as a performance indicator for characterizing Si/SiGe heterostructure p-MOSFET with special consideration of strain and defect effects. The degradation of effective mobility and velocity was observed in nano-scale Si/SiGe p-MOSFETs. This is mainly due to the increased coulombic scattering by the increased doping concentration.
in the channel. The defects and strain relaxation are other two possible mechanisms of mobility degradation. For further down scaling and mobility enhancement of p-MOSFETs, an additional uniaxial strain is desirable for SiGe material with careful optimization of the channel doping.

Thesis Supervisor: Dimitri A. Antoniadis
Title: Professor of Electrical Engineering
To My Parents
for their continued support
and unwavering faith in me

To my lovely wife
Dr. Young Ae Cho
for her love, endless support
and encouragement

To my children
Tony J.S. and Daniel H.S.
for their love, honesty, innocence,
laughter, smiles, hugs and kisses
Acknowledgments

It is a great pleasure for me to thank the many people who made this thesis possible. First, I would like to thank my advisor, Professor Dimitri A. Antoniadis, for his guidance during my research and study at MIT. He was always patient and kind with me - just as he was with people. He encouraged me to go ahead with device fabrication even in the case of frequent failures and little progress. He also gave me the freedom to explore whatever I felt interesting.

I would like to thank the thesis readers Professor Judy L. Hoyt and Professor Akintunde I. Akinwande for giving me comments and suggestions. Judy offered a lot of great ideas for the fabrication of device that would not have been possible without her help. Tayo always encouraged me with insightful questions and suggestions. He also gave back the smile to me whenever I ask a favor.

I would also like to gratefully acknowledge the support of Microsystem Technology Laboratory (MTL) staffs who made MTL as a friendly place to work. I especially thank Dr. Diadiuk, Paul Tieney, Paudely Zamora, Eric Lim, Bob Bicchieri, Bernard Alamariu, Donal Jamieson, Kurt Broderick and Paul McGrath for their advices and help on process development. Thanks to Mark Mondol and Joel Yang for their great help on electron-beam lithography. Thanks also to Mike Hobbs and Bill Maloney for their help with IT questions.

Many people on the 6th floor have contributed their knowledge to my thesis. Thanks to all past members including Andy R., Issac, Ali, Hassan, Ingvar, Maggie, Andy Fan and Cait. I am deeply indebted to present members of Dimitri’s and Judy’s group; Osama,
John, ChangHyun, Hyung-seok, Jamie, Leo, Pouya, Nicole and Meekyung (MK). I especially wish to thank again Leo and Cait for discussions with me on fabrication and characterization of my devices. I would like to thank Gary Riggott for help with epitaxial growth. Thanks also to Luis, Tae-woo, Donghyun, Sefa, Ling, Bin for their friendship. I would also like to thank Jungwoo and Will (Jinwook) for their knowledge and passion from which I learned a lot.

I also want to acknowledge Samsung Electronics for awarding me a fellowship. I would like to thank vice president Kinam Kim for all his help and guidance during my five years at Samsung Elec. R&D center. Dr. Gitae Jeong, Dr. Gwanhyeobk Koh, Dr. Dongwon Shin and Dr. Sujin Ahn, how can I forget your help in my hours at Samsung?

Finally, I have been blessed by a wonderful family. Thanks to my parents for their continued support and unwavering faith in me. I would like to express my deepest gratitude and love to my lovely wife Dr. Young Ae Cho for her love, endless support and encouragement. Thanks to my children Tony J.S. and Daniel H.S. for their love, honesty, innocence, laughter, smiles, hugs and kisses.
# Table of Contents

1. Introduction .................................................................................................................. 19
   1.1. Thesis Goal ................................................................................................................... 20
   1.2. Outline ......................................................................................................................... 20

2. Theory ................................................................................................................................ 23
   2.1. SiGe alloy and Hole Effective mass .............................................................................. 23
   2.2. Scattering mechanism .................................................................................................. 27
   2.3. Effective mobility extraction ....................................................................................... 29
   2.4. Carrier Velocity ............................................................................................................ 33

3. Process design and fabrication ......................................................................................... 35
   3.1. Si/SiGe heterstructure for high mobility pMOSFET ..................................................... 36
   3.2. Nickel Silicide process .................................................................................................. 38
   3.3. Ramp rate dependency of NiSi formation ..................................................................... 40
   3.4. Nickel Fully Silicide Gate Material .............................................................................. 43
   3.5. Nickel Silicide for low resistance source/drain ............................................................ 50
   3.6. Nano scale gate patterning using e-beam lithography .................................................. 54
   3.7. Halo implantation for short channel device ................................................................. 56
   3.8. Chapter Summary ........................................................................................................ 60

4. Characterization of Si/SiGe bulk p-MOSFET .................................................................. 61
List of Figures

Figure 2-1 A simple illustration of lattice mismatch leading to compressive strain on SiGe alloy .......................................................... 24

Figure 2-2 Valence band for relaxed Si_{0.3}Ge_{0.7} (a) and strained Si_{0.3}Ge_{0.7} grown on relaxed Si_{0.7}Ge_{0.3} [8]. ........................................................................ 25

Figure 2-3 Effective hole mobility versus electric field: data extracted from various references .................................................................................. 26

Figure 2-4 Schematic diagram of three dominant scattering mechanisms [16] ........ 28

Figure 2-5 Gate-to-channel capacitance and gate-to-bulk capacitance as a function of gate voltage measured from Si/SiGe heterostructure p-MOSFET (T_{ox} = 38 Å). 30

Figure 2-6 Total device resistance versus channel length plot used for dR/dL method. Data are extracted from Si/SiGe p-MOSFET with 11 μm channel width and 38 Å gate oxide .................................................................................. 32

Figure 3-1 Overview of Si/SiGe sub-100 nm pMOSFET developed in this study .......... 35

Figure 3-2 Layer structure and band diagram of Si/SiGe heterostructure layer grown directly on Si substrate .................................................................................................................. 36

Figure 3-3 SIMS profile of strained SiGe layer in 50/25 HOI structure with different RTA condition .................................................................................................................. 37

Figure 3-4 Nickel silicide gate sheet resistance with different annealing temperature for a one minute anneal .................................................................................................................. 39

Figure 3-5 Gate sheet resistance with different annealing time at temperature 400 °C.. 40
Figure 3-6 Cumulative probability of reverse leakage current of schottky diode depending on annealing condition ................................................................. 42

Figure 3-7 Conventional Nickel FUSI process: (a) Spacer and S/D IIP, (b) ILD and CMP, (c) Nickel deposition, (d) Silicidation anneal, (e) Removal of unreacted nickel .................................................................................................................................................. 43

Figure 3-8 CMP-free FUSI gate fabrication. (a) gate stack and oxide spacer; (b) re-oxidation of source/drain; (c) SiN removal; (d) nickel deposition; (e) anneal and Ni removal; (f) source/drain silicidation ................................................................................................................................. 44

Figure 3-9 C-V characteristics of n-MOSFET with n+ poly gate and NiSi FUSI gate of undoped polysilicon ................................................................................................................................. 46

Figure 3-10 Gate workfunction extraction using flat band voltage extracted by numerical simulation of measured CV data. Workfunction difference is due to the source/drain implantation ................................................................................................................................. 47

Figure 3-11 Flat band voltage shift due to the modulation of gate workfunction which is done by predoping poly gate ................................................................................................................................. 49

Figure 3-12 NiSi source/drain silicidation effect on Si/SiGe heterostructure p-MOSFETs. ................................................................................................................................. 50

Figure 3-13 Nickel Silicide Source/drain formation using oxide etch-back process. (a) conventional S/D with thick oxide spacer ( ˜ 1000 Å ) (b) nickel silicide S/D with thinner oxide spacer ( ˜ 500 Å ) ................................................................................................................................. 51

Figure 3-14 Junction diode characteristics of nickel silicide source/drain with different nickel deposition ................................................................................................................................. 52
Figure 3-15 Contact resistance comparison, Silicide source/drain activation temperature ................................................................. 53
Figure 3-16 90 nm size Polystyrene Latex Spheres used for focusing the electron beam.54
Figure 3-17 Process sequence of nano patterning by mixing e-beam and i-line stepper lithography. (a) electron beam lithography (b) photo resist patterning (c) etching the nitride hard mask (d) HSQ and PR strip (e) etching poly using nitride hard mask.......................................................... 55
Figure 3-18 SEM view after etching gate with HSQ hard mask ............................................. 56
Figure 3-19 2D contour plot for phosphorous concentration extracted from TSUPREM process simulation......................................................................................................... 57
Figure 3-20 Doping profile of 100 nm p-MOSFET extracted using TSUPREM process simulator. (a) Longitudinal doping profile at 10 nm under the gate oxide. (b) Vertical doping profile under the oxide spacer.......................................................... 58
Figure 3-21 Experimental data: V_th roll-off characteristics: (a) with different halo tilt angles (b) with different halo doping conditions at tilt angle 20° ............... 59
Figure 3-22 Experimental data of Si/SiGe hetero-structure short channel device transfer characteristics......................................................................................................... 59
Figure 4-1 Layer structure for Si/SiGe directly grown on silicon bulk......................... 62
Figure 4-2 SIMIS profile of P and Ge in the Si/SiGe heterostructure after full CMOS process.................................................................................................................. 63
Figure 4-3 Comparison of CV data form measurement and DESSIS simulation: (a) C-V characteristics with different doping profile shown in (b) .............. 64
Figure 4-4 The comparison of two layer structures with different high doped region: (a) phosphorous in-situ doping (b) highly doped Sb substrate wafer.............. 65
Figure 4-5 Threshold voltage roll-off characteristic with different layer structure under the SiGe channel. ................................................................. 66
Figure 4-6 Low field mobility enhancement with Sb substrate by reducing dopant pileup. ............................................................................................................. 67
Figure 4-7 Threshold voltage roll-off characteristics with different thickness of intrinsic buffer layer and doping level of highly doped silicon substrate........... 68
Figure 4-8 Hole effective mobility as a function of substrate doping level............. 69
Figure 4-9 Total gate-to-channel capacitance with different frequency................... 71
Figure 4-10 Intrinsic gate to channel capacitance with denoising process on short channel devices (Width = 10 μm, Tox = 4 nm). ........................................... 72
Figure 4-11 Total source/drain resistance as a function of gate-to-source/drain voltage extracted from the Si/SiGe p-MOSFET with $V_{th} = 1V$. .................. 73
Figure 4-12 Schematic of intrinsic parasitic component of MOSFET. $C_{pad}$ is the pad capacitance and not shown in this figure. ........................................ 75
Figure 4-13 Total gate capacitance as a function of drawn channel length: (a) channel length defined by e-beam lithography for short channel (b) optical lithography for long channel........................................................................ 76
Figure 4-14 Total gate capacitance as a function of drawn channel length: the slope of dashed line is extracted from long channel devices............................ 77
Figure 4-15 Electron beam lithography pattern skew between drawn and actual size. 10 KV acceleration voltage with 250 μC/cm$^2$ in dose.............................. 78
Figure 4-16 Effective channel length extraction plot using capacitance method. Open symbol represent corrected gate size using C-L plot ........................................ 79

Figure 4-17 Comparison of effective mobility before and after capacitance correction .... 80

Figure 4-18 Illustration of two MOSFETs method for effective mobility extraction ...... 81

Figure 4-19 Comparison of effective mobility extracted by two different methods: solid symbol is dR/dL method and open symbol is two MOSFETs method........ 83

Figure 4-20 Low field hole mobility versus effective channel length. The mobility is extracted by two MOSFETs method at inversion charge Ni = 5x10^{12} \text{ cm}^{-2} . 85

Figure 4-21 DIBL characteristics of Si/SiGe pMOSFETs used for velocity extraction... 87

Figure 4-22 Hole velocity as a function of effective channel length. Si/SiGe p-MOSFET with 55% Ge fraction: (a) velocity versus DIBL, (b) Velocity versus effective channel length ............................................................................ 88

Figure 4-23 Illustration of the location of virtual source ........................................ 89

Figure 4-24 Compact model fitting result for 88 nm Si/SiGe pMOSFET.................... 90
List of Tables

Table 1 Comparison of silicidation technique ................................................................. 38
Table 2 Summary of Nickel Silicide process ..................................................................... 53
Table 3 S/D implantation conditions .................................................................................. 56
Table 4 Mobility and velocity comparison of experimental and model .............................. 89
1. Introduction

CMOS intrinsic device performance has increased by 17% per year down to the 90 nm node by scaling gate length combined with source/drain and channel engineering [1]. For further performance enhancement beyond the 90 nm node, the intrinsic carrier transport properties in the channel material have been engineered to improve carrier velocity and continue to increase transistor performance. Many previous studies showed electron and hole mobility enhancement using the process induced strain of the bulk Si channel but higher mobility in channel materials is required to continue scaling down to 10 nm gate length [2].

SiGe is an attractive semiconductor material for strain engineering MOSFET devices. Compressive strain in SiGe changes its band structure by splitting valence bands and hence increases hole mobility [3]. Compressively strained SiGe can be achieved by various ways that give either uniaxial or biaxial stress [4]. Among these many ways, the pseudomorphic SiGe layer grown directly on Si bulk substrate has benefits over the conventional SiGe layer grown on relaxed SiGe virtual substrate such as process simplicity and reduced Ge fraction for the same strain. The hole mobility in long channel p-MOSFET increases as the Ge content increases in the SiGe channel layer with peak enhancement 4X over Si reference [5]. However, Si-Ge inter-diffusion increases exponentially with strain and Ge composition even for relatively low thermal budgets and thus makes it difficult to have higher Ge concentration in the channel [6]. Therefore, thermal budget for Si/SiGe heterostructure is one of the most important constraints on the
peak mobility improvement in the pseudomorphic SiGe layer grown directly on Si bulk substrate.

Besides the channel structure, source/drain engineering is also important for advanced deep sub-micrometer CMOS devices in order to maintain special source/drain structures such as halo and shallow junction structure. Defects created by source/drain implantation can impact carrier mobility and they should be minimized for high mobility channel material. Therefore choice of dopant and thermal process should be carefully considered to maximize mobility gain of sub-100 nm p-MOSFET.

1.1. Thesis Goal

The goal of this thesis is to give a fabrication guideline for Si/SiGe heterostructure sub-100 nm p-MOSFET including optimization of pseudomorphic SiGe heterostructure, the formation of fully silicided metal gate, source/drain silicidation and nano-scale gate patterning using e-beam lithography. It also covers the extraction methodology of fundamental MOSFET carrier parameters such as carrier mobility and velocity, effective channel length, source/drain and contact resistance of short channel Si/SiGe heterostructure devices. Finally it discusses short channel effects of p-MOSFET with compressively strained SiGe as a channel material.

1.2. Outline

In chapter 2 general properties of Si and SiGe are reviewed including band structure and hole transport. The conventional mobility extraction method is also described. Chapter 3 introduces process integration scheme including layer structure, nickel silicidation, nano patterning and halo implantation for sub-100 nm devices. Electrical characterization of
Si/SiGe heterostructure device is shown in Chapter 4. This chapter also presents a new mobility extraction method for short channel device. Chapter 5 summarizes the key contribution of this thesis and provides suggestions for future work.
2. Theory

This chapter presents the basic material properties of Si and SiGe to understand hole transport in SiGe layer. First, the concept of compressive biaxial strain of SiGe is presented with effective mass change by band deformation. Next, hole mobility is reviewed with scattering mechanism in SiGe heterostructure. Finally, the parameter extraction methods including effective mobility and velocity are also reviewed.

2.1. SiGe alloy and Hole Effective mass

Silicon and Germanium have similar crystal structure. The lattice constant of Si and Ge are 5.432 Å and 5.658 Å, respectively, which is 4.2 % lattice mismatch. A relaxed Si$_x$Ge$_{1-x}$ alloy has randomly distributed Si and Ge atoms in the lattice with Ge fraction x. The lattice constant of relaxed Si$_{1-x}$Ge$_x$ can be estimated by Vegard’s law which is a linear interpolation of two lattice-parameters [7] and given by

$$a(x) = x \cdot a_{Ge} + (1-x) \cdot a_{Si}$$

where $x$ is Ge fraction in the SiGe alloy, $a_{Ge}$ and $a_{Si}$ are lattice constant of Ge and Si, respectively. A SiGe alloy layer grown pseudomorphically on the relaxed silicon substrate has the compressive strain by matching the in-plane lattice parameter to the underlying silicon substrate as illustrated in Figure 2-1.
Silicon

$\overset{\text{a}_{\text{si}} = 5.431\text{\AA}}{\text{Si}}$

Germanium

$\overset{\text{a}_{\text{Ge}} = 5.646\text{\AA}}{\text{Ge}}$

Relaxed SiGe Alloy

Compressively strained SiGe Alloy

Lattice matched to Si

Hole mobility enhancement

Figure 2-1 A simple illustration of lattice mismatch leading to compressive strain on SiGe alloy.

When the biaxial compressive stress is introduced on the SiGe alloy, the heavy hole and light hole bands split and deform the band structure, which results in reduced interband scattering and in-plane transport mass. Band splitting and deformation are shown Figure 2-2.
Figure 2-2 Valence band for relaxed Si$_{0.3}$Ge$_{0.7}$ (a) and strained Si$_{0.3}$Ge$_{0.7}$ grown on relaxed Si$_{0.7}$Ge$_{0.3}$ [8].

The valence band structure of SiGe alloy is determined by 6-band k.p simulations [9-11] using nextnano [12] with non linear interpolation of valence band parameters described by Rieger and Vogl [13]. The effective mass is then calculated from the band curvature and is given by

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{\partial^2 E(k)}{\partial k^2}$$

2-2

Because of non-parabolic band structure, the effective hole mass is dependent on carrier energy and thus changed by temperature, carrier concentration and doping level. As can be seen in Figure 2-2 (b), the curvature of the heavy hole is similar to that of light hole at the band edge indicating that biaxial strain results in reduction of heavy hole effective mass to close to light hole effective mass in the SiGe material. From the low field mobility equation
\[ \mu = \frac{e \cdot \tau}{m^*} \]

where \( e \) is the electron charge, \( 1/\tau \) is the scattering rate and \( m^* \) is the effective mass, the hole mobility in compressively strained SiGe should increase relative to that of Si or unstrained SiGe both because increased \( \tau \), as will be discussed next, and decreased \( m^* \). Figure 2-3 shows the hole mobility in Si\(_{1-x}\)Ge\(_x\) on Si\(_{1-y}\)Ge\(_y\) heterostructure [3, 14-16]. Here \( x \) and \( y \) are Ge concentration in the channel and in the relaxed SiGe layer. 10X hole mobility enhancement over the Si was observed on the compressively strained pure Ge grown on relaxed Si\(_{0.5}\)Ge\(_{0.5}\) layer [14]. This result indicates the potential of SiGe or Ge as a channel material for future device.

![Figure 2-3](image)

Figure 2-3 Effective hole mobility versus electric field: data extracted from various references.
2.2. Scattering mechanism

The scattering rate $1/\tau$ is the sum of three major scattering mechanisms: phonon scattering, coulombic scattering by ionized impurities, interface roughness scattering at the Si-SiO$_2$ interface. The rate is given by Mattheissen's rule

$$\frac{1}{\tau} = \sum \frac{1}{\tau_n} \tag{2-4}$$

where $\tau_n$ is the momentum relaxation time for the $n^{th}$ scattering mechanism.

Coulombic scattering is dominant at low inversion charge due to the ionized impurities in the channel. This scattering mechanism is important in characterizing short channel devices because highly-doped channel structure is necessary for good electrostatic behavior. As the substrate doping increases, mobility enhancement of the SiGe material decreases at low inversion carrier by coulombic scattering. The coulombic scattering is screened out by the inversion charge as inversion charge increases. This effect is shown in Figure 2-4. According to this diagram, the mobility curve is divided into three regions, coulombic scattering, phonon scattering and surface roughness scattering.
Carriers can exchange the energy with lattice vibration which is modeled by energy quanta called phonons. This carrier-phonon interaction is referred to phonon scattering. As the inversion charge density increases, the phonon scattering becomes dominant scattering mechanism. In compressively strain SiGe, acoustic phonon scattering is dominant because interband optical phonon scattering is reduced due to the large band splitting between the heavy hole and light hole bands [10, 17]. The phonon scattering rate in a 2D quantum well such as a MOSFET inversion layer or heterostructure MOSFET is given by

\[ \frac{1}{\tau_{AC}} = \frac{\pi D_k^2 k_B T_L}{h c_i} \cdot \frac{1}{W_{\bar{a}}} g_{2D}(E) \]
where $D_A$ is the deformation potential for acoustic phonon scattering, $k_B$ is Boltzmann’s constant, $T_L$ is the lattice temperature, $c_l$ is the elastic constant for the material, $W_{\text{fl}}$ is the effective inversion layer thickness and $g_{2D}$ is the 2D density of states.

At high inversion carrier density, the surface roughness scattering is dominant due to the strong transverse electric field. The surface roughness scattering is the elastic scattering of carriers by an imperfect interface between two different materials.

### 2.3. Effective mobility extraction

In order to extract mobility from a nano-scale Si/SiGe heterostructure device, special efforts should be made to eliminate parasitic effects like overlap capacitance, source/drain resistance and channel length uncertainty. Split C-V method [18] and dR/dL method [19] are two widely used methods for MOSFET structures.

The split C-V method uses capacitance data to calculate inversion charge and bulk charges. Figure 2-5 shows gate-to-channel, $C_{ge}$, and gate-to-bulk, $C_{gb}$, capacitance as a function of gate voltage measured from Si/SiGe heterostructure.
Figure 2-5 Gate-to-channel capacitance and gate-to-bulk capacitance as a function of gate voltage measured from Si/SiGe heterostructure p-MOSFET ($T_{ox} = 38 \, \text{Å}$).

The parasitic component of measured capacitance can be easily subtracted or ignored due to the relatively large device area in this case. From the split C-V plot, the voltage-dependent inversion and depletion charges are obtained by integration of $C_{gc}(V_g)$ and $C_{gb}(V_g)$ and given by

$$Q_i(V_g) = \int_{V_{acc}}^{V_g} C_{gc}(v) \, dv, \quad Q_b(V_g) = \int_{V_{acc}}^{V_g} C_{gb}(v) \, dv$$

where the $V_{acc}$ is a gate voltage chosen in accumulation and $V_b$ is the flat band voltage deduced from Maserjian's function [20, 21] or CVC method [22]. The electric field can then be calculated as:

$$E_{eff} = \frac{Q_b + \frac{1}{\eta} \cdot Q_i}{\varepsilon_{si}}$$
where $\varepsilon_s$ is the silicon permittivity and $\eta$ is an empirical factor with values 2 for electron and 3 for hole. Using the MOSFET linear current equation, effective carrier mobility can be given by

$$\mu_{\text{eff}} = \frac{L_{\text{eff}}}{W} \left( \frac{I_d}{V_d - R_{\text{ext}} \cdot I_d} \right) \cdot \frac{1}{Q_i}$$

Where $L_{\text{eff}}$ is the effective channel length, $W$ is the device channel width, $I_d$ is the measured drain current, $V_d$ is the applied drain voltage, $R_{\text{ext}}$ is the external series resistance and $Q_i$ is the inversion charge calculated from split C-V. In order to use split C-V method for short channel device, accurate measurement of these parameters is necessary. Extraction method of these parameters will be discussed in chapter 4.

A dR/dL method extracts mobility using total device resistance versus channel length plot as shown in Figure 2-6. The advantage of this method is that mobility can be extracted directly from short channel devices without the correction of $L_{\text{eff}}$ and $R_{\text{ext}}$. 

31
Figure 2-6 Total device resistance versus channel length plot used for dR/dL method. Data are extracted from Si/SiGe p-MOSFET with 11 μm channel width and 38 Å gate oxide.

The total device resistance of MOSFET can be expressed as:

\[
R_{\text{total}} = R_{\text{FET}} + R_{\text{ext}} = \frac{L_{\text{eff}}}{W \cdot \mu \cdot Q_{\text{inv}}} + R_{\text{ext}}
\]

where \( R_{\text{FET}} \) is the intrinsic channel resistance of device and \( Q_{\text{inv}} \) is the integrated channel charge density per area from long channel device. By taking a derivative of \( R_{\text{total}} \) with respect to \( L_{\text{eff}} \) and rearranging, the mobility \( \mu \) can be expressed as:

\[
\mu = \frac{W \cdot Q_{\text{inv}} \cdot dR_{\text{total}}}{dL_{\text{eff}}}
\]

Determination of the channel length should be accurate and consistent for each size to give correct channel length difference between adjacent two devices. Uncertainty in \( Q_{\text{inv}} \)
at low inversion carrier and the assumption of constant mobility versus $L_{\text{eff}}$ also affect the accuracy of this method. By a linear curve fitting of total resistance versus channel length plot, the extracted mobility is an average of mobility values in those channel lengths. To characterize short channel mobility behavior, mobility should be extracted from narrow range of sizes.

### 2.4. Carrier Velocity

The MOSFET current can be simplified by

$$I_d/W = Q_i \cdot v_s$$  \hspace{1cm} 2-11

where $Q_i$ and $v_s$ are carrier density and velocity near the source edge, respectively [23]. In the long channel device, drive current $I_{on}$ is determined by the product of low field mobility and the electric field near the source. In this transport mechanism, low field carrier mobility is an important parameter in determining drive current. As the gate length scales down to nano-regime, the carrier scattering rate is reduced and thus nonstationary transport is dominant. This phenomenon has been formulated as quasi-ballistic transport by Lundstrom et al [23, 24]. According to this theory

$$v_s = v_{\text{inj}} \cdot \frac{1 - r}{1 + r}$$  \hspace{1cm} 2-12

where $v_{\text{inj}}$ is the injection velocity at the top of the source barrier and $r$ is the back scattering rate which is related to carrier mobility near the source region. In this regime, carrier mobility is still important in determining drive current [25, 26].

It is hypothesized that when the channel length is shorter than 10 nm, no scattering events occur in the channel. In this transport regime, the full ballistic transport is
dominant and the current is directly related to injection velocity [27, 28] and would be given by

\[ \frac{I_d}{W} = Q_i \cdot \nu_{\text{inj}} \]  

To see the performance enhancement in nano scale devices, the intrinsic carrier velocity should be carefully characterized. In this study, the Lochtefeld et al [29] method was used to extract velocity, which gives inversion layer carrier velocity close to the source injection point. The effective velocity on the sub-100 nm Si/SiGe heterostructure pMOSFETs was extracted to evaluate the SiGe material as a performance booster in future devices.
3. Process design and fabrication

Various unit processes for high performance MOSFET were tested to examine carrier transport in sub-100 nm Si/SiGe heterostructure bulk pMOSFET. Figure 3-1 shows an overview of processes developed throughout this study. The development of low temperature process without sacrificing resistance is key issue to maintaining Ge peak concentration in SiGe channel. Low temperature wet oxidation was used for gate oxide. Nickel silicide process was also implemented to minimize gate and source/drain series resistance. To form a sub-100 nm device, the gate was patterned using electron beam lithography with HSQ e-beam resist. Short channel effect was minimized using large tilt angle halo and low energy extension implantation. This chapter provides details of the processes used to build short channel device.

Figure 3-1 Overview of Si/SiGe sub-100 nm pMOSFET developed in this study.
3.1. Si/SiGe heterstructure for high mobility pMOSFET

The layer structure and band diagram are illustrated in Figure 3-2. Epitaxial layers were grown in an Applied Materials low pressure chemical vapor deposition (LPCVD) "Epi-Centura" system. Highly doped p-type silicon was grown at temperature 900 °C on Si substrate to prevent bulk punch-through in short channel device. Intrinsic silicon layer was then grown at 680 °C followed by a 7 nm-thick strained $Si_{1-y}Ge_y$ layer grown at 525 °C with Ge compositions, $y$, with 0.4 and 0.55. A 5 nm-thick silicon cap layer was grown at 600 °C. The final germanium compositions were measured using Secondary Ion Mass Spectrometry (SIMS).

The thermal budget of the process was kept low to minimize the inter-diffusion of the SiGe layer which increases exponentially with strain and Ge composition even for relatively low thermal budgets and thus make it difficult to have higher Ge concentration in the channel [6]. This can be seen from the work of Aberg et. al. [30] shown in Figure 3-3 here. Ge SIMS profiles after 10 sec RTA with different temperature were monitored.
for a strained Si$_{0.5}$Ge$_{0.5}$ channel pseudomorphically grown on a relaxed Si$_{0.75}$Ge$_{0.25}$ layer. 5% of Ge concentration was dropped with 850 °C 10 sec RTA process. The peak Ge concentration in strained SiGe channel directly related to the channel mobility which increased with increasing Ge concentration.

![SIMS profile](image)

Figure 3-3 SIMS profile of strained SiGe layer in 50/25 HOI structure with different RTA condition.

Maintaining the peak Ge composition in the SiGe layer grown on relaxed Si is more difficult due to the high level of strain which increases the diffusivity of Ge. Therefore, the thermal budget for Si/SiGe heterostructure is one of most important constraints on the peak mobility improvement in the pseudomorphic SiGe layer grown directly on Si bulk substrate.
3.2. Nickel Silicide process

To obtain low source and drain series resistance in Si/SiGe p-MOSFET, the low temperature nickel silicidation process is adopted in this work. Nickel silicide (NiSi) has been widely studied in recent years for use as a contact material, and now it has been used as a low resistance metal gate and source/drain material for high performance deep submicron device [31, 32]. The formation of nickel silicide at low temperature is an important aspect because maintaining Ge peak in SiGe channel and source/drain dopant is a key issue in achieving the high hole mobility for Si/SiGe heterostructure devices. As seen in Table 1 nickel silicide can be formed at very low temperature compared to other silicide material like CoSi2 and TiSi, which are also used for silicidation. The nickel silicidation process was done with simple heat treatment at temperatures as low as 400 °C which is low enough to preserve the shallow junctions and hetero-channel structure of high performance devices.

Table 1 Comparison of silicidation technique

<table>
<thead>
<tr>
<th>Process Temp</th>
<th>TiSi₂</th>
<th>CoSi₂</th>
<th>NiSi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st RTP :</td>
<td>620 °C ~ 680 °C</td>
<td>450 °C</td>
<td>400 °C ~ 500 °C</td>
</tr>
<tr>
<td>2nd RTP :</td>
<td>800 °C</td>
<td></td>
<td>Single RTP</td>
</tr>
<tr>
<td>Line width effect</td>
<td>Large</td>
<td>Small</td>
<td>Small</td>
</tr>
<tr>
<td>Thermal Stability</td>
<td>&lt; 850 °C</td>
<td>&lt; 900 °C</td>
<td>&lt; 600 °C</td>
</tr>
<tr>
<td>Silicon Consumption</td>
<td>2.22</td>
<td>3.61</td>
<td>1.83</td>
</tr>
</tbody>
</table>
It is also suitable for low resistance source/drain contact for ultrathin body devices and Si/SiGe-layer-structure devices since nickel consumes less silicon compared to other materials [33].

Nickel Silicidation has three major phases depending on annealing temperature [34]. Figure 3-4 shows gate sheet resistance as a function of annealing temperature of undoped polysilicon while the samples are heated at 100 °C/s and allowed to remain at temperature for 1 min in nitrogen at each annealing temperature. It is expected that this behavior is representative of single-crystal-Si as well.

![Figure 3-4 Nickel silicide gate sheet resistance with different annealing temperature for a one minute anneal.](image)

The three major phases are distinguished based on the film sheet resistance and the low resistivity NiSi film can be formed with a wide process window at temperatures between 400 °C to 750 °C. The first phase of Ni$_2$Si is formed first in the temperature
from 250 °C to 350 °C and is very important for formation of the low resistivity nickel monosilicide. The low resistance phases were formed at temperatures between 400 °C to 750 °C where most of Ni–Si systems are monosilicide NiSi. Up to 750 °C NiSi is stable and changes to silicon rich NiSi$_2$ at temperatures higher than 750 °C. The increase of resistance at temperatures above 750 °C is due to the agglomeration of NiSi film and the formation of high resistance silicon rich phase NiSi$_2$. Considering germanium diffusion in Si/SiGe heterostructure devices and uniformity of low resistivity NiSi layer, the silicidation anneal process between 400 °C and 500 °C is preferred [35].

![Graph](image)

Figure 3-5 Gate sheet resistance with different annealing time at temperature 400 °C.

### 3.3. Ramp rate dependency of NiSi formation

Rapid thermal annealing process has been widely used for CMOS manufacturing to reach the desired temperature in short time minimizing unwanted reaction. However certain response time is necessary for the RTA process to allow the wafer to respond to sudden
change of temperature depending on reaction material. To optimize RTA temperature ramp rate, Schottky diodes were built. 5 nm Ni was deposited on patterned n-type silicon using electron beam evaporator. The silicidation process was done in an RTA machine in nitrogen ambient from room temperature to a final temperature of 450 °C with different ramp conditions. The unreacted nickel was etched away in a H₂SO₄:H₂O₂ = 4:1 solution. Samples are then sintered at 420 °C in hydrogen and nitrogen ambient.

During the first phase of silicidation, stress and dislocation of film can be generated by other phases at temperatures below 350 °C. For example, nickel-rich phases induce mismatch in silicide film and need higher temperature to get the NiSi monosilicide. In order to minimize defect generation, the initial ramp rate of annealing to this temperature should be optimized because a variety of phases can be formed during this step. Junction leakage currents of NiSi Schottky diode were measured to monitor defect generation. Figure 3-6 shows junction reverse leakage current with different initial ramping conditions. One has a ramp rate of 100 °C/sec and others have 10 °C/sec with different holding time at temperature 300 °C. Reverse leakage current density is compared with 10 different positions on each wafer. Lower ramp rate and longer holding time at 300 °C showed decreased junction leakage currents compared to fast temperature ramping.
In the initial ramp cycle nickel diffuses into silicon to form Ni$_2$Si which can be controlled by ramp rate. This nickel rich phase Ni$_2$Si film transforms into the low resistivity monosilicide film NiSi at the second stage of silicidation above 350 °C. If there was too much nickel left in the system at the first stage, the film introduces a large stress at the second stage and can induce leakage current. A higher heating ramp rate results in a shorter thermal response time which in turn results in not enough time to transform Ni$_2$Si.

By controlling heating rate at the first stage, such stresses and defects can be reduced. It also reduces the Ni penetration into the junction depletion region when nickel is used for source/drain silicidation.

Figure 3-6 Cumulative probability of reverse leakage current of schottky diode depending on annealing condition.
3.4. Nickel Fully Silicide Gate Material

The fully silicided (FUSI) gate is a well known metal gate process and offers higher gate capacitance, lower sheet resistance and potentially superior scalability compared to conventional polysilicon and other types of silicide gate material [33]. A conventional method of the FUSI gate transistor is the gate last process which involves CMP (Chemical Mechanical Polishing) process for planarization as shown in Figure 3-7. After the formation of source/drain, the transistor is fully covered with dielectric material (ILD) which is then planarized by etching back to expose polysilicon gates. After the pre-metal cleaning process, nickel is deposited on top of this exposed gate polysilicon and then annealed using the RTA process followed by removal of unreacted nickel.

Figure 3-7 Conventional Nickel FUSI process: (a) Spacer and S/D IIP, (b) ILD and CMP, (c) Nickel deposition, (d) Silicidation anneal, (e) Removal of unreacted nickel

The CMP process was first used for metallization to enable lithography which is sensitive to wafer surface topography. As the process technology was developed CMP is used in transistor formation including shallow trench isolation and metal gate formation.
However, the CMP process can result in defects and thickness variation in chip-to-chip and wafer-to-wafer which is not consistent with the scaling trend of the IC industry. Unlike early technologies that utilized CMP, it is necessary to have a degree of thickness precision and maintain a low defect generation in nano-scale node. In order to use the CMP process in nano-scale technology node, significant efforts in optimizing this process should be done to make advancement in these two areas.

Figure 3-8 CMP-free FUSI gate fabrication. (a) gate stack and oxide spacer; (b) re-oxidation of source/drain; (c) SiN removal; (d) nickel deposition; (e) anneal and Ni removal; (f) source/drain silicidation

In this thesis, the development of a FUSI Nickel Silicidation metal gate transistor without CMP process is presented. Figure 3-8 shows CMP-free silicidation process which uses a nitride sacrificing layer and a low temperature source/drain re-oxidation. After the gate dielectric process, 100-nm-thick undoped amorphous silicon film was deposited by low pressure chemical vapor deposition at 560 °C on the gate oxide and then was implanted with different dopants for workfunction modulation. Silicon nitride film was
deposited to prevent the top of polysilicon from oxidation during the S/D re-oxidation step (b). Source and drain regions were ion implanted and activated with RTP annealing at 800 °C for 10 seconds after LTO oxide spacer was formed and then followed by nitride removal using hot phosphoric acid. The implantation conditions for source/drain will be explained in section 3.7. Nickel film was deposited on exposed poly gate using e-beam evaporation. Before the deposition, samples were cleaned using Piranha solution (H_2SO_4:H_2O_2 = 3:1) and etched back with HF solution to remove native oxide on the polysilicon gate. For the silicidation, rapid thermal annealing was used. The unreacted metal was also removed by Piranha solution (H_2SO_4:H_2O_2 = 4:1~5:1) without damaging the silicided gate. To minimize source and drain series resistance, a second nickel layer was deposited and silicided at the exposed S/D regions in the same manner. The detailed process will be listed in Appendix. The capacitance-voltage (C-V) characteristics of n-MOSFET with FUSI gate of undoped polysilicon is compared to n+ polysilicon gate in Figure 3-9.
Figure 3-9 C-V characteristics of n-MOSFET with n+ poly gate and NiSi FUSI gate of undoped polysilicon.

The flat band voltage shift indicates the change of workfunction of gate materials. Nickel FUSI gate n-MOSFET also shows less poly depletion than n+ polysilicon gate in inversion region, indicating metallization of undoped polysilicon gate and CET (Capacitance Equivalent Thickness) reduction. With fully silicided gate material, a reduction of 3-5 Å in CET can also be achieved.
Figure 3-10 Gate workfunction extraction using flat band voltage extracted by numerical simulation of measured CV data. Workfunction difference is due to the source/drain implantation

For the extraction of flat band voltage nickel FUSI n-MOS capacitors and transistors with different gate oxide thickness (3.0, 4.6, and 6.6 nm) were fabricated. Undoped polysilicon was deposited at 560 °C and followed by nickel deposition and one step rapid thermal annealing process at 450 °C 1 minute. Figure 3-10 shows the flat band voltage of n-MOSFET extracted by numerical simulation of measured CV data with different thickness of gate oxide. Assuming that the all the fixed charges are confined close to oxide and silicon interface, the graph of flat band voltage as a function of oxide thickness gives workfunction difference between gate and substrate at the intercept of the y-axis. The metal workfunction of NiSi gate can be found using

$$V_{FB} = \Phi_{MS} + \frac{Q_{ox}T_{ox}}{\varepsilon_{ox}}$$
where the $\Phi_B$ is the substrate Fermi potential, $E_g$ is the band gap energy and $\chi$ is the electron affinity of semiconductor. The flat band voltage and doping level of each device was extracted from the NCSU CVC program [22] by fitting the measured C-V characteristics. The effective charge density also can be calculated from the slope of line. Extracted workfunctions of two different structures, capacitor structure and transistor structure, are 4.66 eV and 4.61 eV, respectively. The difference of these workfunction is due to the heavy ion implantation for source/drain after patterning the gate. The segregation of the dopants from sidewall to the silicide interface during the silicidation is found to be the cause of the flat band voltage shift [36, 37].

In developing short channel devices, one of the main issues is the selection of proper gate material with desired workfunction [38]. Deep-scaled bulk CMOS technology generally requires two different gate workfunctions for NMOS and PMOS within about 0.2eV from the corresponding band edges. The workfunction of NiSi is close to the middle of the Si bandgap energy, which is suitable for use with SiGe p-MOSFET [39]. Bandgap and strain engineering devices need more emphasis on threshold voltage control using workfunction modulation because of their channel structures. The nickel FUSI gate process is attractive because it has advantages such as tunable workfunction and low temperature silicidation. The workfunction of the nickel FUSI gate of MOSFETs can be controlled by doping the polysilicon gate material with different impurities prior to silicidation [40].
Figure 3-11 Flat band voltage shift due to the modulation of gate workfunction which is done by predoping poly gate.

Figure 3-11 shows the flat band voltage shift for different ion implantation on LPCVD polysilicon gates. The shift of flat band voltages can be explained by impurity segregation into the interface between the NiSi gate and the gate oxide. Impurities segregate out of the silicide because of their low silicide solubility [41-43]. The flatband voltage was shifted -0.15V for As and P implantation regardless of ion implanted species or condition. This is due to the saturation of workfunction of NiSi gate material. For boron doped sample, it showed huge changes in threshold voltage which is due to the boron penetration into channel region. The range of workfunction modulation was found to be rather limited.
3.5. Nickel Silicide for low resistance source/drain

Traditionally, the junction depth of CMOS transistor has been controlled by ion implantation of proper dopants into source/drain region after formation of gate. To prevent short channel effect in sub-100 nm channel transistors, shallow junctions are required where the limited dopants sheet-density increases source/drain resistance. In order to prevent performance degradation by parasitic resistance RC-delay, low resistance source/drain is necessary [44].

![Graph showing source/drain resistance](image)

Figure 3-12 NiSi source/drain silicidation effect on Si/SiGe heterostructure p-MOSFETs.

As discussed earlier, the source/drain series resistance of the Si/SiGe p-MOSFET was reduced in this work using the nickel silicidation technique. Figure 3-12 shows the reduction of total source/drain resistance by factor of 4 compared to a non-silicided device. The S/D doping conditions were as follows: phosphorous dose of $2 \times 10^{13}$ cm$^{-2}$ was implanted at 40 KeV with two symmetries at tilt angle 20° for halo implantation. Boron
dose of $10^{14}$ cm$^{-2}$ was implanted at 3 KeV with two symmetries at 7° tilt angle for S/D extension. After the spacer formation, B dose $4 \times 10^{15}$ cm$^{-2}$ was implanted at 7 KeV with tilt angle 0° for the deep source/drain junction. A single 800 °C 10 sec RTA anneal was used for dopant activation. The NiSi source/drain sample was cleaned and silicided with 3 nm nickel film deposited on top of source/drain region. Considering the change of sheet resistance of p+ junction from 210 Ω/sq to 11 Ω/sq, 75% of total source/drain resistance in NiSi S/D p-MOSFET comes from the lightly doped source/drain extension under the spacer which is 50% without the silicidation.

Figure 3-13 illustrates the formation of NiSi source/drain which has shorter effective S/D extension length compared to conventional S/D (a). Nickel is deposited and annealed after the oxide spacer is etched back in 50:1 HF solution. NiSi is then formed on the S/D extension region and it reduces total resistance. The P+ S/D sheet resistance of 210 Ω/sq which was measured from the sample without the silicidation was reduced to
11 Ω/sq. Oxide etch-back time, thickness of nickel film and implantation condition for S/D extension should be carefully optimized to reduce junction leakage current.

The junction diode current-voltage characteristic was measured after source/drain silicidation and was compared to conventional p+ junction. Before the silicidation step, nickel was evaporated with two different thicknesses, 10 nm and 3 nm.

![Figure 3-14 Junction diode characteristics of nickel silicide source/drain with different nickel deposition](image)

As shown in Figure 3-14, the junction diode characteristic of 3 nm nickel deposition is very similar to p+/n junction without silicidation. For the sample with 10 nm nickel deposition shows Schottky junction diode characteristics which has lower barrier height and increased reverse junction leakage current. Contact resistance was also reduced by nickel silicide as shown in Figure 3-15. Electrical parameters are summarized and compared to conventional source/drain structure in Table 2.
Figure 3-15 Contact resistance comparison, Silicide source/drain activation temperature

Table 2 Summary of Nickel Silicide process

<table>
<thead>
<tr>
<th>Material</th>
<th>Gate Rs</th>
<th>Workfunction</th>
<th>Polycrystalline</th>
<th>Poly depletion</th>
<th>CMP Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate Rs</td>
<td>450</td>
<td>4</td>
<td>Ω/sq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P+ Rs</td>
<td>210</td>
<td>11</td>
<td>Ω/sq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S/D Rsd</td>
<td>2K</td>
<td>550</td>
<td>Ω-um</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact Rc</td>
<td>3x10^{-6}</td>
<td>1x10^{-7}</td>
<td>Ω-cm^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workfunction</td>
<td>-</td>
<td>4.6eV</td>
<td>Undoped poly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poly depletion</td>
<td>-</td>
<td>no</td>
<td>no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMP Process</td>
<td>-</td>
<td>no</td>
<td>no</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.6. Nano scale gate patterning using e-beam lithography

Electron beam lithography was used for patterning nano size gate using The Raith 150 system with hydrogen silsequioxane (HSQ) for e-beam resist [45]. HSQ has been used for low-k dielectric material for semiconductor application. This material changes chemical properties similar to SiO$_2$ by electron injection and serves good hard mask during the etching step. Polystyrene latex spheres were used as focus specimen instead of using gold particle for CMOS process compatibility. Figure 3-16 shows SEM top view of 90 nm latex particles which are easily removed with subsequent cleaning process.

![SEM image of 90 nm latex spheres](image)

Figure 3-16 90 nm size Polystyrene Latex Spheres used for focusing the electron beam.

To reduce e-beam writing time, active gate lines were separated from contact pads. After writing the active gate using e-beam, photo resist was coated and exposed using i-line stepper for long channel device gates and gate contact pads both of which have large area. Figure 3-17 shows key sequences of the two-step lithography which was used for this work.
First, the 4% solid HSQ was spun at 1800 RPM and baked at 120 °C for 2 minutes on a hot plate. Nano patterns are defined with low energy e-beam lithography. In this work, the acceleration voltage of 10 KV was used for e-beam writing and dose was optimized at 100 nm size with 250 μC/cm². The exposed HSQ resist was developed for 4 minutes in LDD-26w followed by metal cleaning. This metal cleaning step was performed because of potential metal cross contamination during the e-beam lithography process. After metallic cleaning process such as Piranha and SC-1, metal particles were
reduced down to the desirable level of CMOS process. Figure 3-18 shows SEM view of gate lines after etching gate. Gate channel length down to 30 nm was well defined.

![Figure 3-18](image)

(a) (b)

Figure 3-18 SEM view after etching gate with HSQ hard mask.

3.7. Halo implantation for short channel device

In order to reduce the short channel effect (SCE), large tilt-angle halo implantation technique was used [46, 47]. After etching the gate, halo, S/D extension and deep S/D were formed with the conditions listed in Table 3.

<table>
<thead>
<tr>
<th>Dopant</th>
<th>Dose</th>
<th>Energy</th>
<th>Tilt</th>
<th>symmetry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphorous</td>
<td>40 KeV</td>
<td>40 KeV</td>
<td>20°</td>
<td>x2</td>
</tr>
<tr>
<td>Boron</td>
<td>3 KeV</td>
<td>3 KeV</td>
<td>7°</td>
<td>x2</td>
</tr>
<tr>
<td>Boron</td>
<td>7 KeV</td>
<td>7 KeV</td>
<td>0°</td>
<td>-</td>
</tr>
</tbody>
</table>

Traditionally, arsenic has been used for halo implantation but arsenic generates more defects than phosphorous which is not desired for the strained channel material. A single 800 °C 10 sec RTA anneal was used for dopant activation after deep S/D implantation.
2D contours of phosphorous halo concentration are shown in Figure 3-19 where the contours are generated by process simulation using TSUPREM. Doping profiles are extracted from TSUPREM result and used in DESSIS electrical simulation.

![Figure 3-19 2D contour plot for phosphorous concentration extracted from TSUPREM process simulation.](image)

To minimize mobility degradation by impurity scattering, the peak of phosphorous concentration is located under the channel where it is effective for preventing bulk punch-through. Figure 3-20 shows doping profile of 100 nm gate length p-MOSFET extracted from TSUPREM process simulation. Longitudinal doping profiles are extracted 10 nm away from the silicon-oxide interface where the SiGe channel is formed. Because of the large tilt-angle implantation, the doping concentration at the middle of the channel is still high. Vertical doping profiles are extracted under the oxide spacer.
Figure 3-20 Doping profile of 100 nm p-MOSFET extracted using TSUPREM process simulator. (a) Longitudinal doping profile at 10 nm under the gate oxide. (b) Vertical doping profile under the oxide spacer.

Based on simulation results, sub-100 nm Si/SiGe heterostructure p-MOSFET was built with e-beam lithography and NiSi process. Figure 3-21 shows threshold voltage roll-off characteristics of nano scale p-MOSFETs with different halo conditions. Large tilt angle and high dose of phosphorous halo implantation shows better $V_{th}$ roll-off characteristics but also affects on hole mobility. By implementing halo structure and RTA process, we could get a functional device down to $L_{eff}$ 37 nm with swing 130 mV/dec and $V_{th}$ -0.7V. The extraction of effective channel length will be discussed in section 4.2. Figure 3-22 shows device transfer characteristic of short channel devices with different channel length.
Figure 3-21 Experimental data: $V_{th}$ roll-off characteristics: (a) with different halo tilt angles (b) with different halo doping conditions at tilt angle 20°.

Figure 3-22 Experimental data of Si/SiGe hetero-structure short channel device transfer characteristics.
3.8. Chapter Summary

A detailed process for fabrication of sub-100 nm Si/SiGe heterostructure device was discussed. To maximize the Ge concentration in the SiGe channel, low temperature NiSi process was developed with CMP-free FUSI gate process and low resistance source/drain. Electron beam lithography with HSQ e-beam resist was developed and used to form a sub-100 nm device. By using large tilt angle halo and low energy extension implantation, the short channel effect was minimized and reasonably scaled MOSFETs were fabricated.
4. Characterization of Si/SiGe bulk p-MOSFET

Electrical characterization of Si/SiGe heterostructure p-MOSFET is discussed in this chapter. First, the effect of dopant in the SiGe channel and intrinsic layer is studied in detail. Unwanted phosphorous dopant was observed in the SiGe channel layer, which results in the loss of mobility enhancement from compressive SiGe channel material. To reduce the dopant in the SiGe channel, different heterostructure was tested. Next, a new channel length and mobility extraction methods were developed to investigate the mobility dependency on channel length at sub-100 nm dimension. Gate length dependent mobility behavior extracted using this method was shown in this chapter. A compact model is also used to check the short channel mobility and velocity, and excellent agreement with experimental data is shown.

4.1. Pseudomorphic Si/SiGe Heterostructure

Growing the strained SiGe channel directly on bulk Si simplifies the layer growth process. It is also possible to have lower Ge concentration for certain amounts of strain which result in less Si-Ge inter-diffusion during the high temperature process. By lowering Si-Ge inter-diffusion in the strained layer, the peak Ge concentration can be maintained and thus have mobility enhancement. Figure 4-1 shows the layer structure used in this work. Epitaxial layers were grown in a low pressure chemical vapor deposition (LPCVD) system. A 300-nm-thick n-type silicon is grown at 1080 °C on n+ silicon substrate for surface treatment. A 2 μm-thick highly doped n-type silicon layer was then grown at 900 °C with different phosphorous concentrations from $10^{16}$ cm$^{-3}$ to $10^{18}$ cm$^{-3}$. This n-type
silicon layer forms the well structure of p-MOSFET and affects the electrical properties of Si/SiGe p-MOSFET. A 30 nm-thick intrinsic layer was then grown at 680 °C to separate the channel layers from the underlying 2 μm-thick highly doped layer. A 7 nm-thick pseudomorphic SiGe with 55% Ge concentration was then grown at 525 °C to serve as hole channel region. 5 nm-thick intrinsic Si capping layers were then grown at 600 °C to allow for good gate oxide formation.

![Layer structure for Si/SiGe directly grown on silicon bulk.](image)

**Figure 4-1 Layer structure for Si/SiGe directly grown on silicon bulk.**

### 4.1.1. Dopant pile up in SiGe channel

In the transition of layer growth, an abrupt doping change from the heavily doped phosphorous layer, within the 30 nm-thick intrinsic layer is desired. The abrupt doping profile and thinner intrinsic layer are necessary for deeply scaled sub-100 nm devices with high mobility and suppression of short channel effect. However, there is a redistribution of phosphorous in the intrinsic layer during the growth because of a strong affinity of phosphorous for the silicon surface [48]. SIMS profile of this structure after MOSFET processing is shown in Figure 4-2.
Figure 4-2 SIMIS profile of P and Ge in the Si/SiGe heterostructure after full CMOS process.

Gradual decrease of phosphorous was seen in the intrinsic layer from the interface between high-doped and intrinsic layer to the surface. Phosphorous was also observed in the SiGe channel layer, which results in the loss of mobility enhancement in SiGe channel. Impurities in these layers create coulomb scattering centers and hence they degrade low field carrier mobility. In Figure 4-3, C-V characteristics of Si/SiGe heterostructures are shown using DESSIS simulation with different doping profiles along with measurement data. In this test structure, the thickness of the intrinsic buffer layer is 50 nm and the highly doped region has 1E18 phosphorous concentration. Phosphorous doping profile in Profile 1 is matched to doping profile extracted from SIMS result except in SiGe region. Profile 2 has phosphorous peak of 7x10^{17} \text{cm}^{-3} in SiGe channel in addition to Profile 1. Profile 3 and profile 4 are step and constant doping profiles,
respectively. Well matched result with experimental measurement data was shown with doping profile 2 which has phosphorous peak in SiGe layer. This result also indicates the auto doping effect during the SiGe layer growth which is grown even lower temperature.

![Figure 4-3 Comparison of CV data from measurement and DESSIS simulation: (a) C-V characteristics with different doping profile shown in (b)](image)

This phosphorous impurity near the channel region can degrade the carrier mobility but also serves as punch-through stopper for short channel devices. Therefore, optimization of the layer structure and growth should be done to maximize mobility enhancement in sub-100 nm heterostructure devices while maintaining a good electrostatic integrity.

### 4.1.2. **SiGe grown on highly doped Sb substrate**

In this work different types of layer structures are tested to reduce the dopant redistribution in the intrinsic layer. The Intrinsic silicon buffer layer and the SiGe channel
layer are grown directly on the highly doped Sb substrate to prevent dopant pileup in the SiGe channel area. Instead of using phosphorous for highly doped region, Sb doped substrate was directly used for high doping material. This structure reduces the redistribution of impurities in the intrinsic layer and recovers the low field mobility loss. The thickness of the intrinsic layer also can be reduced by eliminating impurities in this layer, which results in improvement of the short channel effect. Figure 4-4 illustrates two different structures tested in this work.

![Figure 4-4](image)

Figure 4-4 The comparison of two layer structures with different high doped region: (a) phosphorous in-situ doping (b) highly doped Sb substrate wafer

To reduce dopant pileup near the SiGe channel region, the silicon intrinsic layer and the SiGe were grown directly on highly doped Sb wafer. Different thickness of intrinsic buffer layer was tested to achieve the same electrostatic behavior as phosphorous in-situ doped substrate. By eliminating high temperature in-situ doping process on Sb wafer, the intrinsic silicon buffer layer has fewer impurities and hence the thickness can be reduced.
without the adverse mobility effect. Threshold voltage roll-off characteristics are shown in Figure 4-5 with different substrate structures.

![Threshold Voltage vs. Effective Channel Length](image.png)

**Figure 4-5** Threshold voltage roll-off characteristic with different layer structure under the SiGe channel.

A 3 nm intrinsic silicon buffer layer on Sb doped wafer showed very similar $V_{th}$ roll-off characteristics with 50 nm intrinsic silicon buffer on phosphorous in-situ doped substrate. This result confirms that there is less dopant distribution close to SiGe channel region even in 3 nm thick intrinsic buffer layer by eliminating high temperature in-situ doping process. Long channel mobility gain was compared in these devices.
Figure 4-6 Low field mobility enhancement with Sb substrate by reducing dopant pileup.

As shown in Figure 4-6, Sb substrate device shows higher mobility at low field region because of low impurity concentration in SiGe channel which results in low coulomb scattering. Mobility enhancement can be achieved by bringing up the highly doped region close to the surface while lowering the impurity concentration in SiGe channel.

4.1.3. The effect of layer thickness, doping and Ge fraction

Threshold voltage roll-off characteristics and hole effective mobility in long channel devices were observed with different substrate doping concentration and thickness of intrinsic silicon buffer layer. In this study, Ge concentrations of 40% and 55% were used for device fabrication. Figure 4-7 shows $V_{th}$ roll-off characteristics of various device fabrication conditions. Sub-100 nm short channel devices are also shown in this plot.
along with long channel device. Transistors with gate length size longer than 0.4 μm were patterned using i-line stepper and with gate length less than 0.4 μm were patterned by electron beam lithography. A 0.6 V threshold voltage shift was shown by changing the thickness of the intrinsic layer from 50 nm to 30 nm. A 0.3 V shift by changing doping concentration from $10^{17}$ cm$^{-3}$ to $10^{18}$ cm$^{-3}$.

Figure 4-7 Threshold voltage roll-off characteristics with different thickness of intrinsic buffer layer and doping level of highly doped silicon substrate.

Figure 4-8 shows effective hole mobility at $5 \times 10^{12}$ cm$^{-2}$ inversion charge density with different substrate doping conditions and Ge concentration. The mobility was also compared with silicon surface channel reference device and showed 3.5 X mobility enhancements. 55% SiGe with 50 nm intrinsic silicon buffer layer on 2 μm phosphorous $10^{18}$ cm$^{-3}$ in-situ doped structure was used for the sub-100 nm Si/SiGe heterostructure p-
MOSFET. An additional mobility enhancement is expected from Sb substrate as discussed in section 4.1.2.

![Figure 4-8 Hole effective mobility as a function of substrate doping level.](image)

**4.2. Device parameter extraction**

In the analysis of advanced CMOS technologies, the extraction of basic MOSFET parameters is important. In particular, accurate measurement of the effective channel length and the source/drain resistance is necessary to extract effective mobility in short channel devices. There are several methods to determine these parameters based on linear drain current [49, 50]. However these methods are not stable if a wide range of channel length is used and give average values of parameters in the measurement range [19, 51]. Shift and ratio methods [52, 53] increase the accuracy but are limited to 15nm resolution which is not enough for devices in the sub-100 nm range. These methods are based on constant mobility model which is not valid for short channel devices with halo
implantation and strained material. Split C-V method [54, 55] does not rely on this assumption and thus can give higher accuracy of parameters. However, overlap capacitance which includes inner/outer fringe capacitance and pad capacitance has prevented accurate effective channel length extraction in sub-100 nm device [56]. Scaling of device dimension is another obstacle for capacitance based parameter extraction method because of high gate current and poor signal to noise ratio. Detailed parameter extraction method used in this work will be discussed in the following sub sections.

4.2.1. Measurement frequency and de-noising

Capacitance based parameter extraction is a reliable method for short channel device even with non-uniform lateral doping profile. However, the capacitance measurement itself is not easy for nano scale devices because of the signal level and noise of the measurement tool. Figure 4-9 shows the total gate-to-channel capacitance as a function of gate voltage with different measurement frequency on the device with 180 nm gate length and 50 μm channel width. Because of the small dimension, the measurement values are very noisy at low frequency. To obtain a better signal-to-noise ratio, the measurement frequency should be increased until there is little change of average value of capacitance compared to the capacitance at lower frequency. Therefore, the frequency of C-V measurement should be carefully chosen considering device dimension, material properties, parasitic resistance and measurement range. In this thesis 1 MHz was used for devices less than 200 nm gate length.
Another way to reduce the noise is by taking many measurements at same bias condition and then averaging these values. This method is an easy way to have better signal but takes long time to complete a C-V plot. Moreover, staying at high voltage for long time can degrade device especially with thin oxide and thus can change electrical properties. To avoid device degradation, measurement was done by averaging five measurements at each bias condition. The removal of noise from noisy measurement data to obtain the clean capacitance signal was done by the “wavelet shrinkage” method [57]. The idea is to first represent the data in terms of a wavelet basis, then the coefficients that are below a certain threshold are set to zero and the data are reconstructed using reduced set. The threshold determines how much noise is suppressed and the larger the variance of the noise, the larger it should be [58, 59]. Figure 4-10 shows noisy measurement data and the result of wavelet de-noising.
Figure 4-10 Intrinsic gate to channel capacitance with denoising process on short channel devices (Width = 10 μm, Tox = 4 nm).

4.2.2. Source/Drain series resistance

The series resistance reduces the internal MOSFET gate-to-source voltage and thus results in drive current degradation. As the device dimensions scale down, source/drain series resistance becomes a large portion of total device resistance. Therefore, the current degradation due to the series resistance is serious in a short channel device. In order to characterize the intrinsic device performance like mobility and velocity, accurate measurement of series resistance is necessary. Conventionally, source/drain series resistance is extracted from the intersection point of total resistance vs. channel length plot [60] which relies on constant mobility and lateral channel doping [52, 61]. The parameters extracted from these methods are gate-voltage independent, whereas source/drain resistance changes, in fact, by gate bias [62]. Several techniques have been
developed in [63-65] in order to address this issue. Figure 4-11 shows gate bias dependent source/drain resistance for long channel extracted from total resistance vs. channel length plot considering gate-bias dependency and threshold voltage shift.

Figure 4-11 Total source/drain resistance as a function of gate-to-source/drain voltage extracted from the Si/SiGe p-MOSFET with $V_{th} = 1V$.

The gate-bias dependent $R_{SD}$ is more serious in short channel devices due to the shallow junction formation and halo implantation. This can be problematic in extracting effective mobility with respect to channel length because of the nonuniform lateral doping profile. To exclude the source/drain resistance effect on mobility, the two-MOSFETs-method was used in this study and will be discussed later in this section.

4.2.3. Physical channel length correction

As the transistor size shrinks to sub-65 nm, the effective channel length reaches down to half of the physical gate length. In this scale, small errors in extracted effective channel
length can result in misinterpretation of device characteristics such as effective mobility and velocity. Several extraction methods have been proposed based on the transistor’s resistance at low drain voltage [62, 66]. These methods underestimate low field mobility in short channel devices [67, 68]. The capacitance-based method does not rely on an assumption of constant mobility and provides more accurate effective channel length than resistance methods for short channel devices [69, 70].

In using the capacitance method, reducing the uncertainty of physical gate length is crucial for extracting effective channel length. Physical gate length has been measured using top-view of SEM after etching the gate. Accurate extraction of effective channel length needs accurate measurement of physical channel length for every size using SEM to mitigate process variations resulting from e-beam lithography and reactive ion etching process. In this work, a simple method is introduced to correct physical channel length using total capacitance vs. channel length plot. The total gate capacitance is given by

\[
C_{total}(V_g) = C_{ox}(V_g) \cdot W_{eff} \cdot L_{eff} + C_{para} \\
\quad = C_{ox}(V_g) \cdot W \cdot (L - \Delta L) + C_{para}
\]

where \(C_{ox}(V_g)\) is intrinsic gate capacitance per unit area and \(W_{eff}\) and \(L_{eff}\) are effective channel width and length, respectively. \(C_{para}\) includes inner/outer overlap capacitance and pad component as shown in Figure 4-12 and equation below.

\[
C_{para}(V_g) = C_{inov}(V_g) + C_{ov1} + C_{ov2} + C_{pad}
\]
Figure 4-12 Schematic of intrinsic parasitic component of MOSFET. $C_{pad}$ is the pad capacitance and not shown in this figure.

The effective channel width $W_{eff}$ can be equated to the known width $W$ for the wide width devices and $L_{eff}$ can be written as $L - \Delta L$ where $L$ is physical gate length and $\Delta L$ is channel length reduction by S/D overlap. Figure 4-13 shows the total gate capacitance, including parasitic capacitance, as a function of channel length. The channel length used in plot (a) is e-beam drawn sizes and not corrected yet. The channel length in plot (b) is not corrected as well but the correction factor is very small compared to gate length and thus can be ignored.
Figure 4-13 Total gate capacitance as a function of drawn channel length: (a) channel length defined by e-beam lithography for short channel (b) optical lithography for long channel.

If we take derivative total gate capacitance with respect to channel length Eq. 4-1 can be reduced to

\[ k(V_g) \equiv \left. \frac{dC_{\text{total}}}{dL} \right|_{V_g} = C_{ox}(V_g) \cdot W \quad 4-3 \]

For strong inversion, the slope \( k \) at given gate voltage should be the same for both long and short channel device. Because long channel devices have relatively small parasitic capacitance, the capacitance slope \( k \) can be easily found by C-V measurement. The short channel capacitance vs. channel length and the slope, \( k \), extracted from long channel devices are shown in Figure 4-14.
Figure 4-14 Total gate capacitance as a function of drawn channel length: the slope of dashed line is extracted from long channel devices.

As can be seen in this plot there is discrepancy between short channel and long channel C-L plot. This indicates that the physical channel length of short channel devices should be corrected to match with long channel C-L plot which has little variation in channel length and gate oxide thickness. For example, actual physical gate length of 100 nm is smaller than the e-beam drawn size and 180nm is longer than the drawn size. This effect can also be seen in physical gate length vs. e-beam lithography drawn size as shown in Figure 4-15.
Figure 4-15 Electron beam lithography pattern skew between drawn and actual size. 10 KV acceleration voltage with 250 uC/cm² in dose.

The physical gate length was measured using SEM top view of gate polysilicon after etching gate. The dose of electron beam lithography was optimized at 120 nm where drawn size and physical gate length are the same. This skew between actual size and drawn size can be explained by proximity effect of low energy electron beam lithography [71, 72]. This proximity effect corresponds with the discrepancy in C-L plot shown in Figure 4-14.

By using the C-L plot channel length correction, accurate extraction of effective channel length is possible without measuring all sizes of gate length which is susceptible to process condition like e-beam current, HSQ lifetime and etching condition. Gate length can be corrected by monitoring one size which is optimized for e-beam lithography. Figure 4-16 shows intrinsic gate-to-channel capacitance vs. physical gate
length plot which give channel length reduction $\Delta L$ from the x axis intercept of straight line.

![Graph showing gate length vs. capacitance]

Figure 4-16 Effective channel length extraction plot using capacitance method. Open symbol represent corrected gate size using C-L plot.

### 4.2.4. Intrinsic gate-to-channel capacitance

The total gate-to-channel capacitance is the sum of intrinsic and parasitic capacitance. The parasitic component includes gate length independent capacitance below the threshold voltage and this capacitance has gate voltage dependent inner overlap capacitance which couples the gate and source/drain junction through the substrate in the depletion regime [56]. To extract intrinsic gate-to-channel capacitance, gate voltage dependent inner capacitance should be subtracted from total gate capacitance. In this study, two MOSFETs are used for extracting intrinsic capacitance per unit length [73]. The intrinsic gate-to-channel capacitance per unit length is given by

\[
C_{\text{intrinsic}} = \frac{1}{L_{\text{gate}}} \int C_{\text{int}}(V_g) \, dV_g
\]
\[ C'_{gc}(V_g, L) = \frac{C_{gc2}(V_g, L_2) - C_{gc1}(V_g, L_1)}{L_2 - L_1} \]

where \( L_1 \) and \( L_2 \) are physical gate lengths corrected by capacitance vs channel length plot. Two adjacent gate lengths were used to avoid threshold voltage mismatch. The intrinsic gate-to-channel capacitance can be found for each gate length by multiplying \( C'_{gc} \) by the physical gate length. Figure 4-17 shows the comparison of effective mobility, extracted by the split C-V method, after and before inner overlap capacitance correction. The origin of underestimation of effective mobility in uncorrected split C-V method is mainly due to the overestimation of inversion charge by overlap capacitances.

---

Figure 4-17 Comparison of effective mobility before and after capacitance correction

### 4.2.5. Effective carrier mobility

The effective mobility is a key parameter that characterizes the carrier transport in heterostructure devices. For short channel device, we have to consider variation of
effective mobility with gate length. The extraction of basic parameters such as the effective channel length, the source/drain resistance and intrinsic gate-to-channel capacitance of short channel device is important for the calculation of effective mobility. The uncertainty in these basic parameters affects significantly the accuracy of extracted mobility. The dR/dL method [19] is a useful technique to monitor mobility because it extracts mobility without using values of $L_{eff}$ and $R_{sd}$. However, this method has low accuracy at sub-100 nm devices and assumes a constant mobility model which is not verified for modern MOSFET technologies [74]. In this thesis, modified two-MOSFET method was used to extract effective mobility from short channel devices. This method includes the effect of gate voltage dependent source/drain resistance and inner overlap capacitance [73, 75]. Figure 4-18 illustrates the basic idea of two MOSFETs methods.

![Two MOSFETs Method Illustration](image)

Figure 4-18 Illustration of two MOSFETs method for effective mobility extraction

The effective mobility of MOSFET with channel length $L_2$ is given by
\[ I_2 = W \cdot Q_{\text{inv}} \cdot \mu \cdot \frac{\Delta V}{\Delta L} \]
\[ = \int_{V_{\text{oc}}}^{V_g} \frac{C_2 - C_1}{\Delta L} dV \cdot \mu \cdot \frac{\Delta V}{\Delta L} \]

where \( Q_{\text{inv}} \) is the mobile channel sheet charge density which is calculated by integrating intrinsic gate-to-channel capacitance. \( \Delta V \) is the inner channel voltage drop through \( \Delta L \) which is determined by \( L_2 - L_1 \). The effective mobility is then given by

\[ \mu = \frac{I_2}{\int (C_2 - C_1) dV \cdot \frac{\Delta L^2}{\Delta V}} \]

The internal voltage drop along the length \( \Delta L \) can be calculated using external drain voltage \( V_D \) and current \( I_1 \) and \( I_2 \). If we put \( R_{SD} \) as external source/drain resistance and \( r_{ch} \) as channel resistance per unit length, the external voltage \( V_D \) is given by

\[ V_D = I_1 \cdot (R_{SD} + r_{ch} \cdot I_1) \]
\[ = I_2 \cdot R_{SD} + I_2 \cdot r_{ch} \cdot L_1 + \Delta V \]

and thus the internal voltage drop \( \Delta V \) is given by

\[ \Delta V = R_{SD} (I_1 - I_2) + I_1 \cdot r_{ch} (I_1 - I_2) \]
\[ = (I_1 - I_2) \cdot (R_{SD} + L_1 \cdot r_{ch}) \]
\[ = (I_1 - I_2) \cdot \frac{V_D}{I_1} \]

The channel length difference can be calculated using the total capacitance vs. channel length plot as discussed in section 4.2.4 where the slope \( k \) is given by

\[ k(V_g) \equiv \frac{dC_T}{dL} = C_{\text{ox}} (V_g) \cdot W \]

then channel length difference \( \Delta L \) and internal voltage drop are
\[ \Delta L = \left[ \frac{\Delta C_{\text{total}}}{k} \right]_{V_{g \text{max}}} \quad , \quad \Delta V = (I_1 - I_2) \cdot \frac{V_D}{I_1} \] 4-10

The channel length difference \( \Delta L \) is extracted at strong inversion where source/drain resistance and inner overlap capacitance have less effect on gate voltage. The final effective mobility can be given by

\[
\mu = \frac{I_1 \cdot I_2}{\int (C_2 - C_1) dV} \times \left( \frac{\Delta C_{\text{total}}}{k} \right)_{V_{g \text{max}}} \cdot \frac{(I_1 - I_2) \cdot V_D}{(I_1 - I_2) \cdot V_D} \] 4-11

Note that capacitance \( C_1 \) and \( C_2 \) are measurement data including parasitic capacitance. Gate voltage dependent source/drain resistance is also included in this equation by canceling out with currents \( I_1 \) and \( I_2 \). All parameters used in this equation can be determined by direct measurement of short channel devices.

Figure 4-19 Comparison of effective mobility extracted by two different methods: solid symbol is dR/dL method and open symbol is two MOSFETs method
Figure 4-19 shows the effective mobility extracted by two different methods, dR/dL and two MOSFETs. Six different channel lengths ranging from 80 nm to 180 nm are used for the dR/dL method and two MOSFETs with gate length 98 nm and 118 nm are used for two MOSFETs method. The mobility extracted by the dR/dL method is higher than the mobility extracted by two-MOSFET method. This is because the dR/dL method includes devices with longer channel while the two-MOSFET method uses only two short channel MOSFETs which have lower mobility than long channel devices. The mobility extracted by two MOSFETs methods is reliable in wide range of inversion charge because it includes gate bias dependent source/drain resistance which is constant for the dR/dL method.

4.2.6. Gate length dependent mobility

To suppress the short channel effect, channel doping concentration needs to be increased. This doping is by a combination of in-situ channel doping during the layer growth step, and halo implantation after patterning the gate. Due to the increased doping concentration as the halos of two sides (source and drain) merge in short channel devices, the effective mobility is expected to be reduced by coulomb scattering and high vertical electric field. The effective hole mobility versus the effective channel length characteristics is shown in Figure 4-20. The hole mobility is extracted by two MOSFETs method at three different positions on the wafer at inversion charge density $5 \times 10^{12}$ cm$^{-2}$. The effective channel length is extracted using the capacitance method after physical channel length correction.
Figure 4-20 Low field hole mobility versus effective channel length. The mobility is extracted by two MOSFETs method at inversion charge $N_i = 5 \times 10^{12} \text{ cm}^{-2}$.

The mobility extracted in sub-100 nm devices is lower than the mobility extracted on long channel devices with channel length 3 to 5 µm, which is 250 cm²/Vs. This can be explained by increased channel doping concentration due to the halo implantation. However, in order to have mobility degradation to 100 to 150 cm²/Vs, the channel doping concentration on short channel device should be higher than $10^{19} \text{ cm}^{-3}$, which can be extrapolated using Figure 4-8, effective mobility versus doping concentration plot. Such high concentrations are not very likely in these devices. Other possible mobility degradation factors are non-uniform defects and process induced strain changes. Defects can be generated during the ion implantation process and cured at RTA process. In this study, limited heat treatment, 800 °C 10 sec RTA, was used to maintain Ge peak in the
channel. This might be not enough to cure damages created during the ion implantation process. Further work is required to understand the degradation mechanism.

**4.2.7. Effective velocity**

Regardless of deep scaling of channel length, the drain saturation current is limited by carrier velocity injected from the source into the channel [76, 77]. Therefore, the carrier virtual source velocity is the main driving force for future nano-scale MOSFETs and continuous increase of carrier velocity is necessary to maintain the performance trend of MOSFETs [78, 79]. In this study, we extracted hole velocity on nano-scale Si/SiGe heterostructure pMOSFETs to see the potential velocity benefit over the Si devices.

Traditionally, carrier velocity has been extracted using the transconductance method [80] which is given by

\[ V_{\text{eff}} = \frac{g_{mi}}{W C'_{ox}} \]  

where \( C'_{ox} \) is the gate oxide capacitance per unit area in inversion, \( g_{mi} \) is the peak transconductance corrected for source/drain parasitic resistance, \( W \) is the device width and \( V_{\text{eff}} \) is the effective carrier velocity. However, in order to see the limit of device performance, carrier velocity injected from the source is important. In this study, the \( V_{id} \) method introduced by Lochtefeld et al [29] was used to extract velocity, which gives inversion layer carrier velocity closer to source injection point than the transconductance method. Carrier velocity \( V_{id} \) is given by
where the $\Delta V_t$ is threshold voltage shift due to the DIBL effect and threshold voltage roll-off characteristics, $I_dR_s$ is voltage drop on source resistance and $C'_{gsd}$ is gate capacitance per unit area measured on long channel device.

**DIBL characteristics of nano-scale Si/SiGe pMOSFETs fabricated through this work are shown in Figure 4-21. Devices with effective channel length down to 60 nm show DIBL less than 180 mV/V.**

---

Effective carrier velocity was extracted by using the $\nu_{id}$ method. Figure 4-22 shows extracted hole velocity as a function of DIBL (a) and effective channel length (b). The result indicates that the effective hole velocity is still below 36% of the thermal injection velocity of Si and 25% that of 55% SiGe [81, 82].
As shown in Figure 4-22 (b), the effective velocity starts to decrease below 100 nm in channel length. This is most likely related to mobility degradation in short channel device which was shown in Figure 4-20. For further down scaling of pMOSFETs using SiGe material, the channel doping should be carefully optimized considering mobility, velocity and electrostatic integrity.

4.3. Compact model vs Experimental data

A simple semi-empirical model for short channel MOSFET developed by A. Khakifirooz and D. A. Antoniadis (unpublished) was used to compare with measurement result. This compact model uses only eight parameters which are $v_x$, $\mu$, $R_s$, gate oxide capacitance, swing, DIBL factor, off state leakage current and effective channel length. The last five
parameters can be directly extracted from measurement data. In this model, current is modeled by the product of charge density times the carrier velocity at "virtual source", i.e. at the location of the top of the energy barrier. This is illustrated in Figure 4-23

![Virtual Source Diagram]

Figure 4-23 Illustration of the location of virtual source.

and the current is defined at this point and given by

$$I_{D}/W = Q_{\text{v.s.}} v_{v.s} F_s$$  \hspace{1cm} 4-14

where the virtual-source charge density can be approximated by the empirical function given by

$$Q_{\text{v.s.}} = C_g n_{\text{v.s.}} \ln \left( 1 + \exp \frac{V_{\text{GS}} - (V_T - \alpha \phi_s F_s)}{n \phi_s} \right)$$  \hspace{1cm} 4-15

Functions $F_s$ and $F_I$ are saturation and inversion transition functions which smooth the transition between two different operation regions. Effective mobility and velocity for 88 nm Si/SiGe p-MOSFET were fitted using optimization tool kit in MATLAB software. Mobility and velocity are listed in Table 4 and fitting results are shown in Figure 4-24.

Table 4 Mobility and velocity comparison of experimental and model

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility (cm^2/Vs)</td>
<td>~ 160</td>
<td>186</td>
</tr>
<tr>
<td>Velocity (x10^6 cm/sec)</td>
<td>4</td>
<td>3.8</td>
</tr>
</tbody>
</table>
Figure 4-24 Compact model fitting result for 88 nm Si/SiGe pMOSFET

4.4. Chapter Summary

In this chapter, electrical measurement of Si/SiGe heterostructure p-MOSFET was described. The dopant pileup in the SiGe channel and intrinsic layer was shown using SIMS profile and device simulation. Loss of mobility enhancement from compressive SiGe channel material was observed and attributed to unwanted dopant pileup. Different layer structures were fabricated and tested to reduce the dopant in the SiGe channel. Finally, a new mobility extraction method was developed and used to obtain the channel length dependency of hole mobility with sub-100 nm in size. A compact model was also used to evaluate the short channel mobility and velocity, and excellent agreement with experimental data was shown.
5. Summary and future work

In this final chapter, the thesis is summarized. It also provides the key contribution and suggestions for future work.

5.1. Thesis Summary

The goal of this thesis was to give a fabrication guideline for Si/SiGe heterostructure sub-100 nm p-MOSFET. Various pseudomorphic SiGe heterostructures were tested along with short channel device electrostatic performance.

A Nickel silicide process was studied for fully silicided metal gate and source/drain silicidation. A new and simple FUSI gate process was developed to minimize process variation and poly depletion effect. For a high quality nickel mono silicide, a two step ramping technique was used, which reduces residual stress and defects in the film. Source/drain resistance was effectively reduced by nickel silicidation. By etching back the spacer oxide, the resistance of S/D extension was also reduced while the deep source/drain region away from the channel.

Low temperature process is essential to maintain Ge profile in heterostructure devices. However, annealing defects generated during the implantation step is also important for mobility enhancement and reduction of off-state leakage current in short channel devices. Several different RTA conditions are tested in this study. 700 °C 10 sec RTA was not enough to anneal defects in source/drain, which resulted in large junction leakage current. This leakage current is originated from the SiGe area due to the narrow bandgap. In this study, 800 °C 10 sec was used to maintain Ge peak in the channel and S/D activation. Short channel effect is another factor for optimizing RTA condition.
Nano scale gate patterning using e-beam lithography was successfully implemented with halo implantation to achieve sub-100 nm Si/SiGe pMOSFETs. For halo pocket implantation phosphorous was used instead of arsenic to reduce implantation damage. The peak of the pocket implantation was located under the SiGe channel to maintain mobility gain.

Extraction methodologies of fundamental MOSFET carrier transport properties were covered. By using a two-MOSFET method, the mobility was extracted over a wide range of inversion charge at very localized gate length. Mobility degradation was observed in sub-100 nm Si/SiGe pMOSFETs but there is still mobility enhancement compared to Si. The velocity measurement indicates that the effective hole velocity is still below 36% of the thermal injection velocity of Si and 25% that of 55% SiGe. The velocity and mobility fitted by a compact model showed excellent agreement with experimental data.

5.2. Contributions

1. Nickel Silicide process was developed for SiGe short channel device
   - CMP-free nickel fully silicide gate process
   - Low resistance NiSi source/drain
   - Optimization of silicidation condition
   - Cross contamination test for CMOS process compatibility

2. Si/SiGe heterostructure sub-100 nm p-MOSFET was built
   - Electron beam lithography set up for ICL process
   - Halo implantation optimization
   - Si/SiGe channel structure and doping optimization

3. Characterization of sub-100nm Si/SiGe p-MOSFET
• Channel length correction method for short channel device
• Mobility extraction method
• Nano scale Si/SiGe p-MOSFET mobility
• Model parameter optimization

5.3. Suggestions of Future Work

• Identification of short channel mobility degradation
• Optimization of layer grow considering dopant pileup
• Short channel device on Sb substrate
• Uniaxial strain applied to biaxial compressive strained SiGe
• Optimization of halo implantation for better short channel integrity
## Appendix A

### Nano-scale Si/SiGe p-MOSFET Fabrication Flow

<table>
<thead>
<tr>
<th>Step</th>
<th>Method</th>
<th>Station</th>
<th>Conditions</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCA Cleaning</td>
<td>ICL</td>
<td>RCA station</td>
<td>Piranha (5min) + 50:1 HF 15sec + SC2 5min</td>
<td></td>
</tr>
<tr>
<td>LTO Deposition</td>
<td>ICL</td>
<td>6C</td>
<td>Rec: 400 53A SPK, 50min</td>
<td></td>
</tr>
<tr>
<td>Tox monitoring</td>
<td>ICL</td>
<td>UV1280</td>
<td>Target = 3000A</td>
<td></td>
</tr>
<tr>
<td>Coat PR</td>
<td>ICL</td>
<td>Coater 6</td>
<td>Rec: T1HMDS, Standard coat</td>
<td></td>
</tr>
<tr>
<td>Pattern Active</td>
<td>ICL</td>
<td>i-stepper</td>
<td>Rec: JK-ACT</td>
<td></td>
</tr>
<tr>
<td>Develop and PR reflow</td>
<td>ICL</td>
<td>Coater 6</td>
<td>Rec: JK-FLOW (Standard Dev and 170C 1min)</td>
<td></td>
</tr>
<tr>
<td>Etch Active Oxide</td>
<td>ICL</td>
<td>AME5000</td>
<td>Rec: JK-FOX (leave 200A on active)</td>
<td></td>
</tr>
<tr>
<td>Tox monitoring</td>
<td>ICL</td>
<td>UV1280</td>
<td>Target = 200A</td>
<td></td>
</tr>
<tr>
<td>Ashing PR</td>
<td>ICL</td>
<td>Asher</td>
<td>Rec: 3min</td>
<td></td>
</tr>
<tr>
<td>Piranha</td>
<td>ICL</td>
<td>PM Clean</td>
<td>Premetal Clean (Green)</td>
<td></td>
</tr>
<tr>
<td>HF dip (50:1)</td>
<td>ICL</td>
<td>PM Clean</td>
<td>Target: 200A + over etch</td>
<td></td>
</tr>
<tr>
<td>Gate Oxidation</td>
<td>ICL</td>
<td>6D</td>
<td>600C WET 5D, 3 hrs 40 min, ETox=40 A</td>
<td></td>
</tr>
<tr>
<td>Gate Poly</td>
<td>ICL</td>
<td>6A or 6B</td>
<td>6A:N+ doped, 6B: Undoped Target = 1000A</td>
<td></td>
</tr>
<tr>
<td>SiN Deposition</td>
<td>ICL</td>
<td>DCVD</td>
<td>Rec: SiN700A Target 700A</td>
<td></td>
</tr>
<tr>
<td>Back side etch</td>
<td>ICL</td>
<td>AME5000</td>
<td>Poly etch</td>
<td></td>
</tr>
<tr>
<td>HSQ Spin coat</td>
<td>Heid</td>
<td>Spin Coater</td>
<td>500rpm 5s + 1800rpm 45sec</td>
<td></td>
</tr>
<tr>
<td>pre-bake</td>
<td>Heid</td>
<td>hot plate</td>
<td>120C 2min (careful with x-contamination)</td>
<td></td>
</tr>
<tr>
<td>e-beam lithography</td>
<td>SEBL</td>
<td>Raith 150</td>
<td>Acc. Voltage: 10 KV, Dose: -250 uC/cm²</td>
<td></td>
</tr>
<tr>
<td>e-resist develop</td>
<td>TRL</td>
<td>wet station</td>
<td>LDD 26W 4min</td>
<td></td>
</tr>
<tr>
<td>Piranha cleaning</td>
<td>TRL</td>
<td>wet station</td>
<td>degiganter bath and handler</td>
<td></td>
</tr>
<tr>
<td>Coat PR</td>
<td>ICL</td>
<td>Coater 6</td>
<td>Rec: T1HMDS Standard</td>
<td></td>
</tr>
<tr>
<td>Pattern Gate</td>
<td>ICL</td>
<td>i-stepper</td>
<td>Rec: JK-GP</td>
<td></td>
</tr>
<tr>
<td>Develop</td>
<td>ICL</td>
<td>Coater 6</td>
<td>Rec: Dev6</td>
<td></td>
</tr>
<tr>
<td>Etch SiN</td>
<td>ICL</td>
<td>AME5000</td>
<td>Rec: JK_SiN</td>
<td></td>
</tr>
<tr>
<td>Source/Drain Extension</td>
<td>Ashing PR</td>
<td>ICL</td>
<td>Ahser</td>
<td>Rec: 3min</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>-----</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>Etch HSQ</td>
<td>TRL</td>
<td>wet station</td>
<td>HF</td>
<td></td>
</tr>
<tr>
<td>Etch Poly</td>
<td>ICL</td>
<td>AME5000</td>
<td>Rec: JK_SOFTGP</td>
<td></td>
</tr>
<tr>
<td>Ion implantation, Halo</td>
<td>OUT</td>
<td>Innovion</td>
<td>P 40KeV 2E13x2 Tilt=20</td>
<td></td>
</tr>
<tr>
<td>Ion implantation, LDD</td>
<td>OUT</td>
<td>Innovion</td>
<td>B 3 KeV 1E14x2 Tilt=7</td>
<td></td>
</tr>
<tr>
<td>Blue Piranha</td>
<td>ICL</td>
<td>PM Clean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green Piranah</td>
<td>ICL</td>
<td>PM Clean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCA Cleaning</td>
<td>ICL</td>
<td>RCA station</td>
<td>Piranha (5min) + 50:1 HF 15sec + SC2 5min</td>
<td></td>
</tr>
<tr>
<td>LTO Deposition</td>
<td>ICL</td>
<td>6C</td>
<td>Rec: 400 53A Spike 20min</td>
<td></td>
</tr>
<tr>
<td>Tox monitoring</td>
<td>ICL</td>
<td>UV1280</td>
<td>Target = 1000A</td>
<td></td>
</tr>
<tr>
<td>Etch Spacer</td>
<td>ICL</td>
<td>AME5000</td>
<td>Rec: JK-SPACER</td>
<td></td>
</tr>
<tr>
<td>Tox monitoring</td>
<td>ICL</td>
<td>UV1280</td>
<td>Target = 100A</td>
<td></td>
</tr>
<tr>
<td>Ion implantation, P+</td>
<td>OUT</td>
<td>Innovion</td>
<td>B 7 KeV 4E15 Tilt=0</td>
<td></td>
</tr>
<tr>
<td>Blue Piranha</td>
<td>ICL</td>
<td>PM Clean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green Piranah</td>
<td>ICL</td>
<td>PM Clean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S/D anneal reoxidation</td>
<td>RCA Cleaning</td>
<td>ICL</td>
<td>RCA station</td>
<td>Piranha (5min) + 50:1 HF 15sec + SC2 5min</td>
</tr>
<tr>
<td>RTP</td>
<td>ICL</td>
<td>RTP</td>
<td>800C 10sec</td>
<td></td>
</tr>
<tr>
<td>S/D Reoxidation</td>
<td>ICL</td>
<td>5D</td>
<td>600C 10min</td>
<td></td>
</tr>
<tr>
<td>FUSI</td>
<td>SiN etch</td>
<td>ICL</td>
<td>Hot Phos.</td>
<td>160C 10min</td>
</tr>
<tr>
<td>HF dip</td>
<td>ICL</td>
<td>PM Clean</td>
<td>check spacer</td>
<td></td>
</tr>
<tr>
<td>Nickel Deposition</td>
<td>ICL</td>
<td>e-beam</td>
<td>600A 1A/sec</td>
<td></td>
</tr>
<tr>
<td>RTA Anneal</td>
<td>ICL</td>
<td>RTA2</td>
<td>300C 3min + 450C 30sec</td>
<td></td>
</tr>
<tr>
<td>Etch Nickel</td>
<td>TRL</td>
<td>wet station</td>
<td>4:1 Piranha Visual inspection</td>
<td></td>
</tr>
<tr>
<td>S/D Silicidation</td>
<td>HF dip</td>
<td>TRL</td>
<td>wet station</td>
<td>expose S/D</td>
</tr>
<tr>
<td>Nickel Deposition</td>
<td>ICL</td>
<td>e-beam</td>
<td>40A 1A/sec</td>
<td></td>
</tr>
<tr>
<td>RTA Anneal</td>
<td>ICL</td>
<td>RTA2</td>
<td>300C 3min + 450C 30sec</td>
<td></td>
</tr>
<tr>
<td>Etch Nickel</td>
<td>TRL</td>
<td>wet station</td>
<td>4:1 Piranha Visual inspection</td>
<td></td>
</tr>
<tr>
<td>Backend process</td>
<td>ILD deposition</td>
<td>ICL</td>
<td>PECVD</td>
<td>5000A</td>
</tr>
<tr>
<td>Coat PR</td>
<td>ICL</td>
<td>Coater 6</td>
<td>Rec: T1HMDS Standard</td>
<td></td>
</tr>
<tr>
<td>Process</td>
<td>Vendor</td>
<td>Tool</td>
<td>Recipe</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>--------</td>
<td>---------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>Develop ICL Coater 6</td>
<td>ICL</td>
<td>Coater 6</td>
<td>Dev6</td>
<td></td>
</tr>
<tr>
<td>Etch contact Dry</td>
<td>ICL</td>
<td>AME5000</td>
<td>JK-CNT leave oxide 200 A</td>
<td></td>
</tr>
<tr>
<td>Etch contact Wet</td>
<td>TRL</td>
<td>wet station</td>
<td>BOE 1min</td>
<td></td>
</tr>
<tr>
<td>Ashing PR</td>
<td>ICL</td>
<td>Asher</td>
<td>Rec: 3min</td>
<td></td>
</tr>
<tr>
<td>Pre-metal cleaning</td>
<td>TRL</td>
<td>Acid hood</td>
<td>Piranha + HF</td>
<td></td>
</tr>
<tr>
<td>Metal deposition front</td>
<td>ICL</td>
<td>endura</td>
<td>Ti 1000A + Al 1um</td>
<td></td>
</tr>
<tr>
<td>Metal deposition back</td>
<td>ICL</td>
<td>endura</td>
<td>Ti 500A + Al 5000A</td>
<td></td>
</tr>
<tr>
<td>Coat PR</td>
<td>ICL</td>
<td>Coater 6</td>
<td>T1HMDS Standard</td>
<td></td>
</tr>
<tr>
<td>Pattern Metal</td>
<td>ICL</td>
<td>i-stepper</td>
<td>JK-MT</td>
<td></td>
</tr>
<tr>
<td>Develop</td>
<td>ICL</td>
<td>Coater 6</td>
<td>Dev6</td>
<td></td>
</tr>
<tr>
<td>Etch Metal</td>
<td>ICL</td>
<td>Rainbow</td>
<td>Standard</td>
<td></td>
</tr>
<tr>
<td>Ashing PR</td>
<td>ICL</td>
<td>Ahser</td>
<td>Rec: 3min</td>
<td></td>
</tr>
<tr>
<td>Sintering</td>
<td>TRL</td>
<td>3A</td>
<td>420C 30min</td>
<td></td>
</tr>
</tbody>
</table>
References


68. Romanjek, K., et al., Characterization of the effective mobility by split C(V) technique in sub 0.1 μm Si and SiGe PMOSFETs. Solid-State Electronics, 2005. 49(5): p. 721-726.


