Electrical and Material Properties of Strained Silicon/Relaxed Silicon Germanium Heterostructures for Single-Electron Quantum Dot Applications

Chiao-Ti Huang

A Dissertation
Presented to the Faculty of Princeton University in Candidacy for the Degree of Doctor of Philosophy

Recommended for Acceptance by the Department of Electrical Engineering
Adviser: Professor James C. Sturm

June 2015
Abstract

A single-electron quantum dot device is an ideal environment to demonstrate the concept of a spin-based quantum bit, a promising candidate to realize a quantum computer. Two-dimensional electron gases (2DEGs) in silicon/silicon germanium heterostructures have been considered as a potential platform to fabricate single-electron quantum dots for spin manipulations because silicon has an inherently longer spin coherence time. Then two different types of silicon 2DEGs, modulation-doped 2DEG and enhancement-mode undoped 2DEG, are discussed. The efforts to improve both 2DEGs into a better material system for quantum computing application are the main focus of this thesis.

A severe leakage issue of the Schottky gating on a modulation-doped 2DEG is resolved by successful suppression of phosphorus surface segregation. A high breakdown voltage is thus achieved in a Schottky gated modulation-doped 2DEG without significant gate leakage current. Implant isolation as an alternative for lateral electrical isolation in a modulation-doped 2DEG at 4.2 K is also successfully demonstrated. It preserves surface planarization and prevents the leakage issues through the corners of etched mesas. The best implant conditions for effective isolation and better thermal stability are examined and determined. The quality of these doped 2DEGs is verified to be unaffected by the implant isolation process.

The transport property of an enhancement-mode 2DEG is significant for a spin-based quantum bit. Various mobility-limiting factors in our undoped 2DEGs grown by RTCVD are identified. Efforts to alleviate these scattering mechanisms lead to mobility as high as 400,000 cm²/Vs and the critical density as low as $3.2 \times 10^{10}$ cm⁻² at 4.2 K. A tunable screening effect on remote charges at silicon/oxide interface is found to greatly improve the transport properties of thin-cap enhancement-mode 2DEGs, which compensates the detrimental influences from the remote charges at the interface, and thus remains the capability for a sharp electron patterning from
the top gates. In addition, theoretical and experimental work on the effect of the regrowth interface in undoped 2DEGs is demonstrated as well.
Acknowledgements

This has been really a long journey. Luckily, I have finally reached the destination of my doctoral life. I wouldn’t be able to make it without assistances from many people. First, I would like to thank Princeton University and Department of Electrical Engineering for providing such an exciting opportunity and comfortable environment for me to pursue my Ph.D. degree. I always remember how excited I was when I received admission with full financial support from Princeton University, just as the world was slammed by the financial crisis in 2009. When I was an undergraduate, I aimed to do my doctoral research in the U.S. and live here with my wife, Chi-Hua. Thanks to Princeton University, I had a chance to achieve my dream.

My adviser, Professor James C. Sturm, is undoubtedly the first person I would like to say thank you to. I am deeply indebted to Prof. Sturm for his stimulating insights and inspiring advises that have guided me through the long journey of my Ph.D.. His perpetual enthusiasm and unwavering supports have always enlightened me, especially when I felt perplexed and lost direction in the past six years. I would also like to thank Professor Barry Rand and Professor Stephen Chou for their precious advice and opinions on my thesis. Furthermore, I would like to thank Professor Jason Petta, Professor Mansour Shayegan and Prof. Stephen Lyon for their guidance in the field of low-dimensional physics in semiconductors and novel quantum computing. A special thanks goes to Professor Daniel Tsui for his continuous support and warm encouragement on my doctoral research.

This work could not be completed without assistance from our collaborators. I would like to express my gratitude to Professor Leonid P. Rokhinson and his students at Purdue University for their helps on low-temperature measurements, Dr. Wei Pan and Dr. Tzu-Ming Lu at Sandia National Laboratories for their useful advice on my experimental results. I also would like to thank HRL Laboratories and researchers there for their support on our wafer growth, and Amberwave Systems Inc. for provid-
ing us high quality relaxed SiGe buffer substrates. Here at Princeton, I am obliged to Professor Mansour Shayegan and his student Hao Deng for their prompt helps on magneto-resistance measurements, and Professor Jason Petta and his student Ke Wang and post-doc Dr. Christopher M. Payette for their stimulating discussions on spin-based quantum dot devices.

I am also thankful to all the members I’ve met in Sturm group. I am so excited to be one of the certificated RTCVD heroes with Keith Chung, Sushobhan Avasthi and Jiun-Yun Li. In particular, I am grateful to Jiun-Yun for his guidance on both research and life. Without his help, it would have been a very painful process for me to get my life started in the U.S.. I also want to say thank you to Sturm group members more senior than me: Bahman Hekmatshoar, Yifei Huang, Kevin Loutherback, Noah Jafferis and Ting Liu for their assistance and advice, and to Sturm group members junior than me: Ken Nagamatsu, Warren Rieutort-Louis, Josue Sanz-Robinson, Janam Jhaveri, Yasmin Afsar, Yu Chen, Alexander Berg, and Levent Aygun for their warm company. In particular, a special thanks must go to Janam Jhaveri for his time on proofreading my thesis. Finally, I would like to say thank you to the group members in my year: Bhadri Visweswaran, Amy Wu and Joseph DSilva. I will never forget the happiness and sadness we have shared since we stepped into the Princeton campus together six years ago. Hey Bhadri! Thank you so much for your support on the CVD maintenance since Jiun-Yun left. I really appreciate it!

I am obliged to all PRISM staff for their unwavering support. I would like to thank Dr. George Watson (Pat), Dr. Mikhail Gaevski, Dr. Conrad Silvestre, Dr. Yong Sun, Dr. Sen Liu, Bert Harrop and Joe Palmer for providing us such a good cleanroom to finish our graduate work. I would also like to thank Dr. Nan Yao and Gerald Poirier for their assistances on material characterizations at IAC. In addition, I am grateful to Sheila Gunning, Kim Hegelbach and Carolyn Arnesen for their helps on plenty of complicated administrative processes.
Taiwanese friends from Princeton Association of Taiwanese Students (PATS) gave us a big warm welcome when we first stepped into the U.S.. A big thanks goes to Yen-Ting Chiu, Kung-Ching Liao, Ya-Chin Chiou, Sophia Liao, Wen-Yi Wu, Yu-Wei Chen, Hao-Ting Wu, Carole-Jean Wu, Yu-Yuan Chen, Brandon Chen, Yue-Kai Huang, Vicky Cheng, Amelia Huang, Kaleb Huang, Ting-Jung Lin, Yin Wang, Tzu-Ming Lu, Jun-Wei Chuah, Michelle Wong, Christine Chuah, Jiun-Yun Li, Amy Dai, Irene Li, Leah Li, Chia-Chun Lin, Yao-Chao Tu, Yu-Cheng Tsai, Huai-An Chin, Jen-Tang Lu and Yao-Wen Yeh. Your warm company made our Princeton life joyful and colorful. The memories with you guys at Princeton are always invaluable gifts for us. A special thanks must go to Huai-An Chin for his supports on the RTCVD maintenance in the last two years.

Finally, I would like to express my deepest gratitude to my family. Since I was a child, my parents always supported every decision I made without any complaints. Their perpetual love and care undoubtedly gave me great courage to overcome so much hard time in the past six years. Without your spiritually and financially assistance, I wouldn’t have been able to finish my Ph.D.. In addition, I would also like to thank my sister and brother-in-law for taking care of our parents when I was at a foreign place 8,000 miles away from home for so long. I am also obliged to my 2-year-old daughter, Bera. The happiness and surprises she brought to us made me understand that this world is indeed so colorful and meaningful. (Of course, her naughtiness also forced her parents to learn their limit of how little sleep they could endure.) Last but not least, I would like to attribute all my success and accomplishments to my beloved, Chi-Hua. She always gave me the warmest hug when I was in difficult times. I am the luckiest man in the world, to have her love.
To my beloved wife, Chi-Hua
Contents

Abstract ................................................................. iii

Acknowledgements .................................................. v

List of Figures ......................................................... xv

1 Introduction ......................................................... 1

1.1 Motivation ......................................................... 1

1.2 Thesis Outline .................................................. 3

2 Two Dimensional Electron Gases in Si/SiGe Heterostructures 7

2.1 Introduction ...................................................... 7

2.1.1 Introduction to a 2DEG ......................................... 7

2.1.2 Introduction to the RTCVD at Princeton University ...... 9

2.2 Characteristics of Silicon 2DEGs ............................... 13

2.2.1 Strain-Induced Band Offset in Si/SiGe Heterostructures .. 13

2.2.2 SiGe Virtual Substrate ......................................... 15

2.2.3 Types of Strained Silicon 2DEGs ............................ 16

2.3 Modulation-Doped Strained Silicon 2DEG ..................... 16

2.3.1 Introduction .................................................... 16

2.3.2 Layer Structure and Epitaxial Growth of a Doped 2DEG ... 17

2.3.3 A Quantum Device on a Doped 2DEG ..................... 19

2.4 Enhancement-Mode Undoped Strained Silicon 2DEG ......... 21
2.4.1 Introduction .................................................. 21
2.4.2 Layer Structure and Epitaxial Growth of an Undoped 2DEG . 22
2.4.3 A Quantum Device on an Undoped 2DEG ..................... 23

3 High Breakdown Voltage Schottky Gating of Doped Silicon 2DEGs 25
3.1 Introduction ..................................................... 25
3.2 Suppression of Phosphorus Surface Segregation .................. 26
3.3 Device Fabrication .............................................. 27
3.4 Gate Leakage Test at 4.2 K ......................................... 28
3.5 Depletion of Doped 2DEGs ...................................... 29
3.6 Quantum Point Contact Test ................................... 30
3.7 Summary ......................................................... 34

4 Implant Isolation of Silicon 2DEGs at 4.2 K 36
4.1 Introduction ..................................................... 36
4.2 Sample Growth and Test Device Fabrication ....................... 37
4.3 Implant Conditions .............................................. 39
4.4 Ultra High Sheet Resistance at 4.2 K ............................. 41
4.5 Thermal Stability ................................................ 41
4.6 2DEG Quality Check .......................................... 44
4.7 Summary ......................................................... 44

5 Enhancement-Mode Undoped Silicon 2DEGs 46
5.1 Motivation ....................................................... 46
5.2 Models for Mobility Limitation .................................. 47
5.2.1 Scattering from Background Charged Impurities .......... 48
5.2.2 Scattering from Remote Charged Impurities ............... 49
5.2.3 Scattering from Si/SiGe Interface Roughness ............... 51
5.2.4 Other Possible Scattering Mechanisms ...................... 52
8 Conclusions

8.1 Conclusions .......................................................... 127

8.2 Future Work .......................................................... 130

8.2.1 Unknown Mobility-Limiting Factors ....................... 130

8.2.2 Isotopically-Enriched Undoped 2DEGs for a Longer Spin Coherence Time ................. 131

A Publications and Presentations .................................... 133

A.1 Journal Articles and Conference Papers ....................... 133

A.2 Conference Presentations ......................................... 134

B Fabrication of Undoped Silicon 2DEGs .......................... 136

B.1 Growth of Undoped Silicon 2DEGs ............................ 136

B.1.1 Baking Before Growth on Si and SiGe Substrates ........ 136

B.1.2 Wet Cleaning for Growth Substrates ....................... 137

B.1.3 Baking and Carrier Wafer Coating ........................ 138

B.1.4 100-mm Carrier Wafer for SiGe Buffer Pieces .......... 138

B.1.5 Standard Layer Structure of an Undoped 2DEG and Its Growth Recipe ..................... 139

B.1.6 A Typical SIMS for a High-Mobility Undoped 2DEG .... 140

B.2 Processes for Enhancement-Mode Undoped 2DEGs .......... 141

B.2.1 Full Processes .................................................. 141

B.2.2 ALD Chamber Cleaning and Deposition Recipes ........ 142

C Non-Standard Wafer Growth ....................................... 144

C.1 Growth on a Thin SiGe Buffer (400 μm) ...................... 144

C.1.1 Option 1: 125-mm Carrier Wafer ......................... 145

C.1.2 Option 2: 100-mm Carrier Wafer ......................... 147

C.1.3 Option 3: 75-mm Quartz Wafer Stand ..................... 153
C.2 Growth on a Thick SiGe Buffer (680 µm) . . . . . . . . . . . . . . . . 154

C.2.1 Dicing an 150-mm SiGe Buffer into a 100-mm Wafer . . . . . 154

C.2.2 Growth Temperature Calibration for a Thick SiGe Buffer . . 154

C.2.3 Growth Uniformity and Mobility Results . . . . . . . . . . . . 156

Bibliography 159
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The schematic of rapid thermal chemical vapor deposition (RTCVD) at Princeton University. (Image courtesy of P. V. Schwartz)</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Original published data of normalized transmission versus temperature for 1.3-μm and 1.55-μm lasers for lightly doped n-type (513 μm) and p-type (493 μm) silicon wafers.</td>
<td>11</td>
</tr>
<tr>
<td>2.3</td>
<td>The data of average lamp power versus temperature calibrated by normalized transmission of two infrared lasers with the extrapolation for temperatures &gt;750 °C. In addition, the lamp powers dropped 6% after a new SCR unit was installed. The common reactor baking power and SiGe buffer baking power used before and after the SCR replacement are both labeled.</td>
<td>12</td>
</tr>
<tr>
<td>2.4</td>
<td>(a) The valley splitting in strained silicon results in a quantum well at the strained Si layer between two relaxed SiGe layers. (b) Electrons tend to accumulate in out-of-plane valleys due to their lower energies.</td>
<td>14</td>
</tr>
<tr>
<td>2.5</td>
<td>(a) A layer structure and (b) a band diagram of a modulation-doped strained silicon 2DEG. Some of electrons from the intentionally doped layer accumulate in the strained silicon quantum well and form a 2DEG.</td>
<td>18</td>
</tr>
</tbody>
</table>
2.6 (a) A layer structure of a modulation-doped 2DEG with Schottky splitting gates. A 2DEG can be patterned by the Schottky gates with negative biases. (b) A three dimensional view for a typical quantum dot device fabricated on a modulation-doped 2DEG. 

2.7 (a) A layer structure and (b) a band diagram of a typical enhancement-mode undoped strained silicon 2DEG. Electrons are capacitively induced in the strained silicon channel by the metal gate with positive biases. (The metal gate is not shown here, and the oxide is partially shown for brevity). 

2.8 (a) The layer structure of an enhancement-mode undoped 2DEG with depletion gates inserted between an oxide and a surface silicon cap. Electrons in the strained silicon channel can be patterned by these depletion gates. (b) A three dimensional view of a quantum dot device fabricated on an enhancement-mode 2DEG.

3.1 The schematic of atomic layer structures near the surface during the epitaxial growth for a doped 2DEG. (Image courtesy of Jiun-Yun Li)

3.2 A SIMS profile for the sample with a cap layer grown at 575 °C. The phosphorus bump in the Si 2DEG is a SIMS artifact.

3.3 The phosphorus segregation can be greatly suppressed by lowering cap layer growth temperature from 575 °C to 525 °C. The phosphorus surface concentration is reduced as low as 2×10^{16} cm^{-3}. The P peak at surface is a typical artifact resulted from the surface effect in a SIMS measurement.

3.4 Schottky gate leakage test at 4.2 K. The inset shows the cross section of a test device.
3.5 Depletion test at 4.2 K of the sample with a cap layer grown at 525°C. Both drain and gate leakage currents are shown in log scale in the inset. .......................................................... 31

3.6 Successful QPC test at 4.2 K without any leakage. A wide transient region is observed. The inset shows the schematic of the QPC test structure. ......................................................... 32

3.7 A SEM image of a typical QPC test device. .................................................. 33

3.8 The channel between G₀ and G₁ remains on at -2 V but the channel between G₀ and G₅ can be fully shut off. .......................................................... 34

4.1 The layer structure of a test sample. The layers above the horizontal dotted line were grown by RTCVD. To pattern the 2DEG, only one of implant isolation or conventional mesa isolation by RIE is used on a single sample, but both are illustrated in this figure for brevity. .... 38

4.2 (a) The schematic of a test device (not to scale). It consists of 1 Hall bar structure, 1 set of connected contacts and 1 set of isolated contacts. The cross section of the set of isolated alloy contacts indicated by the white dot line is shown in (b). ........................................ 39

4.3 Simulation of implanted species and resulting vacancy distribution in a 2DEG structure by the Stopping and Range of Ions in Matters (SRIM) software. Two-step implantation (1×10¹⁴ cm⁻² @ 30 keV+1×10¹⁴ cm⁻² @ 60 keV) with implanted species (a) Ar⁺ and (b) Si⁺ is used in this simulation. The target is assumed to be implanted at 0 K. .......... 40
4.4 Isochronal (1 hour) annealing behavior of sheet resistances of 2DEG samples at 4.2 K implanted with (a) Ar\(^+\) and (b) Si\(^+\) with three various doses, 5\(\times\)10\(^{11}\) cm\(^{-2}\), 1\(\times\)10\(^{13}\) cm\(^{-2}\) and 1\(\times\)10\(^{14}\) cm\(^{-2}\) at 30 keV and 60 keV.

Also shown are the range of sheet resistance of the starting 2DEGs, and the experimental instrumental limitation. The estimated error of sheet resistance is ±18% for (a) and 27% for (b).

4.5 Comparison of 2DEG quality after implant isolation or RIE process by the ratio of the mobility at 4.2 K measured from an implant-isolation-defined Hall bar to a RIE-defined Hall bar for a given implant and anneal. The absolute mobility values at 4.2 K with a unit of 10\(^3\) cm\(^2\)/Vs are shown in parentheses with the mobility for implant isolation before the mobility for mesa isolation. The estimated error of mobility measurement is ±5%.

5.1 The schematic of background charged impurities in an undoped silicon 2DEG.

5.2 The schematic of remote charged impurities at the oxide/silicon interface in an undoped silicon 2DEG.

5.3 The schematic of Si/SiGe interface roughness in an undoped silicon 2DEG.

5.4 The layer structure of a typical undoped enhancement-mode strained silicon 2DEG.

5.5 The top view of a standard Hall bar geometry used in this study. The magenta regions are Cr/Au metal stacks deposited in the same evaporation. The Cr/Au metal contacts are deposited directly on the silicon surface (etched holes of the aluminum oxide).
5.6 A typical linear $V_G$ dependence of 2D electron density in an undoped enhancement-mode silicon 2DEG. The inset shows the channel turn-on with increasing gate voltages. .............................................. 56

5.7 A schematic of two equivalent capacitors in series between metal gate and a conducting 2DEG in strained silicon channel. ................ 57

5.8 Typical Hall measurement data measured from 2DEGs grown in our RTCVD with theoretical fitting curves of three main possible mobility-limiting mechanisms. .................................................. 58

5.9 The schematic of potential fluctuation caused by remote charges at the silicon/oxide interface in an undoped 2DEG. ....................... 60

5.10 (a) A 2DEG conducts as a metal in the channel when the electron density is high (b) A 2DEG acts as an insulator because electrons are localized in the valleys of fluctuations when the electron density is too low. ................................................................. 61

5.11 The modified mobility model by a MIT model (green curve) fit the data of the sample with a 27-nm SiGe cap, especially in the low density regime. 62

5.12 A discrepancy between data points of the sample with a 55-nm SiGe cap and the modified model by the MIT model. The modified mobility model is the green line, while the original model is in black. ............ 63

5.13 The SIMS profile of a typical undoped 2DEG grown in our RTCVD before a new gas supply system was introduced. Clear high concentrations of both phosphorus and boron were observed. ............. 65

5.14 The schematic of the new gas supply panel and the gas supply assembly. Phosphine is separately transported from other process gases until a half meter away from the reactor entry to avoid mixture and contamination. ................................................................. 66
5.15 Extremely low concentrations of both phosphorus and boron in grown layers after a new gas panel and a gas supply assembly were installed. Both doping levels hit the detection limits of the high-precision SIMS analysis.

5.16 Mobility curves from samples with various SiGe cap thicknesses. Mobility is enhanced with a thicker SiGe cap, but saturate at 400,000 cm$^2$/Vs.

5.17 The highest mobility and lowest 2D electron density extracted from samples with various SiGe cap layers. The critical density gets lower in the sample with thicker SiGe cap thickness due to weaker potential fluctuations.

5.18 Mobility curves measured from samples with an aluminum oxide deposited at three different temperatures: 150 °C, 225 °C and 300 °C.

5.19 Mobility curves from samples annealed at different temperatures (450 °C, 600 °C and 750 °C) after ALO deposition.

5.20 The shifts in threshold voltages imply that charges are trapped at the interface after annealing.

5.21 Mobility curves from samples whose strained silicon layers were grown at different temperatures.

5.22 An AFM image and its calculated RMS roughness on a complete undoped silicon 2DEG.

5.23 An AFM image and its calculated RMS roughness from a reference sample whose layer structure only consists of the relaxed SiGe buffer and the strained silicon QW (channel).

5.24 Mobility curves labeled by individual growth temperature and its RMS surface roughness.
5.25 Extracted highest mobility and lowest 2D density versus surface roughness.

5.26 Images acquired by Normaski Microscopy after samples were dipped into modified Schimmel solutions. (a) Plenty of dot-like threading dislocations are distributed all over the sample whose $T_{\text{Si, Cap}}$ and $T_{\text{Si, QW}}$ were 575 °C and 700 °C respectively. (b) Much fewer etch pits are seen in the sample whose silicon layers were both grown at 625 °C.

5.27 The EPD measured from images by Normaski microscopy versus silicon growth temperature.

5.28 The highest mobility and lowest 2D density versus etch pit density from five different samples.

5.29 The geometries of (a) a standard Hall bar with a 300-µm spacing between $R_{xx}$ probes (denoted as a 300-µm Hall bar) and (b) a small Hall bar with a 10-µm spacing between $R_{xx}$ probes (denoted as a 10-µm Hall bar).

5.30 The comparison between mobility measured from a standard Hall bar and two small Hall bars fabricated in the sample from the same growth run.

6.1 The four-stage behavior in Hall electron density observed in all three samples (Data here are from the sample with a 14-nm SiGe cap layer). The dashed line shows the theoretical maximum $n_{2D}$ from the self-consistent Schrodinger-Poisson (SCSP) simulation.

6.2 (a) With a small gate bias, electrons accumulate in the buried QW first. (b) Even with a large gate bias, no electrons populate the surface QW due to a high critical density for metal-insulator transition (MIT) in that layer.
6.3 The comparison of buried electron density \( (n_{\text{buried}}, \text{red}) \) and surface electron density \( (n_{\text{surface}}, \text{blue}) \) with increasing gate voltages in both thermal equilibrium with the contacts between two 2DEGs and with the surface channel not in equilibrium. A sudden collapse in \( n_{\text{buried}} \) as the gate voltage increases and the corresponding increase in \( n_{\text{surface}} \) bring the system back to thermal equilibrium.

6.4 (a) no electrons are induced in the surface due to a high critical density of MIT for the surface QW. (b) At a higher gate voltage, electron tunneling from the buried QW towards the surface raises the density above the metal insulator transition point, leading to a current flowing from the contacts into the surface layer (c). (d) By Gauss’s law, the buried electron density must be reduced as the surface density increases at a fixed gate voltage. The system is then switched from non-equilibrium to thermal equilibrium.

6.5 Electrons exist at both the surface and the buried QW, with the same Fermi level in both layers. Now the system is in thermal equilibrium.

6.6 The dependence of Hall mobility on Hall electron density measured at 4.2 K for all three samples with different stages labeled. The gate voltage steps for data points at stage II and stage III/IV are 0.03-0.15 V / 0.1-2 V, respectively for all three samples. The measurement sequence is indicated by dashed lines (from stage II to stage III).

6.7 (a) The reduction in minimum \( n_{2D} \) due to the screening effect by the surface electron layer. (b) The increasing surface electron density screens the scattering from remote charges and thus enhances the buried electron mobility.
6.8 Surface electron mobility (calculated based on the parallel conduction model) versus the surface electron density (calculated in Sec. 6.3.2). The right Y axis shows the conductance ratio of the surface QW to the buried QW.

6.9 The four-stage behavior in density of the sample with a 20-nm SiGe cap measured by Hall measurement at 4.2 K (red). The sample was then dipped to 0.3 K for magneto-resistance transport measurement. Two representative gate biases used at 0.3 K were highlighted in blue.

6.10 The magneto-resistance transport measurements at 0.3 K were shown versus magnetic field up to 15 T at (a) 4 V (stage II) and (b) 9 V (stage III).

6.11 (a) $R_{xx}$ versus the inverse of magnetic field from the sample with a 20-nm SiGe cap at 9 V in the range of 0.5 T to 2.5 T. (b) The Fourier transform of the curve in (a) shows three peaks. The ratio of frequencies show the different degeneracies, but they all represent the same conducting channel and a single electron density of $2.92 \times 10^{11}$ cm$^{-2}$.  

6.12 An identical four-stage behavior in density was also observed in samples with thicker SiGe caps (90 nm in this case).

6.13 Mobility data of the 2DEG sample with a thick SiGe cap (90 nm). No significant screening effect was observed.
6.14 The modulation of electron density under two scanning directions of gate voltages. Electrons are first accumulated in the QW as the gate voltages ramp up at stage II (solid symbols, step 1) until a collapse occurs. The density then drops to a fixed value with further increasing biases (open symbols, step 2). Finally, the density remains constant as the gate voltages ramp down, followed by a linear decrease when the gate voltage crosses the boundary of two stages (open symbols, step 3). (This reference sample has a 75-nm SiGe cap layer and a 145-nm SiGe buffer).

6.15 The transport property of a typical undoped 2DEG with a thick SiGe cap under a regular voltage ramp-up at stage II (solid symbols) and a voltage ramp-down at stage III (open symbols).

7.1 A typical SIMS of an undoped 2DEG shows very high densities of oxygen and carbon at the regrowth interface. The integral 2D densities of oxygen and carbon are $6 \times 10^{13}$ cm$^{-2}$ and $4 \times 10^{14}$ cm$^{-2}$, respectively. A typical baking power for SiGe relaxed buffers is 20% for 5 minutes, which is roughly equal to 850 °C.

7.2 Band diagrams of an undoped 2DEG under a 0.2 V bias with the Fermi level pinned at four different positions: 0.045 eV below $E_C$ (green), 0.25 eV below $E_C$ (purple), mid-gap (brown) and 0.045 eV above $E_V$ (orange).

7.3 The theoretical maximum $n_{2D}$ in equilibrium (with different biases) by the simulation with different pinning conditions. The inset shows the layer structure used in the simulation. (The band gap of Si$_{0.72}$Ge$_{0.28}$ is set as 1.04 eV).
7.4 The experimental maximum $n_{2D}$ at stage III, from samples whose buffer layers are between 150 nm and 190 nm, versus corresponding SiGe cap thickness. Simulations with a fixed buffer thickness (190 nm) and various $E_F$ pinning conditions are shown in green ($E_C-E_F=0.045$ eV), purple ($E_C-E_F=0.25$ eV) and brown ($E_F$ at midgap) solid lines and an orange dash line ($E_F-E_V=0.045$ eV).

7.5 The experimental maximum of $n_{2D}$ at stage III from samples whose SiGe cap layers are between 20 nm and 180 nm versus corresponding SiGe buffer layer thickness. The simulations were done with a fixed buffer thickness (190 nm). The simulation curve with the pinning position ($E_C-E_F=0.045$ eV) in green fits the data best.

7.6 Transport properties of samples with various SiGe buffer layer thicknesses. Solid and open symbols represent data measured in stage II (scan up) and stage III (scan down), respectively.

7.7 Highest mobility and critical densities (the lowest $n_{2D}$) at both stage II (scan up) and stage III (scan down, as shown in Fig. 6.14) extracted from samples with various thicknesses of SiGe buffer layers. The blue solid and blue open symbols are the lowest $n_{2D}$ at stage II and stage III, respectively.

7.8 Mobility curves from samples with various baking powers. Solid and open symbols represent data measured in stage II (scan up) and stage III (scan down), respectively.

7.9 Highest mobility and critical densities at both stage II (scan up) and stage III (scan down) extracted from samples with various baking powers. The lowest $n_{2D}$ at stage II and stage III are blue solid and blue open symbols, respectively.

B.1 A typical SIMS measured on dirty epi-layers grown on a silicon substrate.
B.2 (a) The top view and (b) the cross section along the red line of a standard 100-mm carrier wafer.

B.3 A typical SIMS of an undoped 2DEG with group record high mobility.

B.4 A mask image that describes the four photolithography steps for enhancement-mode undoped 2DEG fabrication.

C.1 The picture of a 125-mm carrier held by a 125-mm quartz wafer stand with a 75-mm wafer on it.

C.2 The picture of 75-mm SiGe buffer carried by a 125-mm carrier wafer after a test growth. The color rings labeled by yellow lines represent the thickness nonuniformity.

C.3 The schematic of a 75-mm SiGe buffer and a silicon temperature control piece held by a 125-mm carrier wafer. The pieces sent out for SIMSs are highlighted.

C.4 The thickness comparison between pieces cut from different locations on the 75-mm SiGe buffer held by a 125-mm carrier wafer.

C.5 The schematic of a 75-mm wafer on a 100-mm carrier wafer. The 75-mm is used to calibrate growth temperatures.

C.6 The normalized transmission of two infrared lasers calculated based on Eq. C.1 for different wafer thicknesses.

C.7 The SIMS analyses for two separate pieces cut from a 75-mm SiGe buffer carried by a 100-mm carrier wafer. The peaks in the silicon channel could be artifacts.

C.8 The thickness non-uniformity of epitaxial layers grown on a 75-mm SiGe buffer carried by a 100-mm carrier wafer.

C.9 The mobility data measured at 4.2 K from center pieces and pieces near the edge. The center piece shows a group record high mobility.
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.10</td>
<td>The thickness percentages of epi-layers grown on a 75-mm silicon substrate carried by a 100-mm carrier wafer.</td>
</tr>
<tr>
<td>C.11</td>
<td>The picture of a diced 100-mm SiGe buffer from a 150-mm SiGe wafer. The ring along the edge of the diced wafer is the etched pattern from the quartz ring of Samco 800 etcher, showing the quartz ring perfectly covered the edge of the diced wafer during the etching.</td>
</tr>
<tr>
<td>C.12</td>
<td>The pieces cut for SIMS analyses from a diced 100-mm SiGe buffer after a standard undoped 2DEG growth.</td>
</tr>
<tr>
<td>C.13</td>
<td>The SIMS analyses for the pieces with different distances from the wafer center of a diced 100-mm SiGe buffer.</td>
</tr>
<tr>
<td>C.14</td>
<td>The thickness non-uniformity of epi-layers grown on a diced 100-mm SiGe buffer.</td>
</tr>
<tr>
<td>C.15</td>
<td>The mobility measured from pieces with different distances from the wafer center of a diced 100-mm SiGe buffer. The highest mobility at 4.2 K is 200,000 cm²/Vs.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation

Since the concept of a transistor was first demonstrated in 1947 by John Bardeen, Walter Brattain and William Shockley [1], and the first commercial silicon-based transistor was presented in 1954 by Texas Instruments [2], silicon has dominated our life for over a half century. The electronic products made from silicon are ubiquitous, from advanced medical instruments to consumer electronics, and from private data storage centers to public communication systems. More and more advanced applications such as smart phones and the internet of things (IOT) that either have or will have huge impacts on our current life style are all enabled by a small silicon chip that consists of billions of silicon transistors. Over the past four decades, semiconductor industries have followed the famous Moore’s law which describes that the number of transistors in an integrated circuit (IC) chip doubles approximately every two years. The denser transistors in an IC chip leads to higher computation performance and lower power consumption. Today, an Intel 15-core Xeon IvyBridge-EX central processing unit (CPU) possesses over 4.3 billion transistors [3]. The scaling of the size of a single transistor has been the rule of thumb in silicon industry to allow more tran-
sistors to be packed into a small IC chip. For example, the semiconductor technology node has progressed from the scale of 100 \( \mu \text{m} \) in the early 1970s to today's 14 nm, which translates to a near 10,000 times reduction in a transistor dimension and \( 10^8 \) in area.

Unfortunately, a notorious short channel effect plagues scaled transistors. The gate of a transistor loses its effective control of the channel switching as a transistor shrinks in size. A novel device, the fin field-effect transistor (FinFET), with metal gates covering three sides of a protruding channel currently predominates most state-of-the-art semiconductor microprocessors because of its superior gate control capability \[4\]. Nevertheless, the soaring capital investment in both research and development (R&D) along with exponentially increased technical difficulties for next-generation technology such as the 10-nm node or the 7-nm node may foresee the end of scaling of the conventional silicon transistor.

Many alternatives have been proposed in past decades as a potential candidate for next-generation computation. Among them, the concept of quantum computing drew many scientists' attention when it was first proposed in the early 1980s. Peter Shor at AT&T Bell Laboratories then developed the first quantum algorithm in 1994 and set a guideline for future research towards the realization of quantum computation \[5\]. Quantum computing itself provides us a promising way to implement fundamental computation at a near atomic level. More importantly, it exploits the advantages of quantum superposition and quantum entanglement between two basic states, enabling a novel concept of quantum bit (Qubit) that can store much more information than a conventional bit, like a metal-oxide-semiconductor field-effect-transistor (MOSFET). Possible methods to implement quantum computing include superconducting devices \[6\], linear optics \[7\], and charge-based semiconductor quantum dots (QD) \[8\]. In the late 1990s, Daniel Loss and David DiVincenzo proposed to utilize the intrinsic spin of an electron in a semiconductor quantum dot as a basic
unit for novel logic computation. The abundant knowledge about electron spin in semiconductors and high compatibility with the facilities in today’s silicon industries greatly increase feasibility of spin-based quantum computation. This novel spin-based quantum bit in a semiconductor motivated all the work done in this thesis.

There is still a long way for quantum computing technology to become mature. The low operation temperature could also limit its future applications for commercial products. Notwithstanding these issues, the theoretically-predicted computation speed that quantum computing may have to offer will keep exciting researchers throughout the world. We can expect that the uncountable efforts scientists put in the quantum computing field will not stop until our lives are deeply and unconsciously influenced by it someday in the future.

1.2 Thesis Outline

This thesis starts with a brief introduction to the two-dimensional electron gas (2DEG) realized in Si/SiGe heterostructures in Ch. 2. The benefits of the Si-based material system for single-electron quantum dot devices as a spin-based quantum bit are elaborated. Band engineering by means of strain and relaxation in Si/SiGe layers is introduced as well as the growth of such Si/SiGe heterostructures in our RTCVD system. The advantages and disadvantages of modulation-doped and enhancement-mode undoped strained silicon 2DEGs in terms of compatibility for quantum dot applications are respectively proposed. The focus is then shifted to improving drawbacks of both 2DEGs, making a silicon 2DEG a better platform for quantum computing applications.

Phosphorus surface segregation from the intentional doping layer in a modulation-doped 2DEG causes an undesirably high phosphorus concentration in cap layers. Its resultant detrimental gate leakage current and low breakdown voltage always leave a
question mark on the possibility of a modulation-doped 2DEG as a candidate for a single-electron quantum dot device. In Ch. 3 we identify the phosphorus segregation mechanism in relaxed SiGe layers during epitaxial growth. A lower growth temperature is experimentally proven to effectively preserve the hydrogen coverage on a surface layer to be grown and thus block the surface segregation paths of phosphorus atoms in a sub-surface layer. A two order of magnitude reduction in surface phosphorus concentration enables a successful depletion test of Schottky splitting gates on top of a modulation-doped 2DEG without any significant gate leakage. A quantum point contact device as a charge sensor is then fabricated to choose the best distance between a sensor and a dot, which paves the road towards a quantum dot device.

The conventional etching isolation on a modulation-doped 2DEG has been proved as an effective way to isolate a mesa with size of around 30 $\mu m^2$, on to which small Schottky splitting gates could be deposited. However, the corner-induced gate leakage and conformability issue of either thin oxide layers or metal gates to cover step edges limit its usefulness. Reliable lateral electrical isolation which preserves surface planarization can be achieved by ion implantation (Ch. 4). The low operation temperature (4.2 K or below) and the low thermal budget of a quantum dot device fabrication make implant isolation feasible in silicon-based materials. Appropriate implant recipes are chosen for an effective isolation that is resistant to post-implant annealing up to 550 °C in both the doping layer and strained silicon channel. The 2DEG qualities in terms of electron mobility of implanted samples are assured to remain as good as those of intact samples.

In Ch. 5 we discuss the importance of electron mobility and critical electron density in an enhancement-mode undoped 2DEG. Both are viewed as essential parameters that show the cleanliness of a 2DEG system and its adequacy for a delicate quantum dot device. Efforts toward a high electron mobility or low critical density have been initiated with a careful identification of possible scattering mechanisms
in our 2DEG system. Efforts to alleviate these mobility-limiting factors, such as a
decisive modification of the gas panel of our CVD system to reduce background phos-
phorus concentration, have led to mobility enhancement and critical density reduc-
tion. The significances of growth temperature and the layer thickness of a Si/SiGe
heterostructure in the transport properties are specifically emphasized here with a
detailed systematic investigation.

An inevitable compromise between sharp electron confinement from top deple-
tion gates versus high mobility and low density comes from the SiGe cap thickness.
The transport properties of a 2DEG could get severely deteriorated if the distance
between the electron channel and oxide/silicon interface gets shortened in order to
sharpen electrical potentials from gates for 2DEG patterning. In Ch. 6 a screening
effect on remote scattering sites at the interface enabled by a newly-formed surface
electron layer is unveiled in a series of enhancement-mode undoped strained silicon
2DEGs. The introduction of a surface electron layer with ultra low mobility is initi-
ated by a tunneling of electrons from the channel towards the interface, followed by
a positive feedback process that pulls the 2DEG system from non-equilibrium back
to thermal equilibrium. The surface electrons act as a shield to screen out the detri-
mental interface charges, leading to a great improvement in transport properties of
enhancement-mode undoped 2DEGs especially with thin SiGe caps.

In Ch. 7 the importance of the regrowth interface (bottom interface) of undoped
silicon 2DEGs is discussed. High carbon and oxygen concentrations at this interface
could lead to the pinning of the Fermi level. The most likely pinning position of the
Fermi level is identified to be near the conduction band minimum, by means of the
comparison of simulation results and experimental data. With varying the SiGe buffer
thickness (vary the distance between the regrowth interface and the silicon channel)
and the baking power for the regrowth surface (vary the quality of the regrowth
interface), the effect of the regrowth interface on transport properties of undoped 2DEGs is examined as well.

Finally, several future work relevant to this work are discussed in Ch. 8 after a brief conclusion for this thesis.
Chapter 2

Two Dimensional Electron Gases
in Si/SiGe Heterostructures

2.1 Introduction

2.1.1 Introduction to a 2DEG

A two-dimensional electron gas (2DEG) in a semiconductor material system, as its name suggests, describes an ultra-thin electron layer existing in a semiconductor with energy confinement in one dimension. Because of its unique low-dimensional nature, a 2DEG system enables many avenues for fundamental research in semiconductor physics. The most common 2DEG in semiconductor technology is the surface electron layer in an n-type metal-oxide-semiconductor field effect transistor (MOSFET) under the strong inversion condition. A high positive voltage expels most of majority carriers (holes) away from the oxide/silicon interface and bends the conduction band ($E_c$) down to populate electrons to form a 2DEG at the surface. On the other hand, a 2DEG located in a quantum well (QW) that could be realized in semiconductor heterostructures, such as AlGaAs and GaAs or strained Si/SiGe, is verified to have much better transport properties because such a 2DEG is separated away from the
oxide/silicon interface or GaAs surfaces which are full of charged sites. Many different types of transistors have been explored in the past decades based on high-mobility 2DEGs in semiconductor heterostructures. For instance, a modulation-doped field effect transistor (MODFET) was fabricated on Si/SiGe heterostructures [10], while high electron mobility transistor (HEMT) was realized in III-V materials [11].

A decade ago, the superior transport properties of 2DEGs in semiconductor heterostructures drew physicists' attentions as a potential platform to implement the quantum bit concept. Researchers in the quantum computing field intend to confine a single electron in a semiconductor quantum dot. The spin direction of such a single electron could represent information, and due to the superposition and entanglement of the electron spins, this spin-based quantum bit (or qubit) can store much more information than a conventional bit. Since a 2DEG in semiconductor material systems has already confined electrons in the out-of-plane direction, methods to isolate a 2DEG laterally, such as etching isolation or Schottky top gates, enable us to fabricate a quantum dot and possibly confine a single electron in it for the following spin manipulation.

The research on a spin-based qubit in a semiconductor quantum dot was first initiated in III-V 2DEGs. The higher mobility due to its low electron mass and the defect-free films due to lattice-matched heterostructures in GaAs/AlGaAs are attractive for isolating electrons with fewer scattering sources. However, the inherently strong interaction between electrons and nuclei, also known as hyperfine coupling, in this material system is detrimental to the spin coherence time (or dephasing time, $T_2^*$) [12, 13]. For example, a very short dephasing time $T_2^*$ ~ 7 ns in III-V materials was observed [14]. Recently, a 2DEG in the Si/SiGe heterostructure became the most popular material system to fabricate a spin-based quantum dot, as the naturally abundant isotopic $^{28}$Si has zero hyperfine coupling. A much longer dephasing time (for instance, ~360 ns in [15]) is thus allowed in the Si-based material system,
facilitating the spin manipulation process. Therefore, in this thesis, we mainly focus on the silicon-based 2DEG system for quantum computing applications. Research on two different types of silicon 2DEGs (modulation-doped 2DEGs and enhancement-mode 2DEGs) are both discussed. In contrast to the spin physics, we emphasize the epitaxial growth and characterization of Si/SiGe heterostructures. Also, we investigate the 2DEG transport properties under various process treatments, and analyze how transport properties can be improved and optimized to fit the requirement for spin-based qubit applications.

2.1.2 Introduction to the RTCVD at Princeton University

All Si/SiGe heterostructures used in this thesis were grown using the rapid thermal chemical vapor deposition (RTCVD) system at Princeton University. As its name suggests, this CVD system possesses the capability to vary temperature between 500 °C to 1000 °C within seconds. The growth reactor RTCVD is a 150-mm quartz tube surrounded by a reflection assembly, and heated by 12 parallel 6-kW tungsten lamps underneath it (Fig. 2.1). A 100-mm wafer is carried by a quartz wafer holder and transferred into the reactor through a load-lock chamber, which is separated from the main growth chamber by a gate valve to avoid contamination from air during the loading process. Available process gases carried by ultra pure hydrogen (99.9999%) in our system include silicon precursors: silane (SiH₄), disilane (Si₂H₈), dichlorosilane (DCS); germanium precursor germane (GeH₄); n-type and p-type dopant precursors: phosphine (PH₃) and diborane (B₂H₆). Their flow rates are controlled by metal-sealed mass flow controllers (MFCs). Each gas is individually guided by an injection valve to either the reactor for epitaxy or to a waste line where a burn box operated at 850 °C burns out excess gases.

The temperature of the wafer in the reactor is controlled by a feedback system. Two infrared lasers with 1.3-µm and 1.55-µm separately modulated by function gen-
Figure 2.1: The schematic of rapid thermal chemical vapor deposition (RTCVD) at Princeton University. (Image courtesy of P. V. Schwartz [16])

Lasers, emit from a fiber on the top of the reactor towards the wafer in it. The transmission through the wafer of the 1.3-µm and 1.55-µm lasers, measured by a photodetector and a lock-in amplifier, are very sensitive to a specific range of temperature: 500 °C-625 °C and 675 °C-750 °C, respectively. The ratio of transmission at a target temperature to that at room temperature is called normalized transmission, and depends on the absorption coefficient and the wafer thickness. The absorption of these two infrared lasers in silicon consists of two different processes: band to band absorption and free carrier absorption [17, 18, 19]. For a lightly-doped silicon substrate, it was found that the dominant mechanism for 1.3-µm photons is band to band absorption, which gets stronger with an increasing temperature due to bandgap narrowing. In contrast, 1.55-µm photons are mainly absorbed by free carrier absorption, which is enhanced by a higher temperature due to the increasing free carrier...
concentrations. The normalized transmission data published in [17] are shown in Fig. 2.2.

A control PC calculates an instantaneous normalized transmission and feeds it back to the lamp power, to keep this normalized transmission equal to that corresponding to the target temperature. The controllable temperature range (500 °C-750 °C) is right the range we commonly use for Si or SiGe epitaxy. Beyond this range, a constant lamp power is used to replace the precise temperature reading for higher temperature baking (>750 °C), and an estimated temperature for any baking power could be roughly extracted based on those regular growth temperatures. The data
Figure 2.3: The data of average lamp power versus temperature calibrated by normalized transmission of two infrared lasers with the extrapolation for temperatures >750 °C. In addition, the lamp powers dropped 6% after a new SCR unit was installed. The common reactor baking power and SiGe buffer baking power used before and after the SCR replacement are both labeled.

of average experimental lamp powers versus corresponding temperatures calibrated by two infrared lasers are shown in Fig. 2.3 with the extrapolations for temperatures higher than 750 °C. Note that the replacement of the silicon control rectifier (SCR) power controller of RTCVD on December 2013 resulted in a 6% drop in lamp powers for all measurable temperatures.
2.2 Characteristics of Silicon 2DEGs

2.2.1 Strain-Induced Band Offset in Si/SiGe Heterostructures

To prevent scattering from the oxide/silicon interfaces or bare sample surfaces, a 2DEG buried underneath the sample surface is preferable for the manipulation of electron spin. A quantum well for electrons in a Si/SiGe heterostructure is attractive to hold a 2DEG because we can easily bury this 2D electron layer as deep as we want to alleviate the influence from the sample surfaces. It has been well-known that the strains in either silicon or SiGe layer shift the energy position of the conduction band minimum or the valence band maximum, and lead to the splitting of valley degeneracy [20]. For example, a biaxial compressively strained SiGe layer between relaxed Si layers gives us a band offset in the valence band, resulting in a quantum well in strained SiGe layer for holes, while the band offset at the conduction band minimum is so small that no electrons are confined in the SiGe layer. In this type of Si/SiGe heterostructure, holes created by absorption of a laser accumulate in the quantum well and recombine with electrons moving in the conduction band. The ratio of non-radiative and radiative recombination (photoluminescence, PL) strongly depends on the impurity concentration (carbon or oxygen) and defects in epi-layers. Thus the PL measurement in such Si/SiGe heterostructures has been used as a qualitative measure for the quality of epitaxial films [21].

A conduction band ($E_c$) offset can be achieved when a silicon layer is biaxial tensile strained on a relaxed SiGe layer (Fig. 2.4a) [22]. This biaxial tensile strain shifts the average energy level of the conduction band minima in silicon, and also splits its 6 fold degeneracy into two groups: Four in-plane valleys ($\Delta_4$) are lifted up, while two out-of-plane valleys ($\Delta_2$) are lowered. A quantum well for electrons thus forms in a strained silicon layer that is sandwiched between two relaxed SiGe layers despite Si
normally having a larger bandgap than SiGe. The Ge fraction in relaxed SiGe layer determines the strength of the strain in the silicon layer, and thus affects the energy separation between $\Delta_4$ and $\Delta_2$. Therefore, the conduction band offset between a relaxed SiGe layer (regular six-fold degenerate $E_c$) and a tensile strained silicon layer (lowered $\Delta_2$) could be estimated based on the equation: \[ \Delta E_c = E_g(\text{sSi}) + \Delta E_v - E_g(\text{SiGe}) = -0.35y - 0.35y^2 + 0.12y^3 \] (2.1)

where $E_g$ is the band gap of the two materials, while $y$ is Ge fraction of the relaxed SiGe layers. Since electrons tend to accumulate in $\Delta_2$ because of its lower energy level, the effective in-plane electron mass is thus reduced from 0.26 $m_0$ down to 0.19 $m_0$, equal to the transverse effective mass of a relaxed silicon layer with 6-fold $E_c$ degeneracy (Fig. 2.4b). This lower electron mass together with the mitigated inter-valley scattering due to the valley splitting both enhance the mobility for electrons in the strained silicon QW.
2.2.2 SiGe Virtual Substrate

It is always a challenging issue to have a high quality relaxed SiGe substrate to introduce tensile strain in pseudomorphic silicon films. A SiGe layer could be relaxed if it is intentionally grown on top of a silicon substrate with considerable thickness due to severe lattice mismatch. However, a high density of threading dislocations may nucleate and distribute all over the SiGe film to relieve the strain caused by lattice mismatch, greatly degrading SiGe film quality. To improve the quality of a relaxed SiGe layer, the concept of graded SiGe buffer was introduced by Currie et al. in 1998 [23]. On top of a regular silicon substrate, a fairly thick (few microns) SiGe layer is epitaxially grown with a gradual increase in Ge fraction. A typical ramping rate for Ge fraction is 10%/µm. The graded SiGe buffer introduces the lattice mismatch in a slow manner, thus preventing detrimental accumulation of threading dislocation defects. In addition, these initial threading dislocations at the Si/SiGe interface help relieve the strain from 0% Ge to a target fraction. Further nucleation for more threading dislocation defects to relieve strain thus becomes unnecessary. Subsequently, another thick SiGe layer (on the order of a micron) with constant Ge fraction (which is the target Ge fraction) is grown to cap the graded SiGe buffer layer. This buffer layer with constant Ge fraction effectively buries existing threading dislocations in the graded layer, and also blocks their possible propagation towards the surface, leading to an extremely low threading dislocation density at the final SiGe surface.

This kind of high quality SiGe substrate consisting of a bottom silicon substrate, a graded SiGe buffer and a relaxed SiGe buffer with a constant Ge fraction is the so-called SiGe virtual substrate. In this study, all Si/SiGe heterostructures were grown on commercial SiGe virtual substrates from Amberwave Inc. In more detail, the virtual substrate used in this study comprises a p-type (100) silicon substrate, a -3-µm SiGe graded buffer up to 28% Ge, and a -1-µm relaxed SiGe buffer with 28%
Ge. To remove the inevitable cross-hatch patterns, a signature shown on the surface of a typical relaxed SiGe buffer caused by the accumulation of dislocation defects, a chemical mechanical planarization (CMP) service was done by Axus Technology. A layer thinner than 500 nm was then removed, and the surface RMS roughness was improved from 7.4 nm down to 0.2 nm.

2.2.3 Types of Strained Silicon 2DEGs

Thanks to the quantum well for electrons in strained Si/SiGe heterostructures, electrons can be confined as a two-dimensional electron gas (2DEG). More importantly, this 2DEG is buried away from the surface of the heterostructure by a relaxed SiGe cap layer with an adjustable thickness. Based on the way that electrons are supplied into the quantum well, two types of silicon 2DEGs are discussed: modulation-doped 2DEGs (electrons are supplied from a n$^+$ doping layer) and enhancement-mode 2DEGs (electrons are induced by a gate). Details for both types of 2DEGs are presented in Sec. 2.3 and Sec. 2.4 respectively.

2.3 Modulation-Doped Strained Silicon 2DEG

2.3.1 Introduction

The modulation-doped 2DEG was first realized in III-V material systems [24]. To exploit high mobility in GaAs system [25], a modulation-doped GaAs/AlGaAs 2DEG is gated to form a high-electron-mobility transistor (HEMT). The electrons in the channel are supplied by an intentionally doped layer few nanometers away from the GaAs channel. By tuning the distance between the channel and the doping layer (or as we call it, a spacer), an electron density in this modulation-doped structure is determined. The top gate then acts as a switch to either deplete electrons to shut off the channel or leave electrons there to keep the channel on.
Since the discovery of the conduction band offset in strained Si/SiGe material system, a layer of phosphorus-doped SiGe was introduced in the relaxed SiGe cap to mimic the modulation-doped 2DEG structure in III-V system. In the early 1990s, scientists invested much effort to investigate modulation-doped strained silicon 2DEG [26, 27, 28]. Exploratory research was focused on n-type modulation-doped strained silicon 2DEGs because of its potential to make a modulation-doped field effect transistor (MODFET), which has higher electron mobility than a regular bulk MOSFET. In addition, higher electron density could be obtained by optimizing the doping concentration and the spacer thickness, providing another knob that can tune up the drive current in a transistor. As a result, intensive research in the mid-1990s enhanced the low-temperature (<4 K) electron mobility in modulation-doped strained silicon 2DEG from 19,000 cm²/Vs up to 500,000 cm²/Vs [29, 30]. Even though the MODFET lost its significance eventually in the late 1990s because of the costly SiGe substrate and the inevitable defects in relaxed SiGe layers, that research has established a solid knowledge foundation for its new application as a platform to implement quantum computing and spin manipulation in the early 2000s.

In this thesis, work related to modulation-doped strained silicon 2DEG is elaborated in Ch. 3 and Ch. 4. Details about its layer structure and the concept of a single-electron quantum dot device will be introduced in the next two sub-sections.

### 2.3.2 Layer Structure and Epitaxial Growth of a Doped 2DEG

A typical layer structure of a modulation-doped strained silicon 2DEG is illustrated in Fig. 2.5a. All 2DEG samples used in this thesis were grown by our RTCVD on top of commercial SiGe virtual substrates with Ge fraction - 28%. Layer structures of 2DEGs used in this thesis may vary for different purposes, but a standard layer structure and growth recipe will be addressed here for reference. Carried by a 3 lpm
ultra-clean hydrogen flow, silane (10% in argon) and germane (0.8% in hydrogen) were injected together into the growth reactor for SiGe epitaxial growth or silane was injected alone for Si epitaxial growth. The flow rates of precursors could be tuned to grow SiGe layers with various Ge fractions. Standard Si and SiGe growth temperatures were 625 °C and 575 °C, respectively, at a fixed pressure - 6 torr. Doped layers were realized by injecting phosphine (100 ppm in hydrogen) along with silane and germane.

A relaxed Si$_{0.72}$Ge$_{0.28}$ buffer layer (~150 nm) was first grown on top of a virtual substrate. The purpose of this buffer layer is to separate the next strained silicon layer as an active electron channel away from the regrowth interface, which may have detrimental defects. A 10-12 nm strained silicon layer was then grown, with a thickness thinner than the critical thickness to keep this layer fully strained [31]. Next, a thin SiGe spacer with various thicknesses (10-30 nm) was grown to separate the channel from the following doping layer. A 10-nm phosphorus-doped SiGe layer was introduced with high phosphorus concentration (>10$^{18}$ cm$^{-3}$) as an electron supply layer. Subsequently, a 20-40 nm relaxed SiGe cap layer was grown without intentional
doping, in which the phosphorus concentration is able to ramp down to avoid high phosphorus concentration at the surface that may potentially cause severe leakage in Schottky gates. Finally, a thin strained silicon layer (~3-4 nm) was grown to cap the SiGe surface.

The band diagram of a modulation-doped 2DEG is shown in Fig. 2.5b. The theoretical electron density allowed in a given modulation-doped 2DEG structure could be calculated based on electrostatics [32]. In practice, an electron density is tuned experimentally by the SiGe spacer thickness and phosphorus doping concentration. For example, a lower electron density could be achieved with a thicker SiGe spacer and a lighter phosphorus doping.

2.3.3 A Quantum Device on a Doped 2DEG

Modulation-doped 2DEGs were the first candidate for QD fabrication in Si/SiGe material system. Several approaches to make a quantum dot device on modulation-doped 2DEG have been demonstrated in the past two decades. Klein et. al. isolated a quantum dot device by etching other conducting paths away [33]. The etching depth must be deeper than the depth where the strained silicon channel is buried for an effective lateral isolation. However, the drawback from the etching is obvious: The defects on the etched sidewalls may deplete the electrons in the channel, especially in such fine features on the order of tens of nanometers, and block the electrical conduction. In addition, an extremely well-controlled anisotropic etching process is necessary for etching-defined QD devices. Any significant undercuts could lead to the severe damage of such devices.

Another alternative, the Schottky split-gate technique, has been widely adopted by several groups to make QD devices instead [34, 35]. Different from etching isolation, the splitting Schottky gates electrically isolates a 2DEG. A series of delicate metal gates are directly deposited on a modulation-doped 2DEG. With proper nega-
Figure 2.6: (a) A layer structure of a modulation-doped 2DEG with Schottky splitting gates. A 2DEG can be patterned by the Schottky gates with negative biases. (b) A three dimensional view for a typical quantum dot device fabricated on a modulation-doped 2DEG. (This is the model of a real device in [36, 37])

Due to the simplicity of this process, more functional gates are allowed as either tunneling barrier controls or charge sensors (Fig. 2.6b). These Schottky gates on n-type Si/SiGe heterostructures are usually made with high workfunction metals, such as palladium (Pd) or platinum (Pt) to maximize the Schottky barrier and reduce gate leakage. However, in practice, Schottky gated QD devices may suffer from a high gate leakage. This gate leakage may result from the high phosphorus concentration at the surface due to phosphorus surface segregation or other non-idealities. A high surface electrical field caused by a high phosphorus level leads to tunneling through the Schottky barrier from the metal gate to the semiconductor. The work about suppression of phosphorus surface segregation will be discussed in Ch. 3 and a large improvement in breakdown voltage of such Schottky gating on a modulation-doped 2DEG will also be demonstrated.
2.4 Enhancement-Mode Undoped Strained Silicon 2DEG

2.4.1 Introduction

Although most researchers’ efforts have been put on modulation-doped 2DEGs in early development of single-electron quantum devices, people started to realize the potential drawbacks of modulation-doped 2DEGs. First, the gate leakage problem degrades the depletion capability and stability of Schottky split gates. Second, it is very challenging to tune the electron density at a relatively low level (~10^{11} \text{cm}^{-2}), which is desirable to deplete a 2DEG into a single electron. Finally, it was found that ionized phosphorus in the doping layer can cause serious potential fluctuations to the electrons in the strained silicon channel. The interference from ionized charges could degrade electron mobility and its spin coherence time.

An enhancement-mode 2DEG was first demonstrated in III-V material system [38]. Then, it was copied to the strained Si/SiGe material system very soon [39]. Since the Si/SiGe heterostructure here is nominally undoped, a gate stack is deposited on top of it to capacitively induce electrons into the strained silicon channel, making it resemble a MOSFET. The benefit of an enhancement-mode undoped 2DEG is clear: the absence of ionized charges could transform this 2DEG into a cleaner system for spin manipulation. In addition, the capability to tune electron density at a low level by simple bias adjustment provides another freedom for layer structure and QD device design.
2.4.2 Layer Structure and Epitaxial Growth of an Undoped 2DEG

Fig. 2.7a shows a layer structure of an enhancement-mode undoped strained silicon 2DEG, which is very similar to that of modulation-doped 2DEGs except for the absence of the intentional doping layer. In our experiments, four Si/SiGe layers were all grown on a commercial SiGe virtual substrate. A 150-nm SiGe layer was first grown, followed by a 10-nm strained silicon channel. A relaxed SiGe cap layer was then grown on the channel with various thicknesses. The varied SiGe thicknesses represent different distances between the electrons in the strained silicon channel and ionized charged impurities at the oxide/silicon interface, which allows us to observe the effect from interface charges on 2DEG quality, especially the electron mobility and 2D density. Last, the whole heterostructure was capped by a thin strained silicon cap with 3-4 nm thickness. The gas precursors, growth temperatures, growth flow rates and growth pressures for standard enhancement-mode undoped strained silicon 2DEGs are all the same as those of modulation-doped 2DEG described in Sec. 2.3.2.

A 90-nm aluminum oxide was then deposited on an as-grown undoped 2DEG by atomic layer deposition (ALD) as an insulator. A Cr/Au stack was finally thermally evaporated onto the insulator as a gate for electron induction.

With a sufficient positive voltage on the metal gate, the quantum well in the strained silicon channel is lowered below the Fermi level ($E_F$). Electrons from the lateral ohmic contacts flow into the strained silicon QW and form a 2DEG (Fig. 2.7b). The SiGe cap layer separates the 2DEG away from the scattering of interface charges, potentially enabling high electron mobility. Here we note that, ideally, the thickness of strained silicon surface cap should always be kept very thin to avoid its ground state from touching the Fermi level earlier than that of strained silicon channel, because any 2DEG induced in the surface QW has a poor quality due to its proximity to interface charges, similar to the 2DEG in a bulk silicon MOSFET.
2.4.3 A Quantum Device on an Undoped 2DEG

A similar Schottky split-gate technique was utilized to deplete electrons and pattern a 2DEG into a QD device on an enhancement-mode undoped 2DEG. The main difference comes from the process complexity. Since an enhancement-mode 2DEG has an insulator and a metal gate on top of the Si/SiGe heterostructures, these depletion gates are inserted between the oxide and the Si cap layer to enable 2DEG effective patterning (Fig. 2.8a and 2.8b). (We note here that recently a very thin aluminum oxide ~3 nm is deposited underneath the depletion gates to prevent possible leakages [15]). From the viewpoint of process flow, the fine splitting gates must be deposited before oxide and metal (here we call it as a universal gate) deposition, which could lead to a risk to contaminate the oxide deposition chamber. In addition, the oxide thickness must be thick enough to prevent any leakage between the universal gate and the depletion gates. Its thickness is essential to avoid leakages from contacts (not shown here) to the universal gate as well.
Figure 2.8: (a) The layer structure of an enhancement-mode undoped 2DEG with depletion gates inserted between an oxide and a surface silicon cap. Electrons in the strained silicon channel can be patterned by these depletion gates. (b) A three dimensional view of a quantum dot device fabricated on an enhancement-mode 2DEG [15, 36, 37].

The absence of any intentional doping atoms leads to extremely low Schottky gate leakage for the splitting gates. However, its depletion capability is now limited by the SiGe cap thickness, which is also the distance between a 2DEG and charged impurities at the interface. A thinner SiGe thickness is preferred because the electrical potential from the depletion gates to the strained silicon channel is sharper. A precise electron confinement is thus possible in the nanometer scale. However, Coulomb forces between charged impurities and electrons in the channel become stronger with a decreasing SiGe cap thickness. Electron mobility could be significantly degraded due to the scattering from those charges. The compromise between these two issues comprises the backbones of Ch. 5, Ch. 6, and Ch. 7. More details about enhancement-mode undoped strained silicon 2DEG will be discussed in those three chapters.
Chapter 3

High Breakdown Voltage Schottky Gating of Doped Silicon 2DEGs

3.1 Introduction

A Schottky split-gate technique on a Si/SiGe modulation-doped two-dimensional electron gas (2DEG) with negative biases has become a common way to define lateral quantum dot arrays as we described in Ch. 2. A negative bias on the gate should fully deplete the 2DEG to pattern it into quantum dots. However, the well-known surface segregation of phosphorus from the doped electron supply layer into the following SiGe/Si capping layers during growth can cause a high electric field at the sample surface when a reverse bias is applied. This high field leads to the tunneling of electrons from the metal to the semiconductor, resulting in undesirable gate leakage, which degrades the reliability of the split gate technique.

In this section, we demonstrate a large reduction in gate leakage by suppression of phosphorus surface segregation during sample growth. In collaboration with Jiun-Yun Li, we successfully reduced the surface phosphorus level down to $2 \times 10^{16}$ cm$^{-3}$ by lowering the growth temperature of the SiGe cap layer. A much higher breakdown
voltage (-7 V) was achieved in the sample with such a low surface phosphorus level, compared to -2 V from the sample with a high surface phosphorus concentration (1×10^{18} \text{ cm}^{-3}). This great improvement in breakdown voltage thus provides a wide window to deplete a 2DEG fully, or shut off the channel, without significant gate leakage.

### 3.2 Suppression of Phosphorus Surface Segregation

The surface segregation of phosphorus is not a new issue in silicon epitaxial growth. It was found that, during growth, phosphorus atoms preferentially jump from the subsurface layer to surface sites because surface sites have lower energy [41]. In the growth of a modulation-doped silicon 2DEG, phosphorus atoms from the intentionally doped layer keep switching crystalline sites with surface host atoms (Si or Ge), leading to a high concentration at the sample surface (Fig. 3.1). For example, growing cap layers at 575 °C in RTCVD results in extensive phosphorus surface segregation, giving a phosphorus surface concentration after a nominally undoped SiGe cap layer thickness of 20 nm about 1×10^{18} \text{ cm}^{-3} (Fig. 3.2).

In the CVD growth process, hydrogen is commonly used as a carrier gas. It covers the growth surface layer by forming Si-H or Ge-H bonds [42, 43], which could break easily under high temperature growth. Li et al. [40] reported that the hydrogen coverage on the growth surface may change the relative energy of surface and subsurface atoms, making phosphorus less energetically favored at the surface. To keep the hydrogen surface coverage high, the growth temperature for layers grown after n-SiGe supply layer is intentionally lowered (<575 °C). This high hydrogen coverage ratio successfully suppresses the segregation of phosphorus atoms towards the surface sites, leaving a relatively low P concentration at the sample surface. For example,
Figure 3.1: The schematic of atomic layer structures near the surface during the epitaxial growth for a doped 2DEG. (Image courtesy of Jiun-Yun Li [44])

the sample with its SiGe cap layer grown at 525 °C (50 °C lower than the sample shown in Fig. 3.2) shows a surface P level at $2 \times 10^{16}$ cm$^{-3}$ (Fig. 3.3), which is almost two orders of magnitude lower than the case shown in Fig. 3.2.

3.3 Device Fabrication

Prior to putting complicated split gates on a modulation-doped 2DEG to make a quantum device, a simple Hall bar with a Schottky gate across it was first made to check gate leakage current and test depletion capability of the gate. This set of devices was made both on samples with high (Fig. 3.2) and low (Fig. 3.3) surface P levels for comparison. After epitaxial growth by rapid thermal chemical vapor deposition (RTCVD), 2DEGs were then etched as Hall bars and AuSb was deposited as contacts followed by 450 °C annealing for 10 minutes. Pd was finally deposited across the Hall bar to form a Schottky gate. A simple schematic of a device cross section is shown
Figure 3.2: A SIMS profile for the sample with a cap layer grown at 575 °C. The phosphorus bump in the Si 2DEG is a SIMS artifact.

The mobility of 2DEG samples used in this study is above 200,000 cm²/Vs and 2D electron densities are in the range of $5 \times 9 \times 10^{11}$ cm⁻².

### 3.4 Gate Leakage Test at 4.2 K

Negative biases applied on splitting Schottky gates of a doped 2DEG deplete electrons in the channel to pattern them. However, before we test the depletion capability of Schottky gates, we need to assure that the gate leakage current (reverse current in the Schottky diode) is negligible at least in a wide enough range of negative biases at 4.2 K. In the case of the sample with the cap layer grown at 575 °C (blue circles in Fig. 3.4), a high electrical field near the surface due to high surface P concentration results in considerable gate leakage at a very small negative bias (-1 V) and its breakdown voltage is smaller than -2 V. In contrast, a very high breakdown voltage
Figure 3.3: The phosphorus segregation can be greatly suppressed by lowering cap layer growth temperature from 575 °C to 525 °C [40]. The phosphorus surface concentration is reduced as low as $2 \times 10^{16}$ cm$^{-3}$. The P peak at surface is a typical artifact resulted from the surface effect in a SIMS measurement. (-7 V) was achieved in the sample with the cap layers grown at 525 °C (red squares in Fig. 3.4). A wide range of negative biases without significant gate leakage enabled us to examine the depletion capability of this Pd Schottky gate next.

### 3.5 Depletion of Doped 2DEGs

A depletion test was then done in the sample with the SiGe cap layer grown at 525 °C. With a small voltage (-1 mV) between source and drain, a current flowed in the channel at zero gate bias. To test the efficiency of the Pd Schottky gate, we gradually increased the negative bias applied on the Schottky gate deposited across the channel. When the bias was less than -0.5 V, the drain current was still on and its level is around 130 nA, while the gate leakage current was 0.1 nA, which is the
detection limit of this measurement (Fig. 3.5). When the bias was increased over -0.5V, the channel was shut off sharply. The drain current plummeted down to ≈0.1 nA, the same level as the gate leakage current (the detection limit). No significant gate leakage was detected up to -2 V, which is consistent with the results from previous gate leakage test in the same sample.

3.6 Quantum Point Contact Test

A quantum point contact (QPC) is commonly integrated into quantum-dot devices as a charge sensor (Fig. 2.6). In contrast to the measurement of the current through the dot, which might be destructive to fine features, the QPC technique measures the tiny conductance change through the channel between QPC and the outer gate of quantum dots. The conductance change of a QPC has been confirmed to be very sensitive to its
Figure 3.5: Depletion test at 4.2 K of the sample with a cap layer grown at 525 °C. Both drain and gate leakage currents are shown in log scale in the inset.

electrostatic environment nearby and effective for measuring the number of electrons in dots \[45, 46, 47\].

The conductance through a QPC is quantized due to the narrow channel width. People noticed that a QPC is very sensitive especially as its quantized conductance is being turned on, where a small change in gate voltage leads to a considerable shift of conductance. Therefore, to maximize the sensitivity of a QPC, the width of the channel (The distance between QPC and the outer gate of a quantum dot) becomes essential. If the channel width is too wide, gates may not be able to completely shut off the channel. On the other hands, if the width is too narrow, the channel could be shut off too sharply. This resulting narrow bias window increased the difficulty in QPC operation. In the following QPC test, we designed five different sets of gates with various spacing, 80, 130, 180, 230 and 280 nm (inset of Fig. 3.6). By measuring
the conductance through these five different sets of gates, the best gap width to sense remote charges from a nearby quantum dots can be determined [45] (Fig. 3.6).

A QPC device was fabricated on the sample with very low gate leakage across a wide range of negative biases (Fig. 3.7). This QPC device sat on an etched mesa (dark grey regions), which extended conducting 2DEG regions out to eight ohmic contacts at the end. The fine gates, defined by ebeam-lithography on the mesa, were connected by eight other photo-lithography-defined Pd (wide bright regions) contacts climbing over the mesa for subsequent electrical controls.

To observe the change in conductance through the channels with various widths, negative biases were applied on $G_0$ and one of $G_1$ to $G_5$ simultaneously with all other gates grounded. A typical QPC conductance measurement is shown in Fig. 3.6, where

Figure 3.6: Successful QPC test at 4.2 K without any leakage. A wide transient region is observed. The inset shows the schematic of the QPC test structure.
$G_0$ and $G_5$ were biased with $G_1$ to $G_4$ grounded. Note that darker color means lower conductance. In this case, the black regions on the left bottom (while the negative biases on $G_0$ and $G_5$ were both higher than -1 V) show no conduction either through the channel between two gates or under any gate. In contrast, once one of gates was biased below around -0.6 V, the channel is fully open and very high conduction (bright yellow regions) was observed.

However, as we mentioned above, QPC reaches its highest sensitivity when the conductance is half of the quantized conduction. That means two gates should be biased properly so that the conduction between them is in orange regions. In this case with $G_5$ (the channel width is 80 nm), the bias window for both gates is around 0.3 V (-0.7 V to -1 V), which is fairly wide. Here we fixed $G_0$ at -1 V and compared the conductance between $G_0/G_1$ and $G_0/G_5$ (Fig. 3.8). For the gap width -280 nm (between $G_0$ and $G_1$), the current flowing in the gap could not be closed for any voltages. The high sensitivity was observed with the narrowest gap width -80 nm.
Figure 3.8: The channel between $G_0$ and $G_1$ remains on at -2 V but the channel between $G_0$ and $G_5$ can be fully shut off.

(between $G_0$ and $G_5$) where the current in the gap could be turned completely on and off within small gate voltages (Fig. 3.8).

3.7 Summary

In this chapter, we demonstrated suppression of phosphorus segregation by lowering growth temperature ($525 \, ^\circ C$) for the SiGe cap layer in a modulation-doped 2DEG. The resulting low surface P concentration reduced the electrical field at the surface under reverse bias, leading to a high breakdown voltage of Schottky gates. The wide range of negative biases without significant gate leakage enabled us to perform a successful depletion test where the Schottky gate shut off the 2DEG channel effectively. A QPC test device was then fabricated to determine the best channel width between
QPC and outer gate of a quantum dot to reach its highest sensitivity as a charge sensor.
Chapter 4

Implant Isolation of Silicon 2DEGs at 4.2 K

4.1 Introduction

As we discussed in Ch. 2, the state-of-the-art single-electron quantum-dot devices based on the Si/SiGe material system are typically fabricated in modulation-doped (depletion-mode) or undoped (enhancement-mode) 2-D electron gases (2DEGs) [15, 34, 48]. 2DEGs are usually electrically isolated by mesa etching. However, the mesa edges can cause problems for subsequent fabrication steps, such as the application of electron-beam resist for submicrometer gates to form quantum dots and thin metal step coverage. In addition, in enhancement-mode devices, high electrical fields in the gate insulator above the corner of the etched mesa may cause breakdown of the insulator or leakage currents.

Ion implantation for lateral electrical isolation (implant isolation) on III-V materials has been a well-known technique for several decades [49]. Ion bombardment creates deep defect levels, and these defects trap free electrons and pin the Fermi level near midgap, resulting in high resistivity. This process not only provides excellent
electrical isolation but also preserves the planarity of the surface. However, relatively few papers have focused on implant isolation in Si-based devices [50], because the resulting high-resistivity regions cannot sustain post-isolation high-temperature processes (>1000 °C) common in silicon technology. Furthermore, the resistivity of the intrinsic silicon (Fermi level at mid-gap) is \( \sim 2 \times 10^5 \, \Omega \cdot \text{cm} \) at room temperature, which is not high enough for most applications. However, the processing of Si/SiGe-based quantum devices is often constrained to be below 600 °C to avoid Si/Ge interdiffusion [51], which could be low enough to avoid annealing of implant damage. Besides, the typical low operation temperature (4.2 K or less) of such quantum devices provides much less thermal energy for electrons to escape from implant-induced defects, so that resistivities much higher than those in silicon at room temperature should be possible.

In this chapter, we demonstrate implant isolation of modulation-doped Si 2DEG structures characterized at 4.2 K. Heavily-doped 2DEGs were used to examine the isolation capability as a worst case (high electron density of \( -10^{12} \, \text{cm}^{-2} \)). The thermal stability was tested for different post-implant annealing temperatures up to 650 °C. The 2DEG quality (electron mobility) of samples processed with implant isolation was compared with ones with conventional mesa isolation by reactive ion etching (RIE). All the work of this chapter is summarized in [52].

### 4.2 Sample Growth and Test Device Fabrication

The layer structure of the modulation-doped Si 2DEGs used in this study is shown in Fig. 4.1. Their Hall mobility, electron density, and sheet resistance at 4.2 K are in the range of 80,000 to 150,000 cm²/Vs, 0.8 to 1.6×10¹² cm⁻², and 30 to 80 Ω/□, respectively. The structures were all grown on Si₀.₇₂Ge₀.₂₈ graded buffer substrates by rapid thermal chemical vapor deposition (RTCVD) between 575 °C and 625 °C.
test device consists of a set of separated Ohmic contacts to test the implant isolation, a set of connected Ohmic contacts to measure original resistivity from a 2DEG, and a Hall bar structure (in a single 2DEG region) with Ohmic contacts to measure electron mobility and density (Fig. 4.2).

After 2DEG growth, a 400-nm screen silicon dioxide layer was first deposited by plasma-enhanced chemical vapor deposition (PECVD) at 250 °C. Ohmic contact regions to the 2DEG were then defined by photolithography and diluted HF wet etching of the oxide. 1% Sb-doped Au was thermally evaporated followed by lift-off. Annealing at 450 °C for 10 minutes to form Ohmic alloyed contacts was performed before the ion implantation except for samples later annealed at 550 °C or 650 °C, where the contacts were formed after 550 °C or 650 °C steps. The areas to be isolated were then defined by photolithography and diluted HF wet etching of the oxide.
4.3 Implant Conditions

Isolation tests were done by implanting separately argon (Ar\textsuperscript{+}) and silicon (Si\textsuperscript{+}) ions into different samples. Si and Ar were chosen due to their electrical neutrality in the Si/SiGe material system. Two implant energies were necessary in all regions because both the 2DEG channel and the doping layer (if doped above the metal-insulator transition level\cite{53}) could conduct electricity at low temperature. A simulator, the Stopping and Range of Ions in Matters (SRIM)\cite{54}, was used to simulate the profile of implanted species and the resulting Si and Ge vacancies. This information helped us to choose two ion implantation energies: 30 keV and 60 keV, which created defects near the depth of the shallow doping layer (30 keV) and near the deeper strained Si 2DEG channel (60 keV), respectively (Fig. 4.3).

Silicon amorphized by a high dose implant (-5×10\textsuperscript{14} cm\textsuperscript{-2} for Si\textsuperscript{+} into Si at 40 keV)\cite{55} can be recrystallized by solid phase epitaxy (SPE) at a temperature as low as 500 °C, which would lead to poor thermal stability of the damage\cite{56}. Further, SiGe alloys are even more easily amorphized by ion implantation than Si due to weaker bonding between Si and Ge atoms\cite{57}. Therefore, doses well below 5×10\textsuperscript{14} cm\textsuperscript{-2} were used. Ar\textsuperscript{+} or Si\textsuperscript{+} was then implanted at room temperature with three
Figure 4.3: Simulation of implanted species and resulting vacancy distribution in a 2DEG structure by the Stopping and Range of Ions in Matters (SRIM) software [54]. Two-step implantation ($1 \times 10^{14}$ cm$^{-2}$ @ 30 keV + $1 \times 10^{14}$ cm$^{-2}$ @ 60 keV) with implanted species (a) Ar$^+$ and (b) Si$^+$ is used in this simulation. The target is assumed to be implanted at 0 K.
different test doses: 5×10^{11} \text{ cm}^{-2} \text{ (denoted as low dose)}, 1×10^{13} \text{ cm}^{-2} \text{ (medium dose)} and 1×10^{14} \text{ cm}^{-2} \text{ (high dose)}, and each dose was implanted at two energies, 30 keV and 60 keV.

In addition, Hall bars defined by conventional mesa isolation by RIE were also made for 2DEG quality comparison. Ohmic contacts were made on one Si_{0.72}Ge_{0.28} graded buffer substrate separately without implant isolation to test its resistivity at 4.2 K for reference.

### 4.4 Ultra High Sheet Resistance at 4.2 K

The isolation capability at 4.2 K was examined by two-point measurements between two implant-isolated 2DEGs. The sheet resistances of isolated regions for all six implant conditions are all above 1×10^{12} \text{ Ω}/\square, which is ten orders of magnitude higher than the original 2DEG sheet resistances (Fig. 4.4). In some cases, the sheet resistance is as high as 1×10^{13} \text{ Ω}/\square, close to the instrumental limit. Since the sheet resistance of the Si_{0.72}Ge_{0.28} graded buffer substrate at 4.2 K was also high (-5×10^{12} \text{ Ω}/\square), the remaining conduction might occur either in the SiGe graded buffer or in the implanted regions. In any case, it is clear that the implant isolation for all three doses is extremely effective.

### 4.5 Thermal Stability

Thermal stability issues could arise when we integrate the implant isolation technique into a device fabrication process, such as insulator deposition and contact annealing. Aluminum oxide deposited by atomic layer deposition (ALD) is a common insulator used on enhancement-mode Si 2DEG devices due to its high quality and low deposition temperature (≤ 300 °C) [58]. To test the thermal stability at the oxide deposition temperature, samples were annealed at 300 °C for 1 hour and their sheet resistances
Figure 4.4: Isochronal (1 hour) annealing behavior of sheet resistances of 2DEG samples at 4.2 K implanted with (a) Ar\(^+\) and (b) Si\(^+\) with three various doses, \(5 \times 10^{11}\) cm\(^{-2}\), \(1 \times 10^{13}\) cm\(^{-2}\) and \(1 \times 10^{14}\) cm\(^{-2}\) at 30 keV and 60 keV. Also shown are the range of sheet resistance of the starting 2DEGs, and the experimental instrumental limitation. The estimated error of sheet resistance is \(\pm 18\%\) for (a) and \(27\%\) for (b).
were measured at 4.2 K (Fig. 4.4). For samples implanted with medium and high
dose, the sheet resistances stay high at $-1 \times 10^{13} \, \Omega/\square$ regardless of implant species,
while for ones implanted with a low dose, the sheet resistances drop to $3 \times 10^{10} \, \Omega/\square$
and $3 \times 10^{9} \, \Omega/\square$ for Ar$^+$ and Si$^+$ implant respectively. Since implantation with the
lowest dose ($1 \times 10^{11} \, \text{cm}^{-2}$) produces the fewest defects (mostly point defects), the
damage can easily be annealed and thus the sheet resistance decreases. In addition,
the higher sheet resistance of low dose Ar$^+$-implanted sample than Si$^+$-implanted one
after 300 °C annealing might be explained by a higher damage level caused by Ar$^+$
due to its heavier atomic mass.

Antimony-doped gold is commonly used on both depletion-mode and enhancement-
mode Si/SiGe devices to form Ohmic alloyed contacts after annealing [58]. However,
in enhancement-mode devices, the necessity of an overlap between the accumulated
electrons in the channel under the gate and contacts for the continuity of conduc-
tion makes the requirement of a flat contact surface crucial to prevent possible
leakage. Because alloyed contacts have rough surfaces, contacts made by n-type
ion implantation before the deposition of the gate insulator are preferred [15, 48].
For heavily-phosphorus-implanted contacts, the activation of implanted phosphorus
occurs at relatively low temperature ($\geq 500$ °C) by SPE [56]. Hence, the thermal
stability of implant isolation samples was again tested after 550 °C annealing for 1
hour. Sheet resistances of the medium and high dose samples at 4.2 K are still as
high as $1 \times 10^{13} \, \Omega/\square$ (Fig. 4.4). After further annealing at 650 °C for 1 hour, the
sheet resistances of the medium and high dose samples at 4.2 K drop nine orders of
magnitude. Therefore, if an annealing temperature for the contact over 550 °C is
desired, it should be done before implant isolation.
4.6 2DEG Quality Check

Any possible degradation of the 2DEG quality due to spurious radiation during ion implantation was investigated based on Hall mobility measurements by standard low frequency lock-in techniques at 4.2 K. Fig. 4.5 shows the ratio of mobility measured from implant-isolation-defined Hall bars to RIE-defined ones (with absolute mobility values shown in parentheses) on the same CVD growth and annealing conditions. Implanted samples without annealing (open symbols) do not show any significant mobility degradation. Even with the highest dose used in this study ($1 \times 10^{14}$ cm$^{-2}$), the 2DEG quality remains unaffected. Mobility ratios of samples experiencing both implant isolation and 550 °C annealing are also shown in Fig. 4.5 (closed symbols). Except for one sample, the mobility ratios after 550 °C annealing are near 100%. We attribute the low mobility ratio for the medium dose Ar$^+$-implanted sample to a non-uniformity in the 2DEG growth between different wafer pieces used for the Hall bars.

4.7 Summary

Successful lateral electrical isolation of silicon two-dimensional electron gases (2DEGs) at liquid helium temperature (4.2 K) by ion implantation is demonstrated. The sheet resistance of the implanted regions can be achieved as high as $1 \times 10^{13}$ Ω/□ at 4.2 K. Thermal stability up to 550 °C makes the technique compatible with most subsequent processing steps to fabricate silicon quantum devices. It has also been confirmed that the 2DEG quality is not degraded by the ion implantation, based on a comparison of Hall mobility of implant-isolated samples with conventional reactive-ion-etching (RIE)-defined samples.
Figure 4.5: Comparison of 2DEG quality after implant isolation or RIE process by the ratio of the mobility at 4.2 K measured from an implant-isolation-defined Hall bar to a RIE-defined Hall bar for a given implant and anneal. The absolute mobility values at 4.2 K with a unit of $10^3 \text{ cm}^2/\text{V}s$ are shown in parentheses with the mobility for implant isolation before the mobility for mesa isolation. The estimated error of mobility measurement is ±5%.
Chapter 5

Enhancement-Mode Undoped
Silicon 2DEGs

5.1 Motivation

The silicon-based material system is attractive for the implementation of single-electron quantum dot (QD) devices for quantum computing applications owing to the longer spin coherence time of electrons in silicon compared to that in III-V compounds [58, 59]. Current research interests are focused more on enhancement-mode quantum devices in undoped silicon two-dimensional electron gases (2DEGs) due to the absence of ionized dopants, which are possible sources of disorder and potential fluctuations in Si 2DEGs [39, 60, 61]. In an enhancement-mode structure, a strained-silicon quantum well (QW) which confines a 2DEG is buried below the surface to reduce the scattering from remote charges from the semiconductor surface. Weaker Coulomb forces due to both the absence of ionized dopants and a longer distance between surface charges and electrons lead to a cleaner 2DEG system where the interference to spin manipulation could be mitigated and electron mobility could be enhanced.
The laws of physics do not show a direct link between electron mobility and electron spin coherence time, but qualitatively speaking, the electron mobility does reflect the detrimental influence from charged impurities in a 2DEG system. In addition, the easier process steps for a Hall bar device and the faster low-temperature Hall measurements (which measures mobility) compared to spin relaxation experiments do provide a beneficial measure of the cleanliness of a strained silicon 2DEG system.

To confine a single electron in a quantum dot, a low minimum electron density in the 2DEG enables larger tolerances in both the size of a quantum dot device and the bias window for depletion gates. Furthermore, a 2D electron density is also an essential parameter to reflect a 2DEG quality because, theoretically, the lowest mobile electron density that is measurable at liquid helium temperature is strongly related to the density of remote charges at the oxide/silicon interface. To sum up, higher electron mobility and lower measurable electron density of an enhancement-mode strained silicon 2DEG qualitatively represent the good quality of a 2DEG system which may yield a longer spin coherence time. A Hall measurement at 4.2 K is a fast and easy measurement to obtain these properties.

5.2 Models for Mobility Limitation

To enhance electron mobility of undoped silicon 2DEG samples grown in our RTCVD, understanding the mobility-limiting mechanisms is the first step. Historically, much work was done to figure out the dominant scattering mechanisms on electron mobility in a MOSFET [62]. The silicon/silicon dioxide interface was identified as the key role that limits electron mobility, especially at high vertical electrical field, in addition to the phonon scattering at room temperature. Relevant scattering mechanisms for a 2DEG in GaAs/AlGaAs systems were also widely discussed in 1980s [63, 64, 65]. Thanks to these theoretical analyses along with plenty of experiments in the past
few decades, the most updated electron mobility record of a 2DEG in GaAs/AlGaAs material system has been improved to as high as near 40,000,000 cm$^2$/Vs at 0.3 K [66].

The discovery of electron confinement in strained silicon between two relaxed SiGe layers by Abstreiter et al. [22] in 1985 boosted the electron mobility record in Si/SiGe material system. In less than 10 years, the low-temperature electron mobility in a doped Si 2DEG was improved up to 500,000 cm$^2$/Vs, which was two orders of magnitude higher than before [30]. However, no systematic theoretical analysis on electron-limiting mechanisms in the Si/SiGe 2D electron system had been proposed until 1993. Monroe et al. [67] first analyzed possible scattering mechanisms that limit the mobility enhancement in the Si/SiGe system based on the previous discussions in silicon MOSFET and GaAs material systems. In the following section, we will use this literature to introduce the most dominant scattering mechanisms in detail.

5.2.1 Scattering from Background Charged Impurities

We can easily rule out phonon scattering because the operating temperature of quantum-dot devices based on silicon 2DEGs is commonly below 4.2 K (liquid helium temperature), where the phonon scattering is definitely negligible.

The first possible scattering mechanism limiting the electron mobility at 4.2 K in our samples could be scattering from background charged impurities in the 2DEG layer itself, which is usually denoted as background scattering [67]. The predicted electron mobility influenced by background scattering is:

$$
\mu_{bs} = \frac{g_v \frac{3}{2} g_s \frac{3}{2} e n_{2D} \frac{1}{2}}{4\pi \frac{1}{2} \hbar N_b}
$$

(5.1)

where $g_v$ and $g_s$ are valley degeneracy and spin degeneracy, respectively. $N_b$ is the density of background charged impurities while $\mu_{bs}$ is the mobility limited by background scattering. Because of the high Fermi level in the 2DEG, donors may be
occupied and neutral. Neutral donors can also be strong scattering sites. For simplicity, when discussing and calculating background scattering in our modeling, we use Eq. 5.1 and set $N_b$ to be the total number of background impurities, and ignore their charge states.

The background charged impurities are those impurities existing in the strained silicon channel where a 2DEG sits (Fig. 5.1). Due to their proximity to mobile electrons, the Coulomb forces from charged impurities can easily scatter electrons in the channel and thus deteriorate mobility. A higher density of background charged impurities leads to lower electron mobility. Furthermore, with electron density increasing, more electrons can screen each other from the scattering of background charged impurities. This so-called self-screening effect on background screening allows higher electron mobility at higher electron density, with a theoretical power law of 0.5.

### 5.2.2 Scattering from Remote Charged Impurities

Another major scattering mechanism is the scattering from remote charged impurities, usually denoted as remote scattering. In addition to charged impurities in the strained...
silicon channel, charged impurities could be incorporated into grown layers during the growth, intentionally or unintentionally. For a doped silicon 2DEG, an intentional n-doped SiGe layer is designed to supply electrons into the strained silicon channel a few tens of nanometers away from the doping layer. This ionized impurity sheet with fairly high density (~10^{11} \text{ cm}^{-2} or greater) results in strong Coulomb forces, scattering electrons in the channel strongly. For a conventional doped 2DEG in GaAs system with ultra-high mobility, remote scattering from the electron supply layer can be the dominant mobility-limiting mechanism.

In an undoped enhancement-mode silicon 2DEG, no intentionally doped layer exists anymore. Instead, charged impurities at the interface between the oxide and the strained silicon cap become the main sources of the remote scattering (Fig. 5.2). Although the predicted mobility model limited by remote scattering in [67] was proposed for a doped 2DEG sample, we could still apply it into an undoped 2DEG case due to the same nature of the scattering sources:

$$\mu_{rs} = \frac{16\pi^{1/2}g_{e}^{1/2}g_{s}^{1/2}e\hbar n_{2D}^{3/2} s^{3}}{hN_{r}} \quad (5.2)$$

where $N_{r}$ is the 2D density of remote charged impurities, and $\mu_{rs}$ is the mobility limited by remote scattering. $s$ is the distance between silicon channel and the location of the remote charged impurity sheet. Since the SiGe cap thickness is usually much thicker than Si cap, the former usually represents $s$ directly.

It is quite straightforward to understand from Eq. 5.2 that Coulomb forces from charged impurities could be reduced by making the SiGe cap thicker or the density of remote charged impurities lower. More detailed discussions about how to improve mobility limited by remote scattering in an undoped sample are in Sec. 5.6. Here we can also see the self-screening effect is described in this theoretical equation. The principle of self-screening on remote scattering is basically the same as the one on background scattering, but remote scattering has a higher power law, 1.5, meaning
that the theory predicts self-screening is more effective for remote scattering than background scattering.

5.2.3 Scattering from Si/SiGe Interface Roughness

The scattering from Si/SiO$_2$ interface roughness (denoted as interface roughness scattering) limits the electron mobility in a conventional silicon MOSFET in the high density regime. Despite very low interface charge densities in today's advanced MOSFETs, the inherently different material nature between silicon and oxide causes roughness [69], reducing the mobility of electrons moving along this interface. Fortunately, electrons in an undoped silicon 2DEG are located in a buried silicon channel, hence the interface between silicon and oxide is not an issue anymore. Instead, they move along the interface between the strained silicon layer and the relaxed SiGe cap layer (Fig. 5.3). The Si/SiGe interface is expected to be superior to Si/SiO$_2$ interface in terms of the roughness not only because Si and SiGe have similar material properties, but Si and SiGe layers were epitaxially grown successively, without exposure out of the vacuum. The predicted mobility model limited by the scattering from interface roughness is as follows:
\[ \mu_{irs} = \frac{e^5 g_v^3 g_s^3}{192\pi^4 h(e\epsilon_0)^2(\partial V/\partial z)^2(\Lambda\Delta)^2 n_{2D}} \]  

(5.3)

where \( \Lambda \) and \( \Delta \) are characteristic length of the interface roughness and rms roughness, respectively. \( \partial V/\partial z \) is a perturbation factor and \( \mu_{irs} \) is the mobility limited by interface roughness scattering. In contrast to previous two scattering mechanisms, the mobility model limited by interface roughness scattering is reduced as 2D electron density increases. This is consistent with the experimental data from Si MOSFETs [70]. More electrons come in the channel from ohmic contacts as the higher positive gate voltage is applied, where a higher vertical electrical field confines electrons in the quantum well closer to the Si/SiGe interface, and thus enhances interface roughness scattering on mobility [71].

### 5.2.4 Other Possible Scattering Mechanisms

The strained silicon 2DEG system is much more complicated than a silicon MOSFET or GaAs 2DEG system because of the lattice mismatch of Si and Ge. In order to strain silicon to create the conduction band offset, relaxed SiGe layers are essential for a silicon 2DEG. Even state-of-the-art relaxed SiGe buffer layers contain various dislocation defects, including threading dislocations and misfit dislocations [23, 72].

---

Figure 5.3: The schematic of Si/SiGe interface roughness in an undoped silicon 2DEG.
Therefore, the scattering from these dislocations may also influence electron mobility in a 2DEG system. Moreover, in practice, electrons in the strained silicon channel are not perfectly confined in the strained silicon channel due to finite barrier height. The tail of electron wave function may spread into the nearby relaxed SiGe layer, where alloy scattering from the random SiGe alloy could occur. By taking all scattering mechanisms into account, we can write the total theoretical mobility by Eq. 5.4.

\[
\frac{1}{\mu_{\text{total}}} = \frac{1}{\mu_{bs}} + \frac{1}{\mu_{rs}} + \frac{1}{\mu_{irs}} + \cdots 
\]  

(5.4)

In the following section, we will compare our experimental data with the total effect from the three different scattering mechanisms described here to see which mechanism dominates in 2DEG samples grown in our RTCVD.

5.3 Measurement of Enhancement-Mode Undoped Silicon 2DEGs

5.3.1 Device Fabrication and Measurement Setup

To study the electron transport properties of a strained silicon 2DEG at low temperature, an enhancement-mode Hall bar device is fabricated (Fig. 5.4) to measure mobility and 2D electron density at liquid helium temperature (4.2 K). First, a 200-nm silicon dioxide layer was deposited by plasma-enhanced chemical vapor deposition (PECVD) on the as-grown structure as an implant mask. The designated regions for Ohmic contacts were then defined by photolithography and wet etching of the silicon dioxide. A 3-step phosphorus implantation was conducted to assure the implanted species were deep enough to form contacts from the surface to the strained Si QW. Samples were then annealed at 600 °C for 1 hour to activate the implanted phos-
Figure 5.4: The layer structure of a typical undoped enhancement-mode strained silicon 2DEG.

Next, a 90-nm aluminum oxide was deposited by atomic layer deposition (ALD) at 300 °C as a gate insulator. Parts of the n⁺ implanted regions were exposed by photolithography and wet etching of the aluminum oxide for metal deposition for contacts. After another photolithography step, a Hall-bar-shaped metal gate was finally formed by evaporating a 2-nm chrome and a 200-nm gold layer along with metal deposition on the exposed n⁺ contact regions in the same evaporation.

We note that the metal gate must overlap with the n⁺ contact regions to keep the continuity of conduction from the contacts to the 2DEG (Fig. 5.4). The process flow depicted above is standard for all samples discussed in this chapter unless otherwise noted. The top view of a standard enhancement-mode Hall bar device used in this study is shown in Fig. 5.5.

Hall measurements were then conducted at liquid helium temperature (4.2 K) with a magnetic field up to 2 T. The current source into the Hall bar is AC modulated with a low frequency (11 Hz) and a current level of 100 nA.
5.3.2 Parallel Capacitor Plate Model

When the gate voltage ($V_G$) is above the threshold voltage, electrons are capacitively induced to form a 2DEG, leading to the onset of channel conductance (The inset of Fig. 5.6). A typical gate voltage ($V_G$) dependence of the 2D electron density ($n_{2D}$) was thus obtained (Fig. 5.6). Since there are two strained silicon quantum wells (strained silicon cap and strained silicon channel) in a typical enhancement-mode Si/SiGe 2DEG structure, a simple and quick parallel-capacitor-plate method is commonly used to identify the location of a gate-induced 2DEG instead of the time-consuming Shubnikov-de-Haas oscillation measurement [32]. Let's assume this conducting 2DEG is in the strained silicon channel first. Thus between a metal gate and this conducting 2DEG, the oxide is equivalent to a capacitor (denoted as $C_{ox}$), which is in series with another capacitor from Si and SiGe cap layers (denoted as $C_{sc}$). Given the thickness of the aluminum oxide (with a relative dielectric constant of 9) , Si cap and SiGe cap shown in Fig. 5.7 the equivalent theoretical capacitance calculated from this parallel capacitance plate model is $6.1 \times 10^{-8}$ F/cm$^2$, which is very
Figure 5.6: A typical linear $V_G$ dependence of 2D electron density in an undoped enhancement-mode silicon 2DEG. The inset shows the channel turn-on with increasing gate voltages.

close to the experimental capacitance ($5.3 \times 10^{-8}$ F/cm$^2$) described by the slope of $n_{2D}$ versus $V_G$ in Fig. 5.6. This consistency strongly suggests that the electrons induced by the bias on metal gate are located in the strained silicon channel, not the silicon cap.
Figure 5.7: A schematic of two equivalent capacitors in series between metal gate and a conducting 2DEG in strained silicon channel.

5.4 Transport Property of Undoped Silicon 2DEGs

5.4.1 A Brief View of Various Scattering Mechanisms in Our 2DEGs

The Hall electron mobility measured from two typical undoped enhancement-mode silicon 2DEGs grown in our RTCVD system were plotted versus 2D electron density (Fig. 5.8). The layer structures for these two 2DEGs were the same except for the SiGe cap thickness. The highest $\mu$ of samples with a 55-nm (blue circles) and a 27-nm (orange triangles) SiGe cap are $\sim$400,000 cm$^2$/Vs and $\sim$200,000 cm$^2$/Vs, respectively.

To identify the dominant limiting factors on mobility of 2DEGs grown in our RTCVD system, we placed theoretical model curves for the three main scattering mechanisms mentioned in Sec. 5.2 together with our data (Fig. 5.8). Several fitting parameters are chosen to optimize the fittings. The total scattering is also plotted in thick brown lines based on Eq. 5.4.
Over a wide range of 2D electron density values, the dominant scattering mechanism was the scattering of remote charged impurities ($-n_{2D}^{1.5}$). The density of remote charges ($N_r$) chosen here are $7.5 \times 10^{12}$ cm$^{-2}$ and $3.5 \times 10^{12}$ cm$^{-2}$ for the samples with a 55-nm and a 27-nm SiGe cap respectively. Such high densities of interface charges are actually in the range of other experimental data for the interface of an aluminum oxide by ALD and silicon [74]. The scattering of background charged impurities with total impurity density, $N_b = 4 \times 10^{14}$ cm$^{-3}$, measured by secondary ion mass spectroscopy (SIMS), is not significant (See Sec. 5.5). In addition, at high density, interface roughness scattering may account for the saturation of the mobility. The product of the characteristic length of the interface roughness $\Lambda$ and rms roughness $\Delta$ are 100 Å$^2$ and 110 Å$^2$ for the samples with a 55-nm and a 27-nm SiGe cap, respectively.
Except for the theoretical curve of background scattering, whose parameter, $N_b$, was from an experimental measurement, the fitting curves for remote scattering and interface roughness scattering were both based on optimizations of respective fitting parameters. More detailed experiments and quantitative analyses for individual scattering mechanism will be discussed in the following several sections.

5.4.2 Metal-Insulator Transition and 2D Critical Density

The importance of the 2D electron density of a 2DEG in quantum computing applications is no less than the electron mobility. As we mentioned in Sec. 5.1, a lower 2D electron density offers a wider window for the geometry design of a quantum dot, allowing a single electron to be obtained in a larger quantum dot and thus facilitating the process for sub-micron fine features. In Fig. 5.8, the lowest $n_{2D}$ of samples with a 55-nm (circles) and a 27-nm (triangles) SiGe cap are $4.9 \times 10^{10}$ cm$^{-2}$ and $1.1 \times 10^{11}$ cm$^{-2}$ respectively, which are the record low density in the samples previously reported with comparable SiGe cap thicknesses [39, 75].

The lowest $n_{2D}$ in a 2DEG is usually referred to critical 2D electron density, which is theoretically affected by potential fluctuations from charges at scattering sites [60, 61]. In an undoped 2DEG, the randomly distributed scattering sites at the oxide/silicon interface non-uniformly influence the potentials in the remote strained silicon channel, leaving a non-uniform conduction band minimum as shown in Fig. 5.9. If an electron density induced by gate voltage is high enough, a 2DEG conducts as a metal (Fig. 5.10a). On the other hand, below a critical electron density, a 2DEG does not conduct and act as an insulator, where electrons are localized in the valleys of potential fluctuations and can not escape them due to lack of thermal energy at liquid helium temperature (Fig. 5.10b). Therefore, the onset of the channel conduction is also called as Metal-Insulator Transition (MIT). The first measurable density point from Hall measurement is called critical 2D electron density.
5.4.3 Experimental Observation of Metal-Insulator Transition

A typical metal-insulator transition can be very sharp in most of undoped 2DEGs. The channel may be turned on or off suddenly with a very small increment in gate voltage -0.01 V. For example, for the sample with a 55-nm SiGe cap in Fig. 5.8, the channel is suddenly shut off from a fairly conducting status (mobility is $>10,000$ cm$^2$/Vs despite the density as low as $4.9 \times 10^{10}$ cm$^{-2}$) to full insulation with only a 0.01 V gate voltage difference. However, instead of this sharp shut-off, a nonlinear decrease of $\mu$ was observed in the sample with a 27-nm SiGe cap in the low density regime [76] (Fig. 5.8). Similar behavior, a nonlinear drop in mobility in the low density regime,
Figure 5.10: (a) A 2DEG conducts as a metal in the channel when the electron density is high (b) A 2DEG acts as an insulator because electrons are localized in the valleys of fluctuations when the electron density is too low.

was also reported in doped 2DEGs in the GaAs/AlGaAs material system [68]. This drop is possibly described by a metal-insulation-transition (MIT) model, assuming the sources of fluctuations are charged impurities at the aluminum oxide/Si interface (Fig. 5.11). The model used here includes a parameter $A$ [77]:

$$A = \frac{N_r}{16\pi n_{2D}^2 s^2}, \text{ if } 4k_F s >> 1 \quad (5.5)$$

where $s$ is the SiGe cap thickness, $N_r$ is the 2D density of remote charged impurities and $k_F$ is the Fermi wavenumber. Therefore, the mobility limited by the scattering of remote charged impurities can be modified as

$$\mu_{rs}(n_{2D}) = \mu_{rs0}(n_{2D})(1 - A) \quad (5.6)$$

In the high density regime, $A$ is much smaller than 1, meaning that $\mu_{rs}$ is close to $\mu_{rs0}$. The mobility is not degraded because the electron density is high enough to screen the influence of potential fluctuations from remote charged impurities. However, once 2D density decreases (given a fixed $N_r$ and $s$) and the parameter $A$ approaches 1, $\mu_{rs}$ decreases sharply. The original model of the scattering from remote charged impurities is no longer applicable. When $A$ is 1, the mobility drops to zero, and the metal-insulator transition occurs.
Figure 5.11: The modified mobility model by a MIT model (green curve) fit the data of the sample with a 27-nm SiGe cap, especially in the low density regime.

### 5.4.4 Failure of the MIT Model

The sharp decrease of $\mu$ in the sample with a 27-nm SiGe cap is well fitted by the MIT model, which indicates that the remote charged impurities at the aluminum oxide/silicon interface have a strong influence on the sample with such a thin cap layer. However, no similar phenomenon was observed in the sample with a 55-nm SiGe cap (Fig. 5.8). A trial to apply the MIT model mentioned in the previous subsection on this sample showed a great discrepancy in data and the modified mobility curve (green line in Fig. 5.12). The disagreement in data and the modified mobility model by the MIT model could be discussed from several aspects. First, the extracted density of remote charged impurities ($N_r$) based on the theoretical equation (Eq. 5.2) may not be accurate. For example, if the mobility curve shown here is limited by a certain
Figure 5.12: A discrepancy between data points of the sample with a 55-nm SiGe cap and the modified model by the MIT model. The modified mobility model is the green line, while the original model is in black.

unknown mechanism, not simply by remote scattering (despite that the slope of data is very close to 1.5), the extracted $N_r$ ($7.5 \times 10^{12}$ cm$^{-2}$) is then doubtful. Second, it could be arbitrary to equate the density of remote charged impurities causing remote scattering that limits mobility to the one that results in potential fluctuation described in this MIT model. Moreover, we could not rule out the possibility that the theory of remote charge scattering is not perfect and complete for Si/SiGe 2DEGs as well as the MIT model. The primary foundations of both theories are still theoretical calculations and experimental data from GaAs/AlGaAs 2DEGs. Imperfection in crystal structures of SiGe alloys and strained silicon could easily raise the uncertainty and complexity in theoretical predictions.
5.5 Background Scattering Alleviation

5.5.1 High Background Impurity Level in Old Samples

We discussed how the scattering from background charged impurities limits the electron mobility of a 2DEG in Sec. 5.2.1. The density of background charged impurities is critical to determine where the mobility level would be expected if background scattering is dominant in a system. In our RTCVD system, extremely high doses of dopants, such as boron or phosphorus, have been used for decades. Remnant dopant atoms could attach along the inner walls of gas pipes, quartz reactor tubes or even quartz wafer holders. Once a newly-loaded wafer is baked at high temperature for cleaning, those buried atoms may evaporate out and incorporate into the newly-grown epitaxial layers, leading to a high background concentration of unexpected dopants. For example, a SIMS analysis clearly showed very high background concentration of both phosphorus (>2×10^{17} \text{ cm}^{-3}) and boron (>1×10^{16} \text{ cm}^{-3}) in an undoped 2DEG grown in our RTCVD system in the year 2010 (Fig. 5.13). If we put \(N_b = 1\times10^{17} \text{ cm}^{-3}\) into Eq. 5.1 with an electron density in a range of \(10^{11} \text{ cm}^{-2}\) to \(10^{12} \text{ cm}^{-2}\), the predicted mobility is as low as 10,000 cm$^2$/Vs, which is about the same level as our old experimental data.

5.5.2 Efforts to Alleviate Background Scattering

To figure out a way to reduce the density of background charged impurities, the first step is to identify where those undesired dopant atoms exactly come from. A simple way is to load a new quartz wafer stand or replace the old quartz reactor tube with a brand new one, and grow a test structure for a SIMS analysis. Unfortunately, SIMS showed very high background doping levels in all the grown layers, especially phosphorus, given the source gases were expected to have very low impurity background levels. Thus it was concluded that those dopant atoms came into the reactor...
Figure 5.13: The SIMS profile of a typical undoped 2DEG grown in our RTCVD before a new gas supply system was introduced. Clear high concentrations of both phosphorus and boron were observed, along with gases from pre-existing contamination in the old gas supply pipes, where the precursor of phosphorus, phosphine (PH$_3$), had been mixed with other precursor gases and hydrogen for decades.

To avoid phosphorus atoms from mixing into other process gases again, a new gas supply assembly was designed and a new gas supply panel was introduced in collaboration with Jiun-Yun Li as shown in Fig. 5.14. The pipe length from the new gas supply panel to reactor is about 5 meters long. We separated other process gases from phosphine so that they are sent to the gas supply assembly by two different pipes. In addition, the venting paths to process pumps were also separated. Based on this design, phosphine is only mixed with other process gases in the gas supply assembly, which is located at only a small distance (~50 cm) to the entry of the
Figure 5.14: The schematic of the new gas supply panel and the gas supply assembly. Phosphine is separately transported from other process gases until a half meter away from the reactor entry to avoid mixture and contamination.

An undoped silicon 2DEG sample was grown with the new gas supply panel and the gas supply assembly, and was sent to Evans Analysis Group (EAG) for high precision SIMS analysis for phosphorus and boron (Fig. 5.15). We noted that the Ge profile here is only for reference and is not precise because the measurements were specifically optimized for phosphorus or boron. Except for high concentrations at the surface caused by the typical surface effect of a SIMS measurement, phosphorus level plummeted to $1 \times 10^{14}$ cm$^{-3}$ while boron level was reduced to $5 \times 10^{13}$ cm$^{-3}$, both hitting the detection limit and at least getting 1000 times lower than those levels observed in samples grown by the old gas panel. The predicted mobility by such low
Figure 5.15: Extremely low concentrations of both phosphorus and boron in grown layers after a new gas panel and a gas supply assembly were installed. Both doping levels hit the detection limits of the high-precision SIMS analysis.

densities of background charged impurities by Eq. 5.1 is over $10,000,000$ cm$^2$/Vs as the electron density is in the range of $10^{11}$ cm$^{-2}$. Such a high value is much higher than the mobility we measured over the whole range of 2D electron density, implying that the background scattering has become negligible in our system.

5.6 Effects of Remote Scattering

5.6.1 Effects of SiGe Cap Thickness

The scattering from remote charged impurities at the oxide/silicon interface in an undoped silicon 2DEG undoubtedly becomes the major mobility-limiting factor, when the background scattering is alleviated by reducing background doping concentra-
tions. The slope of mobility data in log-log scale in Fig. 5.8 are clearly close to 1.5, which is the theoretical characteristic of the remote scattering. With proper fitting parameters (density of remote charged impurity), the model curve fit a wide range of data (Fig. 5.8). However, the independent measurement of the exact density of interface charged impurities was difficult. Due to the complexity of multiple Si/SiGe heterostructure layers underneath the oxide, plus the limitations of C-V measurements at liquid helium temperature, no meaningful C-V data were obtained yet to allow us to quantitatively discuss the remote charge density and resulting remote scattering.

Nevertheless, a hint from experimental data still strongly supports our argument that remote scattering really plays a crucial role in our samples. A series of Hall measurement data from samples with various SiGe cap thicknesses (14, 20, 40, 80, 120, and 180 nm) were plotted together for comparison in Fig. 5.16. Note that all other growth and process recipes for this batch of samples remained the same. The highest mobility and lowest n_{2D} (critical density) for each sample versus SiGe cap thickness are shown in Fig. 5.17.

A trend that the mobility curves moved upwards and leftwards with thicker SiGe cap thickness is clearly observed, at least for samples with cap layer <80 nm. For example, the highest mobility of the 14-nm cap sample is about 40,000 cm²/Vs, while that of the 80-nm cap sample is 380,000 cm²/Vs, almost 10 times higher than the thinnest sample. Given the surface condition of as-grown wafers, surface cleaning processes and aluminum oxide deposition were controlled in the same conditions, we concluded that the improvement in mobility curves are mainly due to the increase in SiGe cap thickness. Physically, a longer distance between the charged impurities and the silicon channel mitigates Coulomb forces from the interface to the 2DEG, lifting the mobility limitation upwards. Therefore, it is straightforward to think that an extremely high mobility is possible if we continue to increase the thickness of SiGe
cap and bury the silicon channel as deep as possible. Such thinking was substantiated by the recent low-temperature mobility record of Si/SiGe undoped 2DEGs (about 2,000,000 cm$^2$/Vs [78]), which was measured from a sample with a 500+ nm thick SiGe cap layer. However, we noticed that there seems to be an unknown factor that limits the highest low-temperature mobility in our samples at around 400,000 cm$^2$/Vs (Fig. 5.16), regardless of thicker SiGe cap layers (>80 nm). More experiments and trials to figure out the limitations will be discussed in the following sections.

We also note that the critical density gets lower with thicker SiGe cap layers (Fig. 5.17). The possible explanation is that the potential fluctuation in the conduction band minimum of the strained silicon channel, caused by Coulomb forces from remote charged impurities at interface, is reduced by increasing the distance between the interface and the channel. Thus the localization of electrons induced in the chan-
Figure 5.17: The highest mobility and lowest 2D electron density extracted from samples with various SiGe cap layers. The critical density gets lower in the sample with thicker SiGe cap thickness due to weaker potential fluctuations.

nel is less severe than the case with a thinner SiGe cap layer (or a shallower 2DEG), allowing a 2DEG to conduct at relatively low density regime.

5.6.2 Treatments on Oxide/Semiconductor Interface

The quality of the interface between oxide and silicon becomes crucial to electron transport properties of our undoped 2DEG samples because the most dominant mobility-limiting mechanisms, remote scattering, is caused by the density of interface charged impurities (Eq. 5.2). The interface between silicon and thermally grown SiO$_2$ is well known for its ultra-low interface charge densities. However, the deposition temperature for a high quality thermally grown SiO$_2$ can be very high (>1000 °C). Ge and Si may inter-diffuse across the Si/SiGe interface when the process temperature
is higher than 850 °C \[51\]. This inter-diffusion could deteriorate the abruptness of strained silicon quantum well and degrade the 2DEG qualities, through alloy scattering, for example. We thus used atomic layer deposition (ALD) to deposition aluminum oxide (ALO) at 300° for our gate insulators. Aluminum oxide deposited by ALD has been proven to have a good interface quality with bare bulk silicon \[79, 80, 81, 82\]. Great conformity and its low deposition temperature (<300°C) make aluminum oxide deposited by ALD a popular insulator on a Si/SiGe 2DEG.

In the literature \[83, 84\], the common operation temperature of thermal ALD for aluminum oxide (In contrast to plasma-enhanced ALD) varies from 150 °C to 300 °C. The ALD system in our university cleanroom is the Cambridge NanoTech Savannah 100, and its maximum deposition temperature is 300 °C. To examine the deposition temperature effects on the interface quality, and then the 2DEG quality, three temperatures (300 °C, 225 °C and 150 °C) were chosen for a 90-nm aluminum oxide deposited on three separate samples grown in the same growth run (Same SiGe cap thickness). After Hall-bar-shaped metal gate was evaporated on these three devices, Hall measurements were then conducted and mobility curves were plotted in Fig. 5.18 for comparison.

Here we repeat that 300 °C is our standard deposition temperature for most of the 2DEG samples shown in this thesis. The data from the sample with 225 °C ALO (red) showed pretty high electron mobility, but no obvious improvement in both mobility and critical density were observed. In fact, its mobility is even lower than the reference sample (black). Furthermore, a gate leakage even interrupted the measurement on the sample with 150 °C ALO, implying the worse quality of ALO by low temperature deposition in our ALD equipment.

Another common measure to improve interface quality is annealing. Four devices were made on the same sample and processed together until the completion of ALO deposition at 300 °C. Three of these devices were then separately annealed at 450
Figure 5.18: Mobility curves measured from samples with an aluminum oxide deposited at three different temperatures: 150 °C, 225 °C and 300 °C, 600 °C and 750 °C in a forming gas environment in RTA AG for 50-100 seconds. However, the following Hall measurements conveyed an inconclusive result (Fig. 5.19). Although annealing at 750 °C (blue) and 450 °C (pink) slightly lifted up respective mobility curves, annealing at 600 °C (green) somehow degraded its 2DEG quality where its mobility is much lower than the reference data (black). Moreover, a huge shift (a volt to couple of volts) in threshold voltages of annealed samples from that of the reference sample may infer the post-ALD annealing induces charge trapping (positive or negative) at the interface (Fig. 5.20). We note that both the minimum density and the threshold voltage do not correlate with high mobility. Thus our fundamental understanding of the interface quality effect on transport properties is not clear.
Remote Scattering from the Regrowth Interface

The above discussions were mainly focused on the influence of remote scattering from the oxide/silicon interface, but we ignored the bottom interface (regrowth interface) where our growth started. We mentioned in Ch. 2 that both our doped or undoped strained silicon 2DEGs were all grown on top of commercial relaxed SiGe buffers from Amberwave System Inc.. A considerable amount of charged impurities or defects may exist at this regrowth interface. Hence, the remote scattering from this interface may also be taken into account while we try to determine the hidden mobility-limiting factors. Because the work relevant to remote scattering from the regrowth interface depends on a newly-discovered surface screening effect, more details about the bottom interfaces will be discussed later in Ch. 7.
5.7 Effects of Growth Temperature

5.7.1 Effects of Growth Temperature

The growth temperature for the SiGe cap layer is usually constrained to lower than 550 °C to suppress the surface segregation of phosphorus atoms from the electron supply layer in a typical doped 2DEG \[40\]. Despite no intentionally doped layer, growth temperature for an undoped 2DEG is still kept as low as possible to avoid undesired strain relaxation in the strained silicon layers. So as shown in Ch. 2, the standard growth temperatures for our undoped 2DEGs are 575 °C for SiGe and 625 °C for Si. Nevertheless, it is interesting to see how much the growth temperature affects the stability of the strained silicon layer and of the Si/SiGe interfaces, and then how these variables subsequently affect the transport properties of 2DEGs.

Figure 5.20: The shifts in threshold voltages imply that charges are trapped at the interface after annealing.
Here we show a set of Hall measurement data from samples with different growth temperatures for their strained silicon cap and strained silicon quantum well (channel) in Fig. 5.21. For this batch of samples, the thickness of each layer was grown as consistent as possible to avoid any perturbation on mobility curves due to distinct layer structures: 3-4 nm for strained silicon cap, 65-80 nm for SiGe cap, 9-12 nm for strained silicon quantum well and 145-160 nm for SiGe buffer layer. The growth temperature for all relaxed SiGe layers was fixed at 575 °C, while the growth temperatures for strained silicon varied. The actual growth temperatures were labeled in the parentheses with the temperature for the silicon cap (\(T_{\text{Si,Cap}}\)) first, and that for the silicon quantum well or silicon channel (\(T_{\text{Si,QW}}\)) second. The results in Fig. 5.21 are quite inconclusive at first glance, but a sure thing is that a growth temperature for strained silicon layers lower than 625 °C is a necessity to achieve a good mobility curve (highest mobility >200,000 cm\(^2\)/Vs). Furthermore, we will show that the change in growth temperatures for strained silicon layers actually did influence both Si/SiGe interface roughness, interlayer mixing (at the top of the strained Si channel) and the density of threading dislocations.

### 5.7.2 Surface Roughness versus Interface Roughness

Not much literature has focused on Si/SiGe interface roughness scattering in both doped and undoped 2DEGs [85, 86]. The difficulty in the measurement of actual Si and SiGe interface roughness in-situ is one of reasons that has prevented relevant research. Another important reasons for the lack of data in literature is that there is no effective measure to vary the interface roughness in different 2DEGs without significant modification of layer structures. Ismail et al. [28] observed a drop in mobility for a sample with a strained silicon layer which is much thinner than the critical thickness. They attributed that drop in mobility to the roughness between Si and SiGe, but no further experimental data supported this argument. However, by
Figure 5.21: Mobility curves from samples whose strained silicon layers were grown at different temperatures.

means of changing growth temperature, it could become possible to vary the interface roughness without changing the layer structure too much, which may enable us to analyze the effect from interface roughness scattering solely without other scattering mechanisms mingling in.

Since we don’t have in-situ tools in our growth reactor for the roughness measurement, we measured the surface roughness on as-grown 2DEG samples whose strained silicon layers were grown at different temperatures. An AFM result was shown in Fig. 5.22 as an example, and its RMS surface roughness calculated by software, NanoScope Analysis, is 0.229 nm (left figure in Fig. 5.22). A reference sample was also grown in the same temperatures for both the strained silicon and the SiGe buffer layer as the sample shown in Fig. 5.22 but the growth was stopped at the strained silicon channel to expose its surface for roughness measurement. An even
Figure 5.22: An AFM image and its calculated RMS roughness on a complete undoped silicon 2DEG.

Figure 5.23: An AFM image and its calculated RMS roughness from a reference sample whose layer structure only consists of the relaxed SiGe buffer and the strained silicon QW (channel).

A smoother surface with RMS roughness ~0.180 nm was obtained from this reference sample (Fig. 5.23), which ensures us that the surface roughness we measured from a complete undoped 2DEG could be viewed as the worst case for its interface roughness under each temperature condition.
5.7.3 Effects of Si/SiGe Interface Roughness

We labeled the RMS roughness adjacent to the mobility curves from individual samples with unique growth temperatures (Fig. 5.24). The sample with both strained silicon layers grown at 700 °C (orange), around 100 °C higher than the standard growth temperature, has the roughest surface at an RMS of 0.543 nm. Its low mobility in the high density regime could be a hint that interface roughness scattering limits its mobility based on Eq. 5.3. In contrast with it, another sample with both strained silicon layers grown at 575 °C, the lowest temperature we tried here, shows an extremely smooth surface with roughness ~ 0.111 nm and high mobility in the high density regime. However, for the rest of the samples, the surface roughness data does not show a strong correlation with the growth temperatures. Nevertheless, an implication from this set of data is still useful: if high mobility is desired, the growth temperature should be at 625 °C or less.

The different growth temperatures for strained silicon layers lead to the variation of interface roughness in 2DEGs. Compared with growth temperatures, it is more interesting to see if there is any effect of interface roughness on both electron mobility and critical density. The highest mobility and lowest 2D density were then extracted from these samples and plotted with measured roughness from this sample set (Fig. 5.25). The lowest 2D density has no obvious correlation with the roughness, while the highest mobility has a subtle inverse proportion to the surface roughness data, as indicated by the red solid line. This observation basically agrees with the theoretical prediction as discussed in Sec. 5.2.3 and Sec. 5.4.1: Interface roughness scattering influences electron mobility more at high density regime and has negligible effect at low density regime.
5.7.4 Effects of Si/SiGe Interlayer Mixing

The SiGe cap layer after the Si channel was grown at a lower temperature (575 °C) than the silicon channel and the silicon cap. Thus the interlayer mixing, which may result in alloy scattering, could only occur during the silicon cap growth. We observed the sample with the silicon channel grown at 700 °C and the silicon cap grown at 575 °C has a poor transport property (green symbols in Fig. 5.24). However, the sample with the silicon channel grown at 575 °C and the silicon cap grown at 700 °C has a good mobility curve (blue symbols) just like other samples with cap layers grown at lower temperatures. Therefore, we can rule out the interlayer mixing and possible alloy scattering in our undoped 2DEGs.
5.8 Effects of Threading Dislocations

Another possible mobility-limiting mechanism hidden in a Si 2DEG is scattering from threading dislocations. Unlike GaAs/AlGaAs material system, Si and SiGe have severer lattice mismatch (4%), which results in inherently inevitable defects existing in a Si/SiGe heterostructure. For example, a threading dislocation could occur at 60° to the growth surface in a relaxed SiGe layer during the epitaxy, and slide along the interface between the relaxed SiGe and the strained silicon layer on the top of it to relieve accumulated strain (misfit dislocation) [73, 87, 88]. Therefore, the density of threading dislocation and the length of misfit dislocation across the strained silicon channel may bring in an adverse effect on the electron transport properties of silicon 2DEGs.
5.8.1 Defect Etching

Preferential etching is a common technique used to reveal defects in a silicon-based material system. The difference in etching rate between defects and other areas leads to etch pits that could be observed at defect locations. By counting the number of black etch pits in a microscopic image, an etch pit density could be estimated, which is usually referred to the density of threading dislocations in a Si/SiGe heterostructure.

Chromium-ion-containing etching solutions, such as Secco [89] and Schimmel [90], are the most popular etching solutions used for defect revelations. Secco consists of K$_2$Cr$_2$O$_7$ and HF, while Schimmel is a mixture of CrO$_3$ and HF, both diluted in DI water. The chromium-ion-containing acid is responsible for oxidizing surface silicon or SiGe, and HF etches the resulting oxides immediately. The etching rate depends on the concentration of etching solutions. For instance, more diluted solutions have slower etching rate through silicon or SiGe layers, making it easier to control the desired etch depth. To observe threading dislocations produced in the layers grown in our system, the target etching depth could not be thicker than the total thickness of layers on top of regrowth interface, which is typical - 250 nm. However, both Secco and Schimmel were designed to etch bulk silicon, not thin epitaxial layers. In spite of being diluted by DI water, the slowest etching rate that is controllable is still as fast as 1 µm per minute. A modified Schimmel solution [91] was proposed to aim at the defect delineation in thin SiGe epitaxial layers whose thickness ranges from tens to hundreds of nanometers. This modified etching solution consists of 55 %vol CrO$_3$ (0.4M) and 45 %vol HF (49%), and the solution is cooled down to + 2 °C (Ice bath). The proposed etching rate depends on the Ge fraction, but basically falls in the range of 5-15 nm/sec. The etching rate for the sample with 30% Ge fraction proposed by this paper [91], 8 nm/sec, was also verified in our lab. This etching rate is slow enough for us to control our etching depth within 250 nm easily.
We dipped our 2DEG samples grown with different temperatures into modified Schimmel solutions for 25 seconds. The total thickness etched away is about 200 nm, and the etching stopped in the middle of the grown SiGe buffer layers as desired. Etched samples were then examined under Nomarski microscope (differential interference contrast) and images were taken for the following etch pits counting. The density of etch pits observed by eye could vary a lot. For example, the image shown in Fig. 5.26a was from the sample whose surface silicon and buried silicon channel were grown at 575 °C and 700 °C respectively. The dot-like threading dislocations are randomly distributed all over the image with line-shaped misfit dislocations. On the contrary, the image shown in Fig. 5.26b from the sample, whose silicon layers were grown at 625 °C, shows much fewer etch pits, and no line-shaped misfit dislocations were spotted.

5.8.2 Effects of Threading Dislocations on Transport Properties

The etch pit density (EPD) for each sample was statistically counted from several randomly chosen 200 μm² windows and plotted with its corresponding growth temperature with an error bar (Fig. 5.27). Note that we only counted etch pits for the estimation of threading dislocation density. The misfit dislocations observed in Fig. 5.26a were not included in the discussion. Samples that were involved in 700 °C growth, no matter whether for the surface silicon or for the buried silicon quantum well, clearly showed 4-5 times higher EPDs (7-10×10⁴ cm⁻²) than samples grown at lower temperature (2×10⁴ cm⁻²). This strongly implies that high temperature growth induces more threading dislocation defects propagating upwards through the whole grown layers.

An interesting question may arise when these relatively high densities of threading dislocations in some samples were observed: Do those threading dislocations affect
Figure 5.26: Images acquired by Normaski Microscopy after samples were dipped into modified Schimmel solutions. (a) Plenty of dot-like threading dislocations are distributed all over the sample whose $T_{\text{Si, Cap}}$ and $T_{\text{Si, QW}}$ were 575 °C and 700 °C respectively. (b) Much fewer etch pits are seen in the sample whose silicon layers were both grown at 625 °C.
transport property of an undoped 2DEG? To answer it, the highest mobility and the critical density for each sample were extracted again, and plotted versus its corresponding EPD in Fig. 5.28. Surprisingly, the highest mobility is basically independent of the EPD, even if some EPDs are almost 4 times higher than others. To seek possible explanations, we need to go back to review our SiGe relaxed graded buffers. As we mentioned in Ch. 2, thanks to the invention of SiGe graded buffer, the density of threading dislocation has been greatly improved. The state-of-the-art SiGe buffer layer with a CMP treatment has a much lower threading dislocations density (1-2×10^4 cm^-2) than old Si/SiGe heterostructures (10^6 cm^-2). Based on this, the EPD observed in our samples were actually pretty low in terms of the order of magnitudes, regardless of growth temperature. Besides, the mobilities we have been talking about were measured in the regime where the electron density is 10^{10}-10^{12} cm^-2. Compared
with the low number of EPD, mobile electrons with such high density shouldn’t be affected by threading dislocations. Nevertheless, a suspicious correlation between EPD and critical density was highlighted by a blue dotted curve: A higher critical density was observed in the sample with higher EPD. This subtle correlation could be the results of severe potential fluctuations induced by threading dislocations, where charged impurities tend to gather. However, insufficient data points do not allow us to believe this correlation is convincing yet. More experiments are necessary to support this argument.

Figure 5.28: The highest mobility and lowest 2D density versus etch pit density from five different samples.
5.8.3 Effects of Hall Bar Size

The density of threading dislocation could be too low to affect electron mobility significantly at the electron density in the range of $10^{10}$-$10^{12}$ cm$^{-2}$. However, the effect from misfit dislocations or other invisible imperfections in SiGe alloys has not been explored yet. A rough estimation from the image shown in Fig. 5.26a gives us an idea that each misfit dislocation are about 50-150 $\mu$m long, and distances between parallel misfits are around 20-100 $\mu$m. Imaging a case that a long line-shaped misfit dislocation is located coincidentally across the middle of a Hall bar device. This misfit dislocation may hinder electrons from moving in the channel, and thus severely deteriorate the transport property. To decrease the opportunity that a misfit cuts a Hall bar device, scaling down the size of a Hall bar device may be a good choice.

The geometry of a standard Hall bar used in this study was described in Fig. 5.5. The distance between two $R_{xx}$ probes is 300 $\mu$m (Thus denoted as 300-$\mu$m Hall bar), and the width of the main conduction path is 100 $\mu$m (Fig. 5.29a). If compared with the dimension of misfit dislocations revealed by wet etching, it is inevitable that some misfit dislocations must overlap with our Hall bar device. In order to alleviate any scattering from the misfit dislocation (or other invisible crystal imperfection in SiGe alloys), much smaller Hall bar devices were also fabricated with a standard Hall bar device in the sample from the same growth run. The geometry of small Hall bar devices is illustrated in Fig. 5.29b. The distance between two $R_{xx}$ probes is now reduced to 10 $\mu$m (denoted as 10-$\mu$m Hall bar), which is 30 times narrower than that in the standard one. The widths of the $R_{xx}$ probes and main conduction path are narrowed 3.33 times and 10 times compared to before, down to 3 $\mu$m and 10 $\mu$m respectively. The total areas enclosed by two outer $R_{xx}$ probes of the 10-$\mu$m Hall bar is only about 220 times smaller than that of the 300-$\mu$m Hall bar.

Since dislocations are randomly distributed all over the samples, more 10-$\mu$m Hall bar devices give us a higher chance to hit the sweet spot. The Hall measurement
Figure 5.29: The geometries of (a) a standard Hall bar with a 300-µm spacing between $R_{xx}$ probes (denoted as a 300-µm Hall bar) and (b) a small Hall bar with a 10-µm spacing between $R_{xx}$ probes (denoted as a 10-µm Hall bar).

data from two small Hall bar devices (green and blue) were shown in Fig. 5.30 with the data from a standard Hall bar (red) for comparison. Unfortunately, we didn’t see any improvement in mobility curve from the miniaturization of Hall bar devices. Instead, the transport properties were degraded. The possible reason for the mobility degradation might be the non-uniformly distributed remote scattering sites at the interface. When the Hall bar device is large, the effect from the non-uniform potential fluctuation could be averaged out. However, once we shrink the size of the Hall bar, the small but dense gatherings of remote scattering sites which are coincidentally formed on top of the Hall bar may severely disturb the electron conduction through the device. Here we note that small Hall bar devices were made on several different
Figure 5.30: The comparison between mobility measured from a standard Hall bar and two small Hall bars fabricated in the sample from the same growth run.

samples repeatedly, but none of them showed the signs of improvement in transport properties.

5.9 Summary

Different kinds of scattering mechanisms possibly limiting electron mobility of an enhancement-mode strained silicon 2DEG have been individually discussed in this chapter. We first greatly alleviated background scattering by means of a significant reduction in background phosphorus concentration enabled by a brand new gas panel of our CVD system. A 1000 times lower phosphorus background concentration ($<10^{14}$ cm$^{-3}$) showed background scattering is no longer significant. Next, the remote scattering is considered as the dominant mobility-limiting factor for 2DEG samples.
with relatively thin SiGe caps (<80 nm). An improvement in transport property of a 2DEG with a thicker SiGe cap has reassured its importance in both mobility and critical density. However, a saturation in mobility curve observed in samples with SiGe caps thicker than 80 nm implies other hidden mobility-limiting factors in our system. Furthermore, the relationship between minimum density, mobility and positive threshold voltages shifts (trapped electrons) is not clear. In addition, interface roughness scattering along with threading dislocation scattering have been systematically examined with a variation in growth temperatures of strained silicon layers. Unfortunately, none of them explained the mobility saturation. They didn't show a direct and clear influence on either mobility or critical density. Other possible mobility-limiting factors could still exist in our 2DEG system and more experiments are necessary to discover them [92].
Chapter 6

Tunable Screening Effect in Undoped 2DEGs

6.1 Introduction

6.1.1 Motivation

Efforts to enhance electron mobility in an undoped 2DEG were discussed from various aspects in Ch. 5. We also mentioned the possible causes that limit the critical density, the 2D electron density at which the metal-insulator transition occurs. Among various mobility-limiting mechanisms, the scattering from remote charged impurities located at the oxide/silicon interface (remote scattering) is viewed as the most likely mechanism dominating in our system. The density of the remote charged impurities is also predicted to affect the critical 2D electron density in the MIT model mentioned in Sec. 5.4. Intuitive thinking concludes that a thick SiGe cap layer to separate the 2DEG away from the interface can lead to high electron mobility ($\mu$) and low critical 2D electron density ($n_{2D}$). Many experimental results also strongly agree this argument [68 76].
However, to pattern an undoped 2DEG into a QD, a thin SiGe cap layer is preferred to enable patterned top gates to precisely define the lateral extension of the 2DEG, even though a thin cap degrades the 2DEG transport properties. In this chapter, we present an improvement in the 2DEG properties (higher mobility and lower critical $n_{2D}$) in samples with thin SiGe cap layers ($<40$ nm) by introducing a tunable shielding electron layer near the surface. We believe that, at a critical electrical field, the tunneling of electrons from the buried silicon QW to the surface triggers the formation of a barely mobile electron layer near the silicon surface. This surface electron layer effectively screens the remote charge scattering sites, and thus dramatically improves both mobility and critical $n_{2D}$. As part of this work, we introduce the concept of equilibrium versus non-equilibrium in 2DEG densities as well. This work is summarized in [93].

6.1.2 Device Fabrication

The enhancement-mode strained silicon 2DEGs have a similar structure and process flow to that described in Ch. 5. The undoped Si/SiGe heterostructures in this study were grown by rapid thermal chemical vapor deposition (RTCVD) on top of relaxed Si$_{0.72}$Ge$_{0.28}$ virtual substrates. After growing another Si$_{0.72}$Ge$_{0.28}$ relaxed buffer layer (90 to 165 nm) on the top of this starting virtual substrate, an 11-nm strained silicon quantum well (denoted as the buried QW in this chapter for clarity) was then grown to hold the 2DEG. Subsequently, a thin undoped Si$_{0.72}$Ge$_{0.28}$ cap layer (14, 20, or 40 nm) was grown, followed by a 4-nm strained silicon cap layer (denoted as the surface QW) growth (5-nm for the 14-nm cap sample). The actual layer thicknesses may differ by $\pm$ 20%. To contact the buried 2DEG, phosphorus was first implanted in contact regions, followed by annealing at 600 °C. A 90-nm aluminum oxide layer was then deposited by atomic layer deposition as a gate insulator. A chrome/gold stack
was finally evaporated on samples to form both a Hall-bar-shaped gate and metal contacts on the implanted regions.

6.2 Four-Stage Behavior of 2D Electron Density

6.2.1 Observation of a Density Collapse

The Hall measurements with the same setup as before were performed at liquid helium temperature (4.2 K). Four clear stages of the Hall electron density ($n_{2D,\text{Hall}}$) were observed in all samples as the gate voltage was ramped up. Fig. 6.1 shows data for the 14-nm SiGe cap sample as a characteristic example. When the gate voltage is above zero but below a threshold voltage ($V_T$), which is $\sim 2.9$ V, the $n_{2D}$ which is ideally induced in the buried QW is low, leading to insulating behavior (referred to as stage I) due to disorder and potential fluctuations, primarily from remote charged impurities at the oxide/silicon interface [60, 61, 77, 94]. Once the gate voltage supports an electron density above the critical density for the metal-insulator transition (MIT) [95], $3.5 \times 10^{11}$ cm$^{-2}$ in this sample, electrons start to flow from the contacts into the buried QW to form a 2DEG (stage II). The experimental capacitance extracted from the linear dependence of the Hall electron density on gate voltages from 2.9 to 3.5 V in stage II is close to (-90\%) the expected value based on a parallel-plate capacitor model between the 2DEG and the gate. (By $n_{2D,\text{Hall}}$ and $\mu_{\text{Hall}}$ in this chapter, we mean those extracted from the measurements assuming a single transport layer, in this case the buried silicon quantum well.)

With further increase in the positive gate bias, a sharp collapse of $n_{2D,\text{Hall}}$ was clearly observed in all samples when $n_{2D,\text{Hall}}$ reached $-6.0 \times 10^{11}$ cm$^{-2}$, dropping to $2.2 \times 10^{11}$ cm$^{-2}$, much lower than the density originally required to initiate conduction. This new range is referred as stage III. With a 3-4 V further increase of gate voltage, $n_{2D,\text{Hall}}$ then increased only marginally ($<0.3 \times 10^{11}$ cm$^{-2}$), while a simple
Figure 6.1: The four-stage behavior in Hall electron density observed in all three samples (Data here are from the sample with a 14-nm SiGe cap layer). The dashed line shows the theoretical maximum $n_{2D}$ from the self-consistent Schrodinger-Poisson (SCSP) simulation.

C·$\Delta V_G$ calculation would predict an increase of $n_{2D,Hall}$ of $-10^{12}$ cm$^{-2}$. In addition to the reduction in minimum $n_{2D}$ in the sharp transition from stage II to stage III, the electron mobility is also dramatically enhanced (Fig. 6.6). Explaining these effects is the focus of this chapter; we hypothesize the effects are due to the formation of a tunable shielding electron layer at the semiconductor surface which screens the buried 2DEG from the scattering from remote charged impurities at the oxide/silicon interface. Here we note that all data from Hall measurements shown in Ch. 5 are in stage II before the density collapses for clarification.
6.2.2 Non-Equilibrium in Stage II

Since the buried QW (11 nm) is thicker than the surface QW (4-5 nm) in our structures, at flatband the ground state ($E_0$) of surface silicon lies higher than the one in the buried QW. Therefore, as the gate voltage increases, $E_0$ of the buried QW drops to the Fermi level ($E_F$), defined by the contacts, before that for the surface layer, leading to the population of a buried 2DEG (Fig. 6.2a). As the gate voltage increases to induce higher density of mobile electrons in the buried QW, eventually $E_0$ of the surface QW will fall below $E_F$, so that electrons in the surface QW would be expected. With this assumption of thermal equilibrium with the contacts (both densities represented by a single Fermi level), once the surface electron layer forms (blue solid line in Fig. 6.3), a further increase in the gate voltage will lead to an increase only in the surface electron density ($n_{\text{surface, Eq}}$) and the electron density in the buried QW ($n_{\text{buried, Eq}}$) will remain fixed to first order (red solid line), because surface electrons will screen out the electrical field from the gate. However, the close proximity of many scattering charges at the oxide/silicon interface leads to a high critical density for the MIT of the surface layer. Thus when $n_{\text{surface, Eq}}$ would be expected to be at a low value, it is impossible for electrons to flow laterally from the contacts into the surface QW. Furthermore, the low vertical electric field prevents electrons from tunneling from the buried QW to the surface layer. Thus a surface layer cannot form, and the surface layer is not in thermal equilibrium with the contacts, with its ground state $E_0$ substantially below the contact Fermi level (Fig. 6.2b) [48]. Therefore, as the gate voltage is raised, more electrons continue to accumulate in the buried QW, with the system continuing to exhibit stage II behavior (dotted lines in Fig. 6.3) in a non-equilibrium condition. The situation is in some sense analogous to deep depletion in a MOS capacitor. In that case, also an inversion layer should be expected based on equilibrium principles, but there is no mechanism to create one.
6.2.3 Switching from Stage II to Stage III

We propose the sharp collapse in $n_{2D,Hall}$ with a further increase in gate voltage is triggered by electron tunneling, which initiates a positive feedback process (Fig. 6.4a and b). At an electron density of $6 \times 10^{11}$ cm$^{-2}$ in all samples, corresponding to a critical electric field of $10^5$ V/cm if spurious charges are ignored, electrons begin to significantly tunnel through the thin SiGe cap layer into the surface. The surface density then reaches a point where some slow conduction laterally from the contacts...
Figure 6.3: The comparison of buried electron density ($n_{\text{buried}}$, red) and surface electron density ($n_{\text{surface}}$, blue) with increasing gate voltages in both thermal equilibrium with the contacts between two 2DEGs and with the surface channel not in equilibrium. A sudden collapse in $n_{\text{buried}}$ as the gate voltage increases and the corresponding increase in $n_{\text{surface}}$ bring the system back to thermal equilibrium.

As the surface layer forms at a fixed gate voltage, electrons must also flow out of the buried QW to obey Gauss's law (Fig. 6.5). The simultaneous increase in $n_{\text{surface(non-Eq)}}$ (blue dashed line in Fig. 6.3) and decrease in $n_{\text{buried(non-Eq)}}$ (red dashed line in Fig. 6.3) brings the whole system back to thermal equilibrium (stage III).

Note that the time scale for the density collapse, namely the time scale for electrons to flow into the surface layer, could be on the order of five minutes - this is the approximate time between Hall measurements at each gate voltage. Beyond this point, in equilibrium, with more gate voltage we expect an increase mostly in the surface density. Furthermore, if the surface mobility (and thus conductivity) were several orders of magnitude below that of the buried layer, which we believe to initially
Figure 6.4: (a) no electrons are induced in the surface due to a high critical density of MIT for the surface QW. (b) At a higher gate voltage, electron tunneling from the buried QW towards the surface raises the density above the metal insulator transition point, leading to a current flowing from the contacts into the surface layer (c). (d) By Gauss’s law, the buried electron density must be reduced as the surface density increases at a fixed gate voltage. The system is then switched from non-equilibrium to thermal equilibrium.
be the case, a single-layer interpretation of the Hall measurements ($n_{2D,Hall}$ and $\mu_{Hall}$) would continue to represent the properties of the buried layer.

### 6.2.4 Support by a Simulation

To build confidence in our model, a self-consistent Schrodinger-Poisson (SCSP) simulation [96] was utilized to calculate the theoretical maximum $n_{2D}$ in the buried quantum well in thermal equilibrium, which is the constant value that the red solid line in Fig. [6.3] represents at high gate voltage. For samples with 14-nm, 20-nm and 40-nm SiGe caps, these values are $2.7 \times 10^{11}$ cm$^{-2}$, $2.9 \times 10^{11}$ cm$^{-2}$, and $1.9 \times 10^{11}$ cm$^{-2}$ respectively. The $n_{2D,Hall}$ values (representing the buried layer) measured near the end of stage II ($-6.0 \times 10^{11}$ cm$^{-2}$) were much higher than these values, implying the surface layer was indeed not in equilibrium at the end of stage II when the collapse occurs. Furthermore, note the experimental values just after the stage II/stage III (equilibrium/non-equilibrium) transition were $2.5 \times 10^{11}$ cm$^{-2}$, $2.4 \times 10^{11}$ cm$^{-2}$ and $1.7 \times 10^{11}$ cm$^{-2}$ for these three samples, respectively, all in reasonable agreement with the predictions (Fig. [6.6]). Both results support our model that the stage II/III col-
lapse is a switch of the surface layer from non-equilibrium to equilibrium. More details about SCSP simulation in undoped 2DEGs will be discussed in Sec. 7.2.

6.3 Tunable Screening Effect in Thin-Cap 2DEGs

6.3.1 Improved Transport Property

We now discuss the transport properties, and show the dependence of the Hall mobility on the Hall electron density (Fig. 6.6). For each sample two sets of points are shown: closed symbols before the transition and open symbols after it. With no surface layer in stage II, the mobility of each sample increases with density due to the usual self-screening. Because the mobility at a given density increased (and the minimum density decreased) as the separation between the semiconductor/insulator interface and the buried 2DEG increased, it seems clear that the main scattering sites are at the surface (or inside the insulator) [67]. When the system switches back to thermal equilibrium, the new intermediate electron layer near the surface separates the buried 2DEG and the scattering sites, resulting in a strong screening effect on both the minimum n_{2D} and electron mobility of the buried layer. In all samples, after the transition, the samples now conduct well at densities only 60-70% of their previous minimum densities (Fig. 6.7a). Note the small range of Hall electron densities in stage III despite an increase of gate voltage of several volts; this is because new charges go mostly into the surface layer and not the buried layer as expected from Fig. 6.3.

Beyond the density reduction, the screening effect enhances the electron mobility of the buried layer as well (Fig. 6.6). In stage II, the highest electron mobility obtained from samples with 14-nm, 20-nm and 40-nm SiGe caps are 47,000 cm^2/Vs, 153,000 cm^2/Vs and 381,000 cm^2/Vs at high densities (-6.0×10^{11} cm^2), respectively, with the 20-nm cap sample requiring a density of -5.0×10^{11} cm^2 to reach a mobility of
Figure 6.6: The dependence of Hall mobility on Hall electron density measured at 4.2 K for all three samples with different stages labeled. The gate voltage steps for data points at stage II and stage III/IV are 0.03-0.15 V/0.1-2 V, respectively for all three samples. The measurement sequence is indicated by dashed lines (from stage II to stage III).

100,000 cm$^2$/Vs. After the transition both the 14-nm and 20-nm cap samples achieve a mobility at or near 100,000 cm$^2$/Vs at a density of only $2.3 \times 10^{11}$ cm$^{-2}$, and the 40-nm cap sample achieves this benchmark at a density of only $1.6 \times 10^{11}$ cm$^{-2}$. The 20-nm cap sample reaches 196,000 cm$^2$/Vs at only $2.6 \times 10^{11}$ cm$^{-2}$. These densities are well below the metal-insulator transition level for each of the three samples before the transition.

6.3.2 Effect of Surface Electron Density

To emphasize the importance of the surface electron layer on the mobility enhancement in stage III, the relation between $n_{\text{surface}}$ and Hall mobility for all three samples
are shown in Fig. 6.7b. Here, the tunable \( n_{\text{surface}} \) was calculated by starting with the change in electron density in the buried layer at the collapse voltage (adjusted for the slightly different gate capacitances from the gate to the buried versus the surface layers), and then scaling the value up by \( C_{\text{ox}} \Delta V_G \) from that point. In stage III, the increase in gate voltage (2-3 V) results in a considerable increase in \( n_{\text{surface}} \) from around \( 4.0 \times 10^{11} \text{ cm}^{-2} \) to over \( 1.0 \times 10^{12} \text{ cm}^{-2} \), but only a marginal increase in \( n_{\text{buried}} \) (\(< 1.0 \times 10^{11} \text{ cm}^{-2} \)). With the increases in \( n_{\text{surface}} \), we do see the significant increase in Hall mobility (Fig. 6.7b). Here we note that we did look for a theory with the tunable shielding layer to fit, but so far we can not find one. Nevertheless, this result increases our confidence in attributing the enhancement in buried layer mobility to the strong screening by the surface electron layer.

### 6.3.3 Parallel Conduction in Stage IV

If we keep ramping up the bias in the equilibrium mode, a decrease in Hall mobility is eventually seen in all samples (stage IV in Fig. 6.6). Here we set the highest mobility as the boundary of stage III and IV. We calculated the surface electron mobility for
Figure 6.8: Surface electron mobility (calculated based on the parallel conduction model) versus the surface electron density (calculated in Sec. 6.3.2). The right Y axis shows the conductance ratio of the surface QW to the buried QW.

both stage III and stage IV (Take the 14-nm SiGe cap sample as an example again) based on the parallel conduction model described in [32], given $n_{\text{buried}}$ is fixed at the lowest value after the density collapses. The relationship between surface electron mobility and $n_{\text{surface}}$ (calculated in Sec. 6.3.2) is then shown in Fig. 6.8 with the corresponding conductance ratio of the surface QW to the buried QW on the right Y axis. First, the increasing surface electron density leads to higher surface electron mobility in stages III as we expected. Similar to Hall electron mobility, surface electron mobility saturates with increasing $n_{\text{surface}}$ probably because of the interface roughness scattering. Second, due to the increases in both mobility and density, eventually the conductance in the surface layer becomes significant compared to that in the buried electron layer ($<1\%$ at the beginning of stage III to $>10\%$ in stage IV). Therefore, the Hall mobility (based on the total lateral conduction) is decreased as Hall electron density is increased [32].
6.3.4 Magneto-Resistance Transport Property at 0.3 K

In the previous discussions, the fundamental assumption we made is that the conduction of electrons in the surface QW is several orders of magnitude lower than that of electrons in the buried QW because electrons in the surface QW are too close to the unfavorable charged impurities at the interface. Hence, we believe the enhanced transport properties shown in stage III still represents the properties of the buried QW, even though measured $1/R_{xx}$ and $1/R_{xy}$ are the equivalent parallel conductivity, not the conductivity from a single layer. However, direct experimental evidence to show where the 2D electron layer is exactly located is still preferable. Therefore, several magneto-resistance transport measurements for samples shown in Fig. 6.6 were conducted under 0.3 K by Hao Deng in Prof. Shayegan’s group. In contrast to Hall measurements at 4.2 K, magneto-resistance transport measurements at 0.3 K were performed with much larger increment in gate biases due to the long scanning time to increase magnetic field up to 15 T. The data from two representative gate biases, 4 V in stage II and 9 V in stage III from the sample with a 20-nm SiGe cap, are highlighted in Fig. 6.9 in blue.

Clear Shubnikov-de Haas (SdH) oscillations in $R_{xx}$ and quantum Hall plateaus in $R_{xy}$ were observed at both 4 V and 9 V (Fig. 6.10). In stage II, the well-resolved and flat quantum Hall plateaus imply a good quality of the 2DEG. The SdH oscillation that touches zero several times and the single frequency extracted from a Fourier transform on $R_{xx}$ at low field show a fact that no parallel conduction occurs at 4 V. The electron density extracted from the SdH oscillation at low field is $3.87 \times 10^{11}$ cm$^{-2}$. (Note that the saturation of the SdH oscillation peak at 12 T in Fig. 6.10a is an artifact.)

When the sample was biased in stage III, namely when the gate voltage was fixed at 9 V, a more ideal SdH oscillation and flatter quantum Hall plateaus at a lower density $2.92 \times 10^{11}$ cm$^{-2}$ were observed (Fig. 6.10b). More filling factors with high
orders were resolved at low magnetic field in both $R_{xx}$ and $R_{xy}$. No overshoot in the left end of each plateau was observed as in the plateaus at 4 V. Furthermore, a fractional quantum Hall effect was identified at $\nu=4/3$ between two large plateaus for $\nu=1$ and $\nu=2$. The above are all the signs that the 2DEG system at stage III becomes cleaner (because less influenced by scattering sites), most likely due to the screening effect we mentioned in previous sections. In addition, we also made a plot with $R_{xx}$ v.s. $1/B$ from the sample with a 20-nm SiGe cap at low field (0.5 T-2.5 T) at 9 V (Fig. 6.11a). The frequency peaks with ratio of 1, 2 and 4 in the the Fourier transform on $R_{xx}$ v.s. $1/B$ are observed (Fig. 6.11b).
Figure 6.10: The magneto-resistance transport measurements at 0.3 K were shown versus magnetic field up to 15 T at (a) 4 V (stage II) and (b) 9 V (stage III).
Figure 6.11: (a) $R_{xx}$ versus the inverse of magnetic field from the sample with a 20-nm SiGe cap at 9 V in the range of 0.5 T to 2.5 T. (b) The Fourier transform of the curve in (a) shows three peaks. The ratio of frequencies show the different degeneracies, but they all represent the same conducting channel and a single electron density of $2.92 \times 10^{11}$ cm$^{-2}$.

The peak at 3.02 T, 6.2 T and 12.4 T show 4 (2 spin degeneracies and 2 valley degeneracies), 2 (2 valley degeneracies), and 1 degeneracy (all degeneracies are split) at such magnetic fields, and a single electron density can be calculated by $(e/h) \times$ degeneracy $\times$ frequency, where $e$ is the elementary charge and $h$ is the Planck constant \[^{[32]}\). Therefore, all three peaks verify that only single significant lateral conducting path exists at stage III. In other words, our main assumption that Hall measurement data validly interprets transport property of the buried QW at stage III is proven correct. The best ambiguity in our theory can be finally ruled out.

### 6.3.5 Negligible Screening Effect in Thick-Cap 2DEGs

The exciting screening effect greatly improves the transport property of undoped 2DEGs with thin SiGe caps (<40 nm). In fact, an identical four stage behavior in density was also experimentally observed in 2DEGs with thicker SiGe caps (Fig. 6.12).
Figure 6.12: An identical four-stage behavior in density was also observed in samples with thicker SiGe caps (90 nm in this case).

However, the transport data from them were different from the thin cap samples: the significant improvement in mobility and critical density was not seen for these samples. As an example, mobility data at stage II (closed symbols) from a 2DEG with a 90-nm SiGe cap were plotted together with the data from stage III and IV (open symbols) in Fig. 6.13. In stage III, the increase in mobility with increasing density basically follows the data in stage II until it steps into the stage IV where the parallel conduction bends down the effective Hall mobility. An implicit message from the overlap between data in stage II and stage III is that the mobility-limiting mechanisms in undoped 2DEGs with thicker SiGe cap layers may not be remote scattering at the top surface anymore. As a result, the screening of remote charged impurities at the interface by the newly-formed surface electron layer is not effective to enhance electron mobility.
That surface potential fluctuations are not important in samples with thicker SiGe caps (big separation between the 2DEG and the surface) is in the end not surprising. However, the cap thicknesses of interests for quantum computing and other patterning 2DEG applications is under 60 nm, where our discovered tunable screening effect may play an important role.

6.4 Reverse $V_G$ Scanning in Stage III

The unique four-stage behavior of densities in an enhancement-mode undoped 2DEG with thin SiGe cap ($<40$ nm) were analyzed in detail in Sec. 6.2. With the gate voltage ramping up, electrons start to accumulate in the buried QW until a critical electrical field is built up to trigger the whole positive feedback process. The density
collapses and then remains constant (to first order) even if the gate voltage increases
couple volts when the system is in thermal equilibrium (stage III).

An interesting trial to reverse the gate voltage scanning direction has also been
done in these 2DEGs with thin SiGe caps after reaching a high gate voltage in stage
III. Here we take the 14-nm SiGe cap sample shown in Fig. 6.1 as an example again.
When we scan the gate voltage down, \( n_{2D} \) basically remains the same level as we scan
up until the gate voltage approaches the boundary between stage II and III. Near
this boundary, a slight decrease in \( n_{2D} \) can be seen before the buried 2DEG became
insulating again. We note that densities with the reverse bias scanning are not shown
in this figure, but they are basically the same as those with the forward bias scanning
in stage III.

A similar tendency under reverse bias scanning in stage III was observed in sam-
ples with thicker SiGe caps as well. The density data for both scanning directions
from a 2DEG sample with a 75-nm SiGe cap are shown in Fig. 6.14 as an example.
However, the extension of measurements under the gate voltages that are supposed
to be classified into stage II was observed. Furthermore, when the gate voltage ramps
down and passes the boundary of two stages, a linear decrease in \( n_{2D} \) occurs. The
similar slopes extracted from this linear decrease at stage III (scan down) and the
linear increase in \( n_{2D} \) at previous stage II (scan up) imply that we modulated the
2DEG in the buried QW in both scans.

The transport properties measured from this sample both under a regular voltage
ramp-up at stage II (solid symbols) and a ramp-down at stage III (open symbols)
are shown in Fig. 6.15. When ramping down, \( n_{\text{surface}} \) decreases first and \( n_{\text{buried}} \) is
constant, reversing the path in Fig. 6.3 (solid lines). Then, the surface layer becomes
insulating, but with a higher density of electrons than the near zero amount when
ramping up in stage II. Beyond this point, \( n_{\text{buried}} \) starts to decrease. However, there
is now a screening layer trapped at the surface, so we expect improved transport
properties. The lowest $n_{2D}$ under a ramp-down at stage III ($-8 \times 10^{10}$ cm$^{-2}$) is about two times lower than that under a ramp-up at stage II ($-1.6 \times 10^{11}$ cm$^{-2}$).

For unknown reasons, the mobility is not improved during the downwards scan. This is not understood, but suggests again that different limiting factors may control minimum $n_{2D}$ and mobility at high densities.
Figure 6.15: The transport property of a typical undoped 2DEG with a thick SiGe cap under a regular voltage ramp-up at stage II (solid symbols) and a voltage ramp-down at stage III (open symbols).

6.5 Summary

We have discovered a new physical effect: the strong screening of the remote charge scattering sites from the oxide/semiconductor interface in enhancement-mode undoped Si 2DEGs, by introducing a tunable shielding electron layer at the semiconductor surface. When a high density of electrons in the buried silicon quantum well exists, the tunneling of electrons from the buried layer to the surface quantum well can lead to the formation of a nearly immobile surface electron layer. The existence of this surface electron layer and the validity of Hall measurements under this circumstance have both been verified by magneto-resistance measurements at 0.3 K. The screening of the remote charges at the interface by this newly formed surface electron layer results in an increase in the mobility of the buried 2DEG. Furthermore,
a significant decrease in the minimum mobile electron density of the 2DEG occurs as well. Together, these effects can reduce the increased detrimental effect of interface charges as the cap distance for the 2DEG to the surface is reduced for improved lateral confinement by top gates. Finally, we have also showed the density data under both upwards and downwards bias scans. A $2 \times$ lower minimum $n_{2D}$ was observed in stage III as the gate bias ramped down than that in stage II as the gate bias ramped up, while the mobility was not improved at the same time. This again concludes that the limiting mechanisms for minimum $n_{2D}$ and mobility at high densities could be different.
Chapter 7

The Role of the Regrowth Interface on Undoped 2DEG Properties

7.1 Introduction

Because our experiments in Ch. 5 indicated possible remote scattering sites at the bottom interface (regrowth interface) in addition to that at the oxide/silicon interface, we began modeling the 2DEG densities for different pinning positions of the Fermi level at the substrate growth interface. The pinning of the Fermi level could be a reflection of a high number of defect states. Then, experimentally we varied the SiGe buffer thickness, which is the separation from the silicon channel to the regrowth interface, to examine its effect on 2DEG transport properties. A 2DEG with thick SiGe buffers (145 nm or greater) shows higher electron mobility and lower critical densities than that with thin SiGe buffers (70 nm or less). In addition, we also varied the baking condition of the regrowth interface before the epi-layer growth. The electron mobility is independent of the baking temperatures (or baking lamp powers), but a lower baking power reduces the critical density. The electron density of $3.2 \times 10^{10}$ cm$^{-2}$ measured from the sample with a 75-nm SiGe cap in stage III (scan
down) with a 18% baking power for 20 minutes was obtained, which is the lowest value over the published data with the similar SiGe cap thickness.

## 7.2 Fermi Level Pinning

### 7.2.1 Contamination at the Regrowth Interface

A high temperature baking (-1000 °C, or baking lamp power 23% with the new SCR power control unit) at 6 torr with 3 slpm hydrogen flow is necessary to remove the remnant oxide layer or other contamination on the surface of a silicon wafer before any epi-layer growth. However, cleaning the SiGe buffer before growth is always much more difficult mainly because high temperature cleaning (-1000 °C) may relax the SiGe graded buffer and introduce more defects into the substrate. Therefore, the common baking temperature for SiGe buffers before growth is controlled below 900 °C (18%-20% baking power, with the new SCR power control unit). Because of this lower baking temperature, high contamination levels are very common at the regrowth interface of typical undoped 2DEG samples grown in our lab by RTCVD (Fig. 7.1). At the regrowth interface, the oxygen and carbon concentrations are $1 \times 10^{20}$ cm$^{-3}$ and $5 \times 10^{20}$ cm$^{-3}$, respectively, while those at the regrowth interface of a silicon wafer are typically below $1 \times 10^{18}$ cm$^{-3}$ (not shown here). The integral 2D densities of oxygen and carbon in Fig. 7.1 are $6 \times 10^{13}$ cm$^{-2}$ and $4 \times 10^{14}$ cm$^{-2}$, respectively. Such high sheet densities very likely pin the Fermi level ($E_F$) at the regrowth interface, and affect the electron densities in the buried QW in thermal equilibrium. Here we note that P and B background concentrations are both below the SIMS detection limit over all layers, so we believe the Fermi level pinning, if any, is not caused by background dopants at the regrowth interface.
Figure 7.1: A typical SIMS of an undoped 2DEG shows very high densities of oxygen and carbon at the regrowth interface. The integral 2D densities of oxygen and carbon are \(6 \times 10^{13} \text{ cm}^{-2}\) and \(4 \times 10^{14} \text{ cm}^{-2}\), respectively. A typical baking power for SiGe relaxed buffers is 20% for 5 minutes, which is roughly equal to 850 °C.

### 7.2.2 Fermi Level Pinning

To figure out the most likely pinning position in our samples, we calculated the theoretical maximum \(n_{2D}\) in equilibrium (The constant value indicated by the red solid line at high gate voltages in Fig. 6.3) by the Schodinger-Poisson self-consistency (SPSC) simulator based on different pinning conditions [96]. The experimental maximum \(n_{2D}\) in the buried QW in equilibrium of our undoped 2DEG samples (for example, the constant value \(-2 \times 10^{11} \text{ cm}^{-2}\) in stage III at high gate voltages in Fig. 6.14) are also collected for comparison.

The band diagrams of an undoped 2DEG at a bias 0.2 V with four different Fermi level pinning positions (0.045 eV below conduction band minimum \((E_C)\), 0.25 eV below \(E_C\), midgap and 0.045 eV above valence band maximum \((E_V)\)) are shown in Fig. 7.2. The pinning positions are intentionally selected to cover all possible pinning situations. This example 2DEG layer structure consists of a 4-nm Si cap, a 60-nm SiGe cap, a 11-nm Si channel and a 180-nm SiGe buffer. The gate bias 0.2 V is used to show the situation where the ground state \((E_0)\) of surface QW is still above
Figure 7.2: Band diagrams of an undoped 2DEG under a 0.2 V bias with the Fermi level pinned at four different positions: 0.045 eV below \( E_C \) (green), 0.25 eV below \( E_C \) (purple), mid-gap (brown) and 0.045 eV above \( E_V \) (orange).

The Fermi level for all pinning conditions. Other boundary conditions include the workfunction of the gate = 5.1 eV (Au) and the electron affinity of aluminum oxide (ALO) = 1 eV [97]. Under this bias, the buried QW with the Fermi level pinned at 0.045 eV and 0.25 eV below \( E_C \) touches the Fermi level already, with the ground state \( (E_0) \) of the surface QW still above the Fermi level, so electrons only accumulate in the buried QW. The electron density in the buried QW \( (n_{\text{buried}}) \) keeps increasing until the \( E_0 \) of the surface QW is lowered below the Fermi level. The further increase in the gate voltage increases the electron density in the surface QW \( (n_{\text{surface}}) \), with the \( n_{\text{buried}} \) fixed at a constant value, which is denoted as the theoretical maximum \( n_{2D} \) in equilibrium.
We calculated the theoretical maximum $n_{2D}$ in equilibrium with different pinning conditions based on the layer structure described above (Fig. 7.3). The maximum $n_{2D}$ decreases when the pinning position of the Fermi level gets further below $E_C$, because it gets earlier for the $E_0$ of the surface QW to become lower than the Fermi level. For the case with the Fermi level pinned at mid gap or below, the $E_0$ of the surface QW is always lowered below the Fermi level first (earlier than that of the buried QW), and thus no electrons accumulate in the buried QW.

### 7.2.3 Comparison of Simulations and Experimental Data

Now the experimental maximum $n_{2D}$ in stage III from our samples are added for comparison. Since the thicknesses of the SiGe cap layer and the SiGe buffer layer...
Figure 7.4: The experimental maximum $n_{2D}$ at stage III, from samples whose buffer layers are between 150 nm and 190 nm, versus corresponding SiGe cap thickness. Simulations with a fixed buffer thickness (190 nm) and various $E_F$ pinning conditions are shown in green ($E_C-E_F=0.045$ eV), purple ($E_C-E_F=0.25$ eV) and brown ($E_F$ at midgap) solid lines and an orange dash line ($E_F-E_V=0.045$ eV).

both affect theoretical maximum $n_{2D}$, we vary one parameter each time with another one fixed. For the simulations in Fig 7.4 we vary the thickness of SiGe cap layer in the simulation with the SiGe buffer layer fixed at 190 nm for four different Fermi level pinning positions. The experimental data are shown in blue solid symbols, and their SiGe buffer layers are in the range of 150 nm to 190 nm. It is clear that, when the Fermi level is pinned at 0.045 eV below $E_C$, the simulation curve fits our experimental data best. With the Fermi level pinning position moving away from $E_C$, the simulated maximum $n_{2D}$ becomes lower, and eventually no electrons are accumulated in the buried QW because electrons are populated in the surface QW earlier.
Figure 7.5: The experimental maximum of $n_{2D}$ at stage III from samples whose SiGe cap layers are between 20 nm and 180 nm versus corresponding SiGe buffer layer thickness. The simulations were done with a fixed buffer thickness (190 nm). The simulation curve with the pinning position ($E_C - E_F = 0.045 \text{ eV}$) in green fits the data best.

Here we note that since the front boundary condition ($E_C - E_F = \text{metal workfunction} - \text{electron affinity of the insulator}$) and the bottom boundary condition (Four different Fermi level pinning conditions as shown above) are both fixed, a thicker SiGe cap increases the distance between two boundaries, and reduces the built-in electrical field under the zero bias. Therefore, less increase in gate voltage is needed, after electrons start accumulating in the buried QW, to get $E_0$ of the surface QW below the Fermi level. Thus a lower $n_{\text{buried}}$ is obtained in samples with thicker SiGe cap layers.

Fig 7.5 shows the experimental maximum $n_{2D}$ (red solid symbols with SiGe cap thicknesses in the range of 20 nm to 180 nm) with the simulation curves in which we
vary the SiGe buffer thickness with a fixed SiGe cap thickness at 90 nm. Again, the simulation curve with the Fermi level pinned at 0.045 eV below $E_C$ fits our experimental data best. All other pinning conditions do not allow electrons to accumulate in the buried QW over the wide range of SiGe buffer layer thicknesses, because $E_0$ of the surface QW is always lowered below the Fermi level earlier than that of the buried QW.

Many other factors also affect the simulation results of the theoretical maximum $n_{2D}$ in equilibrium. For example, the theoretical maximum density strongly depends on background doping concentration in an undoped 2DEG. Higher background doping density may provide unexpected electron accumulation at both silicon QWs. However, based on our SIMS results, we have ensured that the concentration of background impurities in our undoped 2DEG is below $1 \times 10^{14}$ cm$^{-3}$ (Fig. 5.15). Such low doping levels have been verified to cause no effect on the simulation results.

### 7.3 Effects of Regrowth Interface on 2DEG Properties: Thickness of SiGe Buffer

Various mobility-limiting mechanisms have been discussed in Ch. 5, and remote scattering was viewed as the most dominant scattering mechanism that limits mobility in our system. However, we also pointed out in Sec. 5.6 that there seems to be an upper limit that stops the mobility curves from moving upwards and limits the highest mobility to 400,000 cm$^2$/Vs, irrespective of the increasing thickness of SiGe cap above the 2DEG over 40 nm (Fig. 5.16). Interface roughness scattering and the effect of threading dislocations were examined, but none of them were identified as the real mobility-limiting factor.

We note that most samples in earlier experiments have similar thickness of SiGe buffer layer (~165 nm), which is also the distance from the channel to their bottom
interfaces (regrowth interfaces). In Sec. 7.2.1, high concentrations of contamination (Fig. 7.1) were observed at the regrowth interface, which also cause remote scattering. In other words, there may be two sheets of remote charged impurities existing in our 2DEGs. To verify this argument, we grew a set of samples with various thicknesses of the SiGe buffer layer from 15 nm to 290 nm, with the distance between the surface and the channel fixed (i.e., a fixed SiGe cap layer -75 nm). Fig. 7.6 shows transport properties of these samples.

A clear degradation in mobility is indeed seen in the samples with a buffer below 145 nm down to 35 nm. Further, the channel couldn't even be turned on in the sample with a 15-nm buffer layer. However, mobility curves of samples with a buffer layer
Figure 7.7: Highest mobility and critical densities (the lowest n$_{2D}$) at both stage II (scan up) and stage III (scan down, as shown in Fig. 6.14) extracted from samples with various thicknesses of SiGe buffer layers. The blue solid and blue open symbols are the lowest n$_{2D}$ at stage II and stage III, respectively.

of 145 nm or greater are still limited at the same level with the highest mobility at about 400,000 cm$^2$/Vs.

In addition, a thicker buffer layer also helps with the reduction of critical densities, both at stage II and stage III. For example, the critical densities (the lowest n$_{2D}$) at stage II (scan up) and stage III (scan down) of the sample with a 290-nm buffer are about $1.2 \times 10^{11}$ cm$^{-2}$ and $4.5 \times 10^{10}$ cm$^{-2}$, which are two times lower than those of the sample with a 35-nm buffer ($2.1 \times 10^{11}$ cm$^{-2}$ and $1.3 \times 10^{11}$ cm$^{-2}$) (Fig. 7.7).
7.4 Effects of Baking Power at the Regrowth Interface

To improve the quality of the regrowth interface, we varied the baking power of the regrowth interfaces before growing any epi-layers. Except for the baking temperature, all other parameters were intentionally kept the same, including a 75-nm SiGe cap layer and a 145-nm SiGe buffer layer. As we mentioned in Sec. 2.1.2, the percentage of full lamp power (baking power) will be used instead of temperature in the following discussion because all the baking temperatures are higher than 750 °C, which is the upper limit of the temperature measurement system in our RTCVD.

A series of transport properties from samples with various baking powers from 18% to 22% (with the new SCR power control unit) are shown in Fig. 7.8. Baking with 20% power for 5 minutes is our standard cleaning recipe for samples shown in Sec. 6.4. Again, all the highest mobilities and critical densities at stage II (scan up) and stage III (scan down) of these samples are extracted for comparison (Fig. 7.9). Mobility curves in Fig. 7.8 do not move upwards, with the highest mobilities still below 400,000 cm²/Vs, regardless of the baking temperatures. The highest mobilities remain at the same level except for a small drop in mobility of the sample with 21% baking power for unknown reasons.

However, we do see an improvement in critical density at both stage II (scan up) and stage III (scan down) with a decreasing baking power. The critical density at stage III measured from the sample with 18% baking is as low as 3.2×10¹⁰ cm⁻², five times lower than that measured from the sample with 22% baking (1.6×10¹¹ cm⁻²). To the best of our knowledge, this extremely low 2D electron density of 3.2×10¹⁰ cm⁻² in an undoped silicon 2DEG with a 75-nm SiGe cap is the lowest value among all the published data with a similar SiGe thickness.
It is still unclear why a lower baking power leads to a lower critical density. It could stem from distinct $E_F$ pinning positions at the regrowth interface after different temperature baking. Further reducing baking power below 18% leads to a difficulty in cleaning the regrowth interface. In fact, it has been very difficult to clean the regrowth interface at even 18% baking power for 5 minutes. The surface of an as-grown sample with an 18% baking was completely hazy after the epitaxy for a 2DEG structure, so we think a low baking temperature may not be effective to remove native oxides or remnant water molecules on a relaxed SiGe substrate. Thus, we lengthened the baking time from the regular 5 minutes to 20 minutes to enhance the cleaning process at 18% baking power. The extended baking time did help regrowth interface cleaning and provided us with a small clean area on an as-grown 2DEG to make a
Figure 7.9: Highest mobility and critical densities at both stage II (scan up) and stage III (scan down) extracted from samples with various baking powers. The lowest $n_{2D}$ at stage II and stage III are blue solid and blue open symbols, respectively.

Hall bar device out of it, but roughly 75% of the sample surface was still hazy, making a low baking process impractical.

7.5 Summary

The Fermi level pinning position at the regrowth interface in our 2DEGs has been identified to be very likely near the conduction band minimum, because with Fermi level pinned near $E_C$, the theoretical maximum 2D electron densities predicted by the simulation showed a good match with experimental data. The remote scattering from bottom interface (regrowth interface) was then examined by means of comparison in mobility curve and critical densities at both stage II (scan up) and stage III (scan
down). A thicker SiGe buffer layer leads to higher electron mobility and lower critical density. Therefore, a buffer layer thicker than 150 nm is recommended for an undoped 2DEG growth to obtain better transport properties. In addition, a lower baking power at the regrowth interface reduces the minimum $n_{2D}$. A density of $3.2 \times 10^{10}$ cm$^{-2}$ was obtained for a sample with a 75-nm SiGe cap thickness, which is the lowest value among the published data with a similar SiGe cap thickness. However, the upper limit of the highest electron mobility (400,000 cm$^2$/Vs) under all conditions still strongly suggests the existence of some other mobility-limiting mechanisms.
Chapter 8

Conclusions

8.1 Conclusions

The realization of a spin-based quantum bit in the form of a single-electron quantum dot device motivated the work in this thesis. The benefits of silicon, such as its inherently long spin coherence time and high compatibility with facilities in industry, make silicon-based material favorable to fabricate single-electron quantum dot devices for spin manipulation. We take advantage of the conduction band offset resulting from the tensile strain in a silicon layer grown pseudomorphically between relaxed SiGe layers to confine a 2DEG in such a quantum well. Both modulation-doped 2DEGs and enhancement-mode undoped 2DEGs were explored and the feasibility as a platform to isolate a single electron in a quantum dot was also evaluated in this thesis. Introductions to both types of 2DEGs and their practical quantum dot devices were given in Ch. 2, combined with an elaboration of our epitaxial growth system and growth details for Si/SiGe heterostructures used in this study.

In Ch. 3, successful suppression of the detrimental phosphorus surface segregation observed in modulation-doped strained silicon 2DEGs were demonstrated. By means of lowering the growth temperature for the SiGe cap layer from 575 °C down to
525 °C after a doping layer growth, a high hydrogen surface coverage was preserved, which blocked the segregation paths for phosphorus atoms in the sub-surface layer. Experimentally, this 50 degree temperature difference enabled a two order of magnitude reduction in surface phosphorus concentration \(10^{18} \text{ cm}^{-3}\) to \(10^{16} \text{ cm}^{-3}\) in doped 2DEGs with a SiGe cap layer thinner than 40 nm. As a result, the lower surface electrical field greatly reduced the gate leakage current in a Schottky gated 2DEG. The resulting tremendous improvement in the breakdown voltage of those Schottky gates (-2 V to -7 V) enabled us to fabricate quantum point contacts (QPCs) without any significant leakage.

A demonstration of the lateral electrical isolation of doped 2DEGs at 4.2 K by ion implantation of silicon and argon was shown in Ch. 4. As an alternative for electrical isolation of quantum devices fabricated on modulation-doped 2DEGs, implant isolation in contrast to conventional mesa etching, preserves surface planarization and thus prevents corner-induced gate leakages. The sheet resistance at 4.2 K in an as-implanted 2DEG was as high as \(1 \times 10^{13} \Omega/\Box\). The stability of damage induced by ion implantation was stable up to 550 °C, high enough for common processes for quantum devices in Si/SiGe heterostructures, such as the aluminum oxide deposition by ALD at 300 °C and the implant contact annealing at 550 °C. Finally, the 2DEG quality in terms of electron mobility was experimentally proven to remain intact after implantation, which convinces us of the reliability of implant isolation for 2DEG-based quantum devices.

The major scattering mechanisms that may limit electron mobility and critical electron density of an enhancement-mode undoped silicon 2DEG grown by RTCVD were first identified, theoretically and experimentally, in Ch. 5. Efforts to alleviate various mobility-limiting factors, such as the influence from charged impurities, successfully allowed for mobility at 4.2 K as high as 400,000 cm²/Vs and the critical density as low as \(4 \times 10^{10} \text{ cm}^{-2}\) in undoped 2DEGs with a SiGe cap as thin as 60 nm.
In addition, various experiments such as the variation of silicon growth temperature in the range of 575 °C to 700 °C have been conducted as well. The different growth temperatures as expected resulted in desirable variations in both Si/SiGe interface roughness and threading dislocation density in a series of undoped 2DEGs, which revealed the rarely known influence from the interface roughness scattering and the scattering from threading dislocations. These experiments showed that low silicon growth temperatures (625 °C or lower) for silicon QWs in an undoped 2DEG are preferred for better transport properties.

In Ch. 6, we described the discovery of a strong tunable screening effect, which is resulted from an induced surface electron layer in enhancement-mode undoped 2DEGs with thin SiGe caps (<40 nm). It has two major benefits: much enhanced mobility and reduced critical density. The argument that the formation of this surface electron layer switched a 2DEG system from non-equilibrium back to thermal equilibrium was in good agreement with the simulation results. The dominant scattering sources at the oxide/silicon interface were effectively screened by those shielding surface electrons. The resultant weaker Coulomb forces and potential fluctuations explained the impressive improvement in transport properties of undoped 2DEGs with thin SiGe caps, in which the capability for sharp electron patterning was preserved. An improved critical density was also observed after we reversed the scanning direction of the bias at high gate voltages, because the shielding electron layer is trapped at the oxide/silicon interface. A 2× lower critical density (\(-8\times10^{10} \text{ cm}^{-2}\)) in stage III (scan down) than that (\(-1.6\times10^{11} \text{ cm}^{-2}\)) in stage II (scan up) was observed in an example sample, but the mobility remained the same. This again concludes electron mobility and critical density may have different limiting mechanisms.

Extensive investigations into the bottom interface (regrowth interface) have been done in Ch. 7 by means of experiments and theoretical simulations. The Fermi level in an undoped 2DEG was found to be very likely pinned near the conduction
band minimum. The effect of remote scattering from the regrowth interface was also discussed. With a thick SiGe buffer layer (290 nm), a low critical density around $4 \times 10^{10} \text{cm}^{-2}$ was observed with electron mobility as high as 400,000 cm$^2$/Vs. However, mobility gets decreasing with a thinner SiGe buffer layer (75 nm or less) because the remote scattering from the regrowth interface becomes stronger. In addition, with a low baking power for the regrowth interface (18% for 20 minutes), a density of $3 \times 10^{10} \text{cm}^{-2}$ was observed in the sample with a 75-nm SiGe cap in stage III when a gate voltage ramped down, which is the lowest value among the published data with a similar SiGe cap thickness.

\section*{8.2 Future Work}

\subsection*{8.2.1 Unknown Mobility-Limiting Factors}

Much effort regarding quantum computing has built a solid foundation in the past few decades. We may have seen the dawn of the quantum computing era, even though there is still a long way to go for the debut of a quantum computer. We have measured electron mobility at 4.2 K as high as 400,000 cm$^2$/Vs in our undoped 2DEGs that may provide a promising platform to realize a spin-based quantum bit. However, we are still not able to establish the real mobility-limiting factors hidden in our samples at the low density regime. Various scattering mechanisms such as remote scattering, background scattering, interface roughness scattering and scattering from threading dislocations have been evaluated, but eventually ruled out as our mobility ceiling. Nevertheless, dislocation defects other than threading dislocations in relaxed SiGe layers, which can’t easily be revealed by wet etching, could still play a role in our 2DEGs. Huang et al. \cite{78} reported that higher electron mobility could be achieved by growing Si/SiGe heterostructures with a lower Ge fraction. They claimed that the lower density of threading dislocations in a 2DEG with a lower Ge fraction enabled
the high electron mobility, though we have verified that scattering from threading
dislocations in our samples is insignificant. However, some other invisible dislocation
defects may still exist in SiGe layers due to the imperfection of crystalline structure
and mismatched lattice constant. It is worth trying to grow Si/SiGe heterostructures
on a SiGe virtual substrate with lower Ge fractions, hopefully with lower defect
densities. We may see the effect, either significant or insignificant, from Ge fraction
on electron mobility easily.

Silane became our main silicon precursor five years ago when we introduced the
new gas supply system. We avoided using dichlorosilane (DCS) to avoid corrosive
issues. However, without the additional in-situ cleaning by chlorine atoms in DCS,
oxygen and carbon levels are at least two times higher in both Si and SiGe epi-layers,
compared to the growth using DCS, after this precursor replacement. It is still unclear
if these incorporated contamination atoms are scattering sources. More experiments
are necessary to identify their charge types combined with the investigation of their
potential influence on electron mobility.

8.2.2 Isotopically-Enriched Undoped 2DEGs for a Longer
Spin Coherence Time

As we mentioned above repeatedly, silicon-based materials are favorable for spin-
based qubits because of their naturally weaker hyperfine interaction and resultant
longer spin coherence time. Enhanced mobility of an undoped 2DEG represents a
great reduction in the density of charged impurities both in the epi-layers and at the
oxide/silicon interface. This is may lead to a longer spin coherence time because of
fewer scattering sources that may interfere with spin stability. Recently, spintronics
experts noticed the importance of silicon and germane isotopes and how they affect
the spin coherence time. For example, silicon consists of three stable isotopes with
their natural abundances as 92.2% $^{28}$Si, 4.7% $^{29}$Si and 3.1% $^{30}$Si [98]. Among them,
only $^{29}\text{Si}$ has non-zero nuclear spin, which scatters electron spin and thus reduces spin coherence time. Li et. al. [99] grew a strained silicon channel in undoped 2DEGs with isotopically enriched silane. The concentration of isotope $^{28}\text{Si}$ with zero nuclear spin was successfully increased up to 99.72%, while isotope $^{29}\text{Si}$ was lowered down to 0.08%, which translates to a spin dephasing time around 2 $\mu$s [100], six times higher than that measured in a silicon QD device grown by the natural silane. Although electrons used for spin manipulation are located in strained silicon channel, the nearby SiGe layers are also considered to be another source responsible for spin dephasing because $^{73}\text{Ge}$ (7.8%) (1 out of five stable Ge isotopes) possesses non-zero nuclear spin, similar to $^{29}\text{Si}$. A undoped silicon 2DEG in the Si/SiGe heterostructures grown by both isotopically enriched silane and germane is thus promising to achieve unprecedentedly long spin coherence time, that may greatly advance the feasibility of a spin-based qubit.
Appendix A

Publications and Presentations

A.1 Journal Articles and Conference Papers


**A.2 Conference Presentations**

1. C.-T. Huang, J.-Y. Li, and J. C. Sturm, "Very low electron density in undoped enhancement-mode Si/SiGe two-dimensional electron gases with thin SiGe cap layers,” *223rd Meeting of the Electrochemical Society*, Toronto, ON, Canada, May 12-17 (2013)


3. J.-Y. Li, C.-T. Huang, and J. C. Sturm, ”Extremely sharp phosphorus turn-off slope and effect of hydrogen on phosphorus surface segregation in epitaxially-grown relaxed Si$_{0.7}$Ge$_{0.3}$ by RTCVD,” *International SiGe Technology and Device Meeting*, Berkeley, CA, June 4-6 (2012)

4. C.-T. Huang, J.-Y. Li, and J. C. Sturm, "High breakdown voltage Schottky gating of doped Si/SiGe 2DEG system enabled by suppression of phosphorus
surface segregation,” *International SiGe Technology and Device Meeting*, Berkeley, CA, June 4-6 (2012)


Appendix B

Fabrication of Undoped Silicon 2DEGs

B.1 Growth of Undoped Silicon 2DEGs

B.1.1 Baking Before Growth on Si and SiGe Substrates

To reduce possible water vapors or oil vapors back-streamed from the main pump, high temperature baking is necessary before growing any epi-layers. We mentioned in Sec. 2.1.2 that the temperature control system of RTCVD can only precisely control temperatures up to 750 °C. Temperatures higher than that can only be controlled by a fixed lamp power, whose corresponding temperature can be calculated based on the extrapolations shown in Fig. 2.3. After the replacement of the SCR unit, the standard baking power is 23%. A typical baking run consists of (1) 23% baking at 250 torr for 20 minutes with 4 slpm H₂ flow and (2) 23% baking at 6 torr for 1 minute with 3 slpm H₂ flow. We flow 0.5 slpm H₂ after a baking cycle is done for 7 minutes to allow the reactor to cool down. After cooling down, we can start another baking cycle if more baking cycles are necessary.
Figure B.1: A typical SIMS measured on dirty epi-layers grown on a silicon substrate.

For a newly-installed, brand new quartz reactor tube, 100× or more baking cycles are recommended. If we grow layers in a dirty reactor, the epi-layers grown on silicon substrates would be still visually clean, but contamination levels in the films would be extremely high (for example, $10^{20}$ cm$^{-3}$ and $10^{18}$ cm$^{-3}$ for oxygen and carbon, respectively in Fig. B.1). However, if we try to grow layers on SiGe substrates with high contamination levels in the reactor, no epi-layers would be grown on it and the resultant sample surface must become very hazy. Therefore, it is important to grow a test run to make sure all layers grown on SiGe substrates are clean before growing real samples.

**B.1.2 Wet Cleaning for Growth Substrates**

Before growth, a 100-mm silicon carrier wafer, silicon temperature control pieces and SiGe buffer pieces are all cleaned by the standard pirahna solution ($\text{H}_2\text{SO}_4:\text{H}_2\text{O}_2=1:2$) for 20 minutes followed by a 2-minute diluted HF dip (DI water: 49% HF=100:1). Blow the silicon temperature control piece and the SiGe buffer pieces dry carefully, and load them onto the 100-mm silicon carrier wafer. Then we can load the 100-mm
carrier wafer onto the 100-mm quartz wafer stand in the load-locked chamber for growth.

**B.1.3 Baking and Carrier Wafer Coating**

Even if the reactor is used frequently, we always start with a single baking cycle as we described above without any wafers in it. After a baking, we may also choose to bake the carrier wafer alone (20% baking at 6 torr with 3 slpm H\textsubscript{2} for 10 minutes) and coat a thin silicon layer on it (13% lamp power at 6 torr with 50 sccm SiH\textsubscript{4} and 3 slpm H\textsubscript{2} for 10 minutes) without loading any small pieces of SiGe buffers. This way you can also bake out any remnant water vapors on the carrier wafer to avoid contamination on real samples. In addition, if the carrier wafer is new, this baking and coating procedure must be done before any growth. Here we also note that we do not bake or clean the carrier wafer between growth runs in the same growth day. We unload small Si temperature control piece and SiGe pieces and load a new set of samples onto the carrier for the next growth run in the load-locked chamber without taking it out.

**B.1.4 100-mm Carrier Wafer for SiGe Buffer Pieces**

The SiGe buffers used in our experiments are 200-mm wafers provided by AmberWave Systems Inc.. Since our reactor is designed for 100-mm wafers, we have to dice them into 1.05-cm\textsuperscript{2} pieces by the dicing saw in the cleanroom (ADT proVectus 7100) so that we can load them on a 100-mm carrier wafer into the reactor. The design of a 100-mm carrier wafer is shown in Fig. B.2. Because SiGe buffer pieces are thicker (700 µm) than a regular 100-mm silicon wafer (525 µm), whose thickness is standard thickness for temperature calibrations based on the transmission of infrared lasers in our RTCVD (Sec. 2.1.2), we dice a regular 100-mm silicon wafer into small pieces with the same size as SiGe buffer pieces and load it in the center of the carrier wafer.
Figure B.2: (a) The top view and (b) the cross section along the red line of a standard 100-mm carrier wafer.

for the temperature control. Here we note that a thick and lightly-doped wafer (1 mm thick FZ wafer) is preferred to make a carrier wafer. It can provide deeper recesses, which can hold small pieces tightly to avoid samples from falling during the loading process. The light doping of the carrier wafer prevents any contamination caused by the autodoping during the high temperature baking. See more details in Ph.D thesis of Kun Yao [101].

B.1.5 Standard Layer Structure of an Undoped 2DEG and Its Growth Recipe

A typical layer structure of an undoped strained silicon 2DEG consists of a 150-nm Si$_{0.72}$Ge$_{0.28}$ buffer, a 10-nm strained silicon channel, a 60-nm Si$_{0.72}$Ge$_{0.28}$ cap layer and a 3-4 nm strained silicon cap. The standard growth recipe for this structure is as follows: (1) 20% baking for 5 minutes (2) Grow the SiGe buffer layer with 62 sccm GeH$_4$ and 50 sccm SiH$_4$ at 575 °C for 25 minutes (3) Grow the strained silicon channel
with 200 sccm SiH$_4$ at 625 °C for 4 minutes (4) Grow the SiGe cap layer with 62 sccm GeH$_4$ and 50 sccm SiH$_4$ at 575 °C for 10 minutes (5) Grow the strained silicon cap with 200 sccm SiH$_4$ at 625 °C for 1.5 minutes. The growth pressure for all steps is 6 torr, with a constant 3 slpm H$_2$ flow. The typical growth rates for Si$_{0.72}$Ge$_{0.28}$ and silicon under the growth conditions mentioned above are 6 nm/minute and 2.5 nm/minute, respectively.

### B.1.6 A Typical SIMS for a High-Mobility Undoped 2DEG

A typical SIMS for an undoped 2DEG with group record mobility (400,000 cm$^2$/Vs) is shown in Fig. [B.3](#). The oxygen levels in SiGe and strained silicon channel are $1 \times 10^{19}$ cm$^{-3}$ and $4 \times 10^{17}$ cm$^{-3}$, respectively, while the carbon levels in SiGe and strained silicon channel are $2 \times 10^{17}$ cm$^{-3}$ and $1 \times 10^{17}$ cm$^{-3}$, respectively. The background phosphorus and boron concentrations are below the detection limit for all layers. The detection limit for P and B in a regular SIMS analysis is $4 \times 10^{15}$ cm$^{-3}$ and $1 \times 10^{16}$ cm$^{-3}$, respectively.
B.2 Processes for Enhancement-Mode Undoped 2DEGs

B.2.1 Full Processes

The full process to fabricate a typical undoped 2DEG after growth consists of four photolithography steps (Fig. B.4). Four devices can be made on a single 1 cm×1 cm sample. The details of the full process are described as follows: (1) The first lithography followed by a dry etching by Samco RIE800iPB defines a set of alignment marks for following processes. The etching recipe is recipe 4 with 25 etching loop counts to create about 200-nm deep marks (Fig. B.4a). (2) After stripping photoresist, we deposit a 200-nm silicon dioxide layer as an implant mask by PECVD 790. The deposition recipe is standard cleanroom public recipe (0SiO₂) for 10 minutes. (3) The second lithography step followed by a wet etching of oxide by BOE 10:1 defines the patterns for n⁺ contact ion implantations (Fig. B.4b). (4) Attach all samples on a 100-mm silicon wafer and send it to Leonard Kroko Inc. for phosphorus ion implantation. The implant recipe we use is a 3-step implantation consisting of 5×10¹⁵ cm⁻² at 30 keV, 5×10¹⁵ cm⁻² at 60 keV and 5×10¹⁵ cm⁻² at 100 keV. The tilt angle is standard 7 degree without any substrate heating. (5) After implantation, 1-hour annealing at 600 °C in N₂ by Thermco Brute IV, tube 3 is conducted to activate implanted phosphorus. (6) A 90-nm aluminum oxide is then deposited by atomic layer deposition (Cambridge NanoTech, Savannah 100). The detailed recipe for aluminum oxide deposition by ALD along with chamber cleaning procedure will be addressed later. (7) The third lithography followed by a wet etching of the aluminum oxide by BOE 10:1 exposes the contact regions (Fig. B.4c). (8) The fourth lithography defines the Hall-bar-shaped metal gates and the contact regions. After lithography, a 3-nm Cr and a 200-nm Au are evaporated onto samples for both metal gates and contacts (Fig. B.4d). The evaporation is done
by Edwards thermal evaporator in C428. Here we note that the E-beam evaporator for Au deposition has to be avoided because the radiation from the electron beam may damage the aluminum oxide underneath, causing an issue to measure such 2DEGs.

(9) Four devices are diced by the dicing saw for the following low-temperature Hall measurements.

B.2.2 ALD Chamber Cleaning and Deposition Recipes

The atomic layer deposition (Cambridge NanoTech, Savannah 100) is used to deposit the aluminum oxide layer in our undoped 2DEGs. After many depositions, flakes may appear and attach on both the lid and the chamber. If you notice some flakes on your samples after the deposition, it is the time to clean the chamber. To clean ALD, you have to contact Joe Palmer first to put the equipment offline for about 3 days. The cleaning processes are as follows: (1) Turn off all the heaters and wait
until all heaters are at room temperature. (2) Use Scotch-Brite to scrub the chamber and the lid until all the flakes are gone. (3) Use Textwipes with methanol to wipe the chamber several times. (4) You can also replace the O-ring, refill DI water bottle and replace empty TMA bottle at this time, but Joe Palmer should be there. (5) Blow the chamber dry with a nitrogen gun. (6) Increase the temperatures for all heaters. 300 °C for center heater, 250 °C for edge heater, 80 °C for precursor heater and 150 °C for the rest of heaters. Flow 100 sccm N₂ carrier gas and let it bake for a half day. (7) Coat the chamber with a regular deposition recipe without loading any samples in for over 1000 cycles (8) Run a test deposition on a silicon wafer. Measure the thickness and refractive index by the ellipsometer (Gaertner L3W16). The normal deposition rate is 0.9 Å per cycle while the normal refractive index is 1.64.

The recipe of aluminum oxide deposition by this ALD has four parameters: Precursor (0 is water and 1 is TMA), pulse time, exposure time and pump time (the unit is second for all time parameters). Three steps of the recipe are (1) [0, 0.01, 0, 5] and [0, 0.01, 0, 5] for 10 cycles. In this step we repeat water exposure to ensure the sample surface is covered by hydroxyl bonds (-OH) (2) [0, 0.1, 1, 7] and [1, 0.1, 1, 7] for 10 cycles. Here we lengthen the pulse time, exposure time and pump time to make sure the precursors have enough time to complete the reaction for the first few monolayers. (3) [0, 0.01, 0, 5] and [1, 0.01, 0, 5] for 890 cycles. The number of cycles can be adjusted to meet any thickness requirement. Total 900 cycles can give us an 80-nm aluminum oxide with a purple to dark blue color.
Appendix C

Non-Standard Wafer Growth

In addition to the absorption coefficient, the transmissions of two infrared lasers we use to calibrate our growth temperature strongly depend on the wafer thickness. This is the reason why we need to use a 100-mm carrier wafer to carry small pieces of SiGe buffers with a standard silicon piece in the middle. However, the main drawback of using a carrier wafer is that the available 2DEG sample areas are very limited for each growth. If we want to grow epi-layers on a full SiGe buffer wafer with any non-standard thickness, the issue of temperature calibration and control has to be overcome.

C.1 Growth on a Thin SiGe Buffer (400 µm)

One of our outside collaborator has 75-mm SiGe buffer with a 400-µm thickness. To take both loading process and temperature calibration into account, we have several options available to grow epi-layers on these non-standard substrates.
C.1.1 Option 1: 125-mm Carrier Wafer

If we stick to the small silicon temperature control piece for temperature control, we have to make a carrier wafer that can hold the 75-mm wafer along with 1 cm×1 cm silicon piece. The 100-mm wafer is way too small for this option, so we turn to 125-mm wafer (1 mm thick). The design is shown in Fig. C.1. This 125-mm carrier wafer consists of two recesses. The first recess with a hole in the middle is similar to that in a regular 100-mm carrier wafer and designed to hold a small silicon piece for temperature control. The large 10-side polygons are etched (about 500 µm deep) by Samco 800 to hold a 75-mm SiGe buffer.

Here came some difficulties if we choose option 1. First, etching a 125-mm wafer by Samco 800 needs to vent the etcher and modify the setting inside the chamber. We have to contact Pat Watson for this modification, and the etcher has to be set offline for 3 days. Second, a new design for a quartz wafer stand for 125-mm wafer is required because our regular wafer stand is designed for 100-mm wafers (See Fig. C.1). (Technical Glass Products (TGP) helps us to fabricate quartz wafer stands and quartz reactor tubes.) The third issue is the most fatal one: loading difficulty. Although
the smallest diameter of the loading path in our RTCVD is 150-mm, it is still very difficult to safely load a 125-mm wafer into the reactor because there are too many blind spots where the carrier can hit the inner walls of the chambers. Fourth, since the temperature control piece is at the front position, the landing position in the reactor for the wafer stand is different from the standard one. A new calibration for the wafer landing position has to be done to make sure lasers going through the silicon piece.

A test growth was done on a 75-mm silicon substrate carried by a 125-mm carrier wafer. The picture of the 75-mm silicon substrate after growth showed several clear color rings (Fig. C.2). The color rings represent the non-uniformity of the thickness of the epi-layers. We sent out four pieces of samples for SIMS analysis, and space between each piece is 1 inch (Fig. C.3). The thickness information extracted from SIMS are listed in Fig. C.4. Referring to the temperature control piece (where is the center of the temperature distribution in the reactor), the thickness of each layer gets thinner when the location of the piece gets further away from the center because
Figure C.3: The schematic of a 75-mm SiGe buffer and a silicon temperature control piece held by a 125-mm carrier wafer. The pieces sent out for SIMSs are highlighted.

<table>
<thead>
<tr>
<th></th>
<th>625°C (Si)</th>
<th>575°C (sSiGe) Ge: 19-20%</th>
<th>700°C (Si)</th>
<th>600°C (sSiGe) Ge: 14-15%</th>
<th>750°C (Si)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC (nm)</td>
<td>27 (100%)</td>
<td>55 (100%)</td>
<td>41 (100%)</td>
<td>39 (100%)</td>
<td>78 (100%)</td>
</tr>
<tr>
<td>F (nm)</td>
<td>24 (89%)</td>
<td>70 (127%)</td>
<td>30 (73%)</td>
<td>44 (113%)</td>
<td>58 (74%)</td>
</tr>
<tr>
<td>M (nm)</td>
<td>19 (70%)</td>
<td>58 (105%)</td>
<td>24 (59%)</td>
<td>35 (90%)</td>
<td>50 (64%)</td>
</tr>
<tr>
<td>B (nm)</td>
<td>8 (30%)</td>
<td>23 (42%)</td>
<td>11 (27%)</td>
<td>15 (38%)</td>
<td>28 (36%)</td>
</tr>
<tr>
<td>Expected Thickness (nm)</td>
<td>30</td>
<td>40</td>
<td>35</td>
<td>40</td>
<td>70</td>
</tr>
</tbody>
</table>

Figure C.4: The thickness comparison between pieces cut from different locations on the 75-mm SiGe buffer held by a 125-mm carrier wafer.

the growth temperature there is lower. This terrible thickness uniformity makes this option impossible to grow useful epi-layers on 75-mm SiGe buffers for experiments.

C.1.2 Option 2: 100-mm Carrier Wafer

We can also use 100-mm wafer as a carrier to hold 75-mm SiGe buffers, but we have to overcome the temperature calibration issue. The design of this 100-mm carrier
wafer is similar to the previous 125-mm carrier wafer. A 10-sides polygon recess is etched in the middle of a 100-mm wafer to hold a 75-mm SiGe buffer as shown in Fig. C.5.

Now we have to deal with the temperature calibration. Based on the following equation, we can calculate the normalized transmission of two lasers for different wafer thicknesses [17]:

\[
(NT_{ta}) = (NT_{tb})^{\frac{tb}{ta}}
\]  

(C.1)

where \(ta\) and \(tb\) are the thickness of the wafer a and wafer b for temperature calibration. \(NT_{ta}\) and \(NT_{tb}\) are the normalized transmission for wafers with thickness \(ta\) and \(tb\), respectively.

The calculated normalized transmissions for a 400-\(\mu\)m, 525-\(\mu\)m (standard) and 680-\(\mu\)m thick silicon wafer for several common growth temperatures are compared in Fig. C.6. For a 75-mm SiGe buffer, we still stick to the standard growth temperatures.
for undoped silicon 2DEG growth (575 °C and 625 °C for Si$_{0.72}$Ge$_{0.28}$ and Si growth, respectively.)

First we worried that the 3-µm thick SiGe buffer has a smaller bandgap, a high intrinsic carrier density and many defects, all of which may additionally absorb infrared signals and affect the accuracy of the temperature calibration. To solve that, we etched about 4-5 µm away in the middle of the 75-mm buffer layer (1 cm×1 cm). This non-planar wafer surface raises a concern about the photoresist uniformity for the following processes after growth, so a simple test was done by spinning a layer of photoresist (5214) on an etched 75-mm wafer and measuring the thickness of the photoresist by NanoSpec. We could see four streaks radiating from four corners of the etched square. The thickness of resist right at the corner (the thickest part) is 1.442 µm measured by Nanospec. Along the streaks, we measured the point roughly
Figure C.7: The SIMS analyses for two separate pieces cut from a 75-mm SiGe buffer carried by a 100-mm carrier wafer. The peaks in the silicon channel could be artifacts.

15 mm away from the corner, the thickness is 1.418 µm. We also measured a point at uniform area, the thickness is 1.417 µm. Therefore, the thickest point (1.442 µm) is about 1.8% thicker than uniform area, which does not affect too much the following photolithography and patterning. Here we note that for the ebeamresist, the color of the resist became uniform after 170 °C soft baking, so the uniformity of the ebeam-resist should be fine.

Even though the uniformity of the resist is OK, we still try to avoid the etching process because it consumes a considerable area which can not be used after growth. Also, we can avoid the patterns (the square in this case) on a substrate to be grown.
Figure C.8: The thickness non-uniformity of epitaxial layers grown on a 75-mm SiGe buffer carried by a 100-mm carrier wafer.

The patterns on the substrate always cause contamination for the growth, so hazy areas along the patterns are always inevitable. A quick calculation done by Kun Yao [101] showed that the thin layer of SiGe buffer does not affect the normalized transmission too much compared to the silicon substrate with couple of hundreds microns. The only effect is the intensities of lasers at room temperature are slightly lower.

After a typical undoped 2DEG growth (with enriched silicon in the silicon channel) on a 75-mm SiGe buffer carried by a 100-mm carrier wafer without etching a hole, we sent out two pieces (one center piece and one piece 8 mm from the edge) for SIMS, and the results are shown in Fig. C.7. We can still see the thickness non-uniformity, but it is much better than the results of the option 1. The thickness information is extracted and compared in Fig. C.8. The center thickness is very close to the expected thickness. The edge thickness is thinner than the center, but it is still a reasonable undoped silicon 2DEG structure which can be used for the future measurements.
Enhancement-mode Hall bar devices were fabricated on both edge and center pieces to check the 2DEG quality. The Hall measurement data at 4.2 K is shown in Fig. C.9. The highest mobility for two center pieces is 400,000 cm$^2$/Vs, which is our group record. The highest mobility for the edge piece is 200,000 cm$^2$/Vs, lower than that of the center pieces, but it is still fairly high. The reason for reduction in mobility for the edge piece could be the thinner SiGe cap or buffer layers, or the lower baking temperature before the epitaxial growth. Here we note that the peaks of all elements in the silicon channel in the SIMS could be artifacts (Fig. C.7) because no interruption occurred during the growth, and high electron mobility was still observed in both samples.

One additional figure that shows the thickness uniformity for a 75-mm silicon substrate (400 µm thick) on a 100-mm silicon carrier wafer is shown here for reference.
Figure C.10: The thickness percentages of epi-layers grown on a 75-mm silicon substrate carried by a 100-mm carrier wafer.

(Fig. C.10). For SiGe growth, the film thicknesses 17 mm away from the center are still above 90% of film thickness in the wafer center, while they fall to around 70% at the location 34 mm away from the center. In addition, the thickness uniformity of silicon films are worse than SiGe films over the whole 75-mm wafer. Higher silicon growth temperature leads to the worse thickness uniformity.

**C.1.3 Option 3: 75-mm Quartz Wafer Stand**

Another even simpler option is to fabricate a special quartz wafer stand for 75-mm wafer only. This way, the 100-mm carrier wafer is no longer needed, and the loading process would become easier and safer as well.
C.2 Growth on a Thick SiGe Buffer (680 \( \mu m \))

C.2.1 Dicing an 150-mm SiGe Buffer into a 100-mm Wafer

One of our collaborators also required us to grow epi-layers on thick SiGe buffers (680 \( \mu m \)). Such thick SiGe buffers are 150-mm wafers, so before we load it into our reactor, we have to dice them down to 100 mm. Bert Harrop helped me to dice the 150-mm wafers by using ADT dicing saw into to a 20-side polygons, whose peak-to-peak distance is 100 mm, and flat-to-flat distance is 99 mm. The size tolerance of this diced wafer is very small because of following two reasons. First, since the diced wafer is held by a quartz wafer holder when we load it into the reactor, the diameter of this diced wafer can not be smaller than 95 mm. Second, we have to load this wafer into the etcher Samco 800 to etch the central SiGe buffer layer out (will talk about this later), so the wafer size has to fit in the stage inside the Samco 800 chamber. A quartz ring used to cover the top of the wafer to be etched by Samco 800 has a groove with 101-mm diameter. A quick test can be done by putting the diced wafer into the groove to see if the diced wafer can fit in it. We have to do a loading test into Samco 800 as well to see if it can sit in the slot inside the chamber and cover all the helium cooling holes. A successful etching test on a diced wafer is shown in Fig. C.11.

C.2.2 Growth Temperature Calibration for a Thick SiGe Buffer

We can choose to etch the center of this thick SiGe buffer layer down to 525 \( \mu m \), which is our standard thickness for temperature calibration. However, the deep etching could leave the dirty sidewalls very close to our epitaxial areas, leading to hazy regions which reduce the usable sample surfaces. Thus, we decide to remain its thickness for temperature control. Based on the calculation shown in Fig. C.6, we can also calibrate the temperature by a 680-\( \mu m \) thick SiGe buffer. However, since the wafer is thicker
Figure C.11: The picture of a diced 100-mm SiGe buffer from a 150-mm SiGe wafer. The ring along the edge of the diced wafer is the etched pattern from the quartz ring of Samco 800 etcher, showing the quartz ring perfectly covered the edge of the diced wafer during the etching.

than the standard one, the signals of the infrared lasers measured by lock-in amplifiers are smaller. With the SiGe buffer on it, the signals are even smaller. In addition, for a thick wafer, the normalized transmission values for common growth temperatures are smaller too. Therefore, in order to get high enough signals to calibrate temperatures, we still prefer to etch away the SiGe buffer layer for about couple of microns. Here we note that there is always a gap between the normalized transmission of 1.3-μm and 1.55-μm lasers where the temperature can not be precisely controlled (Fig. C.6). The gaps for 525-μm and 400-μm thick wafers are both higher than our regular silicon growth temperature, 625 °C, but that for a 680-μm thick wafer is between 600 °C and 650 °C. Therefore, we lower the silicon growth temperature for thick SiGe buffer down to 600 °C for a better temperature control. (The growth temperature for Si$_{0.72}$Ge$_{0.28}$ is still kept at 575 °C)
Figure C.12: The pieces cut for SIMS analyses from a diced 100-mm SiGe buffer after a standard undoped 2DEG growth.

### C.2.3 Growth Uniformity and Mobility Results

After a regular undoped 2DEG growth, two pieces cut from the diced thick SiGe buffer were sent out for the SIMS analyses (Fig. C.12). The SIMS show again non-uniform thicknesses (Fig. C.13), but the layer structure of the edge piece still gives us a reasonable 2DEG structure. The carbon level in the edge piece is higher than the one closer to the center, which is common in our growth, but the cause is still unknown.

The thickness information is again extracted and compared in Fig. C.14. The SiGe cap thickness percentage of edge piece to the piece closer to the center is less than 50%, which is much worse than that in the undoped 2DEG grown on a 75-mm SiGe buffer (75%) in Fig. C.8. As for the electron mobility, the highest mobility for both edge or center piece are still as high as 200,000 cm$^2$/Vs (Fig. C.15). Even if this value is only half of our group record, it is still high enough for the following physics experiments.
Figure C.13: The SIMS analyses for the pieces with different distances from the wafer center of a diced 100-mm SiGe buffer.

![SIMS Analyses](image)

Figure C.14: The thickness non-uniformity of epi-layers grown on a diced 100-mm SiGe buffer.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Expected Center Thickness</th>
<th>32mm from the Center</th>
<th>8mm from the Center</th>
<th>Thickness Percentage (Refer to the Piece 8mm from the Center)</th>
</tr>
</thead>
<tbody>
<tr>
<td>600°C Si</td>
<td>4nm</td>
<td>&lt;2nm</td>
<td>3nm</td>
<td>~55.5%</td>
</tr>
<tr>
<td>Si_{0.7}Ge_{0.3}</td>
<td>70nm</td>
<td>35nm</td>
<td>83nm</td>
<td>42%</td>
</tr>
<tr>
<td>600°C 28Si</td>
<td>10nm</td>
<td>5nm</td>
<td>9nm</td>
<td>55%</td>
</tr>
<tr>
<td>Si_{0.7}Ge_{0.3}</td>
<td>190nm</td>
<td>120nm</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Figure C.15: The mobility measured from pieces with different distances from the wafer center of a diced 100-mm SiGe buffer. The highest mobility at 4.2 K is 200,000 cm$^2$/Vs.
Bibliography


[82] G. Dingemans, N. M. Terlinden, M. A. Verheijen, M. C. M. van de Sanden, and W. M. M. Kessels, “Controlling the fixed charge and passivation properties of Si(100)/Al$_2$O$_3$ interfaces using ultrathin SiO$_2$ interlayers synthesized by atomic layer deposition,” Journal of Applied Physics, vol. 110, no. 9, p. 093715, 2011.


