Silicon And Silicon-germanium Epitaxy For Quantum Dot Device Fabrications Towards An Electron Spin-Based Quantum Computer

Kun Yao

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: James C. Sturm

September 2009

Abstract

Semiconductor quantum dots are promising candidates as qubits for spin-based quantum computation as they provide highly tunable structures for trapping and manipulating individual electrons. It is the objective of this doctorate thesis to study on the development of silicon and silicon-germanium epitaxy and nanofabrication techniques for quantum dot devices, and the performance level achieved in the silicon/silicon-germanium material heterosystem.

We describe the growth of two-dimensional electron gas structures in strained Si on high-quality SiGe relaxed buffers with low temperature mobility exceeding 10,000 cm^2/Vs, currently limited by the background impurities in our RTCVD system. The modulation of the electron gases using atomic layer deposited Al_2O_3 is also demonstrated. We have developed a wide range of fabrication methods of the electron gases for quantum dot applications including nanolithography and etching techniques optimized for etch selectivity and anisotropy. Feature sizes well under 100 nm can be reliably obtained.

To achieve precise control of exchange coupling of qubits, we present a new concept of parallel 2-D electron gases in a double quantum wells as interaction dimers. A typical value of 0.1 meV for symmetric-anti-symmetric splitting of subbands is predicted by modeling. The signature of inter-well scattering is proved by a negative transconductance effect measured in such structures. The physical realization of such qubit dimers can also enable a novel "flying qubit" scalable architecture for semiconductor-based quantum computers.

The robustness of quantum dot devices is often strongly affected by defect states on the surface arising from the Si/SiO_2 interface. We demonstrate the use of epitaxial regrowth of SiGe for surface passivation, done with thermal cleaning temperatures less than 800 °C and negligible degradation of device performance. Side-gated multiple quantum point contacts are fabricated. They can be used to completely deplete

electrons on a quantum dot with gate leakage less than a few nA. We have also observed periodic single electron tunneling conductance peaks in a single quantum dot transistor with a side-gate-to-dot capacitance of 4.4 aF.

Acknowledgements

First of all, I want to thank Princeton University and the Department of Electrical Engineering for giving me this wonderful opportunity to carry out my research work in the first instance.

I am deeply indebted to my advisor Professor James C. Sturm, whose perpetual enthusiasm, stimulating insight, and constant encouragement that have shaped me throughout the years. I always feel extremely lucky to have had worked with him inside and outside the laboratory. I also wish to thank Professors Jason R. Petta and Antoine Kahn for reviewing this thesis. Furthermore, Professors Sigurd Wagner, Mansour Shayegan, Stephen A. Lyon, Daniel C. Tsui, and Stephen Y. Chou are always willing to help me whenever I need their guide. They have set a role model of a great teacher and mentor.

This thesis could not have been achieved without all the collaborators. I would like to express my gratitude to Professor Leonid P. Rokhinson and his post-doc Dr. Alexander Chernyshov at Purdue University for their low-temperature measurement, Dr. Anthony Lochtefeld and his team at AmberWave Systems for their preparation of SiGe relaxed buffers, and Professor Ya-Hong Xie at UCLA for providing MBE samples. At Princeton, I would like to thank e-beam nanolithography expert Dr. Mikhail Gaevski, TEM imaging and analysis expert Dr. Nan Yao, and Shyam Shankar from Lyon group for his help with ESR measurement. I treasured all the cooperative work experiences that made the research more rewarding for me.

A special thanks goes to all members in Sturm group. Malcolm S. Carroll, Haizhou Yin, Xiang-Zheng Bo, Eric J. Stewart, Rebecca L. Peterson, Keith H. Chung, Weiwei Zheng, Sushobhan Avasthi, and Jiun-Yun Li are all among our RTCVD hall of fame. In particular, I am obliged to Keith for his friendship and loyal support to carry the reactor upon our shoulders for nearly half decade. All other folks, including Iris Hsu, Richard Huang, Ke Long, Troy G. Abe, John A. Davis, David Inglis, Hongzheng

Jin, Bahman Hekmatshoar, Yifei Huang, Noah Jafferis, and Kevin Loutherback, contributed to make the lab a fun place to work. I would also like to thank my colleagues in Tsui group, Gabor Csathy, Yong Chen, Zhihai Wang, Keji Lai, Tzu-Ming Lu, and Dwight Luhman. They are great physicists and have offered me broad perspectives during countless discussions.

My doctoral research would be very painful if I did not have the kind support from PRISM and EE staffs. I am thankful to Dr. Helena Gleskova (now on the faculty of the University of Strathclyde in Scotland), Dr. George Watson, Dr. Conrad L. Silvestre, and Joe E. Palmer for their conscientious maintenance of the cleanroom. I truly appreciate Cathy A. Wertz, Sarah M. Braude, Carolyn M. Arnesen and Sheila R. Gunning for their generous help.

Finally, my deepest gratitude goes to my family for their love and care throughout my life; I am simply impossible without them. I grew up in a typical south China small town and have come a long way. I have no suitable words that can describe the evergreen support from my parents and brother from oversea in the past eight years. Last but not least, to my fiancée Lan, I remember all the moments with your accompanying me during the hard times. Your love enabled me to accomplish this work.

Thank you! Thank You!! Thank YOU!!!

To the victims and heroes of September 11, 2001

On that day I was sitting 50 miles away at Princeton university graduate school orientation in Richardson auditorium in Alexander Hall.

Contents

List of Tables

List of Figures

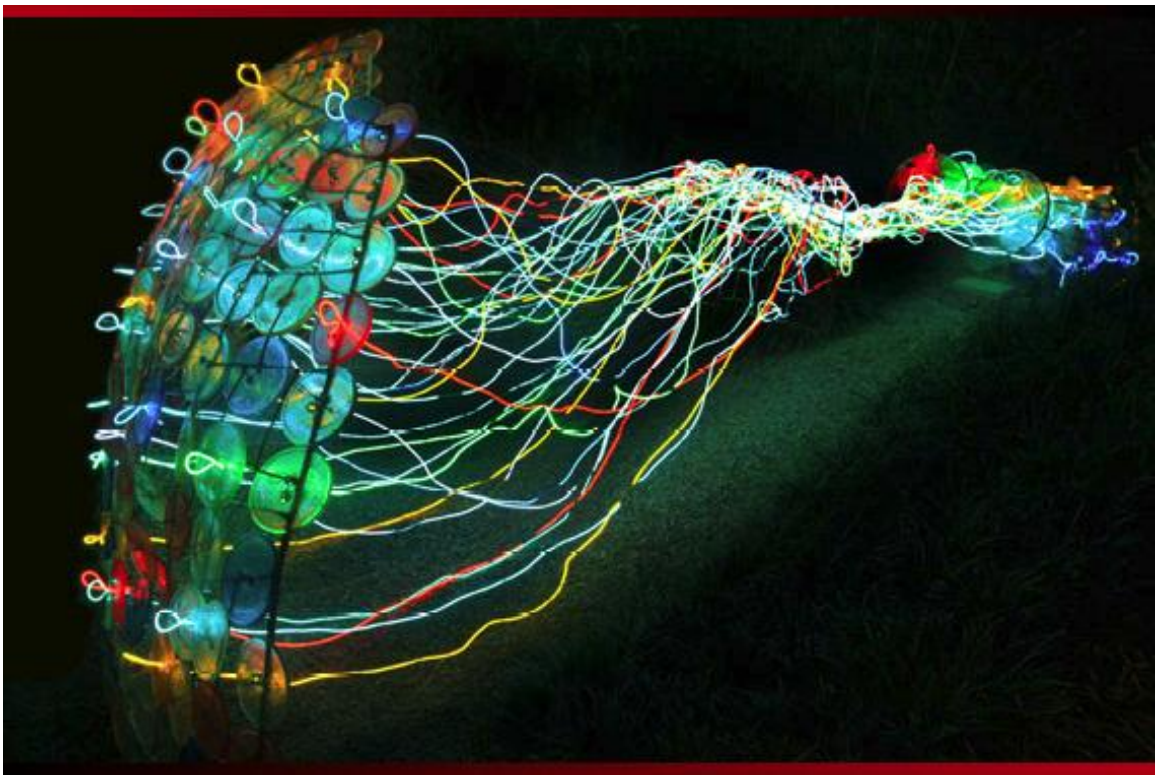


Figure 1: Sensation: Interior View (2006) is an abstract sculpture by Jersey City artist Nancy Cohen that was inspired by discussions with Princeton University President Shirley Tilghman. Tilghman, a leader in the field of molecular biology, collaborated with Cohen and Princeton University Electrical Engineering Professor James Sturm on the artwork, which is an abstraction about the sense of smell and how odors are recognized and remembered. Multi-colored cast resin discs are affixed to a steel armature forming a wall that connects to bulb-shaped structures by vibrant wires. The different colors of discs represent the sensor neurons in the nose that detect different odorant molecules; the wires represent the axonal connections that pass through the skull to the olfactory bulb in the brain, with the neurons from each type of sensor going to their own specific region in the olfactory bulb. (Image courtesy of NSF, from NSF IPAMM final report, 2007)

Chapter 1

Introduction

1.1 Motivation

As the raw material of the information age, silicon has changed the world in a revolutionary fashion over the past 50 years. Interestingly, since silicon is similar to carbon, particularly in its valency, some people have proposed the possibility of silicon-based life. Life itself as we know it could not have developed based on a silicon biochemistry. However, the impact of silicon-based microelectronics on our life is apparent all around us. And the microelectronics universe itself is still expanding dramatically.

Moore's law has been the most important and most famous benchmark for developments in silicon technology. It describes that the number of transistors on a chip will double about every two years. The integrated circuit industry has kept that pace for nearly 40 years. As of today, Intel's next-generation Itanium processors will have 2-billion transistors [1]. CMOS scaling aggressively approximates the molecular scale in less than 10 years, transistors would eventually reach the limit of miniaturization at nearly atomic levels. Then the physical laws that govern the behavior and properties of the circuit will be inherently quantum mechanical in nature, not classical any more. On April 13, 2005, Gordon Moore himself stated in an interview that the law

cannot be sustained indefinitely [2]:

> In terms of size [of transistor] you can see that we're approaching the size of atoms which is a fundamental barrier, but it'll be two or three generations before we get that far - but that's as far out as we've ever been able to see. We have another 10 to 20 years before we reach a fundamental limit. By then they'll be able to make bigger chips and have transistor budgets in the billions.

This is why we begin our journey at the level of atoms and electrons. Quantum computing, as its name suggests, may eventually allow computing to surpass the atomic level size restrictions. It represents the most promising possible final destination beyond the microelectronics roadmap. A quantum computer makes direct use of distinctively quantum mechanical phenomena to realize a fundamentally new mode of information processing. There are a number of quantum computing candidates, including those based on superconductors, Bose-Einstein condensates, quantum optics, and many others. Our work will focus on the "spintronic" quantum dot in silicon concept. This technology exploits the intrinsic spin of electrons and its associated magnetic moment, which is also the origin of the term "spintronics". A clear advantage of this route is that our experimental technique is compatible with existing complementary metal oxide semiconductor (CMOS) technology, making integration of quantum dots in silicon chips feasible.

1.2 A Brief History of Quantum Computing

The idea of a computational device based on quantum mechanics was first explored in the 1970's and early 1980's by a small number of visionaries, such as Charles H. Bennett of the IBM Thomas J. Watson Research Center, Paul A. Benioff of Argonne National Laboratory in Illinois, David Deutsch of the University of Oxford, and

Richard P. Feynman of the California Institute of Technology. The idea of a quantum gate was introduced, the basic possibilities of quantum algorithms were set forth, quantum communication (in the form of quantum cryptography) was well developed, and some rudimentary ideas of how quantum computing could be implemented were considered.

In 1994, Peter Shor, a research and computer scientist at AT&T's Bell Laboratories in New Jersey, devised the first quantum algorithm [3]. Shor's algorithm harnesses the power of quantum superposition to rapidly factor very large numbers (on the order $\sim 10^{200}$ digits and greater) in a matter of seconds, much faster than is possible on conventional computers. With this breakthrough, quantum computing was transformed from a mere academic curiosity into a world interest. There is now a fast growing list of potential quantum tasks such as cryptography, error correcting schemes, quantum teleportation, etc. that show even more desirability of experimental implementations of quantum computing [4]. There is a remarkably long list of physical systems that have been proposed for potential realizations. Several significant experimental examples include: trapped ions [5], cavity QED [6], nuclear magnetic resonance [7], superconducting devices [8, 9, 10, 11, 12]. The Loss-DiVicenzo proposal that would use coupled quantum dot arrays as qubits [13] for a semiconductor-based quantum computer has attracted many researchers in solid-state physics field.

To consider how a practical quantum computer can be built, David DiVincenzo of the IBM Thomas J. Watson Research Center gave a simple list of five requirements: [14]

1. The machine should have a collection of qubits.

2. It should be possible to set all the memory bits to 0 before the start of each computation.

3. The error rate should be sufficiently low.

4. It must be possible to perform elementary logic operations between pairs of bits.

5. Reliable output of the final result should be possible.

To summarize the challenges from the perspective of an engineer, one needs to be able to build a physical system with a collection of well-characterized quantum two-level systems (qubits). Furthermore, the decoherence time of these qubits should be long (compared to the "clock time"), so that the qubits are to a high degree isolated from coupling to the environment so as to not decohere while performing operations on them. Electron spins in silicon are promising because of their long decoherence times, but electrical gating schemes for doable logic operations and measurements of individual qubits still seem challenging.

The Loss-DiVincenzo proposal uses the intrinsic spin of the electrons in coupled quantum dots as the basic carrier of information. Desired operations are effected by the gating of the tunneling barrier between neighboring dots. At this point, devices capable of quantum computing must be constructed so that theory can be put to test. This relies on simultaneous further advances in the experimental techniques of semiconductor nanofabrication, magnetic semiconductor synthesis, single electron electronics, and scanning probe techniques.

At present, quantum computing and quantum communication technology remains in its infancy, especially on the hardware side. However there is a vast amount of versatility as well. Eventually time will tell whether any of these efforts will actually provide a successful route to a quantum computer. Its future undoubtedly lies in the profound effect it will have on the lives of all mankind.

1.3 Thesis Outline

Chapter 2 gives a brief review of the properties of the strained Si/SiGe heterostructures and the means to exploit the strain status in the layers for band engineering.

This chapter also contains details of Si and SiGe epitaxy at Princeton by rapid-thermal CVD (RTCVD).

Chapter 3 is dedicated to two-dimensional electron gas (2DEG) in modulation-doped Si/SiGe heterostructures, which is the physical embodiment of single electrons in quantum dots. Transport properties of electrons and modulation of heterojunction potentials are discussed. A theoretical model is studied to understand the limits on low-temperature electron mobilities.

In chapter 4 we extend our study to parallel 2DEGs in double quantum wells. Such a structure can serve as interaction dimer between two adjacent qubits in a "flying qubit" architecture for quantum computing. For this purpose, we investigate the band structures and the epitaxial growth of double quantum wells.

Chapter 5 presents a thorough review of quantum dot fabrication methods. Various nanolithography and etching methods are critically studied and compared. Examples of successful device applications are given. The expected limitations and benefits of different fabrication options are assessed.

Chapter 6 is focused on Si/SiGe epitaxial regrowth, a novel concept and technique for silicon-based quantum dot surface passivation with ideally zero defects and interface states. A low-temperature cleaning and growth sequence is developed to achieve conformal crystalline passivation over nanopatterned device surface on a wafer scale. This chapter also demonstrates the electrical compatibility of the regrowth technique with the existing 2DEG structures.

The growth, fabrication and epitaxial passivation techniques of Si/SiGe structures are the key to the realization of silicon-based quantum computer architectures. Chapter 7 covers the device aspects of these achievements. Simple quantum device applications including quantum point contact (QPC) and a single quantum dot are presented. We also discuss the possible role of such quantum devices in future quantum computers.

Finally in chapter 8 a summary of the contributions of this thesis is combined with a brief discussion of some possible future directions of research towards quantum computers.

Chapter 2

Si/SiGe Epitaxy

2.1 Introduction

Silicon-based heterostructures have come a long way from the discovery of strain as a new and essential parameter for band structure engineering, to the present state of electron and hole mobilities enhancement which surpass those achieved in the traditional Si/SiO_2 structures. Germanium can be considered as a kind of 'natural' choice for silicon-based heterostructures: the two group-IV elements silicon and germanium crystallize in the same diamond lattice, and form random $Si_{1-x}Ge_x$ alloys of arbitrary composition. By means of these heterostructures, the band structure can be tuned within a wide margin. Table 2.1 lists the most important physical properties of silicon and germanium. In addition, their structural and chemical properties are very similar, which eases epitaxial growth and the application of standard Si CMOS technologies.

The obvious advantages of Si/SiGe heterostructures were recognized at an early stage of research, with the first report on a Si/SiGe superlattice appearing already back in 1975 [15]. After more than 30 years of research and development in this field, SiGe is commonly used as a strain-inducing layer for CMOS transistors in modern

Table 2.1: Selected physical properties of unstrained bulk silicon and germanium at room temperature.

		Silicon	Germanium
Crystal structure		diamond	diamond
Lattice constant (Å)		5.431	5.657
Dielectric constant		11.9	16.2
Direct bandgap (eV)		3.40	0.80
Indirect bandgap (eV)		1.12(Δ)	0.66(L)
Electron mass (m_0)	m_t	0.19	0.08
	m_l	0.91	1.59
Bulk mobility ($cm^2V^{-1}s^{-1}$)	Electrons	1450	3900
(T = 300 K)	Holes	505	1800

technology. Since its introduction at the 90-nm node, strain has become a central performance enhancement element for the standard CMOS flow. With the scaling of the thickness of gate dielectric in silicon CMOS devices, channel mobility in MOSFETs is trending towards lower values due to higher vertical fields [16]. Embedded SiGe source/drain (S/D) was first used in production for 65-nm PMOS. In the 45-nm method, the combined impact of higher Ge fraction in the embedded S/D and the strain enhancement from poly gate removal allow for a 1.5× higher hole mobility compared to 65-nm, despite the scaling of the transistor pitch (contact-poly-contact spacing) from 220 nm to 160 nm.

A key benefit of using SiGe in quantum computing applications comes from the Si/SiGe modulation-doped heterostructures. Modulation-doped structures were first conceived by Dingle in 1978 [17]. In such structures, the active layers consist of an undoped channel for the mobile carriers, an undoped spacer layer that separates the ionized dopants from the channel, and a doping layer. The carriers are confined at the heterojunctions to form a two-dimensional electron/hole gas (2DEG/2DHG). High mobilities are realized in the 2-D gas because the thick spacer layers significantly reduce Coulomb scattering at the ionized impurities of the doping layer. In the spin-qubit scheme for quantum computing we have considered, we are particu-

larly interested in quantum dots created by lateral confinement in a 2DEG. Because individual quantum states are accessible in such quantum dot, we can trap single electrons. The precise control of the electron number is accomplished using the well-documented Coulomb blockade effect.

2.2 The Strained Si/SiGe Heterostructures

2.2.1 Structural Properties

As the lattice constant of germanium is about 4.2% larger than that of silicon, the lattice constant of a bulk SiGe alloy can be estimated by Vegard's rule which uses a linear interpolation of the parameters of the end-point elements of Si and Ge:

$$a(Si_{1-x}Ge_x) = a(Si) + [a(Ge) - a(Si)]x, \tag{2.1}$$

where x represents the fraction of germanium atoms.

Let's first consider the case that a layer of pure silicon is deposited on top of bulk SiGe alloys. When the silicon film is thin, the in-plane silicon lattice tries to stretch and line up with the SiGe. The thin pseudomorphic (meaning that lattice mismatch is accommodated by strain in the thin film on a lattice-mismatched substrate) silicon becomes tensilely "strained". Fig. 2.1 illustrates the formation of such strained thin films.

The thin films cannot relax, because the elastic energy stored in such a homogeneously strained layer is lower than the elastic energy associated with the local distortion around a misfit dislocation. Also as a result of the strain, the perpendicular lattice constant of the silicon, $a_{i\perp}$, will decrease:

$$a_{i\perp} = a_i \left[1 - D^i \left(\frac{a_{i\parallel}}{a_i} - 1\right)\right], \tag{2.2}$$

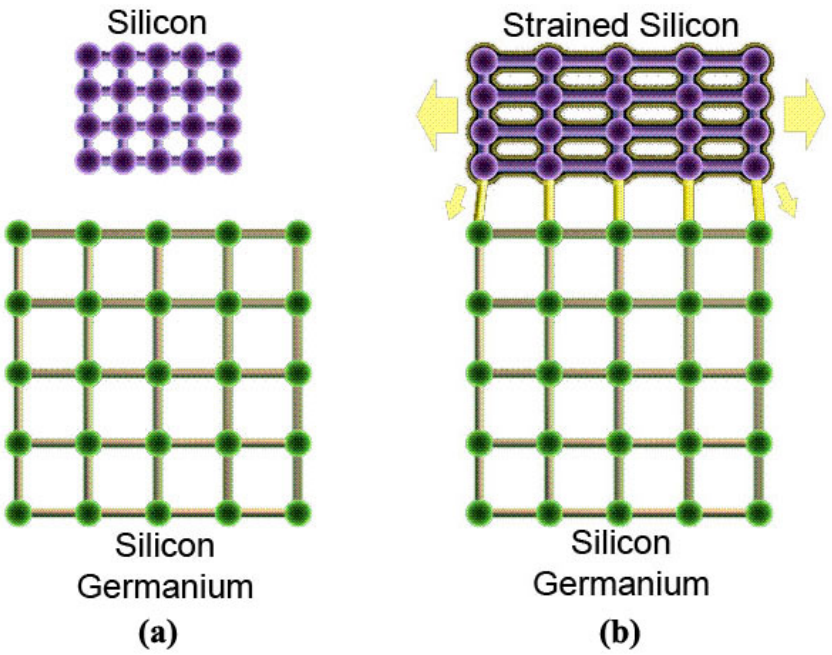


Figure 2.1: Schematic lattice structure of thin silicon deposited on silicon germanium substrate: (a) the SiGe lattice constant is larger compared to that of bulk Si; (b) the pure Si lattice attempts to line up with the SiGe lattice, which causes the Si to become tensilely strained. (Image courtesy of IBM, www.ibm.com)

where a_i denotes the cubic (unstrained) lattice constant of the film, and D^i is a constant that depends only on the elastic constant c^i_{11} and c^i_{12} of the respective material (in the above case, silicon):

$$D^i_{(001)} = 2\left(\frac{c_{12}}{c_{11}}\right). \tag{2.3}$$

So far we have discussed in detail the heterostructures of stained Si on SiGe. The argument for thin pseudomorphic SiGe layers deposited on Si bulk is very similar, except for that SiGe layers are under compressive strained to maintain pseudomorphic bonding with the Si substrate.

It has been shown in Fig. 2.1 that in order to exploit strained silicon, the strain-defining SiGe substrate has to be realized. Bulk SiGe can be ruled out, both because of the growth problems of pulling homogeneous SiGe crystals, and because such substrates would jeopardize the Si/SiGe system's compatibility with existing silicon tech-

nologies. It is necessary to employ silicon substrates and grow a relaxed, intermediate SiGe buffer layer.

For relaxed buffers as well as for pseudomorphic layers the most relevant material parameter is the critical thickness t_c [18], an equilibrium parameter at which strain relaxation by the generation of misfit dislocations can commence. When the strained film thickness exceeds t_c, misfit dislocations become energetically favorable, and provide partial strain relaxation of the film. By using low-temperature epitaxy techniques, such as molecular beam epitaxy (MBE) or chemical vapor deposition (CVD), one can also grow pseudomorphic SiGe films above the equilibrium critical thickness on bulk Si. This is because a metastable range exists in which the nucleation and propagation of misfit dislocations is kinetically suppressed. Fig. 2.2 shows the three regimes labeled 'stable', 'metastable' and 'relaxed' as a function of Ge fraction x [19].

However, as there are many nucleation sites from the threading dislocations in a relaxed "virtual" substrate, in practice there is little metastable critical thickness for heterostructures grown on such substrates, such as tensilely strained silicon on relaxed SiGe buffers on a silicon substrate.

2.2.2 Band Alignment

For the very large-scale integration (VLSI) community, the whole point of growing heterostructures is the opportunity to manipulate the behavior of carriers through band engineering. The bands of Si and SiGe alloys are strongly affected by strain, and experimental data are available both for unstrained bulk SiGe alloys and for pseudomorphic compressively strained SiGe films on Si(100) substrates. The strain-induced heavy-hole/light-hole splitting also leads to a splitting of the valence band in strained SiGe. Fig. 2.3 shows the bandgap values against Ge fraction x [20].

This thesis is focused on strained Si on SiGe relaxed buffer heterostructures. The

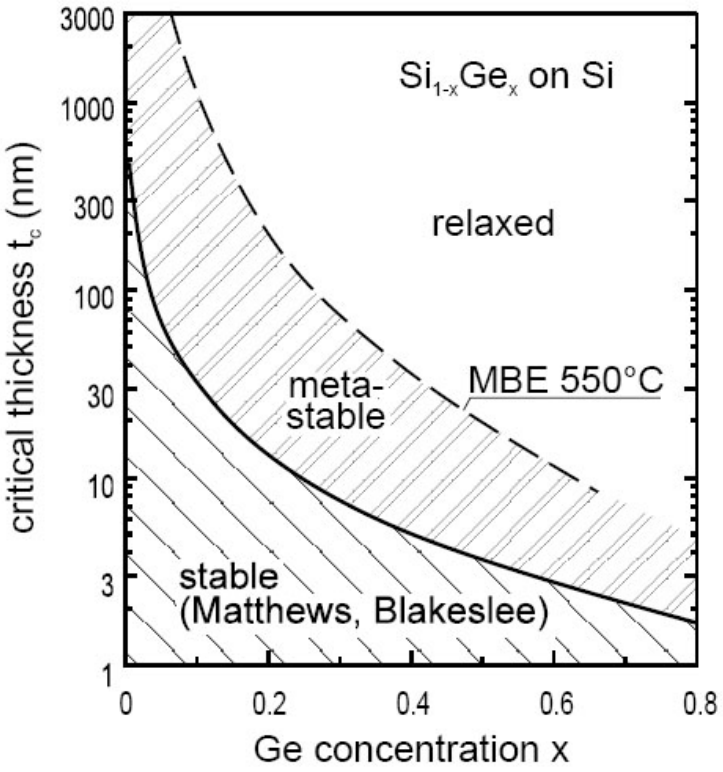


Figure 2.2: [19] Critical thickness against fraction for $Si_1 - xGe_x$ on Si. The lowest curve gives the theoretical limit in thermal equilibrium, whereas the experimental curve is for a metastable layer grown at 550 °C by MBE.

pseudomorphic Si/SiGe interface is commonly used as a quantum well to confine electrons. Such $Si/Si_{1-xs}Ge_{xs}$ heterojunctions are of type II for all value of x_s. Fig. 2.4 illustrates the valley splitting and allows predictions for arbitrary Si/SiGe heterojunctions with respect to band ordering and band offsets [19]. Throughout this thesis we will discuss the $Si/Si_{0.7}Ge_{0.3}$ system unless a different x_s is specified. A complete description of the band alignment can be calculated based on the local density functional and *ab initio* pseudopotentials [22]. In summary, the hydrostatic strain component leads to an overall downward shift of the average valence band and lowers the Δ and L band energies. The uniaxial strain component only splits the Δ conduction bands and the degenerate light/heavy hole valence band edges and leaves the weighted average positions of these bands unaffected. For the conduction band,

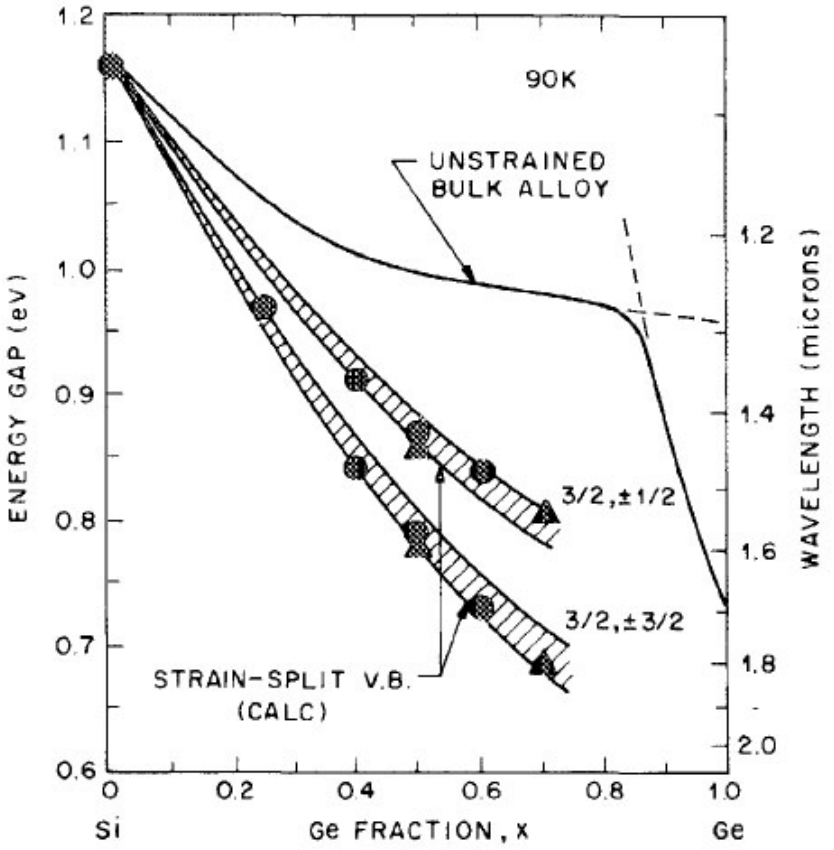


Figure 2.3: [20] Summary of energy-gap values of SiGe alloys, both unstrained bulk and strained SiGe on Si(100) substrate, at 90 K after correcting for quantum well shifts (circles = 75Å wells; triangles = 33Å wells). The double points at the same values of x correspond to a splitting of the valence band. The unstrained bulk alloy data are from optical absorption measurement by Braunstein [21].

the six-fold valley degeneracy in bulk silicon is lifted. The overall lowest conduction band is always the Δ_2 level in the active Si layer. Analytical fittings for the band offsets that will be used for numerical simulations can be expressed as [23]

$$\Delta E_v(x_s) = -0.238x_s + 0.03x_s^2, \tag{2.4}$$

$$\Delta E_c(x_s) = -0.35x_s - 0.35x_s^2 + 0.12x_s^3, \tag{2.5}$$

where $x_s < 0.85$, the negative signs indicate that the conduction and valence bands of strained Si are lower than those of the relaxed $Si_{1-x_s}Ge_{x_s}$ substrate.

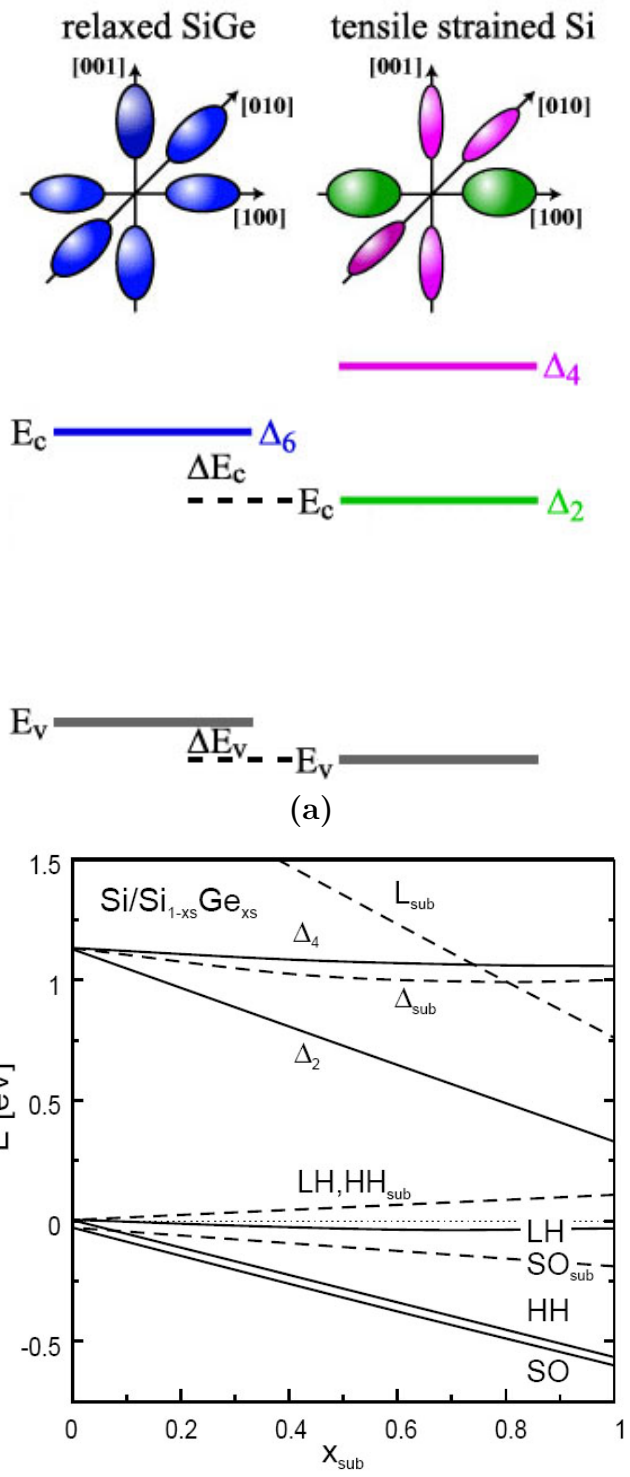


Figure 2.4: (a) The splitting of Si conduction bands in tensilely strained Si/SiGe heterostructures. (b) [19] Variation of the relevant silicon conduction and valence bands of a tensilely strained Si/SiGe heterostructure as a function of relaxed SiGe substrate composition x_{sub}.

The strain-induced band alignment also has strong effects on electron transport in strained silicon. The high electron mobilities realized in Si/SiGe heterostructures are of great interest for application in high performance CMOS as well as for quantum computation. The mechanism of mobility enhancement is well understood. For tensilely strained Si on SiGe substrate, the conduction band minimum lies in the silicon Δ_2. At low temperatures, the electrons only populate in the lower minima. The in-plane effective mass is now reduced to the transverse effective mass, which is only

$$m^* = m_T = 0.19\ m_0. \tag{2.6}$$

Besides the lower in-plane effective mass, the lifting of degeneracy also helps suppression of intervalley scattering. The enhanced electron mobilities and as a result, enhanced mean free paths in these structures have made a variety of transport physics experiments possible. One prominent example is the observation of the fractional quantum hall effect (FQHE) in an n-type Si/SiGe heterostructure [24]. To date, the highest low-temperature electron Hall mobility ever reported in strained Si is around 8.0×10^5 cm^2/Vs [25]. This has approached those in the GaAs/AlGaAs system to within a factor of 50. At room temperature, however, phonon scattering dominates and generally muffles the mobility. Still a 2DEG enhances the mobility by almost a factor of two over that of pure bulk silicon (2600 vs. 1450 cm^2/Vs) [26].

2.3 Si and SiGe Epitaxy at Princeton

2.3.1 Overview of the RTCVD Epitaxy

The Si/SiGe epitaxy for this thesis was performed in a custom-built rapid thermal chemical vapor deposition (RTCVD) apparatus at Princeton [27]. The reactor includes a load-locked, cold-wall quartz tube. A 100-mm silicon wafer is loaded and

supported by a quartz stand. It is heated from underneath by a bank of twelve 6-kW tungsten-halogen lamps. The process gas flows are adjusted by mass flow controllers. Three silicon precursors are available: dichlorosilane (or DCS, SiH_2Cl_2), silane (SiH_4, 10% in argon mixture), disilane (Si_2H_6, 10% in hygrogen mixture). Germane (GeH_4, 0.8% in hydrogen mixture) is used as the germanium precursor. Diborane (B_2H_6, 20 ppm diluted in hydrogen) and phosphine (PH_3, 100 ppm diluted in hydrogen) are added for *in situ* doping in both Si and SiGe growth. Unreacted exhaust gases are handled by a burnbox/scrubber equipment. The burnbox operates at around 850 °C to assure a thorough oxidation of the effluent. The hot treated gases are passed through a water recirculation tank for cooling and removal of residual reactor gases and particles.

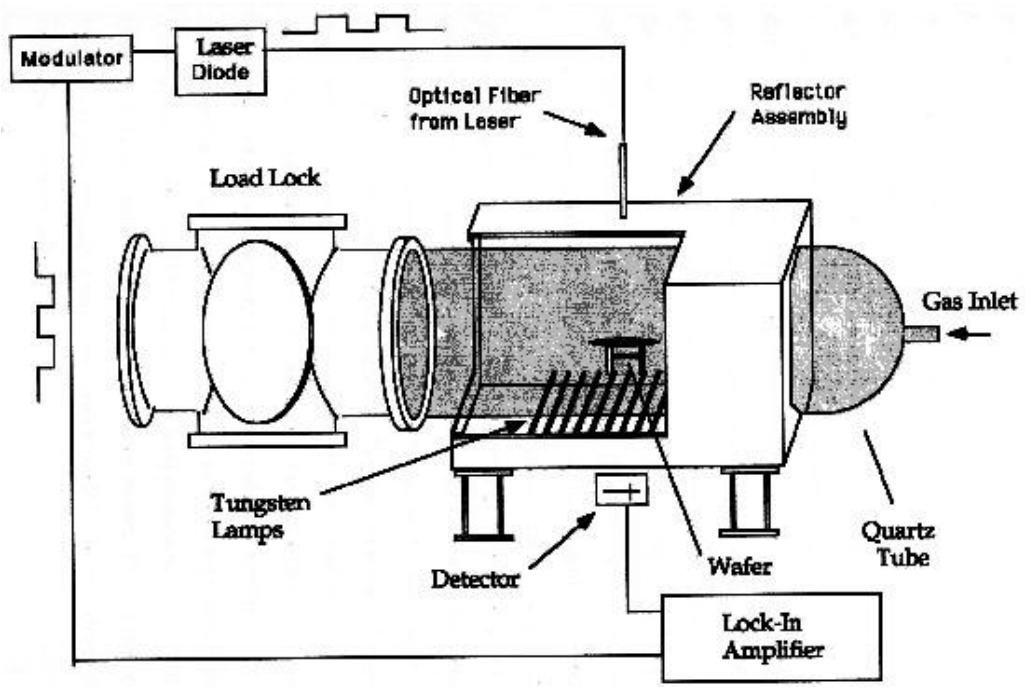


Figure 2.5: Schematic of Princeton RTCVD system used in this thesis. (Image courtesy of P. V. Schwartz [28])

Typical epitaxial growth conditions are at a pressure of 6 Torr with 3 standard liters per minute (slpm) hydrogen carrier flow. Under these conditions the growth surface is hydrogen-terminated. In this susceptor-free reactor, the silicon temperature

is measured by infrared absorption [29]. Two semiconductor lasers at 1.30 and 1.55 μm are coupled into a common fiber. The transmission is measured using lock-in amplifier techniques. A feedback loop controls the lamp power for accurate temperature control. The RTCVD system has been shown capable of growing high quality Si/SiGe layers on a 100 Å scale with an interface abruptness on the order of 10 Å, which is essential for quantum computing applications. Next we will discuss the Si/SiGe epitaxial RTCVD growth in more details, as well as to learn the prospects and limitations of this system.

2.3.2 The Use of Commercially-available SiGe Relaxed Buffers

We obtained our SiGe relaxed buffer from AmberWave Systems (see also www. amberwave.com), one of the semiconductor industry's leading suppliers of strained silicon technology. From the perspective of epitaxial growth, the growth conditions of relaxed buffers of a few microns thick and the thin modulation-doped layers can be quite different. High temperatures typically above 1000 °C are desirable for relaxed buffer growth to enable simultaneous improvements in both dislocation density and growth rate. The growth optimization is focused on high throughput on a large wafer scale. In contrast, optimization for the growth of high-mobility heterostrucures often leads to a contrary requirement. The layers are usually grown at low temperatures ($\sim$ 550 - 750 °C) for precise control of the thin layer thickness and interface abruptness. The typical sample dimension is around 1 cm for magneto-transport experiments at liquid helium temperatures. Therefore, it is difficult for a single system to grow both SiGe relaxed buffers and the modulation-doped heterostructures. We are among the first to use commercially available SiGe relaxed buffers for Si/SiGe modulation-doped heterostructures in academic research. The suitability of such applications is demonstrated.

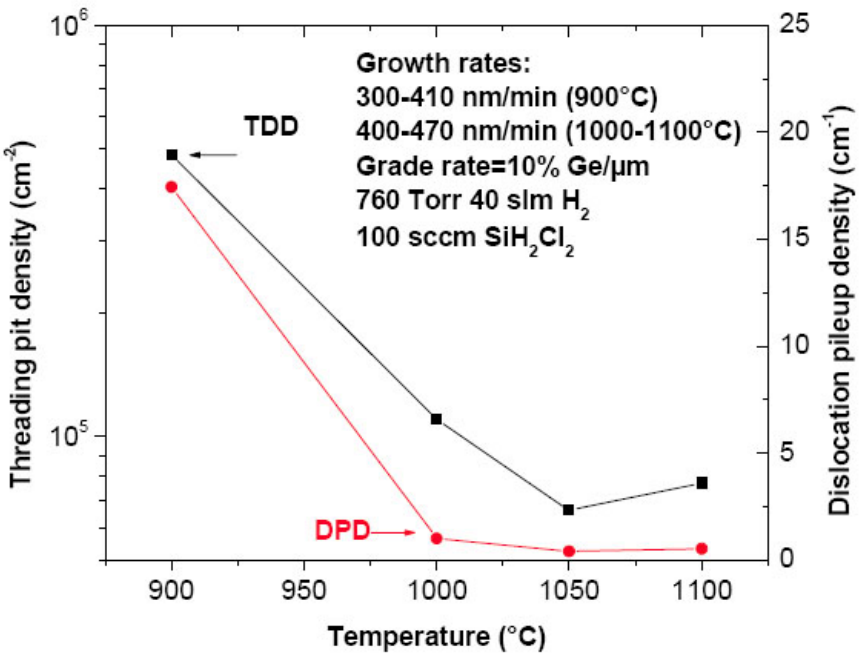


Figure 2.6: [30] Effect of temperature on AmberWave graded buffer threading dislocation density (TDD) and dislocation pileup density (DPD), determined by etch pit density and plan view TEM, respectively.

AmberWave Systems has developed a novel, high quality SiGe graded buffer growth process using $GeCl_4$ [30]. The use of the new germanium precursor enabled previously unattainable growth temperatures and growth rates. The chlorine component also can reduce parasitic deposition on the reactor chamber walls. Fig. 2.6 shows the effect of temperature in their system while maintaining high growth rates. Normally their growth procedure is as follows:

1. Start with 200-mm silicon substrates, either heavily or lightly doped depending on the applications;

2. Grow an undoped linearly-graded SiGe buffer at 10% germanium per micron;

3. Grow an undoped SiGe cap layer at the final Ge content, approximately 2 μm thick;

4. (optional) Chemical mechanical polishing (CMP) of the surface, to remove cross-

hatch roughness.

In this work three kinds of SiGe relaxed buffers from AmberWave Systems were used. Table 2.2 summarizes the parameters of these buffers. The use of CMP to eliminate cross-hatch patterns on relaxed SiGe buffers is illustrated in Fig. 2.7. The polished surface has a roughness RMS of 5.7 Å which is about ten times smoother than the as-grown surface. However the surface roughness has nearly no measurable effect on the quality of quantum wells growth on top, the electrical quality of quantum well samples grown on polished and unpolished buffers from the same growth were identical.

Table 2.2: Summary of AmberWave Systems SiGe relaxed buffers.

Label	AW2L	AW3H	AW3L
Substrate doping	(p-type) lightly	(n-type) heavily	(p-type) lightly
Ge content	20%	30%	30%
Total buffer thickness (μm)	4	5	5
Surface polishing	CMP	CMP	none

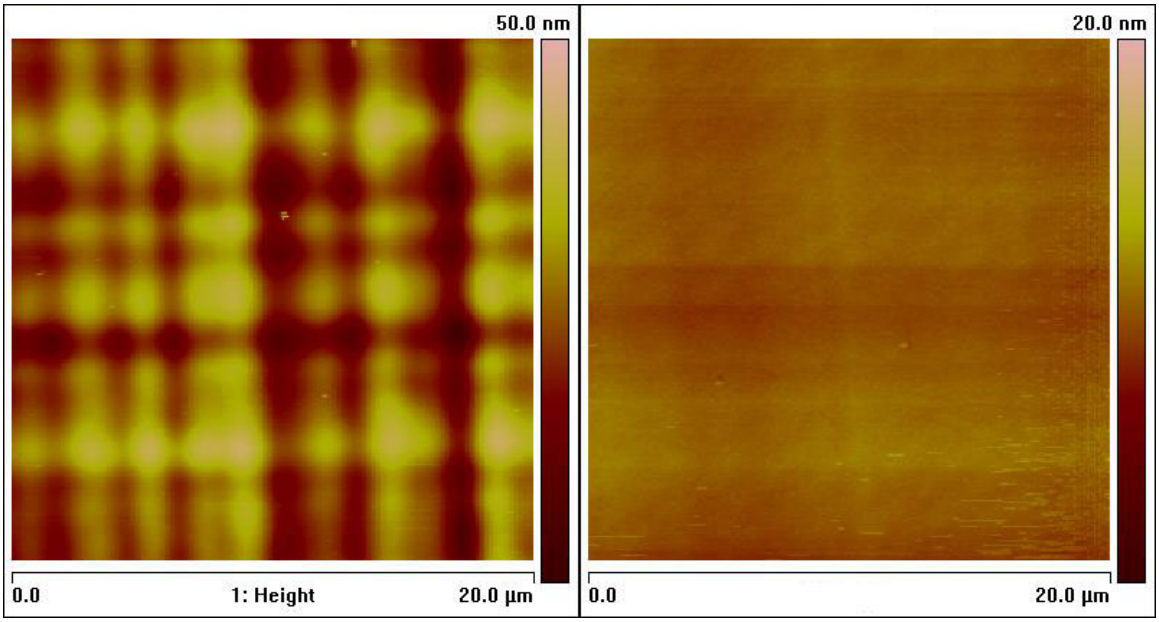


Figure 2.7: Top-view AFM images of: (left) AW3L SiGe relaxed buffers with cross-hatch patterns, surface RMS = 7.05 nm; (right) AW3H SiGe relaxed buffers after polishing, surface RMS = 0.57nm.

Since the Princeton RTCVD system is designed for 100-mm wafers, the 200-mm wafers were first diced into 1×1 cm squares using a Kulicke & Soffa wafer dicing saw (model 982-6) so they can be loaded with support of a 100-mm carrier wafer. The special carrier wafer has 5 etch-defined recessed holes to fit the small samples, as shown in Fig. 2.8.

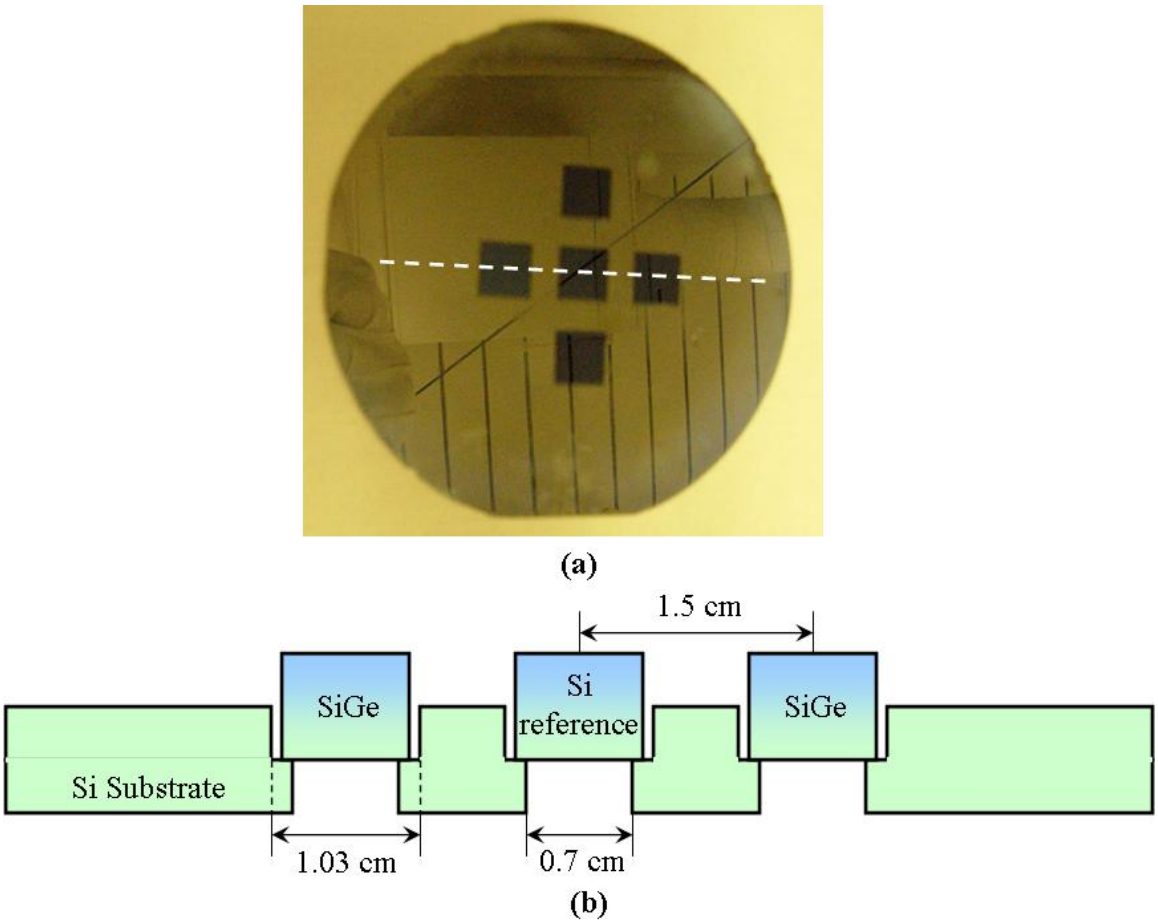


Figure 2.8: (a) A picture of 5 1×1 cm square pieces loaded on a 100-mm carrier wafer used for RTCVD growth. (b) A schematic view of cross section along the dashed line in (a) showing the dimensions of recessed holes.

2.3.3 Infrared Absorption of Si and SiGe

Since in our CVD the growth temperature is inferred by measuring the transmission of infrared through the center of a 100-mm wafer, the effects of the heavily doped

substrate as well as SiGe buffers have to be evaluated first.

For any silicon wafer with known thickness d and doping level n (or p), the system utilizes a single variable, normalized transmission (denoted by $t(T)$), to measure the temperature. Ignoring any change with temperature in the fractional power transmitted at the air-silicon interface [31], the normalized transmission will depend on the wafer thickness and material absorption coefficient:

$$t(T) = e^{-(\alpha(T)-\alpha(NT))d}, \tag{2.7}$$

where NT stands for normalization temperature, which is taken as room temperature. Sturm et al. presented analytical expressions for near-infrared absorption in silicon [29]. The absorption proceeds predominantly by two process: valence band to conduction band transitions and by free carrier absorption.

$$\alpha(T) = \alpha_{BG}(T) + \alpha_{FC}(T). \tag{2.8}$$

Here we only give the final results for the sake of simplicity. For the bandgap absorption

$$\alpha_{BG}(h\nu, T) = \sum_{i=1}^{2}\sum_{j=1}^{2}(-1)^j \frac{\alpha_i \left[h\nu - E_g(T) + (-1)^j k\theta_i\right]}{exp\left[(-1)^j\theta_i/T\right] - 1}\ cm^{-1}, \tag{2.9}$$

$\alpha_1(E)$ and $\alpha_2(E)$ represent absorption from the transverse acoustic and optical phonons, respectively.

$$\begin{aligned}
\alpha_1(E) &= 0.504\sqrt{E} + 392(E-0.0055)^2, E \geq 0.0055 \\
&= 0.504\sqrt{E}, 0 \leq E < 0.0055 \\
&= 0, E < 0, \\
\alpha_2(E) &= 18.08\sqrt{E} + 5760(E-0.0055)^2, E \geq 0.0055
\end{aligned} \tag{2.10}$$

$$= 18.08\sqrt{E}, 0 \leq E < 0.0055$$
$$= 0, E < 0, \tag{2.11}$$

where α is in units of cm^{-1} and E is in electronvolts. The temperature dependence of the silicon bandgap is

$$E_g(T) = E_g^0 - \frac{4.73 \times 10^{-4} T^2}{635 + T} \ eV. \tag{2.12}$$

The second mode, free carrier absorption coefficient can be calculated as

$$\alpha_{FC}(T) = n(T)\sigma_n(T) + p(T)\sigma_p(T), \tag{2.13}$$

where the two cross sections are given by

$$\sigma_n(T) = 1.01 \times 10^{-12} T\lambda^2 \ K^{-1}, \tag{2.14}$$
$$\sigma_p(T) = 0.51 \times 10^{-12} T\lambda^2 \ K^{-1}. \tag{2.15}$$

The electron and hole concentrations n and p depend on n_i in silicon according to the relationship

$$\sqrt{np} = n_i = 3.87 \times 10^{16} T^{3/2} exp\left(-(0.605 - 7.1 \times 10^{-10}\sqrt{n_i/T})/kT\right) \ cm^{-3}. \tag{2.16}$$

Two lasers with wavelengths of 1.30 and 1.55 μm are used in temperature range from 500 to 800 °C. It has been found that in lightly doped silicon, nearly all of the absorption at 1.30 μm proceeds by a band-to-band process due to bandgap narrowing at these temperatures ($\alpha_{BG}(T) \gg \alpha_{FC}(T)$), while at 1.55 μm free carrier absorption dominates ($\alpha_{BG}(T) \ll \alpha_{FC}(T)$). In practice, for standard 100-mm silicon wafers with

light substrate doping, we monitor 1.30 μm transmission for temperatures up to 650 °C then switch to 1.55 μm for up to 800 °C. Fig. 2.9 shows the original published data of normalized transmission over the temperature range. The slopes of the curves in their useful range (for example, 1.30 μm at 600 °C and 1.55 μm at 700 °C) are both around 3%/°C. This means if we assume a conservative estimate of error in transmission measurement to be $\sim \pm 10\%$, the temperature error will be $\sim \pm 3$ °C.

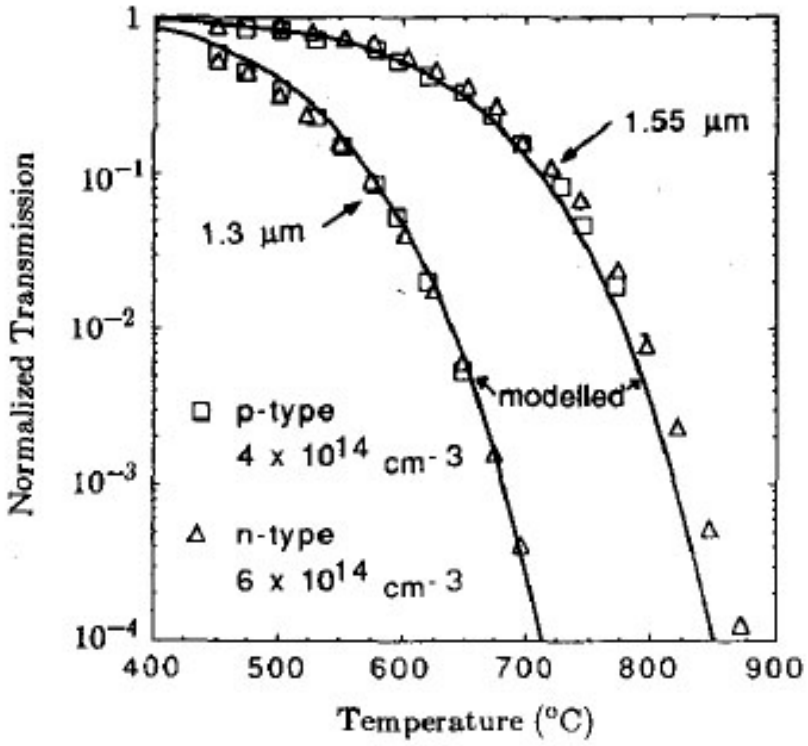


Figure 2.9: [29] Data of normalized transmission vs. temperature for 1.30 and 1.55 μm, for lightly doped <100> n-type (7.5 Ω·cm, thickness = 513 μm) and p-type (37 Ω·cm, thickness = 493 μm). The data have been adjusted to reflect a 500-μm thickness. For comparison, also presented are the model results for the n-type wafer.

There are least two potential problems caused by the use of AmberWave Systems substrates in our temperature measurement. First, the 200-mm starting silicon wafers are 725 μm thick, 50% more than the standard 100-mm wafers. Also the SiGe relaxed buffers cause extra absorption. To extend the usefulness of our existing normalized transmission models, we modify it to reflect the above changes. In particular, the

following rather simplified assumptions are made to incorporate the effect of SiGe layers:

1. The only different parameter when applying the above calculations of infrared transmission to SiGe is the reduced bandgap. All other effects such as the band splitting and change in the phonon-assisted bandgap transitions due to a random alloy are ignored. The reduction of $Si_{1-x}Ge_x$ bulk alloy bandgap is obtained as [32]

$$\Delta E_g = E_{Si} - E_{SiGe} = 0.43x - 0.206x^2 \ eV. \tag{2.17}$$

2. The linear graded SiGe buffer is treated approximately as a single-step relaxed layer with its final Ge content and half of its real thickness. As a result, a SiGe relaxed buffer consisting of a linear graded buffer of thickness d_{graded} and a uniform cap layer of thickness d_{cap} will have an "effective" thickness of

$$d_{SiGe} = d_{graded}/2 + d_{cap}. \tag{2.18}$$

3. We also ignore the power reflected at the Si/SiGe interface, since the two materials are very close to each other in nature and the Ge content change is gradual. In other words, the absorptions in Si and in SiGe are additive, the total normalized transmission is:

$$t(T) = e^{-(\alpha_{Si}(T) - \alpha(NT)_{Si})d_{Si}} e^{-(\alpha_{SiGe}(T) - \alpha_{SiGe}(NT))d_{SiGe}}. \tag{2.19}$$

Under these assumptions the normalized transmission of AmberWave Systems buffers is calculated. As an example, Fig. 2.10 shows the model for a 725 μm Si substrate plus a total of 3 (graded) + 2 (cap) μm $Si_{0.7}Ge_{0.3}$ buffer, so the effective SiGe

layer thickness used is 3.5 μm. The dotted line shows the model for the starting Si substrate only. The substrate doping was chosen to be 1×10^{15} cm^{-3} p-type. One finds out that overall the effect of the SiGe buffer is about 10% reduction in transmission at 1.30 μm and 5% reduction at 1.55 μm. The difference can be explained by noting that the reduced bandgap will have a stronger impact on band-to-band absorption, which is a more dominant mechanism at 1.30 μm.

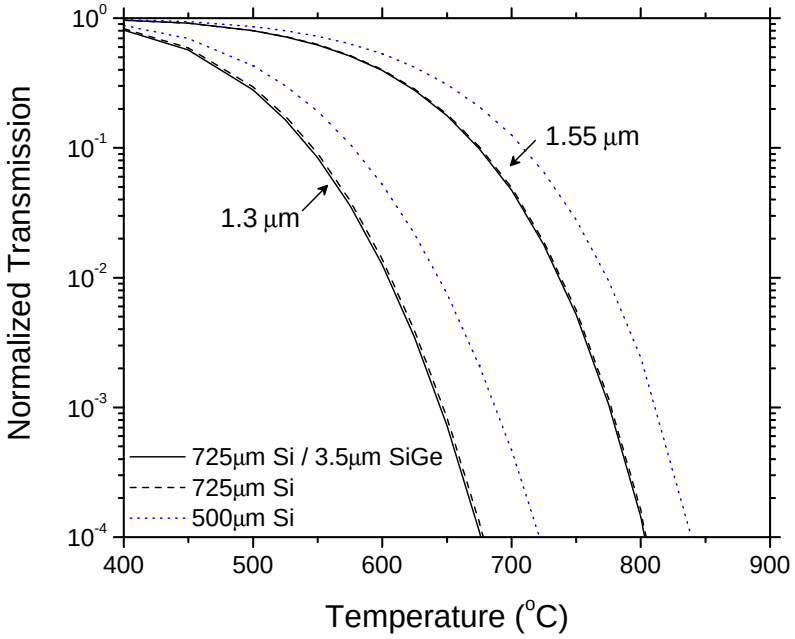


Figure 2.10: Model of normalized transmission vs. temperature for 1.30 and 1.55 μm, for lightly doped <100> p-type (1×10^{15} cm^{-3}, thickness=725 μm) plus a $Si_{0.7}Ge_{0.3}$ buffer (thickness = 3.5 μm). Also presented are the model results for a 500-μm thick silicon.

The slope of the 1.30 μm curve at 600 °C is now steeper, $\sim$ 4.5%/°C. This is better for our temperature control, as now we can achieve $\pm$2°C accuracy with 10% measured signal variation. However, this change in slope comes at a cost: the actual received signal is also about 75% weaker than that previously from the 100-mm Si wafer. The upper temperature limit with the 1.30 μm is around 650 °C. Measuring higher

temperature is difficult due to the weak transmission, which is subject to large error bars caused by electrical noise. Alternatively, we can switch to 1.55 μm transmission at these temperatures. The slope is about 1.5%/°C indicating an estimate of ±6°C error. For the 1.55 μm in the temperature range 700 - 750 °C, both the slope and the actual signal strength remain at a reasonable level. We can still achieve $\sim$ 3%/°C accurary with strong received signals.

So far we have only considered substrates with light doping (10^{15} cm^{-3}). In such cases Fig. 2.10 can be referred to find the desired normalized transmission set points. Physically, the moderate and heavily doped cases are expected to be very different because of the significant free-carrier absorption even at room temperature. The calculation shows that a heavily doped substrate indeed does have a very low normalized transmission at elevated temperatures. We decide the best practice to circumvent such difficulties is to load a regular lightly-doped Si reference piece in the center of a carrier wafer, while placing other samples around the outside of the carrier wafer. Empirical data shows that the temperature nonuniformity within 2 inches from the center of the carrier wafer is less than 5 °C.

For all growth on AW relaxed buffers described in this thesis, the thick Si/SiGe heterostructure samples were placed on a recess hole that is 1.5 cm away from the center, while a 500-μm thick silicon sample was placed in the center for temperature control, as previously shown in Fig. 2.8. The growth on four outside pieces on the perimeter is uniform. This method gives a repeatable temperature so that certain growth conditions may be reproduced.

2.3.4 Growth Rates and Doping Profiles

The successful epitaxy of Si/SiGe heterostructures, especially the structures designed for quantum computing applications, rely on precise knowledge of growth rates and doping profiles. Various historical data are available for our RTCVD reactor, yet

none are based on SiGe relaxed buffers. We performed calibrations with a series of carefully designed growth sequences and their secondary ion mass spectroscopy (SIMS) analysis. All SIMS analysis presented in this thesis were prepared by Evans Analytical Group at Hightstown, New Jersey.

The Si/SiGe epitaxy is performed at temperatures between 550 and 700 °C. A hydrogen background (3 slpm flow at 6 Torr) is used. DCS is the preferred silicon precursor, as the deposition is more selective and hence reduces the parasitic coating of reactant byproducts on the reactor walls. However, we also find that disilane growth at lower temperature can achieve similar growth rates with lower phosphorus background impurities. So data using disilane as precursor are also measured. Fig. 2.11 summarizes the calibration results for SiGe layers growth using various combinations of gas mixtures. From the growth calibration data, for example, one should use a gas mixture of 26 sccm DCS / 1.8 sccm GeH_4 at 625 °C or 25 sccm Si_2H_6 / 4 sccm GeH_4 at 575 °C to grow $Si_{0.7}Ge_{0.3}$ films.

Another critical control parameter for successful heterostructure growth is the doping level in doped layers. For n-type modulation-doped Si/SiGe heterostructures, we are particularly interested in phosphorus doping profiles in SiGe layers, either as intentional dopants or background impurities.

Fig. 2.12 shows the phosphorus doping levels obtained in $Si_{0.7}Ge_{0.3}$ films by *in situ* PH_3 doping. By comparing the two silicon precursors, we find that the phosphorus incorporation is in general higher for a given phosphine flow in SiGe layers grown with DCS than with disilane. This might also contribute to the higher background impurity level seen in DCS growth. With a DCS precursor it is very difficult to achieve *in situ* doping levels below 10^{19} cm^{-3}. Note however the phosphorus doping level does not attain its steady state value immediately when the phosphine is switched on, a typical transient region of a few nanometers was observed in historical SIMS profiles. Thus in the thin supply layer of a few nanometers SiGe, the doping level may be

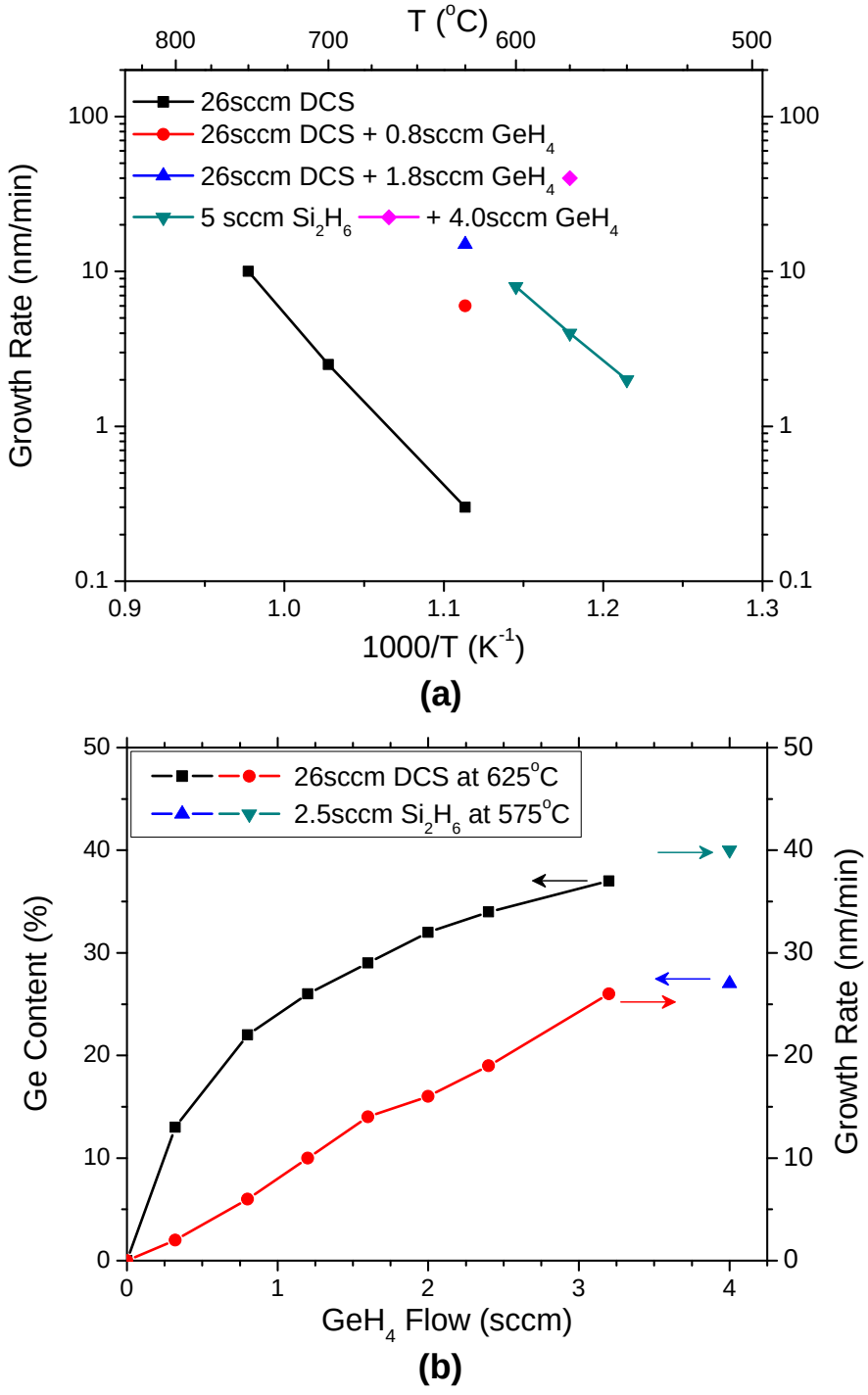


Figure 2.11: Growth rate and Ge content of SiGe layers using DCS and disilane at different temperatures and gas flow rates.

lower than the steady state value from the figure.

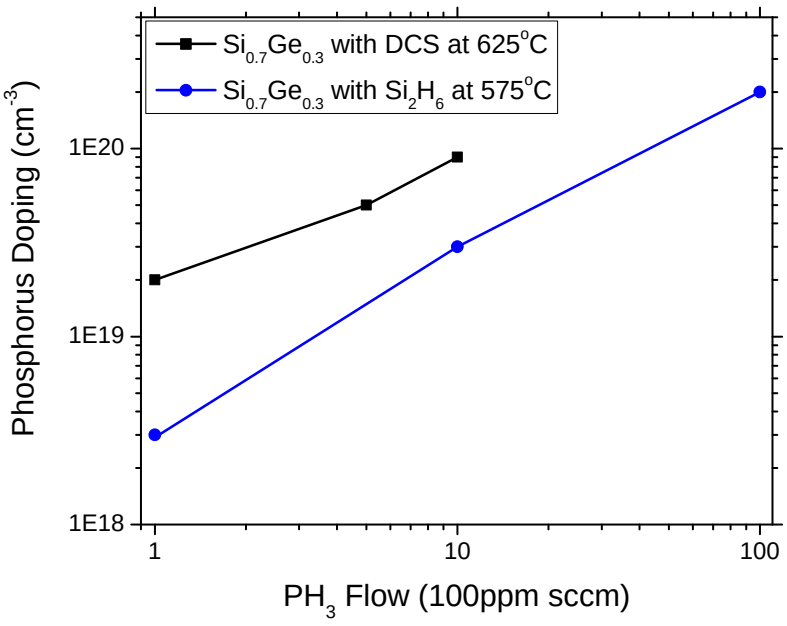


Figure 2.12: N-type doping level in $Si_{0.7}Ge_{0.3}$ vs. phosphine flow (100 ppm in hydrogen) layer grown at two different conditions.

2.4 Summary

Lattice-mismatched Si/SiGe heterostructures can form pseudomorphic layers under a certain temperature-dependent critical thickness. The effects of strain on both the band structure and the band offsets have been reviewed. Band engineering through the strain adjustment has spearheaded both silicon device applications and understanding in mesoscopic physics. In particlar, we are concerned with the tensilely strained silicon on SiGe relaxed buffers for their usefulness in confining electrons. To explore these opportunities, epitaxial silicon and silicon-germanium layers are grown by RTCVD at Princeton. High quality commercially available SiGe relaxed buffers are integrated in our experiments for the growth of thin heterostructure layers. General

growth issues such as infrared absorption for accurate temperature control, calibration of growth rates and doping profiles have been addressed. The rest of this thesis will be dedicated to the growth and applications of Si/SiGe heterostructures.

Chapter 3

The Two-dimensional Electron Gas in Strained Silicon

3.1 Introduction

The two-dimensional electron gas (2DEG), formed by employing modulation-doped Si/SiGe heterostructures, is a very important low-dimensional system for electronic transport. It is the core of a field-effect transistor, which goes by many acronyms including modulation-doped field-effect transistor (MODFET) and high electron mobility transistor (HEMT). The silicon metal-oxide-semiconductor field-effect transistor (MOSFET) is perhaps the most common electronic device, with holes or electrons trapped in an inversion layer at the Si/SiO_2 interface. Since the first observation of the modulation doping effect in Si/SiGe grown by MBE [33], considerable research involving 2DEG has been done, and much continues to this day. The 2DEG offers a mature system of potentially high mobility electrons, especially at low temperatures. These enormous mobilities offer a test bed for exploring fundamental physics, as well as a single figure of merit for the overall growth quality.

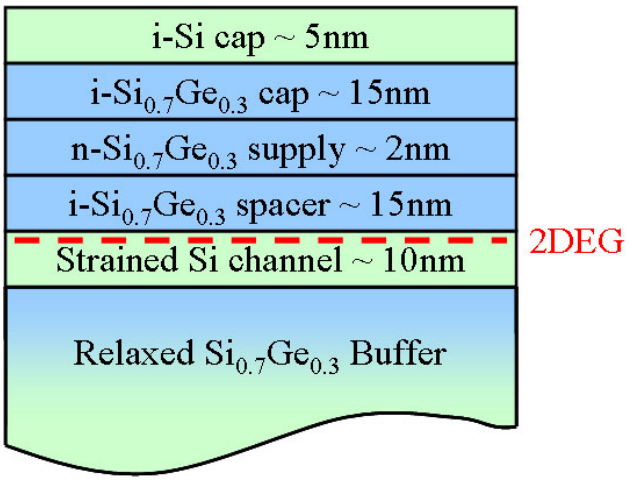
Fig. 3.1 shows the structure for a typical 2DEG in strained Si/SiGe and the

associated band diagram. It has the following layers, starting from the substrate and going up as the direction of growth:

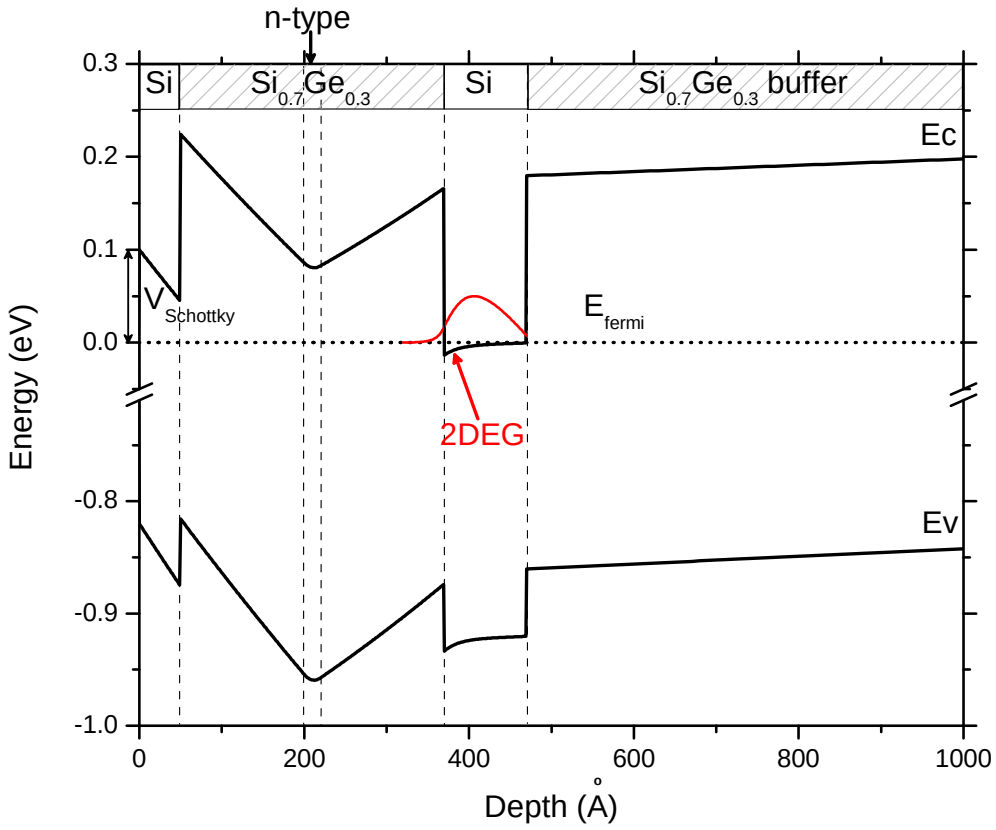
1. Relaxed SiGe buffer layer, to induce tensile strain in silicon;
2. Strained Si channel for the electrons, undoped;
3. SiGe spacer of thickness s, undoped, to separate the ionized dopants from the channel;
4. SiGe doped layer of thickness d and doping N_D (heavily n-type), possibly a monolayer in δ-doped material;
5. SiGe and a very thin Si cap. A metal gate may be deposited on top, which is used to tune the potential of the quantum well.

In the band diagram, we assumed a doping level of 5×10^{18} cm^{-3} in the supply layer. Other materials parameters and calculation procedures can be found in Appendix B and in [23]. Ideally the doping level and surface potential were designed such that the lowest subband in the doping level is located above the Fermi level. Therefore at low temperature, free electrons exist only in the channel, and occupy only the lowest subband lying above the Δ_2 level. For reasons of clarity, only the Δ_2 bands in silicon are plotted. Another important parameter in the band structure is the surface potential $V_{Schottky}$, which will be sensitive to the top surface states/defects. This is not a problem in top-gated structures, in which the surface potential can be tuned continuously.

The one hurdle that has been severely impeding the enhancement of mobilities in strained silicon 2DEG's is the availability of high quality buffer layers. Fig. 3.2 shows the evolution of published Hall mobilities from various publications [25]. Relaxed SiGe buffer layers on Si substrates (also referred to as *virtual substrates*) are relatively thick, since a high degree of relaxation and low defect densities are required.



(a)



(b)

Figure 3.1: (a) A typical structure of modulation-doped layers on a relaxed SiGe buffer layer. (b) The band diagram shows a type-II band alignment with the electrons confined in the tensilely strained Si channel. The surface potential is assumed so that $V_{Schottky} = -0.1$ V. The electron wave function is plotted in red (in arbitrary units).

In the first few years of SiGe buffer layers growth people used a single-step technique, which consists of growing a constant-composition SiGe layer with a thickness far exceeding the critical thickness [41]. Such buffer layers are associated with very high threading dislocations penetrating throughout the growth. The graded buffer technique has been developed to overcome the problems. It employs a linear Ge gradient throughout initial buffers, followed by a final layer with constant Ge composition. The graded buffers have led to a significant enhancement of the mobilities. Further improvements of the relaxed buffer layers aiming towards device applications have been made over time. In the past few years, with progress in chemical mechanical polishing (CMP) [42] and high temperature growth process, atomically flat relaxed buffers with 10^5 cm^{-2} threading dislocations became commercially available [30]. Efforts will continue for optimization of graded buffers both with respect to relaxation and surface morphology.

Before we move to the experimental results, I would like to emphasize the important role of 2DEG in quantum computing. In the Loss-DiVicenzo proposal [14], free electrons are used as a vehicle for information. The Si/SiGe 2DEG appears to be a particularly promising host. Electrons are confined in quantum dots created by lateral confinement of a 2DEG. The mobility is also essential, as it directly defines the mean free path which affects quantum point contacts (QPC) and other single electron phenomena, which may be used as part of quantum dot devices.

3.2 Characterization of the 2DEG's

3.2.1 Sample Structures

Two samples will be discussed as examples of 2DEG's in modulation-doped heterostructures. Sample #3996 was grown on 20% Ge relaxed buffers AW2L, and #4736 was grown on 30% Ge relaxed buffers AW3L. The AW3L susbstrates were not

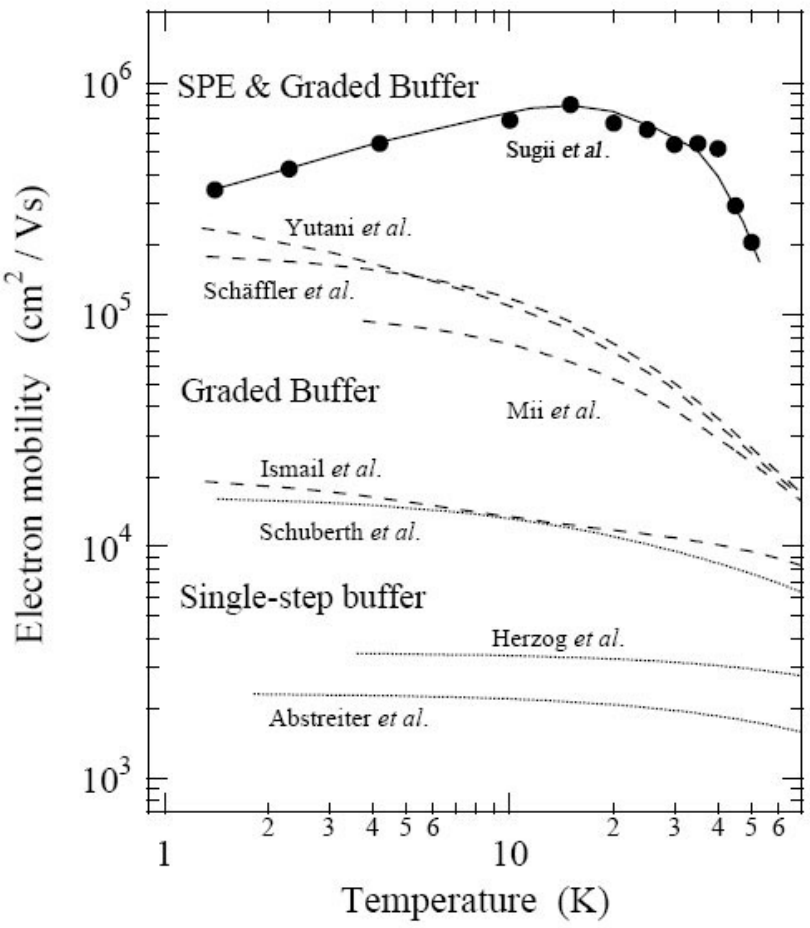


Figure 3.2: [25] Temperature dependences of electron mobility for a Si/SiGe heterostructure with a graded buffer layer ($0 \leq x_s \leq 0.2$) from literature [34, 35, 36, 37, 38, 39, 40]. Dotted lines are samples with a single-step buffer layer; dashed lines are samples with a graded buffer layer; solid line uses MBE combined with solid-phase epitaxy (SPE).

polished therefore have cross hatch patterns on the surface. The AW3H buffers with smooth surface were grown on heavily-doped starting Si. This has caused the diffusion of dopants into the SiGe buffers leading to excessive substrate leakages, which will be discussed in the next Section 3.2.2.

The samples used in growth are 1×1 cm squares resting on a 100-mm silicon carrier wafer. Prior to the loading, the surface is cleaned by a standard wet clean at room temperature in a H_2SO_4/H_2O_2 (7:3) solution for 15 minutes followed by a highly diluted HF (1:1000) dip for 2 minutes. After the wafer is transferred to the

chamber through the nitrogen-purged load lock, the chamber undergoes an *in situ* hot bake process with 5 slpm H_2 at 6 Torr at 900 °C for two minutes. The H_2 flow is then reduced to 3 slpm at the same pressure for growth. 26 sccm SiH_2Cl_2 is used for Si growth at either 700 or 750°C. GeH_4 is added for SiGe growth at 625 °C. The GeH_4 (0.8% in hydrogen) flow rates for $Si_{0.8}Ge_{0.2}$ and $Si_{0.7}Ge_{0.3}$ are 100 and 225 sccm, respectively. The PH_3 flow rate for doping in the supply layer is 2 sccm (100 ppm in hydrogen). Fig. 3.3 shows the nominal layer structures for both samples.

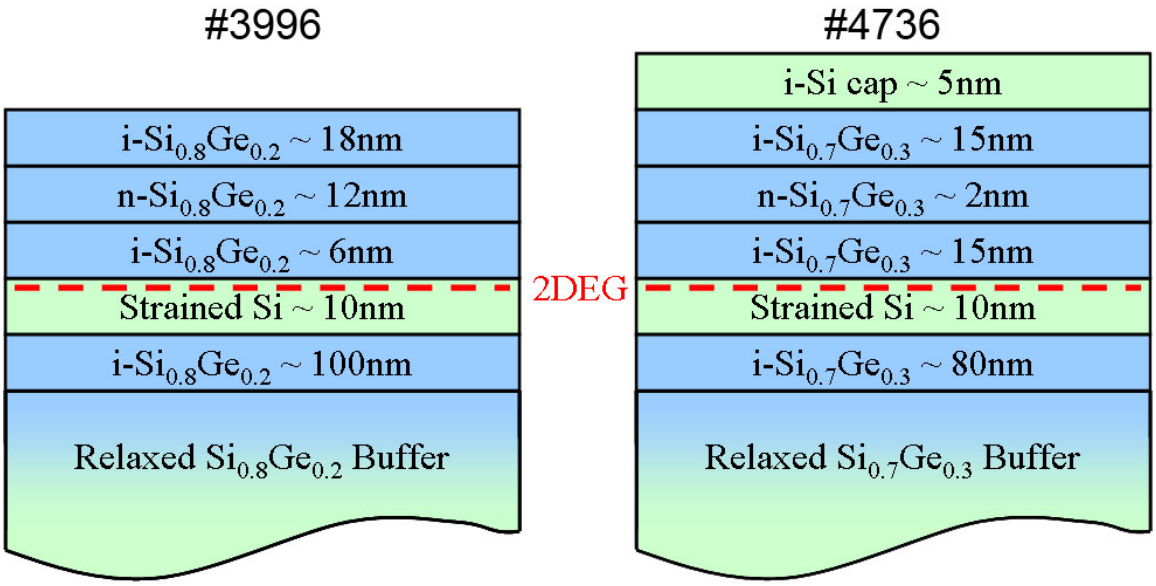


Figure 3.3: Layer sequences for the two 2DEG samples #3996 and #4736. The n-type doping level is $\sim 10^{19}$ cm^{-3} in both samples; grown with DCS as in Fig. 2.12

For processing, Hall bars are defined by optical lithography and etched by reactive ion etching (RIE) using a gas mixture of CF_4/O_2 in a PlasmaTherm 720 SLR Series system. Ohmic contacts are made to the channel by Au:Sb (1% antimony) evaporation, lift-off, and subsequent annealing at 430 °C for 10 minutes. For sample mouting, gold wires were bonded by soldering indium to the contact pads for external electrical measurements. Both samples show low-temperature Hall mobilities of around 10,000 cm^2/Vs at liquid helium temperatures 4.2 K and below. Sample #3996 has a 2-D electron density of 2.2×10^{12} cm^{-2} and #4736 has 0.96×10^{12} cm^{-2}. At present, the

mobility in these samples is limited by the high background phosphorus doping in the silicon ($\sim 10^{17}$ cm^{-3}), which also contributes to the high 2-D density. The background doping effect will be discussed in more details in Section 3.3.

3.2.2 Effect of Substrate Doping in Relaxed SiGe Buffers on Leakage

We have demonstrated the successful use of commercially available relaxed SiGe buffers in Princeton RTCVD system for strained Si/SiGe 2DEG. However, all samples grown on one particular batch of buffers, AW3H, have shown large substrate leakage even at liquid helium temperatures. We suggest that the substrate doping (arsenic) contributes to leakage current in relaxed buffers if the starting Si substrate is heavily doped.

In our measurement the substrate leakage is defined as current from the top 2DEG Hall bar metal contacts to the substrate. The substrate contact is a gold wire bonded to the backside of the sample. Fig. 3.4 shows a strong dependence of leakage current on the substrate doping at T = 4.2 K. Samples grown on both AW3H and AW3L were compared. For the heavily-doped Si substrate, the leakage current is in the μA range even with only a few hundred mV applied. Such substrate shorting makes not only side gating of nanostructures in the 2DEG impossible, but also prevents back gating of the 2DEG by applying a substrate voltage. However if a lightly-doped Si substrate is used, the leakage is only in the pA range, up to applied voltages of $\pm$30 V before breakdown. These two otherwise identical 2DEG structures were grown at less than 800 °C as described before, a thermal budget that does not cause significant dopant diffusion in silicon. The only other high temperature step was the growth of relaxed SiGe buffers at AmberWave systems, which was above 1000 °C.

To understand the origin of the substrate leakage, the doping profiles of the arsenic-doped substrates and SiGe relaxed buffers were analyzed by SIMS, as shown

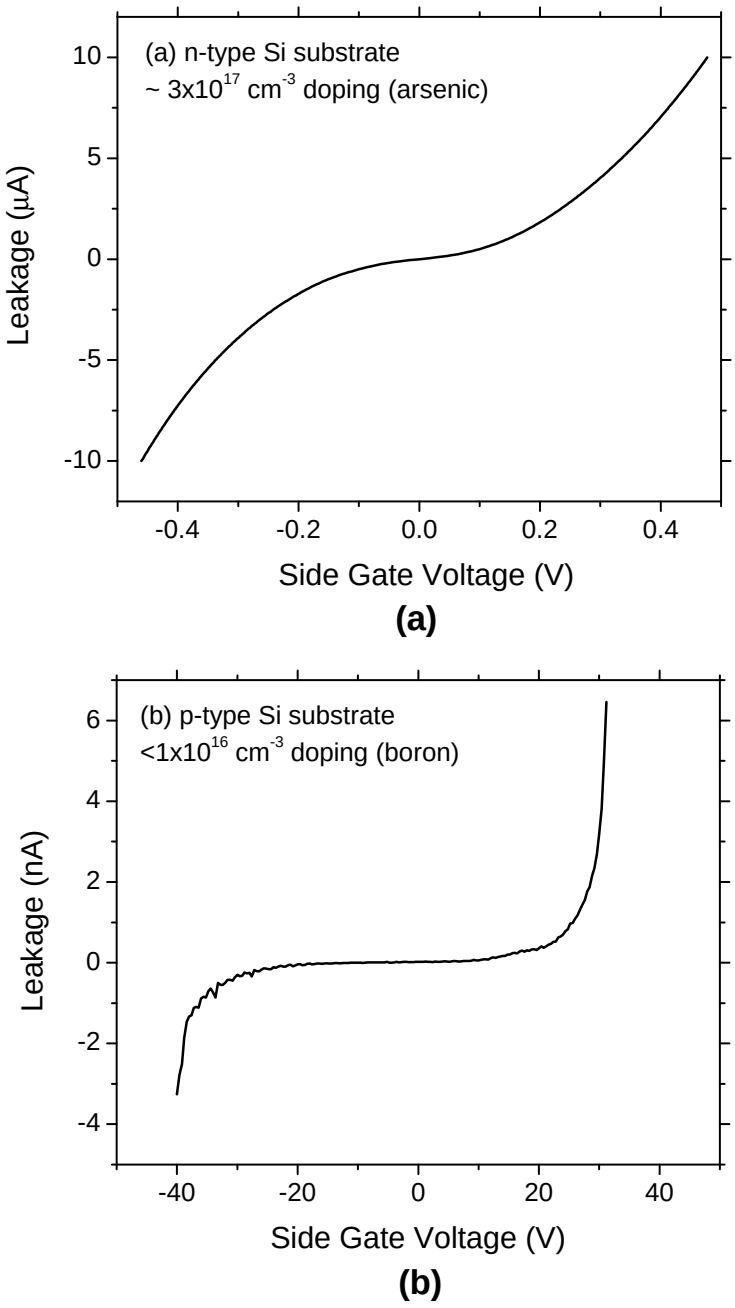


Figure 3.4: I-V curves from the 2DEG Hall bar mesa through the SiGe buffers to a back substrate contact for samples starting on (a) AW3H, a heavily n-type doped (arsenic) and (b) AW3L, a lightly p-type doped (boron) silicon substrates. Note the μA scale in (a) and nA scale in (b).

in Fig. 3.5. We found that during the high-temperature relaxed buffer growth, As diffused through the whole SiGe buffers. Assuming a diffusion length of $\sim$ 5 μm, we can extrapolate the As diffusivity in $Si_{0.7}Ge_{0.3}$ to be about 6×10^{-11} cm^2/s at T = 1000 - 1100 °C, more than three orders magnitude higher than the intrinsic As diffusivity in Si. As diffusion in relaxed SiGe has been studied in the past using ion-implantation method [43, 44], and typical diffusivity for As in $Si_{0.7}Ge_{0.3}$ at T = 1000 °C was reported to be in the mid 10^{-13} cm^2/s range. However, in these previous works As was implanted into the top SiGe surfaces where the dislocation density is low. Our result suggests that when As diffused from the starting Si substrate through the graded buffers with higher threading dislocation densities, the diffusion was further enhanced and penetrated the SiGe relaxed buffers. Similar leakage mechanism caused by enhanced dopant diffusion near misfit dislocations was also reported in the strained Si MOSFET on a SiGe relaxed buffer substrate [45]. Even in pure silicon, since the early days of bipolar transistor technology, the very fast dopant diffusion through threading dislocations in epitaxy was well known to form emitter-collector shorts or "pipes" [46, 47]. Two main mechanisms have been suggested, the fast diffusion of impurities down the dislocation cores and the enhanced diffusion of impurities to an enhanced vacancy concentration caused by climbing or interacting dislocations.

In addition to the enhanced As diffusion in the relaxed SiGe buffers, we suggest that arsenic segregates to the dislocations. Since the average As doping level in our buffers was $\sim 3 \times 10^{17}$ cm^{-3}, lower than the Mott metal-insulator transition level, for the dopants in the buffers to conduct at low temperatures (T < 4.2 K) there must be regions with an effective doping level that exceeds the Mott level. We suggest that a possible explanation is localized As segregation, which has also been studied in silicon by several groups [48, 49].

To avoid substrate leakage, we use only AW3L SiGe relaxed buffers grown on lightly-doped Si substrates ($< 1 \times 10^{16}$ cm^{-3}) for strained Si 2DEG and quantum dot

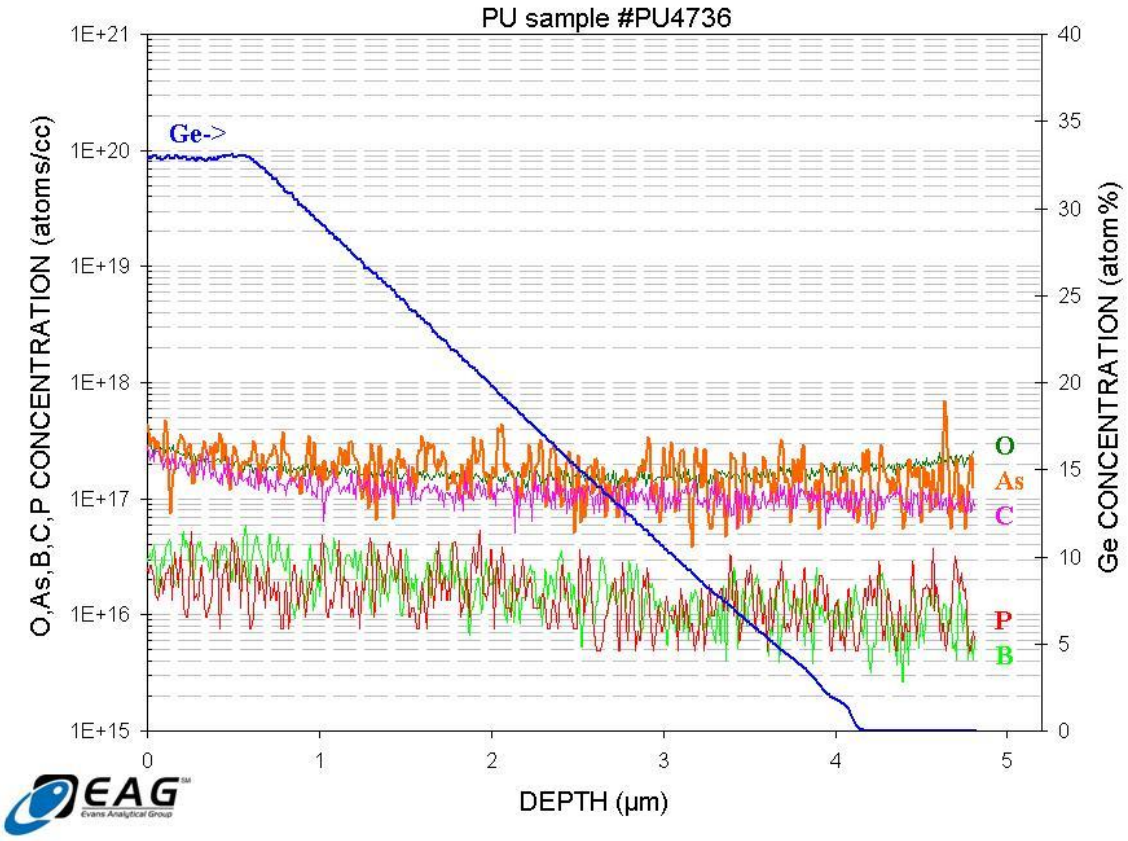


Figure 3.5: SIMS analysis of AW3H substrate showing the starting Si substrate, the linearly graded SiGe buffer, and the uniform $Si_{0.7}Ge_{0.3}$ buffer. The growth temperature of the relaxed SiGe buffer is over 1000 °C.

applications. With low substrate doping, we achieved successful gating and surface passivation, which will be covered in later chapters of this thesis.

3.2.3 Magneto-transport Properties

To further study the transport properties of the electron gases, magneto-transport measurements are taken in a closed-cycle He-3 refrigerator with superconducting magnet. These measurements were performed through collaboration with Professor Leonid Rokhinson's laboratory at Purdue University. As an example, Fig. 3.6 shows the results of sample #4736 in a magnetic field up to 8 Tesla. Well-defined Shubnikov-deHaas (SdH) oscillations and quantized Hall plateaus are observed. Also

from the Fourier transform of the oscillations in reciprocal field, we see only a single oscillation period. This confirms that only the lowest subband is occupied.

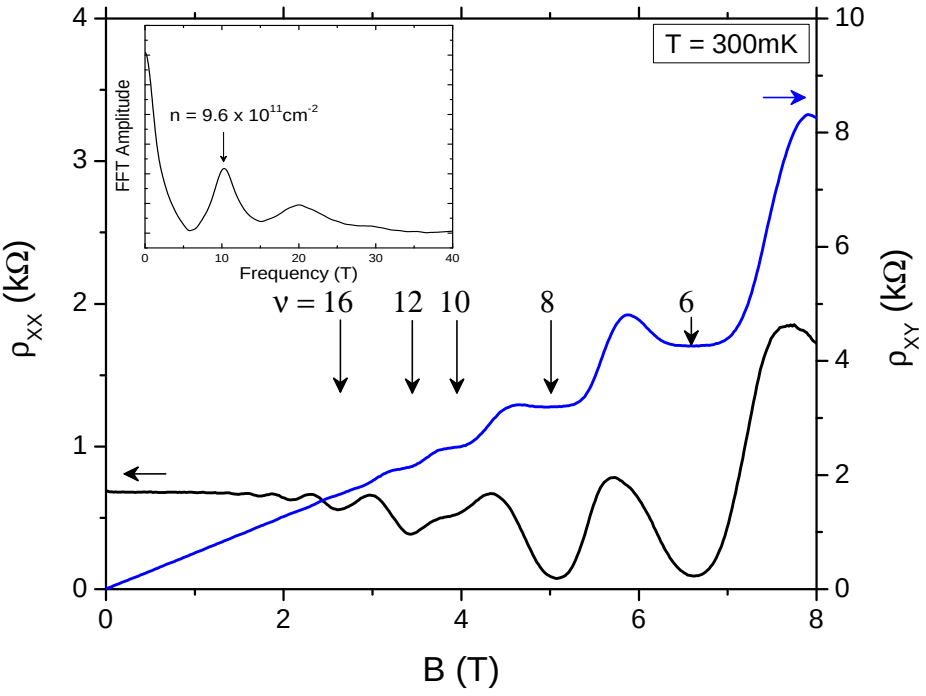


Figure 3.6: Longitudinal resistance ρ_{XX} and Hall resistance ρ_{XY} vs. magnetic field B at T = 300 mK, showing the integer quantum Hall effect. The filling factor ν is marked by arrows. The inset shows the Fourier spectrum of the longitudinal resistance.

In high fields we also observed a clear integer quantum Hall effect. At integral filling factors of ν, where $\rho_{XY}^{-1} = (e^2/h)\nu$, there are broad plateaus at the corresponding values of the Hall resistance that coincide with the minima in the longitudinal resistance. This occurs whenever the Fermi level lies between Landau levels. The change of these filling factors also indicates the system's degeneracy. At low fields, the total degeneracy is 4, resulting from the two-fold valley degeneracy (Δ_2) and the spin degeneracy,

$$g_{low} = g_v g_s = 4. \tag{3.1}$$

Beyond 4 T the Zeeman splitting lifted the spin degeneracy, as a result we observed filling factor of 10 and its changing by consecutive even integers, which indicates that

$$g_{high} = g_v = 2. \tag{3.2}$$

From the onset field of Zeeman spin splitting, we can also estimate the Landau level broadening in the 2DEG. Since the bulk Si has an effective g-factor of 1.99, if we ignore the small enhancement of g-factor in strained Si due to exchange interactions, the Landau level broadening

$$\Gamma = g\mu_B B \approx 0.46\ meV, \tag{3.3}$$

where $\mu_B = e\hbar/2m_0$ is the Bohr magneton.

Magneto-transport measurements on the 20% Ge 2DEG sample #3996 yielded very similar results. The filling factors are larger due to the higher electron density, and also the Landau level broadening is smaller.

3.2.4 Electron Spin Resonance of Electrons in Si and SiGe

In recent years, considerable efforts have been put towards spin manipulation in modulation-doped Si/SiGe heterostructures, due to its important role in quantum computing. To study the possibility of g-factor tuning in our 2DEG, we investigated the anisotropy of g-factor and electron spin resonance (ESR) linewidth in these quantum wells.

There are two popular mechanisms to manipulate spins in semiconductors. One is the Bychkov-Rashba (BR) effect [50]. The BR effect is caused by broken mirror symmetry induced by structure and/or an applied electric field (structure induced asymmetry, SIA). As an alternative mechanism, g-factor tuning has been proposed. In such a scheme spins are assumed to be selectively manipulated by a resonant radio

frequency field.

We first grew a 30% Ge 2DEG sample #4754, which is very similar to #4736. The ESR spectrum is measured through collaboration with Professor Stephen A. Lyon's group at Princeton University. Fig. 3.7 shows the spectrum, a shift in g-factor ($\Delta g = 0.0002$) with respect to the direction of external field is clearly present. To interpret this anisotropy, we need to consider the BR effect, which accounts for the lowest order of spin-orbit interaction (SOI). The SOI causes zero-field spin splitting and can be decribed by an effective magnatic field term:

$$\mathbf{H}_{BR} = \frac{2\alpha_{BR}k_F}{g_0\mu_B}\mathbf{e}_k \times \mathbf{e}_z, \tag{3.4}$$

where k_F is the electron momentum at Fermi level, $\mathbf{e}_k$ is the direction of electron velocity, and $\mathbf{e}_z$ refers to the growth direction (001). The BR parameter α_{BR} is a material constant dependent on structure that reflects the strength of SOI.

In thermal equilibrium, all velocity directions are isotropically distributed. Averaging the resulting total field yields an anisotropy of the resonance field. In an external magnetic field H_0, the effective BR field will cause the measured g-factor depending on the field orientation (angle θ, with respect to $\mathbf{e}_z$):

$$g = g_0\left[1 + \frac{H_{BR}^2}{4H_0^2}(1 + cos^2\theta)\right]. \tag{3.5}$$

This leads the g-anisotropy shown in the spectrum:

$$\Delta g = |g(0^\circ) - g(90^\circ)| = g_0\frac{H_{BR}^2}{4H_0^2}. \tag{3.6}$$

Our measured g-anisotropy is similar to numbers reported by other groups [51], which is about 0.0002 for a pure silicon channel. We also evaluated that $\alpha_{BR} = 1.58 \times 10^{-13}$ eV·cm from g-factor using equation (3.6), while the the reported value is about $5{\times}10^{-13}$ eV·cm in silicon.

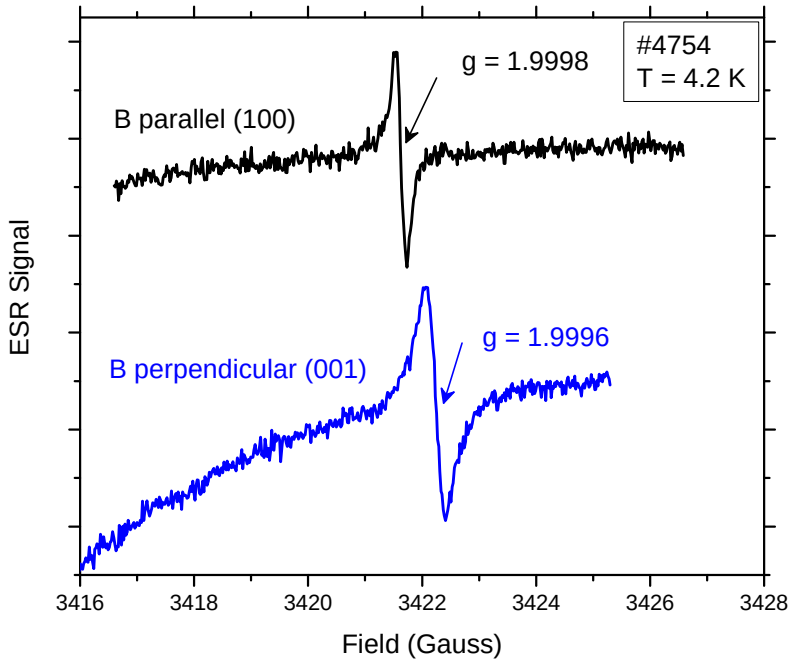


Figure 3.7: The electron spin resonance spectrum of the Si 2DEG sample #4754 in both perpendicular and in-plane magnetic field.

In their work an increase in both g-anisotropy and ESR linewidth with Ge content in the channel was observed. It can be explained by an increase of SOI with more Ge. To confirm this, we also grew a 2DEG sample #4759 with a $Si_{0.95}Ge_{0.05}$ channel and otherwise identical structure to #4754 and measured the ESR spectrum. Fig. 3.8 is the result. With the added 5% Ge in channel, one clearly finds that the g-factor shift becomes larger ($\Delta g = 0.0022$), and the ESR peak lindwidth is broadened.

The ESR signal linewidth ΔH for in-plane field is influenced by the BR effect and can be calculated as below:

$$\Delta H \approx \frac{\pi \alpha_{BR}^2 n_s}{\hbar g_v g \mu_B} \tau_k, \tag{3.7}$$

where τ_k is the momentum relaxation time. To the first order, both Δ_g and Δ_H

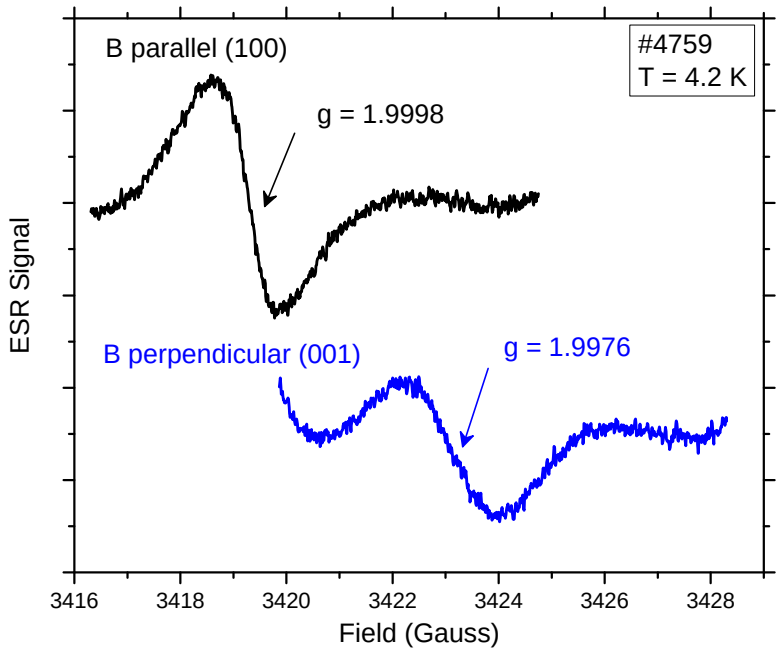


Figure 3.8: The electron spin resonance spectrum of the $Si_{0.95}Ge_{0.05}$ 2DEG sample #4759 in both perpendicular and in-plane magnetic field.

increase proportionally to the square of the BR parameter α_{BR}. Therefore, adding a small amount of Ge to the Si quantum well channel shifts the ESR to smaller g-factors because of the stronger SOI in SiGe. This effort can provide an alternative way for g-factor tuning in quantum dots. The growth of such germanium added channel is also compatible with our existing RTCVD growth routines.

3.3 Theoretical Mobility Models

3.3.1 Calculation of Mobility Limited by Coulomb Scattering Mechanisms

To better understand and estimate the ultimate mobilities that can be achieved in n-type modulation-doped Si/SiGe heterostructures, many theoretical studies were

performed to account for the influence of various scattering mechanisms. Typical mechanisms include Coulombic interactions with remote impurities or background dopants, scattering due to threading dislocations in the buffers, and alloy scattering. With our current growth conditions, high background doping level is likely to limit the achievable mobility. Hence we will focus on Coulomb scattering mechanisms and present a quantitative study within a simple framework published by AT&T Bell Laboratories [53].

Consider free 2-D electrons in a normalization area A in the x-y plane. The electrons are scattered by potential energy V(z,$\mathbf{r}$) from the impurities, where $\mathbf{r}$ is the in-plane position. Define the initial and final states to be the plane waves $\phi_i(z)$ and $\phi_f(z)$. Using an isotropic effective mass m*, the elastic scattering rate is calculated by Fermi's golden rule within the Born approximation

$$\begin{aligned}
\tau^{-1} &= \frac{2\pi}{\hbar}\sum_f |\langle f|V|i\rangle|^2\,\delta(E_f - E_i) \\
&= \frac{2\pi m^*}{\hbar^3|k_f|}\int \frac{A}{(2\pi)^2} d^2\mathbf{k}\delta(k_f - k_i) \\
&\quad \times \left|\int dz\phi_f^*(z)\phi_i(z)\int \frac{d^2\mathbf{r}}{A}\times exp[i(\mathbf{k}_f - \mathbf{k}_i)\cdot\mathbf{r}]V(z,\mathbf{r})\right|^2 . \qquad (3.8)
\end{aligned}$$

Next we represent the scattering potential V(z,$\mathbf{r}$) by its 2-D Fourier transform, or the power spectrum

$$V_{eff}(\mathbf{q}) = \int exp(i\mathbf{q}\cdot\mathbf{r})\frac{d^2(\mathbf{r})}{A}\int_{-\infty}^{\infty} V(\mathbf{r},z)\phi_f^*(z)\phi_i(z)dz, \qquad (3.9)$$

$$S(\mathbf{q}) \equiv \lim_{A\to\infty} AV_{eff}^2(\mathbf{q}). \qquad (3.10)$$

Thus the elastic scattering rate is

$$\begin{aligned} \tau^{-1} &= \frac{2\pi m^*}{\hbar^3 |k_f|} \int \frac{A}{(2\pi)^2} d^2\mathbf{k} \delta(k_f - k_i) V_{eff}^2 \\ &= \frac{m^*}{2\pi\hbar^3} \int S(\mathbf{q}) d^2\mathbf{k} \delta(k_f - k_i). \end{aligned} \tag{3.11}$$

The elastic scattering geometry is show in Fig. 3.9. In the degenerate, low-temperature limit, both $\mathbf{k}_i$ and $\mathbf{k}_f$ are the Fermi wave vector. The above integral over the final wave vector is confined by the delta function to the Fermi surface.

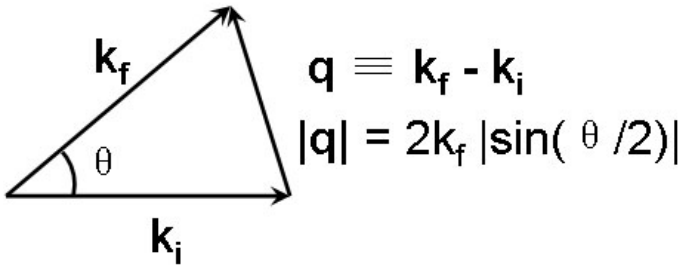


Figure 3.9: Elastic scattering geometry in two-dimensions. $\mathbf{k}_i$ and $\mathbf{k}_f$ refer to the initial and final wave vectors. θ is the scattering angle.

If the potential fluctuations are isotropic, so that S($\mathbf{q}$) is independent of the angle θ, and $dq/d\theta = k\sqrt{1 - q^2/4k^2}$. The total relaxation time is

$$\begin{aligned} \tau^{-1} &= \frac{m^*}{2\pi\hbar^3} \int_0^{2\pi} d\theta S(\mathbf{q}) \\ &= \frac{m^*}{\pi\hbar^3} \frac{1}{k} \int_0^{2k} \frac{S(q)dq}{\sqrt{1 - q^2/4k^2}}. \end{aligned} \tag{3.12}$$

The mobility reflects the momentum relaxation rate, which includes the fraction of the momentum lost, $1 - \cos(\theta) = q^2/2k^2$,

$$\tau_m^{-1} = \frac{m^*}{\pi\hbar^3} \frac{1}{k} \int_0^{2k} \frac{S(q)dq}{\sqrt{1 - q^2/4k^2}} \frac{q^2}{2k^2}. \tag{3.13}$$

Next we need to include the screening for each wave vector $\mathbf{q}$, which causes the

screened potential reduced by

$$\epsilon(\mathbf{q}) \equiv \frac{V(\mathbf{q})}{V_0(\mathbf{q})} = \frac{q}{q+q_s}, \tag{3.14}$$

where

$$q_s \equiv \frac{e^2 DOS}{2\epsilon\epsilon_0} = \left[g_v g_s (m^*/m_0)/\epsilon\right](e^2 m_0/4\pi\epsilon_0\hbar^2). \tag{3.15}$$

For electrons in tensile-strained silicon, the valley and spin degeneracy g_v and g_s are both 2. With the above equations (3.13) and (3.15), various scattering mechanisms can be evaluated with the integral in the form of the screened power spectrum of potential. In particular, we are interested in the Coulombic scattering, which is caused by ionized impurities either from the remote dopant layer or from uniform background charges.

First we will consider the remote impurity scattering. The supply layer is modeled as a thin layer at $z = -h$ from the top of the strained Si channel ($z = 0$). Monroe et al. showed that the power spectrum of the charge sheet at its own plane is

$$S(\mathbf{q}, z = -h) = N_D d_{doping} \left(\frac{e^2}{2\epsilon\epsilon_0 q}\right)^2. \tag{3.16}$$

And away from the sheet the potential decays exponentially

$$S(\mathbf{q}) = \frac{e^4 N_D d_{doping}}{\left[2\epsilon\epsilon_0(q+q_s)\right]^2} e^{-2qh} f(q), \tag{3.17}$$

where

$$f(q) = \left(\int_{-\inf}^{\inf} dz |\phi(z)|^2 e^{-qz}\right)^2. \tag{3.18}$$

And $f(q)$ is close to unity if h is replaced with the effective setback h_{eff} between the centroid of the dopant sheet and the centroid of the wave function. The density of states gives an approximate 2DEG wavefunction width of about a quarter of the

Bohr radius $a_B = (4\pi\epsilon\hbar^2)/(me^2)$.

$$h_{eff} = d_{doping}/2 + d_{spacer} + \frac{\pi\epsilon\hbar^2}{m^* e^2}. \tag{3.19}$$

The final result is that, for remote impurity scattering

$$\mu_{remote} = \frac{e\tau_m}{m^*} \approx \frac{16\sqrt{\pi g_v g_s e^2 n_{2D}^3}\, h_{eff}^3}{\hbar N_D d_{doping}}. \tag{3.20}$$

The predicted mobility increases rapidly with both effctive setback thickness and increasing density. The Fermi wave vector increases with the increasing density, so the integral in (3.13) will be more dominated by smaller scattering angles. As an estimate, for an effective setback $h_{eff} = 15\ nm, n_{2D} = 6\times10^{11}\ cm^{-2}, N_D d_{doping} = 1\times10^{12}\ cm^{-2}$, the calculated μ=190,000 cm^2/Vs, more than one order of magnitude higher than the highest mobility that we have seen. Therefore the Coulombic scattering from remote impurity is not our mobility limitation mechanism.

Next we will consider the effect of background impurities. The background impurities are treated as a uniform charge distribution extended to the whole space.

$$\begin{aligned} S(\mathbf{q}) &= \frac{N_{background} e^4}{\left[2\epsilon\epsilon_0(q+q_s)\right]^2} \int_{-\infty}^{\infty} dz e^{-2q|z-\overline{z}|} \\ &= \frac{N_{background} e^4}{\left[2\epsilon\epsilon_0(q+q_s)\right]^2} \frac{1}{q}. \end{aligned} \tag{3.21}$$

For the well-screened case, $q \ll q_s$, the integral (3.13) yields

$$\mu_{background} = \frac{e\tau_m}{m^*} = \frac{\sqrt{g_v^3 g_s^3 e^2 n_{2D}}}{4\sqrt{\pi}\hbar N_{background}}. \tag{3.22}$$

The above equation shows that the mobility will decrease inversely proportional to the background impurity level.

The total electron mobility is related to the two Coulomb scattering mechanisms

by the following equation:

$$\frac{1}{\mu_{total}} = \frac{1}{\mu_{remote}} + \frac{1}{\mu_{background}}. \tag{3.23}$$

3.3.2 The Effect of Background Impurities

One of the major drawbacks of our RTCVD system is the high background phosphorus doping level due to the reactor history. By using dichlorosilane as the silicon precursor at growth temperature between 600 - 750 °C, SIMS analysis shows a typical background P level of around 3×10^{17} cm^{-3} in Si. The P level usually increases even more with the addition of germanium, perhaps due to the lower heat of formation of GeP than that of SiP phosphide (6 vs. 15 kcal/mole) [52]. Levels of around 1×10^{18} cm^{-3} are typical in SiGe. Fig. 3.10 shows a SIMS analysis of multiple Si/SiGe quantum wells. Note the background phosporous level rises in SiGe layers. As a result, our achieved electron mobility is limited by scattering from such ionized impurities.

In order to evaluate the detrimental effect of background doping level on the 2DEG mobility, we calculated electron mobilities in #4736 for different background doping levels using equations (3.20), (3.22) and (3.23). To provide a more accurate calculation, we also take into account that 2-D electron density will depend on the background impurities in any given structure. This can be modeled by solving one-dimensional Poisson's equation self-consistently. We used a free program written by G. L. Snider at Notre Dame [54] and modified its material parameters to include strained Si and SiGe bulks. For simplicity, a constant background doping of phosphorous is assumed in all grown layers. For a fixed background level, we first extracted the 2-D electron density in strained silicon using Snider's program, then mobility components were calculated accordingly.

Fig. 3.11 shows the results for the mobility calculation. From the lower figure we can see that the two scattering mobility curves cross at a background doping level

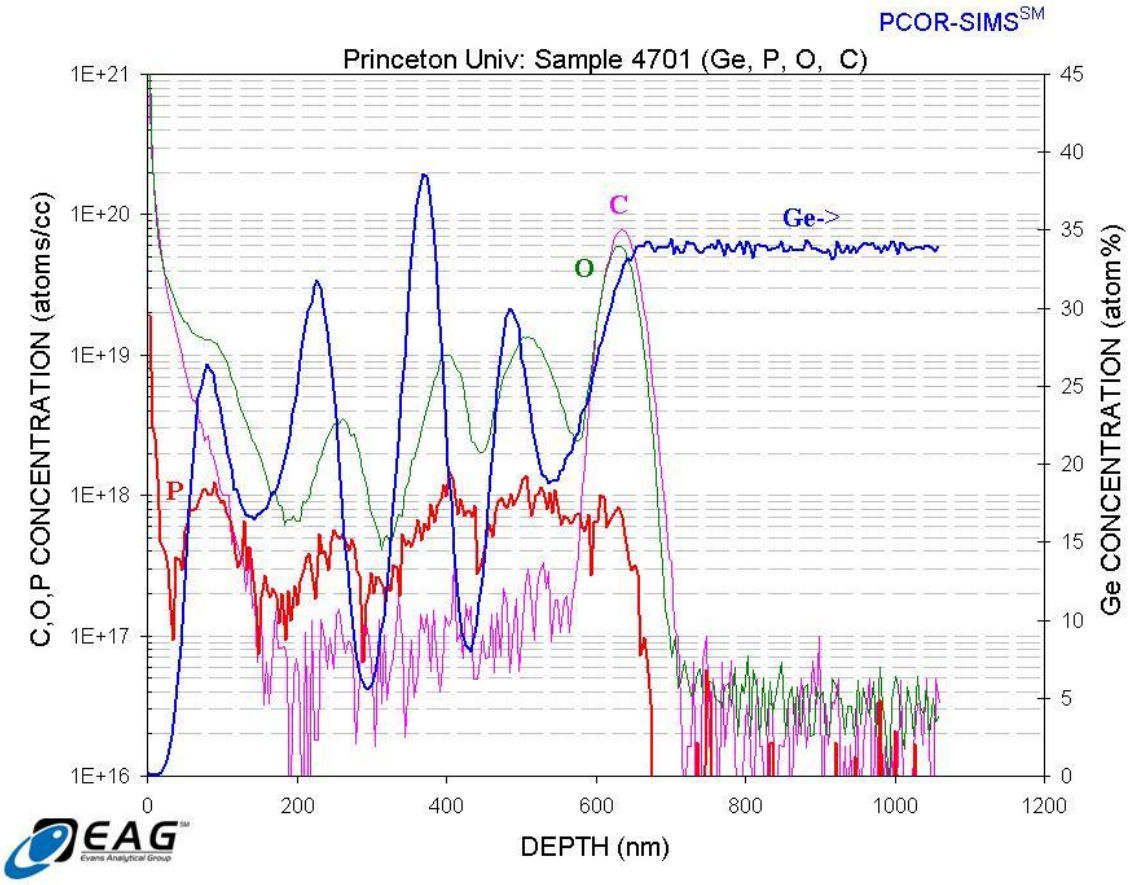


Figure 3.10: SIMS analysis of sample #4701 showing multiple Si quantum wells on $Si_{0.7}Ge_{0.3}$ relaxed buffers. The growth temperatures for Si and SiGe are 625 °C and 700 °C, respectively. The sample was not doped intentionally with phosphorus.

of 8×10^{15} cm^{-3}. In our case, the high background impurity level is clearly the limiting scattering mechanism. The measured density and mobility of sample #4736 both indicate a background doping level around 1 - 3×10^{17} cm^{-3}. Therefore, our experimental mobility of 10,000 cm^2/Vs is consistent with the theoretical calculation.

Another important quantity when evaluating 2DEG mobility is the Dingle ratio, τ_t/τ_s, defined as the ratio of the transport scattering time τ_t to the single-particle elastic relaxation time τ_s. A large ratio indicates that long-range Coulombic scattering at remote impurities is dominant, preferentially small-angle scattering occurs. On the other hand, if isotropic, short-range scattering events such as background impurities or interface charge scattering occurs, the momentum loss factor (1-cos θ) in the τ_t will

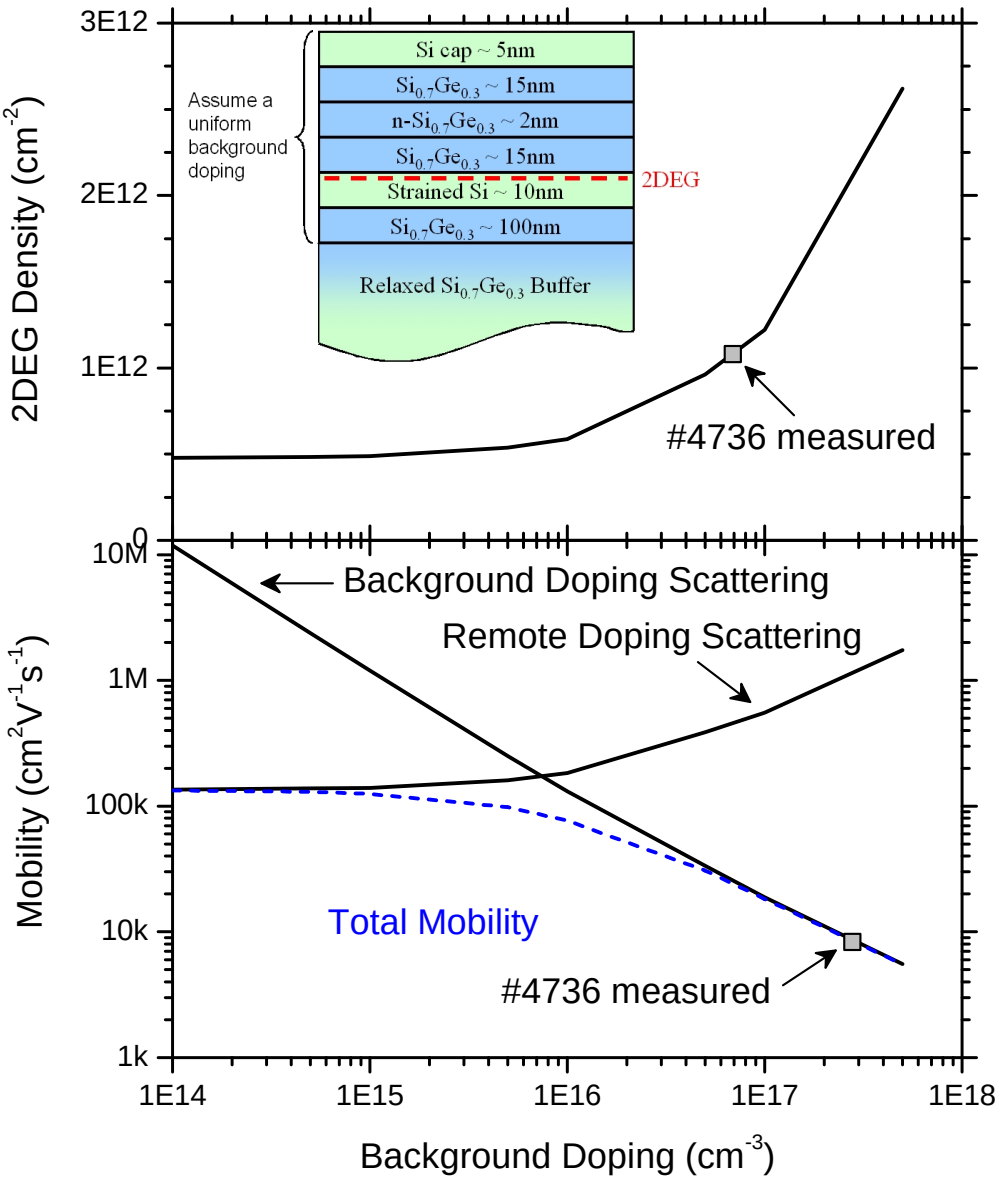


Figure 3.11: Calculated 2DEG densities and electron mobilities limited by Coulombic scattering mechanisms vs. uniform background impurity levels. The inset shows the layer structures assumed in the calculation.

drop. Hence the ratio will be close to unity. For our sample #4736, the τ_t can be estimated from the mobility and is around 1.08 ps. The τ_s is deduced from the onset magnetic field of SdH oscillations B_{on}

$$\tau_s \omega_B = 1 \;\Rightarrow\; \tau_s = \frac{1}{\omega_B} = \frac{m^*}{eB_{on}}. \tag{3.24}$$

For a B_{on} = 1.22 T as shown in Fig. 3.6 and $m^* = 0.19\ m_0$, τ_s = 0.89 ps. So the Dingle ratio is about 1.22, indeed very close to unity.

3.3.3 Efforts Towards Lower Background Impurities

The high background impurity level is an important problem to solve in our RTCVD. There are several ongoing concerns of possible means to reduce the background doping level.

Normally the background level decreases with higher growth temperature if all other conditions remain the same. In our experiments we found that 2DEG mobility increased from 5,000 to 10,000 cm^2/Vs after raising the silicon growth temperature from 700 to 750 °C. Fig. 3.12 shows the effect of growth temperature on background level [55]. However raising the growth temperature has its limitations. First, higher temperature will increase the growth rate, making it more difficult to control the thickness and abruptness of thin Si/SiGe layers. Second, the increased dopant diffusion will make it difficult to control the sharpness of doping profiles.

Using alternative silicon and/or germanium precursors is a promising solution. Several silicon precursors including disilane (Si_2H_6) and neopentasilane (Si_5H_{12}) are being installed and calibrated in our RTCVD system [56]. Fig. 3.13 shows a SIMS analysis of multiple Si/SiGe quantum wells using Si_2H_6 and GeH_4. We find a reduction of phosphorous background level in SiGe layers compared to the previous results in Fig. 3.10, especially at the Si/SiGe interface and in the SiGe spacer layer, which is

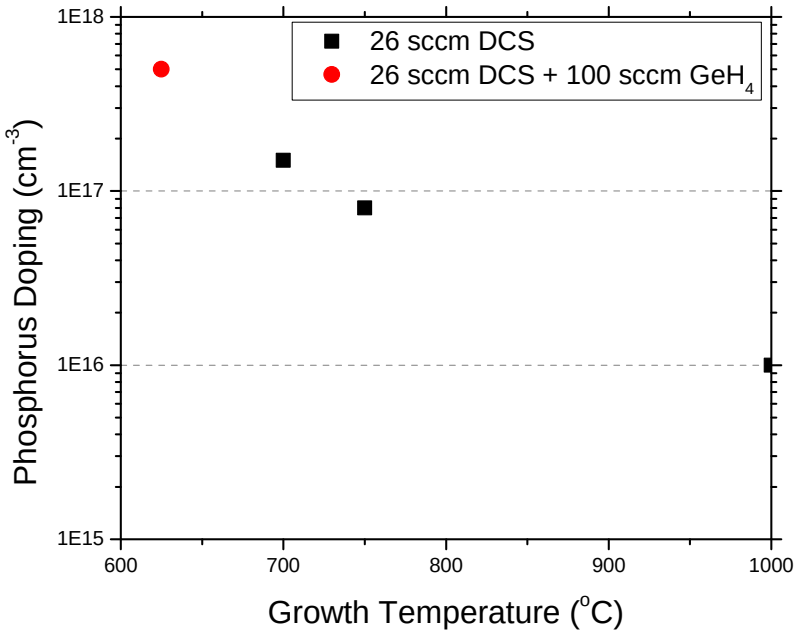


Figure 3.12: Background phosphorus doping levels in Si and SiGe layers vs. growth temperature.

just above the centroid of 2DEG. The background doping in Si seems to be about the same or slightly lower compared to the growth with dichlorosilane.

3.4 Modulation of Si/SiGe 2DEG Electron Density with Top Gating

3.4.1 Al_2O_3 High-k Gate Dielectric by Atomic Layer Deposition

The modulation of electron density in the Si/SiGe 2DEG is of great interest because the fabrication of complex low dimensional devices relies upon such control. Unlike in III-V semiconductor systems, Schottky gates on Si/SiGe heterostructures have proven

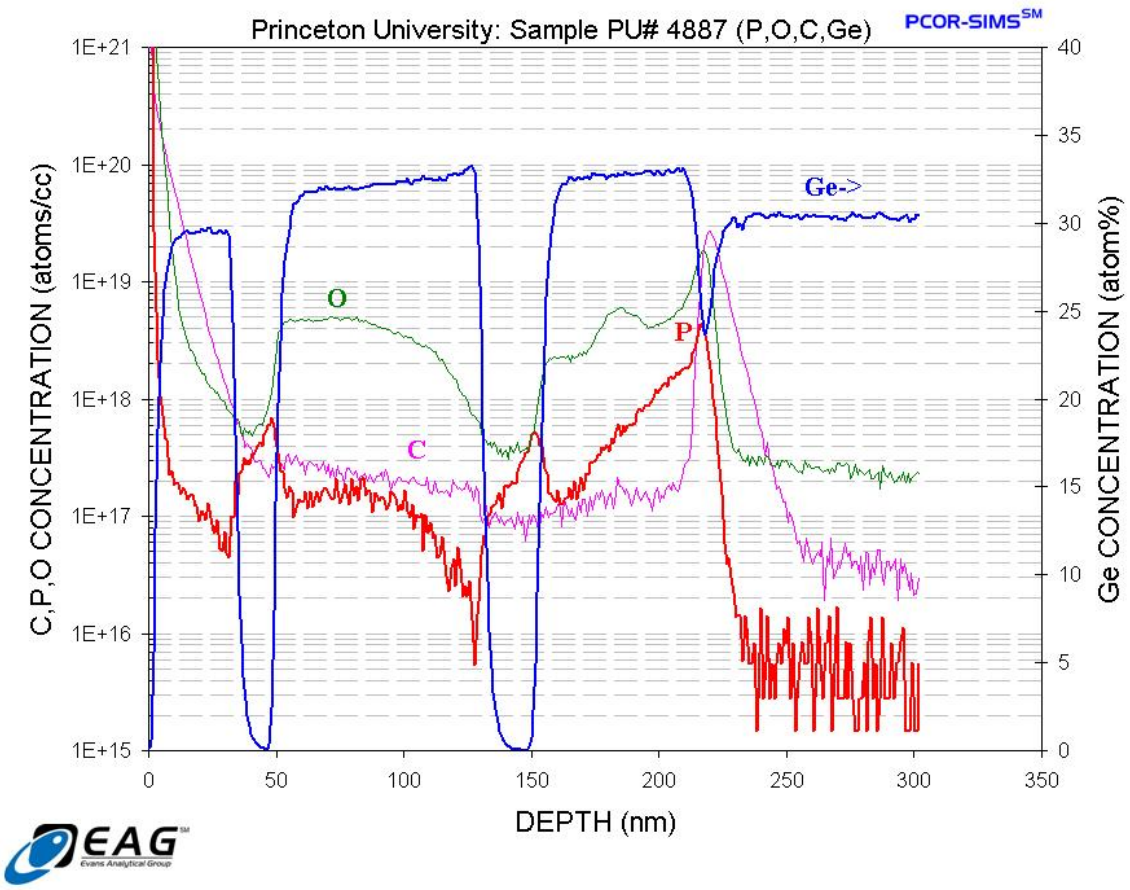


Figure 3.13: SIMS analysis of sample #4887 showing multiple Si quantum wells on $Si_{0.7}Ge_{0.3}$ relaxed buffers. The Si and SiGe layers are both grown at 575 °C using Si_2H_6 as the silicon precursor.

leaky [57]. It was argued that the large leakage was caused by dopant segregation at the surface or threading dislocations. Recently palladium metal top gates have been successfully applied to quantum dot gate geometries [58, 59]. However, the leakage problem over large area remains poorly understood.

As an alternative, effective gating by gate dielectrics have been proposed [60]. Atomic-layer-deposited (ALD) gate dielectric shows superior characteristic compared to plasma-enhanced CVD (PECVD) silicon dioxide and metal Schottky gate in terms of leakage and interface trap density. A variety of ALD thin films have been intensively studied as high-k gate dielectrics for CMOS applications. At Princeton, we can deposit ALD Al_2O_3 with a state-of-the-art Cambridge NanoTech model Savannah

100 reactor. This process for Si/SiGe 2DEG gating was first developed by Lai et al. in Professor Daniel C. Tsui's group at Princeton [61].

The principle of ALD is based on sequential pulsing of chemical precursor vapors, a cycle of which forms one atomic layer. This generates pinhole free coatings that are extremely uniform in thickness, even deep inside pores, trenches and cavities. As an example of Al_2O_3 deposition, one trimethyl aluminum (TMA) and one H_2O vapor pulse form each cycle, and deposit a monolayer of Al_2O_3 of approximately 0.9Å in thickness. The two reaction steps in each cycle are:

$$Al(CH_3)_{3(gas)} + : Al{-}O{-}H_{(solid)} \rightarrow : Al{-}O{-}Al(CH_3)_{2(solid)} + CH_4, \quad (3.25)$$

$$2H_2O_{(gas)} + : O{-}Al(CH_3)_{2(solid)} \rightarrow : Al{-}O{-}Al(OH)_{2(solid)} + 2CH_4. \quad (3.26)$$

Fig. 3.14 shows such ALD cycles. In addition to its gating capacity, ALD Al_2O_3 can also be used as surface passivation for future device packing, as it is robust against surface reactions and moisture absorption over large area.

3.4.2 Modulation of Si/SiGe 2DEG Electron Density

For device fabrication, a 2DEG sample #4748 was first grown and subsequently etched into a Hall bar. After a wet chemical clean of sample surface, a 90-nm thick Al_2O_3 layer was deposited at a substrate temperature of 300 °C with 1000 cycles. Prior to the ALD process, the native oxide of the Si substrate was removed by diluted HF dip. The first 20 cycles are pulsed with only H_2O vapor to form well-defined chemical native silicon oxide. Then each complete cycle has two precursor exposure pulses and a 5-sec N_2 purge in between. The growth rate at 300 °C is confirmed to be 0.9 Å per cycle using ellipsometry (model Gaertner Scientific L3W16). Contact holes are wet etched through the oxide. The etch rate of Al_2O_3 in 1:10 buffered oxide etch (BOE) solution is 4 Å/s. Ohmic contacts are then made by thermal evaporation,

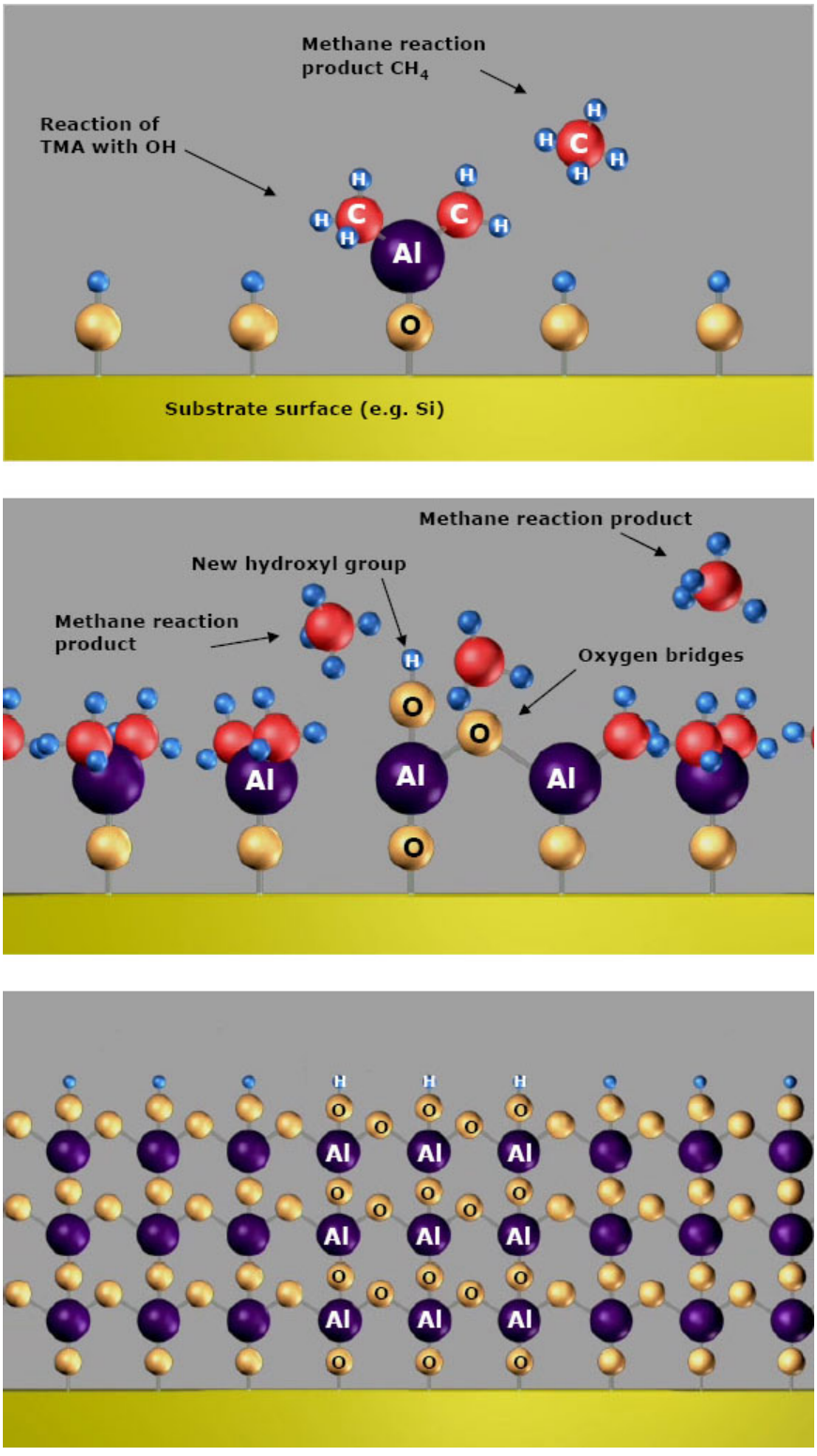


Figure 3.14: ALD cycle for Al_2O_3. Two reaction steps in each cycle as well as completed monolayers are shown. (Image courtesy of Cambridge NanoTech, www.cambridgenanotech.com)

lift-off and annealing using the same process described before. Finally, an aluminum gate is defined by lift-off on the gate dielectric. Fig. 3.15 shows the schematic view of the finished device structure.

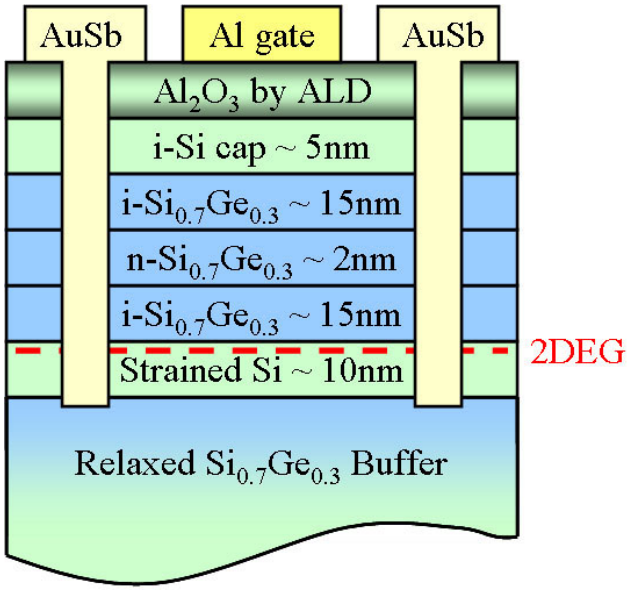


Figure 3.15: Schematic view of an n-channel Si/SiGe MOSFET with ALD Al_2O_3 as gate dielectric.

After the gate fabrication, the sample was mounted in a He-3 refrigerator. The gate leakage is negligible (< 20 pA) within the entire gate voltage scan range, as shown in Fig. 3.16. Low-temperature magneto transport traces were taken with applied gate voltages. Fig. 3.17 shows the traces at different gate voltages. The electrons are depleted with more negative gate voltage applied as in a depletion-mode n-channel MOSFET. The reduction in electron density is reflected in both the change of Hall slope and the SdH oscillation period. To take a closer look at the SdH and quantum Hall effect, Fig. 3.18 shows the SdH oscillations and the corresponding filling factors at the resistance minina at two difference gate voltages. At both gate voltages the behavior is very similar to the 2DEG #4736 without top gating. At low field, the filling factor ν changes by 4 indicating a total four-fold degeneracy, at field beyond 4 T only the two-fold valley degeneracy remains.

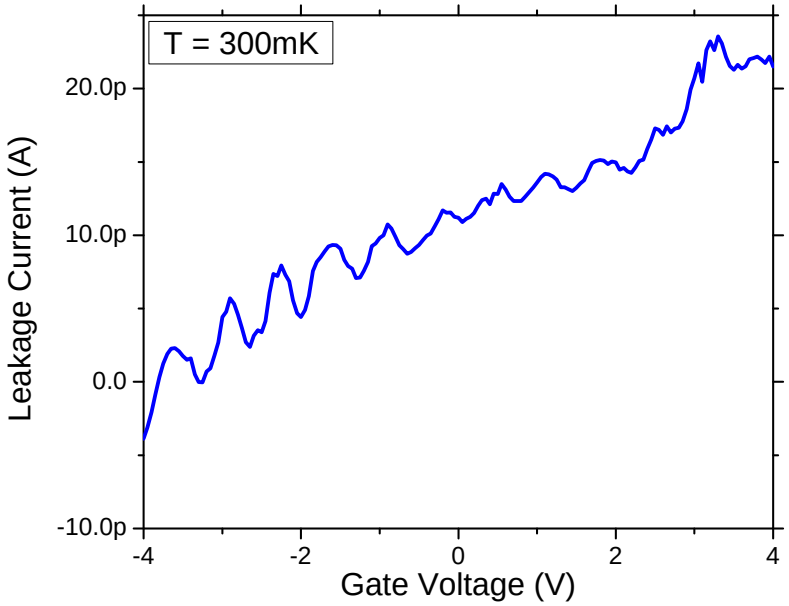


Figure 3.16: Gate leakage current through the ALD Al_2O_3 vs. applied voltage.

Also as a result of the depleted 2DEG with negative gate voltage, the electron mobility is reduced. Based on our previous equations (3.20) and (3.22), the elastic Coulombic scattering rates strongly depend on Fermi wave vector. If we assume that the background impurity scattering is the limitation mechanism, then the $\mu_{background}$ should be proportional to the Fermi wave vector, or the square root of electron density. Another observation from the magneto transport measurement is that the onset magnetic field of SdH oscillations also shifts to the right as more negative gate voltage is applied, indicating a shorter single-particle elastic relaxation time τ_s. As a result of the reduction in both relaxation times τ_t and τ_s, the Dingle ratio does not change much with the gate voltage. Fig. 3.19 shows the measured results of both electron density and mobility as a function of the applied gate voltage.

To quantitatively study the gating characteristics, we calculated the electron density by incorporating a simple MOS structure into the 1-D Poisson's equation solving scheme we mentioned in the previous section. The ALD Al_2O_3 is modeled with rela-

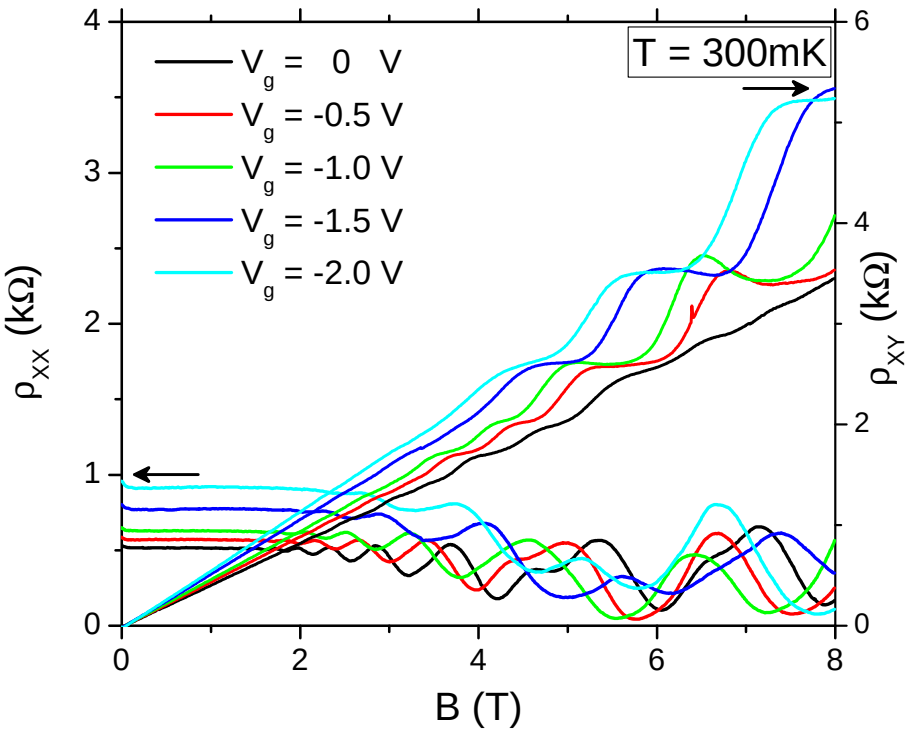


Figure 3.17: Longitudinal resistance ρ_{XX} and Hall resistance ρ_{XY} vs. magnetic field at T = 300 mK, with an applied top-gate voltage V_g from 0 to –2 V.

tive dielectric constant of 9.0, band gap of 7.0 eV, and conduction band offset to Si of 2.5 eV [62]. No oxide-silicon interface charge states are assumed. The calculated result is also included in Fig. 3.19. An inflection point at a gate voltage of about –1 V is predicted by the modeling. However, it is only weakly observed in the experimental data. For gate voltages larger than the threshold, our model shows that the accumulated electrons will tend to stay in the undoped Si and SiGe cap layers, so the 2DEG density in Si channel remains almost constant. Below the threshold, electrons are depleted by the gate voltage as a linear function with a slope of 3.24×10^{11} cm^{-2}/V, or an equivalent capacitance of 0.052 μF/cm^2. The predicted capacitance between the gate and channel is estimated to be 0.068 μF/cm^2, by using ideal parallel plates

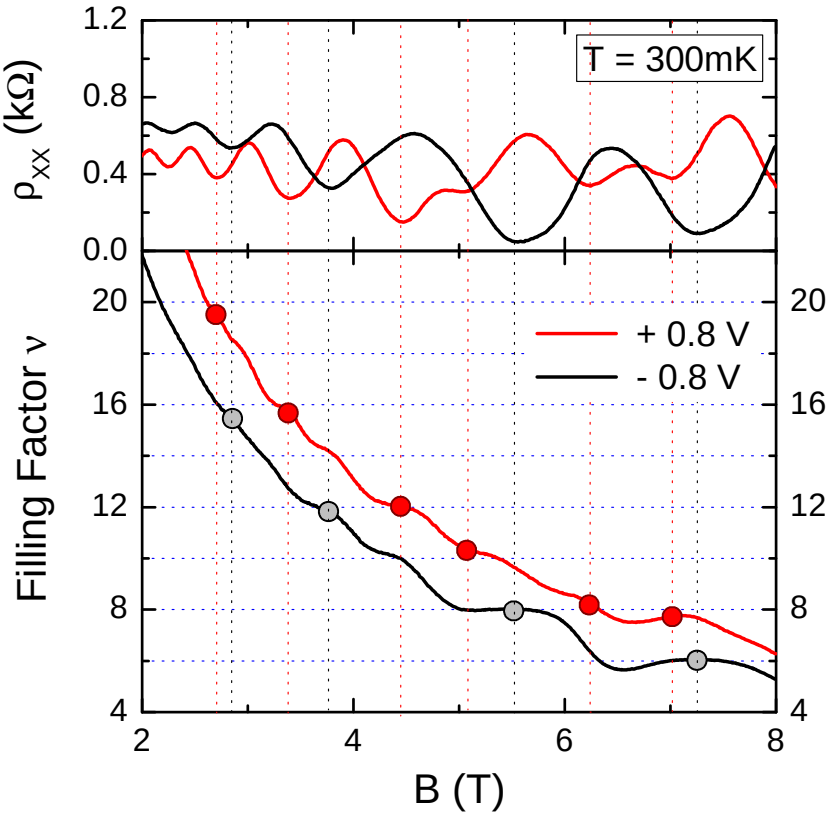


Figure 3.18: Shubnikov-de Hass oscillations and quantum Hall effect at integer filling factors at T = 300 mK, with applied top gate voltages of +0.8 and –0.8 V.

separated by 90 nm Al_2O_3 and 40 nm $Si_{0.7}Ge_{0.3}$.

Due to the high initial electron density, the 2DEG density remained in the low 10^{11} cm^{-2} range even with a high negative gate voltage ($\sim$ –4 V) applied. The channel resistivity was in the insulating regime ($\rho > h/e^2$) with such high voltages as the mobility decreased fast.

In conclusion, we have demonstrated Al_2O_3 as an excellent gate dielectric for modulation of electron density in a Si/SiGe 2DEG. The electron density can be depleted linearly by applying a negative gate voltage. Neither gate leakage nor hysteresis in either direction was observed.

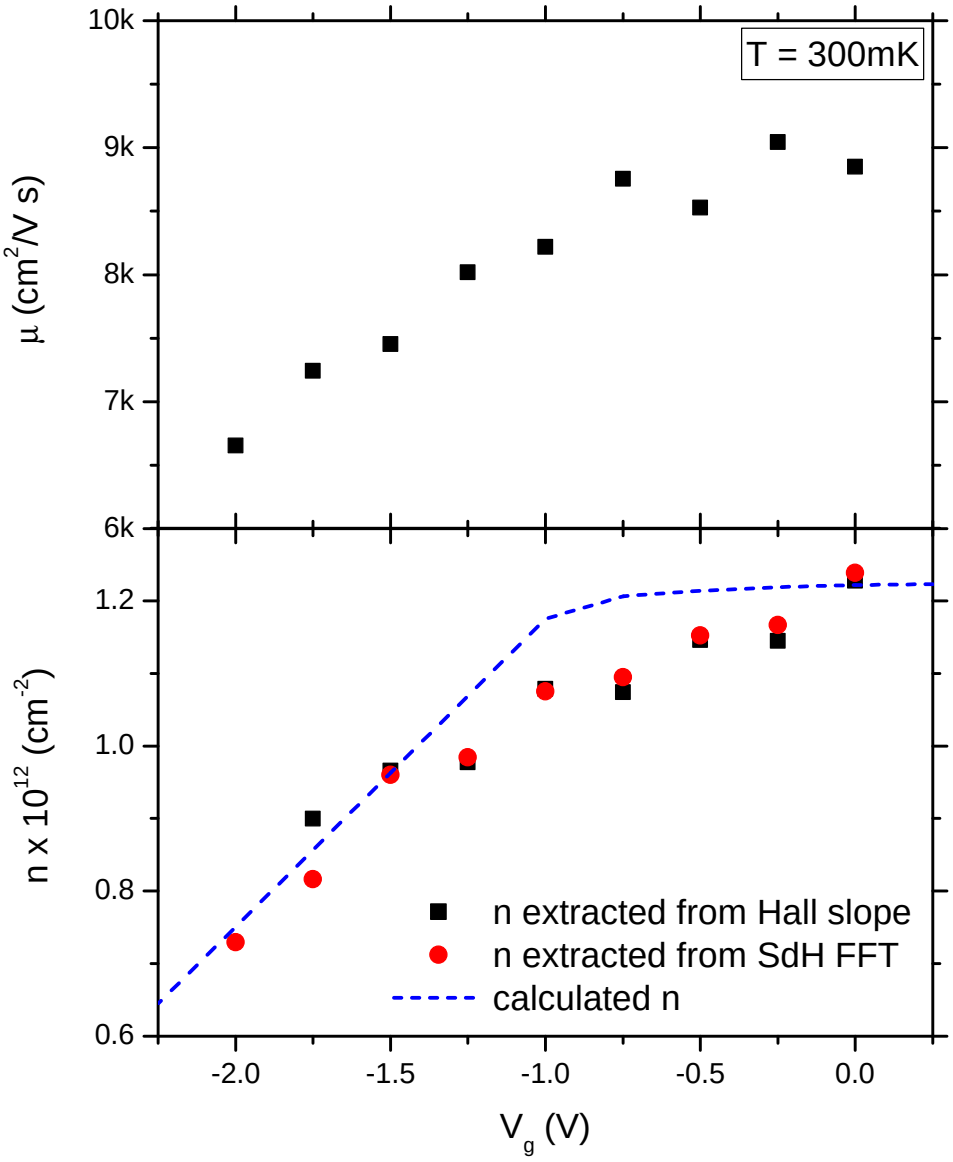


Figure 3.19: 2DEG density and mobility vs. applied gate voltage at T = 300 mK. Three methods for extracting the electron density are plotted: from Hall slope, from SdH oscillation period, and from solution of 1-D Poisson's equation.

3.5 Summary

Band engineering with modulation-doped Si/SiGe heterostructures enables the realization of two-dimensional electron gases, which has been the primary vehicle for research on semiconductor based quantum computing. In such systems 2-D electrons are trapped in a quantum well formed at the heterojunction, in which the low temperature mobilities are limited by Coulombic scatterings or interfaces rather than phonons.

High quality 2DEG's have been successfully employed on both 20% and 30% Ge commercially available relaxed buffers. Electron mobility around 10,000 cm^2/Vs is achieved. We first studied the electrostatics to get estimates of band diagrams as well as electron densities. Then we calculated the mobility of a 2DEG by considering the screening from Coulomb interactions. High background phosphorus impurity levels are our current mobility limiting mechanism. Several methods for improving the mobility have been proposed.

Electron spin resonance provides an accurate way to study the g-factor in the Si/SiGe 2DEG systems. The g-anisotropy in external magnetic fields can be tuned by adding a small amount of Ge into the Si channel, which can provide a mechanism to manipulate electron spins. Growth of such 2DEG with up to 5% Ge is demonstrated.

Finally we developed an experimental technique to implement ALD oxide as gate dielectrics on the modulation-doped heterostructures. As an example, ALD Al_2O_3 can be deposited at low temperature with digital thickness control. The 2DEG in the depletion regime with negative gate voltage applied showed well-behaved density tuning with negligible leakage current. The successful modulation of Si/SiGe 2DEG opens up new ways to explore low dimensional physics as well as device applications in previously inaccessible regimes.

Chapter 4

Parallel Two-Dimensional Electron Gases in Double Quantum Wells

4.1 Introduction

2DEG systems in semiconductors have been an incubator for both new physics and revolutionary technologies. Considering the many breakthroughs in single quantum well 2DEG's, it is quite natural to extend the study towards double or multiple quantum well systems. The additional degree of freedom results in rich physics not present in a single-layer system. One recent prominent example is the quantum cascade laser (QCL) [63]. One-dimensional multiple quantum well confinement in QCLs leads to the splitting of the band of allowed energies into a number of discrete subbands, which makes possible a population inversion between two subbands. Although QCLs were first fabricated in the III-V semiconductor systems, the intersubband transitions are independent of the relative electron and hole momenta across the bandgap, thus can be applied to indirect semiconductor systems such as silicon as well. Since the first Si/SiGe quantum cascade emitter was demonstrated in 2000 [64], Si/SiGe laser has been an active research topic.

A double quantum well (DQW) incorporating a 2DEG in each quantum well is among the simplest structures exhibiting significant interaction effects between two layers of 2DEG's. For example, one of the observed phenomena, which is directly related to the wave nature of electrons, is the wave-function coupling between two closely spaced parallel 2DEG's [65]. When the interlayer separation is comparable to the intralayer distance of the individual electrons in each 2DEG, interlayer Coulomb interactions are then just as important as intralayer ones and the system allows collective phases that do not exist in the individual layers. DQW systems in GaAs/AlGaAs heterostructures have been widely studied both theoretically and experimentally. Many fascinating new physical properties are reported, these include: quantized Hall effect (QHE) when electrical currents flow in parallel through the two wells [66], a giant enhancement of the zero bias interlayer tunneling conductance [67, 68], the vanishing of both the longitudinal and Hall resistances when equal but oppositely directed currents flow in the two layers [69, 70], and much more.

No experiments on DQW systems in Si/SiGe heterostructures exist to our best knowledge. It is the purpose of this chapter to study such DQW systems. Not only because they are as promising for the study of many-body physics as those already demonstrated in the III-V systems, but also because they are potential building blocks for quantum information processing, or in other words, the "Lego block" of our silicon-based quantum computer.

Let us consider the Loss-DiVicenzo proposal again. In their scheme the spin of single electron confined in quantum dot is used as qubits. Single-qubit operations can be achieved by local magnetic fields or g-factor engineering. Two-qubit operations are based on interactions resulting from exchange coupling (J) between adjacent qubits. Fig. 4.1 shows the interaction between two single-electron dots. When J is large, the two wavefuctions will overlap so operations such as swap can be performed. Such a quantum device involving two-qubit interactions is also referred as a quantum dimer.

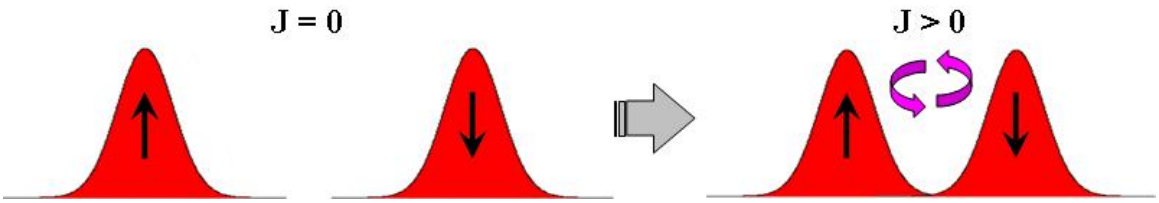


Figure 4.1: Exchange interactions between two spin qubits: when J = 0, the two qubits are uncoupled; when J > 0, the two qubits are coupled and two-qubit operations can be performed.

The conventional method to realize spin exchange coupling is to use double quantum dot. Fig. 4.2 (a) shows such a double quantum dot device. Two quantum dots are connected to each other, and their coupling is controlled by two lateral finger gates. The strength of the interaction depends on the gate voltages and is very sensitive to the gate voltage noise. Quantum-gate mechanisms in double quantum dot were studied extensively [71, 72]. However, experiments were not feasible until very recently [73]. Alternatively, if we could grow and fabricate quantum dots based on DQW systems, a quantum dot dimer would need only one dot that contains two parallel 2DEGs, as shown in Fig. 4.2 (b). It potentially offers a much neater design. In addition, our epitaxial regrowth technique (see Chapter 6 for a complete discussion) can provide 3-D confinement to both qubits and dimers and reduce the number of gates. The exchange interaction depends only on the growth structure hence is better controlled. Ideally each qubit/dimer requires no gate. An overall top gate is only needed to balance the two well densities in the dimer if they are different.

In collaboration with Professor Leonid Rokhinson's group, our group is among the first to propose the use of Si/SiGe DQW for silicon-based quantum computers [74]. It enables a scalable architecture for semiconductor-based quantum computing. We call it the "flying qubit" architecture, in a sense that electrons are physically shuffled in real 3-D space other than sitting in 2-D channels. Fig. 4.3 depicts this idea. Controllable parallel 2DEG's in Si/SiGe DQW systems are used as the interaction dimers and form the backbone in such an architecture.

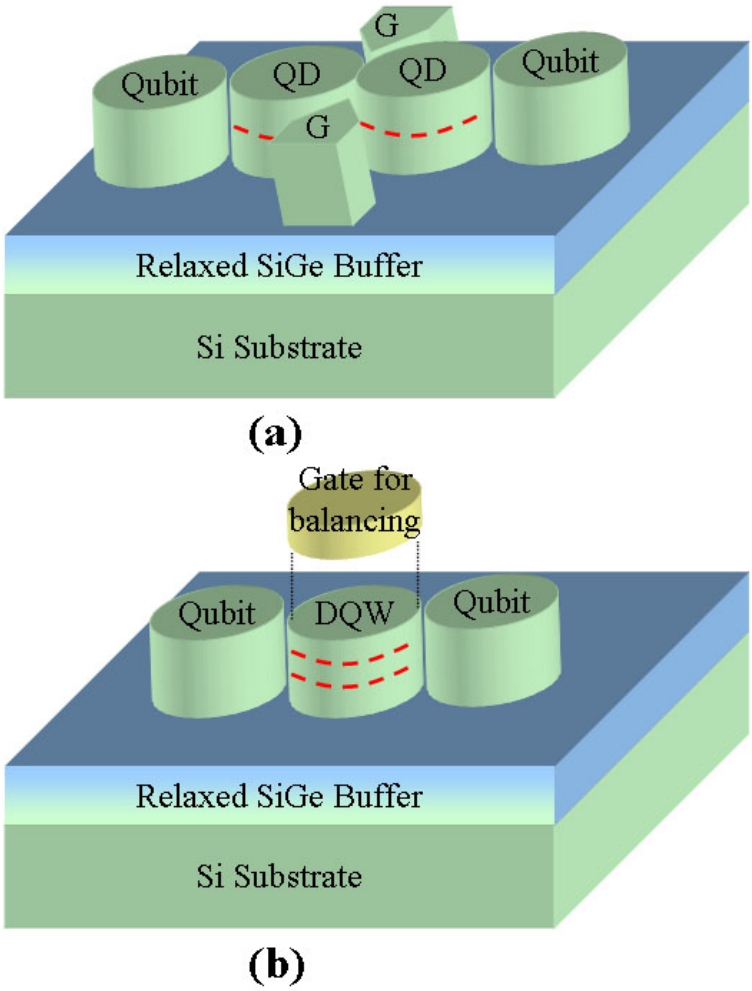


Figure 4.2: Schematic view of two quantum dimer fabrication schemes: (a) a coupled double quantum dot, the exchange coupling is controlled by two side finger gates; (b) a single quantum dot of double quantum wells, the exchange coupling is controlled by primarily the as-grown structure of the DQW and not by a gate voltage.

4.2 Modeling of the Double Quantum Well Systems

4.2.1 Design of Double Quantum Well Systems

In the following section we will study two types of DQW system design. The first design has two asymmetric quantum wells with doping supply placed only on one side. In the second design, the two quantum wells are symmetric with the same

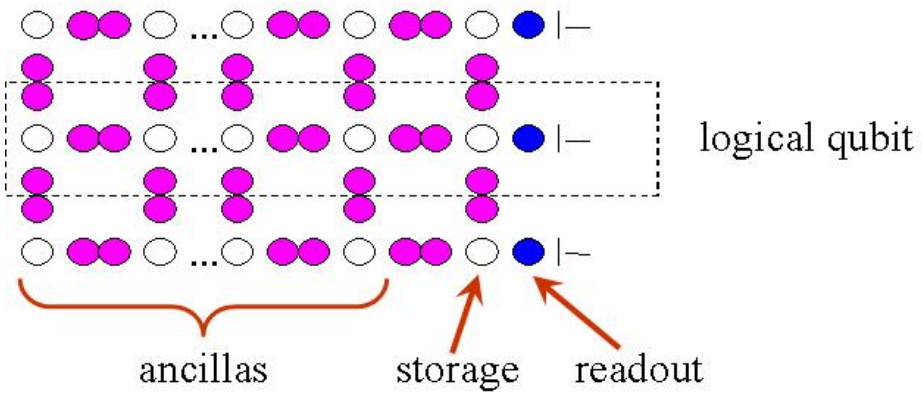


Figure 4.3: The "flying architecture" map for a quantum computer. Each qubit (white circle) is coupled to four identical qubits in the array via a quantum dimer (adjacent pink circles).

doping supply and spacer layers from both sides. Fig. 4.4 shows the detailed layer structures that will be studied. No silicon cap is added to avoid parasitic effects in band calculation due to the extra quantum well at the surface. In the asymmetric design, it is necessary to make the top quantum well thinner so that more carriers can be transferred to the bottom well for balancing the electron densities in the two 2DEG channels.

For the band diagram and electron density calculation, we used the same program and material parameters that were presented in Chapter 3. To include the quantum mechanics effects, we solved both the Poisson and Schrödinger equations self-consistently in the DQW region. Considering the simple one-dimensional scenario, the underlying equations to find the conduction band and electron states are:

$$-\nabla \cdot \epsilon_s(x)\nabla\phi(x) = -\frac{\partial}{\partial x}\epsilon_s(x)\frac{\partial}{\partial x}\phi(x) = q\left[N_D(x) - n(x)\right], \tag{4.1}$$

$$-\frac{\hbar^2}{2}\nabla \cdot \frac{1}{m^*(x)}\nabla\psi(x) + V(x)\psi(x) = -\frac{\hbar^2}{2}\frac{\partial}{\partial x}\frac{1}{m^*(x)}\frac{\partial}{\partial x}\psi(x) + V(x)\psi(x) = E\psi(x), \tag{4.2}$$

where ϕ is the electrostatic potential, ϵ_s is the dielectric constant, N_D and n are the ionized donor and electron concentrations, ψ is the wave function, E is the energy,

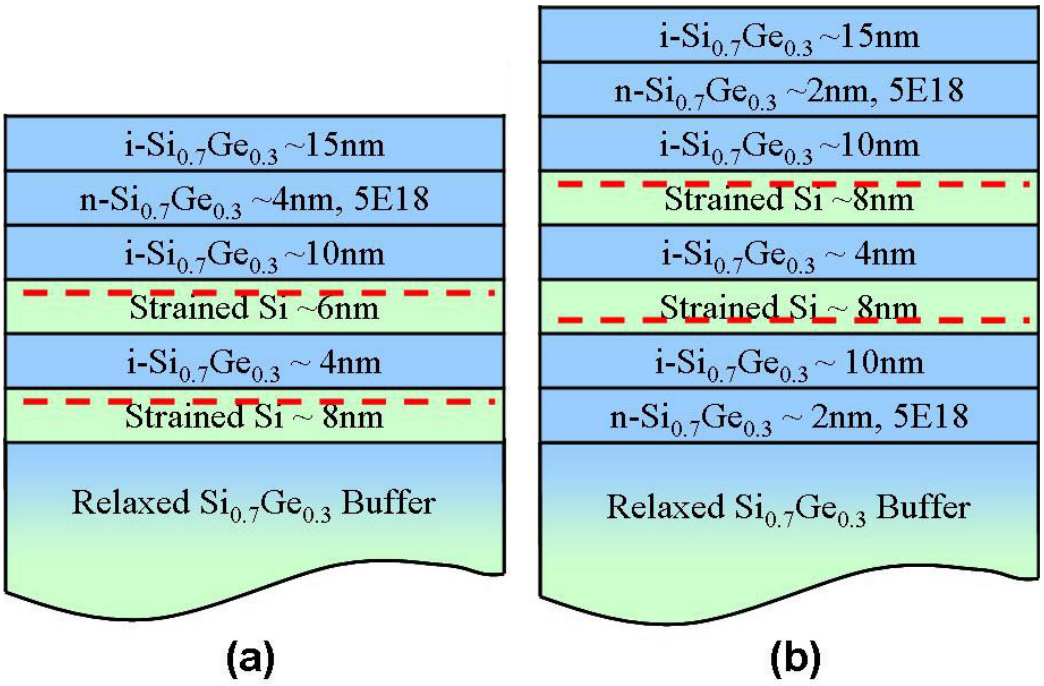


Figure 4.4: Schematic view of layer structures of two double quantum well systems: (a) asymmetric DQW with only one supply layer, (b) symmetric DQW with double supply layers. The red dashed line indicates a 2DEG.

and V is the potential energy, which is simply equal to the conduction band energy. The above two equations are related by:

$$V(x) = V_{CB}(x) + \phi(x), \tag{4.3}$$

$$n(x) = \sum_{k=1}^{m} \psi_k^*(x)\psi_k(x) \times \int_{E_k}^{\infty} \frac{\sqrt{2m^*}}{\pi\hbar\sqrt{E - E_k}} \left[\frac{dE}{1 + exp(E - E_k/kt)}\right], \tag{4.4}$$

where V_{CB} represents the conduction band edge potential at zero doping, the electron density is calculated by the summation over all the subbands from solving equation (4.2).

In all the following calculations, SiGe refers to $Si_{0.7}Ge_{0.3}$. For simplicity, a moderate level of 5×10^{16} cm^{-3} background doping of phosphorous is assumed in all grown Si and SiGe layers. The temperature is T = 4 K.

4.2.2 Asymmetric Double Quantum Wells

The first structural design contains two asymmetric quantum wells with doping supply from the top. The top channel is very thin to compensate for the fact that it is closer to the supply layer, in an effort to balance the electron densities in the two wells. The structure shown in Fig. 4.4 (a) is used for calculation. The Schrödinger equation is used between depth 50 Å and 800 Å from the top, outside this region only the Poisson's equation is solved assuming Boltzmann statistics. The model and materials parameters described in Appendix B are used. The boundary condition at the top surface is defined by a barrier height between the Fermi level and the conduction band $V_{Schottky} = E_f - E_c$. Fig. 4.5 shows the electron densities in each quantum wells and the total density with different barrier height.

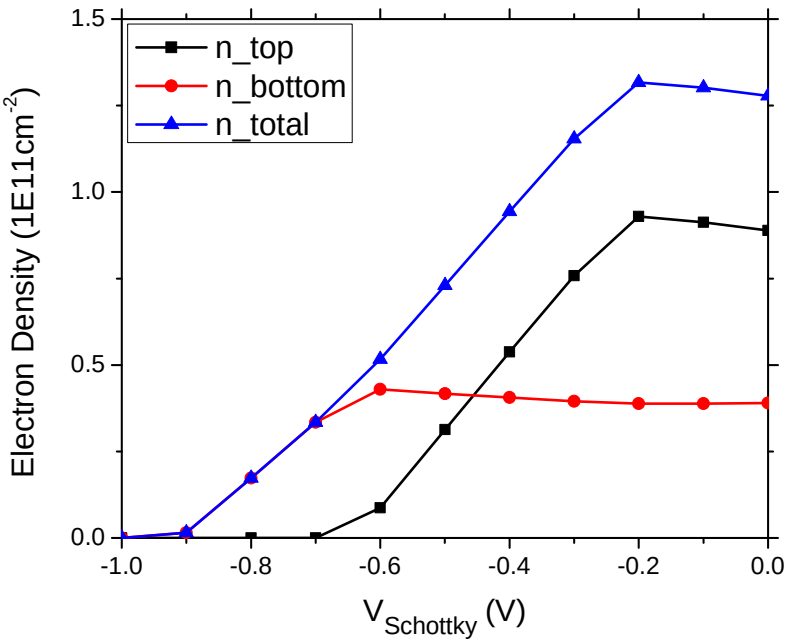


Figure 4.5: Electron densities in asymmetric DQW system with different surface Schottky barrier height applied.

The calculation shows that the balance of the two 2DEG densities occurs when the

surface Fermi level is at mid-gap with $V_{Schottky} = -0.45$ V. In the accumulation region the density in the bottom well remains fairly flat, only the density in the top well increases with more applied positive voltage. When the Fermi level at the top surface is raised close to the conduction band ($|V_{Schottky}| < 0.2$ V), the electron density in the DQW saturates. Further increasing the Schottky voltage will populate electrons only into the top SiGe layers outside the quantum wells. In the depletion region, after the top well is completely depleted, the density in the bottom well decreases at a slower rate, as a result of the larger spacing between the bottom 2DEG channel and the top gate.

Fig. 4.6 shows the band alignment diagrams for when the double well densities are balanced ($V_{Schottky} = -0.45$ V) and when the top quantum well is depleted ($V_{Schottky} = -0.8$ V). Since we are only interested in electrons, for simplicity only the conduction band edge is plotted. The Fermi level deep in the substrate buffer layers is pinned close to the conduction band edge due to background impurities, so the band alignment in the bottom well only varies a little compared to the dramatic change in the top well. Adding back-gating should greatly help to adjust the band alignment in the bottom well and keep the double well electron densities balanced at any total density level.

4.2.3 Symmetric Double Quantum Wells

The symmetric double quantum wells require two parallel 2DEG's with identical spacer and supply layers. Although conceptually simple, it can be very difficult for growth control due to the out-diffusion of dopants from the bottom supply layer towards top layers, which not only reduces the effective spacer thickness of the bottom well but also can cause increased background impurity doping.

The calculated electron densities are shown in Fig. 4.7. The Schrödinger equation is used in the region between 50 Å and 1000 Å from the top in the structure (Fig. 4.4 (b)). Since the bottom well has its own supply layer from the substrate side,

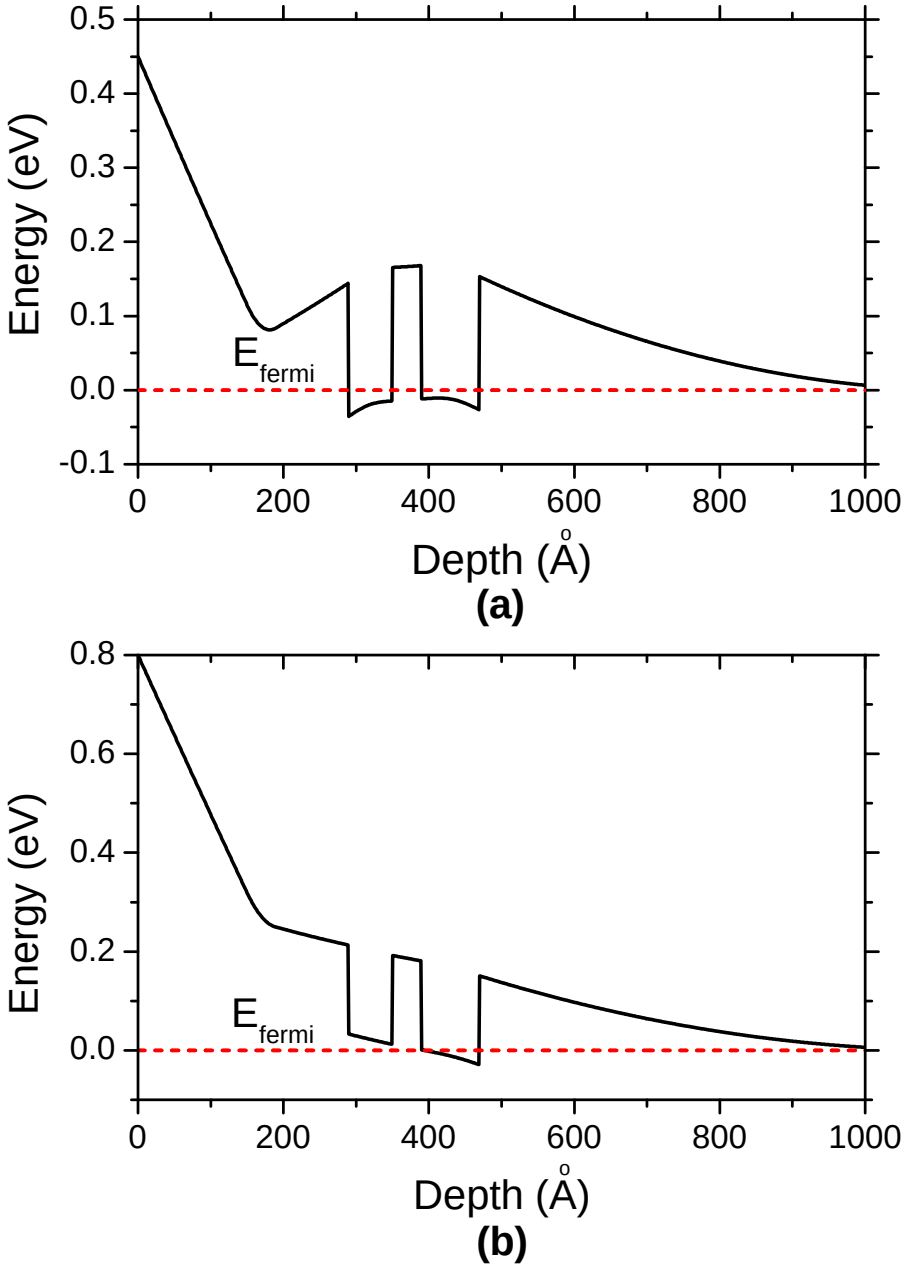


Figure 4.6: Conduction band diagrams in asymmetric DQW design with surface potentials at (a) $V_{Schottky} = -0.45$ V, (b) $V_{Schottky} = -0.8$ V.

the DQW band alignment should be symmetric when no surface Schottky barrier is applied and the electron densities in the two quantum wells are balanced. The top well is completely depleted when the surface Fermi level is at mid-gap with around $V_{Schottky} = -0.5$ V. The carriers in the bottom well are depleted much slower when more negative Schottky barrier is applied. There are still some electrons left in the bottom well ($\sim 0.5 \times 10^{11}$ cm^{-2}) even when $V_{Schottky}$ is as high as –1 V. In the accumulation region, the densities in the DQW remain almost balanced over a wide range as long as the top surface potential is close to conduction band edge. No more electrons are added to the DQW when the surface barrier passes the conduction band edge ($V_{Schottky} > 0$), and the density in top well cannot be tuned to much higher than that in the bottom well.

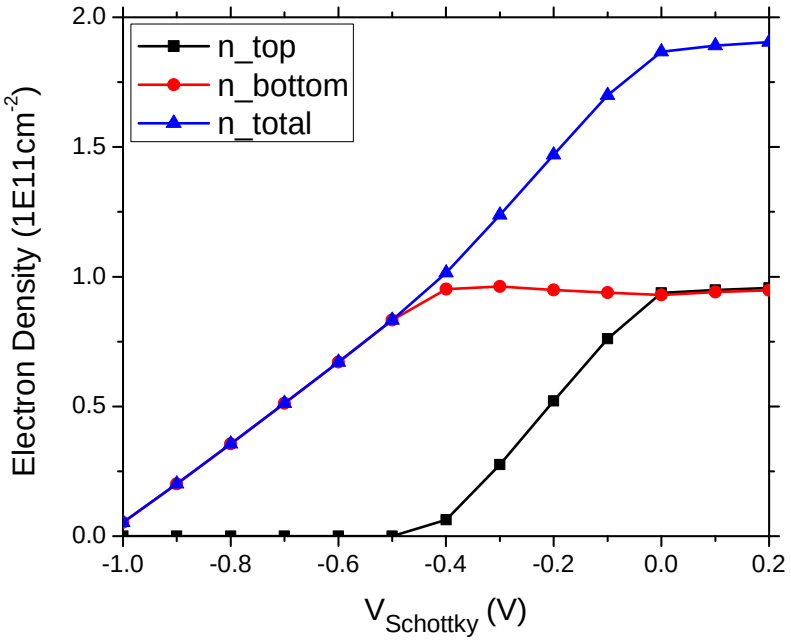


Figure 4.7: Electron densities in symmetric DQW system with different surface Schottky barrier heights applied.

We choose $V_{Schottky} = 0$ and $V_{Schottky} = -0.8$ V to check the energy band alignment

for balanced and depleted DQW, as shown in Fig. 4.8. Compared with the band alignment in the asymmetric DQW system, the Fermi level in the layers below the bottom well converges to the conduction band more quickly due to the bottom supply layer. Therefore without back-gating it is even more difficult to tune the density in the bottom well.

A very important feature of the symmetric DQW system is the symmetric and anti-symmetric split behavior. We can show this by calculating the lowest two quantized states assuming a flat-band condition with no dopants. Fig. 4.9 shows the wave functions of such lowest two quantized states. Clearly there is a large tunneling coupling between the two wells. In fact, the quantized states are no longer stationary states, they form symmetric and anti-symmetric subbands which are characterized by the energy gap Δ_{SAS}. The symmetric subband is of lower energy. This is directly analogous to the formation of a pair of bonding and anti-bonding orbitals in a hydrogen molecule. For the separation of 4 nm between the two silicon quantum wells in our model, our calculation predicts a $\Delta_{SAS} = 0.1$ meV, which is comparable to typical values of 0.1 - 1 meV measured in III-V compound semiconductor systems [75].

4.2.4 Comparison of Double Quantum Well Schemes

So far we have calculated the conduction band alignment, 2DEG densities and electron subband structures for both asymmetric and symmetric DQW schemes. The density balance of the two 2DEG channels can be achieved and controlled through a variable electric field induced by top-gating. A back gate is desirable to adjust band alignment below the bottom well and to allow independent control of electron densities in the two wells.

The asymmetric double quantum wells using only one top supply layer are easier to grow. This design offers a wider range of electron density in the top well that

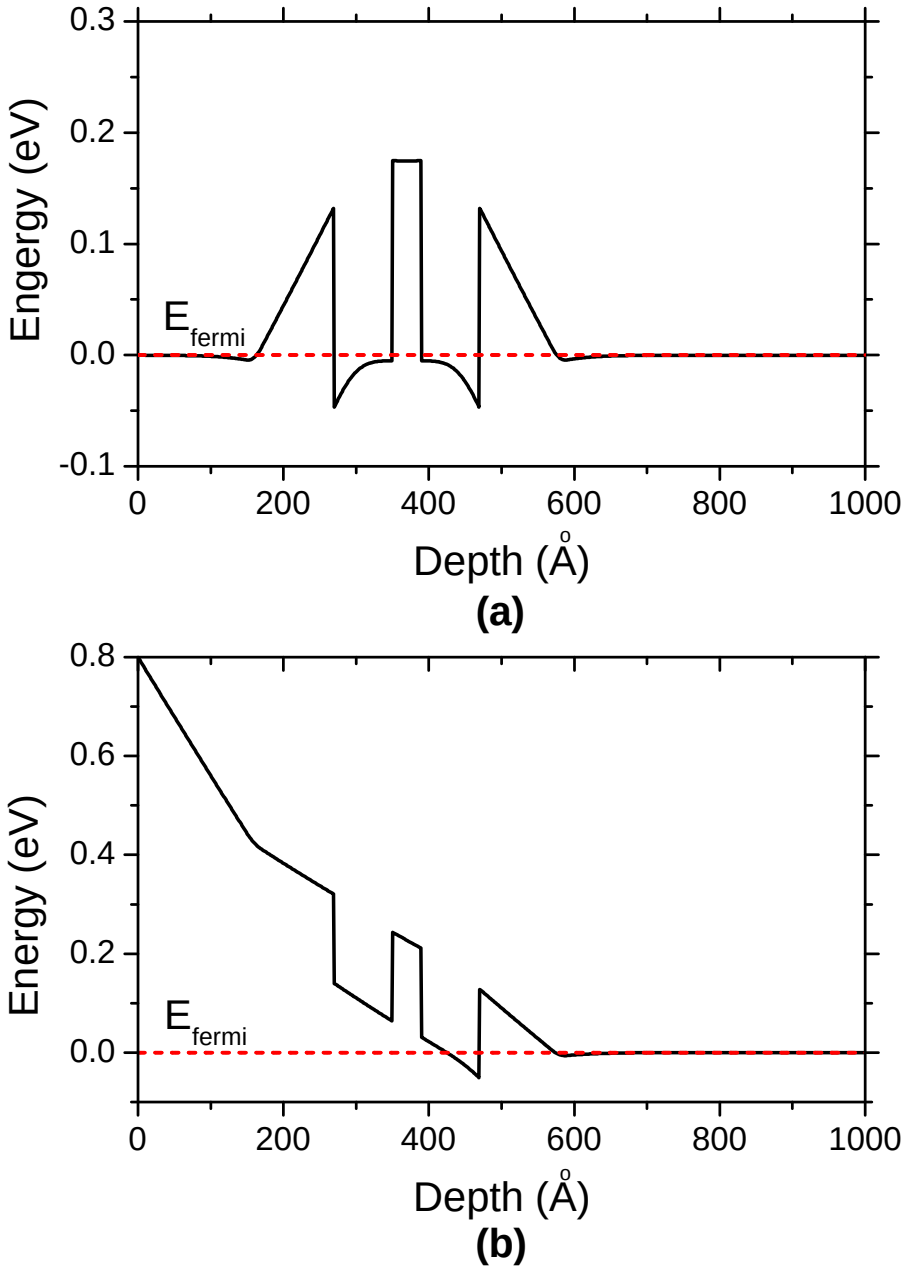


Figure 4.8: Conduction band diagrams in symmetric DQW designs with surface potentials at (a) $V_{Schottky} = 0$, (b) $V_{Schottky} = -0.8$ V.