STRAINED SILICON AND SILICON-GERMANIUM QUANTUM DEVICES BY CHEMICAL VAPOR DEPOSITION

JIUN-YUN LI

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDANCY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

ELECTRICAL ENGINEERING

ADVISOR: JAMES C. STURM

SEPTEMBER 2013

Abstract

Strained SiGe band-to-band tunneling (BTBT) devices and strained Si two-dimensional electron gases (2DEGs) are promising for low-power and quantum computing applications. The objective of this dissertation is to pursue the fundamental understanding of BTBT in strained SiGe films and electron transport properties in strained Si.

We report the first quantitative study of BTBT in strained p^+-SiGe/n^+-Si heterojunctions and p^+-SiGe/n^+-SiGe homojunctions at forward and reverse biases. Negative differential resistance (NDR) at forward bias is clearly observed for each device, with the highest observed peak current density of 10^4 A/cm^2. In reverse bias, a BTBT current density of 10^6 A/cm^2 is measured and a model comparison with good agreement is also presented. Furthermore, we demonstrate that the precise modeling of reverse-biased BTBT devices requires the observation of NDR in forward bias.

The surface segregation of phosphorus in relaxed SiGe films is studied with an extremely sharp phosphorus turn-off slope of 6 nm/decade reported. This enables effective Schottky gating on a depletion-mode device of a Si two-dimensional electron gas (2DEG). We also investigate the effect of surface hydrogen on phosphorus segregation. A phenomenological model for this segregation is proposed to explain the experimental results with good agreement.

A 2DEG with a record high mobility of 522,000 cm^2/V-s in an isotopically enriched ^{28}Si quantum well is presented. The estimated electron dephasing time of ~ 2 μs is presented. We investigate the effects of different layers in a Si 2DEG structure on

electron mobility and conclude that the remote impurity charges are the dominant source for electron scattering. The reduced segregation of phosphorus enables an inverted modulation-doped Si 2DEG with extremely high mobility of 470,000 cm^2/V-s. For the first time second subband occupancy was achieved in a Si quantum well.

Acknowledgements

First, I would like to thank my thesis advisor, Professor James C. Sturm for his guidance and support throughout my doctoral study in Princeton University. His encouragement and constant supports were essential to the completion of this thesis. Furthermore, his approach to advising students has been an invaluable asset for us. I am also very thankful to Professor Sigurd Wagner and Professor Jason Petta for spending time to read this thesis and providing many valuable comments to improve the level of this thesis. I would also like to thank Professor Stephen Chou and Professor Claire Gmachl for their time to serve on my FPO committee.

Special thanks to Professor Daniel Tsui for his encouragement, which was enlightening and inspiring. I greatly appreciate Professor Leonid Rokhinson at Purdue University for his help on magneto-resistance measurements of Si 2DEG samples and valuable inputs. I would also like to thank Professor Jason Petta for stimulating discussions on Si 2DEG work, and Dr. Steve Koester, Dr. Amlan Majumdar, and Dr. Isaac Lauer at IBM Watson Research Center for the stimulating discussions on tunneling work. I also like to thank Drs. Malcolm Carroll, James Ohlhausen, and Tzu-Ming Lu at Sandia National Laboratories for their support on SIMS analysis of isotopically enriched ^{28}Si 2DEG. Special thanks to Dr. Tzu-Ming Lu for his inputs on the work of Si 2DEGs.

I would like to thank the PRISM staff: Dr. Pat Watson, Dr. Mikhail Gaevski, and Joe Palmer for their excellent support. Special thanks to Pat for taking his time to read this thesis while he was on a trip! It's been fun and pleasure to work in PRISM cleanroom. I also appreciate the great assistance from Bob Hengstebeck at EAG to offer

very efficient SIMS analysis. This work cannot be done without his support.

I would like to thank the members in Professor Sturm's group: Kun Yao, Hongzheng Jin, Bahman Hekmatshoar, Sushobhan Avashi, Yifei Huang, Ting Liu, Chiao-Ti Huang, Amy Wu, Ken Nakamatsu, Bhadri Lalgudi Visweswaran, Warren Rieutort-Louis, Josue Sanz-Robinson, Janam Jhaveri, and Yasmin Afsar. Special thanks to Chiao-Ti Huang because in most of my phd life in Princeton, he was the only helper on RTCVD maintenance.

Finally, I would like to thank my family, especially to my wife. Without her tireless and continuous support, I would not have any chance to finish this thesis. She spends all her time to raise our two adorable daughters Irene and Leah together and always comforts me. She is not only a mother of two, a wife, but also the soul mate to me.

To my parents, thanks for being so supportive. Thank you brother to take care of them while I have been ~ 8,000 mile away from home. To my cute daughters, dad kept promise and finally graduated. Let's go home! Grandparents, we're coming home!

Contents

Lists of Tables

Lists of Figures

Chapter 1 Introduction

1.1 Motivation for Strained Si and SiGe Quantum Devices

Silicon has been dominating the semiconductor industry for over forty years because of the widely used Si metal-oxide-semiconductor field-effect transistor (MOSFET), the basic unit of a logic device. Si MOSFETs are probably the most fabricated electronic devices; for example, in 2011, there were more than 10^8 computers sold [1]. Assuming there were 10^9 transistors in a CPU such as Intel Core i7 [2] for each computer, at least 10^{17} Si MOSFETs were fabricated in a year.

The great success of Si MOSFETs may be attributed to Moore's law, which has successfully predicted the pace of transistor development over recent decades. However, to follow Moore's law further, several critical issues such as gate leakage, junction leakage, low transconductance, interconnect capacitance, and subthreshold conduction [3] must be solved. Furthermore, as a device is further scaled down, quantum mechanics must be used for nano-devices, and novel processing techniques have to be employed to create atom-sized structures in such small devices.

In the quantum regime, several interesting proposals for the next generation of computing devices have been proposed, such as tunneling transistors [4] and quantum computing [5]. A tunneling field-effect transistor (TFET) was first proposed by Baba [6] in 1992. The major benefit of this device is the low leakage current in the OFF state due to the sharp subthreshold slope, which is crucial for low-power applications.

Furthermore, because its structure is similar to a MOSFET and its fabrication steps are compatible with those of MOSFETs [7], remarkable TFET efforts have been made in the past decade.

On the other hand, for better system performance, quantum computing has been suggested to out-perform the classical computation by Shor [8]. He proposed a quantum algorithm with a much faster computation speed, as high as 10^{200} bits/second, by harnessing the power of quantum superposition in a quantum system. There are several candidates for the implementation of quantum computing, such as nuclear magnetic resonance [9], superconducting devices [10], and double quantum dots in semiconductors [11]. Quantum dot devices have drawn increasing attention from physicists and material scientists recently since the successful demonstrations of coherence control of quantum bits in GaAs [12] and Si [13]. The scalability and compatibility with semiconductor technology make quantum dots more feasible and promising than other proposed devices.

1.2 Thesis Organization

In this thesis, we first report band-to-band tunneling in strained SiGe for TFET applications. Then we present the electron transport properties in strained Si two-dimensional electron gas for quantum dot applications.

In chapter 2, band-to-band tunneling in strained p^+-SiGe/n^+-Si heterojunctions and p^+-SiGe/n^+-SiGe homojunctions, prepared by chemical vapor deposition (CVD), is investigated, and the experimental results are presented with a model comparison. This experimental verification of the model allows the device designers to predict the

dependence of band-to-band tunneling on germanium fraction and can be used in device simulators for the prediction of TFET performance and the relevant devices.

In chapter 3, the surface segregation of phosphorus in relaxed SiGe films grown by CVD is investigated. The effect of surface hydrogen on phosphorus segregation is explored and a phenomenological model including surface hydrogen is proposed. The model is in good agreement with the experimental results. A record sharp turn-off slope for phosphorus of 6 nm/dec is achieved.

A two-dimensional electron gas (2DEG) in strained Si is studied in chapter 4 for quantum dot applications. First, the basic physics of a 2DEG is reviewed. Efforts towards achieving high electron mobility in strained Si are described. The experimental results and the effects of layer structure on electron mobility are also presented.

For quantum computing applications, spin decoherence must be reduced for better device performance. ^{29}Si induces spin decoherence due to its nuclear spin via hyperfine interactions. In chapter 5, we present our work on the growth of 2DEGs in isotopically enriched ^{28}Si quantum wells. Furthermore, we report extremely high electron mobility in an inverted modulation-doped Si 2DEG grown by low-temperature epitaxy.

Finally in chapter 6, a brief summary is presented to conclude this thesis, and is followed by a few suggestions for future work.

Chapter 2 Band-to-Band Tunneling in Strained p^+-SiGe/n^+-Si Heterojunctions and p^+-SiGe/n^+-SiGe Homojunctions

2.1 Motivation

Tunneling is a fundamental quantum mechanical process. While quantum mechanics was developed in the 1920's, the first tunneling device was demonstrated in 1958 by Esaki in a heavily doped p-n junction of Ge [14]. Since then, semiconductor quantum devices, such as resonant tunneling diodes [15], quantum cascade lasers (QCLs) [16], and tunneling field-effect transistors (TFETs) [17], have been of great interest to scientists and device engineers. In the past five years particularly, extensive studies on TFETs have been conducted because of their potential for low-power applications [4].

Unlike a conventional metal-oxide-semiconductor FET (MOSFET), a TFET is operated by switching its tunneling junction on and off between the source and channel

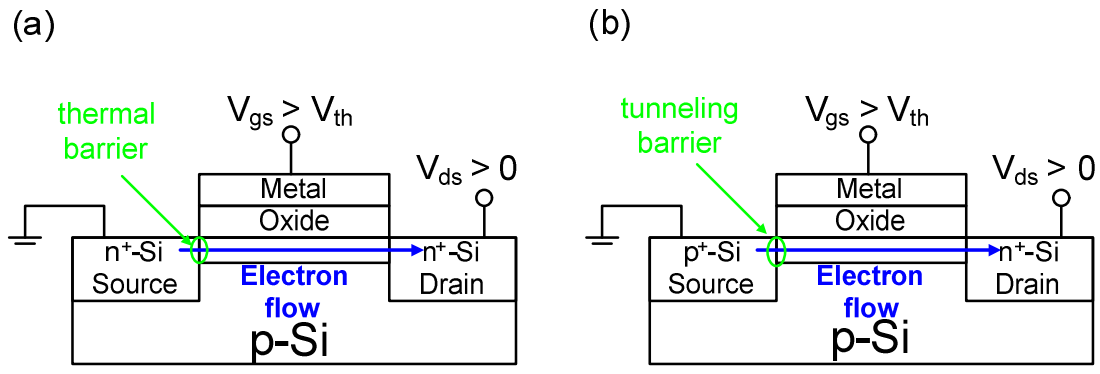

Fig. 2.1 Schematics of (a) n-type MOSFET and (b) n-type TFET in the ON state.

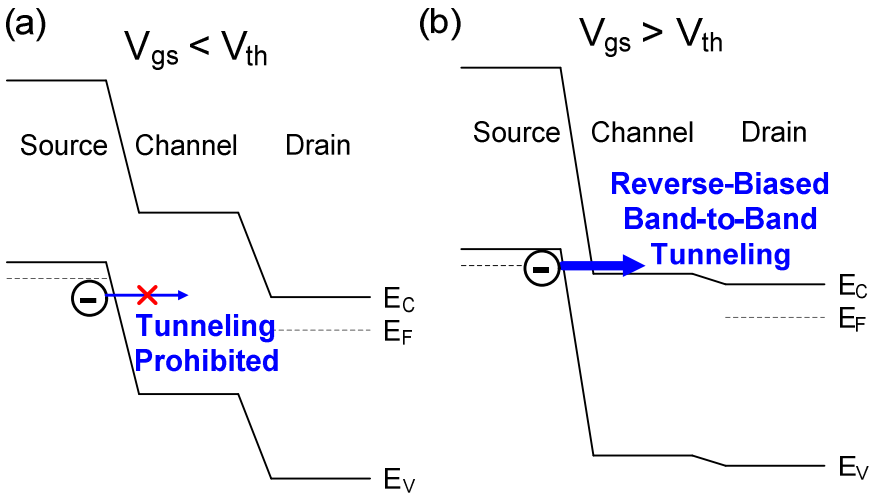

Fig. 2.2 Band diagrams of n-type TFET at (a) OFF state and (b) ON state.

regions. The device structures of an n-type MOSFET and an n-type TFET are illustrated in Fig. 2.1. A p^+-source is used in an n-type TFET, instead of an n^+-source in an n-type MOSFET. If the gate voltage (V_g) is below the threshold voltage (V_{th}), there is ideally no available state for electrons below the Fermi level, in the valence band of the p^+-source, to tunnel into the bandgap of the channel region for current conduction (Fig. 2.2 (a)). Furthermore, the distance for electrons in the source region to directly tunnel to the conduction band of the drain region is fairly large, leading to a very small leakage current at OFF state. On the other hand, when V_g is larger than V_{th}, the tunneling barrier between the source and channel regions is greatly reduced, such that electrons in the valence band of p^+-source can tunnel for current conduction (Fig. 2.2(b)). Because of this tunneling nature, the subthreshold slope can be smaller than the thermal limit of 60 mV/decade in a Si MOSFET at room temperature [18].

Si TFETs demonstrate great promise for low-power applications because of its sharp subthreshold slope. However, the ON state current in a Si TFET is fairly low

because of the large tunneling barrier resulting from the large bandgap of Si, limiting the potential for device performance (i.e. switching speed). SiGe has been considered a candidate because of its smaller bandgap energy and the compatibility of Si technology [19]. Furthermore, theoretical work has predicted that the subthreshold slope can be particularly sharp [20]. Two types of SiGe TFETs were proposed: (i) a heterojunction of p^+-SiGe source with a Si channel [20] and (ii) a homojunction of p^+-SiGe source with a SiGe channel [21]. In the following sections, we present the work on band-to-band tunneling in both p^+-SiGe/n^+-Si heterojunctions and p^+-SiGe/n^+-SiGe homojunctions, which serves as a baseline for TFET device modeling.

2.2 Fundamentals of Band-to-Band Tunneling

2.2.1 Negative Differential Resistance

The most important feature of band-to-band tunneling in a heavily doped p-n junction is the presence of negative differential resistance (NDR) in forward bias [22]. To observe NDR, two requirements have to be met. Firstly, a small tunneling barrier resulting from a large electric field (usually > 10^8 V/m) in the p-n region is required for tunneling to occur. Secondly, the Fermi level (E_F) must be located within the valence band in the p-type region and in the conduction band in the n-type region (Fig. 2.3); otherwise, a monotonic increase of tunneling current with applied voltage would be observed, such as Zener tunneling in reverse bias. Usually, in a degenerate doped p-n junction, both requirements are satisfied. For example, if the doping levels are 5×10^{19} cm^{-3}, the electric field can be as large as 3×10^8 V/m by a device simulator [23],

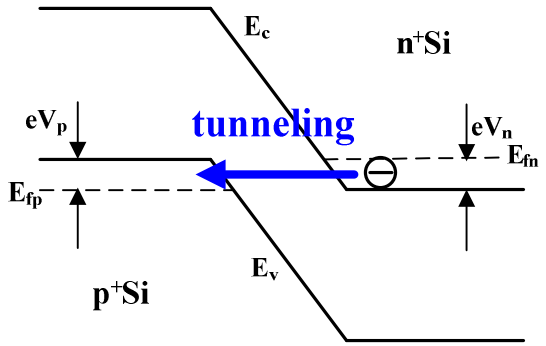

Fig. 2.3 A degenerately doped p-n junction in Si at a small forward bias. eV_n and eV_p represent the energy difference between the Fermi level in the n-type region and the conduction band edge, and the Fermi level in the p-type region and the valence band edge, respectively.

leading to a short tunneling distance of ~ 10 nm. At this doping level, E_{fp} is 88 meV below the edge of the valence band in the p-type region, and E_{fn} is 94 meV above the conduction band edge in the n-type region, at small forward biases.

In Fig. 2.4, the band diagrams (top) and the associated I-V curves (bottom) of a degenerately doped tunneling diode are plotted. Initially, in reverse bias, Zener tunneling occurs with the electrons in the valence band of the p-type region tunneling to the conduction band of the n-type region (Fig. 2.4(a)). As the negative bias is increased, the electric field in the p-n junction increases. Therefore, the tunneling barrier for electrons becomes smaller, leading to a higher current.

At forward bias, the tunneling direction reverses. The electrons in the conduction band of the n-type region now tunnel to the available states in the valence band of the p-type region at the same energy level. At a certain bias (V_2), a band of electron energy in the n-type conduction band, aligned with a band of the unoccupied states in the p-type valence band, leads to a peak current in forward bias (Fig. 2.4(b)). As the applied voltage is increased slightly, some electrons at the highest energy levels in

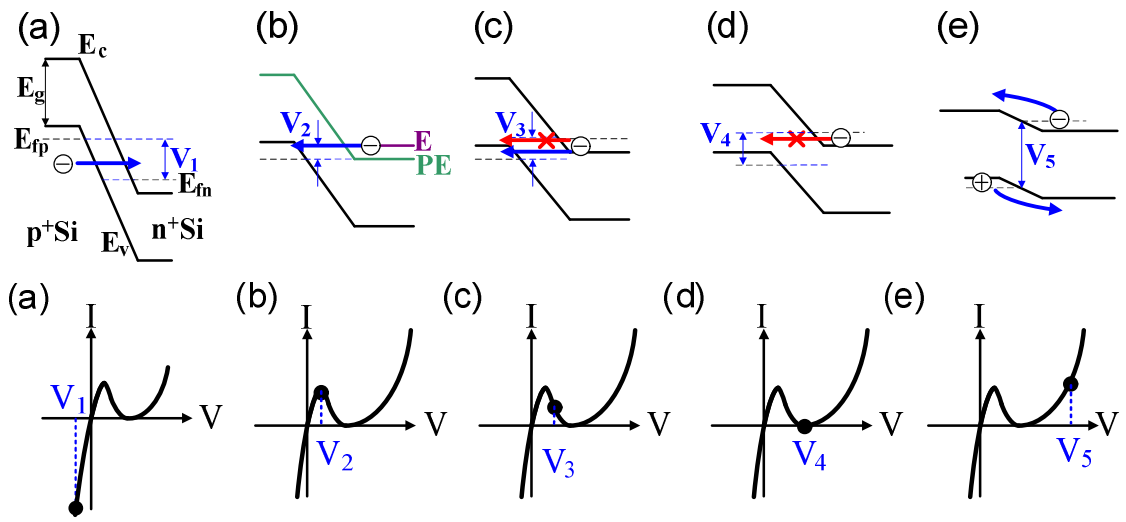

Fig. 2.4 Band diagrams (top) and I-V curves (bottom) of a tunneling diode at different biases.

the n-type conduction band cannot tunnel to the p-type valence band for current conduction, since there is no available state at the same energy levels in the bandgap of the p-type region (red arrow in the top part of Fig. 2.4(c)). As a result, the current starts to drop (Fig. 2.4(c)). As the entire band filled with electrons in the n-type region is completely "uncrossed" with all available states in the valence band of the p-type region, tunneling ceases and the current becomes zero (Fig. 2.4(d)). When the applied bias is further increased, more electrons (holes) in the conduction (valence) band of n-type (p-type) region have enough energy to move towards the junction, with normal diode characteristics (Fig. 2.4(e)).

Typically, the WKB approximation (Wentzel-Kramers-Brillouin method) is used to estimate the tunneling probability T_t [24]:

$$T_t \approx \exp\left[-2\int_{x_p}^{x_n} |k(x)|\,dx\right], \tag{2.1}$$

where $|k(x)|$ is the absolute value of the wavevector of the electron in the tunneling

barrier, and x_p and x_n are the positions of the valence band edge and conduction band edge in the p-type and n-type regions, respectively. The wavevector is

$$k(x) = \sqrt{\frac{2m^*}{\hbar^2}(PE - E)}\,, \tag{2.2}$$

where m^* is the effective mass of the tunneling electron, PE is the potential energy, and E is the electron energy (top of Fig. 2.4 (b)). The tunneling probability is

$$T_t \approx \exp\left(-\frac{\pi\sqrt{m^* E_g^3}}{2\sqrt{2}e\hbar F_{field}}\right)\exp\left(-\frac{2E_\perp}{E_{eff}}\right), \tag{2.3}$$

where E_g is the bandgap energy, F_{field} is the peak electric field in the p-n junction, $E_\perp$ is the energy associated with momentum perpendicular to the tunneling direction, and E_{eff} is the measure of transverse momentum as given by [24]

$$E_{eff} \equiv \frac{4\sqrt{2}e\hbar F_{field}}{3\pi\sqrt{m^* E_g}}. \tag{2.4}$$

In thermal equilibrium, the net tunneling current I_{net} in a tunneling diode from the p-type region to the n-type region is the difference between tunneling current of $I_{p\to n}$ and $I_{n\to p}$, as given by

$$I_{p\to n} \equiv B\int_{E_C}^{E_V} F_V(E)n_V(E)T_t[1 - F_C(E)]n_C(E)dE\,, \tag{2.5}$$

$$I_{n\to p} \equiv B\int_{E_C}^{E_V} F_C(E)n_C(E)T_t[1 - F_V(E)]n_V(E)dE\,, \tag{2.6}$$

where B is a pre-factor, $F_C(E)$ and $F_V(E)$ are the Fermi-Dirac distribution functions, T_t is the tunneling probability and is assumed to be equal for both directions, and $n_C(E)$ and $n_V(E)$ are the density of states in the conduction band and valence band respectively.

Thus, the peak tunneling current can be calculated as [25]

$$I_{peak} = A\frac{em^*}{4\pi^2\hbar^3}\exp\left(-\frac{\pi\sqrt{m^*E_g^3}}{2\sqrt{2}e\hbar E_{field}}\right)E_{eff}\cdot D\,, \tag{2.7}$$

$$D \equiv \int[F_C(E)-F_V(E)]\cdot\left[1-\exp\left(-\frac{2E_S}{E_{eff}}\right)\right]dE\,, \tag{2.8}$$

where A is the device area, D is the effective density of states, E_S is the smaller of E_{fp} and E_{fn}, which are the energies of the hole and electron, measured from the edges of the p-type valence band and n-type conduction band respectively.

To enhance tunneling current for TFET applications, a large electric field and a small bandgap energy are preferred. By increasing the p-type doping level in the source region of an n-type Si TFET, a higher ON state current is expected. Moreover, by replacing Si with SiGe in the source region, the tunneling barrier between the source-channel junction becomes smaller due to the smaller bandgap energy of SiGe than Si. Therefore, the tunneling can be boosted further [7, 21]. Consequently, in this chapter we focus on the dependence of band-to-band tunneling (BTBT) on the p-type doping level and Ge fraction, in both forward and reverse biases. Experimental results and model comparisons will be presented in the following sections.

2.2.2 Defect-Assisted Tunneling

In the preceding section, the fundamentals of tunneling were introduced qualitatively and quantitatively, based on an assumption that the p-n junction is free of defects. The derivation of the tunneling model was proceeded by assuming an ideal p-n

junction without any defects. However, in practice, defects exist in the junction and affect the tunneling characteristics. For all reported tunneling junctions made by degenerately doping, an "excess" current at biases between negative differential resistance and the onset of the thermal current was observed [22], and is defined as defect-assisted tunneling (DAT) current in this thesis. Chynoweth *et al*. suggested that this excess current results from an additional tunneling mechanism via junction defects [26] at biases, where band-to-band tunneling is prohibited due to the complete misalignment of energy bands of the n-type and p-type regions.

By intentionally introducing defects into the junction, Sah [27] and Weaver [28] confirmed the existence of defect-assisted tunneling, and showed that NDR disappeared when this excess tunneling current swamps the BTBT current. Thus, for an accurate measure of BTBT in a p-n junction, the defect-assisted tunneling component has to be eliminated (or heavily reduced) so that NDR can be observed.

In Fig. 2.5, a tunneling junction with defect states in the bandgap is illustrated [26]. At this bias, the electron band in the n-type region is completely uncrossed with the valence band in the p-type region, and is still far away from the normal diode operation;

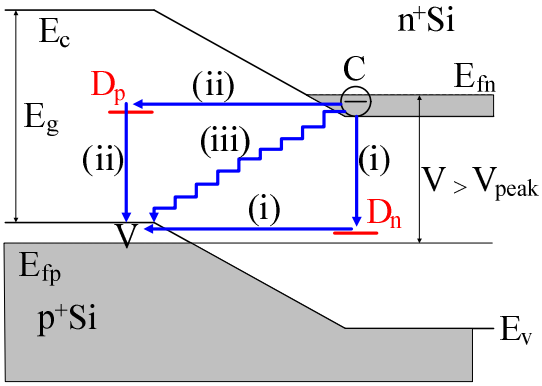

Fig. 2.5 Band diagram in a degenerately doped p-n junction, which shows the possible paths of electron tunneling from the n-type conduction band to the p-type valence band, via defects in the bandgap [26]. The shadowing areas represent the energy bands filled with electrons.

thus, there is no tunneling. With the presence of defects in the junction, electrons in the conduction band of the n-type region can actually move to the valence band of the p-type region via several paths, such as (i) C to D_p to V, (ii) C to D_n to V, or (iii) C to V directly. For (i), electrons first tunnel to the defect state D_p and recombine with holes in the valence band. Conversely, for (ii) electrons move to the defect state D_n first, and then tunnel to the valence band. The last route for defect-assisted tunneling involves the existence of continuous defect levels between the conduction band and the valence band, where an electron loses its energy by these continuous transitions.

For simplicity, Chynoweth *et al*. assumed that path (ii) would dominate and that the limiting step is the tunneling from D_n to V, but not the recombination of C to D_n [26]. The defect-assisted tunneling (DAT) current is given by

$$I_{DAT} = KD_x \exp\left\{-\alpha_x \left[E_g - eV + 0.6\left(E_n + E_p\right)\right]\right\}, \tag{2.9}$$

where K is a pre-factor, D_x is the defect density, α_x is a constant, and V is the applied

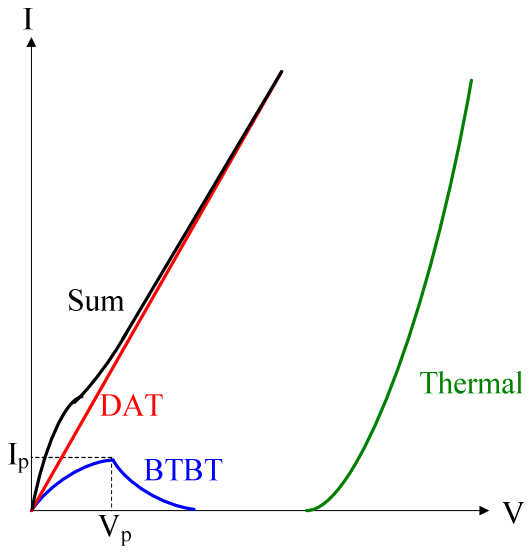

Fig. 2.6 Three main current components in a tunneling diode in practice: BTBT current (blue), DAT current (red), and thermal current (green), with their sum also shown (black). There is no NDR, because DAT dominates over BTBT due to high defect density in the p-n junction.

voltage. The main factors of DAT current are the defect density, bandgap energy, and the applied voltage (electric field). With a smaller bandgap or a larger bias (stronger electric field), the tunneling probability from D_n to V becomes higher, with the result that I_{DAT} increases. Between the peak voltage and the onset of the thermal current, the current never falls to zero because of the presence of defect-assisted tunneling. Fig. 2.6 shows I-V curves of a tunneling diode in practice with high defect density, where DAT swamps BTBT at small biases so that NDR disappears.

2.3 Band-to-Band Tunneling in p^+-SiGe/n^+-Si Heterojunctions

2.3.1 Device Fabrication

For tunneling FET applications, SiGe has been suggested to replace Si for higher tunneling currents due to its smaller bandgap energy [20]. In this section we focus on band-to-band tunneling in p^+-SiGe/n^+-Si heterojunctions. Since the level of n-type dopant is limited to the level of 10^{19} cm^{-3} due to the low incorporation [29], for the heavily doped n^+-Si layer, we used ion implantation of phosphorus with multiple steps to achieve a high doping level of 2×10^{20} cm^{-3} (Table 2.1). Due to the high-dose implantation, Si substrates became amorphous and defective. Thus, an annealing step is required to re-crystallize the Si substrates and to remove the defects by implantation.

Table 2.1 Steps of ion implantation of phosphorus for n^+-Si layer

Species: Phosphorus	**Step 1**	**Step 2**	**Step 3**	**Step 4**
Energy (keV)	15	40	80	120
Dose (cm^{-2})	5×10^{14}	7×10^{14}	1×10^{15}	2×10^{15}

Prior to being annealed, the implanted Si (100) wafers of phosphorus level of ~ 10^{20} cm^{-3} were cleaned by H_2SO_4:H_2O_2 (2.5:1) for 15 minutes followed by diluted HF (1:100) in deionized (DI) water for 2 minutes to remove the residual oxide. Then the wafers were loaded into the CVD reactor for high temperature annealing (700 ~ 1050 ^oC) for 5 minutes, followed by p^+-SiGe epitaxial growth. The growth pressure was 6 torr. The temperatures for SiGe growth were 625 ^oC for Ge fractions of 14, 21, and 27%, and 575^oC for 35 and 39%. The gas precursors were dichlorosilane (SiH_2Cl_2) and diluted germane (0.8% GeH_4 in hydrogen) for SiGe growth. Diluted diborane (100 ppm in hydrogen) was used for in-situ p-type doping. The thicknesses of the SiGe films were kept below the critical thickness for each Ge fraction to avoid strain relaxation, which could induce dislocation defects in the junction. After the epitaxial growth, the wafers were mesa-etched by reactive-ion etching (RIE) to isolate the p-n junction and to define the device area of 25 μm × 25 μm. The etching gases were CF_4 of 50 sccm and O_2 of

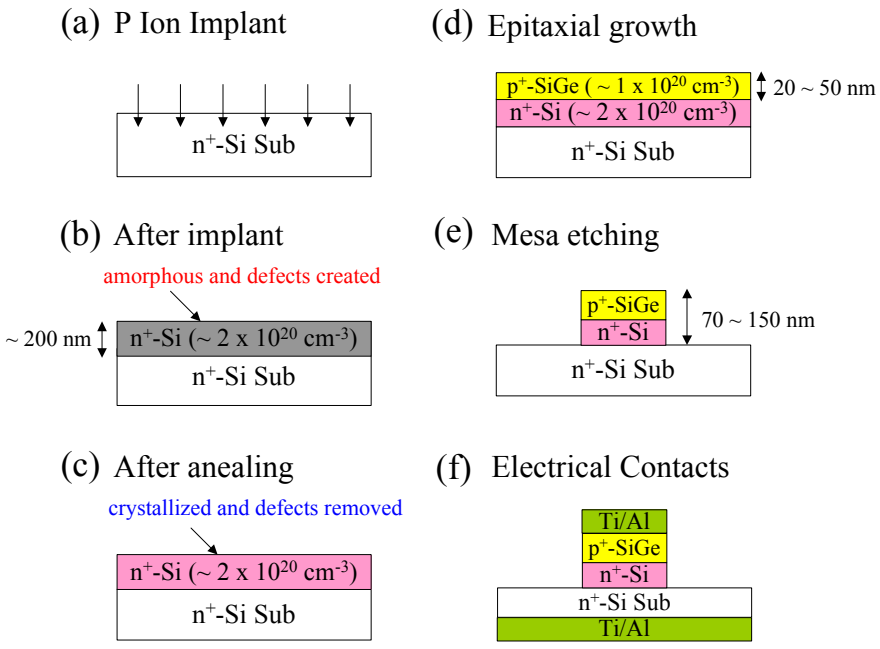

Fig. 2.7 Fabrication steps of p^+-SiGe/n^+-Si heterojunction tunneling diodes.

10 sccm. The chamber pressure was 100 mtorr and the RF power was 100 W. The etching rate of SiGe films was ~ 100 nm/min. Lastly, Ti/Al was deposited on the top and bottom surfaces for electrical contacts. The details of the fabrication steps are illustrated in Fig. 2.7. I-V measurements were performed at room temperature.

Table 2.2 Growth parameters of p^+-SiGe layers

Ge Fraction (%) and growth temp	**Gas flow rates (sccm) (DCS/GeH_4/B_2H_6)**	**Growth Rate (nm/min)**	**Thickness of p^+-SiGe (nm)**	**Doping level (cm^{-3})**
14 (625°C)	26/50/100	3	25	1.0×10^{20}
21 (625°C)	26/100/125	6.5	44	1.3×10^{20}
27 (625°C)	26/200/167	10	20	1.8×10^{20}
35 (575°C)	26/300/250	6.8	16	2.5×10^{20}
38 (575°C)	26/400/350	5.7	14	0.8×10^{20}

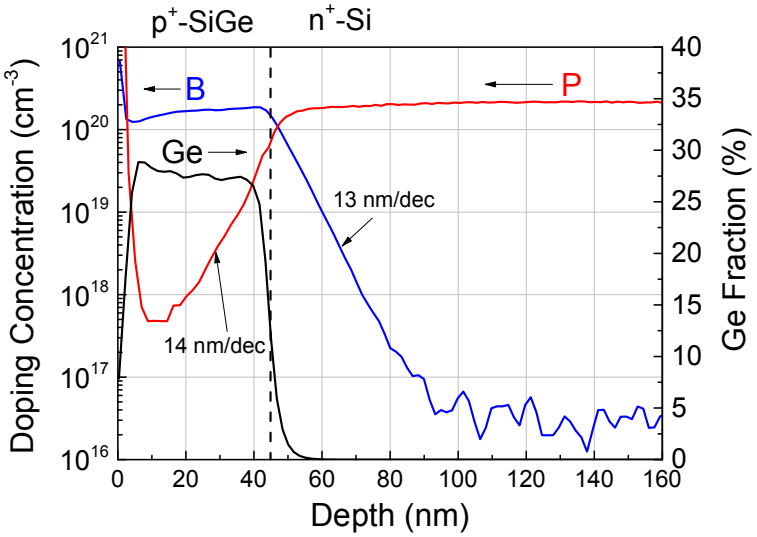

Fig. 2.8 SIMS analysis of a p^+-$Si_{0.73}Ge_{0.27}$/n^+-Si heterojunction tunneling diode (sample #5150). The slopes of B trailing edge and P leading edge are 13 and 14 nm/decade respectively.

The information of Ge fraction, gas flow rates, growth rates, layer thicknesses, and doping levels were collected by secondary ion mass spectrometry (SIMS) and listed

in Table 2.2. A typical SIMS profile is shown in Fig. 2.8. For this device, the annealing temperature was 950°C, and the phosphorus level after annealing is 2×10^{20} cm^{-3} and flat. The boron level is ~ 1.5×10^{20} cm^{-3}. The slopes of the boron trailing edge and phosphorus leading edge are 13 nm/decade and 14 nm/decade respectively. The former is believed to be an artifact of SIMS due to the knock-on effect [30], and the later is caused by surface segregation during epitaxy [31], which will be discussed in chapter 3.

2.3.2 Effects of Annealing Temperature on Negative Differential Resistance

We now study the effects of annealing temperature on the presence of negative differential resistance in the implanted tunneling diodes. Several p^+-$Si_{0.73}Ge_{0.27}$/n^+-Si tunneling diodes with different annealing temperatures to remove the junction defects were fabricated and were measured at room temperature, with their J-V curves shown in Fig. 2.9. The boron and phosphorus levels are 1.5×10^{20} cm^{-3} and 2×10^{20} cm^{-3} respectively. With 700 °C annealing, NDR was not observed because its defect-assisted tunneling current is higher than the band-to-band tunneling current at small forward biases, due to the lack of defect annealing. To observe NDR, a higher annealing temperature is required to remove the defects. For the device annealed at 800 ~ 950°C, NDR was observed. For 1050°C annealing there was no NDR, most likely because strong dopant diffusion at such high temperatures reduces the junction abruptness and the electric field, leading to a greatly reduced tunneling current despite the lower defect density. As a figure of merit, the peak-to-valley current ratio (PVCR) at forward bias is usually used to indicate the junction quality in a tunneling diode [22]. As shown in Fig. 2.9, a best PVCR of 2 was achieved in the device annealed at 900°C.

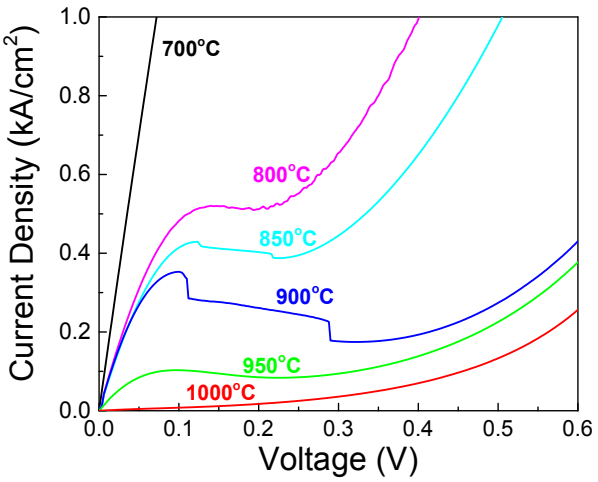

Fig. 2.9 J-V curves of p^+-$Si_{0.73}Ge_{0.27}$/n^+-Si heterojunction tunneling diodes at different annealing temperatures.

Note that for some devices, a bump or hump of current in the region of NDR was observed (e.g., diodes annealed at 850 °C and 900 °C in Fig. 2.9). We confirmed this was due to the oscillations in the measurement circuit [32], which includes an Agilent 4155C, electrical cables and probe station, and the tunneling devices. When the oscillations occur, a dc I-V measurement contains ac components resulting from the resonance of the entire circuit. As the tunneling diode enters the NDR region biased at a voltage slightly larger than the peak voltage, ac components of current are induced by the ac voltage swings. Near the peak voltage, any ac voltage swing results in the reduction of the tunneling because of the fewer available states in the p-type valence band at the same energy levels of the electrons in the n-type conduction band due to the ac voltage deviations from the peak voltage. Thus, a dc I-V measurement gives a sharp drop in current near the peak voltage, and no other physics is involved.

To further investigate the effects of annealing temperature on BTBT and DAT

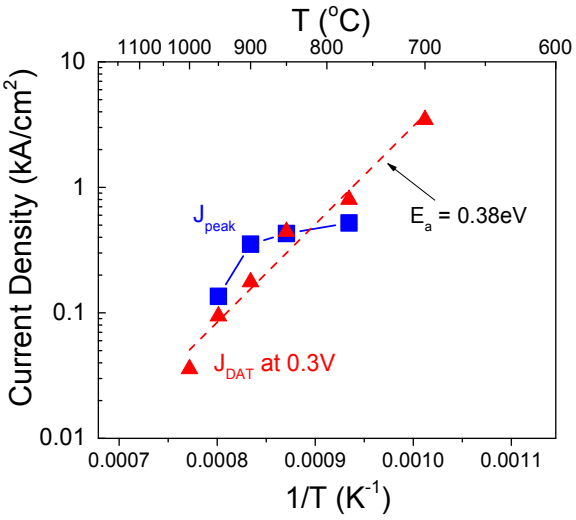

Fig. 2.10 J_{peak} and J_{DAT} (at 0.3 V) of p^+-$Si_{0.73}Ge_{0.27}$/n^+-Si heterojunction tunneling diodes vs. annealing temperature.

current, we plotted the peak current density (J_{peak}) and the DAT current density (J_{DAT}) at 0.3 V versus annealing temperature in Fig. 2.10. For DAT, as the annealing temperature was increased, J_{DAT} monotonically decreased due to fewer junction defects. For BTBT, as the annealing temperature was increased, J_{peak} first decreased slightly, before dropping sharply at T > 900°C until no NDR was observed at 1050°C. In practice, the peak current represents the sum of the BTBT and DAT components at the peak voltage. For the annealing temperature of 800 ~ 900°C, J_{DAT} decreases with annealing temperature, so J_{peak} decreases as well, assuming to first order the BTBT currentis constant. At 1050°C, despite the high efficiency of defect removal, the abruptness of doping profiles is reduced even more due to strong dopant diffusion, leading to a smaller electric field and less tunneling. Thus, J_{peak} drops further until NDR is gone.

2.3.3 Effects of Electric Field and Bandgap Energy on Band-to-Band Tunneling

To observe NDR in implanted p^+-SiGe/n^+-Si heterojunctions, a post-annealing

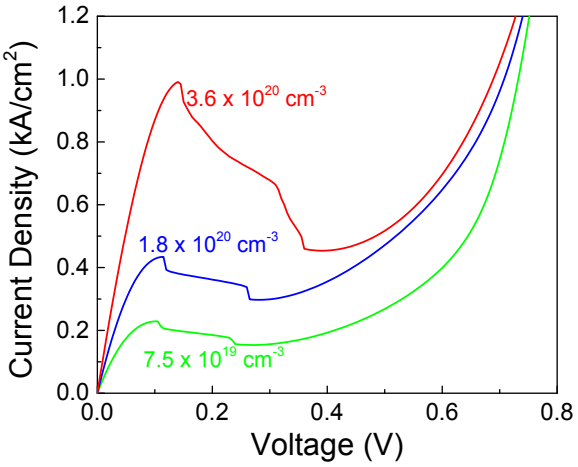

Fig. 2.11 J-V curves of p^+-$Si_{0.73}Ge_{0.27}$/n^+-Si heterojunction tunneling diodes of three different boron concentrations.

step at certain temperatures is required. In the last section, we found that the best annealing temperature was 900 ^oC which gave the highest PVCR. Thus, for all heterojunctions devices of different boron levels and Ge fractions investigated in this section, 900^oC annealing was applied.

Firstly, p^+-$Si_{0.73}Ge_{0.27}$/n^+-Si diodes of a fixed n-type doping (phosphorus) of 2 $\times$ 10^{20} cm^{-3}, and three different p-type doping levels (boron) at 0.75, 1.8, and 3.6 $\times$ 10^{20} cm^{-3}, were fabricated. The J-V curves are shown in Fig. 2.11. For each device, NDR was clearly seen. As the boron level increases, the resulting higher electric field reduces the tunneling barrier, so the peak tunneling current increases from 0.2 to 1 kA/cm^2. While BTBT current increases with the boron level, DAT current also increases because of its tunneling nature.

Next, we present the results of the dependence of tunneling on Ge fraction in p^+-$Si_{1-x}Ge_x$/n^+-Si diodes with various Ge fractions of 14, 21, 27, 35, and 38%. The

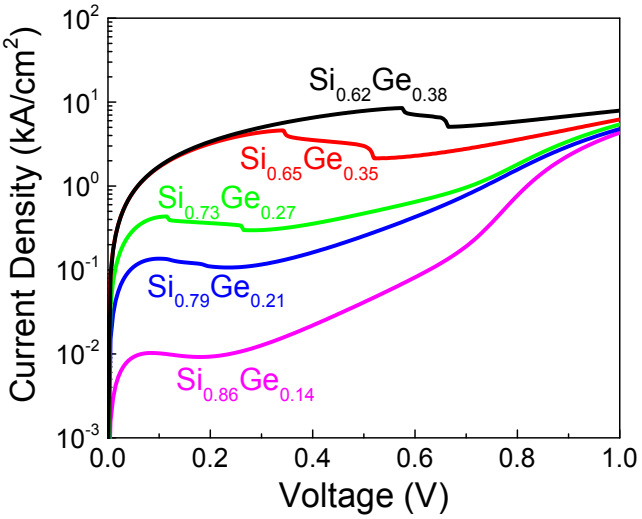

Fig. 2.12 J-V curves of p^+-$Si_{1-x}Ge_x$/n^+-Si tunneling diodes of different Ge fractions.

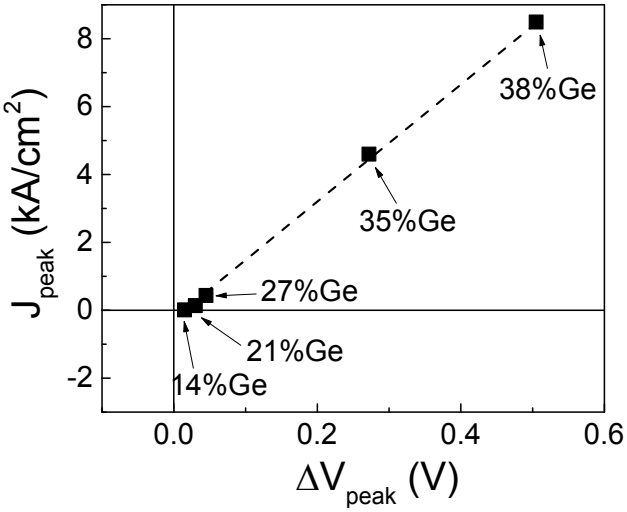

Fig. 2.13 J_{peak} vs. ΔV_{peak} for different Ge fractions. The linear dash line represents the spreading resistance of n^+-Si substrate.

annealing temperature for these devices was 900°C, and J-V curves are shown in Fig. 2.12. As the Ge fraction increases, the tunneling barrier reduces due to the smaller bandgap energy, leading to stronger tunneling. As the Ge fraction increases from 0.14 to

0.39, the peak tunneling current density increases by a factor of 10^4 from 0.01 to 9 kA/cm^2, which is believed to be the highest reported among all Si-based tunneling diodes by CVD. The peak voltage shifts to a larger value with Ge fraction because of the series resistance. In those devices, the series resistance is dominated by the spreading resistance in the Si substrate, so it is approximately constant for each device. Thus, the shift of the peak voltage (ΔV_{peak}) is proportional to the peak current density and increases with the Ge fraction (Fig. 2.13).

2.4 Band-to-Band Tunneling in p^+-SiGe/n^+-SiGe Homojunctions

2.4.1 Device Fabrication

In this section, we discuss band-to-band tunneling in p^+-SiGe/n^+-SiGe homojunctions. Similar to the fabrication steps of p^+-SiGe/n^+-Si heterojunction tunneling diodes, the Si substrates were prepared by the same ex-situ cleaning steps, before being loaded into the CVD reactor. Unlike for the heterojunction devices, both p^+ and n^+ layers were epitaxially grown by in-situ doping CVD. Without implant steps, radiation damage could be eliminated completely. Furthermore, the abruptness of the doping profiles could be preserved, since post-annealing was no longer required before epitaxial growth. Before growth, high temperature baking at 850^{o}C for 5 minutes was performed to remove the residual oxide on the Si substrates. SiH_2Cl_2 and diluted GeH_4 (0.8% in H_2) were the precursors for SiGe growth. Diluted phosphine (PH_3) and diborane (B_2H_6) (both 100 ppm in H_2) were used for n-type and p-type in-situ doping.

Table 2.3 Growth parameters of p^+-SiGe and n^+-SiGe layers

Ge Fraction (%) (growth temp)	Gas flow rates (sccm) (DCS/GeH_4/B_2H_6/PH_3)	GR (nm/min)	Thickness of p^+-/n^+-SiGe (nm)	Critical thickness (nm)	Doping levels of p^+-/n^+-SiGe (cm^{-3})
14 (625°C)	26/50/100/100	3	32/18	450	$1.3/0.7\times10^{20}$
21 (625°C)	26/100/125/125	6.5	44/22	180	$1.3/1.1\times10^{20}$
27 (625°C)	26/200/167/167	10	25/16	80	$1.8/1.2\times10^{20}$
35 (575°C)	26/300/250/250	6.8	15/12	35	$3.5/2.4\times10^{20}$

After baking, an n^+-SiGe layer was immediately grown, followed by the deposition of a p^+-SiGe layer by fast switching between the doping gases. The diodes with Ge fractions of 14, 21, and 27% were grown at 625°C with thicknesses of 50, 66, and 40 nm, and the diode with a Ge fraction of 35% was grown at 575°C with a thickness of 30 nm. The layer thicknesses, Ge fraction, and doping concentrations were determined by SIMS (Table 2.3). Metastable critical thicknesses of strained SiGe layers on Si substrates for Ge fraction of 14, 21, 27, and 35% are 450, 180, 80, and 35 nm, respectively [33]. Therefore, we expected the layers to be biaxially compressively strained, pseudomorphic to the Si substrates. Lastly, square mesas were dry-etched with an area of 25μm × 25 μm. Ti/Al was evaporated on the bottom of the wafers for the n-side ohmic contact. For the p-side ohmic contact, Ti/Al was patterned on top of the mesa by a combination of photolithography and lift-off with a contact area of 23 μm × 23 μm between Ti/Al and the mesa.

2.4.2 Forward-Biased Band-to-Band Tunneling

First, we studied the effect of p-type doping levels on BTBT by varying the boron concentration (N_A: 1.7 to 5.1 × 10^{20} cm^{-3}) with a fixed phosphorus level of N_D ~

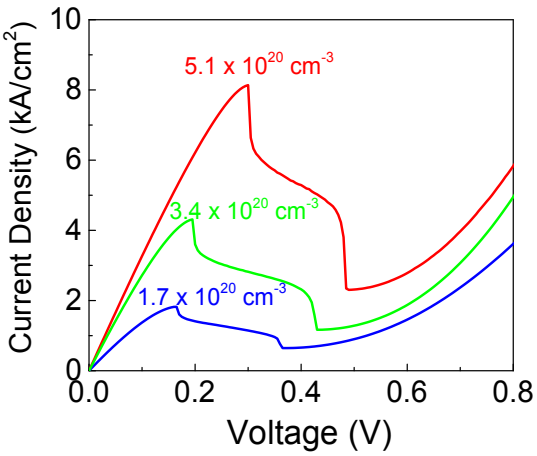

Fig. 2.14 J-V curves of p^+-$Si_{0.73}Ge_{0.27}$/n^+-$Si_{0.73}Ge_{0.27}$ homojunction tunneling diodes of three different boron levels.

1.0×10^{20} cm^{-3} in $Si_{0.73}Ge_{0.27}$ tunneling diodes. NDR is clearly seen from the J-V curves of these three devices (Fig. 2.14). The peak tunneling current density in forward bias ($J_{peak,FB}$) increases with the boron concentration from 1.7 to 8 kA/cm^2, due to the reduced tunneling barrier. The peak voltage is shifted to a larger value because of the series resistance. A best PVCR of 3.6 is achieved with a peak tunneling current density of 8.2 kA/cm^2, an indication of the high quality of the tunneling devices by in-situ doping CVD.

Next, to study the effect of Ge fraction, strained p^+-SiGe/n^+-SiGe homojunction tunneling diodes, with four different Ge fractions of 14, 21, 27, and 35%, were grown and fabricated. J-V curves are shown with NDR clearly seen in Fig. 2.15. $J_{peak,FB}$ increases from 0.03 A/cm^2 to 8.2 kA/cm^2 as the Ge fraction increases from 0.14 to 0.35 because of the smaller bandgap energy. While the increasing J_{peak} at forward bias is evidence of the effect of Ge fraction on tunneling, the effect of different p-type doping levels must be isolated to accurately extract the dependence of tunneling on Ge fraction.

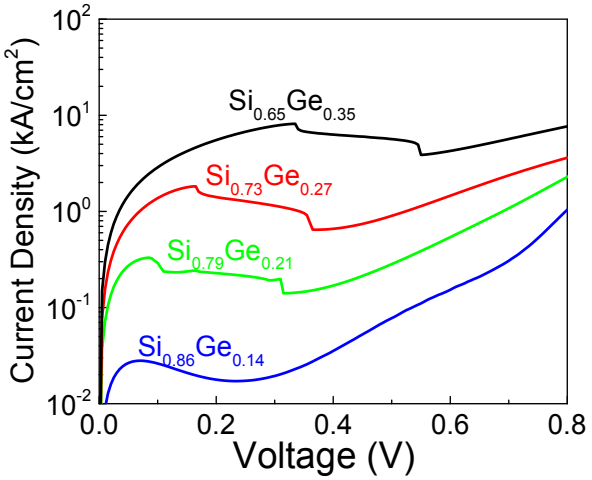

Fig. 2.15 J-V curves of p^+-$Si_{1-x}Ge_x$/n^+-$Si_{1-x}Ge_x$ homojunction diodes with four different Ge fractions of 14, 21, 27, and 35%.

To understand how the forward-biased BTBT current is affected by SiGe bandgap energy, we used a empirical model of bandgaps by Robbins *et al*. based on their photoluminescence measurements at 4 K [34], with a subtraction of 50 meV for the difference between their measurements at 4 K and ours at room temperature [35], to convert the measured Ge fractions into bandgap energies:

$$E_g(300K) = 1.17 - 0.896x + 0.396x^2 - 0.05 \quad \text{(eV)}, \tag{2.10}$$

where x is the germanium fraction in $Si_{1-x}Ge_x$ alloys. We then used a correction procedure to separate out the effect of doping profiles (shown schematically in Fig. 2.16), which were not the same for all Ge fractions. (E.g. N_A varied from 1.2 to 3.6 $\times$ 10^{20} cm^{-3} and N_D varied from 0.7 to 2.5 $\times$ 10^{20} cm^{-3}). In summary, we used Eqs. (2.4), (2.7), and (2.8) to predict how the peak tunneling current density scales with the electric field, and then adjusted the data points to reflect a single dopant profile at different Ge fractions. First we calculate the peak voltage (V_{peak}) and its peak electric field (F_{field}), by

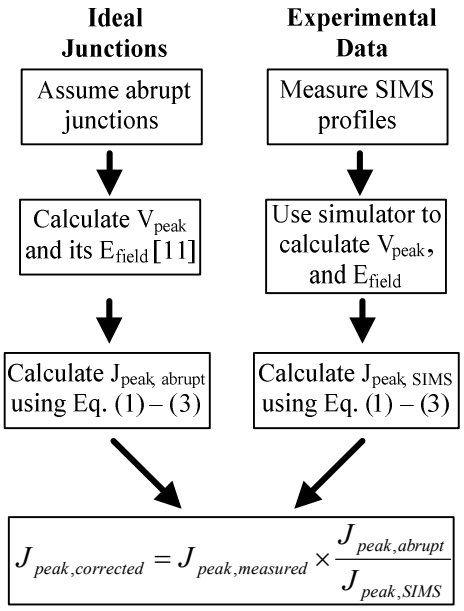

Fig. 2.16 Correction procedure of peak tunneling current density for each Ge fraction to a single set of fixed doping levels.

assuming perfectly abrupt doping profiles and no series resistance, with a single set of fixed doping levels of $N_A = 1.2 \times 10^{20}$ cm^{-3} and $N_D = 0.7 \times 10^{20}$ cm^{-3} for each Ge fraction. We calculate the peak current density ($J_{peak,abrupt}$) using Eqs. (2.4), (2.7), and (2.8). Next, we use the actual doping profiles measured by SIMS to calculate V_{peak} and F_{field} using a device simulator. We then calculate the peak current density ($J_{peak,SIMS}$) at that field. The ratio of these two current densities gives a correction factor to be applied to the experimental data for the adjustment of doping levels for all Ge fractions, so a comparison of data with a single set of doping profiles ($N_A = 1.2 \times 10^{20}$ cm^{-3} and $N_D = 0.7 \times 10^{20}$ cm^{-3}) can be made. After measuring the I-V curves of the devices to obtain $J_{peak,measured}$, we can calculate the corrected peak current density ($J_{peak,corrected}$) using the relationship of

$$J_{peak,corrected} = J_{peak,measured} \cdot \times \frac{J_{peak,abrupt}}{J_{peak,SIMS}}. \qquad (2.11)$$

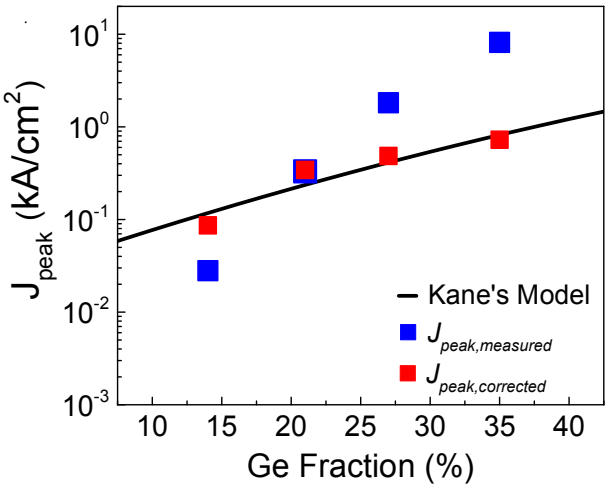

Fig. 2.17 J_{peak} vs. Ge fraction. Blue squares are the experimental data, red squares represent corrected data to a single set of fixed doping levels (N_A = 1.2 $\times 10^{20}$ cm^{-3} and N_D = 0.7 $\times$ 10^{20} cm^{-3}), and the solid line is the model prediction based on Eqs. (2.4), (2.7), and (2.8).

Fig. 2.17 shows the experimental ($J_{peak,measured}$) and corrected peak current density ($J_{peak,corrected}$) as a function of Ge fraction (blue and red squares), and also a theoretical prediction for a single set of doping levels assuming abrupt profiles (solid line). For a Ge fraction of 0.14, $J_{peak,corrected}$ is larger than $J_{peak,measured}$ because the actual doping profiles were not abrupt, resulting in a smaller electric field. For higher Ge fractions however, $J_{peak,measured}$ was larger than $J_{peak,corrected}$ because of the higher doping levels for those devices. No adjustable parameters were used in the correction process. Significantly, there is a good agreement between the slope of the theoretical calculation of peak tunneling current and data corrected to a single set of doping levels versus bandgap. This confirms that Eqs. (2.4), (2.7), and (2.8) can be used to predict the band-to-band tunneling of p^+-SiGe/n^+-SiGe homojunctions at current density up to ~ 10 kA/cm^2 in forward bias.

2.4.3 Reverse-Biased Band-to-Band Tunneling

2.4.3.1 Effects of Defect-Assisted Tunneling

The operation of TFETs relies on BTBT under reverse bias, also known as Zener tunneling [36]. In reverse bias, there is no simple clear feature such as NDR in forward bias which can be used to confirm that the observed current is due to BTBT. For example, defect states in the bandgap at the junction can lead to defect-assisted tunneling (DAT) [26], in which an electron first tunnels from the valence band of the p-type region to a defect state in the bandgap of the p-n junction, and then tunnels from the defect state to the conduction band of the n-type region (Fig. 2.18). Because each step of this process has a much lower tunneling barrier than direct BTBT, the two-step DAT process can easily swamp the direct BTBT. To investigate the effects of the DAT process, we examined tunneling in both forward and reverse biases in p^+-SiGe/n^+-Si heterojunction tunneling diodes. The preparation steps of the heterojunction tunneling diodes were introduced in section 2.3.1.

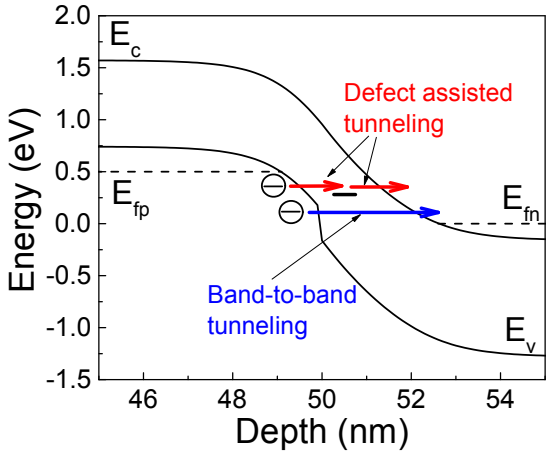

Fig. 2.18 Schematic of a band energy diagram for a p^+-$Si_{0.73}Ge_{0.27}$/n^+-Si heterojunction to show the processes of band-to-band tunneling and defect-assisted tunneling in reverse bias.

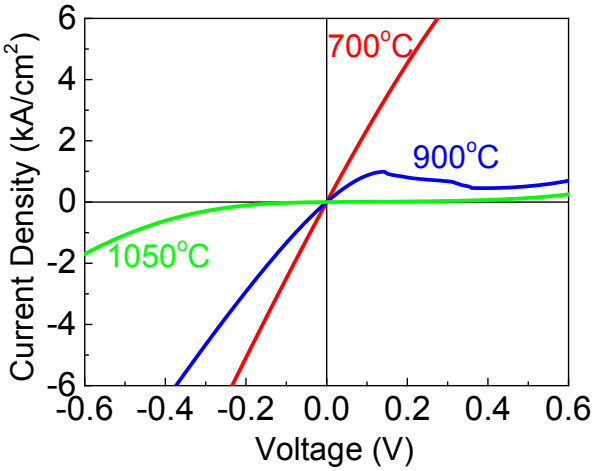

Fig. 2.19 J-V curves of p^+-$Si_{0.73}Ge_{0.27}$/n^+-Si heterojunction tunneling diodes with three annealing temperatures, showing the importance of the presence of NDR in forward bias on the BTBT current in reverse-bias.

For a relatively low annealing temperature (700 °C), a fairly high current with ohmic characteristics was observed at both forward and reverse biases, with no evidence of NDR (Fig. 2.19). With 900°C annealing, a lower current with NDR at forward bias was observed, with a lower current in reverse bias as well. We hypothesize that for the device with 700 °C annealing, the current in forward and reverse biases was dominated by the DAT process due to the incomplete annealing of the implanted damages, which swamped the true BTBT current. The defect density (and thus the DAT process) was reduced by 900 °C annealing, so that NDR at forward bias and the true BTBT current density could be observed. In the sample annealed at 1050°C, NDR disappeared with a much lower current density at both forward and reverse biases, because the diffusion of dopants at 1050°C reduces the junction abruptness and the electric field. The main message is that if NDR was not observed (e.g. at 700°C), the observed current in both forward and reverse biases has a large DAT component, and cannot be used as a true

measure of BTBT current. Consequently, we strongly suggest that to use reverse-biased tunneling data for the calibration of a BTBT model, a demonstration of NDR at forward bias is necessary to exclude the possibility of a dominant contribution of a defect-assisted tunneling current.

2.4.3.2 Effects of Series Resistance

Zener tunneling (reverse-biased BTBT) in silicon-based devices was in the past usually characterized at moderate doping levels (~ 10^{18} cm^{-3}). At such doping levels, the current density is very low (~ 10^{-4} kA/cm^{-2} [36, 37]), so the effect of series resistance could be ignored. However, for SiGe TFETs, operation at electric fields > 10^{7} V/m is desired, leading to a high current level. Thus, the series resistance effect such as current crowding must be considered for a precise calibration of BTBT in SiGe tunneling diodes. Guo *et al*. suggested that by scaling down the mesa width (W) of the diodes, current

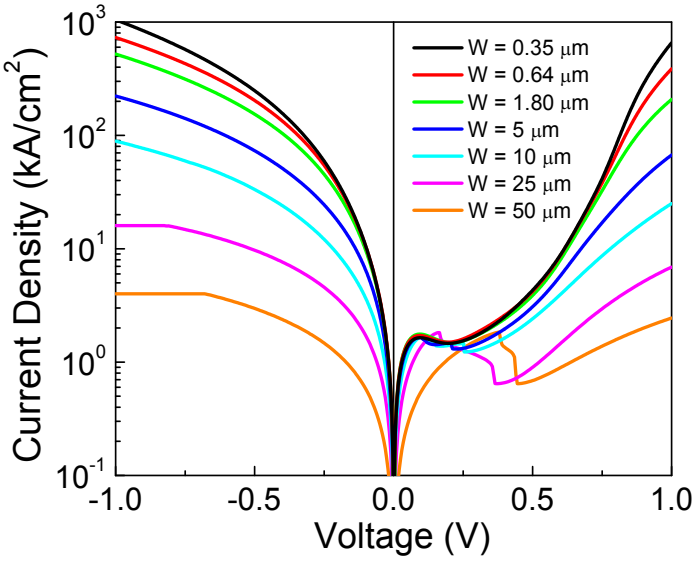

Fig. 2.20 J-V curves of $Si_{0.73}Ge_{0.27}$ homojunction tunneling diodes with different mesa widths.

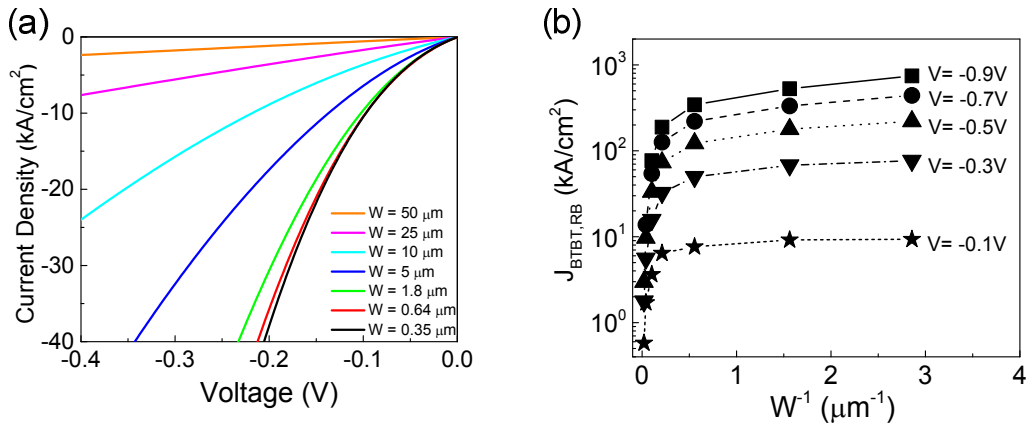

Fig. 2.21 (a) An enlarged view of J-V curves in Fig. 2.20 at small reverse biases, and (b) reverse-biased current density vs. W^{-1} at V = -0.1, -0.3, -0.5, -0.7, and -0.9 V, respectively.

crowding can be eliminated [38]. Therefore, we fabricated several $Si_{0.73}Ge_{0.27}$ homojunction p^+/n^+ diodes with mesa widths ranging from 50 μm to 0.35 μm using a combination of photolithography and electron-beam lithography. The growth parameters and layer structures are listed in Table 2.3. The current density at forward bias is shown in Fig. 2.20, with clear NDR for each device. The average of $J_{peak,FB}$ is 1.73 kA/cm^2 and the deviations are within 5% for all mesa sizes, confirming a negligible contribution of the leakage current via the mesa edges [39]. In reverse bias, the current density approaches a constant level as the mesa width decreases (Fig. 2.21(a) and (b)). Because (i) NDR is clearly seen at forward bias and (ii) the effect of series resistance was eliminated by scaling down the mesa width, we believe the plateaus of the current density Fig. 2.21(b) represent the true BTBT current density in reverse bias.

2.4.3.3 Comparison of Experimental Results and Models

We now seek to present BTBT in reverse bias as a function of the Ge fraction and electric field for device designers to model TFETs and related devices. Similar to

the peak current density at forward bias, the reverse-biased BTBT current ($J_{BTBT,RB}$) also strongly depends on the electric field and bandgap energy. An analytical form of $J_{BTBT,RB}$ in reverse bias based on Fair's model [36] is

$$J_{BTBT,RB} = \frac{\sqrt{2m^*}e^3 E_{field} V}{4\pi^3 \hbar^2 \sqrt{E_g}} \cdot \exp\left(\frac{-4\sqrt{2m^* E_g^3}}{3e\hbar F_{field}}\right), \tag{2.12}$$

where V is the applied voltage. The tunneling probability of electrons depends on the tunneling distance. In a semiconductor p-n junction, the tunneling distance is inversely proportional to the peak electric field [24], so $\log(J_{BTBT,RB})$ in Eq. (2.12) depends on the inverse of the electric field. The parameter of $J_{BTBT,RB}/V$ is plotted versus the electric field (F_{field}) for different Ge fractions in Fig. 2.22, along with theoretical calculations based on Eq. (2.12). For our devices, the electric field was varied by adjusting the reverse bias from 0.1 to 0.9 V. As in forward bias, the measured SIMS doping profiles

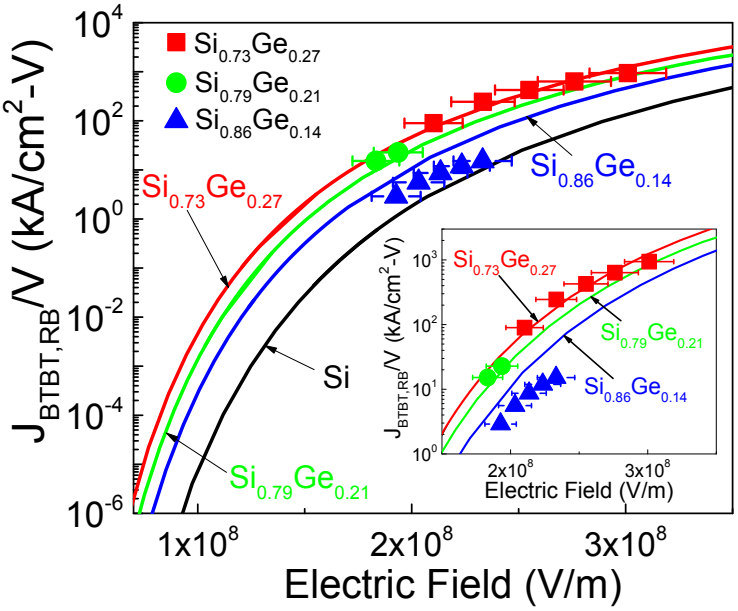

Fig. 2.22 Reverse-biased BTBT current density per volt ($J_{BTBT,RB}/V$) vs. electric field in SiGe BTBT. Symbols are the experimental data, and multiple lines represent the model predictions based on Eqs. (2.10) and (2.12). An inset provides an enlarged view for high electric fields of 2 to 3 $\times$ 10^8 V/m.

and the device simulator were used to calculate the electric field for all devices. The horizontal error bars in the electric field, which resulted from an estimated uncertainty of ±15 % in the doping levels, are also presented in Fig. 2.22.

At small electric fields (< 1×10^8 V/m), Eq. (2.12) predicts that $J_{BTBT,RB}/V$ rises sharply with F_{field}. On the other hand, at large electric fields, $J_{BTBT,RB}/V$ increases much more slowly. For example, at low fields ($F_{field} = 5 \times 10^7$ V/m), $J_{BTBT,RB}/V$ of $Si_{0.73}Ge_{0.27}$ (red line in Fig. 2.22) increases by a factor of 2×10^5 as F_{field} increases by 60%. However, at $F_{field} = 2 \times 10^8$ V/m, it only increases by a factor of 30. Our data of reverse-biased tunneling conductance density (current density/voltage) at electric fields of > 2×10^8 V/m, which are between 3 – 1000 kA/cm^2-V, are in close agreement with the model predictions (the inset in Fig. 2.22). This is fortuitous as our data points of $J_{BTBT,RB}/V$ are three to five orders of magnitude higher than those for Si BTBT at low electric fields (< 1×10^8 V/m) [36, 37], where Eq. (2.12) has been applied.

To isolate the effect of bandgap energy, $J_{BTBT,RB}/V$ vs. Ge fraction was plotted at $F_{field} = 2 \times 10^8$ in Fig. 2.23 by interpolating between points in Fig. 2.22, along with the model of Eqs. (2.10) and (2.12). The horizontal error bars of ± 6 % variations in the peak electric field of Fig. 2.22 were used to estimate the resulting errors in the BTBT current. Within the uncertainty in current introduced by the variation in electric field, the results show that Eqs. (2.10) and (2.12) can be used to model the dependence of $J_{BTBT,RB}$ on Ge fraction. More complete modeling might include the effect of heavy doping on bandgap energy, and the effects of strains on the effective density of states and the effective mass of electrons, which are not considered in this work.

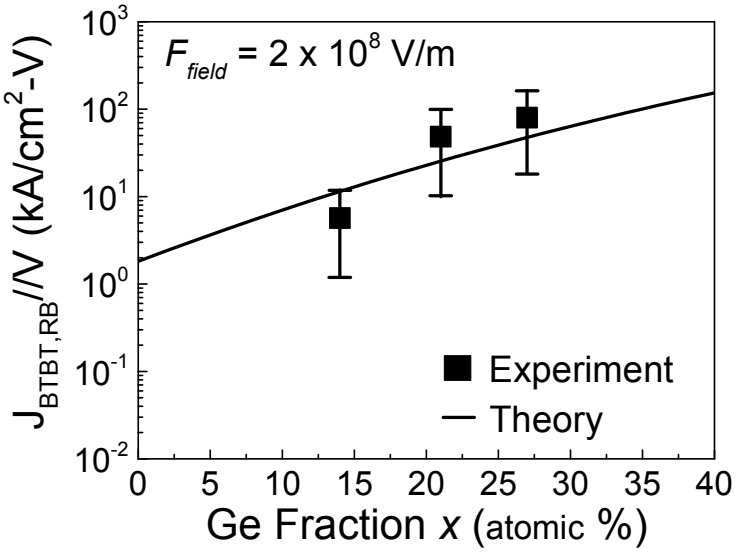

Fig. 2.23 $J_{BTBT,RB}$ vs. Ge fraction at F_{field} of 2 × 10^8 V/m. Points are the experimental data with error bars, representing the effects of the variations in electric field (the horizontal error bars in Fig. 2.22).

2.5 Summary

SiGe-based TFETs were proposed as replacements for Si TFETs because higher tunneling current is expected for the smaller bandgap. We studied two types of SiGe tunneling junction: p^+-SiGe/n^+-Si heterojunctions and p^+-SiGe/n^+-SiGe homojunctions, to establish the relationship between the experiment results and models. To observe NDR in implanted p^+-SiGe/n^+-Si heterojunction tunneling diodes, annealing at high temperature is required to reduce defect-assisted-tunneling (DAT) via the junction defects introduced during the implant process for n^+-Si layer. For low-temperature annealing, DAT swamps band-to-band tunneling, so no NDR was observed. For high-temperature annealing, there was also no NDR observed because of the reduced electric field resulting from strong dopant diffusion. For a Ge fraction of

0.35, a peak tunneling current density of 8.2 kA/cm^2 in a homojunction tunneling diode was reported, which is the highest for all Si-based tunneling diodes grown by CVD. For p^+-SiGe/n^+-SiGe homojunction tunneling diodes by in-situ doping CVD, defect-assisted tunneling can be greatly reduced with an improved peak-to-valley current ratio of 3.6.

For both types of tunneling junctions, the effects of the electric field and Ge fraction on BTBT at forward bias and bias bias were investigated. Increasing the doping levels and Ge fraction enhances tunneling. We compared our experimental results in forward bias with Kane's model [25] and the reverse-biased experimental data with Fair's model [36], both in good agreement. This suggested the adequacy of Kane's model for predicting BTBT in strained SiGe junctions, at least up to a current density level of 10^4 A/cm^2 and 10^6 A/cm^2 in forward and reverse biases, respectively.

We qualitatively demonstrated the importance of the presence of NDR in reverse biased BTBT. With NDR observed in each device and the elimination of series resistance by scaling down the device, the true measure of reverse-biased BTBT in SiGe p^+/n^+ homojunction was identified for the first time.

Chapter 3 Surface Segregation of Phosphorus in Relaxed $Si_{0.7}Ge_{0.3}$ Layers Epitaxially Grown by Chemical Vapor Deposition

3.1 Introduction

As device dimensions are scaled down, a sharp profile of dopants is becoming a key factor to realize nano-scale semiconductor devices such as tunneling diodes [40], tunneling field-effect transistors [20], and two-dimensional electron gases (2DEGs) in a modulation-doped Si/SiGe heterostructure [41]. 2DEGs are of particular interest for quantum dot (QD) applications. A QD is usually fabricated on a 2DEG, with top metal Schottky depletion gates used to isolate a single electron in the underlying 2DEG layer. However, the strong surface segregation of n-type dopants in a relaxed SiGe epitaxial film can cause a high dopant concentration at the surface, resulting in high gate leakage current and ineffective gating. Therefore, a sharp turn-off slope of n-type dopants is necessary.

Although a turn-off slope of 2-3 nm/dec for antimony was reported in Si epitaxial films grown by molecular beam epitaxy (MBE) [42], it has been difficult to obtain such abrupt profiles for phosphorus and arsenic, the most common n-type dopants in chemical vapor deposition (CVD) systems [43]. Several works were reported to reduce phosphorus segregation in Si by ex-situ cleaning (13 nm/dec) [44] or by introducing substitutional carbon atoms into Si epitaxial films (11 nm/dec) [45].

However, the former approach requires a growth interruption, which may introduce contaminants into the growth interface. For the latter, the control over carbon atoms into substitutional sites is critical since the interstitial carbon could degrade device performance due to their midgap energy states [46].

In this chapter, we report an extremely sharp phosphorus turn-off slope of 6 nm/dec in relaxed $Si_{0.7}Ge_{0.3}$ films without any ex-situ cleaning step or introduction of carbon into the epitaxial films. We found that the hydrogen coverage on the surface during the growth plays an important role in the suppression of phosphorus segregation in the CVD process at low temperatures (500 ~ 600 ^{o}C). Finally, a phenomenological model is proposed to explain the effect of surface hydrogen on phosphorus segregation.

3.2 Two State Model

In our work, a matrix of Si and Ge atoms and surface hydrogen complicate the analysis of phosphorus segregation. In a simpler case of phosphorus in Si (100), a segregation energy of 0.64 eV, which was defined as the energy difference of phosphorus in the surface and bulk layers, was firstly reported by Nützel *et al*. [47] from SIMS results. This work used a so-called two-state model (TSM) [48] of atoms moving between the surface and sub-surface layers in Si. Later, by temperature programmed desorption (TPD), Cho *et al*. [49] reported a segregation enthalpy of 0.86 eV, which is essentially the same as the segregation energy defined in the TSM. The P coverage in Nützel's work [47] and Cho's work [49] was larger than 0.1 monolayer (ML), and the major Si surface structures with P surface coverage > 0.1 ML was previously by Yu *et al*. [50] as a mixture of Si–Si, Si–P, and P–P dimers. Sen *et al*. [51] used density functional

theory to predict the different favorable sites for surface phosphorus atoms at coverage above and below 0.13 ML. Thus, the work in [50, 51] might not be directly relevant to our experimental results since the integrated phosphorus doses in our samples are at most 5×10^{12} cm^{-2} (~ 0.01 ML). Other works on Sb [52], As [53], and Ge [54, 55] surface segregation in Si have also been modeled by using a TSM. In addition, those works all ignored any temperature dependence of attempt frequency and used the segregation energy to reflect all temperature effects. Thus, in this chapter, we also use a modified TSM to investigate phosphorus segregation in a more complicated structure of relaxed $Si_{0.7}Ge_{0.3}$ layers.

A two-state model (TSM) describes the dopant segregation as an exchange process of P atoms and host atoms (Si or Ge in this study) between the surface layer and the sub-surface layer (Fig. 3.1). The rate equations governing this exchange process between those two layers are

$$\frac{dn_1}{dt} = -r_{12}n_1(1-n_2) + r_{21}n_2(1-n_1), \tag{3.1}$$

$$\frac{dn_2}{dt} = -r_{21}n_2(1-n_1) + r_{12}n_1(1-n_2), \tag{3.2}$$

$$r_{12} = \nu e^{-\frac{E_1}{kT}}, \tag{3.3}$$

$$r_{21} = \nu e^{-\left(\frac{E_1+\Delta E_{surf}}{kT}\right)}, \tag{3.4}$$

where n_1 and n_2 are the normalized concentrations of phosphorus in the sub-surface layer (layer 1) and the surface layer (layer 2), r_{12} and r_{21} are the jumping rates of phosphorus from the sub-surface layer to the surface layer, and vice versa. E_1 is the

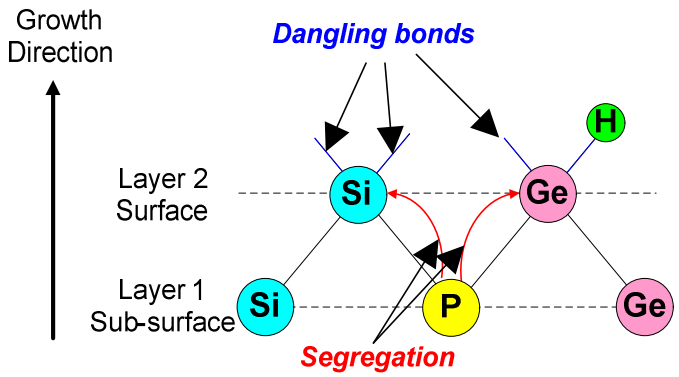

Fig. 3.1 Schematic of atomic layer structures near the surface during the epitaxial growth.

activation energy barrier facing phosphorus in the sub-surface layer, ΔE_{surf} is the segregation energy, which represents the difference of activation barriers of layer 1 and layer 2, and v is the attempt frequency (Fig. 3.2). We assume a single attempt frequency independent of temperature as previous reports suggested for phosphorus segregation in Si [47, 48, 49, 53, 54, 55]. It is assumed that P atoms below the sub-surface layer are trapped and cannot diffuse during the time of the growth. Assuming n_1 and $n_2 << 1$, the differential rate equations can be solved and an analytical form of phosphorus turn-off slope x_0 (nm/decade) is given by [56]

$$x_0 = \frac{a_0 \ln 10}{4} \frac{1}{\ln\left(1+e^{\frac{-\Delta E_{surf}}{kT}}\right) - \ln\left[1-e^{-[r_{12}+r_{21}]\frac{a_0}{4GR}}\right]}, \qquad (3.5)$$

where a_0 is lattice constant and GR is the growth rate.

According to the TSM, the surface segregation occurs because P atoms at the sub-surface layer tend to stay in the surface layer due to the lower energy level in the surface layer (Fig. 3.2). At thermal equilibrium (low growth rates), the surface segregation is determined by the ratio of the phosphorus concentrations at the surface

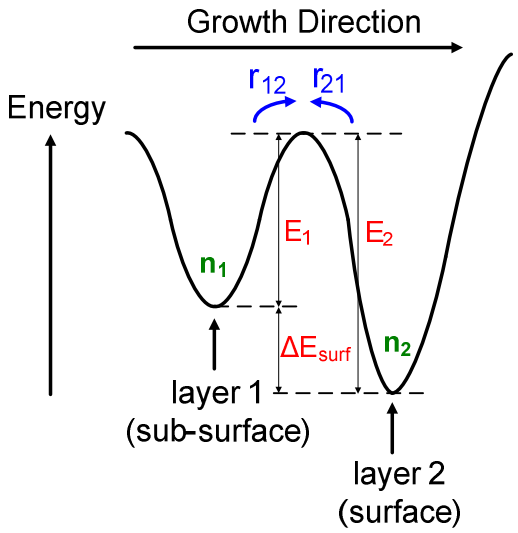

Fig. 3.2 Schematic of phosphorus energy near the surface during the epitaxial growth in a two-state model. Layer 2 represents the surface layer and layer 1 (sub-surface layer) represents the next layer below the surface.

and sub-surface layers (n_2/n_1), which only relies on the temperature and the segregation energy (ΔE_{surf}) through the first term in the denominator of Eq. (3.5). At lower temperatures, if still in equilibrium, more P atoms are trapped in the surface layer than in the sub-surface layer due to the lower energy state of the former, so the segregation is stronger. On the other hand, in the kinetic-limited regime of high growth rates, P atoms in the sub-surface layer cannot reach the equilibrium with those in the surface layer. Therefore, phosphorus will be trapped in the sub-surface layer and the limiting factor is its activation energy barrier E_1. As the temperature is reduced, the probability of phosphorus jumping across the barrier to the surface layer is smaller because of the lower kinetic energy of phosphorus. As a result, phosphorus segregation is reduced. This physical limit has been applied to reduce phosphorus segregation in Si (100) grown by MBE at temperature below 500 ^{o}C [47], and the best turn-off slope (4 nm/dec) for P in

Si grown by a combination of scanning tunneling microscopy and MBE was reported by room-temperature growth [57].

3.3 Failure of Two-State Model

3.3.1 Experimental Results

In this work, Si (100) substrates (for strained SiGe) and polished relaxed $Si_{0.7}Ge_{0.3}$ virtual substrates with a graded $Si_{1-x}Ge_x$ ($0 < x < 0.3$) buffer layer grown on Si (100) substrates (for relaxed SiGe) were used to study phosphorus segregation. Prior to being placed into the reactor, substrates were cleaned by the following steps: 5 min in diluted HF (1%), 15 min in H_2SO_4: H_2O_2 (2.5:1), followed by 2 min in diluted HF (1%). Then the samples were heated to 850 ^oC in hydrogen gas at 6 torr to remove the residual oxide before the epitaxial growth starts. The gas precursors were diluted silane (10 % in argon) and GeH_4 (0.8 % in hydrogen) for Si and SiGe growth, and a diluted phosphine (100 ppm in hydrogen) was the doping gas. The test structure for phosphorus segregation is as follows: first, a 20-nm undoped $Si_{0.7}Ge_{0.3}$ buffer layer was grown followed by a 10-nm n-type $Si_{0.7}Ge_{0.3}$ layer doped with phosphorus of peak level between 10^{18} and 10^{19} cm^{-3}. Both layers were grown at 575 ^oC and the growth rate was 5 nm/min. Then an undoped $Si_{0.7}Ge_{0.3}$ cap layer was grown at 500 ^oC ~ 600 ^oC to study the effect of growth temperature on phosphorus segregation, with its thickness between 30 to 150 nm. To investigate the effect of growth rate, we varied the growth rate of the $Si_{0.7}Ge_{0.3}$ cap layers between 0.1 to 30 nm/min by adjusting the partial pressures of silane and germane. The Ge fraction in the SiGe cap layer was between 0.28 to 0.30.

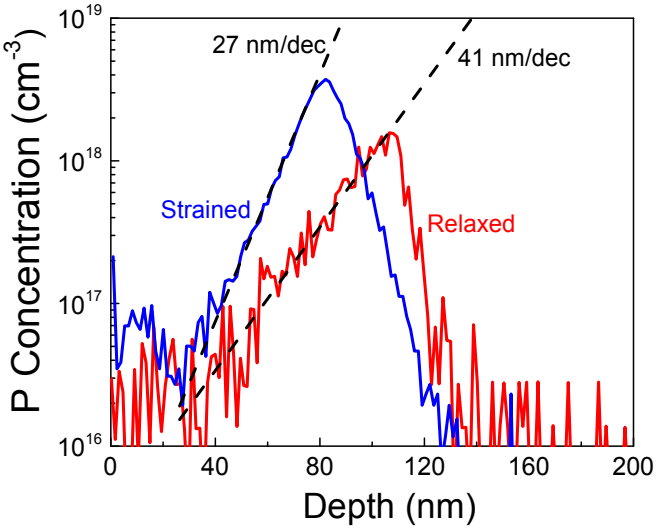

Fig. 3.3 Phosphorus profiles in strained and relaxed $Si_{0.7}Ge_{0.3}$ layers grown at 575 ^{o}C under the same growth conditions [56]. The growth rates of strained and relaxed $Si_{0.7}Ge_{0.3}$ layers are 8 nm/min and 6 nm/min, respectively.

Last, a thin Si cap layer of 4 nm was grown at 625 ^{o}C, with the growth rate of 2.5 nm/min. The films were subsequently characterized by SIMS to determine the phosphorus profiles and the growth rates.

Most prior works of phosphorus segregation in SiGe were done in compressively strained SiGe layers [53], not in the relaxed SiGe layers required for a modulation-doped 2DEG, i.e. a higher conduction band edge in the SiGe layer than in the Si layer. Thus, we compared the phosphorus profiles in strained and relaxed SiGe films first. We found that the segregation is much worse in relaxed SiGe films than in strained SiGe films (Fig. 3.3). For a growth temperature of 575 ^{o}C and growth rate of 5 nm/min, the turn-off slopes of phosphorus in strained and relaxed films are 27 and 41 nm/decade, respectively [56]. The fundamental reasons for this difference are unknown and a further study is required. We then focused on phosphorus segregation in the

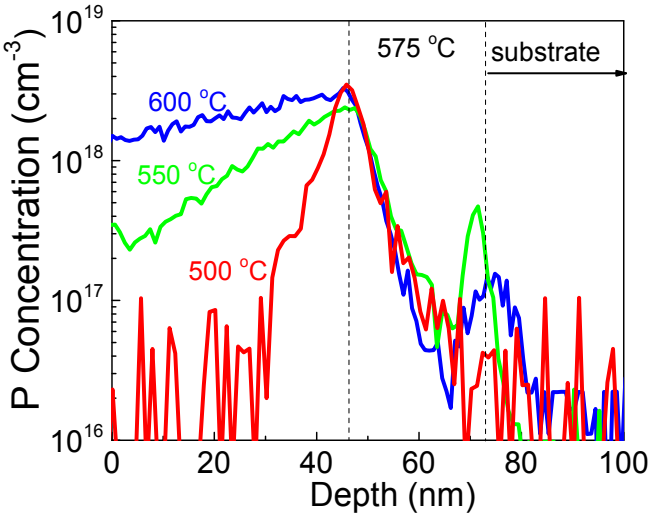

Fig. 3.4 Phosphorus profiles in relaxed $Si_{0.7}Ge_{0.3}$ layers grown at 500, 550, and 600 °C. Phosphorus supply was turned off at the depth of 45 nm. P turn-off slopes were 127, 40, and 9 nm/decade for 600, 550, and 500 °C, respectively [56].

relaxed $Si_{0.7}Ge_{0.3}$ films.

Phosphorus profiles measured by SIMS for relaxed SiGe layers grown at 500, 550, and 600 °C are shown in Fig. 3.4. At a depth of 45 nm, the phosphine supply for the doped layer was turned off and growth was continued without interruption. As the growth temperature is reduced, the surface segregation is reduced, with the phosphorus turn-off slope declining from 127 nm/dec at 600 °C to 9 nm/dec at 500 °C. By adjusting the gas flow rates of silane and germane at 500 °C, an extremely sharp slope of 6 nm/min was obtained with the growth rate of 0.08 nm/min, which we believe is the sharpest reported turn-off slope of phosphorus in relaxed SiGe films.

Our experimental data (points) of phosphorus turn-off slope vs. growth rate at different temperatures (500 °C to 600 °C) and theoretical curves based on the TSM are shown in Fig. 3.5. At 600 °C the segregation is near the transition between thermal

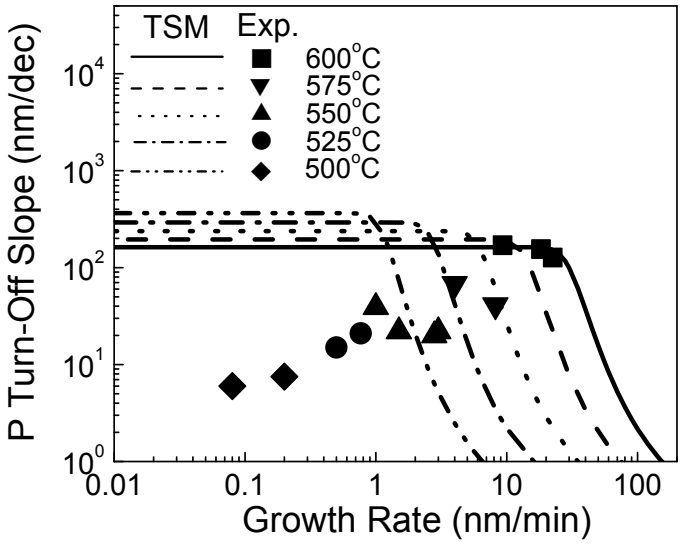

Fig. 3.5 Phosphorus turn-off slope vs. growth rate for different temperatures. Experimental results (points) and the theoretical prediction (lines) are presented for comparison [56].

equilibrium and kinetic-limited regime. Thus, $\Delta E_{surf} = 0.47$ eV can be fitted by assuming the experimental results were in the regime of thermal equilibrium. Furthermore, as Nützel *et al*. [58] suggested that the attempt frequency would be between 10^{11} to 10^{13} Hz, we selected $\nu = 1 \times 10^{12}$ Hz to fit the data at 600 °C and found $E_1 = 1.84$ eV. Despite a good match between the experimental data and the theoretical curve at 600 °C, there is a large discrepancy between them at 500 °C to 575 °C. The low dependence of the segregation on growth rate suggests the data at low temperatures were in the equilibrium regime, not kinetically limited. However, the phosphorus slopes are much sharper at lower temperatures, in contrast with what would be expected from the TSM in equilibrium. This discrepancy cannot be resolved by simply adjusting the fitting parameters.

3.3.2 Effects of Hydrogen on Phosphorus Segregation

In the CVD process, hydrogen is used as a carrier gas and it is well established that hydrogen could cover the surface by forming Si–H [59] or Ge–H bonds [60]. At high temperatures, most of those bonds break easily and hydrogen desorbs, so the surface coverage of hydrogen is nearly zero. At low temperatures, however, the thermal energy is too low to break the Si–H or Ge–H bonds efficiently, so hydrogen will cover most of the surface layer. Our data show that the phosphorus turn-off slope is nearly constant with growth rate at a fixed temperature, suggesting that it is in the regime of thermal equilibrium, with the segregation then depending on ΔE_{surf}.

At the heart of our model, we assume phenomenologically that the presence of surface hydrogen changes the relative energy of P atoms in the surface and sub-surface layers such that the segregation energy ΔE_{surf} is reduced. Thus, at lower temperatures, the segregation will be suppressed due to higher hydrogen coverage on the surface. Because of two types of surface sites (with or without H), in principle we could model the problem with two segregation energies and a fraction of phosphorus segregation to each site. However, the two energies would probably depend on the local numbers of Si or Ge atoms, leading to too many parameters. Prior works by MBE and TPD used an effective segregation energy [54, 55] to investigated the effect of hydrogen on Ge segregation into Si (100), which we follow in this study. Thus, we treat ΔE_{surf} as a single effective parameter which varies with hydrogen coverage as the temperature changes.

By the introduction of the ideas of an effect of hydrogen on phosphorus segregation, a fit between the experimental data and the model is obtained by using ΔE_{surf} as a fitting parameter at different temperatures or hydrogen pressures (Fig. 3.6),

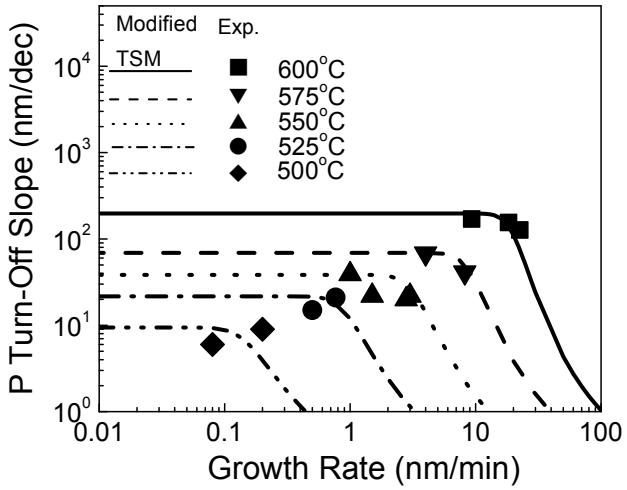

Fig. 3.6 Comparison of experimental data and modified TSM of phosphorus turn-off slope vs. growth rate by including the effect of surface hydrogen on the segregation energy (E_{surf}) [56]. The hydrogen pressure was 6 torr.

with ΔE_{surf} plotted vs. temperature in Fig. 3.7. At hydrogen pressure of 6 torr, the reduced segregation at lower temperatures originates from a smaller segregation energy ΔE_{surf}. The decrease of ΔE_{surf} with decreasing temperature will reduce the ratio of phosphorus populations in the surface vs sub-surface layers (n_2/n_1), resulting in a reduction of the segregation. On the other hand, ΔE_{surf} increases with temperature because of less hydrogen coverage at higher temperatures. The reduced segregation energy of P in $Si_{0.7}Ge_{0.3}$ with more surface hydrogen coverage follows the trend of Ge segregation in Si [54, 55, 61].

To confirm that the main effect of segregation reduction comes from a change of hydrogen coverage rather than a change of attempt frequency or other effects with temperature, we adjusted hydrogen pressure to vary its surface coverage [62] at 575 ^{o}C (Fig. 3.7). As expected, with higher hydrogen pressure (23 torr), the segregation is

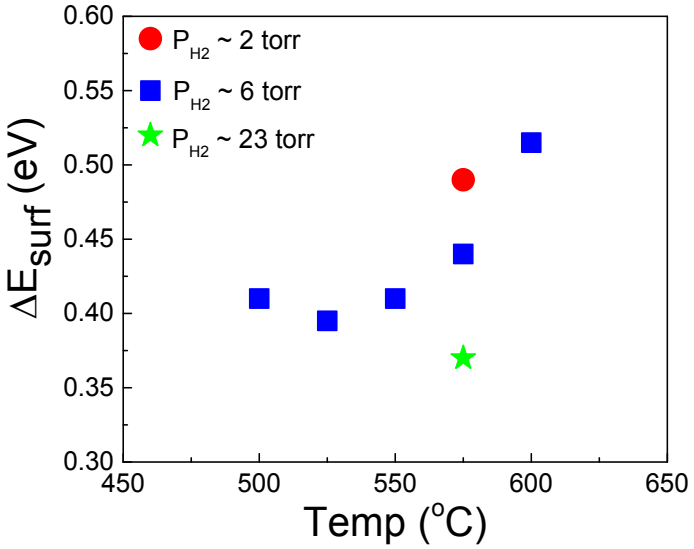

Fig. 3.7 Segregation energy (E_{surf}) vs. growth temperature. At 575 °C, E_{surf} for hydrogen pressure of 2, 6, 23 torr are also plotted to confirm the effect of surface hydrogen on P surface segregation [56].

suppressed, leading to a lower effective segregation energy. The opposite trend is shown with a lower hydrogen pressure of 2 torr. The effective segregation energy obtained in this work ranges between 0.37 ~ 0.52 eV, which is lower than those published for P in Si of 0.67 ~ 0.86 eV [47, 49] without hydrogen coverage. The difference could be explained by the presence of surface hydrogen on $Si_{0.7}Ge_{0.3}$ surface in this work, and the SiGe matrix in our work vs. Si matrix in other work.

3.4 Summary

We studied the surface segregation of phosphorus in $Si_{0.7}Ge_{0.3}$ films grown by CVD. The segregation in relaxed layers is stronger than in strained layers. Furthermore, the segregation in relaxed SiGe layers is reduced as the temperature is decreased due to

the higher surface coverage of hydrogen. We proposed a phenomenological model to explain the effect of surface hydrogen on phosphorus segregation, where surface hydrogen changed the bonding structure of host atoms near the surface and reduced the segregation energy ΔE_{surf} as the growth temperature is decreased. Thus, the segregation is suppressed at lower temperatures. An extremely sharp phosphorus turn-off slope of 6 nm/dec in relaxed $Si_{0.7}Ge_{0.3}$ layers grown at 500 ^{o}C was also reported, enabling effective Schottky gating on a modulation-doped Si two-dimensional electron gas (chapter 4).