Base Doping Effects and Design of Si/SiGe/Si Heterojunction Bipolar Transistors

Željka Matutinović Krstelj

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

November 1994

© Copyright 1994 by Željka Matutinović Krstelj.

All rights reserved.

Mojim dragim roditeljima, Katarini i Željku

To my dear parents, Katarina and Želimir

Abstract

Recent advances in high speed performance of Si/Si$_{1-x}$Ge$_x$/Si heterojunction bipolar transistors (HBT's) and the possibility of their integration into standard silicon bipolar technology have been the focus of attention among Si-based heterojunction devices. This thesis focuses on Si/Si$_{1-x}$Ge$_x$/Si HBT's, specifically issues related to process integration, the design of these devices and empirical DC modeling.

The devices in this work were grown by Rapid Thermal Chemical Vapor Deposition (RTCVD). The quality of epitaxial material and interfaces was studied in a wide pressure range by x-ray reflectivity (XRR), photoluminescence, electrical performance of p-type resonant tunneling diodes, and x-ray diffraction (XRD) of a superlattice. An upper limit to interface roughness of below 5Å is established by XRR and XRD.

Apart from high gain, low noise and high output resistance, Si/Si$_{1-x}$Ge$_x$/Si HBT's offer low intrinsic device delays (high f_T) due to germanium (bandgap) grading in the narrow epitaxial base and low parasitic base resistances due to heavy base doping, essential for high speed circuit performance. When integrated into Si technology, processing needs to be adjusted to reduced thermal cycles (below 800°C) to prevent strain relaxation and to minimize base dopant diffusion. A heavily doped base in a bipolar transistor can lead to a p^+–n^+ base-emitter junction. An upper limit to the doping on the lighter doped side of the junction of 5×10^{18} cm^{-3} is established before the onset of significant parasitic tunneling current. Hall and drift lateral hole mobilities are measured in a wide range of base p-type dopings and germanium concentrations. The first empirical model for effective bandgap narrowing for minority carrier transport in the p-Si$_{1-x}$Ge$_x$ base over a wide range of base dopings and Ge concentrations, extracted from the room temperature collector current measurements, is presented. The DC design trade-off between the base sheet resistance and gain is modeled. Minority carrier diffusion length is measured for the first time in p-type $Si_{1-x}Ge_x$ as a function of doping.

Finally, a new vertical transport device in the Si–based material system, a symmetric electron resonant tunneling diode, is demonstrated for the first time. The anomalous temperature behaviour of the lowest bias resonance is explained by a phonon-absorption-assisted model.

Acknowledgments

I wish to thank all the people whose help has contributed to this thesis.

First of all, Prof. Jim Sturm has been the best advisor that a graduate student could have, and I consider myself very lucky for having had a chance to be a part of his research group for the past five years. I appreciate his guidance, support and encouragement.

All the faculty members of the Electronic Materials and Devices group deserve to be mentioned, but special thanks should go to Prof. Sigurd Wagner and Prof. Stephen Forrest for taking their time to read this thesis.

I should also mention the faculty and my colleagues at the Department of Electronics of the School of Electrical Engineering in Zagreb, especially Prof. Petar Biljanović, who have supported me in the desire to pursue my graduate career at Princeton.

My colleagues in Prof. Sturm's group made our lab a pleasant place to work. Erwin Prinz, Peter Schwartz and Xiadong Xiao were great teachers for me who always had time to answer my stupid questions. Cheewee Liu maintained the RTCVD system. My dear friend Venki Venkataraman was always there when I needed him, solving "generals" questions on the beach in California, ice-skating on Carnegie Lake and doing Hall measurements to meet my deadline... It was also fun working with Peter Garone, Mi Qun, Anthony St.Amour, Louis Lanzerotti, Arvind Reddy, Chung-Chih Wu, Keith Jackson and Kern Rim. I will also remember all of our fellow graduate students in the Department of Electrical Engineering. Special thanks go to Steve Sherman who helped me in trying to do some last minute experiments and Hari Manoharan who developed the XRD simulation program that I used.

Dr. Eric Chason at Sandia National Laboratories contributed a great deal to this thesis by doing XRR measurements. Dr. H. C. Liu at NRC, Canada processed a run of my resonant tunneling diodes. I am also grateful for having had the opportunity to collaborate with T. Y. Chiu, James Sung and Mark Liu from AT&T. Dr. Charles Magee at Evans East did SIMS measurements on some of my samples. I would also like to thank TMA for providing MEDICI for my 2D device simulations and particularly my friend Ognjen Milić for helping me out with any problems.

This thesis would never have been accomplished without the support of my family and friends in Croatia. Being so far away from the hardships that have come across them, all I had to worry about was my work. Yet, my mind has been with them throughout all of these years. Although they have been going through very hard times, all I could hear were words of support and encouragement for myself. I could only wish that some day I'll be able to be such a role-model for my children as my parents have been for me.

Finally, my husband Igor has been going through all of my good and bad moods in the past five years and always found a way to make me laugh.

Contents

Abstract iv

Acknowledgments vi

1 Introduction 1

1.1 Motivation 1

1.2 Thesis outline 2

2 Integration of Si/SiGe/Si HBT's into Si-technology 4

2.1 Introduction 4

2.2 Si/SiGe/Si heterojunction bipolar transistors 4

2.3 Development of high speed Si/SiGe/Si HBT's, historic overview . . . 6

2.4 The Si/SiGe heterojunction 8

2.5 Speed of Si/SiGe heterojunction bipolar devices and circuits 12

2.5.1 Improvement of ECL gate delay via HBT structure - simulation 16

2.6 Process integration 22

2.6.1 Boron outdiffusion from SiGe base 24

3 RTCVD Growth and Interface Quality of Si/SiGe Heterojunctions 32

3.1 Introduction 32

3.2 Rapid Thermal Chemical Vapor Deposition (RTCVD) 33

3.3 Composition and growth rate at various growth pressures 36

3.4 Photoluminescence measurements 38

3.5 X-ray reflectivity measurements 38

3.6 Resonant tunneling diodes 45
3.7 Low pressure limit to interface roughness 50
3.8 Conclusions 52

4 Tunneling Current in $p^+ - n^+$ Junctions 55

4.1 The importance of tunneling currents in heavily doped p-n junctions for bipolar transistors 55
4.2 Experimental results 57
4.3 Discussion 64
4.4 Conclusions 68

5 Lateral Hole and Vertical Electron Currents in $Si_{1-x}Ge_x$ Bases 70

5.1 Introduction 70
5.2 Device fabrication 71
5.3 Majority carrier properties 73
5.4 Effective bandgap measurements and model 77
5.5 Collector current vs. base resistance model 92
5.6 Summary 93

6 Minority Carrier Properties in $Si_{1-x}Ge_x$ 96

6.1 Introduction 96
6.2 Methods to measure minority carrier diffusion length and mobility . . 97
6.3 Structure optimization for measurements of minority carrier parameters in strained SiGe alloys 99
6.4 Device fabrication 105
6.5 Diffusion length measurements 109
6.6 Summary 116

7 **Electron Resonant Tunneling in Si/$Si_{1-x}Ge_x$ Heterostructures** **117**

7.1 Introduction 117

7.2 Device fabrication 118

7.3 Results 120

7.4 Phonon-absorption-assisted model 125

7.5 Summary 131

8 **Conclusions and Suggestions for Further Research** **133**

8.1 Conclusions 133

8.2 Suggestions for future work 134

A **Growth and Processing Details** **136**

A.1 Graded base HBT samples 136

A.2 Samples for interface roughness study 137

A.3 P-type resonant tunneling diodes 138

A.4 Flat–base HBT samples 139

A.5 Lateral SiGe channel devices 139

A.6 N-type resonant tunneling diodes 140

A.7 Processing of resonant tunneling diodes 141

B **Publications and Presentations Resulting from this Thesis** **145**

References **147**

List of Tables

3.1 Summary of XRR samples and results: a) growth parameters, b) XRR results 54

3.2 Simulated and measured ratios of XRD intensities of superlattice peaks. Simulation with no mixing and including mixing at the interface was compared to data. The best fit for interface width is between 1 Å to 2 Å. 54

4.1 Summary of p^+–n^+diodes: a) doping, b) growth parameters 69

5.1 Summary of HBT samples and relevant parameters (G_B -SIMS is the raw data uncorrected for depletion region effects, G_B -Hall is calculated assuming $r_H = 1$) 94

6.1 Summary of structural parameters of samples designed for measurements of minority carrier properties 107

6.2 Summary of processing runs and implant conditions 110

7.1 Summary of RTD samples 132

A.1 HBT sample #759, base and emitter growth 137

A.2 XRR and PL sample #1334, $Si_{1-x}Ge_x$ and Si cap layer growth . . . 138

A.3 Growth details for superlattice sample #1432 139

A.4 Sample #1297, p-type RTD 140

A.5 Sample #934, flat-base HBT 141

A.6 Sample #1570, lateral transport device 142

A.7 Sample #1072, n-type RTD 144

List of Figures

2.1 Simulated band diagrams of an npn Si/$Si_{0.8}Ge_{0.2}$/Si HBT (solid line) and Si BJT (dashed line) with the same doping levels in forward-active mode under the same bias conditions: $V_{BE} = 0.5V$, $V_{CE} = 1V$. The emitter and collector in both devices is Si 5

2.2 A comparison of strained (a) and relaxed (b) $Si_{1-x}Ge_x$ layers on a silicon <100> substrate 10

2.3 Equilibrium critical thickness of $Si_{1-x}Ge_x$ alloy layer on Si 11

2.4 Bandgaps of strained and unstrained $Si_{1-x}Ge_x$ alloys vs. Ge concentration 12

2.5 Band lineup of strained $Si_{0.8}Ge_{0.2}$ layers on <100> Si-substrate, $\Delta E_C \sim$ 20meV, $\Delta E_V \sim$ 150meV 13

2.6 Schematic diagram of ECL gate used in SPICE simulations, R = 1.66kΩ for low power simulation and R = 500Ω for high power simulation. 18

2.7 Simulated gate delay as a function of intrinsic base sheet resistance for various f_T 's, $A_E = 1 \times 2\mu m^2$, P ~2mW/gate 19

2.8 Simulated gate delay as a function of intrinsic base sheet resistance for various C_{BEO}'s, $A_E = 1 \times 2\mu m^2$, P ~2mW/gate 20

2.9 Simulated gate delay as a function of intrinsic base sheet resistance for various f_T 's, $A_E = 1 \times 2\mu m^2$, P ~8mW/gate 21

2.10 Simulated gate delay as a function of cutoff frequency for various $R_{B,sheet}$, $A_E = 1 \times 2\mu m^2$, P ~8mW/gate 22

2.11 a) Simulated profiles (SUPREM) of an HBT structure as-grown (solid) and after a 30 minutes anneal at 800°C (dashed). b) Simulated zero-bias band diagrams (SEDAN) of the profiles of (a). Note the parasitic barriers formed when B outdiffuses from the $Si_{1-x}Ge_x$ base (dashed). 25

2.12 SUPREM simulation of $Si/Si_{1-x}Ge_x/Si$ HBT doping profiles: as grown (a) and at various annealing conditions (b,c,d). The Ge profile is also shown in (a) 29

2.13 As grown SIMS profile of the device used for annealing experiments (#759) 30

2.14 Measured collector current for various annealing conditions of a device with 350Å thick 8×10^{19} cm^{-3} doped base with 150Å thick BC and BE $Si_{1-x}Ge_x$ spacers (#759) 31

3.1 Schematic diagram of the RTCVD reactor used in ths work 34

3.2 Ge fraction vs. GeH_4 flow for different growth pressures at 625°C. The H_2 was 3 slpm and the dichlorosilane flow was 26 sccm. The uncertainty in Ge fraction is +/- 2%. 36

3.3 Growth rate vs. GeH_4 flow for different growth pressures at 625°C. The H_2 was 3 slpm and the dichlorosilane flow was 26 sccm. The uncertainty in growth rate is +/- 10%. 37

3.4 Typical PL spectra of SiGe samples grown at 6, 60, and 220 torr. Note the no-phonon (NP) peak and transverse optical (TO) phonon replica indicating band-edge luminescence and low defect concentration. . . . 39

3.5 Qualitative diagram that shows the sequence of gas and temperature switching during growth 41

3.6 Typical XRR spectra for structures with single $Si_{1-x}Ge_x$ layers ($x = 0.2-0.3$, $t = 200 - 300$Å) with 150Å Si caps. The reflected intensity is plotted as a function of scattering vector for different pressures. The solid line represents the data and the dashed line is the best fit. . . . 43

3.7 The interface width of the top and bottom $Si_{1-x}Ge_x$/Si interfaces as a function of growth pressure. 44

3.8 Qualitative diagram that explains the effects of gas switching on the thicknesses of graded layers at the top and bottom interface. The thickness of the interface layer is proportional to the cross-hatched area in the lower figure. 46

3.9 a) Desired valence band diagram for resonant tunneling structures grown at various pressures. b) and c) Qualitative band diagrams of structures grown at 60 and 220 torr, respectively, showing the effect of graded interfaces. 48

3.10 RTD I-V curves measured at 80K of samples grown at 6, 60, and 220 torr. The desired structure was the same in all three samples. 49

3.11 X-ray diffraction data of a 20-period $Si_{0.8}Ge_{0.2}$/Si superlattice sample compared to simulation to determine the interface roughness. 52

4.1 Band diagram of a p-n junction illustrating tunneling currents through interface states 56

4.2 Simulated band diagram of Si/Si (solid) and $Si/Si_{0.85}Ge_{0.15}$ (dashed) p^+–n^+junction at $V_a = 0.48V$, $N_A = 5 \times 10^{19}$ cm^{-3}, $N_D = 1 \times 10^{19}$ cm^{-3}. The cross-hatched areas show electron and hole tunneling barriers. . 58

4.3 Room temperature forward bias I-V characteristics of several Si/Si and Si/SiGe devices with the following n-type dopings: (a) 1×10^{19} cm^{-3} (Si/SiGe); (b) 9×10^{18} cm^{-3} (Si/Si); (c) 7×10^{18} cm^{-3} (Si/Si); (d) 3×10^{18} cm^{-3} (Si/Si); (e) 1.5×10^{18} cm^{-3} (Si/Si and Si/SiGe); (f) 1.3×10^{17} cm^{-3} (Si/Si). 60

4.4 Current density vs. n-type doping at (a) $V_a = 0.32$V and (b) $V_a = 0.48$V. J_{ideal} is the calculated ideal current density, $J_{T,ion.impl.}$ corresponds to the tunneling currents in ion implanted junctions and $J_{T,this\ work}$ corresponds to this work. The solid lines are fits to the heavily doped data points (7×10^{18} cm^{-3} - 1×10^{19} cm^{-3}). 62

4.5 Measured current density as a function of doping at T=200K and $V_a = 0.48$V. Tunneling currents of heavily doped devices measured at room temperature are also shown for comparison. 63

4.6 a) Typical current vs. temperature dependence of a heavily doped SiGe/Si device ($N_D = 1 \times 10^{19}$ cm^{-3}, $V_a = 0.32$V, $A = 3.25 \times 10^{-4}$ cm^2). b) Current vs. $Si_{0.85}Ge_{0.15}$ bandgap of the same device. The data confirms the expected shape of excess tunneling current. 65

4.7 Calculated effects of the tunneling current on the gain of $Si/Si_{1-x}Ge_x/Si$ HBT's as the emitter doping is increased. Both curves based on the ion implantation results and this data are given for comparison, as well as the ideal curve (no tunneling). 67

5.1 A typical SIMS profile of a device used in this study. Si and Ge are in arbitrary units. B was contained within the SiGe layer even for the heaviest doped devices. 72

5.2 A typical device structure used in this study. Van der Pauw patterns were fabricated next to transistor devices. 73

5.3 a) Hole lateral Hall mobility as a function of base doping for various Ge concentrations, b) Hole lateral Hall mobility as a function of Ge concentration for different doping levels. 75
5.4 Hall scattering factor for holes as a function of Ge concentration . . . 76
5.5 a) Drift mobility as a function of base doping for various Ge concentrations. The line is the fit to data, b) Drift mobility as a function of Ge concentrations for two doping levels. 78
5.6 Typical Gummel plot of an HBT used in this study. Ge concentration in this device was 23% with the base doping of 3×10^{18} cm^{-3} 81
5.7 Effective bandgap reduction with respect to intrinsic Si vs. Ge concentration 84
5.8 Apparent bandgap narrowing vs. base doping after linear dependence of the bandgap reduction on Ge content has been subtracted 85
5.9 Apparent bandgap narrowing vs. base doping after linear dependence of the bandgap reduction on Ge content has been subtracted compared with data from literature 86
5.10 Qualitative band diagram illustrating the difference between the true and the effective bandgap 88
5.11 Position of the Fermi level with respect to the valence band edge for Si and $Si_{1-x}Ge_x$ 89
5.12 Calculated Fermi-Dirac correction to the bandgap narrowing as a function of base doping for various Ge concentrations at room temperature 90
5.13 The heavy-doping contribution to the true bandgap reduction in strained $Si_{1-x}Ge_x$ calculated from the measured $\Delta E_{G,eff}$ as: $\Delta E_{G,dop,true} = \Delta E_{G,dop} - \Delta E_{G,FD}$. Also shown for comparison are the Si data points of Wagner and the theoretical calculation for p-Si of Jain and Roulston (solid line). 91

5.14 Relative collector current vs. base sheet resistance. The lines correspond to the model, the points are the data 93

6.1 Device structure optimized using MEDICI (2D simulator) for diffusion length measurements 100

6.2 Simulated band diagram of p^+-Si/p^+-$Si_{0.8}Ge_{0.2}$/p^+-Si 101

6.3 Simulated collector current density for a $Si_{0.8}Ge_{0.2}$ channel device (w_{SiGe} = 600Å, N_A(SiGe) = 1×10^{19} cm^{-3}, N_A(Si) = 5×10^{19} cm^{-3}, $w_{Si,cap}$ = 200Å, $w_{Si,buffer}$ = 5μm, w_B = 15μm), and bulk Si and bulk $Si_{0.8}Ge_{0.2}$ devices, (N_A = 1×10^{19} cm^{-3}, $w_{vertical}$ = 20μm, w_B = 15μm in both cases) 103

6.4 Simulated collector current density for a $Si_{0.8}Ge_{0.2}$ channel device ($w_{Si,cap}$ = 200Å, w_{SiGe} = 600Å, $w_{Si,buffer}$ = 5μm) when electron mobilities in Si and $Si_{0.8}Ge_{0.2}$ layers are varied 104

6.5 Effective lifetime as a function of true lifetime for various surface recombination velocities 106

6.6 Device structure used for diffusion length measurements. The collector contact area was $150 \times 150\mu m^2$, the basewidths ranged from 1μm to 30μm. 108

6.7 Photocurrent vs. basewidth for N_A =1.7×10^{18} cm^{-3} in $Si_{0.8}Ge_{0.2}$ layer 111

6.8 Room temperature diffusion length vs.doping in $Si_{0.8}Ge_{0.2}$, published Si data is also shown 113

6.9 Low temperature diffusion length vs.doping in $Si_{0.8}Ge_{0.2}$, also shown published Si data 114

7.1 Schematic conduction band diagram at zero bias. Also shown are the different layers: silicon spacers, w_s = 175Å; $Si_{0.65}Ge_{0.35}$ barriers, w_b = 40 – 70Å; and Si well, w_w = 20 – 50Å. 119

7.2 RTD device structure 121

7.3 I-V curves of a device with 25Å well and 60Å barriers (area=70 × 70μm^2) at various temperatures: (a) 220K, (b) 150K, (c) 80K, (d) 4.2K. . . . 122

7.4 Typical I-V and dI/dV-V curves of a 50Å well/70Å barriers device (area=100 × 100μm^2) with four distinct resonances at 80K 123

7.5 Schematic diagram of phonon-absorption-assisted tunneling indicating phonon absorption prior to tunneling (path $a_1 - a_2$) and after tunneling (path $b_1 - b_2$). 125

7.6 Schematic diagram illustrating direct and phonon-assisted electron tunneling processes in SiGe/Si material system. Phonon-assisted process: $\Delta k_z \approx q \approx 2 \times 0.85\pi/a + G \approx 0.3\pi/a$ 126

7.7 I-V curves of the lowest energy resonance for a 50Å well device at three different temperatures 128

7.8 Integrated tunneling current for the lowest energy resonance vs. 1000/T for: (a) 25Å well device, positive bias, (b) 25Å well device, negative bias, (c) 50Å well device, positive bias. The dashed lines are fits to phonon population curves (a) $\hbar\omega = 14meV$, (b) $\hbar\omega = 16meV$, (c) $\hbar\omega = 12meV$ 129

Chapter 1

Introduction

1.1 Motivation

The $Si_{1-x}Ge_x$ material system has received a lot of attention in recent years because of improved device performance compared to existing silicon devices and because of a wide range of new devices that could potentially be integrated into current silicon VLSI technology. The possibility of bandgap engineering in a Si-based material system has boosted research activities both in unipolar and bipolar devices, as well as in novel device structures such as resonant tunneling devices or modulation doped structures. The bandgap of strained $Si_{1-x}Ge_x$ makes this material an interesting candidate for optoelectronic devices, such as infrared detectors.

Among the wide range of research areas, $Si/Si_{1-x}Ge_x/Si$ heterojunction bipolar transistors have received the most attention and made tremendous progress. The scaling of lateral and vertical dimensions in silicon bipolar technology forces several tradeoffs on the design of bipolar devices. For example, an increase in the base doping desired to avoid punchthrough is limited by a high-gain requirement. Such tradeoffs can be overcome in $Si/Si_{1-x}Ge_x/Si$ HBT's since increased base doping is compensated by germanium in the base to maintain high gain. Furthermore, high base doping implies low base sheet resistance which is important for high speed performance. Speed in heterojunction bipolar devices is even further enhanced by narrow epitaxial bases with graded germanium profiles. $Si/Si_{1-x}Ge_x/Si$ HBT's with peak cutoff frequencies of over 100GHz have been demonstrated [1, 2], as well as a 12-bit Digital to Analog Convertor (DAC) built in $Si/Si_{1-x}Ge_x$ technology operating at 1GHz [3].

The development of high-performance devices has been enabled by advanced growth techniques such as Molecular Beam Epitaxy (MBE) or Chemical Vapor Deposition (CVD). However, the integration of epitaxially grown $Si/Si_{1-x}Ge_x$ layers into silicon bipolar technology is not straightforward. High quality abrupt interfaces are desired for good device performance. Furthermore, $Si/Si_{1-x}Ge_x$ heterostructures are sensitive to high temperature processing which implies modifications in standard silicon processes to lower thermal budgets. Increased base doping may result in the formation of a p^+–n^+ junction at the base-emitter interface, which could affect the device performance. Some of the issues related to process integration of $Si/Si_{1-x}Ge_x/Si$ HBT's are addressed in this thesis.

Although remarkable results in terms of high gain and high speed performance have been demonstrated, experimental data for even DC modeling is still lacking. For example, the effects of heavy doping in $Si_{1-x}Ge_x$ on bandgap narrowing have not been substantially verified experimentally. This thesis describes an experimental study of doping effects on lateral majority carrier transport, important for base resistance, and bandgap narrowing and minority carrier properties in the base, important for accurate modeling of collector current.

1.2 Thesis outline

Chapter 1 gives the motivation for undertaking the work resulting in this thesis.

Chapter 2 provides an introduction to the $Si/Si_{1-x}Ge_x$ material system and heterojunction bipolar transistors (HBT's). It gives a comparison of HBT's and Si bipolar transistors from the application perspective. Further, important issues related to process integration of Si/$Si_{1-x}Ge_x$/Si HBT's are addressed.

Chapter 3 gives the details on Rapid Thermal Chemical Vapor Deposition (RTCVD) used to grow samples in this work. The effects of the growth pressure on the material quality are studied by X-ray reflectivity, photoluminescence and performance of resonant tunneling diodes. A low pressure limit to interface abruptness of epitaxially grown layers by RTCVD is established.

In chapter 4 tunneling in heavily doped p^+–n^+ junctions is studied. Since bases in Si/$Si_{1-x}Ge_x$/Si HBT's are generally more heavily doped than those in Si BJT's, significant tunneling could cause an increase in the base current and degradation of device performance.

In chapter 5 majority carrier properties of $Si_{1-x}Ge_x$ base are studied and a DC collector current model is developed. An empirical model of bandgap narrowing in heavily doped $Si_{1-x}Ge_x$ layers is established and the DC design trade-off between the base sheet resistance and gain is emphasised.

Chapter 6 examines the minority carrier properties in strained $Si_{1-x}Ge_x$. Measurements of minority carrier diffusion length are presented.

A novel vertical transport device in the Si–based material system, an electron resonant tunneling diode, is presented in chapter 7. The conduction band offset for n-type double barrier structures is provided by a graded, relaxed $Si_{1-x}Ge_x$ substrate. The temperature behaviour of resonant features is studied.

The most important results from this thesis are summarized in chapter 8 and future research directions are discussed.

Chapter 2

Integration of Si/SiGe/Si HBT's into Si-technology

2.1 Introduction

This chapter first explains the basic operating principle of an HBT and the physics of strained $Si_{1-x}Ge_x$ material system on <100> Si. It further gives the technical motivation for most of the work in this thesis: the importance of heavy doping in $Si_{1-x}Ge_x$. It is first addressed from the perspective of speed of bipolar devices and digital circuits. Then the limits of integration of heavily doped devices into Si technology are discussed and it is shown how the problems of process integration could be acceptably solved.

2.2 Si/SiGe/Si heterojunction bipolar transistors

The operating principle of a heterojunction bipolar transistor relies on the possibility of varying the bandgap in a bipolar transistor structure in order to increase emitter injection efficiency. In a III-V material system this is usually accomplished by having an emitter with a wider bandgap material than the rest of the structure, such that the barrier for holes injected from the base into the emitter is increased resulting in lower base current and thus higher gain. The same effect is achieved in

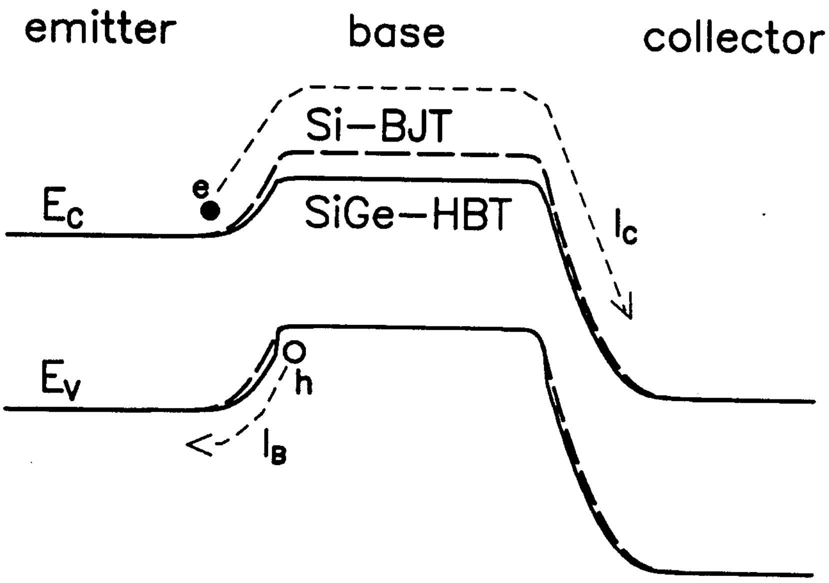


Figure 2.1: Simulated band diagrams of an npn Si/$Si_{0.8}Ge_{0.2}$/Si HBT (solid line) and Si BJT (dashed line) with the same doping levels in forward-active mode under the same bias conditions: $V_{BE} = 0.5V$, $V_{CE} = 1V$. The emitter and collector in both devices is Si

Si/$Si_{1-x}Ge_x$ material system by making the base of a narrower bandgap material (Fig. 2.1) so the potential barrier for electrons injected into the base is decreased. Although the basic physics is very simple, heterojunction bipolar devices were not realized until growth techniques like MBE or CVD were developed to provide the high quality material required for bipolar devices [4].

Fig. 2.1 shows a band diagram of an npn Si/$Si_{0.8}Ge_{0.2}$ HBT in forward-active mode, where the base-emitter junction is forward-biased and base-collector junction is reverse-biased. Also shown in the figure (dashed) is a band diagram of an all-Si device with the same dopings in the emitter, base and collector layers and under the same bias conditions. Since the number of electrons injected from the emitter into the base exponentially depends on the height of the conduction band barrier at the base-emitter junction, the increase in the collector current in the heterojunction device compared to the homojunction device will exponentially depend on the bandgap difference between the two materials, ΔE_G. In the ideal case, the dominant component of the base current is the hole current injected from the base into the emitter. Since the valence band barrier for the holes at the base-emitter junction is of the same height for both devices, the base current should be the same for the same V_{BE}. Therefore, the common-emitter current gain ($\beta = I_C/I_B$) will also exponentially increase with the bandgap difference ($\beta \propto \exp(\Delta E_G/k_B T)$).

2.3 Development of high speed Si/SiGe/Si HBT's, historic overview

Because of the possibility of high gain, high output resistance, high speed performance and relative compatibility with the existing silicon technology $Si/Si_{1-x}Ge_x/Si$ HBT's have received enormous attention in the recent years and impressive results have been obtained.

The first devices were grown by MBE [5, 6, 7, 8, 9, 10, 11]. These devices were characterized by nonideal base currents, most likely due to low minority carrier lifetimes or defects in the epitaxial layers or interfaces, which resulted in poor gain performance. The first devices that clearly demonstrated the exponential increase in gain with the bandgap offset (King et al. [12, 13, 14]) were grown by Limited Reaction Processing (LRP, a non-UHV CVD technique developed at Stanford University [15]). These devices were characterized by high level of oxygen contamination in the layers. It was also shown that the incorporation of misfit dislocations in partially relaxed $Si_{1-x}Ge_x$ base layers leads to poor minority carrier lifetimes, i.e. non-ideal base currents.

The first devices with grading in the base for reduced base transit times integrated into poly-emitter process were grown by UHV CVD at IBM (Patton et al. [16]). These devices had near-ideal base currents. The near-ideal base currents in devices with graded profiles grown by a non-UHV CVD technique, namely RTCVD, were demonstrated by Prinz et al. [17]. Soon afterwards, the first MBE grown devices with ideal base currents followed (Pruijmboom et al. [18]).

Several groups have been working on high speed optimization and process integration [19, 20, 21], the bulk of results being demonstrated by the group at IBM. The next break-through result in development of high-speed HBT's was a cutoff frequency of 75GHz reported by Patton et al. [20]. The first ECL circuit results followed soon afterwards (Burghartz et al. [22, 21]) with sub-30ps delays. The limited ECL improvement over all-Si devices fabricated by the same process was due to high base resistances, both extrinsic and intrinsic, despite high f_T's in the $Si_{1-x}Ge_x$ –base devices. Further improvement in design and technology resulted in the development of low-thermal-cycle SiGe-base ECL BiCMOS technology with f_T's of 50GHz, f_{max}'s of 59GHz and 19ps ECL gate delay (at $\sim$ 8mW) [23].

Gruhle et al. [24] reported the first high speed performance devices grown by MBE with f_T's of 46GHz and f_{max}'s of 53GHz. All of the layers were *in situ* doped. The processing of these devices was not compatible with VLSI Si technology since there were no high temperature steps at all and gold contacts were used. More recently, cutoff frequencies of MBE-grown HBT's with low base resistivities of over 100GHz have been demonstrated [2] and an f_{max} of 90GHz has been reported [25].

The high βV_A products of HBT devices make them interesting for analog applications. Fabricated in the same low-thermal-cycle process with UHV CVD grown bases, devices with SiGe-bases demonstrated f_T's of 113GHz and βV_A products of 48400V compared to f_T's of 73GHz and βV_A of 630 for Si-base devices [1]. Finally, with aggressive grading of the Ge profile in the base, scaled dimensions and reduced base resistance, ECL gate delays of 17ps have been achieved and the first MSI circuit, namely a 12-bit DA converter, has been realized in Si/$Si_{1-x}Ge_x$ technology operating at 1GHz, much faster than any existing Si DAC (Harame et al. [3]).

The Si/$Si_{1-x}Ge_x$ bipolar technology has made tremendous progress in the past years. While the early devices were fabricated with lower base doping levels, the doping was increased and processing optimized for high-speed performance devices, resulting in successful integration of HBT's with poly-emitter process. This chapter provides some of the basics necessary for understanding the advantages and limitations of Si/$Si_{1-x}Ge_x$/Si HBT's and addressed some of the challenges that the development of Si/$Si_{1-x}Ge_x$ technology has been facing. Although remarkable results have been achieved, several effects of heavy doping in strained $Si_{1-x}Ge_x$ still need to be studied. For example, the experimental data to establish the heavy-doping bandgap narrowing in $Si_{1-x}Ge_x$ is still insufficient. This thesis addresses some important issues related to base doping effects on DC device performance, such as tunneling in epitaxial p^+–n^+ junctions and hole mobility, minority carrier properties and bandgap narrowing in heavily doped base layers.

2.4 The Si/SiGe heterojunction

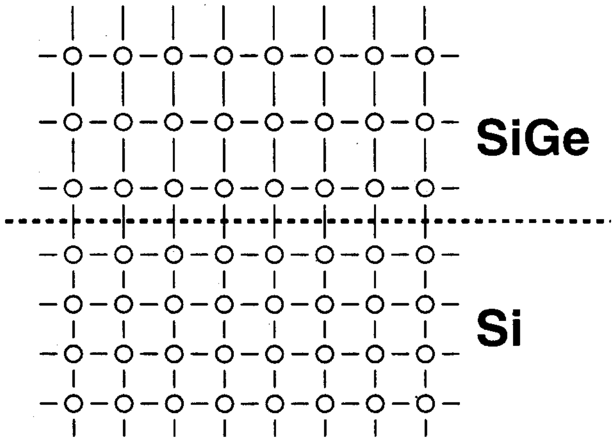
The Ge lattice constant (5.64Å) is larger than that of Si (5.43Å) resulting in a 4% lattice mismatch. This implies two possible ways of growth of $Si_{1-x}Ge_x$ alloy layers on a <100> silicon substrate to accommodate the mismatch. One is a relaxed alloy layer with a lattice constant interpolated between the Si and Ge values. Such growth results in incorporation of misfit dislocations at the Si/$Si_{1-x}Ge_x$ interface (Fig. 2.2.b), since the atoms in the Si substrate are not matched by atoms in the $Si_{1-x}Ge_x$ layer. Such growth is not desired for high quality material. The interface defects could increase leakage currents in p-n junctions and lower minority carrier lifetimes, which are important for good bipolar device performance. It has been shown that the incorporation of misfit dislocations causes non-ideal base currents in Si/$Si_{1-x}Ge_x$/Si HBT's [14].

The other possibility is pseudomorphic growth without dislocations, resulting in formation of a strained layer (Fig. 2.2.a). Such a layer has a lattice constant in the growth plane matching that of the Si substrate, but this is compensated by a larger lattice constant in the growth direction.

For thin $Si_{1-x}Ge_x$ films on <100> Si substrates it is energetically favorable to remain coherently strained. However, if the film thickness is increased, the energy associated with strain increases and above certain thickness it becomes energetically favorable for the material to relax by formation of misfit dislocations. The thickness above which a strained film is likely to relax is called the critical thickness. There are several theories to calculate the critical thickness [26, 27, 28, 29, 30] assuming that the structure is in the equilibrium state of its lowest energy. Fig. 2.3 shows the most commonly used critical thickness for strained $Si_{1-x}Ge_x$ layers calculated by Mathews and Blakeslee [27]. This implies that only very thin defect-free $Si_{1-x}Ge_x$ layers can be epitaxially grown for device applications. However, dislocation-free strained layers of thicknesses above the critical thickness can be grown in a metastable state, where kinetic limitations prevent the structure from relaxing. These layers are temperature sensitive, since they will return to the equilibrium state with energy provided from the heat, by relaxing the strain and generating dislocations. This imposes a high temperature restriction to the processing of Si/$Si_{1-x}Ge_x$ devices.

Both silicon and germanium are indirect bandgap materials, with bandgaps at zero Kelvin of 1.17eV and 0.66eV, respectively. The conduction band minima in silicon are 6-fold degenerate and located along Δ axis. Valence band maxima are located at the center of the Brillouin zone Γ. Up to high germanium concentrations in the alloy layer ($x \sim 0.8$) the band minima in bulk unstrained $Si_{1-x}Ge_x$ alloy are

a) strained:



b) relaxed:

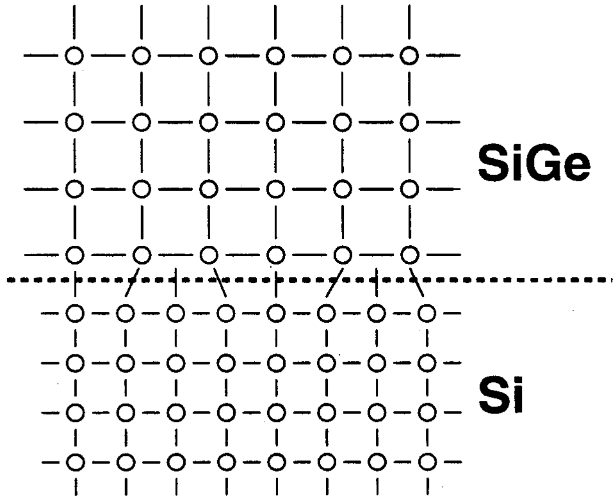


Figure 2.2: A comparison of strained (a) and relaxed (b) $Si_{1-x}Ge_x$ layers on a silicon <100> substrate

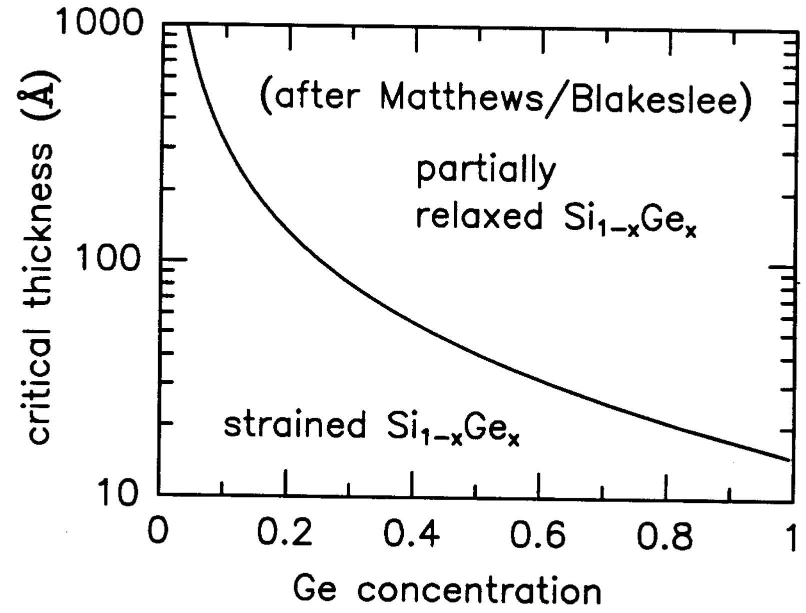


Figure 2.3: Equilibrium critical thickness of $Si_{1-x}Ge_x$ alloy layer on Si

silicon-like, with the bandgap decreasing as Ge content is increased [31]. The strain in $Si_{1-x}Ge_x$ alloy affects the band structure. When grown on <100> Si substrate the $Si_{1-x}Ge_x$ film is under biaxial compressive strain which lifts the degeneracy of conduction and valence bands. Four conduction band minima in the growth plane move down with respect to the other two in the growth direction. In the valence band, the degenerate heavy and light hole bands split with the heavy hole band lying higher. The resulting bandgap is lower than that of an unstrained material [32, 33, 34] and decreases much more rapidly with increasing Ge content. The bandgaps of strained (People et al. [32]) and unstrained (Braunstein et al. [31]) random $Si_{1-x}Ge_x$ alloys as a function of Ge concentration are shown in Fig. 2.4.

The band lineup at $Si/Si_{1-x}Ge_x$ interface is shown in Fig. 2.5. Up to Ge concentrations of ~ 60%, a Type-I band alignment has been calculated [33, 35] and experimentally confirmed [36]. The conduction band offset is small while most of the bandgap offset occurs in the valence band. From the perspective of SiGe HBT's this

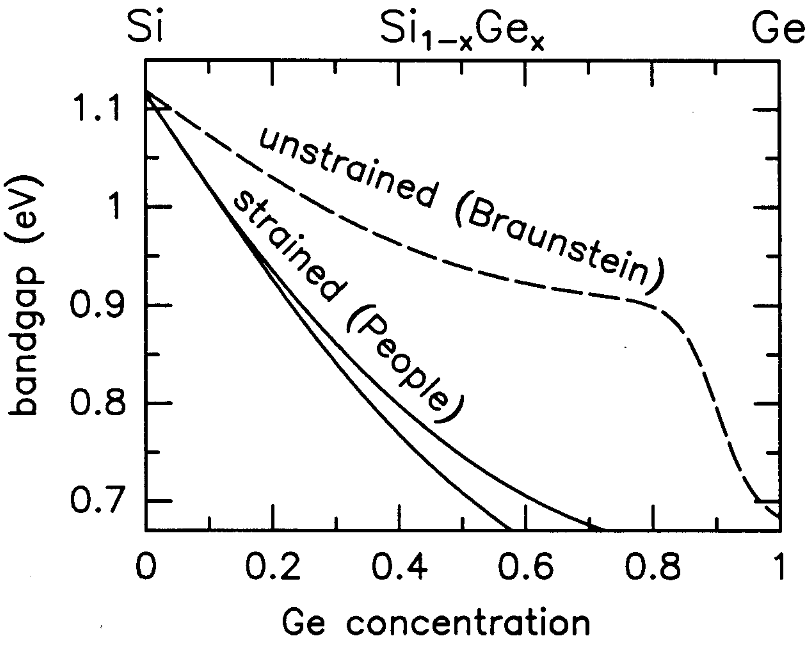


Figure 2.4: Bandgaps of strained and unstrained $Si_{1-x}Ge_x$ alloys vs. Ge concentration

is ideal for npn structures but not for pnp devices, where germanium grading at the interface is necessary to avoid "spike and notch" effects [37].

2.5 Speed of Si/SiGe heterojunction bipolar devices and circuits

One of the most significant advantages of $Si/Si_{1-x}Ge_x/Si$ HBT's over Si bipolar devices is speed. This section outlines the basic issues important for understanding the high frequency performance of bipolar devices and circuits and explains some of the limitations of Si BJT's and possible improvements via HBT structures.

A typical parameter often used to characterize the speed of bipolar transistors is the cutoff frequency f_T. At high frequencies the current gain of a bipolar transistor will decrease. The cutoff frequency f_T is the frequency at which current gain reaches

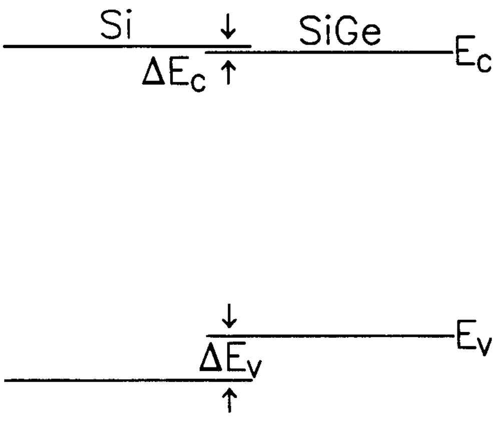


Figure 2.5: Band lineup of strained $Si_{0.8}Ge_{0.2}$ layers on <100> Si-substrate, $\Delta E_C \sim$ 20meV, $\Delta E_V \sim$ 150meV

unity. It is related to the transit time that it takes electrons from the emitter to reach the collector [38]:

$$\frac{1}{2\pi f_T} = \tau_{EC} = \tau_E + \tau_B + \tau_d + \tau_C \tag{2.1}$$

The transit time τ_{EC} is the sum of the transit times across the emitter τ_E, the base τ_B, the base-collector depletion region τ_d, and collector τ_C, and it reflects the vertical device profile. The emitter transit time consists of the emitter storage time and the time associated with charging up base junction capacitances. The emitter storage time is the delay due to hole charge storage in the emitter, and it is usually negligible in heavily doped emitters of high-gain bipolar devices. The main source of emitter delay is represented by the RC constant:

$$\tau_E = \frac{k_B T}{q I_E}(C_{BE} + C_{BC}) \tag{2.2}$$

Since it is inversely proportional to the DC emitter current, at high enough current levels τ_E is not a significant component of the overall delay in a bipolar transistor. The base transit time τ_B is the delay related to electron transport through the neutral base. For devices with flat doping profiles in the base the electrons move solely by diffusion

3.6 Typical XRR spectra for structures with single $Si_{1-x}Ge_x$ layers (x = 0.2–0.3, t =200 – 300Å) with 150Å Si caps. The reflected intensity is plotted as a function of scattering vector for different pressures. The solid line represents the data and the dashed line is the best fit. . . . 43

3.7 The interface width of the top and bottom $Si_{1-x}Ge_x$/Si interfaces as a function of growth pressure. 44

3.8 Qualitative diagram that explains the effects of gas switching on the thicknesses of graded layers at the top and bottom interface. The thickness of the interface layer is proportional to the cross-hatched area in the lower figure. 46

3.9 a) Desired valence band diagram for resonant tunneling structures grown at various pressures. b) and c) Qualitative band diagrams of structures grown at 60 and 220 torr, respectively, showing the effect of graded interfaces. 48

3.10 RTD I-V curves measured at 80K of samples grown at 6, 60, and 220 torr. The desired structure was the same in all three samples. 49

3.11 X-ray diffraction data of a 20-period $Si_{0.8}Ge_{0.2}$/Si superlattice sample compared to simulation to determine the interface roughness. 52

4.1 Band diagram of a p-n junction illustrating tunneling currents through interface states 56

4.2 Simulated band diagram of Si/Si (solid) and Si/$Si_{0.85}Ge_{0.15}$ (dashed) p^+–n^+junction at $V_a = 0.48$V, $N_A = 5 \times 10^{19}$ cm^{-3}, $N_D = 1 \times 10^{19}$ cm^{-3}. The cross-hatched areas show electron and hole tunneling barriers. . 58

4.3 Room temperature forward bias I-V characteristics of several Si/Si and Si/SiGe devices with the following n-type dopings: (a) 1×10^{19} cm^{-3} (Si/SiGe); (b) 9×10^{18} cm^{-3} (Si/Si); (c) 7×10^{18} cm^{-3} (Si/Si); (d) 3×10^{18} cm^{-3} (Si/Si); (e) 1.5×10^{18} cm^{-3} (Si/Si and Si/SiGe); (f) 1.3×10^{17} cm^{-3} (Si/Si). 60

4.4 Current density vs. n-type doping at (a) $V_a = 0.32$V and (b) $V_a = 0.48$V. J_{ideal} is the calculated ideal current density, $J_{T,ion.impl.}$ corresponds to the tunneling currents in ion implanted junctions and $J_{T,this\ work}$ corresponds to this work. The solid lines are fits to the heavily doped data points (7×10^{18} cm^{-3} – 1×10^{19} cm^{-3}). 62

4.5 Measured current density as a function of doping at T=200K and $V_a = 0.48$V. Tunneling currents of heavily doped devices measured at room temperature are also shown for comparison. 63

4.6 a) Typical current vs. temperature dependence of a heavily doped SiGe/Si device ($N_D = 1 \times 10^{19}$ cm^{-3}, $V_a = 0.32$V, $A = 3.25 \times 10^{-4}$ cm^2). b) Current vs. $Si_{0.85}Ge_{0.15}$ bandgap of the same device. The data confirms the expected shape of excess tunneling current. 65

4.7 Calculated effects of the tunneling current on the gain of $Si/Si_{1-x}Ge_x/Si$ HBT's as the emitter doping is increased. Both curves based on the ion implantation results and this data are given for comparison, as well as the ideal curve (no tunneling). 67

5.1 A typical SIMS profile of a device used in this study. Si and Ge are in arbitrary units. B was contained within the SiGe layer even for the heaviest doped devices. 72

5.2 A typical device structure used in this study. Van der Pauw patterns were fabricated next to transistor devices. 73

5.3 a) Hole lateral Hall mobility as a function of base doping for various Ge concentrations, b) Hole lateral Hall mobility as a function of Ge concentration for different doping levels. 75

5.4 Hall scattering factor for holes as a function of Ge concentration . . . 76

5.5 a) Drift mobility as a function of base doping for various Ge concentrations. The line is the fit to data, b) Drift mobility as a function of Ge concentrations for two doping levels. 78

5.6 Typical Gummel plot of an HBT used in this study. Ge concentration in this device was 23% with the base doping of 3×10^{18} cm^{-3} 81

5.7 Effective bandgap reduction with respect to intrinsic Si vs. Ge concentration 84

5.8 Apparent bandgap narrowing vs. base doping after linear dependence of the bandgap reduction on Ge content has been subtracted 85

5.9 Apparent bandgap narrowing vs. base doping after linear dependence of the bandgap reduction on Ge content has been subtracted compared with data from literature 86

5.10 Qualitative band diagram illustrating the difference between the true and the effective bandgap 88

5.11 Position of the Fermi level with respect to the valence band edge for Si and $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ 89

5.12 Calculated Fermi-Dirac correction to the bandgap narrowing as a function of base doping for various Ge concentrations at room temperature 90

5.13 The heavy-doping contribution to the true bandgap reduction in strained $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ calculated from the measured $\Delta E_{G,\mathrm{eff}}$ as: $\Delta E_{G,\mathrm{dop,true}} = \Delta E_{G,\mathrm{dop}} - \Delta E_{G,\mathrm{FD}}$. Also shown for comparison are the Si data points of Wagner and the theoretical calculation for p-Si of Jain and Roulston (solid line). 91

5.14 Relative collector current vs. base sheet resistance. The lines correspond to the model, the points are the data 93

6.1 Device structure optimized using MEDICI (2D simulator) for diffusion length measurements 100

6.2 Simulated band diagram of p^+-Si/p^+-$Si_{0.8}Ge_{0.2}$/p^+-Si 101

6.3 Simulated collector current density for a $Si_{0.8}Ge_{0.2}$ channel device (w_{SiGe} = 600Å, N_A(SiGe) = 1×10^{19} cm^{-3}, N_A(Si) = 5×10^{19} cm^{-3}, $w_{Si,cap}$ = 200Å, $w_{Si,buffer}$ = 5μm, w_B = 15μm), and bulk Si and bulk $Si_{0.8}Ge_{0.2}$ devices, (N_A = 1×10^{19} cm^{-3}, $w_{vertical}$ = 20μm, w_B = 15μm in both cases) 103

6.4 Simulated collector current density for a $Si_{0.8}Ge_{0.2}$ channel device ($w_{Si,cap}$ = 200Å, w_{SiGe} = 600Å, $w_{Si,buffer}$ = 5μm) when electron mobilities in Si and $Si_{0.8}Ge_{0.2}$ layers are varied 104

6.5 Effective lifetime as a function of true lifetime for various surface recombination velocities 106

6.6 Device structure used for diffusion length measurements. The collector contact area was 150 × 150μm^2, the basewidths ranged from 1μm to 30μm. 108

6.7 Photocurrent vs. basewidth for N_A =1.7×10^{18} cm^{-3} in $Si_{0.8}Ge_{0.2}$ layer 111

6.8 Room temperature diffusion length vs.doping in $Si_{0.8}Ge_{0.2}$, published Si data is also shown 113

6.9 Low temperature diffusion length vs.doping in $Si_{0.8}Ge_{0.2}$, also shown published Si data 114

7.1 Schematic conduction band diagram at zero bias. Also shown are the different layers: silicon spacers, w_s = 175Å; $Si_{0.65}Ge_{0.35}$ barriers, w_b = 40 – 70Å; and Si well, w_w = 20 – 50Å. 119

7.2 RTD device structure 121

7.3 I-V curves of a device with 25Å well and 60Å barriers (area=70 × 70μm^2) at various temperatures: (a) 220K, (b) 150K, (c) 80K, (d) 4.2K. . . . 122

7.4 Typical I-V and dI/dV-V curves of a 50Å well/70Å barriers device (area=100 × 100μm^2) with four distinct resonances at 80K 123

7.5 Schematic diagram of phonon-absorption-assisted tunneling indicating phonon absorption prior to tunneling (path $a_1 - a_2$) and after tunneling (path $b_1 - b_2$). 125

7.6 Schematic diagram illustrating direct and phonon-assisted electron tunneling processes in SiGe/Si material system. Phonon-assisted process: $\Delta k_z \approx q \approx 2 \times 0.85\pi/a + G \approx 0.3\pi/a$ 126

7.7 I-V curves of the lowest energy resonance for a 50Å well device at three different temperatures 128

7.8 Integrated tunneling current for the lowest energy resonance vs. 1000/T for: (a) 25Å well device, positive bias, (b) 25Å well device, negative bias, (c) 50Å well device, positive bias. The dashed lines are fits to phonon population curves (a) $\hbar\omega = 14meV$, (b) $\hbar\omega = 16meV$, (c) $\hbar\omega = 12meV$ 129

Chapter 1

Introduction

1.1 Motivation

The $Si_{1-x}Ge_x$ material system has received a lot of attention in recent years because of improved device performance compared to existing silicon devices and because of a wide range of new devices that could potentially be integrated into current silicon VLSI technology. The possibility of bandgap engineering in a Si-based material system has boosted research activities both in unipolar and bipolar devices, as well as in novel device structures such as resonant tunneling devices or modulation doped structures. The bandgap of strained $Si_{1-x}Ge_x$ makes this material an interesting candidate for optoelectronic devices, such as infrared detectors.

Among the wide range of research areas, $Si/Si_{1-x}Ge_x/Si$ heterojunction bipolar transistors have received the most attention and made tremendous progress. The scaling of lateral and vertical dimensions in silicon bipolar technology forces several tradeoffs on the design of bipolar devices. For example, an increase in the base doping desired to avoid punchthrough is limited by a high-gain requirement. Such tradeoffs can be overcome in $Si/Si_{1-x}Ge_x/Si$ HBT's since increased base doping is compensated by germanium in the base to maintain high gain. Furthermore, high base doping implies low base sheet resistance which is important for high speed performance. Speed in heterojunction bipolar devices is even further enhanced by narrow epitaxial bases with graded germanium profiles. $Si/Si_{1-x}Ge_x/Si$ HBT's with peak cutoff frequencies of over 100GHz have been demonstrated [1, 2], as well as a 12-bit Digital to Analog Convertor (DAC) built in $Si/Si_{1-x}Ge_x$ technology operating at 1GHz [3].

The development of high-performance devices has been enabled by advanced growth techniques such as Molecular Beam Epitaxy (MBE) or Chemical Vapor Deposition (CVD). However, the integration of epitaxially grown $Si/Si_{1-x}Ge_x$ layers into silicon bipolar technology is not straightforward. High quality abrupt interfaces are desired for good device performance. Furthermore, $Si/Si_{1-x}Ge_x$ heterostructures are sensitive to high temperature processing which implies modifications in standard silicon processes to lower thermal budgets. Increased base doping may result in the formation of a p^+–n^+ junction at the base-emitter interface, which could affect the device performance. Some of the issues related to process integration of $Si/Si_{1-x}Ge_x/Si$ HBT's are addressed in this thesis.

Although remarkable results in terms of high gain and high speed performance have been demonstrated, experimental data for even DC modeling is still lacking. For example, the effects of heavy doping in $Si_{1-x}Ge_x$ on bandgap narrowing have not been substantially verified experimentally. This thesis describes an experimental study of doping effects on lateral majority carrier transport, important for base resistance, and bandgap narrowing and minority carrier properties in the base, important for accurate modeling of collector current.

1.2 Thesis outline

Chapter 1 gives the motivation for undertaking the work resulting in this thesis.

Chapter 2 provides an introduction to the $Si/Si_{1-x}Ge_x$ material system and heterojunction bipolar transistors (HBT's). It gives a comparison of HBT's and Si bipolar transistors from the application perspective. Further, important issues related to process integration of Si/Si$_{1-x}$Ge$_x$/Si HBT's are addressed.

Chapter 3 gives the details on Rapid Thermal Chemical Vapor Deposition (RTCVD) used to grow samples in this work. The effects of the growth pressure on the material quality are studied by X-ray reflectivity, photoluminescence and performance of resonant tunneling diodes. A low pressure limit to interface abruptness of epitaxially grown layers by RTCVD is established.

In chapter 4 tunneling in heavily doped p^+–n^+ junctions is studied. Since bases in Si/Si$_{1-x}$Ge$_x$/Si HBT's are generally more heavily doped than those in Si BJT's, significant tunneling could cause an increase in the base current and degradation of device performance.

In chapter 5 majority carrier properties of Si$_{1-x}$Ge$_x$ base are studied and a DC collector current model is developed. An empirical model of bandgap narrowing in heavily doped Si$_{1-x}$Ge$_x$ layers is established and the DC design trade-off between the base sheet resistance and gain is emphasised.

Chapter 6 examines the minority carrier properties in strained Si$_{1-x}$Ge$_x$. Measurements of minority carrier diffusion length are presented.

A novel vertical transport device in the Si–based material system, an electron resonant tunneling diode, is presented in chapter 7. The conduction band offset for n-type double barrier structures is provided by a graded, relaxed Si$_{1-x}$Ge$_x$ substrate. The temperature behaviour of resonant features is studied.

The most important results from this thesis are summarized in chapter 8 and future research directions are discussed.

Chapter 2

Integration of Si/SiGe/Si HBT's into Si-technology

2.1 Introduction

This chapter first explains the basic operating principle of an HBT and the physics of strained $Si_{1-x}Ge_x$ material system on <100> Si. It further gives the technical motivation for most of the work in this thesis: the importance of heavy doping in $Si_{1-x}Ge_x$. It is first addressed from the perspective of speed of bipolar devices and digital circuits. Then the limits of integration of heavily doped devices into Si technology are discussed and it is shown how the problems of process integration could be acceptably solved.

2.2 Si/SiGe/Si heterojunction bipolar transistors

The operating principle of a heterojunction bipolar transistor relies on the possibility of varying the bandgap in a bipolar transistor structure in order to increase emitter injection efficiency. In a III-V material system this is usually accomplished by having an emitter with a wider bandgap material than the rest of the structure, such that the barrier for holes injected from the base into the emitter is increased resulting in lower base current and thus higher gain. The same effect is achieved in

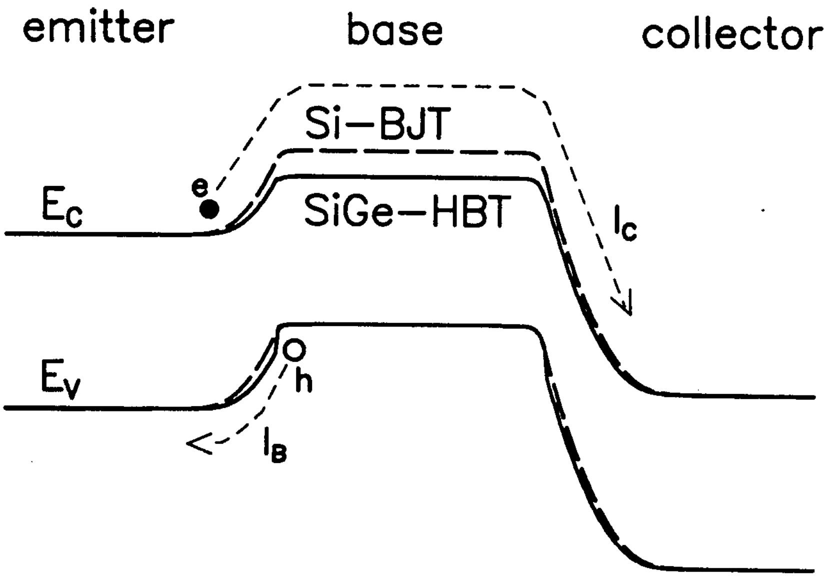


Figure 2.1: Simulated band diagrams of an npn $Si/Si_{0.8}Ge_{0.2}/Si$ HBT (solid line) and Si BJT (dashed line) with the same doping levels in forward-active mode under the same bias conditions: $V_{BE} = 0.5V$, $V_{CE} = 1V$. The emitter and collector in both devices is Si

$Si/Si_{1-x}Ge_x$ material system by making the base of a narrower bandgap material (Fig. 2.1) so the potential barrier for electrons injected into the base is decreased. Although the basic physics is very simple, heterojunction bipolar devices were not realized until growth techniques like MBE or CVD were developed to provide the high quality material required for bipolar devices [4].

Fig. 2.1 shows a band diagram of an npn $Si/Si_{0.8}Ge_{0.2}$ HBT in forward-active mode, where the base-emitter junction is forward-biased and base-collector junction is reverse-biased. Also shown in the figure (dashed) is a band diagram of an all-Si device with the same dopings in the emitter, base and collector layers and under the same bias conditions. Since the number of electrons injected from the emitter into the base exponentially depends on the height of the conduction band barrier at the base-emitter junction, the increase in the collector current in the heterojunction device compared to the homojunction device will exponentially depend on the bandgap difference between the two materials, ΔE_G. In the ideal case, the dominant component of the base current is the hole current injected from the base into the emitter. Since the valence band barrier for the holes at the base-emitter junction is of the same height for both devices, the base current should be the same for the same V_{BE}. Therefore, the common-emitter current gain ($\beta = I_C/I_B$) will also exponentially increase with the bandgap difference ($\beta \propto \exp(\Delta E_G/k_B T)$).

2.3 Development of high speed Si/SiGe/Si HBT's, historic overview

Because of the possibility of high gain, high output resistance, high speed performance and relative compatibility with the existing silicon technology Si/$Si_{1-x}Ge_x$/Si HBT's have received enormous attention in the recent years and impressive results have been obtained.

The first devices were grown by MBE [5, 6, 7, 8, 9, 10, 11]. These devices were characterized by nonideal base currents, most likely due to low minority carrier lifetimes or defects in the epitaxial layers or interfaces, which resulted in poor gain performance. The first devices that clearly demonstrated the exponential increase in gain with the bandgap offset (King et al. [12, 13, 14]) were grown by Limited Reaction Processing (LRP, a non-UHV CVD technique developed at Stanford University [15]). These devices were characterized by high level of oxygen contamination in the layers. It was also shown that the incorporation of misfit dislocations in partially relaxed $Si_{1-x}Ge_x$ base layers leads to poor minority carrier lifetimes, i.e. non-ideal base currents.

The first devices with grading in the base for reduced base transit times integrated into poly-emitter process were grown by UHV CVD at IBM (Patton et al. [16]). These devices had near-ideal base currents. The near-ideal base currents in devices with graded profiles grown by a non-UHV CVD technique, namely RTCVD, were demonstrated by Prinz et al. [17]. Soon afterwards, the first MBE grown devices with ideal base currents followed (Pruijmboom et al. [18]).

Several groups have been working on high speed optimization and process integration [19, 20, 21], the bulk of results being demonstrated by the group at IBM. The next break-through result in development of high-speed HBT's was a cutoff frequency of 75GHz reported by Patton et al. [20]. The first ECL circuit results followed soon afterwards (Burghartz et al. [22, 21]) with sub-30ps delays. The limited ECL improvement over all-Si devices fabricated by the same process was due to high base resistances, both extrinsic and intrinsic, despite high f_T's in the $Si_{1-x}Ge_x$ –base devices. Further improvement in design and technology resulted in the development of low-thermal-cycle SiGe-base ECL BiCMOS technology with f_T's of 50GHz, f_{max}'s of 59GHz and 19ps ECL gate delay (at $\sim$ 8mW) [23].

Gruhle et al. [24] reported the first high speed performance devices grown by MBE with f_T's of 46GHz and f_{max}'s of 53GHz. All of the layers were *in situ* doped. The processing of these devices was not compatible with VLSI Si technology since there were no high temperature steps at all and gold contacts were used. More recently, cutoff frequencies of MBE-grown HBT's with low base resistivities of over 100GHz have been demonstrated [2] and an f_{max} of 90GHz has been reported [25].

The high βV_A products of HBT devices make them interesting for analog applications. Fabricated in the same low-thermal-cycle process with UHV CVD grown bases, devices with SiGe-bases demonstrated f_T's of 113GHz and βV_A products of 48400V compared to f_T's of 73GHz and βV_A of 630 for Si-base devices [1]. Finally, with aggressive grading of the Ge profile in the base, scaled dimensions and reduced base resistance, ECL gate delays of 17ps have been achieved and the first MSI circuit, namely a 12-bit DA converter, has been realized in Si/$Si_{1-x}Ge_x$ technology operating at 1GHz, much faster than any existing Si DAC (Harame et al. [3]).

The Si/$Si_{1-x}Ge_x$ bipolar technology has made tremendous progress in the past years. While the early devices were fabricated with lower base doping levels, the doping was increased and processing optimized for high-speed performance devices, resulting in successful integration of HBT's with poly-emitter process. This chapter provides some of the basics necessary for understanding the advantages and limitations of Si/$Si_{1-x}Ge_x$/Si HBT's and addressed some of the challenges that the development of Si/$Si_{1-x}Ge_x$ technology has been facing. Although remarkable results have been achieved, several effects of heavy doping in strained $Si_{1-x}Ge_x$ still need to be studied. For example, the experimental data to establish the heavy-doping bandgap narrowing in $Si_{1-x}Ge_x$ is still insufficient. This thesis addresses some important issues related to base doping effects on DC device performance, such as tunneling in epitaxial p^+–n^+ junctions and hole mobility, minority carrier properties and bandgap narrowing in heavily doped base layers.

2.4 The Si/SiGe heterojunction

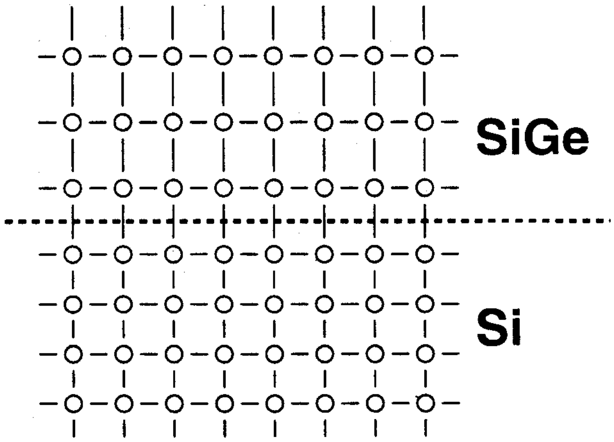
The Ge lattice constant (5.64Å) is larger than that of Si (5.43Å) resulting in a 4% lattice mismatch. This implies two possible ways of growth of $Si_{1-x}Ge_x$ alloy layers on a <100> silicon substrate to accommodate the mismatch. One is a relaxed alloy layer with a lattice constant interpolated between the Si and Ge values. Such growth results in incorporation of misfit dislocations at the Si/$Si_{1-x}Ge_x$ interface (Fig. 2.2.b), since the atoms in the Si substrate are not matched by atoms in the $Si_{1-x}Ge_x$ layer. Such growth is not desired for high quality material. The interface defects could increase leakage currents in p-n junctions and lower minority carrier lifetimes, which are important for good bipolar device performance. It has been shown that the incorporation of misfit dislocations causes non-ideal base currents in Si/Si$_{1-x}$Ge$_x$/Si HBT's [14].

The other possibility is pseudomorphic growth without dislocations, resulting in formation of a strained layer (Fig. 2.2.a). Such a layer has a lattice constant in the growth plane matching that of the Si substrate, but this is compensated by a larger lattice constant in the growth direction.

For thin Si$_{1-x}$Ge$_x$ films on <100> Si substrates it is energetically favorable to remain coherently strained. However, if the film thickness is increased, the energy associated with strain increases and above certain thickness it becomes energetically favorable for the material to relax by formation of misfit dislocations. The thickness above which a strained film is likely to relax is called the critical thickness. There are several theories to calculate the critical thickness [26, 27, 28, 29, 30] assuming that the structure is in the equilibrium state of its lowest energy. Fig. 2.3 shows the most commonly used critical thickness for strained Si$_{1-x}$Ge$_x$ layers calculated by Mathews and Blakeslee [27]. This implies that only very thin defect-free Si$_{1-x}$Ge$_x$ layers can be epitaxially grown for device applications. However, dislocation-free strained layers of thicknesses above the critical thickness can be grown in a metastable state, where kinetic limitations prevent the structure from relaxing. These layers are temperature sensitive, since they will return to the equilibrium state with energy provided from the heat, by relaxing the strain and generating dislocations. This imposes a high temperature restriction to the processing of Si/Si$_{1-x}$Ge$_x$ devices.

Both silicon and germanium are indirect bandgap materials, with bandgaps at zero Kelvin of 1.17eV and 0.66eV, respectively. The conduction band minima in silicon are 6-fold degenerate and located along Δ axis. Valence band maxima are located at the center of the Brillouin zone Γ. Up to high germanium concentrations in the alloy layer ($x \sim 0.8$) the band minima in bulk unstrained Si$_{1-x}$Ge$_x$ alloy are

a) strained:



b) relaxed:

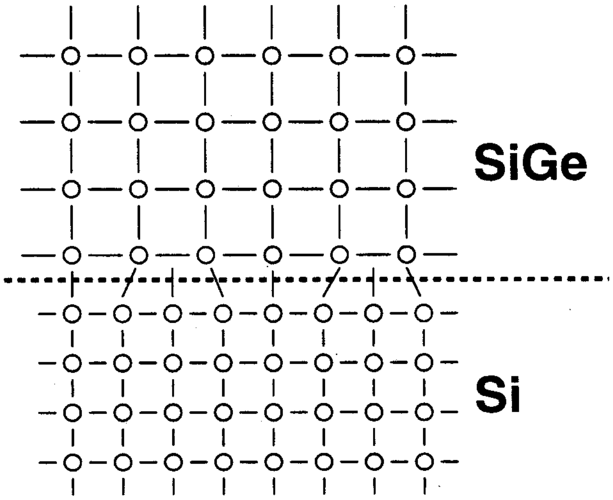


Figure 2.2: A comparison of strained (a) and relaxed (b) $Si_{1-x}Ge_x$ layers on a silicon <100> substrate

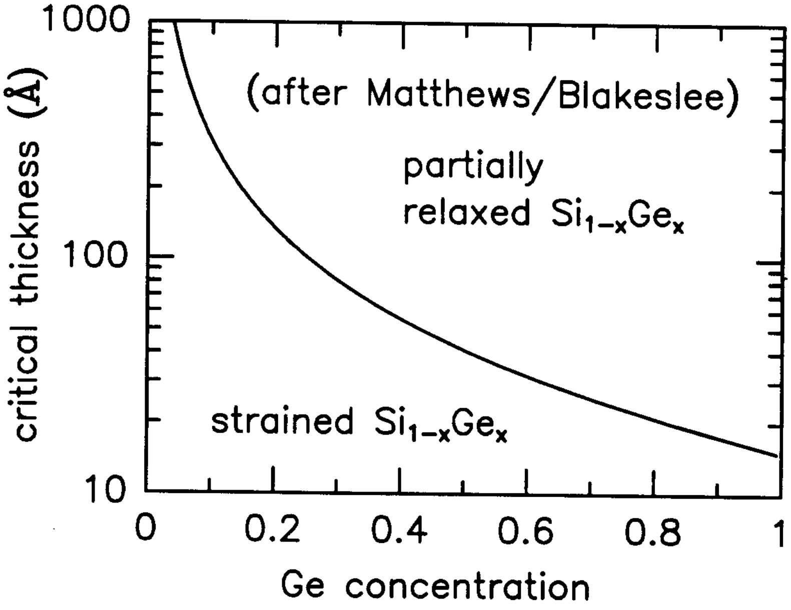


Figure 2.3: Equilibrium critical thickness of $Si_{1-x}Ge_x$ alloy layer on Si

silicon-like, with the bandgap decreasing as Ge content is increased [31]. The strain in $Si_{1-x}Ge_x$ alloy affects the band structure. When grown on <100> Si substrate the $Si_{1-x}Ge_x$ film is under biaxial compressive strain which lifts the degeneracy of conduction and valence bands. Four conduction band minima in the growth plane move down with respect to the other two in the growth direction. In the valence band, the degenerate heavy and light hole bands split with the heavy hole band lying higher. The resulting bandgap is lower than that of an unstrained material [32, 33, 34] and decreases much more rapidly with increasing Ge content. The bandgaps of strained (People et al. [32]) and unstrained (Braunstein et al. [31]) random $Si_{1-x}Ge_x$ alloys as a function of Ge concentration are shown in Fig. 2.4.

The band lineup at $Si/Si_{1-x}Ge_x$ interface is shown in Fig. 2.5. Up to Ge concentrations of ~ 60%, a Type-I band alignment has been calculated [33, 35] and experimentally confirmed [36]. The conduction band offset is small while most of the bandgap offset occurs in the valence band. From the perspective of SiGe HBT's this

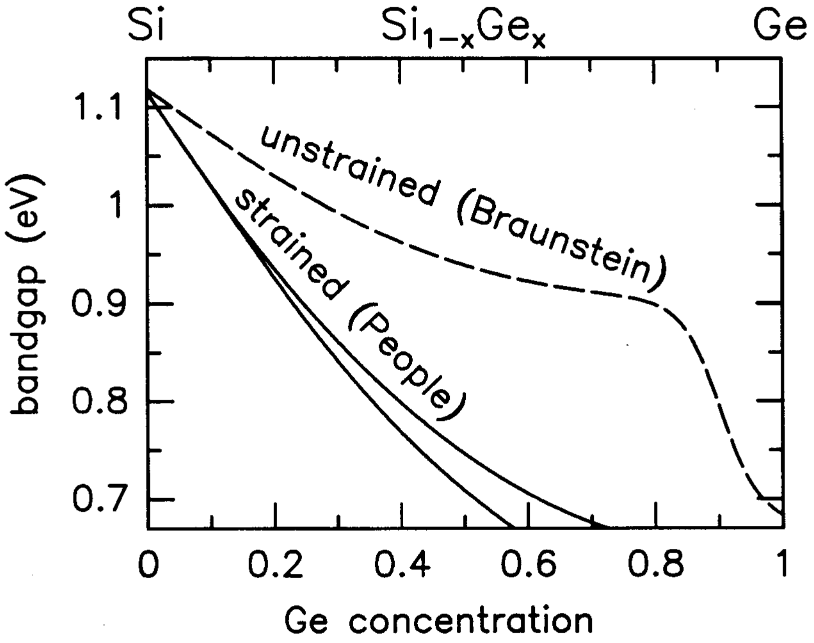


Figure 2.4: Bandgaps of strained and unstrained $Si_{1-x}Ge_x$ alloys vs. Ge concentration

is ideal for npn structures but not for pnp devices, where germanium grading at the interface is necessary to avoid "spike and notch" effects [37].

2.5 Speed of Si/SiGe heterojunction bipolar devices and circuits

One of the most significant advantages of $Si/Si_{1-x}Ge_x/Si$ HBT's over Si bipolar devices is speed. This section outlines the basic issues important for understanding the high frequency performance of bipolar devices and circuits and explains some of the limitations of Si BJT's and possible improvements via HBT structures.

A typical parameter often used to characterize the speed of bipolar transistors is the cutoff frequency f_T. At high frequencies the current gain of a bipolar transistor will decrease. The cutoff frequency f_T is the frequency at which current gain reaches

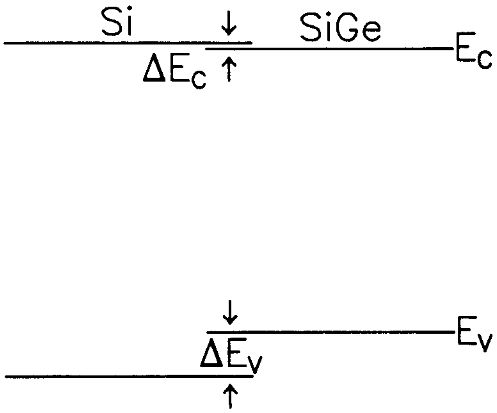


Figure 2.5: Band lineup of strained $Si_{0.8}Ge_{0.2}$ layers on <100> Si-substrate, $\Delta E_C \sim$ 20meV, $\Delta E_V \sim$ 150meV

unity. It is related to the transit time that it takes electrons from the emitter to reach the collector [38]:

$$\frac{1}{2\pi f_T} = \tau_{EC} = \tau_E + \tau_B + \tau_d + \tau_C \tag{2.1}$$

The transit time τ_{EC} is the sum of the transit times across the emitter τ_E, the base τ_B, the base-collector depletion region τ_d, and collector τ_C, and it reflects the vertical device profile. The emitter transit time consists of the emitter storage time and the time associated with charging up base junction capacitances. The emitter storage time is the delay due to hole charge storage in the emitter, and it is usually negligible in heavily doped emitters of high-gain bipolar devices. The main source of emitter delay is represented by the RC constant:

$$\tau_E = \frac{k_B T}{q I_E}(C_{BE} + C_{BC}) \tag{2.2}$$

Since it is inversely proportional to the DC emitter current, at high enough current levels τ_E is not a significant component of the overall delay in a bipolar transistor. The base transit time τ_B is the delay related to electron transport through the neutral base. For devices with flat doping profiles in the base the electrons move solely by diffusion and τ_B only depends on the width of the base region w_B and diffusion coefficient of the minority carriers in the base D_n:

$$\tau_B = \frac{w_B^2}{2D_n} \tag{2.3}$$

The delay due to the transport across the base-collector space charge region τ_d depends on the width of the depletion region x_{BC}. Assuming that the carriers move at saturation velocity v_{sat}, it is given by:

$$\tau_d = \frac{x_{BC}}{2v_{sat}} \tag{2.4}$$

Higher collector doping is desired for narrower depletion widths, i.e. lower transit times, and operation at high current levels without deleterious space charge effects in the base-collector depletion region. However, collector doping is limited by base-collector junction breakdown constraints. The delay associated with the collector RC constant τ_C is negligible in today's bipolar transistors due to the low resistivity of the buried n^+ layer below the collector.

A significant part of τ_{EC} comes from the base transit time τ_B. Typical basewidths in present-day silicon devices exceed 1000Å, and are limited by punchthrough requirements and inability to accurately control the junction depths of implanted and diffused junctions. The typical basewidth of an epitaxial $Si_{1-x}Ge_x$ base is a few hundred Å, and it can be reduced to below ~200Å for high f_T performance [2]. In standard Si devices, an exponential base doping profile with the highest base doping at the base-emitter junction creates a built-in drift field that speeds up the electrons across the base and reduces the base transit time. In such a case minority carriers in the base move both by drift and diffusion, and depending on the magnitude of the built-in field, the factor of two in Eqn. 2.3 should be replaced by η where $\eta > 2$. The same effect in $Si/Si_{1-x}Ge_x/Si$ HBT's can be accomplished by changing the bandgap in the base via a graded Ge profile. If the bandgap is wider, i.e. the Ge concentration is lower, at the emitter side of the base than that at the collector side, an electric field will be created to reduce the base transit time [39]. With narrow epitaxial bases and carefully designed Ge and doping profiles, a tremendous improvement in f_T of $Si/Si_{1-x}Ge_x$ heterojunction devices can be achieved compared to Si devices. $Si/Si_{1-x}Ge_x/Si$ HBT's with f_T's of over 100GHz have been demonstrated recently [1, 2].

A figure of merit often used to describe the performance of microwave transistors is the maximum frequency of oscillation f_{max}. This is the frequency at which the power gain decreases to unity. It reflects not only the intrinsic device profile but also some parasitics, such as base resistance R_B and base-collector capacitance C_{BC}:

$$f_{max} = \sqrt{\frac{f_T}{8\pi R_B C_{BC}}} \tag{2.5}$$

The base-collector capacitance is usually determined by the junction area. The parasitic base resistance consists of an intrinsic and an extrinsic part. The extrinsic base resistance is a parasitic resistance from the base terminal to the active base region and it can be made relatively small ($\sim 10\Omega$ [40]) in state-of-the-art self-aligned bipolar processes. The intrinsic base resistance is the resistance of the active base region and it is determined by the doping profile in the base:

$$R_{B,sh} = \frac{1}{\int_{base} q\mu_p(x)N_A(x)\,dx} \tag{2.6}$$

where μ_p is the in-plane hole mobility in the base. The typical doping at the base emitter junction of Si BJT's is $\sim 10^{18}cm^{-3}$ and it decreases exponentially towards the collector. This resulting base sheet resistivities are of the order of 10kΩ/sq. This is usually a limiting factor for high f_{max} and it is one of the key issues that can be solved by introducing $Si_{1-x}Ge_x$ into the base. In $Si/Si_{1-x}Ge_x/Si$ HBT's, base doping can easily be increased by an order of magnitude while high gain is still maintained. The collector current density of a bipolar transistor with constant doping in the base is given by:

$$J_C = \frac{qD_n\, n_{i,base}^2}{N_A w_B}\, e^{qV_{BE}/k_BT} \tag{2.7}$$

Assuming, to first order, the same minority carrier diffusion coefficients and densities of states in Si and $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ bases one can write:

$$\frac{J_C(\mathrm{SiGe})}{J_C(\mathrm{Si})} = \frac{\beta(\mathrm{SiGe})}{\beta(\mathrm{Si})} = \frac{(N_A w_B)_{\mathrm{Si}}}{(N_A w_B)_{\mathrm{SiGe}}} \, e^{\frac{q\Delta E_G}{k_B T}} \tag{2.8}$$

While high gain can be achieved by a bandgap offset, ΔE_G, the base doping in HBT's can be vastly increased to reduce base sheet resistance and hence, improve f_{max} and high speed circuit performance. Furthermore, the scaling of lateral and vertical dimensions in bipolar technology requires increased base doping to avoid punchthrough, which can be compensated by the $\mathrm{Si}/\mathrm{Si}_{1-x}\mathrm{Ge}_x$ bandgap offset in HBT devices. The tradeoff between the gain and base sheet resistance in $\mathrm{Si}/\mathrm{Si}_{1-x}\mathrm{Ge}_x/\mathrm{Si}$ HBT's will be addressed in more detail in chapter 5.

Although f_T and f_{max} can fairly well characterize the high speed performance of bipolar devices and, to some extent, bipolar circuits, they are insufficient to accurately predict the maximum speed of operation of a bipolar circuit. The gate delay in bipolar digital circuits not only depends on device parameters but also on the current level and load capacitances. Several expressions have been developed to model gate delay in bipolar digital circuits [41, 42]. However, one can still conclude that small intrinsic device delay (high f_T), low parasitic base resistances, low parasitic capacitances, and optimized loads are prerequisites for high speed circuit performance.

2.5.1 Improvement of ECL gate delay via HBT structure - simulation

This section illustrates possible advantages in high speed circuit performance of $\mathrm{Si}/\mathrm{Si}_{1-x}\mathrm{Ge}_x/\mathrm{Si}$ HBT's over Si BJT's by giving an example of how the ECL (emitter-coupled logic) gate delay of an existing bipolar process could be improved by introducing $\mathrm{Si}/\mathrm{Si}_{1-x}\mathrm{Ge}_x$ heterojunction devices. ECL is one of the most commonly used digital circuit families in bipolar technology. The simulations of ECL ring oscillators described in this section are performed using the SPICE circuit simulator. They are based on the BEST II ("BiCMOS-compatible ECL Super-self-aligned Technology") technology developed at AT&T [40, 43, 44, 45]. In a self-aligned process, the base and emitter contacts are processed in the same photolitographic step, and the distance between them is determined by a thin SiO_2 spacer, thus minimizing the extrinsic base resistance. The main advantages of the BEST process are small contact resistances achieved by self-aligned silicides and extremely small extrinsic base resistance (14Ω). The base sheet resistivity of 10kΩ/sq. is limited by relatively low base doping for gain requirement and the basewidth is determined by diffusion profiles of boron and arsenic. Since the base sheet resistance could be reduced in an HBT structure while maintaining sufficient gain, the small extrinsic base resistances make this technology an interesting candidate for high speed performance via an HBT structure [44].

Fig. 2.6 shows a schematic diagram of an ECL gate fabricated in the BEST technology and used in the simulations. The starting point for the simulations was the current process with the measured SPICE parameters: bf = 140 (gain), tf = 6ps (forward transit time), re = 17Ω (emitter resistance), rb = 600Ω (base resistance, includes intrinsic and extrinsic base, $rb = R_{B,ex} + 0.125 \times R_{B,sh} \times W_e/L_e$), cje = 7fF (base-emitter zero-bias junction capacitance), cjc = 3.5fF (base-collector zero-bias junction capacitance), cjs = 15fF (collector-substrate zero-bias junction capacitance). The emitter area was $1 \times 2\mu m^2$. The operation at a low power level (~ 2mW/gate) was simulated using resistances in the circuit (R in Fig. 2.6) of 1.66kΩ, and for a high power level (~ 8mW/gate) the resistances of 500Ω were used [44]. These parameters resulted in the delay of ~ 40ps at low power levels. Fig. 2.7 shows how the gate delay is decreased as a function of intrinsic base sheet resistance for devices with various f_T's (related to forward transit times in SPICE ~ 1/tf). Note that the cutoff frequency in these simulations, defined as 1/tf, does not include emitter delays. The extrinsic base resistance of 14Ω was assumed unaffected by the process modifications needed to incorporate $Si_{1-x}Ge_x$ into the base. It is obvious from the

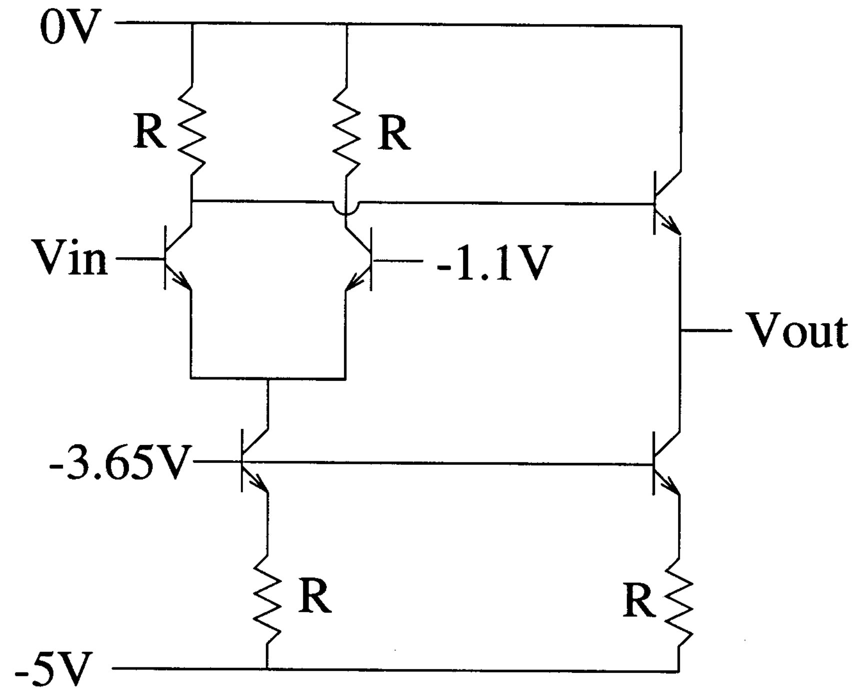

0V
R
R
Vin
-1.1V
Vout
-3.65V
R
R
-5V

Figure 2.6: Schematic diagram of ECL gate used in SPICE simulations, R = 1.66kΩ for low power simulation and R = 500Ω for high power simulation.

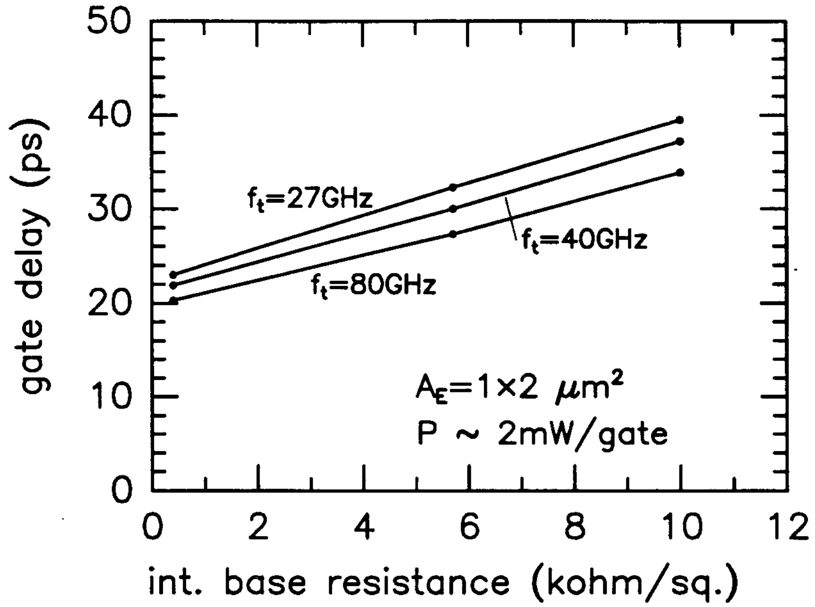


Figure 2.7: Simulated gate delay as a function of intrinsic base sheet resistance for various f$_T$'s, A$_E$ =1 × 2μm^2, P ~2mW/gate

figure that at this power level the main source of the delay lies in the base resistance. Without changing the cutoff frequency of individual devices (~ 27GHz), by reducing the base sheet resistance from 10kΩ/sq. down to 400Ω/sq. the gate delay can be reduced from 40 down to 23ps, almost a factor of two. It is also shown in the figure that even if f$_T$ is increased to 80GHz, if the base sheet resistance is not reduced the delay will only improve to 34ps. So it is the base resistance, not f$_T$, that is critical for high speed performance. Fig. 2.8 is similar, except that base-emitter zero-bias junction capacitance is varied as a parameter. Since the reduction of base resistance requires an increase in the base doping, narrower depletion regions at base-emitter junction could cause an increase in the capacitance. Base-collector capacitance is not that likely to increase since it is dominated by a lighter doped collector side of the junction which is not changed by increased base doping. It is shown in the figure that, for example, if the base sheet resistance is reduced by a factor of two to 5kΩ/sq. and

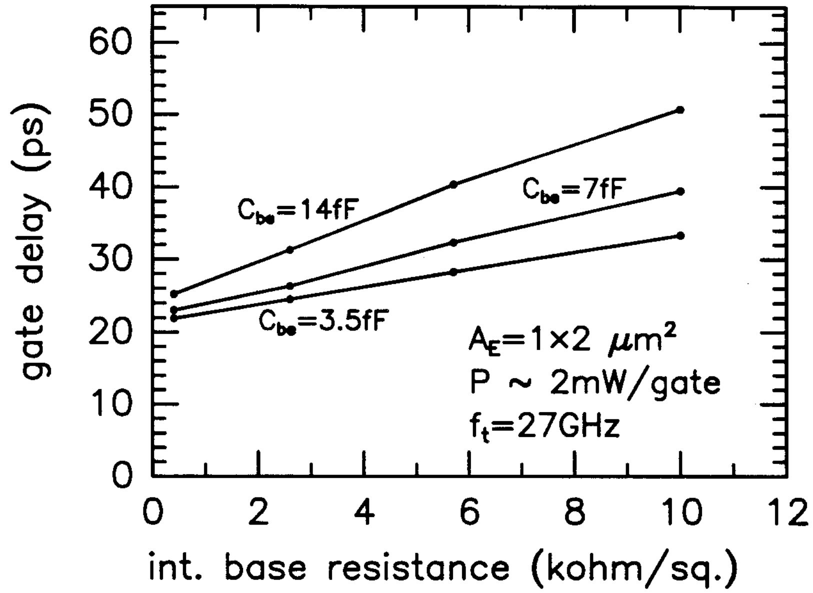


Figure 2.8: Simulated gate delay as a function of intrinsic base sheet resistance for various C_{BEO}'s, $A_E = 1 \times 2\mu m^2$, P ~2mW/gate

base-emitter capacitance is doubled the gate delay is hardly improved at all, while an improvement of ~10ps is possible if the capacitance is not increased. The importance of base resistance on the ECL gate delay is emphasized in figures 2.9 and 2.10 where gate delay is plotted as a function of intrinsic base resistance and f_T, respectively, at higher power level (8mW/gate). For BEST II technology, at this higher power level the gate delay is still limited by base sheet resistance. An improvement from 27GHz to 64GHz in f_T reduces the gate delay from 37ps down to 28ps if the base resistance is not reduced, similar to the 27ps delay achieved if the base sheet resistance is only reduced to 5.7kΩ/sq. Further reduction in base resistance lowers the delay to below 15ps. For example, a gate delay of ~ 13ps could be achieved at the power level of 8mW/gate with f_T's of 35GHz and base sheet resistancies of 400Ω/sq. A cutoff frequency of 35GHz with base resistance of 400Ω/sq. could be easily obtained in $Si_{1-x}Ge_x$ technology, with a basewidth of 500Å, base doping of 5×10^{19} cm^{-3} and

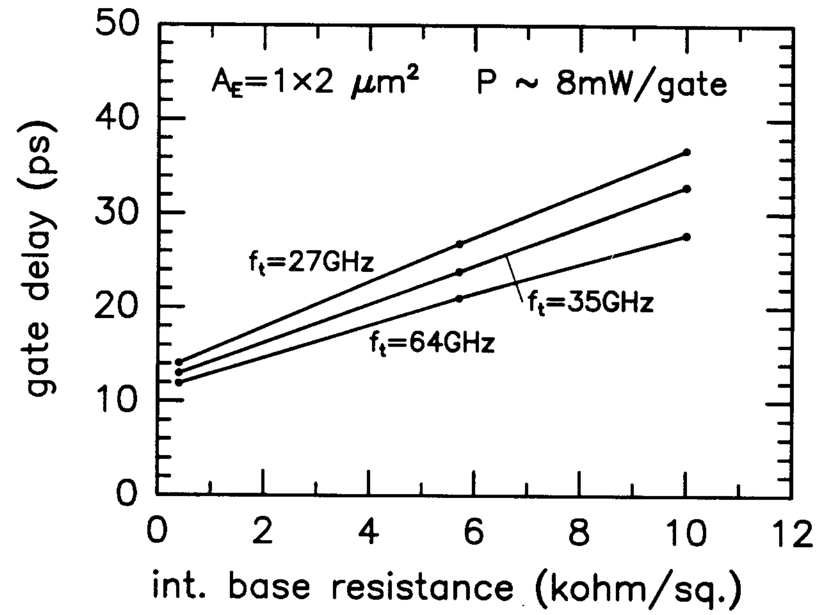


Figure 2.9: Simulated gate delay as a function of intrinsic base sheet resistance for various f_T 's, A_E =1 × 2μm^2, P ~8mW/gate

Ge concentration graded between 12% and 18%. For comparison, a typical basewidth in the BEST technology is ~ 1500Å.

Another advantage of Si/SiGe/Si HBT's over Si BJT's that needs to be mentioned is high output impedance, i.e. Early voltage (V_A), especially important in analog applications. A figure of merit often evaluated is βV_A product, which can be vastly improved with narrower gap material at the base-collector junction, as is the case in HBT's [46].

This section has provided some understanding the limitations for high speed bipolar devices and circuits and addressed what improvements could be achieved by integrating Si/Si$_{1-x}$Ge$_x$ heterojunction into Si bipolar technology. Although largely compatible with Si processing technology, the introduction of Si$_{1-x}$Ge$_x$ HBT's increases the process complexity to some extent and there are several issues that need to be addressed.

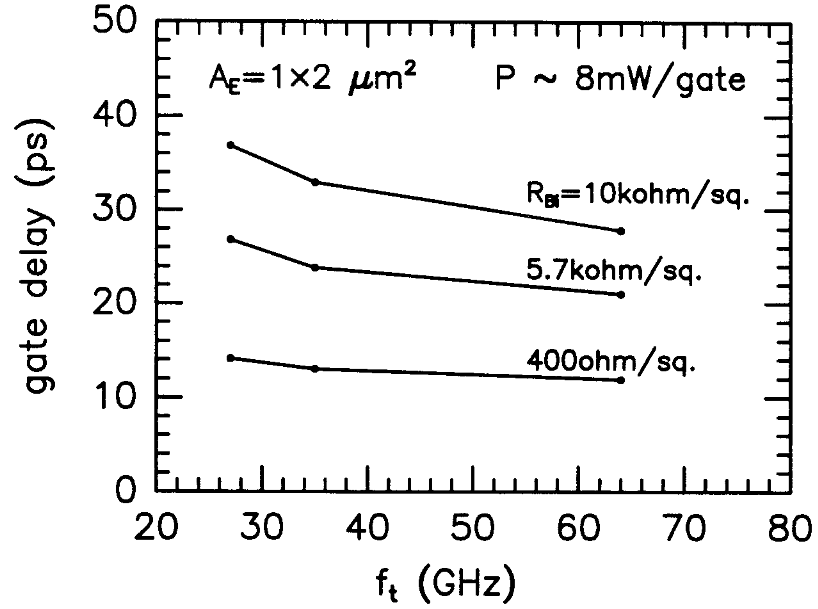


Figure 2.10: Simulated gate delay as a function of cutoff frequency for various $R_{B,sheet}$, $A_E = 1 \times 2\mu m^2$, P ~8mW/gate

2.6 Process integration

The selective growth of high quality epitaxial $Si_{1-x}Ge_x$ layers requires sophisticated growth techniques such as CVD. Since defects could be a source of generation-recombination centers which may cause increased leakage or tunneling currents in p–n junctions and low minority carrier lifetimes, for high performance bipolar devices abrupt and defect-free epitaxial $Si/Si_{1-x}Ge_x$ interfaces are desired. The design of narrow-base high speed device relies on the possibility to precisely control germanium and doping profiles. A detailed study of interface abruptness by Rapid Thermal Chemical Vapor Deposition (RTCVD) is given in chapter 3. Interface widths below 5Å are achievable in present-day CVD reactors.

Defect-free material can be grown only if $Si_{1-x}Ge_x$ films are strained. The thickness of strained $Si_{1-x}Ge_x$ is limited by the critical thickness, as explained in section

2.4. Usually, layers above the critical thickness are grown in the metastable state. Such material is extremely temperature sensitive since a high temperature step could cause strain relaxation and the creation of misfit dislocations. The typical Si process requires high temperature steps for implant anneals, for example, in AT&T's BEST technology a simultaneous base and emitter polysilicon implant anneal requires a temperature of 950°C for one hour. Any high temperature step is potentially undesirable for the stability of $Si_{1-x}Ge_x$ layers.

One of the biggest advantages of $Si/Si_{1-x}Ge_x/Si$ HBT's is that the base doping could be increased (typically $\sim 10^{19}cm^{-3}$) to reduce base resistance while maintaining the high gain. But high base doping near the emitter will increase the base-emitter junction capacitance, which is not desired for high speed performance. The emitters of Si BJT are usually heavily doped for high gain. Heavy doping at both sides of base-emitter junction could cause lower breakdown voltage, higher leakage currents and significant tunneling currents at the junction. The tunneling current in p^+–n^+ epitaxial junctions is studied in detail in chapter 4. These issues impose a constraint on base and emitter doping near the junction.

The boron profile in an epitaxial $Si_{1-x}Ge_x$ base is usually achieved by *in situ* doping (during growth). The targeted profile is often box-like. Since boron diffuses faster than germanium, during any subsequent high temperature steps the boron profile will be degraded resulting in exponential doping tails of B outside the $Si_{1-x}Ge_x$ on both sides. The doping tail at the emitter side of the base will create a retrograde electric field, increasing the base transit time. One needs to take this into account when designing the Ge profile in the base [47], so that the retrograde doping profile is compensated by Ge grading. Furthermore, if the doping tail extends outside the $Si_{1-x}Ge_x$ base layer, parasitic barriers will be formed that degrade both DC [48, 49] and high frequency transistor performance [18, 50]. Prinz et al. [49] have proposed the introduction of undoped $Si_{1-x}Ge_x$ spacer layers in the base, both at base-emitter and base-collector junctions to prevent boron outdiffusion. The thicknesses of such spacer layers should be carefully designed and optimized for subsequent processing. Thinner spacers are desired to keep the total basewidth narrow and base sheet resistance as low as possible while they have to be sufficiently thick to accommodate boron diffusion in the subsequent processing steps. Spacers which are too thick could severely damage high speed performance if the total basewidth or the base resistance are increased. Furthermore, if the p-n junction is formed within the $Si_{1-x}Ge_x$ layer, the emitter storage time becomes no longer negligible [47]. So it is important that the p–n junction coincides with the $Si/Si_{1-x}Ge_x$ interface. A carefully designed spacer layer at the emitter side of the base limits the base doping at base-emitter junction and also solves the problem of base-emitter capacitance and tunneling at the base-emitter junction [51]. The following section deals with the boron outdiffusion problem in more detail.

2.6.1 Boron outdiffusion from SiGe base

The effects of boron outdiffusion are shown in Fig. 2.11. Solid lines in figures 2.11.a and 2.11.b show as-grown doping profiles and the corresponding zero-bias band diagrams [1]. On top of the as-grown profile an anneal of 30 minutes at 800°C is simulated by SUPREM, a process simulation program. The diffusion coefficients for $Si_{1-x}Ge_x$ were assumed the same as in Si. The annealed profile and corresponding band diagram are shown as dashed lines in figures 2.11.a and b. The p–n junctions have moved outside the $Si_{1-x}Ge_x$ base layer resulting in an increased basewidth and conduction band barriers in the corresponding band diagram. These barriers degrade the gain of an HBT and limit high frequency performance, as discussed in the previous section.

---

[1] Band diagrams in this thesis were simulated by SEDAN, a 1D device simulator [52] modified by E.J. Prinz to incorporate $Si/Si_{1-x}Ge_x$ heterojunctions using the values of conduction and valence-band discontinuities calculated by People and Bean [33]

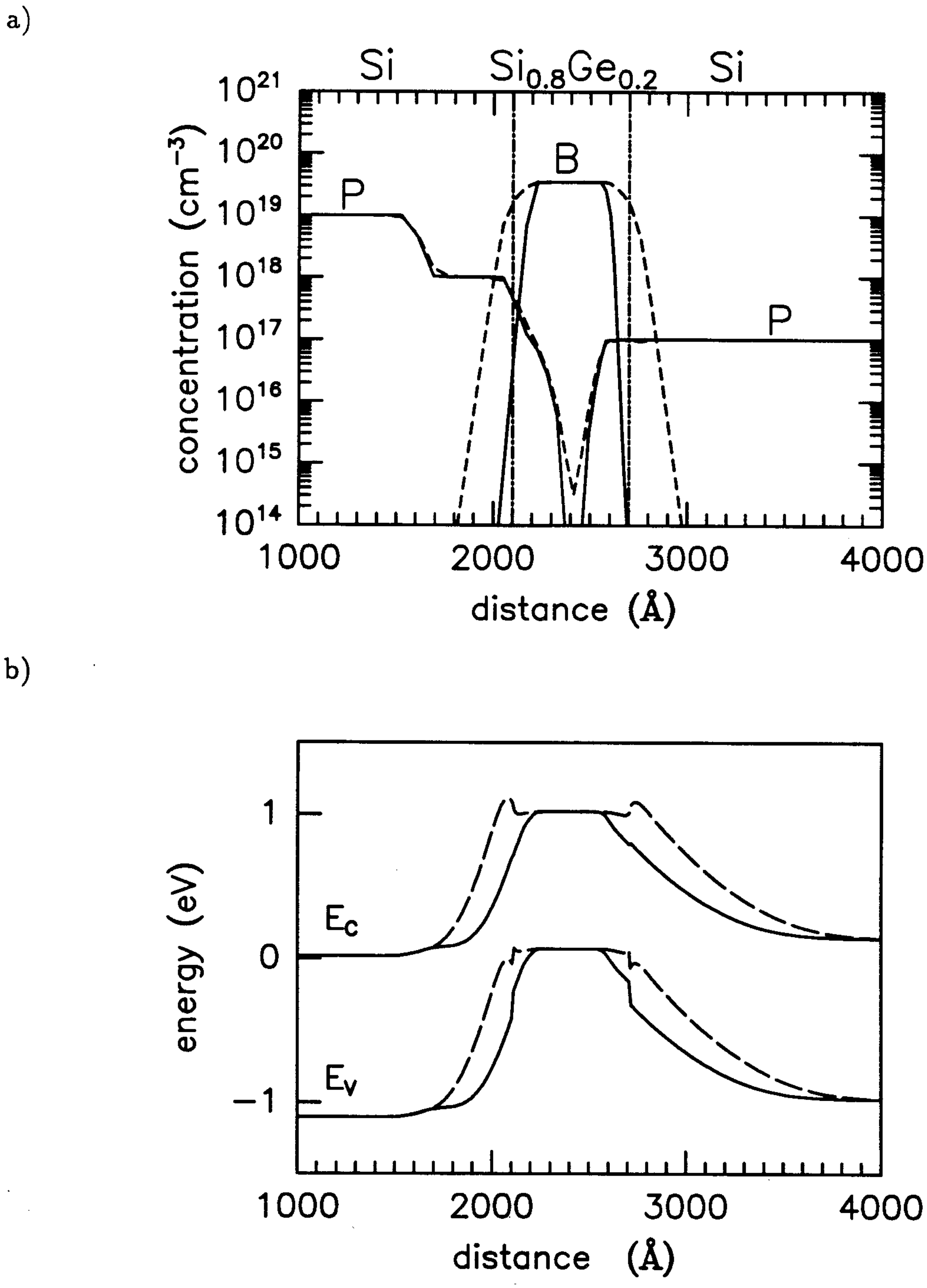


Figure 2.11: a) Simulated profiles (SUPREM) of an HBT structure as-grown (solid) and after a 30 minutes anneal at 800°C (dashed). b) Simulated zero-bias band diagrams (SEDAN) of the profiles of (a). Note the parasitic barriers formed when B outdiffuses from the $Si_{1-x}Ge_x$ base (dashed).

To optimize the thermal cycle for Si/Si$_{1-x}$Ge$_x$/Si HBT processing, we have studied the boron outdiffusion effects both experimentally and by SUPREM simulation, on test structures designed for integration into BEST technology (described in the previous section). The effect of boron outdiffusion is actually somewhat less severe in Si/Si$_{1-x}$Ge$_x$ heterojunctions than predicted by SUPREM simulations with Si diffusion coefficients. This is due to slower diffusion in strained Si$_{1-x}$Ge$_x$ than in Si [53, 54] and boron segregation at Si/Si$_{1-x}$Ge$_x$ interfaces. Further, it has been shown, both theoretically and experimentally that acceptor impurities tend to segregate within Si$_{1-x}$Ge$_x$ layer while donor impurities segregate into Si [55, 56]. Although overestimating the effect, SUPREM simulations are still a useful tool to establish an upper limit to the Si$_{1-x}$Ge$_x$ processing thermal budget.

Fig. 2.12.a shows a simulation of an as-grown profile of a device with measured base sheet resistance of 300Ω/sq. and basewidth of 650Å. The Ge in the base was graded from 12% at the base-emitter junction to 18% within the half of the basewidth to compensate for the retrograde B-profile and obtain the drift field to reduce electron transit time across the base. The 150Å thick spacers within the base were designed to accommodate boron diffusion during an 800°C, 10 minute long anneal. Although the standard BEST technology requires a longer (1 hour) anneal at much higher temperature (950°C) through a series of modifications and implant–and–anneal experiments, it has been shown that the desired profile in the n$^+$poly-emitter could be achieved with an anneal as short as 10 minutes at 800°C using P–implant into polysilicon. The initial base profile (B and Ge) of the same device measured by SIMS is shown in Fig. 2.13. Fig. 2.12.b shows the simulated annealed profile. Boron is still contained within Si$_{1-x}$Ge$_x$ and p–n junctions coincide with the Si/Si$_{1-x}$Ge$_x$ interfaces. Fig. 2.12.c shows simulated profile when the annealing time is increased to 40 minutes. The boron profile clearly extends into Si layers. If the temperature is increased, the effect is even worse, as shown in Fig. 2.12.d. At the temperature of 900°C for only ten minutes the basewidth is almost tripled and p−n junctions are moved far into Si layers.

Similar results have been observed experimentally. Fig. 2.14 shows the collector current of the same device as a function of base-emitter bias (Gummel plot) after various annealing conditions. The effect of the parasitic barriers is obvious as degradation of collector current. Though always ideal (60mV/decade) as expected for a bipolar transistor, the collector current decreases with increased annealing times or for increased annealing temperatures.

We can conclude that although integration of $\mathrm{Si/Si_{1-x}Ge_x/Si}$ HBT's imposes a constraint on thermal processing, the effect of boron outdiffusion could be succefully overcome, even for very heavily doped devices. We have established an upper limit to a thermal cycle of 800°C for 10 minutes before the degradation of HBT performance starts to occur. The same thermal cycle was shown to be sufficient to obtain the desired diffused poly-emitter profile.

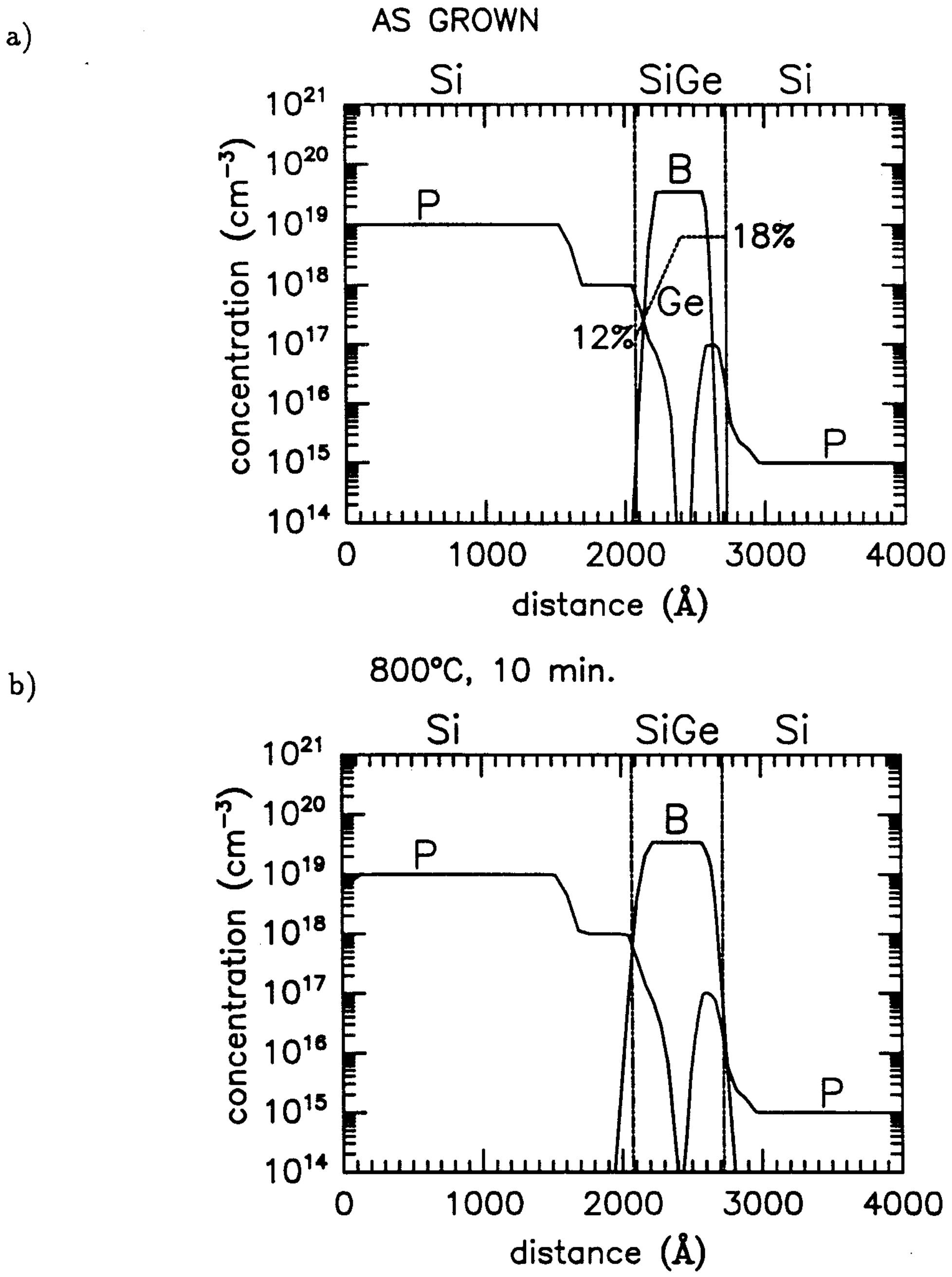
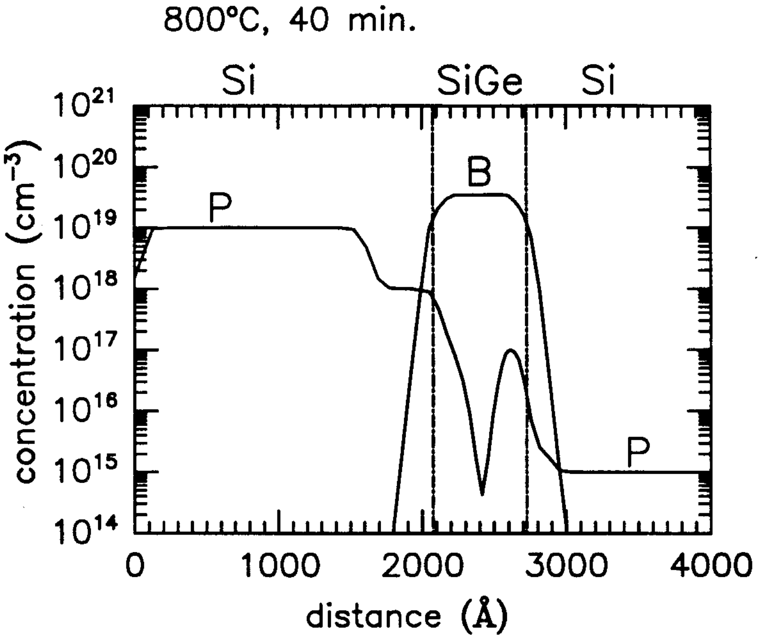


Figure 2.12: SUPREM simulation of Si/Si$_{1-x}$Ge$_x$/Si HBT doping profiles: as grown (a) and at various annealing conditions (b,c,d). The Ge profile is also shown in (a)

c)



d)

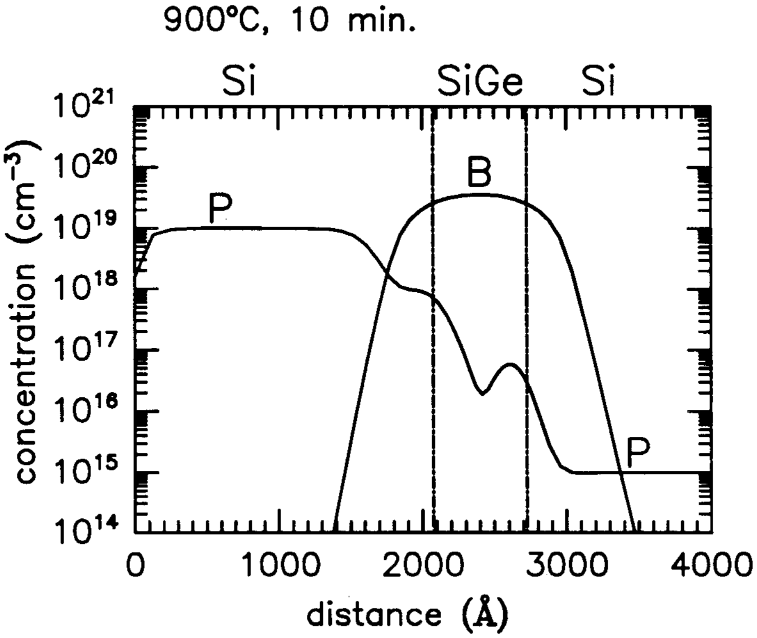


Figure 2.12: SUPREM simulation of $Si/Si_{1-x}Ge_x/Si$ HBT doping profiles: as grown (a) and at various annealing conditions (b,c,d). The Ge profile is also shown in (a)

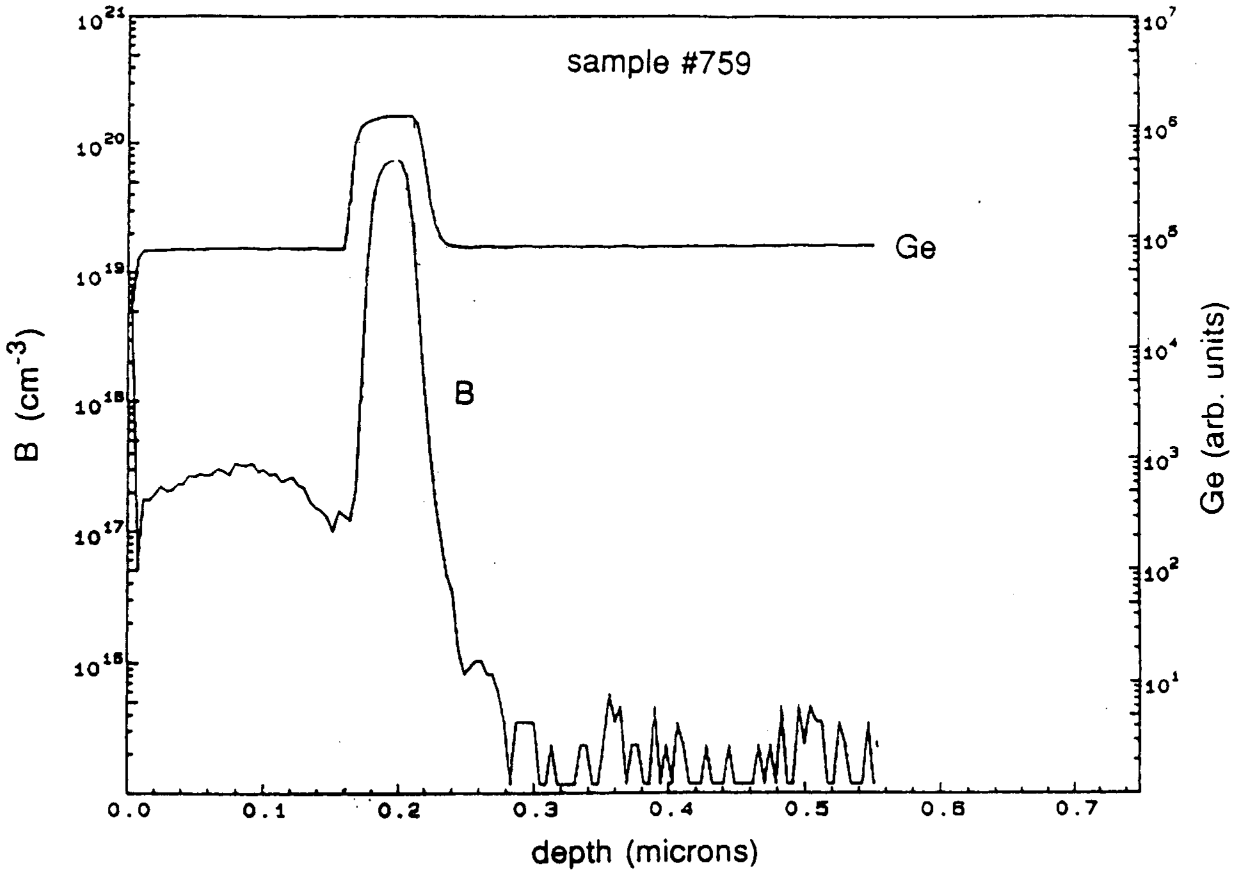


Figure 2.13: As grown SIMS profile of the device used for annealing experiments (#759)

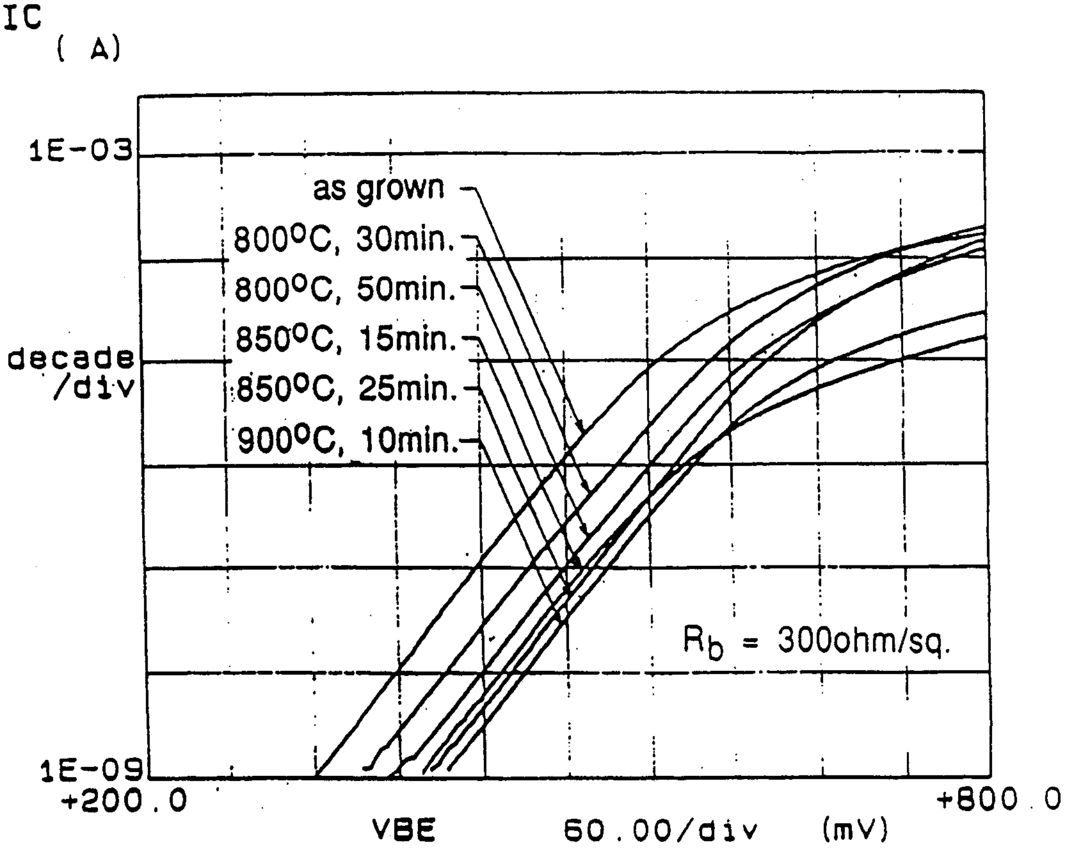


Figure 2.14: Measured collector current for various annealing conditions of a device with 350Å thick 8×10^{19} cm^{-3} doped base with 150Å thick BC and BE $Si_{1-x}Ge_x$ spacers (#759)

Chapter 3

RTCVD Growth and Interface Quality of Si/SiGe Heterojunctions

3.1 Introduction

The growth of thin crystalline Si and $Si_{1-x}Ge_x$ layers for HBT applications by low-temperature chemical vapor deposition (CVD) has been extensively studied in recent years, and high quality material and electronic devices made in these films have been demonstrated. Various growth techniques have been used to produce high quality material. These different techniques operate in a wide range of growth pressures, from mtorr to atmospheric pressure. In Ultra-High-Vacuum CVD (UHV/CVD) reactors, the growth pressure is measured in mtorr range [57, 58], Limited Reaction Processing (LRP) [15, 59] and Rapid Thermal CVD (RTCVD) [60] usually operate in the 1-10 torr range, while Atmospheric Pressure CVD (APCVD) [61, 62] works at atmospheric pressure. The first $Si/Si_{1-x}Ge_x$ HBT's demonstrated were grown by LRP technique [13], while the most recent high speed devices were grown by UHV/CVD [1]. Typically, dichlorosilane (SiH_2Cl_2) or silane (SiH_4) is used as Si source and germane (GeH_4) as Ge source. Abrupt germanium and dopant transitions and high quality defect-free interfaces are critical for the desired performance of $Si/Si_{1-x}Ge_x/Si$ HBT's. However, there have been no systematic studies reported so far relating the growth pressure to the material characteristics and interface quality.

In this chapter the RTCVD system at Princeton is described and the effects of growth pressure on the material quality and interfacial properties of $Si_{1-x}Ge_x$ layers are discussed. The material has been characterized by photoluminescence (PL), X-ray reflectivity (XRR), X-ray diffraction (XRD), and the performance of resonant tunneling diodes (RTD's). An upper limit to interface roughness of below 4Å is established.

3.2 Rapid Thermal Chemical Vapor Deposition (RTCVD)

All samples used for experiments in this thesis were grown in a single Rapid Thermal Chemical Vapor Deposition (RTCVD) system [60]. The RTCVD reactor is schematically shown in Fig. 3.1.

In this system, a single 100-mm silicon wafer is suspended on quartz pins inside a 175mm-diameter quartz tube. A load-lock chamber is separated from the main chamber by a gate valve. During epitaxial growth, the wafer is heated by a bank of twelve 6kW tungsten-halogen lamps. This enables rapid changing and optimizing the wafer temperature for each epitaxial layer. Accurate knowledge of the wafer temperature is crucial since the growth rate of silicon and $Si_{1-x}Ge_x$ alloy layers is a strong function of temperature. The wafer temperature is determined accurately in the range between 550°C and 800°C by measuring infrared light transmission at $\lambda = 1.3\ \mu m$ and $1.5\ \mu m$ through the wafer [63].

The gases used for growing $Si_{1-x}Ge_x$ layers are dichlorosilane ($SiCl_2H_2$), silane (SiH_4), and germane (0.8% GeH_4 in H_2). As dopant sources, diborane (10 ppm B_2H_6 in H_2) and phosphine (70 ppm PH_3 in H_2) are used. All layers are grown with a hydrogen carrier flow of 3 slpm. The hydrogen is purified by diffusion through a palladium cell, and the dichlorosilane is purified by a Nanochem™ cell to remove water vapor and oxygen.

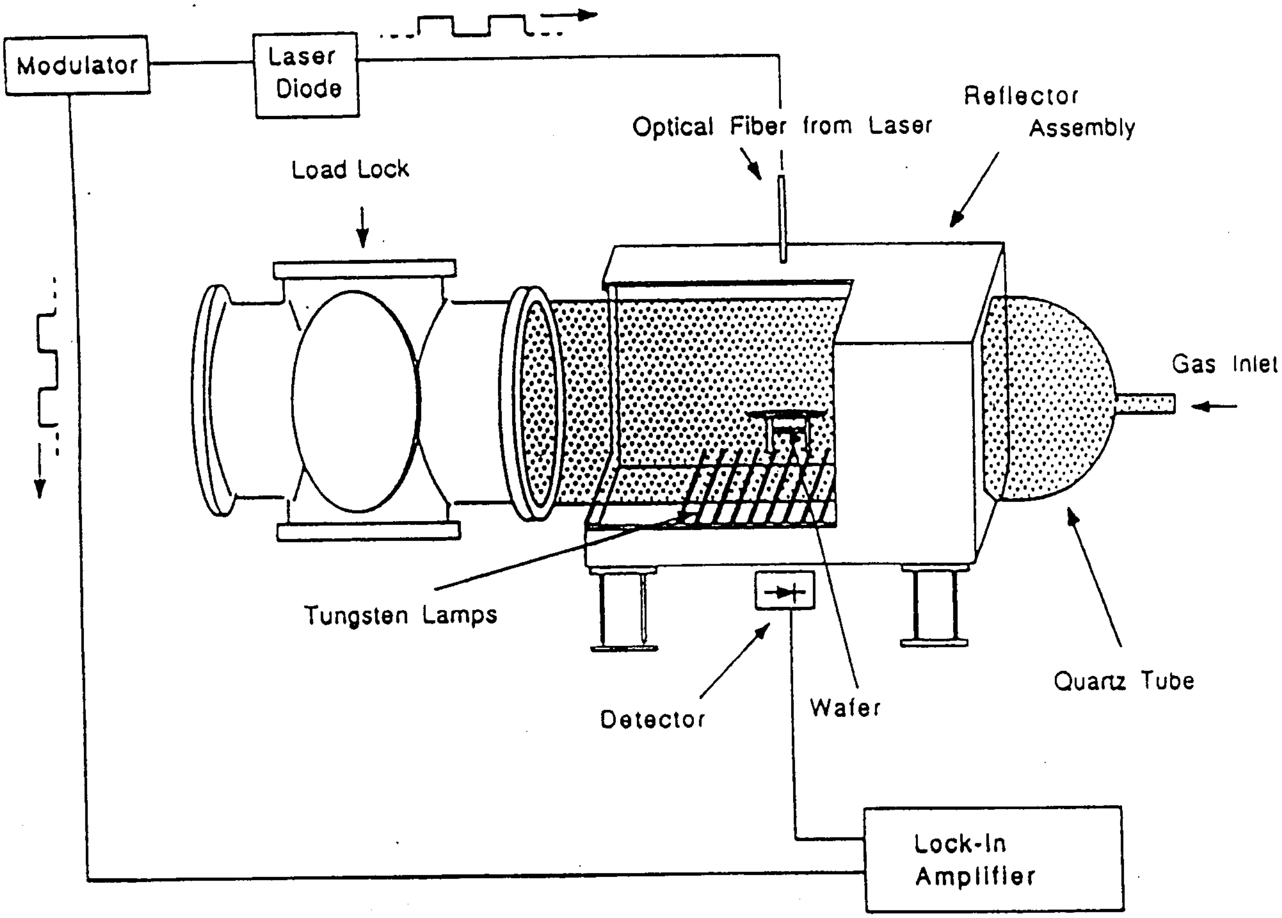

Modulator
Laser Diode
Optical Fiber from Laser
Reflector Assembly
Load Lock
Gas Inlet
Tungsten Lamps
Detector
Wafer
Quartz Tube
Lock-In Amplifier

Figure 3.1: Schematic diagram of the RTCVD reactor used in ths work

Before growing the epitaxial Si or $Si_{1-x}Ge_x$ layers, Si wafers are cleaned in a hot 1:1 solution of $H_2SO_4 : H_2O_2$, rinsed in deionized water, and dipped in a 1:100 solution of HF : H_2O to remove the chemical oxide. The wafers are then blown dry in nitrogen and immediately loaded into the RTCVD reactor. This cleaning procedure resulted in a hydrogen-passivated, oxide-free surface [64].

A typical growth sequence starts with an initial 1 minute clean at 1000°C in hydrogen, followed by a $\sim 1\mu m$ thick buffer layer grown at 1000°C using dichlorosilane as Si source. After the buffer layer a thin Si layer is grown at 700°C, and the temperature is then lowered (with dichlorosilane on) to 625°C for all $Si_{1-x}Ge_x$ growth. The $Si_{1-x}Ge_x$ growth is controlled by turning the GeH_4 source gas (0.8% in H_2) on and off. A run/vent gas line configuration is used so that the gas flows in the mass flow controllers are stabilized before switching reactive gases into the growth chamber. Note that gas switching, as in conventional CVD, and not temperature switching, as in LRP, is used to start and stop the growth of layers.

To study the growth pressure effects on the material quality and interface roughness, the growth pressure was varied between 6 and 220 torr. The pressure was controlled by a butterfly valve on the input of the process pump. Since the hydrogen carrier was fixed at 3 slpm, a higher pressure implied a lower gas velocity in the growth chamber.

For $Si/Si_{1-x}Ge_x/Si$ HBT structures, the growth pressure was fixed at 6 torr. The desired Ge and boron profiles in the base were controlled by germane and diborane flow-rates and switching. The high-temperature buffer layer was used as collector, not *in-situ* doped, but rather using the phosphorus outdiffusion from the n^+ buried layer (devices of chapter 5) or using the fact that the background doping in our high temperature Si layers was $N_D \sim 1 \times 10^{16}$ cm^{-3} (devices of chapter 2). The emitters were the top Si layers grown at 800°C and *in situ* doped with phosphorus.

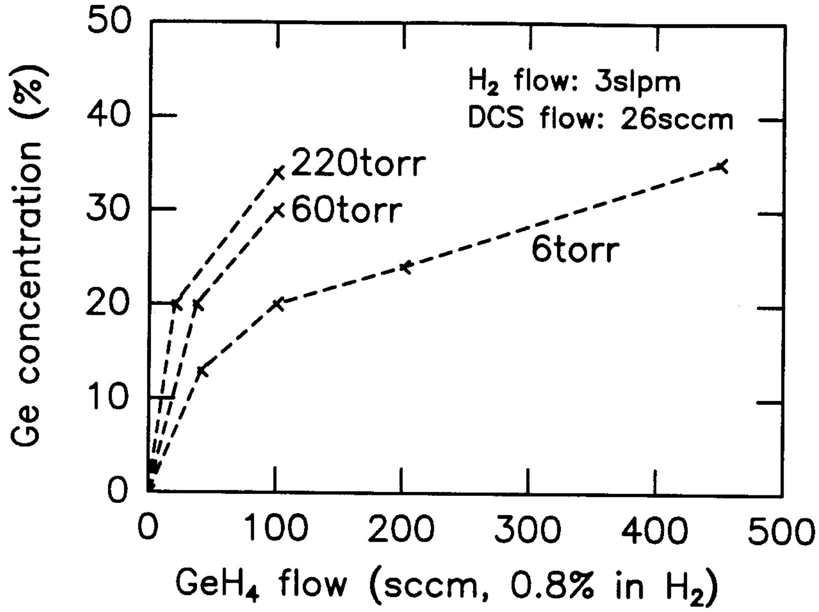


Figure 3.2: Ge fraction vs. GeH$_4$ flow for different growth pressures at 625°C. The H$_2$ was 3 slpm and the dichlorosilane flow was 26 sccm. The uncertainty in Ge fraction is +/- 2%.

3.3 Composition and growth rate at various growth pressures

The composition of thin Si$_{1-x}$Ge$_x$ alloy layers (150Å– 250Å) in the samples grown at various pressures was determined from the bandgap as measured by photoluminescence (assuming ΔE/1%Ge $\sim$ 8.5meV). Thin samples were used to avoid strain relaxation (although thicker than 100 Å to avoid quantum confinment effects). The Ge concentration in the alloy layer (x) increased with increasing GeH$_4$ flow at a given pressure, as shown in Fig. 3.2. Note however, that the Ge fraction increases with pressure for the same gas flow conditions, even though the dichlorosilane/germane ratio is constant. The reason for this effect is not known.

The growth rate in Si$_{1-x}$Ge$_x$ layers grown at various pressures was measured by

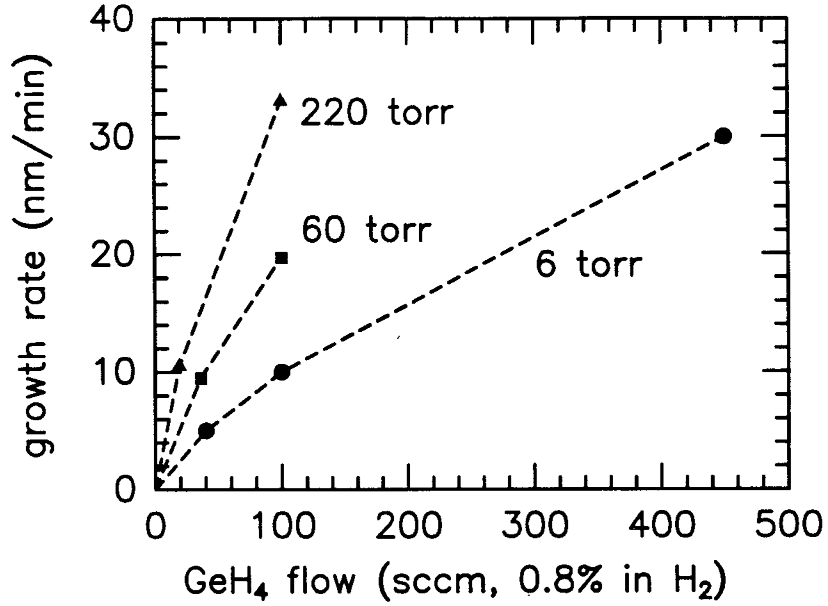


Figure 3.3: Growth rate vs. GeH_4 flow for different growth pressures at 625°C. The H_2 was 3 slpm and the dichlorosilane flow was 26 sccm. The uncertainty in growth rate is +/- 10%.

a bevel-and-stain technique on thick test samples (~ 1000Å). On thin test samples (200 – 400Å) selective chemical etching of the $Si_{1-x}Ge_x$ layer was done and the thickness was measured with a stylus profilometer. The resulting growth rates were consistent with those obtained by X-ray reflectivity measurements described later. The growth rate increased monotonically with GeH_4 flow, but it was higher for the same gas flows at higher pressure, as shown in Fig. 3.3. It is interesting to note that the growth rate was similar for the samples with similar Ge content in the solid, regardless of the growth pressure. For example, for a Ge fraction of x=0.2, GeH_4 source gas flows (0.8% in H_2) of 100, 35, and 19 sccm were required at pressures of 6, 60, and 220 torr, respectively, but the growth rate in all cases was ~ 100 Å/min.

3.4 Photoluminescence measurements

Photoluminescence (PL) results from the radiative recombination of excess electrons and holes when a semiconductor is illuminated by light of photon energy above bandgap. The radiative process always competes with non-radiative recombination processes due to deep levels or traps. Since $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ is an indirect bandgap material, for photoluminescence to be observed a a low level of non-radiative recombinations is required, implying high minority carrier lifetimes [65]. Hence, photoluminescence may be used as a sensitive probe of material quality.

PL spectra were measured at 77K, on 250Å thick, undoped $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ layers ($x = 20-34\%$) grown at various pressures, with a Si cap layer of 150 Å to prevent surface recombination. A typical $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ PL-spectrum at 77K shows a no-phonon (NP) feature at higher energy due to alloy randomness, and a lower energy transverse optical (TO) phonon replica (Fig. 3.4). The energy position of the NP peak is determined by the band-gap of the $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ quantum well, i.e. the Ge concentration.

All the samples clearly showed these two thermally broadened peaks corresponding to NP and phonon-replica lines, as expected for high quality films. The PL intensity was noticeably weaker for samples grown at 220 torr, indicating a lower lifetime. This might be due to increased oxygen incorporation at higher pressure, due to a higher background oxygen partial pressure from trace contamination of the gas sources.

3.5 X-ray reflectivity measurements

The interface roughness of SiGe layers was determined by measuring X-ray reflectivity (XRR) using energy dispersive detection, performed by Dr. E. Chason at Sandia National Laboratories. This technique measures interference due to reflections from different layers. The scattering vector (k) is scanned by keeping the scattering angle fixed and measuring the reflected intensity as a function of energy. XRR is sensitive

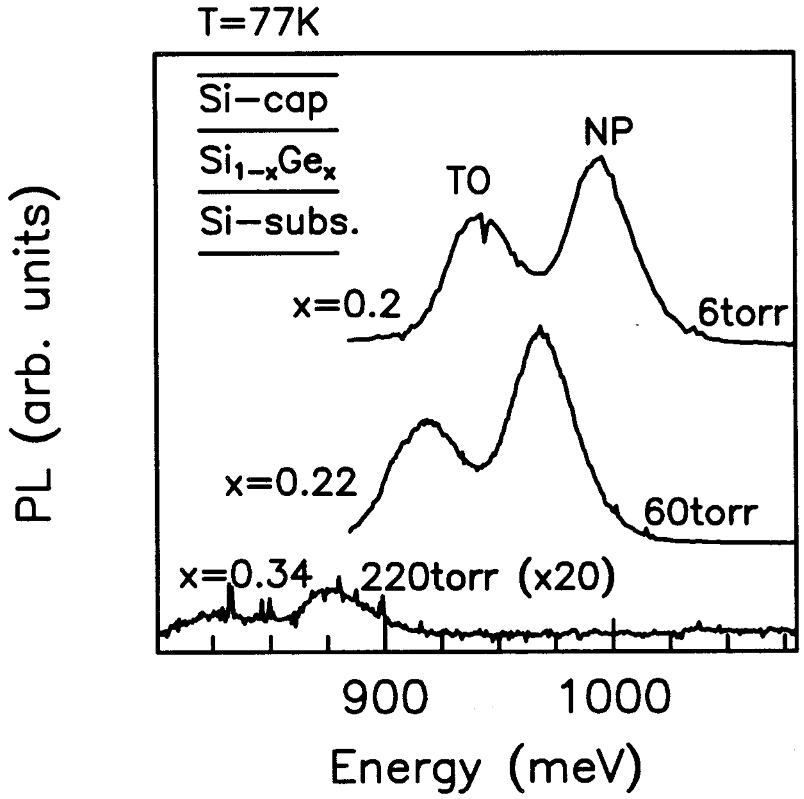
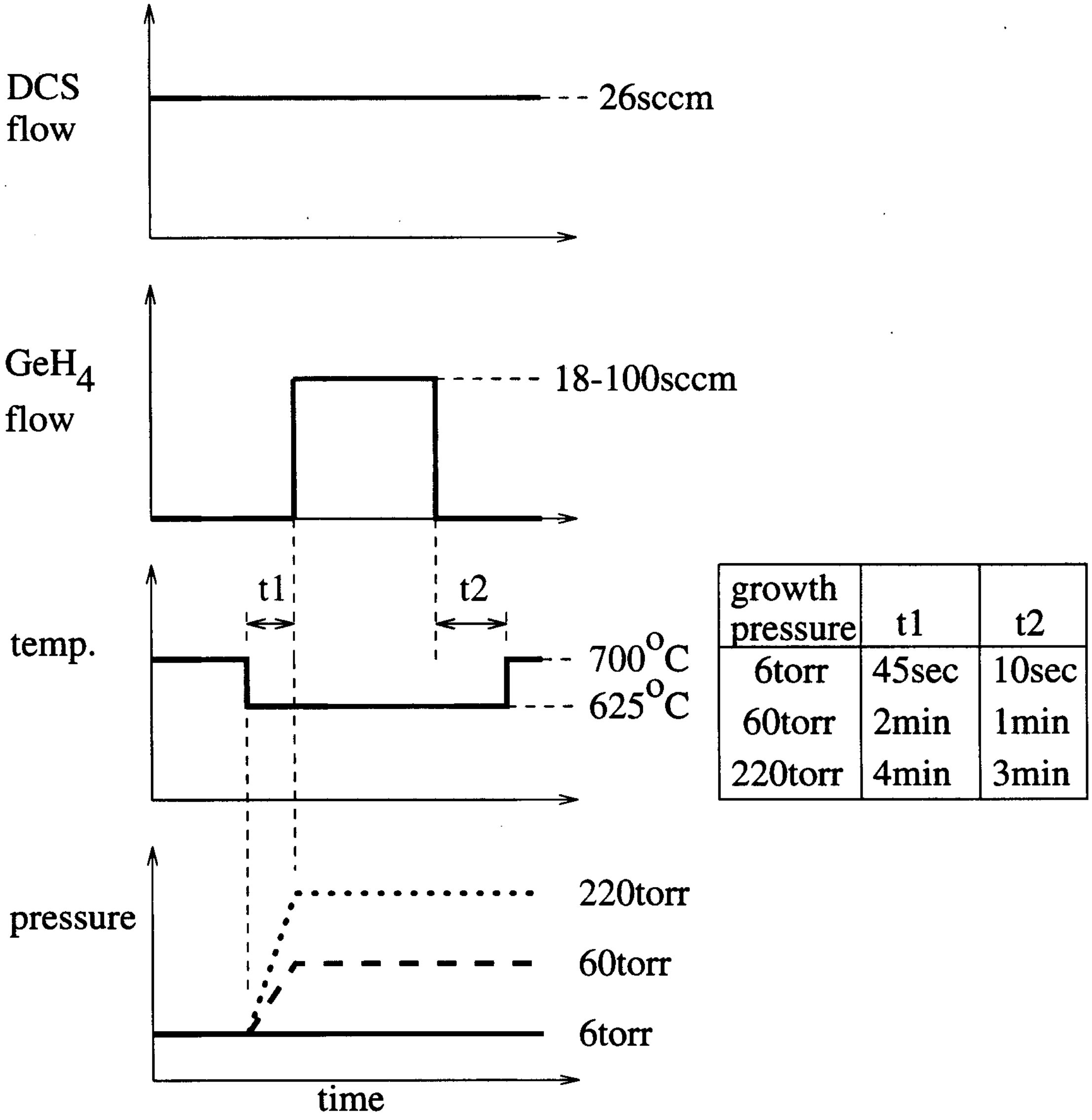


Figure 3.4: Typical PL spectra of SiGe samples grown at 6, 60, and 220 torr. Note the no-phonon (NP) peak and transverse optical (TO) phonon replica indicating band-edge luminescence and low defect concentration.

to the gradient in the electron density (dN/dz) normal to the surface so that the measured intensity is proportional to the square of the magnitude of the the Fourier transform of dN/dz. An interface is characterized by a non-uniform electron density profile in the direction perpendicular to the sample surface, i.e. a peak in the gradient profile. Abrupt interfaces will have a sharper peak in dN/dz, while rough or graded interfaces will exhibit a broader peak. The oscillations in the reflectivity spectrum result from the interference between different layers. The periodicity of the signal is determined by layer thicknesses, and oscillations decay at high "k" due to non-abrupt interfaces between layers. The amplitude of the signal depends on Ge concentration in the layer. The measured signal was fitted with a model where the interface width, layer thicknesses and Ge fraction were fitting parameters. The electron density gradient across the interface was approximated by a gaussian lineshape. The interface width (σ) was determined by a best fit to the data ($dN/dz \propto e^{-z^2/2\sigma^2}$). More details on this technique can be found elsewhere [66]. This method has been used to characterize the interface quality of various materials [66, 67, 68]. Interface abruptness can be degraded by both a gradual transition (i.e. grading) as well as a rough growth surface. However, since both lead to a gradual transition in electron density, grading vs. roughness effects can not be distinguished by this technique.

The samples used for this measurement were the same samples used for measuring the PL spectra. The $Si_{1-x}Ge_x$ layers were $200 - 300$Å thick with $x = 0.2 - 0.3$, followed by 150Å thick Si caps, grown as described earlier. Fig. 3.5 illustrates how the gases, pressure and temperature were switched in time. The dichlorosilane flow was not interrupted when the temperature was lowered from 700°C to 625°C for the $Si_{1-x}Ge_x$ growth, which is generally typical for our growth conditions, since the Si growth rate below 700°C is essentially zero. The delay t_1 in Fig. 3.5 is the time needed for the temperature and the pressure to stabilize (if pressure higher than 6torr), while t_2 is the waiting period for germane to purge out from the reactor tube. Typical XRR


DCS flow
26sccm
GeH_4 flow
18-100sccm
temp.
t1
t2
700°C
625°C
pressure
220torr
60torr
6torr
time

growth pressure	t1	t2
6torr	45sec	10sec
60torr	2min	1min
220torr	4min	3min

Figure 3.5: Qualitative diagram that shows the sequence of gas and temperature switching during growth

spectra and the best fits are shown in Fig. 3.6, for structures grown at 6, 60 and 220 torr. Note the faster decay of the oscillations at higher pressures indicating wider interfaces in the high pressure samples.

In all samples, a best fit to the top surface roughness (the surface of the Si cap) was found to be $\sigma = 3 - 5\text{Å}$, regardless of the growth pressure. This number clearly shows roughness, not grading, since this is the interface between Si and vacuum. The presence of a native oxide on the Si cap was shown to have a negligible effect on the roughness determination by removing it with dilute HF from some samples. Fig. 3.7 shows the interface width vs. growth pressure both for bottom (between the Si substrate and SiGe layer) and top (between the SiGe layer and Si cap) interfaces. No clear dependence of interface width on Ge concentration in the alloy layer was observed. The interface width was found to linearly increase as the growth pressure was increased, for both interfaces. Note, however, that regardless of the growth pressure, the bottom interface is worse (wider) than top. The summary of the samples is given in Table 3.1.

We now propose a model to explain the increasing interface width with pressure and the larger width of the bottom interface. The model is based on gas transients when switching layers, causing a graded interface. When GeH_4 is switched on/off, the partial pressure of the gas in the tube increases gradually, not instantly. Growth during GeH_4 transients results in a graded interface. Since the same H_2 carrier flow was used in all experiments (3 slpm), the gas velocity through the chamber is inversely proportional to pressure, and the residence time of gas in the chamber is linear with pressure. Therefore, at higher pressure, slower transients are expected and thus, a thicker interface. A differentially pumped residual gas analyser (RGA) was used to measure the actual rate of change of germane pressure in the growth chamber (although at a location significantly downstream from that of the silicon wafer). At 6 torr, the time scale for the transient was a few seconds, while it was as large as 60

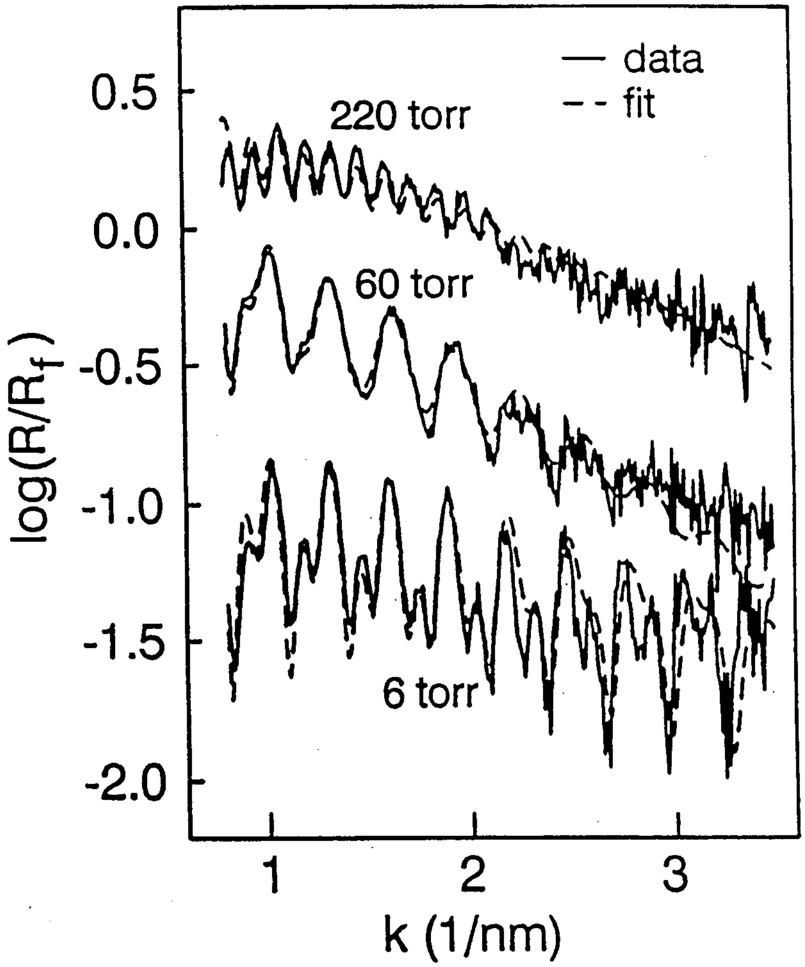


Figure 3.6: Typical XRR spectra for structures with single $Si_{1-x}Ge_x$ layers ($x = 0.2$–0.3, $t = 200 - 300$Å) with 150Å Si-caps. The reflected intensity is plotted as a function of scattering vector for different pressures. The solid line represents the data and the dashed line is the best fit.

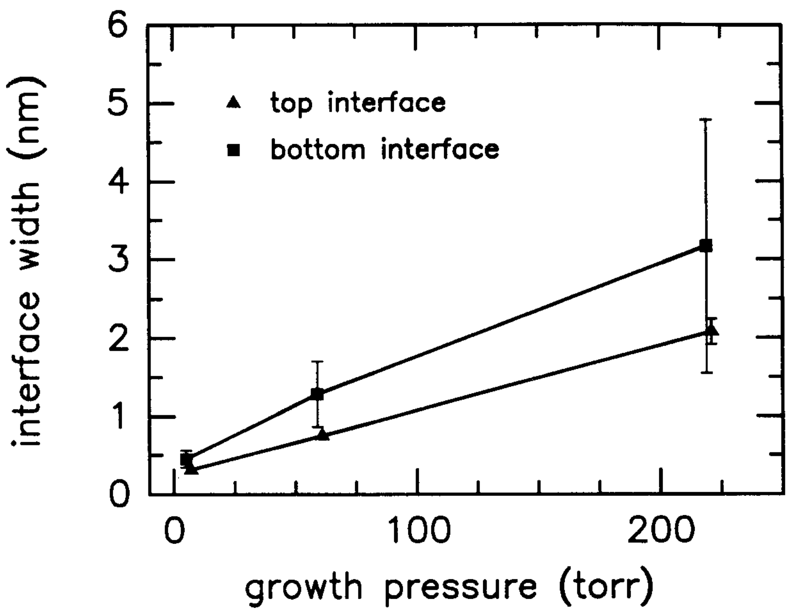


Figure 3.7: The interface width of the top and bottom $Si_{1-x}Ge_x/Si$ interfaces as a function of growth pressure.

seconds at 60 torr. For a chamber volume of ~30l, one calculates similar results (4.7 sec at 6 torr, 47 sec at 60 torr). Using the measured gas transients and known growth rates, the calculation of interface width yielded an order-of-magnitude agreement with the values measured by XRR.

A second interesting feature in the data of Fig. 3.7 is that the lower interface was consistently broader than the top interface. Because of the low growth temperature of $Si_{1-x}Ge_x$ (625°C), this could not be due to excess thermal diffusion at the lower interface. If one were to explain this interface width as due to the roughness of the growth surface, one would need to assume that the original silicon homoepitaxial surface (before the $Si_{1-x}Ge_x$ growth) were rougher than the heteroepitaxial $Si_{1-x}Ge_x$ growth surface. Exactly the opposite is observed in practice, however [69, 70]. Rather, this difference between the two interfaces is due to the dependence of the growth rate on germane flow. The RGA measurement showed that the gas transients had the shape of a decaying exponential, as illustrated in Fig. 3.8. The growth rate vs. time will have a similar shape as the GeH_4 partial pressure, as shown in Fig. 3.3. The interfacial width will be the integral of the growth rate during the transient period (cross-hatched in Fig. 3.8). The lower $Si/Si_{1-x}Ge_x$ interface, which occurs when GeH_4 was turned on, has a large part of the transient in a region of high GeH_4 flow and hence high growth rate, leading to a thick graded region. Similarly, the top interface, occuring when the GeH_4 is turned off, has most of its transient during a period of low GeH_4 flow and low growth rate, leading to a sharper interface. From RGA results and Fig. 3.8, one estimates that the bottom interface should be thicker than the top one by a factor of 1.5−2, consistent with the XRR data.

Therefore, at higher pressures, the interface width is limited by the grading at the interface, due to gas transients, and not the interface roughness. One can estimate the intrinsic roughness of the growth surface, however, by extrapolating the lines in Fig. 3.7. to zero pressure, where gas transients would be zero. In this case one finds an interface roughness of 2−4 Å, similar to the roughness of the top Si surface measured by XRR. This is consistent with the surface roughness of 3 Å (rms, measured by AFM) of a 110Å thick $Si_{1-x}Ge_x$ layer grown at low pressure (~0.11 torr), reported by A.J. Pidduck et al. [70]. Similarly, Dutarte et al. [71] have reported an upper limit to interface width of 4 Å, resolved from TEM micrographs of a $Si/Si_{1-x}Ge_x$ superlattice. The growth pressure was ~ 1.5 torr. The estimated gas transient times were in the range of 1 sec, resulting in interface layers with negligible grading at a low growth temperature (550°C) and relatively low growth rate (50Å/min).

3.6 Resonant tunneling diodes

The performance of resonant tunneling diodes (RTD's) was used to study the effects of interface abruptness on electrical devices. An RTD consists of a $Si_{1-x}Ge_x$ quantum

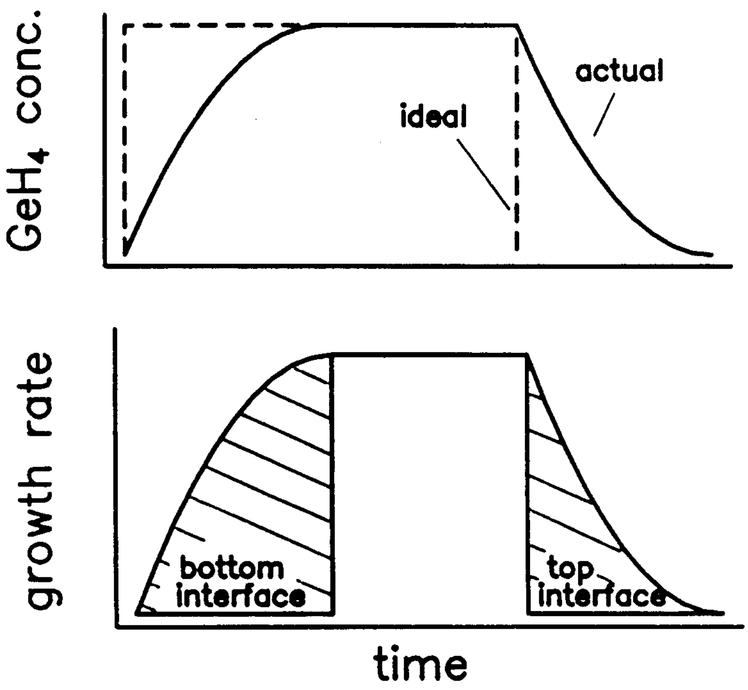
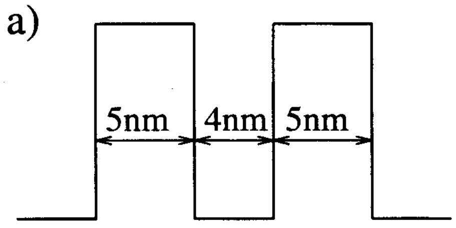

GeH4 conc.
actual
ideal
growth rate
bottom interface
top interface
time

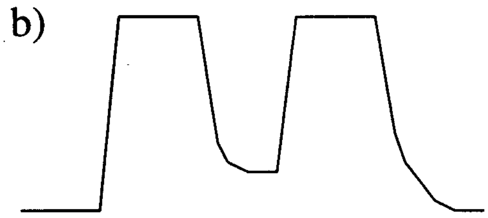
Figure 3.8: Qualitative diagram that explains the effects of gas switching on the thicknesses of graded layers at the top and bottom interface. The thickness of the interface layer is proportional to the cross-hatched area in the lower figure.

well sandwiched between Si barriers and symmetrically doped contact layers on both sides. The resulting I-V curve shows a peak in the current when the device is biased such that the bottom of the band filled with carriers in the emitter aligns with the state in the well. When biased further, such that there is no state to tunnel to, the I-V characteristics show negative differential resistance. The performance of RTD's is very sensitive to well and barrier thicknesses and to the quality of the interfaces. Both p- and n-type RTD's in $Si/Si_{1-x}Ge_x$ heterostructures have been demonstrated and studied by several groups [72, 73, 74, 75, 76].

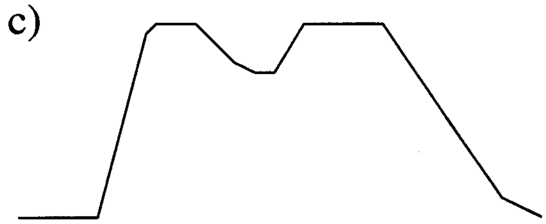
For the study of interface quality, p-type RTD's were fabricated with nominal barrier widths of 50 Å and well widths of 40 Å, (similar structure to that of ref. [72]) at growth pressures of 6, 60, and 220 torr (Fig. 3.9.a). The Ge fraction in the $Si_{1-x}Ge_x$ layers of double-barrier structures was 25-30%. The resulting I-V curves at 80K are shown in Fig. 3.10. The sample grown at 6 torr clearly shows negative differential resistance, as expected. Similar behaviour is observed in the 60 torr sample. The sample grown at 220 torr shows no resonant tunneling behaviour at all, rather the I-V characteristics look like that expected for a single barrier. This is not surprising, considering the X-ray reflectivity results. The interface widths are measured to be only $3-5$ Å at 6 torr, $7-13$ Å at 60 torr, but $20-40$ Å at 220 torr. That means that the actual structure grown at 6 torr is very close to the desired one. Although the transitions between the barriers and the well become more diffuse due to grading at 60 torr (Fig. 3.9.b), the double barrier structure is still clearly resolved. At 220 torr the well between the barriers almost disappears, leading to a single-barrier structure, consistent with the electrical results (Fig. 3.9.c). The performance of double barrier RTD's is thus consistent with the results measured by XRR.



desired



60torr



220torr

Figure 3.9: a) Desired valence band diagram for resonant tunneling structures grown at various pressures. b) and c) Qualitative band diagrams of structures grown at 60 and 220 torr, respectively, showing the effect of graded interfaces.

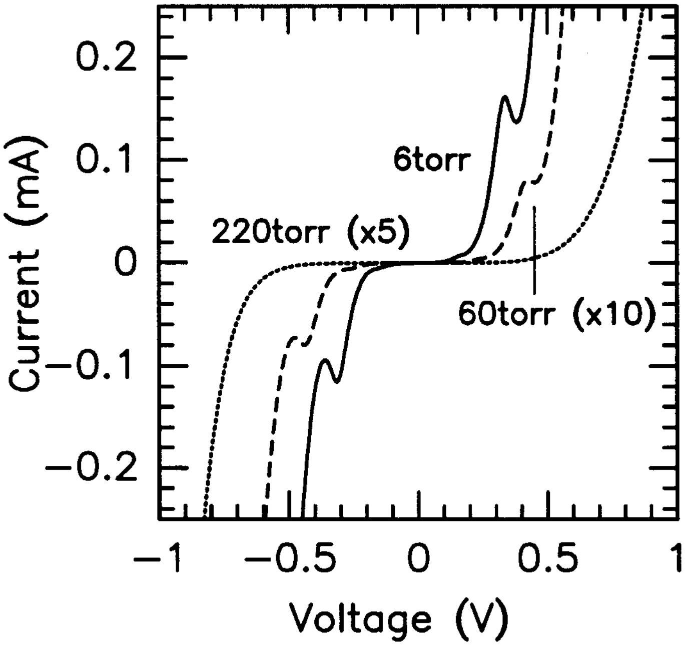


Figure 3.10: RTD I-V curves measured at 80K of samples grown at 6, 60, and 220 torr. The desired structure was the same in all three samples.

3.7 Low pressure limit to interface roughness

So far, we have demonstrated that high quality $Si_{1-x}Ge_x$ layers can be grown in a wide pressure range, but as the growth pressure increases the effects of gas transients become more significant because of longer residence time in the chamber. The interface width was found to be below 5 Å at 6 torr, and increases with increasing growth pressure due to grading. One can take different approaches to avoid the effect of grading due to gas transients and isolate the intrinsic limit to interface abruptness due to roughness. Simply lowering the pressure or increasing the carrier gas flow rate to decrease the gas residence time was not possible due to pumping speed limitations in the RTCVD system at Princeton. Another possibility would be to turn the growth on and off at interfaces by rapid switching of the sample temperature, as done in Limited Reaction Processing. With no growth occuring during the gas switching, which is done at a very low temperature, the effect of gas transients would be removed. A second approach is to lower the growth temperature, and hence, lower the growth-rate, using conventional gas switching as in our experiments. The interface layer will thus be thinner for the same time of gas transient.

Both methods were used to grow test structures at 6 torr. For the LRP approach, the growth was interrupted by reducing the lamp power such that the wafer temperature was too low (below 450°C) for growth during germane transient after the $Si_{1-x}Ge_x$ layer was grown at 625°C. The resulting interface widths measured by XRR were 2.9 Å for the top Si/vacuum interface, 5.9 Å for the bottom $Si/Si_{1-x}Ge_x/Si$ interface, but 0.6 Å for the top $Si/Si_{1-x}Ge_x/Si$ interface where the growth was interrupted. This is even lower than the low pressure limit extrapolated from the XRR results shown in Fig. 3.7.

To further probe the low pressure limit to interface abruptness, the the second approach of low-temperature growth was also probed. The growth rate at 625°C of

$Si_{0.8}Ge_{0.2}$ is 100 Å/min at 6 torr for our growth conditions. When the temperature is lowered to 550°C the growth-rate drops to only 5 Å/min. The measured interface widths at 6 torr were 3-5 Å. If the interface roughness was still degraded by grading, lowering the growth rate during gas transients 20 times would reduce the thickness of the grading due to transients to be negligible. The 77K PL spectrum of the sample where the $Si_{0.8}Ge_{0.2}$ was grown at 550°C is almost identical to the one grown at 625°C with the same Ge content and similar layer thicknesses, showing the high quality of such layers.

The interface width for such 550°C growth was probed by X-ray diffraction (XRD) of a superlattice (SL). Higher order satellite peaks decrease in intensity with mixing (grading) at the interface. By comparing the intensity ratios of the peaks one can determine the upper limit to interface roughness. We grew a 20-period superlattice (sample #1432) where 14 Å thick $Si_{0.8}Ge_{0.2}$ layers were grown at 550°C (using GeH_4 and SiH_2Cl_2), and 28 Å thick Si layers were grown at 625°C using silane. Silane was used instead of SiH_2Cl_2 for the pure Si layers so that a reasonable (12 Å/min) growth rate could be achieved at 625°C. Both transients of turning GeH_4 on and off occurred at 550°C. The XRD data is shown in Fig.3.11. A simulation of the XRD of the same structure was performed and the data was fitted by adding mixing at the interface. This was performed by convoluting the ideal Ge profile with a gaussian lineshape to represent diffusion. The XRD simulation program developed by H. C. Manoharan at Princeton [77] was used. Since other effects, e.g. non-ideal periodicity in the superlattice, could reduce the XRD satellite peak intensities, this method gives an upper limit to the interface width (roughness). XRD results are summarized in Table 3.2. We compared the intensity ratios of the SL peak and the first satellite peak, and the ratio of the first and the second satellite peaks with the data. The best fit gives an upper limit to interface width (σ) of 1−2 Å. Within the experimental error, this is consistent with the low pressure limit of X-ray reflectivity results and

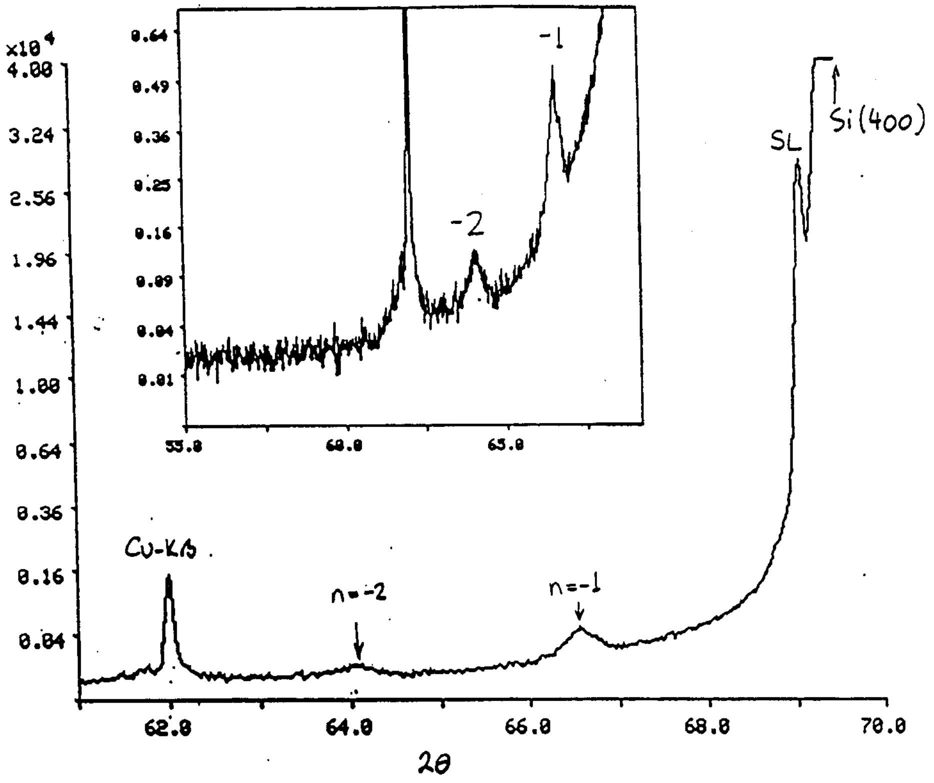


Figure 3.11: X-ray diffraction data of a 20-period $Si_{0.8}Ge_{0.2}/Si$ superlattice sample compared to simulation to determine the interface roughness.

shows that very abrupt interfaces without gas transient effects can be achieved with low–temperature CVD.

3.8 Conclusions

High quality $Si_{1-x}Ge_x$ layers can be grown by RTCVD from 6 to 220 torr, at 625°C, although the lifetime in the samples grown at 220 torr appeared degraded compared to that grown at lower pressure. At higher growth pressures the growth rate increases and the Ge fraction in the films was higher than at low pressure for the same gas flows and temperature. The interfacial width was found to be strongly degraded by gas transients at higher pressures, the lower interface was affected more than the upper Si/$Si_{1-x}Ge_x$ interface. By reducing the temperature to reduce the growth rate (e.g. 550°C), or by interrupting the growth at the interface by decreasing the temperature, the effect of gas transients can be removed while high quality layers can still be grown. In this case an interface roughness below 2 Å for $x=0.2$ was found by XRD fitting and XRR measurements, which is, within experimental error, consistent with the extrapolated XRR low pressure limit of 2-4 Å.

The $Si_{1-x}Ge_x$ layers for HBT's in this thesis were grown at 6 torr. At this pressure high quality layers with abrupt interfaces can be grown (PL and RTD measurements). The interface width at 6 torr is, measured by XRR, found to be 3 - 5 Å. Considering that the typical base widths are of the order $\sim$ 100 Å, and typical undoped $Si_{1-x}Ge_x$ spacer layers are several tens of Å wide, transitions in Ge and dopant profiles of below 10 Å are sufficient so that the effects of interfacial widths may be neglected.

sample number	growth pressure (torr)	growth time (seconds)	GeH_4 flow (sccm)	Ge concentration (%)
1326	6	150	100	20
1334	60	75	100	30
1335	220	35	100	34
1371	60	150	35	20
1372	220	150	19	20
1386	60	150	35	22

sample number	top surface roughness (Å)	top interface roughness (Å)	bottom interface roughness (Å)	$Si_{1-x}Ge_x$ thickness (Å)
1326	2.9-3.1	2.9-3.4	3.4-5.6	225-226
1334	2.9-3.1	5.4-8.3	4.1-6.9	252-253
1335	3.3	23.1	33	222-231
1371	4.5	7.0-7.3	15.6-16	234
1372	4.0	19.5	18-50	244-275

Table 3.1: Summary of XRR samples and results: a) growth parameters, b) XRR results

		I_{-1}/I_{-2}	I_{-0}/I_{-1}
simulation	ideal	4.4	36.0
	$\sigma = 1$Å	4.7	37.1
	$\sigma = 1.5$Å	5.9	40.6
data		5.1	40.8

Table 3.2: Simulated and measured ratios of XRD intensities of superlattice peaks. Simulation with no mixing and including mixing at the interface was compared to data. The best fit for interface width is between 1 Å to 2 Å.

Chapter 4

Tunneling Current in $p^+ - n^+$ Junctions

4.1 The importance of tunneling currents in heavily doped p-n junctions for bipolar transistors

Several studies have shown that the forward current of heavily doped p^+–n^+ silicon junctions is affected by a parasitic tunneling current [78, 79, 80, 81, 82]. This current is not band-to-band tunneling, but rather the tunneling of both electrons and holes to midgap states at the p-n interface, as illustrated in Fig. 4.1. The tunneling current becomes dominant at high doping levels. However, heavily doped layers are desired for good performance of both Si-BJT's and Si/$Si_{1-x}Ge_x$/Si HBT's. In scaled bipolar transistors an increase in the base doping is required to avoid punchthrough, and heavily doped bases are also desired for low base resistance and high Early voltages. High emitter dopings are desired for high emitter efficiency and low series resistance. When the lightly doped side of the base-emitter junction exceeds a certain doping ($> 1 \times 10^{18}$ cm^{-3} according to [78]), the space charge region narrows, the tunneling barrier becomes very small and causes non-negligible tunneling to the midgap states. This tunneling current becomes a significant component of the base current, especially at low forward biases. Although the significance of tunneling currents is more important at lower bias levels and is not a critical issue for digital performance of HBT's, ideal base currents are important for analog applications.

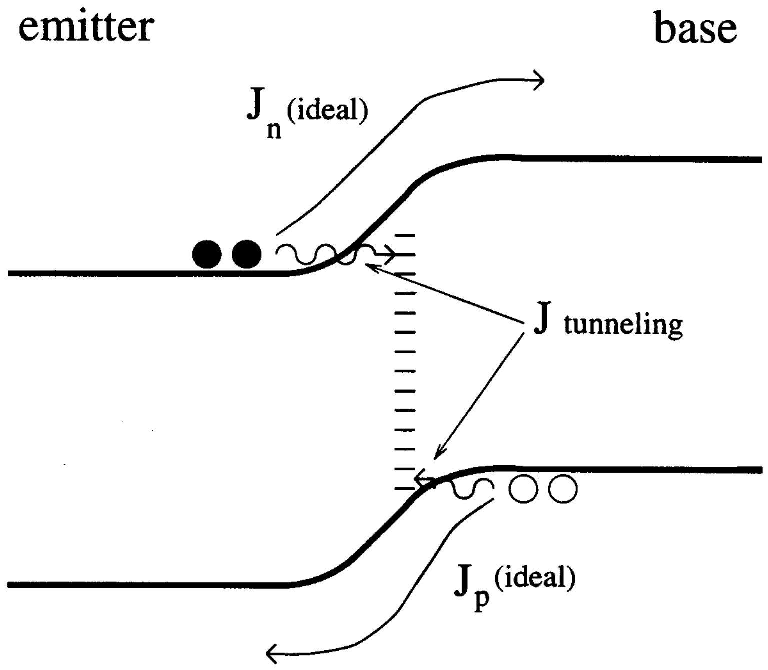


Figure 4.1: Band diagram of a p-n junction illustrating tunneling currents through interface states

The results of del Alamo et. al.[78] determined the doping level at the onset of significant tunneling for ion implanted p-n junctions. However, in Si/$Si_{1-x}Ge_x$/Si HBT's the base-emitter junction is, most likely, between epitaxially grown layers. The onset of significant tunneling does not have to be the same in epitaxially grown junctions, since the tunneling current is mediated by interface states, which could depend on the fabrication technique. The Si/$Si_{1-x}Ge_x$ heterojunction could also affect the interface layer and thus, tunneling current.

In this chapter, tunneling currents in heavily doped Si/Si homojunctions and Si/$Si_{1-x}Ge_x$ heterojunctions grown by RTCVD were studied [83] and the results were compared with the previously reported ion implantation data of ref. [78].

4.2 Experimental results

Epitaxial p^+–n^+(i.e. like a base-emitter) Si/Si and Si/$Si_{0.85}Ge_{0.15}$ layers were grown by RTCVD. Thick silicon n^+ layers (3-13μm) were grown on n-type substrates at high temperature (850°C-1000°C). The dopings ranged from 1×10^{17} to 1×10^{19} cm^{-3}. After the high temperature step, the growth was stopped for 30 seconds and the temperature was lowered to 700°C to prevent outdiffusion and provide an abrupt junction. Thin epitaxial p^+ silicon layers (50nm), doped 5×10^{19} cm^{-3}, were then grown at low temperature (700°C). On some of the samples, a thin p^+(30nm) $Si_{0.85}Ge_{0.15}$ strained epitaxial layer was grown at 625°C (doped 5×10^{19} cm^{-3}) between the n^+ and p^+ silicon layers. This provided a n^+-Si/p^+-SiGe junction. The dopings were measured by spreading resistance and C-V measurements. The values calculated from zero-bias capacitances confirmed the dopings calculated from the slope of $1/C^2$ vs. voltage. This shows that the junctions were abrupt and no excess diffusion occured during the 700°C layer growth. The junctions were isolated by a simple mesa process. Metal (Al) was used as a mask to define the device area by etching in SF_6/CCl_2F_2 plasma. Good ohmic contacts were obtained without annealing.

Room temperature forward bias I-V curves are shown in Fig. 4.3. The effect of the tunneling is clearly evident in the heavily doped devices at low current levels as the large "$n > 2$" current (n represents the ideality factor of the diode). Fig. 4.2 shows simulated band diagrams of Si/Si (solid line) and Si/$Si_{1-x}Ge_x$ junctions (dashed line) under forward bias condition ($V_a = 0.48$V) with $N_A = 5 \times 10^{19}$ cm^{-3} and $N_D = 1 \times 10^{19}$ cm^{-3}. Since the hole and electron currents to the traps must be equal, the current is controlled by the lightly doped side of the junction which has a larger tunneling barrier, as seen in Fig. 4.2. Therefore the tunneling current increases as the doping on the lightly doped side is increased. It should be stressed that this is not the typical space charge recombination current ($1 < n < 2$), which should increase

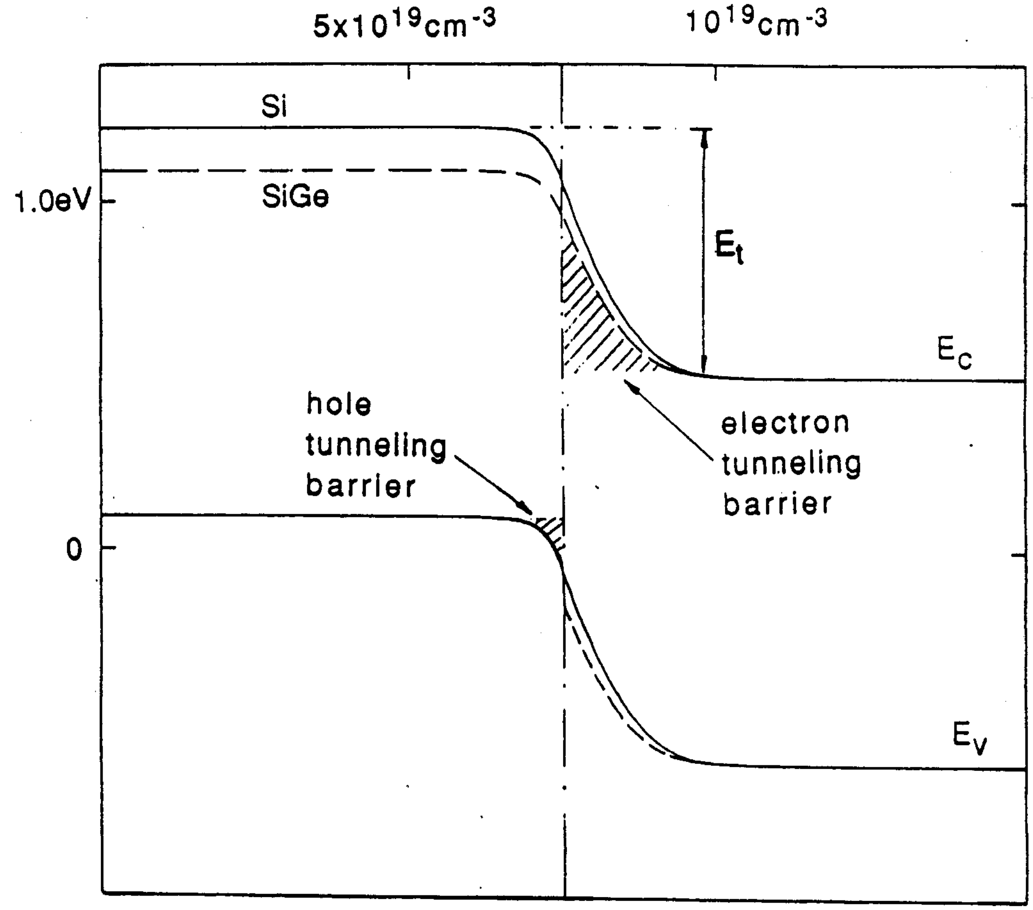

5x10¹⁹cm⁻³
10¹⁹cm⁻³
Si
SiGe
1.0eV
0
Et
Ec
hole
tunneling
barrier
electron
tunneling
barrier
Ev

Figure 4.2: Simulated band diagram of Si/Si (solid) and Si/$Si_{0.85}Ge_{0.15}$ (dashed) p^+–n^+junction at $V_a = 0.48V$, $N_A = 5 \times 10^{19}$ cm^{-3}, $N_D = 1 \times 10^{19}$ cm^{-3}. The cross-hatched areas show electron and hole tunneling barriers.

at low base dopings with larger space-charge regions, but is tunneling current which requires a small space-charge region (heavy doping) to reduce the tunneling barriers. I-V curves among the same area devices with the same doping were uniform, and measurements on the several different area devices (1.4×10^{-2} cm^2 – 9.5×10^{-5} cm^2) confirmed that the peripheral current components were negligible. The flattening of the I-V curves at high voltages is due to the parasitic contact resistances which varied from sample to sample and is not important for this analysis. Significant tunneling was observed at doping levels of the order of 1×10^{19} cm^{-3} for both Si/Si and SiGe/Si devices. Simulations of band diagrams by SEDAN showed a slightly lower tunneling barrier ($\sim$ 100 meV) for electrons in the SiGe/Si devices compared to the Si/Si devices for similar n-doping densities, as seen in Fig. 4.2. Since the n-type side has a substantially larger tunneling barrier for electrons than that for holes on p-side, the electron barrier will control the tunneling recombination rate. No significant difference in the behavior of Si/Si and SiGe/Si devices at the same doping levels was observed at room temperature however. The summary of samples used in this study is given in Table 4.1.

In figures 4.4.a and 4.4.b, the current density is plotted as a function of doping at two different bias levels (0.32V and 0.48V). Also plotted are calculated values of the ideal diode current. The ideal current is the calculated minority carrier injection current taking into account bandgap narrowing at high doping levels according to ref. [84] for n-type and ref. [85] for p-type dopings. The ideal current is similar in all devices since it is dominated by electron injection into the p^+Si layer, and to first order, not affected by n^+ doping level. Previous results for tunneling limits in ion implanted p^+–n^+ junctions [78] are also shown for comparison to our data. At low doping levels, the tunneling current is negligible compared to the ideal current. Similarly, at high biases the tunneling current component is less significant since the bias dependence of tunneling current is weaker than that of the minority carrier

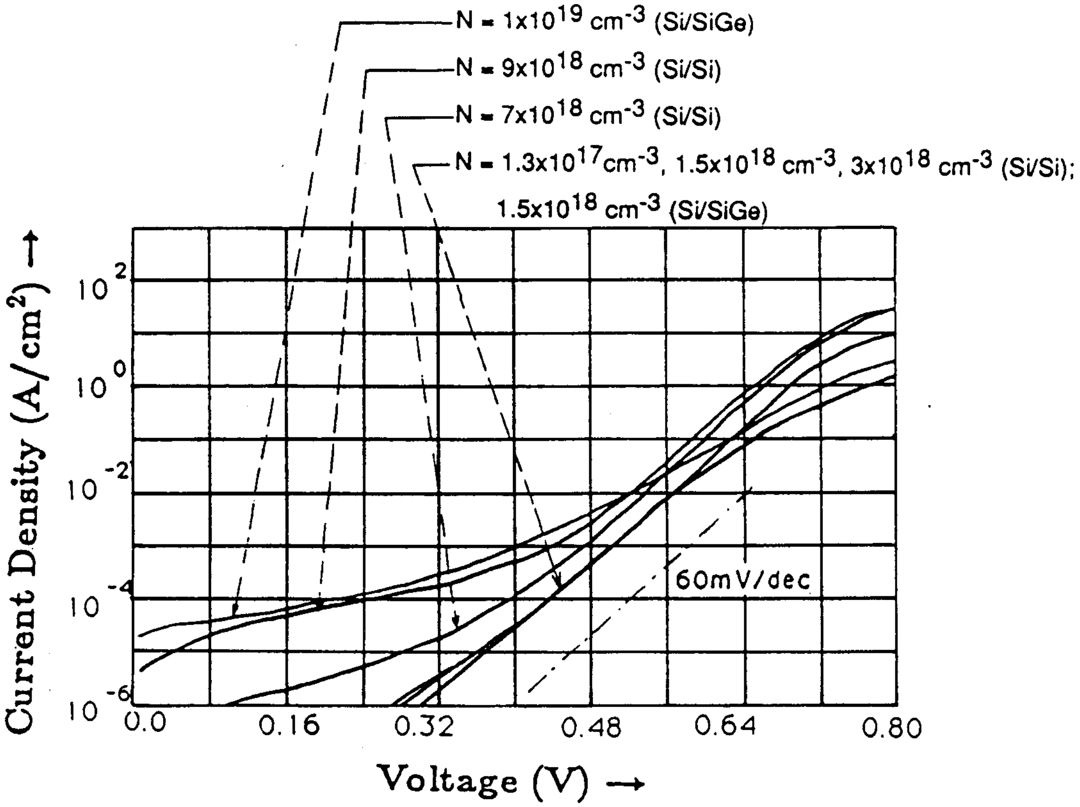


Figure 4.3: Room temperature forward bias I-V characteristics of several Si/Si and Si/SiGe devices with the following n-type dopings: (a) 1×10^{19} cm^{-3} (Si/SiGe); (b) 9×10^{18} cm^{-3} (Si/Si); (c) 7×10^{18} cm^{-3} (Si/Si); (d) 3×10^{18} cm^{-3} (Si/Si); (e) 1.5×10^{18} cm^{-3} (Si/Si and Si/SiGe); (f) 1.3×10^{17} cm^{-3} (Si/Si).

injection current [84] and the ideal current dominates (the tunneling current becomes apparent at higher doping levels). But at lower bias levels tunneling causes a several order of magnitude increase in the junction current of heavily doped devices.

That these high currents were indeed tunneling was confirmed with temperature-dependent measurements. Since the injected currents are vastly reduced at lower temperatures and the tunneling currents are fairly insensitive to temperature, the effect of tunneling is even more significant at low temperatures. Current-voltage characteristics were measured in the temperature range from 175K to 350K. Fig. 4.5 shows current density as a function of doping measured at 200K, as well as room temperature tunneling current of heavily doped devices. The increase with doping is even more obvious than at room temperature since the ideal current is negligible compared to tunneling current. To explain the temperature dependence of tunneling current, we have used the approach of Chynoweth et al. [82]. The tunneling current is proportional to the density of interface states D and tunneling probability P:

$$\mathrm{I_t} \propto \mathrm{D} \times \mathrm{P} \tag{4.1}$$

The tunneling probability for a triangular barrier can be approximated by:

$$\mathrm{P} = \exp(-\alpha \mathrm{E_t}^{3/2}/\mathrm{F}) \tag{4.2}$$

where α is a constant, E_t is the tunneling barrier height, and F is the maximum field in the junction determined by the doping and space-charge region width [82]. E_t is given by:

$$\mathrm{E_t} \approx \mathrm{E_G} - \mathrm{qV_a} + \mathrm{q(V_p + V_n)} \tag{4.3}$$

where $\mathrm{E_G}$ is the bandgap of the p-type material (in case of lighter doped n-side), $\mathrm{V_a}$ is the applied bias and $\mathrm{V_n}$ and $\mathrm{V_p}$ are the potential differences between the Fermi level and conduction band on n-side, and valence band on p-side, respectively. To first order, the temperature dependence of the tunneling current will be determined by

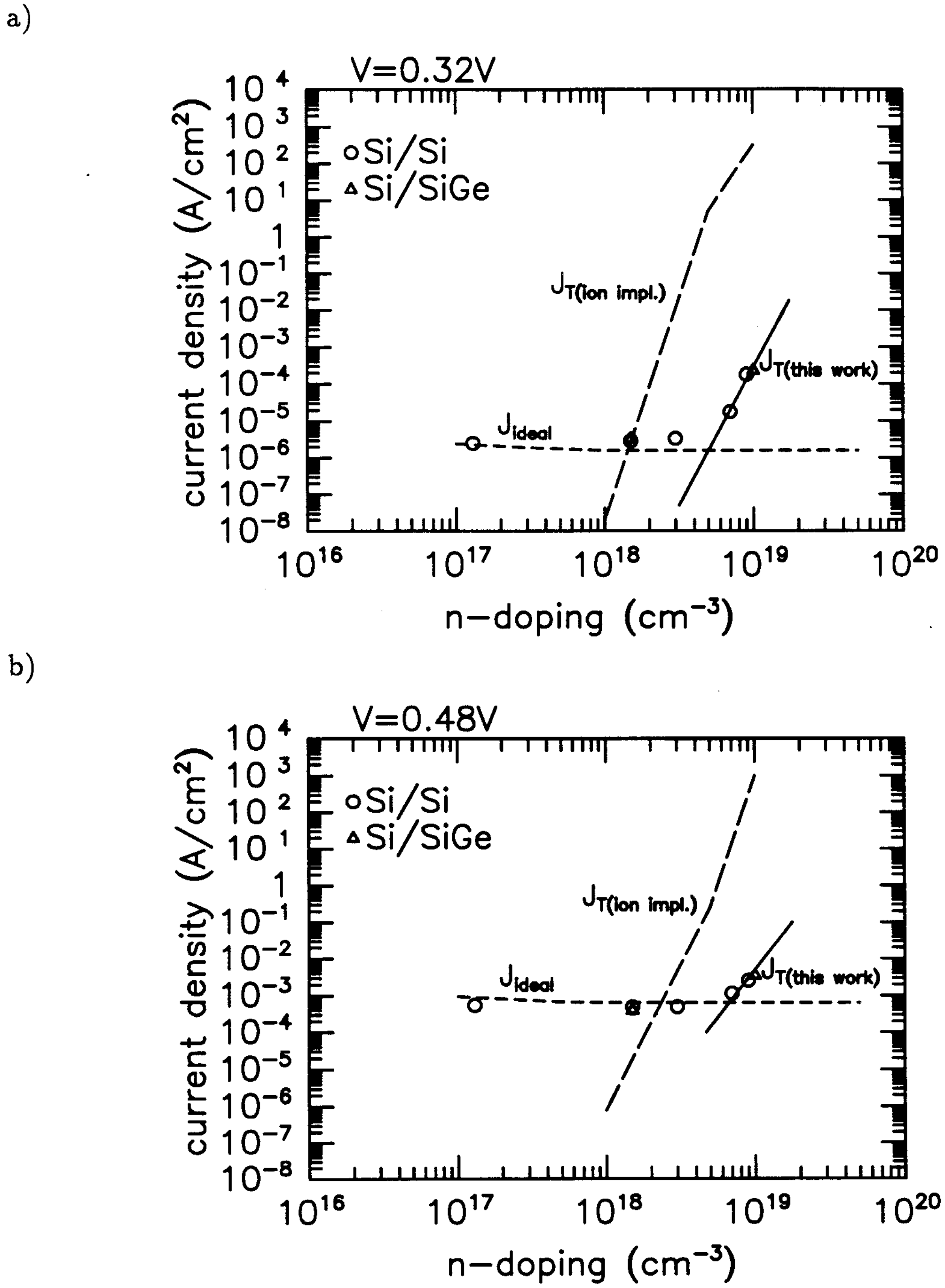


Figure 4.4: Current density vs. n-type doping at (a) $V_a = 0.32V$ and (b) $V_a = 0.48V$. J_{ideal} is the calculated ideal current density, $J_{T,ion.impl.}$ corresponds to the tunneling currents in ion implanted junctions and $J_{T,this\ work}$ corresponds to this work. The solid lines are fits to the heavily doped data points (7×10^{18} cm^{-3} – 1×10^{19} cm^{-3}).

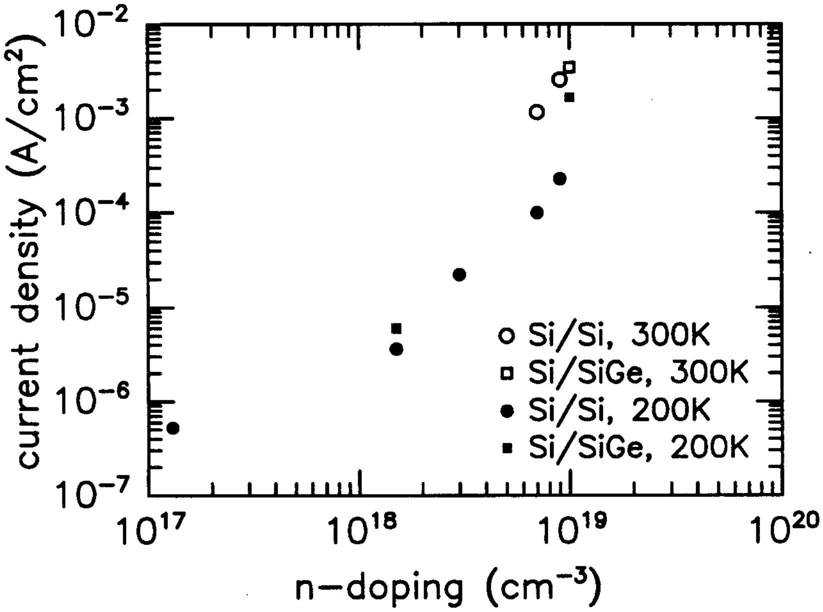


Figure 4.5: Measured current density as a function of doping at T=200K and V_a = 0.48V. Tunneling currents of heavily doped devices measured at room temperature are also shown for comparison.