Optimization, analysis, and fabrication of SiGe heterojunction bipolar transistors

by

Kai Hay Kwok

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical Engineering

Waterloo, Ontario, Canada, 2001

©Kai Hay Kwok 2001

National Library of Canada

Acquisitions and Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada

Acquisitions et services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-60549-3

Canada

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

This thesis deals with the optimization, analysis, and fabrication of silicon-germanium (SiGe) heterojunction bipolar transistors (HBTs).

Two vertical base profile optimization studies for improving the high-frequency performance of SiGe HBTs are presented. In the first study, the Ge profile is optimized for the minimum contribution of the emitter and base delay times to the transition frequency in the low-injection regime. A fixed Ge dose is used as the optimization constraint. Non-quasi-static effects at high frequencies are taken into account. It is shown that the graded Ge profile is more effective than the box Ge profile in minimizing the two delay time contributors for SiGe HBTs with today's typical emitter and base dimensions.

In the second optimization study, the base doping and Ge profiles are optimized for minimum base delay time in low- and high-injection regimes before the onset of Kirk effect. Fixed Ge dose, intrinsic base resistance, and base concentration near the emitter are adopted as optimization constraints. The effect of plasma-induced bandgap narrowing in high injection is considered. An iteration scheme for calculating the base delay time for a wide range of collector current densities is developed. It is shown that the retrograde base doping profile with graded Ge profiles gives the minimum base delay time in both low- and high-injection regimes.

An analysis of the retrograde portion of a base retrograde doping profile in a SiGe HBT is also performed. A closed-form analytical expression of the base delay time is derived with various physical effects taken into consideration. The relative importance of the physical effects is assessed. It is found that the adverse effect of the retrograde portion of the base retrograde doping profile on the base delay time is less pronounced than expected, especially when a high Ge grading exists across the base. It is also shown that the effect of the field dependency of the electron

diffusivity needs to be considered when modelling the base delay time in the SiGe base with a high electric field.

Finally, SiGe HBTs are fabricated by high-dose Ge implantation, Si amorphization, and solid-phase epitaxy. The results from electrical measurements are presented. Although further work is required in this area, transistor action is observed in SiGe HBTs with Si amorphization used.

Acknowledgements

First of all, I wish to express my gratitude to my supervisor, Professor Selvakumar, for his advice and encouragement throughout the course of my study, and for introducing me to the fascinating area of semiconductor devices. I am deeply thankful for all the freedom he granted me in my research.

I would like to acknowledge the financial support from the National Science and Engineering Research Council (NSERC) of Canada and Micronet. I am grateful to the Department of Electrical and Computer Engineering at the University of Waterloo for its fabrication facilities.

My heartfelt appreciation also goes to the following people who have encouraged and supported me in many different ways in the past few years: Ada, Christopher & Gladys, Daniel, Denny & Grace, Dickson, Fenton, Jack, Jasmine, John, Kenneth & Viola, Leith & Kit-Sum, Mike, Patrick, Philip & Rainbow, Puiwing, Rock, Russell, Sandra, Sovannary, Stephen, Suzanne, Tim, Tinnie, Wilfred, William & Donna, and Drs. Andrew Wong & Nancy Wong.

I thank my parents for their undying patience and love. Without them, this work would have never been accomplished. Thank you, Ah pa Ah ma.

Above all, I thank my triune God for giving me life to learn.

To my parents

Wai King Ng &
Shue Yue Kwok

Contents

1 Introduction 1

1.1 A brief history of SiGe HBTs 4

1.2 SiGe heterojunction bipolar transistors 7

1.3 Why SiGe HBTs? 12

1.4 Transition frequency 15

1.5 Outline of the thesis 21

2 Ge profile optimization 23

2.1 Introduction 23

2.2 Theory 27

2.2.1 Profile definitions 27

2.2.2 Emitter delay time 32

2.2.3 Base delay time 47

2.2.4 Non-quasi-static correction 51

2.3 Verification 68

2.3.1 Emitter and base delay times 68

2.3.2 D.C. minority carrier profiles in emitter and base 69

2.3.3 A.C. current gain and total delay time 74

2.4 Results and Discussions 77

2.5 Conclusions 99

3 **Analytical expressions of base delay time** **106**

3.1 Introduction 106

3.2 Regional model 111

3.2.1 Base profile approximation 111

3.2.2 Apparent bandgap narrowing due to heavy doping effects . . 111

3.2.3 Bandgap narrowing due to Ge concentration gradient 112

3.2.4 Doping-dependent diffusivity 113

3.2.5 Velocity saturation at collector-base junction 113

3.2.6 Regional base transit times 116

3.2.7 Electric-field-dependent diffusivity 117

3.2.8 Analytical expressions for base delay time 119

3.3 Results 122

3.4 Verification and comparison 128

3.5 Conclusions 131

4 **Doping and Ge profile optimization** **133**

4.1 Introduction 133

4.2 Theory 136

4.2.1 Assumptions 136

4.2.2 Profile Definitions 138

4.2.3 Solution method 140

4.3 Verification 147

4.4 Low injection 152

4.5 High injection 154

4.6 Conclusions 165

5 Fabrication of SiGe HBTs **167**

5.1 Introduction 167

5.2 Goal of the experiment 170

5.3 Device fabrication 172

5.4 Experimental results 176

5.5 Conclusions 182

6 Conclusions **183**

A General formulation for S'_p **187**

B Derivation for equation (3.26) **192**

C Physical models for electron transport **193**

D Fabrication process **199**

Bibliography **202**

List of Tables

1.1 SiGe BiCMOS processes (Note. f_T and f_{max} values correspond to SiGe HBTs). 6

2.1 Profile details of SiGe test devices. 74

2.2 Profile details of transistor set #1 (†indicates a uniform base profile). . . 78

2.3 Profile details of transistor set #2 (†indicates a uniform base profile). . . 86

2.4 Profile details of transistor set #3 (†indicates a uniform base profile). . . 92

2.5 Profile details of transistor set #4 (†indicates a uniform base profile). . . 97

5.1 Conditions used in the first annealing step (Note: all annealing steps were performed in a N_2 ambient). 171

A.1 Definitions of poly-emitter model parameters 189

D.1 Fabrication process steps. 200

D.2 Fabrication process steps (cont'd). 201

List of Figures

2.1 A schematic of the net doping profile under study (Note. The doping concentration is drawn in a logarithmic scale). 28

2.2 A schematic of the Ge profile under study (Note. The Ge fraction is drawn in a linear scale). 31

2.3 A typical minority hole profile in poly-Si emitter structure used in this study. 36

2.4 Emitter delay time versus emitter junction depth (mono-Si region thickness). Results of Suzuki [110] are the sum of the emitter delay times in poly-Si and mono-Si region. Basu *et al.* 's [112] results were available as individual emitter delay times in mono-Si and poly-Si regions. Infinite recombination velocity at emitter contact (S_m) and uniform emitter doping were assumed in [110,112]. 70

2.5 Emitter delay time versus polysilicon thickness. Results of Suzuki [110] are the sum of the emitter delay times in the poly-Si and mono-Si regions. Basu *et al.* 's [112] results were available as individual emitter delay times in the mono-Si and poly-Si regions. Infinite recombination velocity at emitter contact (S_m) and uniform emitter doping were assumed in [110,112]. 71

2.6 Base and emitter delay time versus base width. Infinite recombination velocity at emitter contact (S_m) and uniform doping in both base and emitter were assumed in [110]. 72
2.7 D. C. minority carrier profiles in neutral emitter and base regions of SiGe HBTs. Comparison of results generated by analytical expressions, derived in this work, and those from the numerical simulator S^3A ($S_m = 10^6$ cm/s, $S = 10^7$ cm/s, $S_p = 0$ cm/s). 75
2.8 A.C. current gain (h_{FE} or β) as a function of the frequency of SiGe test HBTs. Comparing of numerical simulation results from S^3A (symbols) with results generated by analytical expressions, based on the quasi-static assumption (dash lines, (2.68)), and those using non-quasi-static correction (solid lines, (2.106)). 76
2.9 Emitter delay time vs. Ge grading for transistor set #1 (W_{ep} = 0.38 μm, X_{em} = 0.02 μm, X_b = 100 nm). 80
2.10 D.C. current gain as a function of Ge grading for transistor set #1 (W_{ep} = 0.38 μm, X_{em} = 0.02 μm, X_b = 100 nm). 81
2.11 Base delay time as a function of Ge grading for transistor set #1 (W_{ep} = 0.38 μm, X_{em} = 0.02 μm, X_b = 100 nm). 82
2.12 Total delay time (without non-quasi-static correction) as a function of Ge grading for transistor set #1 (W_{ep} = 0.38 μm, X_{em} = 0.02 μm, X_b = 100 nm). 83
2.13 Effective total delay time (with non-quasi-static correction) as a function of Ge grading for transistor set #1 (W_{ep} = 0.38 μm, X_{em} = 0.02 μm, X_b = 100 nm). 85
2.14 Emitter delay time vs. Ge grading for transistor set #2 (W_{ep} = 0.38 μm, X_{em} = 0.02 μm, X_b = 30 nm). 87

2.15 Base delay time as a function of Ge grading for transistor set #2 ($W_{ep} = 0.38$ μm, $X_{em} = 0.02$ μm, $X_b = 30$ nm). 89
2.16 Total delay time (without non-quasi-static correction) as a function of Ge grading for transistor set #2 ($W_{ep} = 0.38$ μm, $X_{em} = 0.02$ μm, $X_b = 30$ nm). 90
2.17 Effective total delay time (with non-quasi-static correction) as a function of Ge grading for transistor set #2 ($W_{ep} = 0.38$ μm, $X_{em} = 0.02$ μm, $X_b = 30$ nm). 91
2.18 Emitter delay time vs. Ge grading for transistor set #3 ($W_{ep} = 0.25$ μm, $X_{em} = 0.15$ μm, $X_b = 30$ nm). 93
2.19 Total delay time (without non-quasi-static correction) as a function of Ge grading for transistor set #3 ($W_{ep} = 0.25$ μm, $X_{em} = 0.15$ μm, $X_b = 30$ nm). 94
2.20 Effective total delay time (with non-quasi-static correction) as a function of Ge grading for transistor set #3 ($W_{ep} = 0.25$ μm, $X_{em} = 0.15$ μm, $X_b = 30$ nm). 95
2.21 Effective total delay time (with non-quasi-static correction) as a function of Ge grading for transistor set #4 ($W_{ep} = 0.25$ μm, $X_{em} = 0.15$ μm, $X_b = 30$ nm, $N_{bo} = 10^{20}$cm^{-3}). 98
2.22 Graded Ge profile leverage factor vs. emitter junction depth for uniformly-doped base profiles (Circles: $X_b = 30$ nm, $N_{bo} = 1.8 \times 10^{19}$cm^{-3}; squares: $X_b = 50$ nm, $N_{bo} = 1.8 \times 10^{19}$cm^{-3}; triangles: $X_b = 100$ nm, $N_{bo} = 7.5 \times 10^{18}$cm^{-3}, total emitter width $= 0.4$ μm, $S = 10^7$cm/s). 100

2.23 Graded Ge profile leverage factor vs. emitter junction depth for retrograde base profile with $\theta = 0.2$ (Diamonds: $X_b = 30$ nm, $N_{bo} = 10^{20}\text{cm}^{-3}$; circles: $X_b = 30$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; squares: $X_b = 50$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; triangles: $X_b = 100$ nm, $N_{bo} = 2 \times 10^{19}\text{cm}^{-3}$, total emitter width = 0.4 μm, $S = 10^7$cm/s). 101

2.24 Graded Ge profile leverage factor vs. emitter junction depth for retrograde base profile with $\theta = 0.5$ (Circles: $X_b = 30$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; squares: $X_b = 50$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; triangles: $X_b = 100$ nm, $N_{bo} = 2 \times 10^{19}\text{cm}^{-3}$, total emitter width = 0.4 μm, $S = 10^7$cm/s). 102

2.25 Graded Ge profile leverage factor vs. emitter junction depth for retrograde base profile with $\theta = 0.8$ (Circles: $X_b = 30$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; squares: $X_b = 50$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; triangles: $X_b = 100$ nm, $N_{bo} = 2 \times 10^{19}\text{cm}^{-3}$, total emitter width = 0.4 μm, $S = 10^7$cm/s). 103

3.1 Doping profile for an exponentially-doped retrograde base. 109

3.2 Minority electron low-field mobility as a function of base concentration. Solid line: Klaassen's results [121, 125] with local $N_d = 10^{16}\text{cm}^{-3}$. Dashed line: fitted by (3.7). 114

3.3 Base transit times in different subregions as a function of the base concentration at the edge of the depletion layer (with all effects considered and $S = v_s = 10^7$cm/s assumed). 123

3.4 Fraction of the base transit time of the retrograde region as a function of the Ge grading across the base (with all effects considered and $S = v_s = 10^7$cm/s assumed). 124

3.5 Base transit time of the retrograde region and the total base transit time as a function of the base concentration at the edge of the depletion layer. Solid lines: our results with all effects considered and $S = v_s = 10^7$cm/s assumed. Dashed lines: Gao *et al.*'s results using (7)–(10) in [175] with $D_n^* = 4.4$ cm^2/s and F = 20 kV/cm (i.e., a Ge grading of 13.5%). Other parameters: $x_1 = 15$ nm, $x_2 = 25$ nm, $W_b = 50$ nm, $N_p = 10^{19}$cm^{-3}, $N_w = 4 \times 10^{17}$cm^{-3}. 126

3.6 Ratio of the base transit time with one effect neglected to the base transit time with all effects considered as a function of the Ge grading across the base ($S = v_s = 10^7$cm/s assumed). 127

3.7 Base transit time as a function of the Ge grading across the base. Darker dashed line: Patton *et al.*'s simulation results [93]. Solid line: $x_1 = 7.4$ nm, $x_2 = 12.5$ nm, $W_b = 26.2$ nm, $N_o = 6.27 \times 10^{18}$cm^{-3}. Dashed line: $x_1 = 8.9$ nm, $x_2 = 14$ nm, $W_b = 27.7$ nm, $N_o = 5.77 \times 10^{18}$cm^{-3}. Dot-dashed line: $x_1 = 10.4$ nm, $x_2 = 15.5$ nm, $W_b = 29.2$ nm, $N_o = 5.31 \times 10^{18}$cm^{-3}. Gao *et al.*'s results are generated using (7)–(10) in [175] with $D_n^* = 4.4$ cm^2/s. 129

4.1 Ge profiles (W_b is arbitrarily set to 100 nm for illustrative purpose) and base doping profiles with intrinsic base resistance of 5k$\Omega/\square$, N_{be} =5 $\times$ 10^{18}cm^{-3},N_{bc}=9 $\times$ 10^{16}cm^{-3}: uniform ($W_b = 29$ nm), Gaussian ($W_b = 54$ nm), exponential ($W_b = 83$ nm), retrograde ($W_b = 26$ $nm, R_p = 13$ nm, N_p =10^{19}cm^{-3}). 139

4.2 Collector current density vs. base-emitter voltage. Suzuki's uniform and Gaussian base: $W_b = 100\ nm, N_{be} = 2 \times 10^{18}\text{cm}^{-3}$. $N_{bc} = 2 \times 10^{16}\text{cm}^{-3}$ is used for Suzuki's Gaussian base [193]. Rinaldi's Gaussian base [210]: $W_b = 100\ nm, y(W_b) = 0.15$, triangular Ge profile, $N_{be} = 10^{18}\text{cm}^{-3}$, $N_{bc} = 5 \times 10^{16}\text{cm}^{-3}$. The variable N is the number of iteration. 149

4.3 (a) Normalized injected electron concentration profiles at $V_{BE} = 0.9\ V$, (b) Base transit time vs. base-emitter voltage. For both (a) and (b), Suzuki's uniform and Gaussian bases [193]: $W_b = 100\ nm, N_{be} = 2 \times 10^{18}\text{cm}^{-3}$. $N_{bc} = 2 \times 10^{16}\text{cm}^{-3}$ is used for the Gaussian base. 150

4.4 Base transit time vs. averaged Ge fraction (defined as $y_{\Delta c}/2$ for triangular Ge profile). Patri & Kumar's profiles [185]: $W_b = 60\ nm, R_b = 5k\Omega/\square$. Their results are generated under low injection without considering electric-field dependency of the diffusion coefficient. Rinaldi's uniform base [210]: $N_b = 10^{18}\text{cm}^{-3}$. The triangular symbols for Rinaldi's base are calculated using equations (3) and (5) in [210], which correspond to the low injection and strong high injection, respectively. Kwok's retrograde base [89]: $W_b = 38.5\ nm, x_1 = 6.5\ nm, x_2 = 11.5\ nm, N_{be} = 6.55 \times 10^{18}\text{cm}^{-3}, N_{bc} = 3 \times 10^{17}\text{cm}^{-3}$, $N_p = 9.4 \times 10^{18}\text{cm}^{-3}$ where x_1 and x_2 define the neutral base region with a constant base concentration of N_p. Kwok's results are produced under low injection. 151

4.5 Base transit time and intrinsic base sheet resistance vs. X_T/W_b. J_n =$10^3 A/cm^2$, $R_b = 5k\Omega/\Box$, $y_e = 0$, N_{be} =5×10^{18}cm^{-3}, N_{bc} =9×10^{16}cm^{-3}. Uniform doping profile ($W_b = 29\,nm, y_{\Delta c}$ =0.1724), Gaussian ($W_b = 54\,nm, y_{\Delta c}$ =0.0926), exponential ($W_b = 83\,nm, y_{\Delta c}$ =0.0602), retrograde ($W_b = 26\,nm, y_{\Delta c} = 0.1923, R_p = 13\,nm, N_p$ =10^{19}cm^{-3}). 153

4.6 Base transit time and intrinsic base sheet resistance vs. Ge dose of a trapezoidal Ge profile with $X_T/W_b = 0.68$ for different doping profiles. J_n =$10^3 A/cm^2$, $R_b = 5k\Omega/\Box$, $y_e = 0$, N_{be} =5×10^{18}cm^{-3}, N_{bc} =9×10^{16}cm^{-3}. Uniform ($W_b = 29\,nm$), Gaussian ($W_b = 54\,nm$), exponential ($W_b = 83\,nm$), retrograde ($W_b = 26\,nm, R_p = 13\,nm, N_p$ =10^{19}cm^{-3}). 155

4.7 Base transit time and intrinsic base sheet resistance vs. collector current density for a triangular Ge profile with different doping profiles. $R_b = 5k\Omega/\Box$, $y_e = 0$, N_{be} =5×10^{18}cm^{-3}, N_{bc} =9×10^{16}cm^{-3}. Uniform doping profile ($W_b = 29\,nm, y_{\Delta c}$ =0.1724), Gaussian ($W_b = 54\,nm, y_{\Delta c}$ =0.0926), exponential ($W_b = 83\,nm, y_{\Delta c}$ =0.0602), retrograde ($W_b = 26\,nm, y_{\Delta c} = 0.1923, R_p = 13\,nm, N_p$ =10^{19}cm^{-3}). . 157

4.8 Base transit time and intrinsic base sheet resistance vs. collector current density for a triangular Ge profile with different doping profiles. Plasma-induced bandgap narrowing is considered. $R_b = 5k\Omega/\Box$, $y_e = 0$, N_{be} =5×10^{18}cm^{-3}, N_{bc} =9×10^{16}cm^{-3}. Uniform doping profile ($W_b = 29\,nm, y_{\Delta c}$ =0.1724), Gaussian ($W_b = 54\,nm, y_{\Delta c}$ =0.0926), exponential ($W_b = 83\,nm, y_{\Delta c}$ =0.0602), retrograde ($W_b = 26\,nm, y_{\Delta c} = 0.1923, R_p = 13\,nm, N_p$ =10^{19}cm^{-3}). . 158

4.9 Base transit time and intrinsic base sheet resistance vs. X_T/W_b. J_n $=10^5 A/cm^2, R_b = 5k\Omega/\square$, $y_e = 0, N_{be}$ $=5 \times 10^{18}$cm^{-3}, N_{bc} $=9 \times 10^{16}$cm^{-3}. Uniform doping profile ($W_b = 29\ nm, y_{\Delta c}$ =0.1724), Gaussian ($W_b = 54\ nm, y_{\Delta c}$ =0.0926), exponential ($W_b = 83\ nm, y_{\Delta c}$ =0.0602), retrograde ($W_b = 26\ nm, y_{\Delta c} = 0.1923, R_p = 13\ nm, N_p$ $=10^{19}$cm^{-3}). 159

4.10 Base transit time and intrinsic base sheet resistance vs. X_T/W_b. Plasma-induced bandgap narrowing is considered. J_n $=10^5 A/cm^2, R_b = 5k\Omega/\square$, $y_e = 0, N_{be}$ $=5 \times 10^{18}$cm^{-3}, N_{bc} $=9 \times 10^{16}$cm^{-3}. Uniform doping profile ($W_b = 29\ nm, y_{\Delta c}$ =0.1724), Gaussian ($W_b = 54\ nm, y_{\Delta c}$ =0.0926), exponential ($W_b = 83\ nm, y_{\Delta c}$ =0.0602), retrograde ($W_b = 26\ nm, y_{\Delta c} = 0.1923, R_p = 13\ nm, N_p$ $=10^{19}$cm^{-3}). . 161

4.11 Ratio of base transit time with one effect neglected to base transit time with all effects considered vs. X_T/W_b for a retrograde doping profile in a SiGe base. J_n $=10^5 A/cm^2$, $y_e = 0, y_{\Delta c} = 0.1923, N_{be}$ $=5\times 10^{18}$cm^{-3}, N_{bc} $=9 \times 10^{16}$cm^{-3}, N_p $=10^{19}$cm^{-3},$W_b = 26\ nm, R_p = 13\ nm$. 162

4.12 Ratio of base transit time with one effect neglected to base transit time with all effects considered vs. Ge fraction at the collector edge of a triangular Ge profile for a retrograde doping profile. J_n $=10^5 A/cm^2$, $y_e = 0, N_{be}$ $=5 \times 10^{18}$cm^{-3}, N_{bc} $=9 \times 10^{16}$cm^{-3}, N_p $=10^{19}$cm^{-3},$W_b = 26\ nm, R_p = 13\ nm$. 163

4.13 Base transit time and intrinsic base sheet resistance vs. Ge dose of a trapezoidal Ge profile with $X_T/W_b = 0.67$ for different doping profiles. Plasma-induced bandgap narrowing is considered. $J_n = 10^5 A/cm^2$, $R_b = 5k\Omega/\square$, $y_e = 0$, $N_{be} = 5 \times 10^{18}\text{cm}^{-3}$, $N_{bc} = 9 \times 10^{16}\text{cm}^{-3}$. Uniform ($W_b = 29\ nm$), Gaussian ($W_b = 54\ nm$), exponential ($W_b = 83\ nm$), retrograde ($W_b = 26\ nm, R_p = 13\ nm, N_p = 10^{19}\text{cm}^{-3}$). 164

5.1 Schematic of a cross-sectional view of Si BJT, SiGe HBT, and SiGe HBT with Si amorphization. 171
5.2 Fabrication process flow. 172
5.3 Simulated as-implanted Ge profile (190 keV and $3.5 \times 10^{16}\text{cm}^{-2}$) from PROFILE CODE. 174
5.4 Simulated net vertical doping profile under the emitter of Si BJTs on wafer #3. 175
5.5 Typical measured Gummel plots of Si BJTs. 177
5.6 Typical measured output characteristics of Si BJT. 178
5.7 Typical emitter-base junction characteristics of SiGe HBTs. 179
5.8 Typical base-collector junction characteristics of SiGe HBTs. 180
5.9 Typical measured output characteristics of SiGe HBT with Si amorphization. 181

A.1 A schematic of current flows at the poly-Si/oxide/mono-Si interfaces. . . 188

Chapter 1

Introduction

The insatiable demand for bandwidth in data networks, together with the rapid growth of the wireless communications market, has been one of major driving forces behind the semiconductor industry over the past few years. Although silicon (Si) microelectronics has over 95% of the global semiconductor market [1], it does face challenges in these two growing application areas. For network applications, optical fiber system protocols SONET OC[1]-3 (155 Mbit/s) to OC-192 (9.95 Gbit/s) are in use today [2]. Future systems will operate at 40 Gbit/s and beyond [3]. Implementation of high-speed circuits for these high-bit-rate optical links requires transistors of both high transition frequency (f_T) and high maximum oscillation frequency[2] (f_{max}). Advances in bipolar and CMOS processes, innovative circuit techniques, and careful layout considerations have allowed Si bipolar junction tran-

[1]SONET stands for Synchronous Optical Network and OC stands for optical carriers.

[2]The maximum oscillation frequency f_{max} is the frequency at which the power gain drops to unity and is given by $f_{max} \sim \sqrt{\frac{f_T}{8\pi R_b C_{jc}}}$ where f_T is the transition frequency; R_b, the total base resistance; C_{jc}, the collector-base capacitance in bipolar transistors.

sistors (BJTs) [4–8] and Si CMOS [9–12] to compete with III-V devices[3] [13–16] in realizing these high-speed circuits for systems operating up to 50 Gbit/s. However, for systems above 50 Gbit/s, III-V devices instead of Si devices have traditionally been the technology of choice [17, 18].

For wireless communications, the high-volume RF[4] market includes several low-power applications such as pagers, cordless and cellular systems (0.8–1.9 GHz), wireless local area network chipsets (2.4 GHz), and short-range radio link 'Bluetooth' products (2.4 GHz). Transceiver circuits for the frequency spectrum 0.8–2.4 GHz have been successfully implemented with both Si BJTs [19–22] and CMOS technologies [23–25]. Although GaAs HEMTs[5] offer the best performance in the frequency range 1-5 GHz [26], the trend towards integration for cost reduction and the successful downscaling of CMOS transistors have opened up this low-end RF market for the cost-effective, highly integrated Si CMOS technology. Recently, various transceiver circuits operating at 5 GHz for wireless local area network applications [27–30] have also been implemented with 0.24 and 0.35 μm CMOS processes.

However, microwave and millimeter wave applications such as devices for road pricing and satellite TV transmission (5–10 GHz), satellite for multimedia access (10–20 GHz), sensors for industry, robotic, and environment observation on earth (20–50 GHz), sensors for automobile collision avoidance or radars (50–100 GHz) have been traditionally dominated by III-V devices [31]. The reasons are quite

---

[3]III-V devices are made of elements from both group III and group V in the periodic table.

[4]Here, the term 'RF' (radio frequency) refers to the frequency spectrum from VHF (30-300 MHz) to the S band (2-4 GHz); the term 'microwave' refers to the frequency spectrum from the C Band (4-8 GHz) to the Ka band (26.5-40 GHz); the term 'millimeter wave' refers to the frequency 40 GHz and above.

[5]HEMTs stands for high electron mobility transistors.

simple. Compared with Si devices, III-V devices such as AlGaAs/GaAs HBTs[6] have superior f_T and f_{max} performance. The high f_T is due to the suppression of holes injected back into the emitter (a consequence of the heterojunction effect) and the much higher electron intrinsic mobility in GaAs ($\sim$ 5.3 times higher than that in Si [32]). The high f_{max} is due to the low base resistance associated with a heavily doped base. Any reduction in the current gain due to the high base concentration can be compensated by the enhanced collector current as a result of the heterojunction effect and high electron mobility. Furthermore, with their high breakdown voltage (due to the large bandgap) and low noise figure (due to the low base resistance), III-V materials make good power amplifiers and low-noise amplifiers. The semi-insulating substrate on which III-V devices are fabricated also gives high-quality passive devices essential to wireless applications. Therefore, despite their low levels of integration, low yield, and high cost, with respect to Si technologies, III-V devices have traditionally been the only contender for high-end wireless applications. This parallels the situation for high-speed applications – for optical systems at 50 Gbit/s and below, Si BJTs and CMOS can compete with III-V devices (just as their RF counterpart can compete with III-V devices for frequencies at 5 GHz and below). However, applications at higher bit rates have been traditionally dominated by III-V devices. It is within this context the silicon-germanium heterojunction bipolar transistor (SiGe HBT) has emerged and evolved.

This thesis is about SiGe HBTs, with emphasis on their high-frequency performance. In this introductory chapter, we first present a brief history of SiGe HBTs and some basics on their operation, including their advantages over other competing technologies. Next, we elaborate on the transition frequency (f_T), as

[6]HBTs stands for heterojunction bipolar transistors.

an important figure of merit to measure the high-frequency performance of SiGe HBTs, and map out alternatives for its improvement. Finally, an outline of the thesis is given.

1.1 A brief history of SiGe HBTs

Unlike III-V based heterojunctions where the lattice mismatch between the two materials is minimal, SiGe/Si heterojunctions have to live with a 4.2% lattice mismatch between Si and Ge atoms [33]. This partly explains why SiGe HBTs came much later than III-V based HBTs since the mismatch can pose problems to the growth of a device-quality Si/SiGe hetero-interface.

In 1984, Bean *et al.* [34] showed that SiGe layers of good quality can be grown on Si substrates by molecular beam epitaxy (MBE) over the full range of Ge concentrations at 600°C. The first working MBE-grown SiGe HBTs were reported in late 1987 [35–37] and early 1988 [38–40]. In 1988, Meyerson *et al.* [41] successfully grew strained SiGe layers up to a Ge fraction of 0.2 at a lower temperature (550°C) by ultra-high vacuum chemical vapor deposition (UHV/CVD), a technique that allows multiple wafer processing. Integrating this new technique with the non-self-aligned polyemitter bipolar process from IBM, Patton *et al.* fabricated two SiGe HBTs with base width of $\sim$ 50-75 nm in 1989. One had a Ge grading from 0 to 18% [42] and the other had a grading from 0 to 14% [43]. Both HBTs exhibited ideal junction characteristics and a current gain of 290, compared with a gain of 100 for the Si control BJTs. Around the same time, Fischer *et al.* [44] from IBM reported a UHV/CVD-grown, non-self-aligned polyemitter SiGe HBT with f_T = 45 GHz. It was the fastest published Si-based transistor at that time. The Ge profile was graded from 0 to 11% over a base of 65 nm to achieve the high f_T. Compared with

the Si control device, the current gain was enhanced by ten times. SiGe HBTs fabricated by other techniques[7] have been also reported [45,46].

In the past decade, the performance of SiGe HBTs has drastically improved. Patton *et al.* [47] reported an n-p-n polyemitter SiGe HBT with a f_T of 75 GHz and a BV_{CEO}[8] of 2 V in 1990. The Ge profile was graded from 0 to 7% over a base of 45 nm. The current gain was $\sim$ 135, a factor of two higher than the Si control BJTs. In 1993, Crabbe *et al.* [48] used a new ultra-low thermal cycle process to achieve a more aggressive base profile design (0-25% Ge over a base of 35 nm, base sheet resistance of 7 k$\Omega/\square$). They measured a f_T of 113 GHz, a current gain of 440 (vs. 70 in Si counterpart), and a BV_{CEO} of 2 V for a UHV/CVD-grown, non-self-aligned SiGe HBT with a collector doping of $1.5 \times 10^{18}\text{cm}^{-3}$. IBM announced the first SiGe BiCMOS[9] technology in 1992 [49] and its manufacturable 0.5 μm self-aligned SiGe HBT technology on 200 mm wafers in 1994 [50] (f_T = 32 GHz, f_{max} = 45 GHz, BV_{CEO} = 3.8 V). The technology was then qualified for manufacturing in 1996 [51] and demonstrated SiGe HBTs with f_T = 47 GHz, f_{max} = 65 GHz, and BV_{CEO} = 3.3 V.

SiGe circuits also began to emerge in the 1990's: emitter-common-logic (ECL) ring oscillator circuits [52,53], digital-to-analog converters (DAC) [54], and a variety of RF circuits [31] such as low noise amplifiers, wideband amplifiers, and oscillators. Recently, SiGe HBT-based multiplexers and demultiplexers for optical systems operating above 50 Gbit/s have been announced [55,56]. Wireless circuits based on SiGe HBTs for applications at 5 GHz and beyond have been demonstrated [57–62].

[7] Please see Chapter 5 for more details.

[8] BV_{CEO} stands for the common-emitter breakdown voltage.

[9] BiCMOS stands for bipolar CMOS.

Reference	L_G [μm]	W_E [μm]	f_T/J_c [GHz/$\frac{mA}{\mu m^2}$]	f_{max}/J_c [GHz/$\frac{mA}{\mu m^2}$]	measured @ [V]	BV_{CEO} [V]
CNET/ST [68]	0.35	0.4	45/1.2	60/1	$V_{CE} = 1.5$	3.6
IMEC [64]	0.35	0.3	50/1.5	80/1.5	$V_{CE} = 3$	3.9
Hitachi [67]	0.25	0.2	76/3.5	180/3.5	unknown	2.5
IBM [65]	0.18	0.18	90/4	90/4	$V_{CB} = 0.5$	2.7
NEC [63]	0.18	0.15	73/5.4	61/6.5	$V_{CB} = 1$	2.6
Lucent [66]	0.16	0.28	58/1.06	102/1.06	$V_{CE} = 3$	3.6

Table 1.1: SiGe BiCMOS processes (Note. f_T and f_{max} values correspond to SiGe HBTs).

Today, a number of 200 mm SiGe BiCMOS processes[10] are available (see Table 1.1). Most of them[11] have managed to integrate SiGe HBTs into existing CMOS processes without compromising the performance of either type of transistor. In particular, IBM has already gone through three generations of SiGe BiCMOS processes (0.5, 0.25, and 0.18 μm). Based on IBM's 0.18 μm SiGe BiCMOS process, Freeman *et al.* [65] demonstrated a SiGe HBT with f_T = 90 GHz, f_{max} = 90 GHz, and BV_{CEO} = 2.7 V. The same process also offers a high breakdown-voltage variant with BV_{CEO} = 5.5 V but lower f_T (25 GHz). In 2000, Caroll *et al.* [66] reported a SiGe HBT with f_T = 58 GHz, f_{max} = 102 GHz, and BV_{CEO} = 3.6 V, using Lucent's 0.16 μm SiGe BiCMOS process. Furthermore, using Hitachi's 0.25 μm SiGe BiCMOS process with SOI substrates, Washio *et al.* [67] demonstrated a SiGe HBT with f_T = 76 GHz, f_{max} = 180 GHz, and BV_{CEO} = 2.5 V. All this effort makes SiGe BiCMOS a promising solution for "system-on-a-chip" designs.

[10]Hashimoto *et al.* [63] and Decoutere *et al.* [64] did not specify the wafer size. The other four are 200 mm processes.

[11]Decoutere *et al.* [64] did not detail the performance of their CMOS devices.

1.2 SiGe heterojunction bipolar transistors

Before explaining why a SiGe HBT can meet the demands in both high-speed data networks and wireless communications, it will be useful to review some of the essential features of its operation.

A Si BJT contains two homojunctions: the emitter-base junction and the base-collector junction. They are called homojunctions because both sides of junction are made of the same material, i.e., Si. A SiGe HBT is similar to a Si BJT except that the Si base layer is replaced by a SiGe base. As a result, two heterojunctions are obtained: Si/SiGe emitter-base junction and SiGe/Si base-collector junction. Compared with the Si emitter and collector, the SiGe base has a smaller bandgap. The amount of bandgap narrowing depends primarily on the total Ge dose and secondarily on the base doping level. Because of the lattice mismatch, the SiGe base is strained if no defect dislocations are induced. Lang *et al.* [69] have found that a strained SiGe layer has a even larger bandgap narrowing. According to People and Bean [70], the bandgap narrowing due to the Ge presence in a strained SiGe layer grown on a (100) Si substrate is $0.74y$ eV where y is the Ge fraction in the SiGe layer. This Ge-induced bandgap narrowing has great impact on the common-emitter current gain (β), f_T, and f_{max} of a bipolar transistor. To explain this, let us consider the collector current density of an n-p-n Si BJT[12]:

$$J_c = \frac{q\ e^{V_{BE}/V_T}}{\int_0^{W_b} \frac{N_b(x)\ dx}{D_n(x) n_{ie}^2(x)}} \tag{1.1}$$

where V_{BE} is the base-emitter voltage; V_T, the thermal voltage; W_b, the neutral base width; N_b, the base concentration; D_n, the electron diffusion coefficient; n_{ie},

[12]For brevity, we assume low level injection and 'zero' electron concentration at the base-collector junction. Please see Chapter 2 for the derivation of a more general expression.

the effective intrinsic carrier concentration; q, the electronic charge. In the case of a Si base, the effective intrinsic carrier concentration n_{ie} is simply the temperature-dependent constant n_i (1.08×10^{10} cm^{-3} at room temperature [71]). If the base layer is made of SiGe, the effective intrinsic carrier concentration in the base becomes[13]:

$$n_{ie}^2 = n_i^2 e^{\Delta E_{g,Ge}/(kT)} \tag{1.2}$$

where $\Delta E_{g,Ge}$ is the Ge-induced bandgap narrowing; k, the Boltzmann constant; T, the temperature. For a uniform Ge profile across the neutral base, the collector current density in (1.1) will be enhanced by a factor $exp[\Delta E_{g,Ge}/(kT)]$. For a Ge fraction of 0.1 (i.e. 10 atomic %), the bandgap narrowing $\Delta E_{g,Ge}$ will be 0.074 eV and the enhancement factor at room temperature will be 17.6. In other words, for the same V_{BE}, a SiGe base with a box Ge profile at 10% will increase the collector current density by more than one order of magnitude (even higher at lower temperatures)! This is called the heterojunction effect[14]. Since the base current, to first order, remains unchanged for the same V_{BE}, an enhancement in J_c will translate to an enhancement in the current gain (β).

From another perspective, the effect of the Ge-induced bandgap narrowing can be considered a reduction in V_{BE} required for obtaining the same collector current density. For example, in a Si BJT, a V_{BE} of 0.75 V is required to obtain a collector current density J_{c1}. By having a SiGe base with a box Ge profile at 10%, a V_{BE}

[13]For illustrative purposes, we neglect heavy doping effects, the Fermi-level shift due to the use of Fermi-Dirac statistics, and the modification of the effective densities of states by the strain. Please see Appendix C for details of these effects.

[14]To be precise, the Ge-induced bandgap narrowing is only one of the two requirements for the heterojunction effect. The other requirement is that the band discontinuity at the Si/SiGe heterointerface, caused by the narrower bandgap in the base, appears mostly in the valence band instead of the conductor band [69]. Therefore, the enhanced electron current flow due to the Ge-induced bandgap narrowing is not hindered by any energy barrier in the conduction band.

of 0.676 V (= 0.75 − 0.074) is now required for driving the transistor at the same current density J_{c1}. This reduction in V_{BE} has a tremendous effect on the total minority charges in the neutral emitter and consequently on the value of f_T. The reason is as follows. According to the law of the junction, the minority carrier concentration at the edge of the depletion layer, on the *emitter* side, is exponentially dependent on the forward bias voltage, V_{BE}. Therefore, at the same J_{c1}, the minority carrier concentration at the depletion layer edge on the emitter side will be decreased by a factor of 17.6 as the V_{BE} for obtaining the same J_{c1} has been reduced by 0.074 V in our example.

By definition, the emitter delay time is equal to the total minority carrier charges in the emitter divided by the collector current density. A reduction in the minority carrier concentration at the depletion layer edge, on the emitter side, will reduce the emitter delay time by the same factor. Since the maximum transition frequency f_T is inversely proportional to the sum of the delay times in different regions of the transistor, a reduction in the minority carrier concentration can translate to an enhancement in f_T. In our example of SiGe HBT with 10% Ge, if we assume that f_T is entirely dominated by the emitter delay time, f_T will be enhanced by a factor of 17.6 for the same current J_{c1} ! Since f_{max} is proportional to the $\sqrt{f_T}$, the enhancement in f_T will translate to an enhancement in f_{max} by a factor of 4.2 for the same base doping and base width.

However, what if f_T is not dominated by the emitter delay time but the base delay time? In this case, there are at least two approaches to make use of a Ge profile to improve the high-frequency performance of the Si BJT. The first approach is to trade the current gain enhancement, due to the heterojunction effect, for an increase in f_T and f_{max}. This can be done by reducing the base width because a thinner base yields a higher f_T that is dominated by the base delay time. The drawback of a

thin base is a large intrinsic base resistance that degrades f_{max}. Therefore, the base concentration has to be significantly increased in order to reduce the intrinsic base resistance to a level such that either the same or a higher f_{max} is obtained. Overall, f_T and/or f_{max} become higher at the expense of a lower current gain enhancement[15] or, in some cases, a lower current gain[16]. In addition, the Early voltage and noise figures are improved by the increased base concentration and lower base resistance, respectively.

A slight modification of the first approach is to trade the current gain for a higher f_{max} only without changing the base width. The figure f_{max} is increased only by a reduction in the base resistance that comes from a higher base concentration. The current gain, obtained from the heterojunction effect, can offset the reduction due to the increased base concentration. Although the emitter delay time will be different as the current gain is changed, f_T remains practically the same as it is dominated by the base delay time. Similar to the case where f_T is dominated by the emitter delay time, the key to improvement by the first approach is the heterojunction effect.

The second approach for making use of Ge is to grade the base Ge profile in a way that enhances the electron transport across the base. The Ge fraction is increased from a low value near the emitter to a higher value near the collector. The resulting bandgap narrowing gradient induces an aiding electric field in the base and consequently reduces the base delay time. As f_T is dominated by the base delay time, a lower base delay time translates to a higher f_T and, in turn, a higher

[15]The current gain enhancement is reduced because the factor by which the base concentration is increased should be larger than the factor by which the base width is reduced if a lower intrinsic base resistance is required. Please see (1.1) for the relation between the base width, the base concentration, and the current gain.

[16]In digital applications, high f_T and f_{max}, but not very high β, are required.

f_{max} for the same base width and concentration. Another benefit of the positive Ge concentration gradient is a high Early voltage, which is important for analog circuit designs. In short, the key to improvement by this approach is the Ge grading.

The questions of which is a better approach and when f_T is dominated by the base delay time, are important and often debated ones[17]. Obviously, if a very large or unlimited amount of Ge can be incorporated into the base, these questions become unimportant as one can adopt either one or both approaches to the fullest extent. However, this cannot be the case in a SiGe HBT because of the inevitable lattice mismatch between the Si and Ge atoms. When the accumulated strain due to the lattice mismatch is too large (as more Ge atoms are added into the base), it is more energetically favorable for the strain to be accommodated by forming defect dislocations, which severely degrade the device performance. Although it has been shown that the critical Ge dose[18] can be raised by growing the SiGe layer at lower temperatures, the layer will become metastable and may still relax upon subsequent thermal processing [33].

For conventional polysilicon-contacted processes where an 850°C furnace anneal is followed by a rapid thermal annealing step at 900-1000°C, a metastable SiGe layer will still stand a chance of dislocation formation if the Ge dose is too high. In state-of-the-art SiGe BiCMOS processes (e.g. 0.25 and 0.18 μm) where the CMOS thermal cycle can be even higher than the bipolar one [72], issues concerning the

---

[17]These questions can be phrased differently in terms of determining the optimal Ge profile, e.g., "Is a box Ge profile better than a graded Ge profile?" In some way, part of the work in this thesis has been motivated by these questions.

[18]The critical Ge dose is defined as maximum number of Ge atoms per unit area that can be incorporated into the layer without introducing defect dislocations during the growth. This parameter is closely related to the more popular measure, critical thickness. In practice, the critical Ge dose or critical thickness is a function of a number of process variables such as growth temperature, wafer preparation conditions, contamination control, etc., in a particular SiGe growth process.

SiGe film stability are even more crucial. Thus, it becomes important to be able to make the *best* use of the limited amount of Ge that is allowed in a given process. In other words, an optimal vertical profile design can have great impact on the high-frequency performance of SiGe HBTs in polysilicon-contacted bipolar processes or SiGe BiCMOS processes [65].

1.3 Why SiGe HBTs?

Why choose SiGe HBTs over Si BJTs, Si CMOS, and III-V devices? As mentioned above, the improved performance of today's Si BJTs and CMOS has allowed them to compete with III-V devices in applications such as high-speed optical systems operating at 50 Gb/s and below, and the low-power RF market of 5 GHz and below. The continuing improvement in Si process technologies will likely enable them to expand into applications of higher speeds and frequencies in near future [73]. Nevertheless, SiGe HBTs still have distinct advantages over Si BJTs and CMOS and consequently are very competitive in these applications.

For low-power RF applications, transistors are not always biased at current densities for maximum f_T or f_{max}. For example, in designing low-noise amplifiers, the bias current is determined by meeting the requirement for low noise figure and low power dissipation. The transistor is often biased at as low a current density as possible, still consistent with the acceptable noise figure performance [74], in order to reduce power consumption[19] and avoid a unnecessarily high gain that degrades the overall linearity of a receiver. In view of this, a SiGe HBT can perform better than a Si BJT or CMOS because it can achieve a higher f_T even at the same low

[19]Considering that a low-noise amplifier is constantly "on" to detect incoming signals, reducing power consumption is a valid concern.

current density. Since the minimum noise figure is inversely proportional to f_T, a higher f_T at low current operations improves the noise performance. A lower noise figure also comes from the lower base resistance in a SiGe HBT when compared with a Si BJT. Furthermore, SiGe HBTs can achieve an improvement of two orders of magnitude lower than Si CMOS in terms of 1/f noise [1]. This makes SiGe HBTs a better choice for implementing mixer and oscillator circuits as 1/f noise up-converts to phase noise in these circuits.

More importantly, with respect to Si BJTs or Si CMOS, one can obtain the same f_T or f_{max} by biasing SiGe HBTs at a lower current density [74]. In other words, the high-frequency or speed performance of SiGe HBTs can be traded off for a reduced power dissipation in many applications. This is of particular importance to portable applications where the power consumption is a major concern. Furthermore, for power amplifiers, SiGe HBTs have better power added efficiency (PAE) when compared with Si CMOS because of the exponential nature of the $J_c - V_{BE}$ relationship in SiGe HBTs [75]. In addition, SiGe HBTs have higher transconductance, i.e., higher driving capability, than Si CMOS for the same current. This gives SiGe HBTs advantages over Si CMOS in realizing high-speed circuits.

What about III-V devices? There is no doubt that III-V devices outperform SiGe HBTs in terms of f_T and f_{max} for the same base width. However, SiGe HBTs can still compete with III-V devices in optical systems above 50 Gb/s or microwave and millimeter wave applications. The strength of SiGe HBTs over III-V devices is mainly the ability of SiGe HBTs to integrate with the standard low-cost CMOS process (i.e. SiGe BiCMOS). High levels of integration always implies lower overall system power consumption because less power is spent on routing high-frequency signals on and off chips [76]. Moreover, in achieving the same f_T, a lower current density is required in the case of SiGe HBTs than some GaAs-based devices [31].

This again translates to lower power consumption. The cost advantage, as a result of the higher yield, larger wafer size, and higher levels of integration, also makes SiGe a better candidate for applications wherever both technologies can meet the design specification.

Apart from power dissipation and cost, SiGe HBTs have other advantages over III-V devices. For example, SiGe HBTs with their smaller base width than that of III-V HBTs have lower noise figure [77]. Also, they exhibit better 1/f noise behaviour as a result of their higher quality surface passivation [77] (reflected by the lower β_{max}/β ratios in their Gummel plots). This enables oscillators with low phase noise to be made [31]. In addition, the lower turn-on voltage in SiGe HBTs[20] renders them more suitable for low-voltage circuits in portable applications.

To summarize, the main advantage of SiGe HBTs over Si BJTs and CMOS is lower power consumption at the expense of high-frequency performance. While the main advantages of SiGe HBTs over III-V devices are lower cost and lower system power consumption due to higher yield and higher levels of integration. These advantages enable SiGe HBTs to be very competitive in application areas such as high-speed circuits for optical networks and low-power wireless communications. However, for the SiGe HBT technology to stay competitive, several aspects of the technology, especially in the context of SiGe BiCMOS processes, need to be improved. These are high-quality passive components[21], low-loss transmission lines and related substrate effects, device parasitics, integration with future generations of CMOS processes, and vertical profile design. For vertical profile design, achiev-

[20]The only exception is InP-based HBTs, which have been shown to have a lower turn-on voltage than SiGe HBTs [78].

[21]Most recently, using Hitachi's SiGe BiCMOS process on SOI wafers, Washio *et al.* [67] have demonstrated MIM capacitor of 0.7 fF/μm^2 with a Q factor of 13 at 10 GHz and a spiral inductor of 0.9 nH with a Q factor of 20 at 10 GHz.

ing higher f_T, f_{max}, and breakdown voltages are some immediate concerns at the present time. A major part of this thesis attempts to address vertical profile design issues specific to the high-frequency performance of SiGe HBTs.

1.4 Transition frequency

The figures of merit, f_T and f_{max}, are frequently used to measure the high-frequency performance of SiGe HBTs. The two are closely related to each other even though a high f_T does not guarantee a high f_{max} or vice versa. A high collector-base capacitance or a high intrinsic base resistance (due to the use of a small base width without high enough base concentration to increase f_T) may degrade f_{max}. As the vertical dimensions of the transistor (e.g. the emitter depth, the base width, epitaxial collector width) are smaller, the extrinsic or parasitic components (e.g. extrinsic base resistance and extrinsic base-collector capacitance, collector-substrate capacitance, etc.) become more important. Since f_{max} incorporates some of these parasitic components, in particular, the extrinsic base resistance and the extrinsic base-collector capacitance, it is sometimes considered a more important measure than f_T for power gain in small signal and large signal amplifiers, wideband analog amplifiers, and, to a certain extent, nonsaturating logic gates [79]. However, the significant improvement of advanced bipolar and BiCMOS technologies in the past two decades (e.g. polysilicon emitter, self-aligned technology, silicide base contact, deep and shallow trench isolation, etc.) have also successfully reduced these extrinsic components. As a result, f_T is still a meaningful measure of the high-frequency performance of SiGe HBTs.

Based on the quasi-static approach, the transition frequency f_T of a bipolar transistor can be expressed in terms of a number of delay time constants. Each

of them is associated with the charging/discharging of a particular region in the transistor through the collector current in response to a small signal base-emitter voltage, v_{BE}. By definition, f_T is associated with the common-emitter configuration where the collector current is measured at the output port and the base-emitter voltage is applied at the input port. Therefore, each delay term is equal to the change in the amount of charges in the respective region through the collector current divided by the change in the collector current density in response to v_{BE}. The f_T expression can be written as follows [80]:

$$\frac{1}{2\pi f_T} = \frac{C_{je} + C_{jc}}{g_m} + \tau_e + \tau_b + \tau_{bc} + \tau_{be} + C_{jc}R_c \tag{1.3}$$

where C_{je} and C_{jc} are the depletion layer capacitances of the emitter-base junction and the base-collector junction, respectively. The quantity g_m is the small signal transconductance (I_c/V_T). The quantity R_c is the collector series resistance between the collector-base depletion layer and the buried collector layer. To be precise, as explained in Chapter 2, the contribution of the terms τ_b and particularly τ_e to f_T is more complicated than shown in (1.3) because of the non-quasi-static effects at high frequencies. However, for discussion, (1.3) is used as an approximation.

When the base-emitter voltage is modulated by v_{BE}, the depletion layer widths of the emitter-base junction and the base-collector junction will change and the amount of localized charges in the depletion layers will vary accordingly. The first term $\frac{C_{je}+C_{jc}}{g_m}$ is, therefore, the time required for changing these localized charges at the edges of the depletion layers through the collector current in response to v_{BE}. Because of the g_m in the denominator, the first delay term is expected to decrease with increasing J_c.

The v_{BE} modulation also changes the amount of minority free carriers in the

neutral base and emitter regions because the minority carrier concentrations at the depletion layer edges of the emitter-base junction is a function of the base-emitter voltage. Thus, the term τ_e (τ_b), called the emitter (base) delay time, is the time taken for changing the amount of minority free carriers in the neutral emitter (base) regions through the collector current in response to v_{BE}. Although τ_b is also called the base transit time, it *does not* measure the time for a minority carrier to travel across the base. As Varnerin [81] explained, τ_b is called the transit time because the ratio of stored charges to current is the average time spent per carrier in the base. In the low-injection regime, both τ_e and τ_b are not strongly dependent on the collector current but rather on the profile data of the transistor. This is further discussed in Chapter 2.

The term τ_{bc} is the collector-base depletion layer delay time. When V_{BE} is modulated by v_{BE}, J_c will change. Since the concentration of *free carriers* in the collector-base depletion layer on the collector side is $\sim J_c/(qv_s)$ (where v_s is the saturation velocity)[22], the amount of free carriers in the collector-base depletion layer will vary with v_{BE}. Therefore, τ_{bc} measures the time taken for changing the amount of *free carriers* in the collector-base depletion layer through the collector current in response to v_{BE}. Meyer and Muller [82] have shown that τ_{bc} is independent of J_c at low-level injection and is equal to $W_d/(2v_s)$ where W_d is the width of the collector-base depletion layer on the collector side. To be precise, the τ_{bc} expression $W_d/(2v_s)$ accounts for the free carrier change in the collector-base depletion layer *on the collector* side only. However, since the depletion layer width on the base side is generally much smaller than that on the collector side as the base

[22]Here we assume that the collector-base junction is under a reverse bias and the transistor operates before onset of Kirk effect. Therefore, the electric field at the collector-base junction is large enough to drive the carrier through at v_s. Furthermore, we assume no velocity overshoot in the discussion, i.e., no rapid spatial change of the electric field.

doping is typically higher than the collector doping by roughly one to two orders of magnitude, the expression $W_d/(2v_s)$ is a good approximation for τ_{bc}.

The term τ_{be} is the emitter-base depletion layer delay time. It measures the time required for changing the *free carriers* in the emitter-base depletion layer in response to v_{BE}. According to Negus and Roulston [83], the effect of τ_{be} can be taken into account by multiplying C_{je} with a factor F, which is a function of V_{BE}. At low currents, F is approximately 1. At high forward biases, F is approximately 2.5. The last term $C_{jc}R_c$ is required because the collector current flows through R_c out of the output port where the capacitance C_{jc} is connected. This RC time constant is often called the collector charging time.

Let us look at two simulation studies on how these delay terms contribute to f_T at different collector current densities in a Si BJT. Roulston and Hebert [84] simulated a polyemitter BJT with a collector doping of $1.36 \times 10^{16}\text{cm}^{-3}$, base doping of $\sim 5.5 \times 10^{18}\text{cm}^{-3}$, emitter depth of 0.145 μm, metallurgical base width of 0.27 μm, epitaxial collector width of 0.7 μm. At a low current density ($J_c \sim 3.7 \times 10^3$ A/cm^2), they found that the term $\frac{C_{je}+C_{jc}}{g_m}$ was 14% of the total delay time, and τ_e and τ_b were around 25% and 33%, respectively. The terms τ_{bc} and τ_{be} were 9% and 18%, respectively. At a higher J_c ($\sim 1.5 \times 10^4$ A/cm^2) before the onset of Kirk effect, the term $\frac{C_{je}+C_{jc}}{g_m}$ dropped to 7%, while τ_e and τ_b increased to 35% and 39%, respectively. The terms τ_{bc} increased slightly to 11% and τ_{be} dropped to 7%. The term $C_{jc}R_c$ remained around 1.3% at both current densities. After the onset of Kirk effect ($J_c \sim 2.2 \times 10^4$ A/cm^2), τ_e and τ_b continued to rise and other terms (except $C_{jc}R_c$) dropped with increasing J_c.

Another simulation study was performed by Chen *et al.* [85] for a scaled Si BJT with emitter depth of 25 nm, metallurgical base width of 60 nm, and epitaxial collector width of 0.3 μm. The emitter doping was 10^{21}cm^{-3}. The base profile was

graded from $1.5 \times 10^{19}\text{cm}^{-3}$ at the emitter-base junction to 10^{17}cm^{-3} at the base-collector junction. The epitaxial collector doping was 10^{17}cm^{-3}. At a low current density ($J_c \sim 6 \times 10^4$ A/cm^2), the term $\frac{C_{je}}{g_m}$ was 31% of the total delay time whereas the term $\frac{C_{jc}}{g_m}$ was only 1%. The terms τ_e and τ_b were $\sim$ 5% and 27%, respectively. The term τ_{bc} was $\sim$ 35%. At a higher J_c ($\sim 7.5 \times 10^4$ A/cm^2) before the onset of Kirk effect, the term $\frac{C_{je}}{g_m}$ dropped to 28%, while $\frac{C_{jc}}{g_m}$ remained at 1%. The term τ_e remained unchanged and τ_b increased very slightly to 28%. The term τ_{bc} increased to 38%. After the onset of Kirk effect ($J_c \sim 1.5 \times 10^5$ A/cm^2), τ_b rose rapidly and other terms dropped with increasing J_c.

These two studies show the following trends: the terms $\frac{C_{je}}{g_m}$ and τ_{be} are large at low J_c and then drop when J_c is increased. In the current range from low to high (before the onset of Kirk effect), the terms τ_e and τ_b contribute to more than half of the total delay time in a unscaled BJT and around one third in a scaled BJT. Both increase slowly with J_c in this range and τ_b rises rapidly after the onset of Kirk effect. The term τ_{bc} changes slightly (within $\sim$ 10%) before the onset of Kirk effect. It contributes to $\sim$ 10% of the total delay time in a unscaled BJT and around one third in a scaled BJT. The contribution from $\frac{C_{jc}}{g_m}$ and $C_{jc}R_c$ is relatively small when compared with other terms. Overall, $\frac{C_{je}}{g_m} + \tau_{be}$, $\tau_e + \tau_b$, and τ_{bc} are three most important terms. Each of them contributes to roughly one third to f_T before the onset of Kirk effect in a scaled device. Now, the question is: given this information, how do we make use of the Ge and doping profiles to minimize the total delay time? In other words, how do we make use of the Ge and doping profiles to maximize f_T?

According to these trends, at least three types of vertical profile optimization for minimizing the total delay time can be attempted. The first type is the optimization of the emitter and base profiles in the emitter-base depletion layer. This

will minimize $\frac{C_{je}}{g_m}$ and τ_{bc} in the low- and high-injection regimes for scaled devices and in the low-injection regime only for unscaled devices. The second type is the base profile optimization for minimizing τ_b and/or τ_e in the low- or high-injection regime as both delay times are mainly determined by the base profiles (both doping and Ge) before the onset of Kirk effect. The third type is the collector profile optimization [86–88] for minimizing τ_b and/or τ_{bc} above the Kirk current density as these delay times are mainly determined by the doping and Ge profiles in the collector-base depletion layer and in the collector after the onset of Kirk effect. The delay time τ_{bc} should be minimized also at current densities before the onset of Kirk effect. In addition, this type of optimization should maximize BV_{CEO} because the collector doping profile has great influence on BV_{CEO}. A major portion of this thesis focuses on the second type, the base profile optimization.

One final remark on setting optimization constraints for the vertical profile design studies needs to be mentioned. When the shape and the concentration of the doping profile and/or Ge profile are varied during optimization, other important figures of merit or properties of the transistor will be affected. For example, when the width or the concentration of the base profile is changed, the intrinsic base resistance will be different. That is, even though the optimization may lead to a base doping profile which gives the minimum τ_b, the same profile may also cause a high base resistance and consequently a very low f_{max}. Or, it may be found that a box Ge profile, instead of a graded Ge profile, yields a minimum τ_b. However, this result might not be meaningful because if the total Ge dose of graded profile is much higher than that of the box profile, the SiGe HBT with a box profile will be more prone to strain relaxation in subsequent thermal steps. Therefore, for useful optimization results, crucial parameters, such as the intrinsic base resistance, total Ge dose, etc., should be used as constraints when optimizing profiles for minimum

delay times.

1.5 Outline of the thesis

This thesis is composed of studies on the optimization, analysis, and fabrication of SiGe HBTs. A portion of the work presented here has been published [89, 90].

Chapter 2 is a base profile optimization study for minimizing the emitter delay time and base delay time in the low-injection regime. It focuses on how factors, such as emitter dimension, base width, base doping, and base profile shape, alter the performance of different Ge profiles in minimizing the contribution of these two delay times to the transition frequency. A graded Ge profile leverage factor is proposed to compare the effectiveness of Ge profiles of identical dose in minimizing these delay times for each combination of emitter and base doping profiles. For the first time, non-quasi-static effects on the emitter delay time at high frequencies are taken into account in an optimization study of SiGe HBTs.

In Chapter 3, we analyze the effect of the retrograde portion of a base doping profile on the base delay time for a SiGe HBT with a graded Ge profile operating in the low-injection regime. We present for the first time a closed-form analytical base delay time expression which considers the retarding built-in electric field due to the retrograde region, heavy doping effects, the effect of velocity saturation at the collector-base junction, the concentration and field dependencies of the electron diffusivity, and the electric field caused by the Ge-induced bandgap narrowing. We then use this expression to assess the relative importance of these effects in calculating the base delay time. The result of this assessment is useful for modeling purposes.

Chapter 4 is devoted to the base profile optimization for minimizing the base

delay time in both low- and high-injection regimes before the onset of Kirk effect. We present a new iteration scheme which allows the calculation of the base delay time in a wide range of collector current densities. The scheme is then used to determine the optimal set of base doping and Ge profiles for minimizing the base delay time in both regimes of injection. We consider, for the first time in a profile optimization study, the effect of the plasma-induced bandgap narrowing on the base delay time in the high-injection regime.

Chapter 5 describes the fabrication of SiGe HBTs by high-dose Ge implantation, Si amorphization, and solid-phase epitaxy. The impetus behind this attempt is to enhance the ability of SiGe HBT technologies to fully integrate with existing Si CMOS processes. The implantation technique, as a conventional doping step in modern Si processes, not only facilitates integration but also allows *multiple* Ge profiles to be optimized for different applications on the same chip. We present, for the first time, the results of electrical measurements on HBTs fabricated by this technique.

Finally, conclusions of the thesis findings and contributions are given in Chapter 6.

Chapter 2

Ge profile optimization for minimum emitter and base delay times

2.1 Introduction

Since the heterojunction bipolar action of SiGe base HBTs was demonstrated [35], researchers have begun to look for a Ge profile in the base that enhances device properties, in particular, the transition frequency (f_T). In general, two main but different results of this search have been proposed [91, 92]: i) a graded Ge profile, and ii) a box Ge profile. The graded profile is advocated mainly by researchers in IBM, Siemens, NEC, Hitachi, etc. On the other hand, the box profile is favored by researchers from Daimler-Benz, TEMIC, Philips, etc. Although this search has been an ongoing effort for almost a decade, it is still not clear which proposal is better.

Among different figures of merit of SiGe HBTs, the transition frequency f_T has received the most attention for improvement. For example, early in 1990, Patton *et al.* [93] have recognized that in order to maximize f_T, a Ge profile with high Ge content at the emitter-base junction should be used when the emitter delay time (τ_e)[1] dominates the contribution to f_T (as a higher bandgap narrowing can be obtained to increase the current gain and consequently reduce τ_e). On the other hand, when the base delay time (τ_b) dominates, a Ge profile with a large grading should be used to create an aiding electric field to enhance the electron transport. However, the Ge dose in the base was not kept constant when the impact of different Ge profiles on the various time delay terms was investigated in their study. As a result, the Ge profiles were not fairly compared. Similarly, Zhang *et al.* [88] favored the graded profile over the box profile but the Ge dose was not kept constant in their study either. Harame *et al.* [94] derived analytical expressions for SiGe-to-Si ratios of τ_b and τ_e for a uniformly doped base with Ge profiles of different shapes but the same Ge dose. However, they only considered SiGe HBTs with conventional polyemitter contacts, where f_T was assumed to be entirely dominated by τ_b, when comparing different Ge profiles. As expected, they concluded that the graded Ge profile is more effective than the box Ge profile in maximizing f_T.

More recently, Hueting *et al.* [95] emphasized the interplay between the effective base Gummel number, the collector current (or the current gain), and the emitter-base depletion layer delay time τ_{be}. They concluded that it is the base Gummel number, not the Ge grading, which determines the maximum f_T. Therefore, the box Ge profile should be preferred. Using numerical simulations, Richey *et al.* [96] also showed that the box Ge profile is more optimal, since the emitter delay time

[1]The emitter delay time τ_e should be distinguished from the emitter transit time $(t_{e,transit})$, which is defined as the emitter delay time multiplied by the injection-limited DC common-emitter current gain.

becomes important in a *scaled* device. In short, both studies point back to the necessity of not only minimizing the base delay time but also maximizing the current gain as the latter affects the emitter delay time and the emitter-base depletion layer delay time. This is especially true when the base width is further scaled. In fact, a similar approach was taken earlier by Roulston and McGregor [97]. Considering a SiGe base without neutral base recombination and with a fixed Ge dose, they concluded that a box Ge profile would give both a maximum current gain and a minimum emitter delay time. However, they did not report the absolute magnitude of the emitter and base delay times corresponding to different Ge profiles. Only normalized values were given. This makes the comparison of the *sum* of delay times impossible. It should be noted that a box Ge profile which gives the minimum emitter delay time does not necessarily minimize the sum of the base and emitter delay times.

Unlike Roulston and McGregor, Hueting *et al.* and Richey *et al.* showed the effect of different Ge profiles on the final maximum f_T value. Also, they simulated on a more realistic (non-uniform) base doping profile. Their base profiles are more realistic in that when the base is becoming narrower and more heavily doped, the more the base profile deviates from an ideal uniform shape, due to the dopant out-diffusion during subsequent thermal steps in the fabrication. However, one limitation of both studies is that only positively-graded Ge profiles were compared with the box Ge profile. In contrast, Roulston and McGregor included Ge profiles of both positive and negative gradients, even though their work dealt only with a uniformly-doped base. Considering negative Ge gradients is useful because when the total delay time of a transistor is dominated by τ_e, a negative Ge gradient may be used to give the highest possible bandgap narrowing for a given Ge dose, i.e., resulting in a much higher β and a much lower τ_e, and consequently a higher overall

f_T. Another drawback of Hueting *et al.* 's study is that the emitter delay time was not considered.

One common limitation of all the studies mentioned above is that they made the quasi-static approximation when estimating f_T. For transistors with large base width, the base transit time likely dominates in the f_T calculation and the quasi-static approximation will not be of concern. But, when the base width is small and the emitter delay time becomes more significant, the emitter charge partitioning as a non-quasi-static phenomenon should be considered; otherwise, the emitter delay time may be over-estimated and a Ge profile that favors a smaller emitter delay time can be unfairly considered more effective at maximizing f_T.

From this brief review, a few criteria can be summarized for improvement on similar studies. Firstly, the Ge dose of the compared Ge profiles should be identical. Secondly, the sum of the emitter and base delay times instead of the individual normalized delay times should be calculated and compared. Thirdly, both uniform and non-uniform base doping profiles should be considered. Fourthly, both positive and negative Ge gradients should be included in the comparison. Lastly, non-quasi-static effects must be considered, especially in the case of scaled devices. To our knowledge, no studies meeting all these criteria have been reported.

The primary objective of this chapter is not to determine the optimal Ge profile for maximizing f_T. Rather, the purpose is twofold. First, we try to map out important factors which impact the performance of the Ge profile in minimizing the contribution of the emitter and base delay times to the f_T of a SiGe HBT. Second, we attempt to optimize the Ge profile for the minimum contribution of the emitter and base delay times to f_T, for different combinations of emitter and base profiles, in the low injection regime. It is believed that these efforts can deepen understanding of and provide insights on the bigger problem of seeking the optimal

Ge profile for maximizing f_T.

Section 2.2 details the theoretical tools that are required to carry out the optimization. They include the profile definitions, the derivation of expressions for the base and emitter delay times, and the method of estimating the sum of the delay times with non-quasi-static correction. Section 2.3 verifies the expressions presented in 2.2 by comparing them with published transit time models in the literature and with numerical simulation results. Section 2.4 then compares Ge profiles in the SiGe base over different emitter and base doping profile parameters. Finally, conclusions are given in section 2.5.

2.2 Theory

2.2.1 Profile definitions

A typical doping profile for this study is shown in Fig. 2.1. This high-low emitter structure, adopting the principle of doping inversion, is believed to become increasingly important because a higher base doping will be needed to keep the base resistance at a reasonable level and to avoid base punch through when the base width is reduced. Consequently, the concentration of the mono-Si emitter must be reduced so that both the tunneling current and the junction capacitance can be kept within an acceptable level.

The emitter profile is divided into three regions: the uniformly doped poly-Si region, the Gaussian-doped transition region from poly-Si to mono-Si, and the uniformly doped mono-Si region. The transition region can be described as follows:

$$N_{et}(x) = N_{ep}\ exp\left[-\left(\frac{x - W_{ep}}{x_e}\right)^2\right] \qquad , W_{ep} < x < W_{ep} + W_{em1} \qquad (2.1)$$

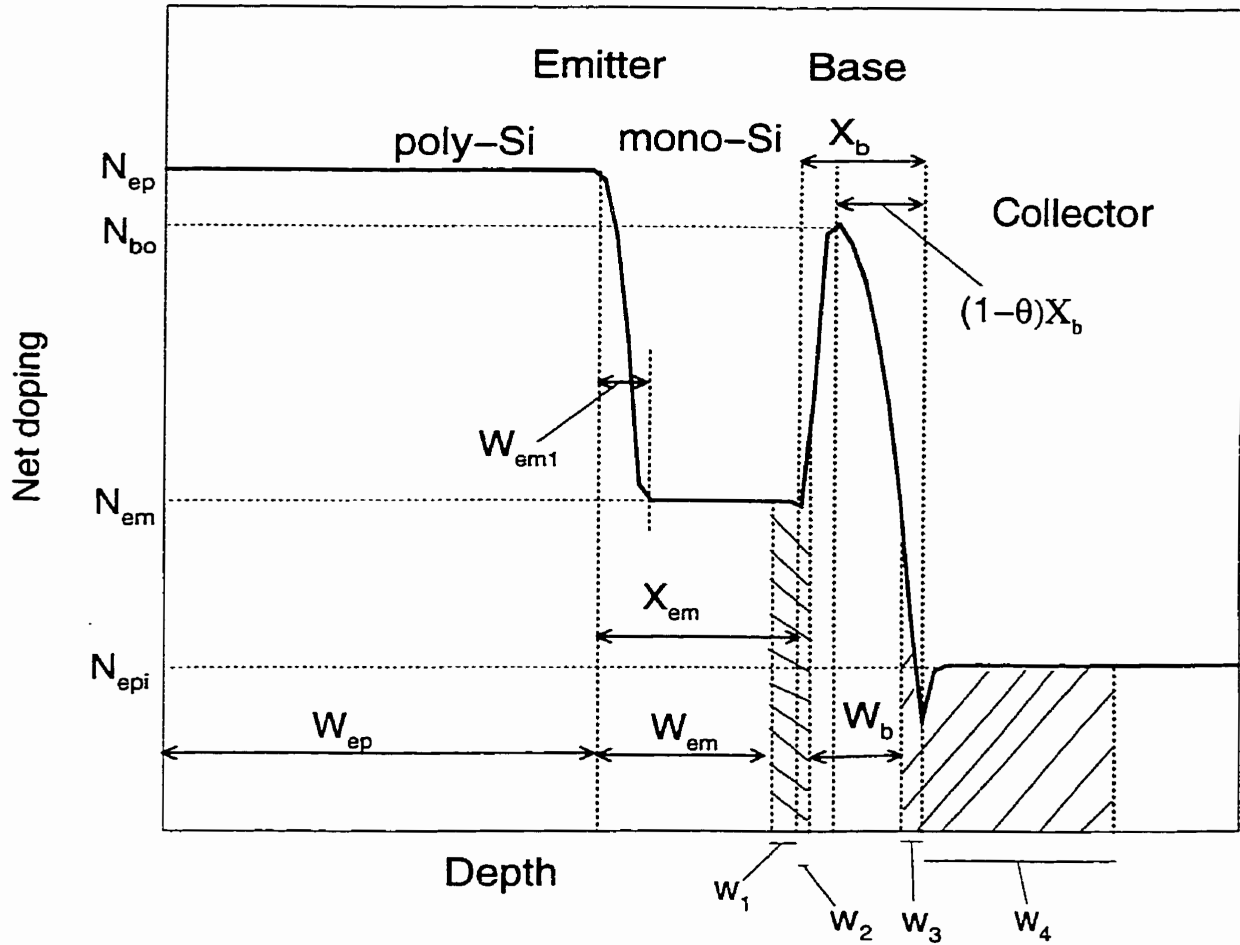

Figure 2.1: A schematic of the net doping profile under study (Note. The doping concentration is drawn in a logarithmic scale).

where

$$x_e = \frac{W_{em1}}{\sqrt{ln\left(\frac{N_{ep}}{N_{em}}\right)}} \tag{2.2}$$

The constants N_{ep} and N_{em} are the concentrations of the poly-Si and the mono-Si, respectively. W_{ep} and W_{em1} are the widths of the poly-Si region and the transition region, respectively. In Fig. 2.1, X_{em} is the metallurgical width of the mono-Si region, which is the mono-Si neutral region (W_{em}) plus the depletion layer width of the emitter-base junction on the emitter side (w_1).

The base profile consists of two regions: a retrograde region and a tail region. The retrograde region is often used to reduce the base concentration near the emitter and hence the tunneling leakage current and the emitter-base junction capacitance can be lowered [93]. However, it retards the electron transport by inducing a retarding electric field. The impact of this retrograde region is studied in detail in Chapter 3. The retrograde base region ($W_{ep} + X_{em} < x < W_{ep} + X_{em} + \theta X_b$) is described as:

$$N_{b1}(x) = N_{bo}\ exp\left[-\left(\frac{x - W_{ep} - X_{em} - \theta X_b}{x_{b1}}\right)^2\right] - N_{em} \tag{2.3}$$

where

$$x_{b1} = \frac{\theta X_b}{\sqrt{ln\left(\frac{N_{bo}}{N_{em}}\right)}} \tag{2.4}$$

N_{bo} is the base peak concentration at the peak location $W_{ep} + X_{em} + \theta X_b$ and θ is the ratio of the retrograde base width to the metallurgical base width (X_b). The

base tail region ($W_{ep} + X_{em} + \theta X_b < x < W_{ep} + X_{em} + X_b$) is:

$$N_{b2}(x) = N_{bo}\ exp\left[-\left(\frac{x - W_{ep} - X_{em} - \theta X_b}{x_{b2}}\right)^2\right] - N_{epi} \qquad (2.5)$$

$$x_{b2} = \frac{(1-\theta)X_b}{\sqrt{ln\left(\frac{N_{bo}}{N_{epi}}\right)}} \qquad (2.6)$$

where N_{epi} is the epitaxial collector concentration. In Fig. 2.1, W_b is the neutral base width, which can be calculated as the metallurgical base width (X_b) minus the sum of the width of the emitter-base depletion layer on the base side (w_2) and the width of the base-collector depletion layer on the base side (w_3).

As mentioned, in order to fairly compare different Ge profiles, the Ge dose of different profiles must be the same, i.e., identical SiGe film stability is maintained. Therefore, the Ge profile in terms of Ge fraction can be described as follows (see Fig. 2.2):

$$y(x) = y_{av} - g\ W_b/2 + g\ x \qquad (2.7)$$

where y_{av} is the average Ge fraction and g is the Ge grading over the neutral base. This definition allows Ge profiles of both positive and negative gradients to be included[2]. For $g = 0$, it describes a box profile of a uniform Ge fraction of y_{av}. In other words, (2.7) represents a family of Ge profiles (box profile, graded profiles with negative Ge gradings, and graded profiles with positive Ge gradings) with identical Ge dose of $y_{av}W_b$.

[2]It should be noted that the same Ge profile definition can be found in [97].

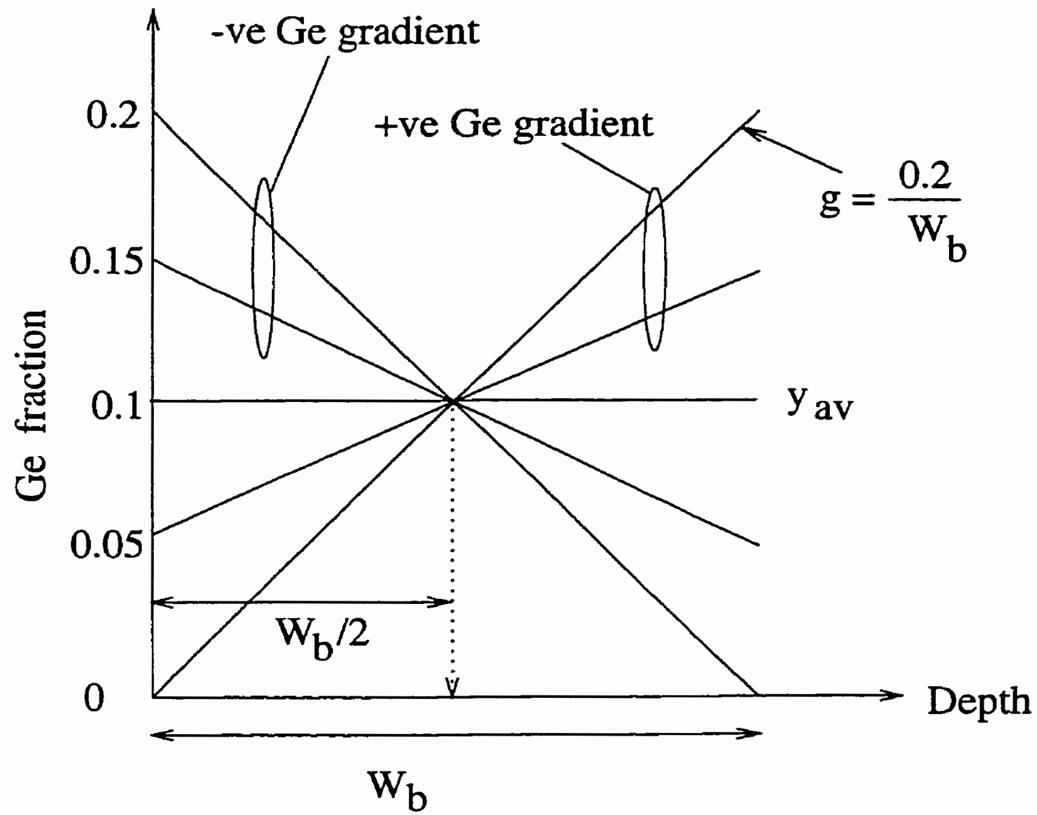

Figure 2.2: A schematic of the Ge profile under study (Note. The Ge fraction is drawn in a linear scale).

2.2.2 Emitter delay time

A brief review

Polysilicon emitter (polyemitter) has been widely used in today's high-speed bipolar and BiCMOS production processes [98]. One of its advantages is the ability to maintain the emitter-base junction capacitance at a reasonable level by allowing a shallow emitter junction to be formed when the lateral device dimension is further reduced. Despite the shallow emitter junction depth, it provides a considerable current gain over conventional bipolar transistors. In light of these advantages, together with the self-aligning features introduced into the production process, polyemitter bipolar transistors are responsible for ushering bipolar technology into VLSI [99].

Models attributing the current gain enhancement to different mechanisms have been published: oxide tunneling [100], thermionic emission due to the oxide barrier [101] or due to the doping pile-up [102], low-mobility transport in the poly-Si region or at the poly/mono-Si interface [103, 104], and some combinations of the above mechanisms [105–109]. However, only a few studies [109–112] were devoted to modelling the emitter delay time in polyemitter transistors. In general, two approaches have been taken to calculate τ_e. The first one is the macroscopic approach which models the effect of the poly-Si region by material parameters at the poly/mono-Si interface and the mobility value in the poly-Si bulk. This approach was first demonstrated by Suzuki [110], adopted by other researchers [113–116], and later improved by Basu *et al.* [112]. The second one is the microscopic approach which was first attempted by Castaner *et al.* [111] and later improved by Rinaldi [109]. As the name implies, this approach not only considers the poly/mono-Si interface properties but also models the effect of the poly-Si region by looking

into its microstructures such as grains and grain boundaries. A brief review of both approaches follows.

Taking into consideration oxide tunneling, reduced mobility in poly-Si and recombination mechanisms, Suzuki [110] reported a unified model with analytical expressions for the injected minority hole current and the emitter delay time τ_e. One minor limitation of Suzuki's model is that it can only be applied to the uniformly doped poly-Si region. Employing Suzuki's model, Chang *et al.* [113] and later Lu *et al.* [114] derived closed-form analytical expressions of τ_e for Si BJT and SiGe HBT, respectively. Chyan *et al.* [115] and Ma *et al.* [116] extended Suzuki's model to the high-injection regime for Si BJT and SiGe HBT, respectively. More recently, Basu *et al.* [112] considered the effect of oxide break-up at the interface between the poly-Si region and mono-Si region and derived new expressions for the emitter delay time in the poly-Si region (τ_{ep}) and that in the mono-Si region (τ_{em}).

Castaner *et al.* [111] reported an alternative method to calculate τ_e under the microscopic approach. The key to their method is to determine the ratio (r) of minority charges stored in the poly-Si region (Q_p) to that in the mono-Si region (Q_m). The emitter delay time can then be obtained as the emitter delay time (with respect to the mono-Si region) multiplied by the factor $(1 + r)$. To calculate Q_p, Castaner *et al.* adopted the multi-grain box model of Yu *et al.* [106], in which the poly-Si region is assumed to consist of n grain boxes and $2n+2$ interfaces (including $2n$ interfaces between grains and grain boundaries as each grain is bound by two grain boundaries, and one interface for the metal contact at one end, and one for the oxide at the other end). The charge in each grain and grain boundary is related to an effective recombination velocity (ERV), which in turn depends on the diffusion length, the mobility, and the density of the interface traps. In short, to calculate Q_p, one first needs to know the charge stored in each grain, i.e., determine the ERV

at each interface, and then sum up the charges in all grain boxes. The difficulties of this method are: i) it is not clear how many grain boxes are required for an accurate calculation of Q_p, ii) a numerical iteration is required as the ERV of each grain depends on the ERV of the next grain, and iii) parameters like the density of the interface traps at the grain boundaries are not well characterized.

Rinaldi [109] improved upon Castaner *et al.* 's method by defining an effective hole diffusion constant and an effective hole diffusion length which encapsulate the properties of the multi-grain poly-Si region. By assuming an infinite number of grains, Q_p can be approximated asymptotically. As such, difficulties i) and ii) of Castaner *et al.* 's method can be avoided. However, since both effective quantities depend on parameters like the density of interface traps at the grain boundary, it still shares the last difficulty of Castaner *et al.* 's method. Another concern is that experimental evidence suggests that only one to three grains exist in the vertical direction [117]. In other words, the assumption of infinite number of grains that is required for the asymptotic approximation is questionable. Furthermore, the asymptotic method is not computationally efficient considering that one quadratic equation for each emitter structure studied needs to be solved to evaluate the asymptotic values before obtaining the two effective quantities.

Derivation

In this section, the macroscopic approach and in particular Suzuki's model, with modifications, is used to calculate τ_e for the following reasons:

- the details of the grain structures in the microscopic approach (e.g. the interface trap densities at grain boundaries) is seldom completely characterized [117],

- the concentration of the mono-Si emitter used in this chapter is uniform and high enough that Suzuki's low-injection model with the drift current neglected in the poly-Si region is still valid,
- the trend in the industry is that a HF dip prior to the polysilicon deposition is often used to intentionally remove the interfacial oxide for reasons of reproducibility [118], and thus modelling the oxide breakup is not necessary,
- in order to compare the effect of a large number of Ge profiles on τ_e and τ_b for bases and emitters with a wide range of dimensions, a simple analytical expression is preferred to the iteration method or the asymptotic method, and
- an analytical expression facilitates the calculation of the charge partitioning factors (which will be discussed later) since it allows a quick determination of the hole profile in both mono-Si and poly-Si emitter regions.

Based on Suzuki's model, a new analytical expression for τ_e will be derived to account for the finite effective recombination velocity at the metal/poly interface (S_m). It should be noted that in Suzuki's and similar works [110, 112–116], a zero hole concentration, i.e., an infinite effective recombination velocity, is always assumed at the metal/poly interface, which slightly simplifies the derivation. However, considering both the concentration and electric-field dependence of the carrier mobility and the form of potential between the space-charge region near the metal contact and the semiconductor, Heasell [119] showed that the effective recombination velocity is lower than 10^7 cm/s in most situations. Fig. 2.3 shows a typical minority carrier profile in the poly-emitter structure that was defined earlier in the chapter.

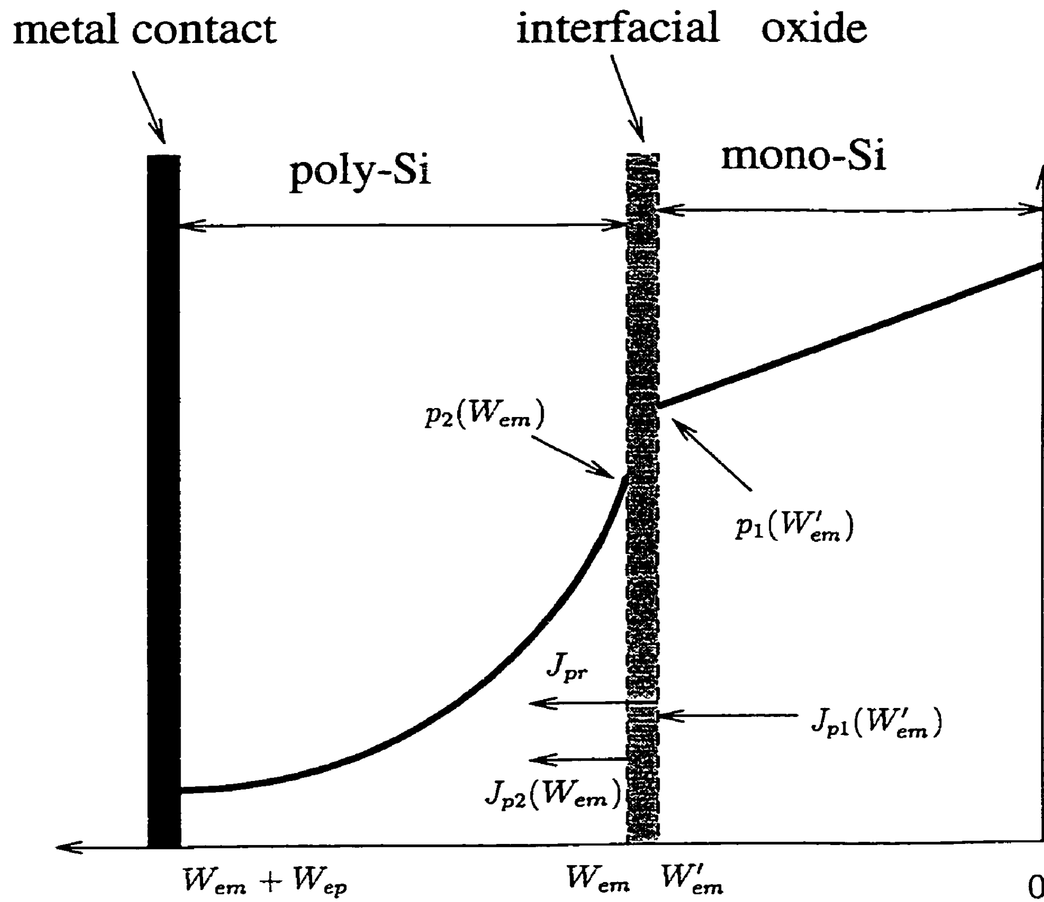

Figure 2.3: A typical minority hole profile in poly-Si emitter structure used in this study.

First of all, the continuity equation in the poly-Si region needs to be solved. Since the doping concentration in poly-Si is often uniform, the drift term in the hole current equation can be ignored and the continuity equation can be turned into the hole diffusion equation as follows:

$$\frac{d^2 p_2(x)}{dx^2} - \frac{p_2(x)}{L_{p2}^2} = 0 \tag{2.8}$$

where $L_{p2} = \sqrt{D_{p2}\tau_{p2}}$ is the hole diffusion length in poly-Si. The general solution to this 2nd order linear differential equation is:

$$p_2(x) = A\ e^{\frac{x}{L_{p2}}} + B\ e^{-\frac{x}{L_{p2}}} \tag{2.9}$$

where A and B are constants to be determined from the boundary conditions.

At $x = W_{em}$, a Dirichlet condition exists (see Appendix A),

$$p_2(W_{em}) = \alpha p_1(W'_{em}) \tag{2.10}$$

where α is the hole tunneling probability, which can be expressed in terms of the hole effective mass in the insulator (m_h^*), the effective potential barrier of the oxide to holes (χ_h) tunneling into the poly-Si from the mono-Si, and the thickness of the interfacial oxide[3] (δ) [105,120]:

$$\alpha = \frac{e^{-b_h}}{1 - C_h kT} \tag{2.11}$$

[3] For generality and the advantage of being able to cross-check with existing τ_e expressions, the following derivation will first assume the presence of an interfacial oxide layer and then set $\alpha = 1$ to ignore its effect in calculations for HF devices.

with

$$b_h = \frac{4\pi\delta}{h}\sqrt{2m_h^* q\chi_h} \tag{2.12}$$

$$C_h = \frac{2\pi\delta}{h}\sqrt{\frac{2m_h^*}{q\chi_h}} \tag{2.13}$$

where h is the Planck constant, k is the Boltzmann constant, and T the temperature.

To account for the finite effective recombination velocity at the metal/poly-Si interface, a Neumann boundary condition is used at $x = W_{em} + W_{ep}$ (this is where our derivation departs from the derivations of Suzuki and others):

$$\begin{aligned} J_{p2}(W_{em} + W_{ep}) &= qS_m p_2(W_{ep} + W_{em}) \\ -qD_{p2}\left.\frac{dp_2}{dx}\right|_{x=W_{ep}+W_{em}} &= qS_m p_2(W_{ep} + W_{em}) \\ \left.\frac{dp}{dx}\right|_{x=W_{ep}+W_{em}} &= -\frac{S_m p_2(W_{ep} + W_{em})}{D_{p2}} \end{aligned} \tag{2.14}$$

Substituting (2.10) and (2.14) into (2.9) yields:

$$A = \alpha p_1(W'_{em}) \cdot \frac{(\frac{1}{L_{p2}} - \frac{S_m}{D_{p2}})e^{-\frac{W_{em}+W_{ep}}{L_{p2}}}}{\left(\frac{1}{L_{p2}} + \frac{S_m}{D_{p2}}\right)e^{\frac{W_{ep}}{L_{p2}}} + \left(\frac{1}{L_{p2}} - \frac{S_m}{D_{p2}}\right)e^{-\frac{W_{ep}}{L_{p2}}}} \tag{2.15}$$

and

$$B = \alpha p_1(W'_{em}) \cdot \frac{(\frac{1}{L_{p2}} + \frac{S_m}{D_{p2}})e^{\frac{W_{em}+W_{ep}}{L_{p2}}}}{\left(\frac{1}{L_{p2}} + \frac{S_m}{D_{p2}}\right)e^{\frac{W_{ep}}{L_{p2}}} + \left(\frac{1}{L_{p2}} - \frac{S_m}{D_{p2}}\right)e^{-\frac{W_{ep}}{L_{p2}}}} \tag{2.16}$$

With (2.15) and (2.16), the hole density given by (2.9) becomes:

$$p_2(x) = \alpha p_1(W'_{em}) \cdot \frac{(\frac{1}{L_{p2}} - \frac{S_m}{D_{p2}})e^{\frac{x-W_{em}-W_{ep}}{L_{p2}}} + (\frac{1}{L_{p2}} + \frac{S_m}{D_{p2}})e^{\frac{W_{em}+W_{ep}-x}{L_{p2}}}}{\left(\frac{1}{L_{p2}} + \frac{S_m}{D_{p2}}\right)e^{\frac{W_{ep}}{L_{p2}}} + \left(\frac{1}{L_{p2}} - \frac{S_m}{D_{p2}}\right)e^{-\frac{W_{ep}}{L_{p2}}}} \tag{2.17}$$

With some algebraic manipulation, $p_2(x)$ can be re-written as:

$$p_2(x) = \alpha p_1(W'_{em}) \cdot \frac{\frac{1}{L_{p2}} cosh\left(\frac{x-W_{em}-W_{ep}}{L_{p2}}\right) - \frac{S_m}{D_{p2}} sinh\left(\frac{x-W_{em}-W_{ep}}{L_{p2}}\right)}{\frac{1}{L_{p2}} cosh\left(\frac{W_{ep}}{L_{p2}}\right) + \frac{S_m}{D_{p2}} sinh\left(\frac{W_{ep}}{L_{p2}}\right)} \tag{2.18}$$

In (2.18), $p_2(x)$ is expressed in terms of $p_1(W'_{em})$. But $p_1(W'_{em})$ is an unknown. To obtain τ_e, both $p_2(x)$ and $p_1(x)$ must be known and integrated over the poly-Si region and the mono-Si region, respectively. The sum of the two integrations is then divided by the sum of the electron and hole currents to obtain τ_e. Therefore, $p_1(W'_{em})$ must be evaluated first to determine $p_2(x)$ and later $p_1(x)$. To do so, we examine the current continuity at the poly/mono-Si interface (see Fig. 2.3 and Appendix A for details):

$$J_{p1}(W'_{em}) = J_{pr} + J_{p2}(W_{em}) \tag{2.19}$$

where $J_{p1}(W'_{em})$ is the hole current injected from the mono-Si region at $x = W'_{em}$, and J_{pr} is the recombination current at the oxide/mono-Si interface, which can be written as:

$$J_{pr} = qS_p p_1(W'_{em}) \tag{2.20}$$

where S_p is the recombination velocity at the oxide/mono-Si interface[4]. The quantity $J_{p2}(W_{em})$ is the hole current at $x = W_{em}$ on the side of the poly-Si region, which can be easily obtained since the electric field is assumed negligible in the uniformly-doped poly-Si region:

$$J_{p2}(W_{em}) = -qD_{p2}\left.\frac{dp_2(x)}{dx}\right|_{x=W_{em}} \tag{2.21}$$

Using (2.18), $J_{p2}(W_{em})$ is found as:

$$J_{p2}(W_{em}) = q\alpha p_1(W'_{em})\frac{S_m + \frac{D_{p2}}{L_{p2}}tanh\left(\frac{W_{ep}}{L_{p2}}\right)}{1 + S_m\frac{L_{p2}}{D_{p2}}tanh\left(\frac{W_{ep}}{L_{p2}}\right)} \tag{2.22}$$

Since $J_{p2}(W_{em})$ can also be written as:

$$J_{p2}(W_{em}) = qp_2(W_{em})S_{poly} \tag{2.23}$$

where S_{poly} is the effective recombination velocity (ERV) relative to the poly-Si bulk region, from (2.10) and (2.22) S_{poly} is then found as:

$$S_{poly} = \frac{S_m + \frac{D_{p2}}{L_{p2}}tanh\left(\frac{W_{ep}}{L_{p2}}\right)}{1 + S_m\frac{L_{p2}}{D_{p2}}tanh\left(\frac{W_{ep}}{L_{p2}}\right)} \tag{2.24}$$

It should be noted that (2.24) is actually the same as equation (10) in Yu *et al.* 's multi-grain box model [106] with the special case of only one grain considered in the poly-Si bulk. The equivalence is not surprising since our macroscopic approach, by

[4]It should be noted that Rinaldi [109] used the symbol S_{is} to name the recombination velocity at the oxide/mono-Si interface, which is more appropriate. However, to facilitate the comparison between the derived expressions and Suzuki's expressions [110], Suzuki's notation of S_p is used here for convenience.

definition, attempts to model the poly-Si region as a "black box" (i.e. a single grain) instead of looking into its microstructures. To calculate S_{poly} in this work, the value of D_{p2} is obtained from Klaassen's recent unified mobility model [121] multiplied by the poly-Si to crystalline-Si majority hole mobility ratio taken from [122] (assuming that this ratio is applicable to the minority carrier mobility). At a doping of $10^{20}\mathrm{cm}^{-3}$, the ratio is ~ 0.57. A more recent result from [123] gives the ratio a value of 0.3. However, it is found that the two values give rise to a difference of less than 5 % in the final results presented in Section 2.4 (except for the current gain, which varies by a factor of two). The value of L_{p2} is taken from [124] (which is around 100 nm). Substituting (2.20) and (2.23) into (2.19), the relation between $p_1(W_{em})$ and the hole current $J_{p1}(W'_{em})$ can be obtained:

$$J_{p1}(W'_{em}) = qp_1(W'_{em})(\alpha S_{poly} + S_p) \tag{2.25}$$

In a more compact form, $J_{p1}(W'_{em})$ can be re-written as:

$$J_{p1}(W'_{em}) = qp_1(W'_{em})S'_p \tag{2.26}$$

where S'_p is the effective recombination velocity at the poly/mono-Si interface (to be precise, the oxide/mono-Si interface, see Appendix A), which is defined as:

$$S'_p \triangleq S_p + \alpha S_{poly} \tag{2.27}$$

with S_p modelling the oxide/mono-Si interface recombination, α modelling the oxide tunneling, and S_{poly} modelling the minority carrier transport in the poly-Si bulk. If only HF devices are considered, i.e., no oxide tunneling effect ($\alpha = 1$), (2.27) reduces to Ning and Issac's two-region model [103]. Equation (2.26) shows that

$p_1(W'_{em})$ can be evaluated if J_{p1} is known.

To evaluate J_{p1} in the mono-Si region, we neglect the recombination there. This is a reasonable assumption because the diffusion length in the mono-Si region is often larger than the typical emitter junction depth[5]. In other words, J_{p1} is a constant. From conventional drift-diffusion transport theory, J_{p1} can be expressed as:

$$J_{p1} = q\mu_p(x)p(x)\mathcal{E}(x) - qD_{p1}(x)\frac{dp(x)}{dx} \tag{2.28}$$

where μ_p and D_{p1} are the hole mobility and diffusion coefficient in mono-Si, respectively. Their values are both concentration and field dependent and can be obtained from Klaassen's mobility model [121] and Caughey-Thomas electric field adjustment [126]. Since the electric field in the low injection regime, $\mathcal{E}$, does not change significantly from its equilibrium value in the mono-Si region [127, 128], it can be determined by setting $J_{p1} = 0$ in (2.28):

$$\mathcal{E} = \frac{V_T}{p_o}\frac{dp_o}{dx} \tag{2.29}$$

where V_T is the thermal voltage and p_o is the equilibrium hole concentration. By substituting (2.29) into (2.28), we obtain:

$$J_{p1} = -qD_{p1}p_o\frac{d}{dx}\left(\frac{p}{p_o}\right) \tag{2.30}$$

---

[5]Using Klaassen's latest lifetime and mobility models [121, 125], for a donor concentration of 10^{20}cm^{-3}, the hole diffusion length is estimated as 0.385 μm, which is much higher than the thickness of today's mono-Si region. This assumption of negligible recombination is even more valid in this study because the doping of the mono-Si region is significantly lower than 10^{20}cm^{-3}.

The same result had been obtained earlier by Selvakumar [129] and del Alamo and Swanson [130] using the definition of a normalized carrier density $u = p/p_o$. The equilibrium hole concentration p_o is related to the effective intrinsic carrier concentration n_{ie} as follows:

$$p_o = \frac{n_{ie}^2}{N_d} \tag{2.31}$$

Considering the bandgap narrowing due to the heavy doping effect, the effective intrinsic carrier concentration can be written as [131]:

$$n_{ie}^2 = n_{io}^2 exp\left[\frac{\Delta E_g(x)}{kT}\right] \tag{2.32}$$

with n_{io} as the intrinsic carrier concentration and $\Delta E_g(x)$ as the apparent bandgap narrowing, which can be calculated using Klaassen's recent unified bandgap narrowing model [132]. Integrating (2.30) over the mono-Si region,

$$J_{p1}\int_0^{W_{em}} \frac{dx}{D_{p1}p_o} = -q\left[\frac{p(W_{em})}{p_o(W_{em})} - \frac{p(0)}{p_o(0)}\right] \tag{2.33}$$

In Fig. 2.3, $x = 0$ is the depletion layer edge of the emitter-base junction on the emitter side and W_{em} is now re-defined as the location of the oxide/mono-Si interface, instead of the poly-Si/oxide interface. Therefore,

$$\frac{p(0)}{p_o(0)} = e^{V_{BE}/V_T} \tag{2.34}$$

where V_{BE} is the applied emitter-base terminal voltage. Applying this boundary condition with expressions (2.26) to (2.33), J_{p1} is finally obtained as:

$$J_{p1} = \frac{q\ e^{V_{BE}/V_T}}{\int_0^{W_{em}} \frac{dx}{D_{p1} p_o} + \frac{1}{S'_p p_o(W_{em})}} \tag{2.35}$$

This is equivalent to equation (8) in Shibib *et al.* 's transparent emitter model [133].

If S_m approaches infinity, (2.27) reduces to:

$$S'_p = S_p + \frac{\alpha D_{p2}}{L_{p2}\ tanh\left(\frac{W_{ep}}{L_{p2}}\right)} \tag{2.36}$$

Then J_{p1} becomes:

$$J_{p1} = \frac{q\ e^{V_{BE}/V_T}}{\int_0^{W_{em}} \frac{dx}{D_{p1} p_o} + \frac{L_{p2}\ tanh\left(\frac{W_{ep}}{L_{p2}}\right)}{p_o(W_{em}) D_{p2} \alpha \left[1 + \frac{S_p L_{p2}}{\alpha D_{p2}} tanh\left(\frac{W_{ep}}{L_{p2}}\right)\right]}} \tag{2.37}$$

This is the same as equation (13) in Suzuki's model [110].

For convenience, J_{p1} is re-written in terms of the effective emitter Gummel number (G_e) as follows:

$$J_{p1} = \frac{q\ e^{V_{BE}/V_T}}{G_e} \tag{2.38}$$

where

$$G_e = g_{em} + g_{ep} \tag{2.39}$$

with

$$g_{em} = \int_0^{W_{em}} \frac{dx}{D_{p1} p_o} \tag{2.40}$$

which is associated with the mono-Si region, and

$$g_{ep} = \frac{1}{p_o(W_{em}) S'_p} \tag{2.41}$$

which is associated with the poly-Si region.

Now with the knowledge of J_{p1}, both $p_1(x)$ and $p_2(x)$ can be readily obtained. To obtain $p_1(x)$, (2.30) is integrated from x to W_{em},

$$J_{p1} \int_x^{W_{em}} \frac{dz}{D_{p1} p_o} = -q \left[\frac{p(W_{em})}{p_o(W_{em})} - \frac{p(x)}{p_o(x)} \right] \tag{2.42}$$

Hence,

$$p_1(x) = \frac{p_o(x) J_{p1}}{q} \left(\int_x^{W_{em}} \frac{dz}{D_{p1} p_o} + g_{ep} \right) \tag{2.43}$$

From (2.18) and (2.26), $p_2(x)$ is obtained as:

$$p_2(x) = \alpha \frac{J_{p1}}{q S'_p} \cdot \frac{\frac{1}{L_{p2}} cosh\left(\frac{x - W_{em} - W_{ep}}{L_{p2}}\right) - \frac{S_m}{D_{p2}} sinh\left(\frac{x - W_{em} - W_{ep}}{L_{p2}}\right)}{\frac{1}{L_{p2}} cosh\left(\frac{W_{ep}}{L_{p2}}\right) + \frac{S_m}{D_{p2}} sinh\left(\frac{W_{ep}}{L_{p2}}\right)} \tag{2.44}$$

Finally, τ_e can be calculated. By definition,

$$\tau_e \triangleq \frac{dQ_e}{d(J_{p1} + J_n)} \tag{2.45}$$

where Q_e is the total hole charges in the neutral emitter, which can be written as:

$$Q_e = q \int_0^{W_{em}} p_1(x)dx + q \int_{W_{em}}^{W_{em}+W_{ep}} p_2(x)dx \tag{2.46}$$

and J_n is the electron minority current injected into the neutral base:

$$J_n = \frac{q\ e^{V_{BE}/V_T}}{G_b} \tag{2.47}$$

where Q_b is defined as effective base Gummel number. Its definition will be given later when the base transit time expression is presented. Since $p_1(x)$ and $p_2(x)$ are proportional to J_{p1} and both J_{p1} and J_n are proportional to e^{V_{BE}/V_T}, (2.45) simply becomes:

$$\tau_e = \frac{Q_e}{J_{p1} + J_n} \tag{2.48}$$

Substituting (2.46) into (2.48), we can divide the emitter delay time into two components: τ_{em} for the mono-Si region and τ_{ep} for the poly-Si region:

$$\tau_{em} = \frac{q \int_0^{W_{em}} p_1(x)dx}{J_{p1} + J_n} \tag{2.49}$$

and

$$\tau_{ep} = \frac{q \int_{W_{em}}^{W_{em}+W_{ep}} p_2(x)dx}{J_{p1} + J_n} \tag{2.50}$$

Substituting (2.43), (2.44), (2.38), and (2.47) into (2.49) and (2.50), one can finally obtain the two components of the emitter delay time:

$$\tau_{em} = \frac{1}{1+\beta_{dc}} \int_0^{W_{em}} p_o \left[\int_x^{W_{em}} \frac{dz}{D_{p1} p_o} + g_{ep} \right] dx$$
$$\tau_{ep} = \frac{1}{1+\beta_{dc}} \cdot \frac{\alpha L_{p2}^2}{S_p' D_{p2}} \cdot \frac{1 - \frac{1}{cosh\left(\frac{W_{ep}}{L_{p2}}\right)} + \left(\frac{D_{p2}}{L_{p2} S_m}\right) tanh\left(\frac{W_{ep}}{L_{p2}}\right)}{\frac{L_{p2}}{D_{p2}} tanh\left(\frac{W_{ep}}{L_{p2}}\right) + \frac{1}{S_m}} \tag{2.51}$$

where β_{dc} is the D.C. common-emitter current gain, here defined as G_e/G_b. It can be seen from (2.51) that τ_e is inversely proportional to the D.C. current gain β_{dc}.

Again, if S_m approaches infinity (i.e. S_p' becomes (2.36)), (2.51) reduces to:

$$\tau_{em} = \frac{1}{1+\beta_{dc}} \int_0^{W_{em}} p_o \left[\int_x^{W_{em}} \frac{dz}{D_{p1} p_o} + \frac{L_{p2} tanh\left(\frac{W_{ep}}{L_{p2}}\right)}{\alpha D_{p2} p_o(W_{em}) \left[1 + \frac{S_p L_{p2}}{\alpha D_{p2}} tanh\left(\frac{W_{ep}}{L_{p2}}\right)\right]} \right] dx$$
$$\tau_{ep} = \frac{1}{1+\beta_{dc}} \cdot \frac{L_{p2}^2}{D_{p2}} \cdot \frac{1 - \frac{1}{cosh\left(\frac{W_{ep}}{L_{p2}}\right)}}{\frac{S_p L_{p2}}{\alpha D_{p2}} tanh\left(\frac{W_{ep}}{L_{p2}}\right) + 1} \tag{2.52}$$

which are the same as equations (22) and (23) in Suzuki's model [110].

2.2.3 Base delay time

Compared with τ_e, the calculation of the base delay or transit time τ_b is more straightforward. The primary task here is to incorporate the effects of non-uniform energy bandgap, due to the presence of Ge, into the conventional expression of τ_b for Si BJTs. A more detailed derivation and background survey can be found in Chapter 3 where a closed-form analytical expression of τ_b is derived for a SiGe

HBT with a retrograde base profile and linear Ge grading. In this section, a more limited derivation of the analytical τ_b expression for a SiGe HBT with arbitrary base doping and Ge profiles is presented.

As the base counterpart of the hole current density (2.30) in the neutral emitter, the electron current density injected into the neutral base can be similarly written as:

$$J_n = qD_n n_o \frac{d}{dx}\left(\frac{n}{n_o}\right) \tag{2.53}$$

where D_n is the electron diffusion coefficient, which is both concentration and field dependent and can be obtained from Klaassen's mobility model [121] with Caughey-Thomas adjustment [126]. The quantity n_o is the equilibrium electron concentration, and n is the electron concentration. To incorporate the influence of Ge in the base, n_o is defined through the effective intrinsic carrier concentration (n_{ie})[6]:

$$n_o(x) = \frac{n_{ie}^2(x)}{N_b(x)} \tag{2.54}$$

where N_b is the base concentration, and

$$n_{ie}^2(x) = n_{io}^2(x)\ exp\left(\frac{\Delta E_g(x)}{kT}\right) \tag{2.55}$$

[6]Although the same notation is used for both the effective intrinsic concentration in the base and in the emitter, it should be understood that the two are not the same because of the Ge presence in the base.

with the apparent bandgap narrowing

$$\begin{aligned}\Delta E_g(x) =& \Delta E_{g,HDE} + \Delta E_{g,Ge} + kT\ ln\left(\frac{N_{c,SiGe}N_{v,SiGe}}{N_{c,Si}N_{v,Si}}\right)\\ &+ kT\left[ln\left(\frac{N_b}{N_{v,SiGe}}\right) - \mathcal{F}_{\frac{1}{2}}^{-1}\left(\frac{N_b}{N_{v,SiGe}}\right)\right]\end{aligned} \tag{2.56}$$

The terms $\Delta E_{g,HDE}$ and $\Delta E_{g,Ge}$ denote the bandgap narrowing due to the doping and the Ge presence, respectively. The last two terms of (2.56) account for the reduced effective density of states due to the Ge and the Fermi-level shift associated with Fermi-Dirac statistics[7], respectively. The details of the physical models for each of these terms are given in Appendix C and the corresponding analytical expressions can be found in Chapter 4. With these physical models, the last three terms of (2.56) can be expressed in terms of the Ge fraction as defined in (2.7). In other words, the effects of the Ge profile are now incorporated through the quantity n_o.

Since the base width for today's optimized HBT is around 100 nm or smaller, which is much smaller than the electron diffusion length at typical base doping levels used for today's SiGe HBTs[8], the neutral base recombination is neglected. In other words, J_n is treated as a constant. Integrating (2.26) from x to the depletion layer edge of the base-collector junction on the base side, i.e., W_b (since the depletion layer edge of the emitter-base junction on the base side is assumed at $x = 0$) yields:

$$J_n \int_x^{W_b} \frac{dz}{D_n n_o} = q\left[\frac{n(W_b)}{n_o(W_b)} - \frac{n(x)}{n_o(x)}\right] \tag{2.57}$$

[7]Since the base doping used in this study can be very high (e.g. 10^{19}cm^{-3}), the Fermi-Dirac statistics correction term is included.

[8]Using Klaassen's latest lifetime and mobility models [121, 125], for an acceptor concentration of 10^{20}cm^{-3}, the electron diffusion length is estimated as 0.487 μm, which is much higher than the thickness of today's base width.

At $x = W_b$, the electron current can be expressed in terms of the junction velocity at the base-collector junction (S). The junction velocity (S) can be reasonably assumed to be the thermal velocity v_s as there exists a high electric field at the base-collector junction:

$$J_n = -qSn(W_b) \tag{2.58}$$

Substituting (2.58) into (2.57) gives:

$$\frac{n(x)}{J_n} = -\frac{n_o(x)}{q}\left[\frac{1}{Sn_o(W_b)} + \int_x^{W_b} \frac{dz}{D_n\, n_o}\right] \tag{2.59}$$

By definition, the base delay time is:

$$\tau_b = \frac{dQ_b}{dJ_n} \tag{2.60}$$

where

$$Q_b = q\int_0^{W_b} n(x)dx \tag{2.61}$$

Since $n(x)$, in turn Q_b, and J_n are all proportional to e^{V_{BE}/V_T} (see (2.59) and the end of the following derivation for J_n), the base delay time can be simply calculated as:

$$\begin{aligned}\tau_b &= \frac{Q_b}{J_n} \\ &= \frac{q\int_0^{W_b} n(x)dx}{J_n} \\ &= q\left(\int_0^{W_b} \frac{n(x)}{J_n}dx\right)\end{aligned} \tag{2.62}$$

Substituting (2.59) into (2.62) with the absolute value of $n(x)/J_n$ taken, one finally obtains:

$$\tau_b = \int_0^{W_b} n_o(x) \left[\frac{1}{S n_o(W_b)} + \int_x^{W_b} \frac{dz}{D_n\ n_o} \right] dx \tag{2.63}$$

This is the same as equation (11) in Suzuki and Nakayama's different, but equivalent derivation [134], which defined τ_b as the the product of the emitter-base diode resistance and electron diffusion capacitance. If the velocity saturation at the base-collector junction is neglected (i.e. S is set to infinity), (2.63) reduces to:

$$\tau_b = \int_0^{W_b} n_o(x) \int_x^{W_b} \frac{dz}{D_n\ n_o} dx \tag{2.64}$$

which is equivalent to Kroemer's τ_b expression [135].

To show that J_n is proportional to e^{V_{BE}/V_T}, one can set x to 0 in (2.57) and then apply the boundary condition at $x = 0$ (i.e. $n(0) = n_o(0)e^{V_{BE}/V_T}$). J_n becomes:

$$J_n = \frac{q\ e^{V_{BE}/V_T}}{G_b} \tag{2.65}$$

which will be the same as the J_n introduced in previous section (2.47) if the effective base Gummel number G_b is defined as:

$$G_b = \int_0^{W_b} \frac{dx}{D_n\ n_o} + \frac{1}{S n_o(W_b)} \tag{2.66}$$

2.2.4 Non-quasi-static correction

Analytical expressions for emitter and base delay times have been derived in the two previous subsections. It is tempting to immediately relate the two delay times to the transition frequency as most, if not all, studies on Ge profile design have

done:

$$f_T = \frac{1}{2\pi\tau_t} = \frac{1}{2\pi(\tau_e + \tau_b)} \tag{2.67}$$

where τ_t is defined in this chapter as the total contribution of the emitter and base delay times to f_T[9]. Equation (2.67) can be easily shown to be based on the current gain model resulting from the conventional charge-control theory:

$$\beta(\omega) = \frac{\beta_o}{1 + j\omega\beta_o(\tau_e + \tau_b)} \tag{2.68}$$

where ω is the radian frequency, the quantity j denotes the imaginary number $\sqrt{-1}$, and β_o is the current gain at zero frequency. This represents the classic 6 dB/octave drop with frequency and is the basis on which f_T can be extrapolated from low frequencies. When $|\beta(\omega)| = 1$, $\omega = 1/\tau_t \sim 1/(\tau_e + \tau_b)$. Since the transition frequency is defined as the frequency at which the short-circuit current gain (i.e., β) drops to 1, (2.67) is obtained.

However, in this subsection, it is shown that (2.68) is not sufficient to describe the high frequency behaviour of transistors where the emitter delay time is comparable to the base delay time, and consequently the contribution of τ_e and τ_b to f_T measured by (2.67) will be inaccurate. Furthermore, it is explained how studies on the non-quasi-static behaviour of the transistors at high frequencies can provide a better estimation of the contribution of τ_e and τ_b to f_T, and in particular, why Hamel's high frequency model [136,137] is adopted in this chapter.

[9] To be precise, it should be noted that there are other delay time terms, as described in Chapter 1, in the overall f_T expression.

Limitation of quasi-static assumption

The term "quasi-static" refers to the situation where the dynamic or time-dependent behaviour of a transistor is deduced from some static or steady-state quantities. A typical quasi-static situation is encountered when Gummel's static charge-control relation [138] is used under the charge-control approach [139] proposed by Beaufoy and Sparkes to predict the A.C. or transient behaviour of a device. This results in an expression similar to (2.68). In (2.68), the base transit time τ_b, originally a static parameter (which can be determined from the D.C. or steady-state solution of the minority carrier as shown in the previous subsection), is used to model the A.C. current gain. As such, τ_b in (2.68) can be considered a quasi-static parameter.

Another well known quasi-static example is the use of Beaufoy and Sparkes' charge control approach in the conventional Gummel-Poon capacitance-based bipolar transistor model [140,141] for modelling the A.C. or dynamic behaviour of bipolar transistors. However, such use of the Gummel-Poon model has been shown to cause inaccuracies, e.g. transconductance phase error [142–144]. These inaccuracies stem from the quasi-static assumption made in the model. In other words, the true time-dependent physical effects of the transistor (hereafter called non-quasi-static effects), in response to the time-varying signals applied to its terminal, are ignored.

In fact, Lindmayer and Wrigley [145] have long pointed out a non-quasi-static effect when they calculated the small-signal base diffusion capacitance of a uniformly doped base. They have observed that even at low frequencies[10] the base diffusion capacitance seen from the junction terminal is not the capacitance calculated based

[10]The phrase "low frequency" refers to the frequency which is considerably less than the reciprocal of the transit time of the minority carrier in a quasi-neutral region.

on the quasi-static assumption:

$$\begin{aligned} C_{db,qs} &= \frac{dQ_b}{dV_{BE}} \\ &= \frac{dQ_b}{dJ_c}\frac{dJ_c}{dV_{BE}} \\ &= \tau_b g_m \end{aligned} \tag{2.69}$$

where V_{BE} is the time-varying base-emitter voltage and g_m is the transconductance. Instead, they found that for a wide-base junction, the capacitance is only one-half of the quasi-static base diffusion capacitance; while for a narrow-base junction, it is only two-thirds. This implies that only a fraction of the "stored" minority charges in the base is reclaimable through the junction ($Q_b/2$ for the wide base and $2Q_b/3$ for the narrow base). That is, not all stored charges respond to the applied small-signal voltage even at low frequencies. Phenomenologically, the stored charge is partitioned between the terminals in a way that only a fraction of them can be reclaimed through the junction during the dynamic operation of the device. This charge partitioning phenomenon can be considered as incorporating the non-quasi-static effect, which is modelled in this particular example by a charge-partitioning factor, α_b (1/2 for the wide base and 2/3 for the narrow base).

A physical interpretation of this non-quasi-static effect is that it takes a finite amount of time for the minority carriers to respond to the time-varying voltage across the junction and then settle to a steady-state value. Therefore, the higher the frequency of the applied voltage, the shorter time the carriers will have to respond and the more the actual carrier distribution will deviate from the steady-state value, i.e., the more invalid the quasi-static approximation will be. In other words, the quasi-static approximation tacitly assumes that the minority carrier can travel at an infinite velocity to respond to the applied voltage instantaneously [144].

Because only a portion of "stored" charges is reclaimable through the junction, the delay time, which is defined as the ratio of the change in "stored" charges to the change in the current density in response to the an applied voltage, may be over-estimated. For the purpose of comparing τ_t's corresponding to different base doping and Ge profiles in this chapter, a quasi-static approximation such as (2.67) will over-estimate the delay time terms that contribute to the overall f_T and the comparison results will become unreliable. Therefore, the question is: how do we account for the non-quasi-static effects when adopting τ_b and τ_e as quasi-static parameters to model τ_t accurately?

A brief review of studies on non-quasi static modelling

Four general approaches have been taken by researchers on modeling non-quasi-static effects of bipolar transistors.

<u>Weighting methodology</u>

The weighting methodology [142, 146–148] weighs the time-dependent continuity equation and/or the current equation with appropriate analytic functions, in order to quantify the charge partitioning phenomenon and obtain a new charge control relation. The first example of this approach is the work by Fossum and Veeraraghavan [142], who simply integrated the time-dependent continuity equation over the neutral base twice (i.e. the weighting function being set to 1) to obtain a new expression for the charge partitioning between the emitter and the base. The main drawback of their work is twofold: i) the recombination is neglected (i.e., the result cannot be applied to neutral region where recombination is significant, e.g. the emitter), and ii) the correct charge partitioning in the case of having a built-in electric field in the neutral region (e.g. regions with non-uniform doping or non-uniform bandgap) is not well defined [142].

Klose and Wieder [146] integrated the current equation instead of the continuity equation from the emitter terminal to the collector to obtain a new charge control relation that predicts the correct charge partitioning even in the case of a non-zero electric field. However, as shown by Hamel and Selvakumar [148], Klose and Wieder's relation is only useful in regions where recombination is negligible [148]. Another limitation is that their relation involves a double integral, which makes it impractical for circuit-level compact modelling. McDonald [147] showed that the expression for charge partitioning can be arrived by taking moments of the continuity equation with arbitrary weighting functions and thus provides a more general framework from which Fossum and Veeraraghavan's model can be derived and from which improvements can be made to handle non-uniformly doped neutral regions. However, he did not show how these arbitrary weighting functions can be deduced. For modelling the non-quasi-static charge partitioning in neutral regions with significant recombination, Hamel and Selvakumar [148] weighted the time-dependent current equation with a position-dependent weighting function (in their case the minority current), resulting in a general charge control relation which predicts the following charge partitioning factor for the emitter:

$$\alpha_e = \frac{\int_o^{W_e} p(x)^2/p_o(x)dx}{e^{V_{BE}/V_T}\int_0^{W_e} p(x)dx} \tag{2.70}$$

where W_e is the width of the neutral emitter region. The significance of this result is that the charge partitioning factor can be solely determined from the static charge distribution $p(x)$ and the profile data $p_o(x)$. The advantage is not only that it is less computationally intensive than Klose and Wieder's result, but it also allows the non-quasi-static effect in a neutral region, where recombination can be significant, to be accurately modelled by a quasi-static parameter such as α_e.

Series expansion method

The second approach relies on truncating the infinite series solution to the time-dependent continuity equation in the time or frequency domains [144, 149–153]. Unlike the first approach that attempts to arrive at a new charge control relation from which the charge partitioning and terminal currents can be obtained, this approach aims to arrive at terminal currents directly; even though expressions for quasi-static parameters, such as the charge partition factor, can often be obtained analytically as secondary results. The goal is to incorporate non-quasi-static effects into the current expressions but not necessarily through the concept of charge partitioning. Using a perturbation expansion to the time-dependent continuity equation with truncation up to the first order term, Hurkx [149] showed that the A.C. behaviour of the transistor can be modelled by small-signal base and collector currents as follows[11]:

$$\begin{aligned} i_b &= \frac{v_{be}}{V_T}\left[I_b + j\omega I_c(\tau_t - (1-\alpha_e)\tau_e)\right] \\ i_c &= \frac{v_{be}}{V_T} I_c[1 - j\omega(1-\alpha_b)\tau_b] \end{aligned} \tag{2.71}$$

where $\tau_t = \tau_e + \tau_b$ and the charge partitioning factor can be expressed as:

$$\alpha_{b,e} = \frac{k\int_0^W [m(x)\int_0^x G(x')dx']dx}{\int_0^W G(x)dx \int_0^W m(x)dx} \tag{2.72}$$

with

$$k = \frac{Wv/D}{Wv/D + G(W)W/\int_0^W G(x)dx} \tag{2.73}$$

[11] It should be noted that in Hurkx's paper, the charge partitioning factor is defined as one minus the charge partitioning factor used here.

where W is the position of the emitter contact in the case of a neutral emitter region or the base edge of the base-collector junction in the case of a neutral base region, v is the velocity at $x = W$, and

$$G(x) = \frac{m(x)}{D_m(x)(n_{ie}(x)/n_{io}(x))^2} \tag{2.74}$$

where $m = n$ (or p), the minority carrier concentration for base (or emitter). Similar to problems of Klose and Wieder's model, Hurkx's model is computationally intensive as it involves double integral in determining α and neglects recombination. Also, it is not very accurate since higher order terms are ignored during the truncation.

Performing a Laplace transform on the continuity equation and then solving the resulting ordinary differential equation in frequency domain by series expansion with truncation, Chen, Lindholm and Wu [144] proposed an inductive model which successfully predicts the frequency dependence of the magnitude of the transconductance for an exponentially-doped base. However, the main limitation is that recombination is neglected in the derivation. Later in another paper, Wu and Lindholm [150] accounted for the emitter recombination and arbitrary base profile and presented an expression for the A.C. input admittance in the emitter region that involves a few triple integrals. The intensive computation required for evaluating the triple integrals is the major drawback of their result. Similar to Hurkx, Hamel [151] employed the rigorous perturbation method but included recombination, and obtained the same charge partitioning factor predicted earlier by Hamel and Selvakumar's charge control relation (see (2.70)). This confirms that the charge partitioning factor can be solely determined from the static charge distribution and profile data.

More recently, Rinaldi [152,153] extended Wu and Lindholm's work to the high injection regime by solving the current and continuity equation with the integral series solution method proposed by Wu and Lindholm. Rinaldi showed that small-signal base and collector currents can be expressed in terms of an infinite polynomial of a complex variable. Results of a few other non-quasi-static models can be obtained using different approximations of the general current expressions. In addition, simple analytical expressions for model parameters, such as the charge partitioning factor and phase shift were given. However, sharing the common problem of all series expansion methods, Rinaldi's model can be very accurate but at the same time very computationally intensive as many high order terms can be included in the solution.

Effective time constant method

The third approach for modelling non-quasi-static effects is characterized by the use of effective time constant(s) to represent the infinite number of poles and zeros in the small-signal expressions for the base and collector current or equivalently the admittance (y_{11}), trans-admittance (y_{21}) and current gain [143,154]. This approach usually provides a highly accurate non-quasi-static model since higher order terms are implicitly included in the effective time constants. It is also compact and computationally efficient provided that the effective time constants are known. Like the second approach, the concept of charge partitioning is not necessarily involved as long as the terminal currents can be modelled accurately. Thomas and Moll [154] showed that β can be represented by an excess phase shift term instead of the product of an infinite number of poles. As such, the magnitude and phase of β can be determined from three amplitude measurements. Seitchik *et al.* [143] then applied the same treatment to y_{21} since both β and y_{21} can be represented by the

product of an infinite number of poles, resulting in the following expressions[12]:

$$y_{11}^E = \frac{g_m \alpha_o}{\beta_o} + \frac{j\omega\alpha_o\tau_f g_m}{1 + j\omega\tau_2} \quad (2.75)$$

where y_{11}^E is the common-emitter admittance and α_o is the common-base current gain at zero frequency,

$$y_{21}^E = g_m \alpha_o \left[\frac{e^{-j\omega\tau_1}}{1 + j\omega\tau_2}\right] \quad (2.76)$$

where y_{21}^E is the common-emitter trans-admittance, and finally

$$\beta = \frac{\beta_o e^{-j\omega\tau_1}}{1 + j\omega\beta_o\tau_f} \quad (2.77)$$

Equations (2.75)–(2.77) show that three important small-signal A.C. quantities can be fully characterized using three effective time constants τ_1, τ_2, and τ_f, where τ_1 and τ_f are related by the excess phase shift (which can be determined experimentally or estimated) and τ_f and $\tau_1 + \tau_2$ can be calculated from other non-quasi-static models. In other words, although models of effective time constant are compact and accurate, they rely on either A. C. measurements or other non-quasi-static models to provide them with reliable time constants.

Extended charge control method

The last approach for modelling non-quasi-static effects is the extended charge control models [136, 137, 155–158]. It extends the charge-control approach in that it models the small-signal A.C. quantities using parameters that can be determined from static device quantities and profile data. In fact, it is the same as the effec-

[12]It should be noted all admittances discussed in this chapter do not include the junction and parasitic admittance since it has been assumed that only delay times associated with the neutral regions contribute to f_T as discussed earlier.

tive time constant approach, except that in this case, the time constants can be expressed in terms of quasi-static parameters such as τ_e, τ_b, α_e, and α_b, which in turn depend only on profile data and static charge distributions. This technique takes advantage of the accuracy of the effective time constant approach and the fact that expressions for *certain* static quantities can be analytically derived from other non-quasi-static models, especially those categorized under the second approach. In short, this approach is the most compact, accurate and computationally efficient among all the approaches described above.

Given that devices with many different Ge and doping profiles will be compared in this chapter, an approach which can accurately model non-quasi-static effects by parameters, calculated from analytical expressions in terms of static charge distributions and profile data in a computationally efficient way, is required. As a result, this extended charge-control approach, and in particular the model proposed by Hamel [136, 137, 156–158], has been chosen for this study. In the following, we give a brief description of how the final expressions of Hamel's non-quasi-static model for the current gain are obtained as they were reported without detailed derivation in the literature.

Non-quasi-static current gain model

Using the notation in two-port network theory, one can express the common-emitter gain β as:

$$\beta = \frac{y_{21}^E}{y_{11}^E} \tag{2.78}$$

since

$$
\begin{aligned}
y_{21}^E &= \left.\frac{i_c}{v_{be}}\right|_{v_{ce}=0} \\
y_{11}^E &= \left.\frac{i_b}{v_{be}}\right|_{v_{ce}=0}
\end{aligned}
\tag{2.79}
$$

where the superscript "E" refers to the common-emitter configuration and i_b, i_c, v_{be} and v_{ce} are the small-signal base current, collector current, base-emitter voltage and collector-emitter voltage, respectively. Therefore, if y_{21}^E and y_{11}^E can be accurately modelled, then an accurate model for β will be obtained. It should be noted that since y_{21}^E is related to i_c, only the neutral base region is considered in its calculation. On the other hand, y_{11}^E is related to i_b which has two major current components: the hole current due to the recombination in the neutral base region (I_{bb}) and the hole current injected into the emitter at the emitter-base junction (I_{be}). Therefore, both the neutral base and the neutral emitter are considered in its calculation. For convenience, y_{11}^E is written as:

$$
y_{11}^E = y_{11b}^E + y_{11e}^E \tag{2.80}
$$

where the subscripts "b" and "e" refer to the neutral base and neutral emitter regions, respectively.

Let us first consider y_{21}^E. We can start from the Seitchik *et al.* 's model (2.76) and then derive expressions for τ_1 and τ_2. Rinaldi [153] showed that the common-base trans-admittance y_{21}^B can be expressed as the reciprocal of the polynomial of the complex variable '$j\omega$' as follows:

$$
y_{21}^B = -\frac{g_m}{1 + j\omega\tau_{D1} + (j\omega\tau_{D2})^2 + \cdots} \tag{2.81}
$$

From two-port network theory, $y_{21}^E = -y_{21}^B$ (assuming y_{22}^B is negligible[13]), therefore,

$$y_{21}^E = \frac{g_m}{1 + j\omega\tau_{D1} + (j\omega\tau_{D2})^2 + \cdots} \tag{2.82}$$

Truncating after the first-order term in the denominator, (2.82) is approximated as:

$$y_{21}^E = \frac{g_m}{1 + j\omega\tau_{D1}} \tag{2.83}$$

Multiplying both numerator and denominator by $e^{j\omega\tau_2}$ and applying series expansions, (2.83) then becomes:

$$y_{21}^E = \frac{g_m e^{-j\omega(\tau_{D1}-\tau_2)}}{1 + j\omega\tau_2} \tag{2.84}$$

Comparing (2.84) and (2.76) yields:

$$\tau_1 + \tau_2 = \tau_{D1} \tag{2.85}$$

if α_o is assumed to be unity. Since at low frequencies, an admittance can be expressed as:

$$y = G + j\omega C \tag{2.86}$$

where G and C are frequency-independent constants corresponding to the small-

[13]This is a good assumption since in the case of a practical bipolar transistor, the output conductance needs to be small.

signal conductance and capacitance, respectively, one can re-write y_{21}^B as:

$$y_{21}^B = -\alpha_o g_m + j\omega(1-\alpha_b)\frac{dQ_b}{dV_{EB}} \tag{2.87}$$

where α_o is the common-base D. C. current gain. Equation (2.87) is the same as equation (3.35) in [159]. Therefore (according to two-port network theory),

$$y_{21}^E = \alpha_o g_m - j\omega(1-\alpha_b)\frac{dQ_b}{dV_{BE}} \tag{2.88}$$

where [148, 151]

$$\alpha_b = \frac{\int_o^{W_b} n(x)^2/n_o(x)dx}{e^{V_{BE}/V_T}\int_0^{W_b} n(x)dx} \tag{2.89}$$

Substituting (2.69) into (2.88) gives:

$$y_{21}^E = g_m[1 - j\omega(1-\alpha_b)\tau_b] \tag{2.90}$$

if α_o is assumed to be unity. This is the same as Hurkx's result (see (2.71)). Comparing (2.90) and (2.83) after series expansion yields:

$$\tau_{D1} = (1-\alpha_b)\tau_b \tag{2.91}$$

From (2.85) and (2.91),

$$\tau_1 + \tau_2 = (1-\alpha_b)\tau_b \tag{2.92}$$

which is equation (4) in [156]. By observing that the excess phase shift ($= \tau_1/\tau_f = \tau_1/\tau_b$, see equation (8) in [143]) is linearly proportional to the quantity $(1-\alpha_b)$,

Hamel [136,156] proposed the following:

$$\tau_1 \approx \tau_2 \approx \frac{(1-\alpha_b)\tau_b}{2} \tag{2.93}$$

Therefore, from (2.93) and (2.76),

$$y_{21}^E = \frac{I_c}{V_T} \cdot \frac{e^{-j\omega(1-\alpha_b)\tau_b/2}}{1+j\omega(1-\alpha_b)\tau_b/2} \tag{2.94}$$

Now, let us consider y_{11}^E. In order to obtain the expression for y_{11b}^E, let's apply Seitchik *et al.* 's results to the neutral base region only, i.e., the hole current will be just the current due to neutral base recombination (I_{bb}) and $\tau_f = \tau_b$. Substituting (2.93) into (2.75) yields:

$$y_{11b}^E = \frac{I_{bb}}{V_T}\left[1 + \frac{j\omega\beta_b\tau_b}{1+j\omega(1-\alpha_b)\tau_b/2}\right] \tag{2.95}$$

where $\beta_b = I_c/I_{bb}$. Note: $\beta_{dc} \triangleq \frac{I_c}{I_{be}+I_{bb}} = \frac{1}{1/\beta_e+1/\beta_b}$.

Considering the low-frequency y_{11}^B in terms of charge-partitioning factors, Hamel [159] wrote:

$$y_{11}^B = g_m + j\omega\left(\alpha_b\frac{dQ_b}{dV_{EB}} + \alpha_e\frac{dQ_e}{dV_{EB}}\right) \tag{2.96}$$

According to two-port network theory, $y_{11}^E = y_{11}^B + y_{21}^B$ (for small output conductance), therefore, (2.96) and (2.87) together gives:

$$y_{11}^E = (1-\alpha_o)g_m + j\omega\left(\frac{dQ_b}{dV_{BE}} + \alpha_e\frac{dQ_e}{dV_{BE}}\right) \tag{2.97}$$

Since $\alpha_o = 1/(1+\beta_o)$ and $\beta_o = I_c/(I_{be}+I_{bb})$, (2.97) becomes:

$$y_{11}^E = \frac{g_m}{\beta_e} + j\omega\alpha_e \frac{dQ_e}{dV_{be}} + \frac{g_m}{\beta_b} + j\omega \frac{dQ_b}{dV_{be}} \tag{2.98}$$

where $\beta_e = I_c/I_{be}$. With the charge-control relations $g_m\tau_e = \frac{dQ_e}{dV_{be}}$ and $g_m\tau_b = \frac{dQ_b}{dV_{be}}$, (2.98) becomes:

$$y_{11}^E = \frac{I_{be}}{V_T}(1 + j\omega\alpha_e\beta_e\tau_e) + \frac{I_{bb}}{V_T}(1 + j\omega\beta_b\tau_b) \tag{2.99}$$

which can be easily shown to be equivalent to Hurkx's result (2.71) on the base current. Comparing (2.99) with (2.80), one can separate y_{11e}^E and y_{11b}^E:

$$y_{11e}^E = \frac{I_{be}}{V_T}(1 + j\omega\alpha_e\beta_e\tau_e) \tag{2.100}$$

and

$$y_{11b}^E = \frac{I_{bb}}{V_T}(1 + j\omega\beta_b\tau_b) \tag{2.101}$$

Since both (2.95) and (2.101) are expressions for y_{11b}^E, which one should be used? Rinaldi [153] has shown that (2.95) includes the second order term in the infinite number of zeros in the numerator of the polynomial of the complex variable whereas (2.101) includes only the first order terms. Therefore, (2.95) is more accurate. Similarly, (2.100) does not contain any high order terms and therefore is not accurate to model the actual y_{11e}^E as frequency increases. From the expression (2.100), it can be readily seen that y_{11e}^E will exhibit the classic 6 dB/octave rise as frequency increases beyond $1/(2\pi\alpha_e\beta_e\tau_e)$. This will fail to explain the 3 dB/octave increase observed at very high frequencies. At very high frequencies, the admittance should

be proportional to $\sqrt{\omega}$ (see equation (2-86) in [145] and equation (4.30) in [160]). In light of this, Hamel [137,158] proposed that (2.100) should be modified as follows:

$$y_{11e}^{E} = \frac{I_{be}}{V_T}\left[1 + \frac{j\omega\alpha_e\beta_e\tau_e}{\sqrt{1 + j\omega(1 - \alpha_e)\beta_e\tau_e}}\right] \tag{2.102}$$

It should be noted that (2.100) and (2.102) only differ in the denominator of (2.102). The square root in the denominator will cause a 3 dB/octave rise when the frequency increases beyond $1/[2\pi(1-\alpha_e)\beta_e\tau_e]$. For frequencies below this, the numerator will dominate and still show the classic 6 dB/octave rise with frequency.

If the denominator is approximated by a series expansion, (2.102) becomes:

$$y_{11e}^{E} = \frac{I_{be}}{V_T}\left[1 + \frac{j\omega\alpha_e\beta_e\tau_e}{1 + j\omega(1 - \alpha_e)\tau_{e,transit}/2}\right] \tag{2.103}$$

where $\tau_{e,transit}$ is the emitter transit time $\tau_e\beta_e$. Observing that (2.103) and (2.95) have the same form, Hamel suggested that (2.95) should also be modified in order to model the 3 db/octave increase at very high frequencies. Therefore, y_{11b}^{E} becomes:

$$y_{11b}^{E} = \frac{I_{bb}}{V_T}\left[1 + \frac{j\omega\beta_b\tau_b}{\sqrt{1 + j\omega(1 - \alpha_b)\tau_b}}\right] \tag{2.104}$$

Finally, the common-emitter current gain can be obtained from (2.78), (2.94), (2.103) and (2.104) and expressed as:

$$\beta(\omega) = \frac{\beta_e e^{-j\omega(1-\alpha_b)\tau_b/2}}{\left[1 + \frac{j\omega(1-\alpha_b)\tau_b}{2}\right]\left\{1 + \frac{j\omega\alpha_e\beta_e\tau_e}{\sqrt{1+j\omega(1-\alpha_e)\beta_e\tau_e}} + \frac{\beta_e}{\beta_b}\left[1 + \frac{j\omega\beta_b\tau_b}{\sqrt{1+j\omega(1-\alpha_b)\tau_b}}\right]\right\}} \tag{2.105}$$

If $I_{bb} \to 0$, $\beta_b \to \infty$, then $\beta_{dc} = \beta_e$ and (2.105) becomes:

$$\beta(\omega) = \frac{\beta_{dc} e^{-j\omega(1-\alpha_b)\tau_b/2}}{\left[1 + \frac{j\omega(1-\alpha_b)\tau_b}{2}\right]\left[1 + \frac{j\omega\alpha_e\beta_{dc}\tau_e}{\sqrt{1+j\omega(1-\alpha_e)\beta_{dc}\tau_e}} + \frac{j\omega\beta_{dc}\tau_b}{\sqrt{1+j\omega(1-\alpha_b)\tau_b}}\right]} \tag{2.106}$$

From (2.106), it becomes clear why the current gain model derived from charge-control theory (e.g.(2.68)) fails to predict f_T when non-quasi-static effects are significant because there is no guarantee that the drop will follow the classic 6 dB/octave slope when the frequency is increased near f_T. Only when $\alpha_e = \alpha_b = 1$ will the classic 6 dB/octave fall-off be guaranteed as (2.106) reduces to (2.68) with $\beta_o = \beta_{dc}$.

In short, (2.106) allows the common-emitter current gain to be accurately modelled at frequencies near f_T and beyond (which will be shown in Section 2.3) in terms of quasi-static parameters $(\tau_b, \tau_e, \alpha_b, \alpha_e)$ that can be determined from static charge distributions and profile data (see (2.63), (2.52), (2.89), and (2.70)). The *effective* total delay time, with non-quasi-static effects taken into account, can then be obtained by setting the magnitude of (2.106) to one and $\tau_f = 1/\omega_t$.

2.3 Verification

In this section, the expressions for τ_e, τ_b, minority carrier profiles in both neutral emitter and base regions, and $\beta(\omega)$ are verified by comparing with published results in the literature and results from numerical simulations.

2.3.1 Emitter and base delay times

The emitter and base delay time expressions ((2.52), (2.63)) are compared with results obtained by Suzuki [110] and Basu *et al.* [112] for $S_m = \infty$. A wide range

of thicknesses for both poly-Si and mono-Si regions is covered. In Suzuki's work, Slotboom and de Graaff's bandgap narrowing model [161], del Alamo *et al.* 's hole mobility [162] and Swirhun *et al.* 's electron mobility [163] were used for the mono-Si emitter and base regions. A hole diffusion length in the poly-Si region of 100 nm was assumed. The ratio of the hole minority mobility in poly-Si region to that in mono-Si was assumed to be 0.3. The recombination velocity S_p at the poly/mono-Si interface was assumed to be 10^4cm/s for HF devices. No Ge was used in the base region of the devices. In Basu *et al.* 's work, the hole diffusion coefficient in mono-Si was assumed to be 1.27 cm^2/s. The ratio of hole mobility in mono-Si to that in poly-Si was taken to be 3. A hole diffusion length of 100 nm was assumed in poly-Si. The quantity S_p was assumed to be zero for HF devices. No Ge was used in the base region of the devices.

In Fig. 2.4, the thickness of the poly-Si region was fixed and the mono-Si thickness was varied from 0 to 300 nm. In Fig. 2.5, the thickness of the mono-Si region was fixed and the poly-Si thickness was varied from 0 to 300 nm. Both figures show that the results based on the τ_e expressions in this work (2.52) agree well with those reported by Suzuki and Basu *et al.* In Fig. 2.6, the emitter and base delay times were plotted against the base width with constant thicknesses for both mono-Si and poly-Si regions. It shows that the results of (2.63) and (2.52) match well with those of Suzuki.

2.3.2 D.C. minority carrier profiles in emitter and base

Since the accuracy of both the delay times and the current gain depend on the accuracy of the carrier profiles, it will be useful to check the minority carrier profiles in both the base and the emitter as calculated by (2.35), (2.43), (2.44), (2.59), and (2.65). The results are compared with a sinusoidal steady-state numerical simula-

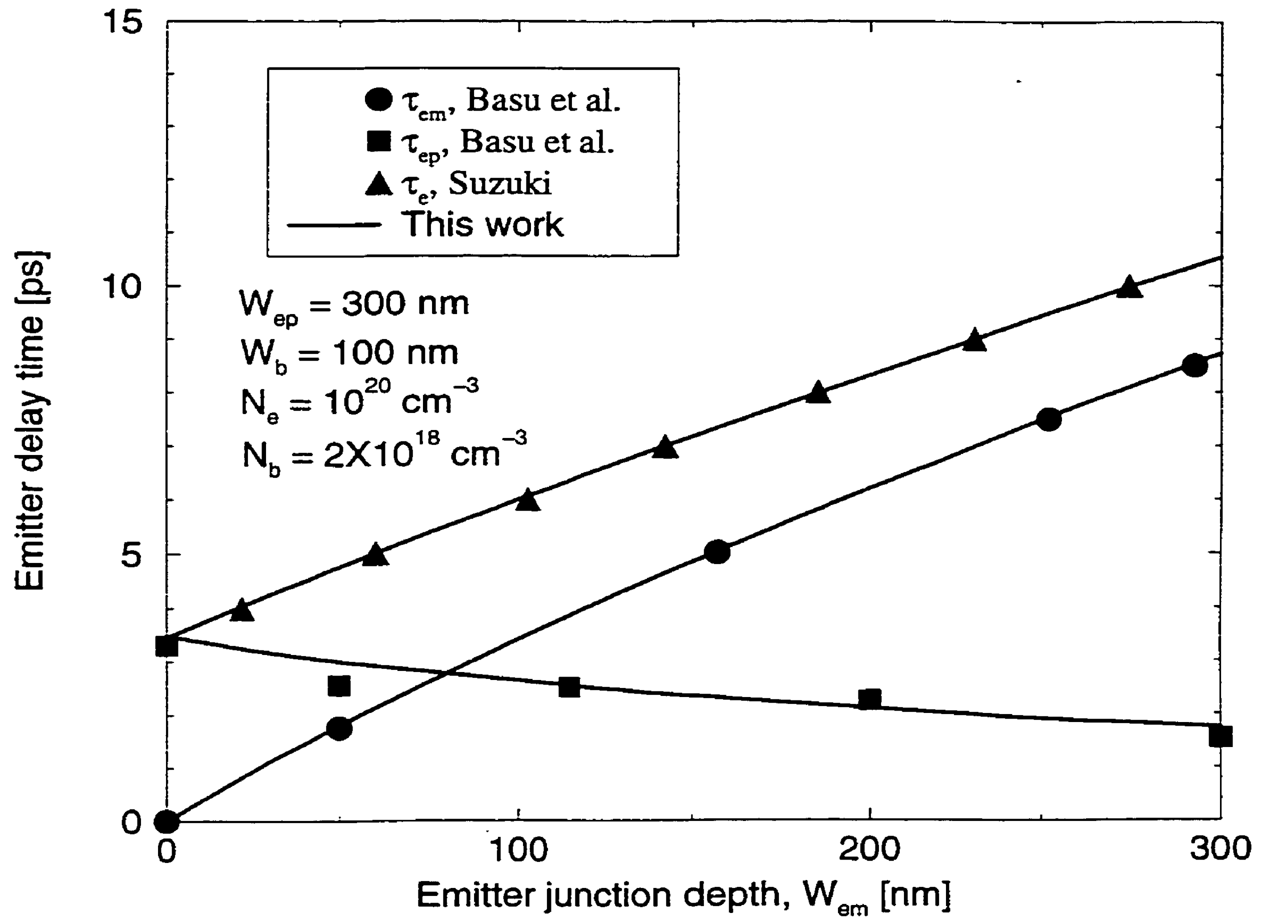

Figure 2.4: Emitter delay time versus emitter junction depth (mono-Si region thickness). Results of Suzuki [110] are the sum of the emitter delay times in poly-Si and mono-Si region. Basu *et al.* 's [112] results were available as individual emitter delay times in mono-Si and poly-Si regions. Infinite recombination velocity at emitter contact (S_m) and uniform emitter doping were assumed in [110, 112].

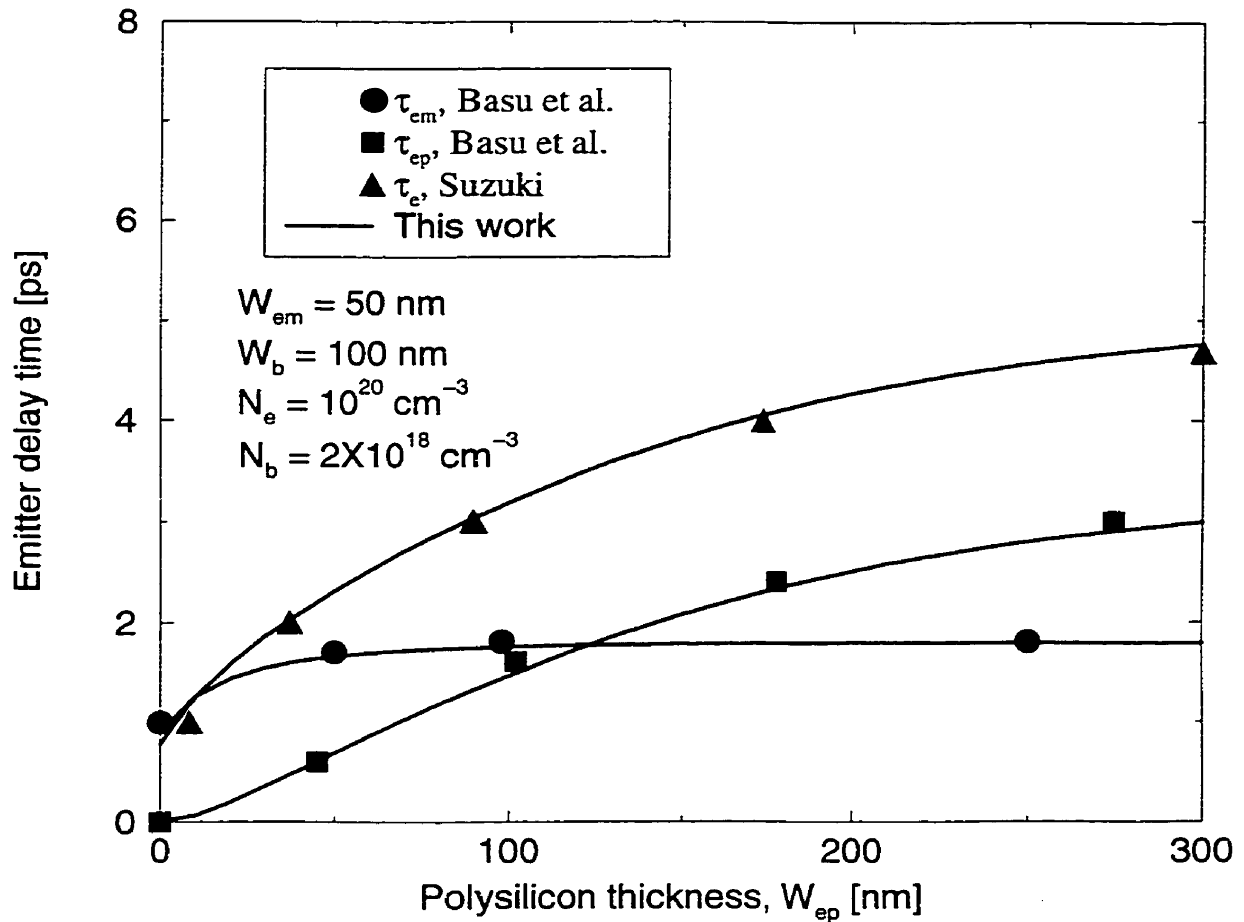

Figure 2.5: Emitter delay time versus polysilicon thickness. Results of Suzuki [110] are the sum of the emitter delay times in the poly-Si and mono-Si regions. Basu *et al.* 's [112] results were available as individual emitter delay times in the mono-Si and poly-Si regions. Infinite recombination velocity at emitter contact (S_m) and uniform emitter doping were assumed in [110, 112].

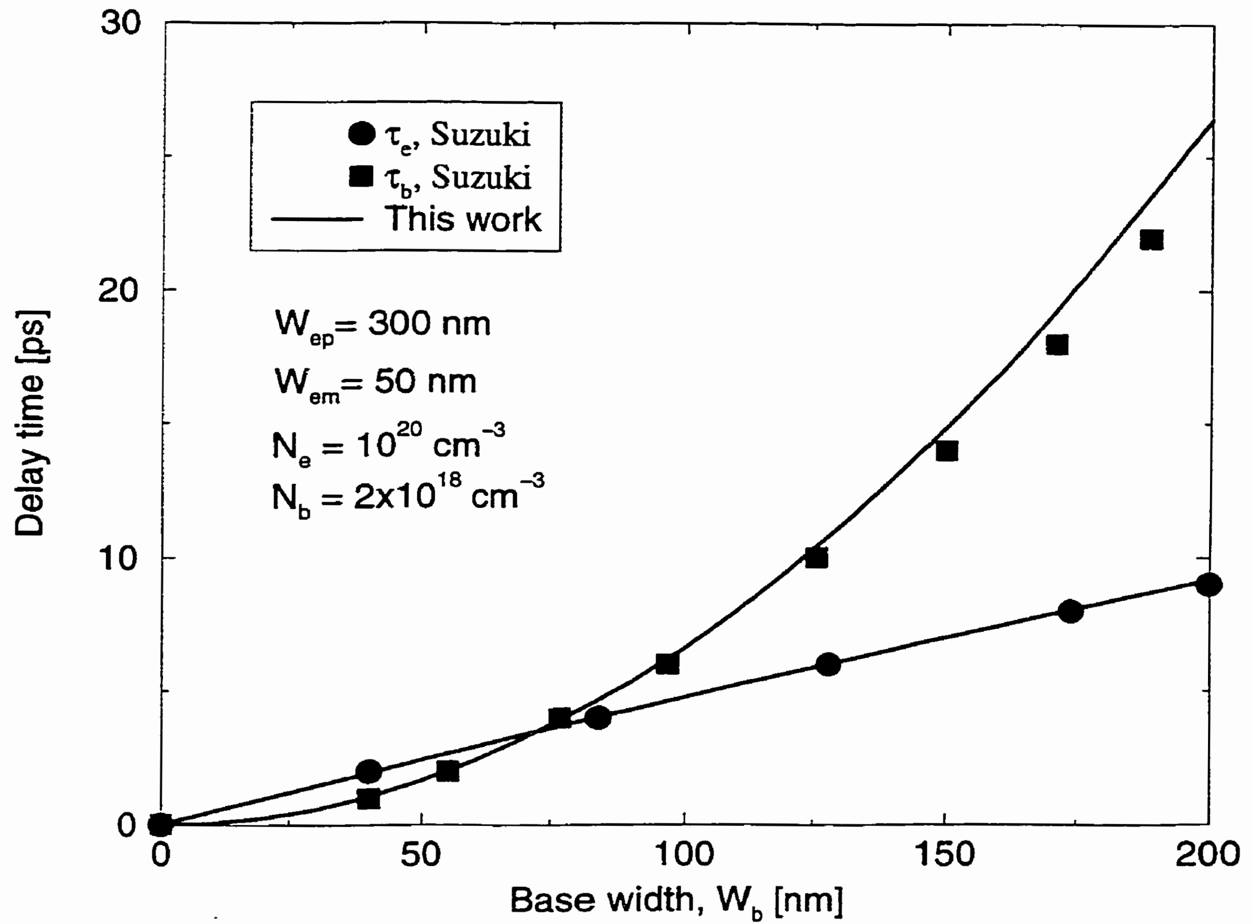

Figure 2.6: Base and emitter delay time versus base width. Infinite recombination velocity at emitter contact (S_m) and uniform doping in both base and emitter were assumed in [110].

tor called S^3A, developed at the University at Waterloo [159, 164, 165]. S^3A solves the 1-D A.C. continuity equation using a finite-difference discretization. Recombination, heavy doping effects, doping and field dependence of diffusion coefficients, finite recombination velocity at edges of neutral regions, and non-uniform bandgap due to the Ge presence are all incorporated in the implementation. Since the effect of recombination velocity at the edges of neutral regions can be easily reflected in the carrier profile, S^3A is an ideal tool for checking the validity of τ_e expression especially in the case where S_m is finite. Furthermore, since S^3A takes recombination into account, the comparison can show the validity of assuming negligible recombination in the mono-Si and neutral base regions, made in the the previous sections.

Four test devices of different base profiles, base widths, mono-Si emitter thicknesses, poly-Si emitter thicknesses, and Ge gradings are used for the comparison (see Table 2.1). Doping levels of the poly-Si, mono-Si emitter, epitaxial collector regions are 10^{20}, 10^{18}, and 10^{17}cm^{-3}, respectively. The Gaussian-doped transition region between the poly-Si and mono-Si emitters has a width of $(X_{em} - w_1)/4$. The parameters used in the comparison are as follows. The recombination velocity at the emitter contact (S_m) is 10^6 cm/s. The recombination velocity at the collector-base junction (S) is 10^7 cm/s. The recombination velocity at the poly/mono-Si interface is set to zero since S^3A does not model the interface recombination. The tunneling probability α is set to unity since only HF devices are studied. Definitions of doping and Ge profiles given at the beginning of this chapter are used. Identical physical models are used by both S^3A and analytical expressions derived in this chapter. A forward bias of 0.7 V is applied across the base-emitter junction and a reverse bias of 2 V across the collector-base junction. Depletion layer widths for the test structures under such biases are calculated by solving the Poisson equation with

Device	W_{ep} [μm]	X_{em} [μm]	N_{bo} [cm^{-3}]	X_b [nm]	θ	w_1 [nm]	w_2 [nm]	w_3 [nm]	$y(0)$	$y(W_b)$
1	0.38	0.02	5×10^{19}	50	0.5	16.64	6.81	10.76	0.1	0.1
2	0.25	0.15	5×10^{19}	50	0.2	17.96	4.09	15.07	0	0.2
3	0.2	0.2	5×10^{19}	30	0.8	16.68	6.67	3.74	0.2	0
4	0.1	0.3	2×10^{19}	100	0.5	16.95	10.86	20.17	0.1	0.1

Table 2.1: Profile details of SiGe test devices.

the depletion approximation using the numerical method suggested by Lin [166].

Fig 2.7 compares results generated by analytical expressions derived in the previous section and those by S^3A for the four SiGe test HBTs specified in Table 2.1. Excellent agreement is obtained for all four devices. The good match of minority carrier concentrations at the edges of the devices indicates that the finite recombination velocities at the emitter contacts and collector-base junction are modelled accurately. Also, the good match implies that the assumption of negligible recombination in the mono-Si neutral region and the neutral base region is justified.

2.3.3 A.C. current gain and total delay time

To check the validity of (2.106), the A.C. current gains for the four test devices specified in Table 2.1 calculated by (2.106) over a wide range of frequencies are compared with results obtained from S^3A (Note: S^3A does not make any quasi-static approximation). Fig. 2.8 compares S^3A simulation results with those obtained from analytical expressions based on the quasi-static assumption ((2.68)) and those from non-quasi-static correction ((2.106)). It shows that the results of the expression, based on the quasi-static assumption, and the results based on non-quasi-static correction agree well with S^3A results from low frequencies to around one order of magnitude above the 3-dB corner frequency, for all devices. However, when the fre-

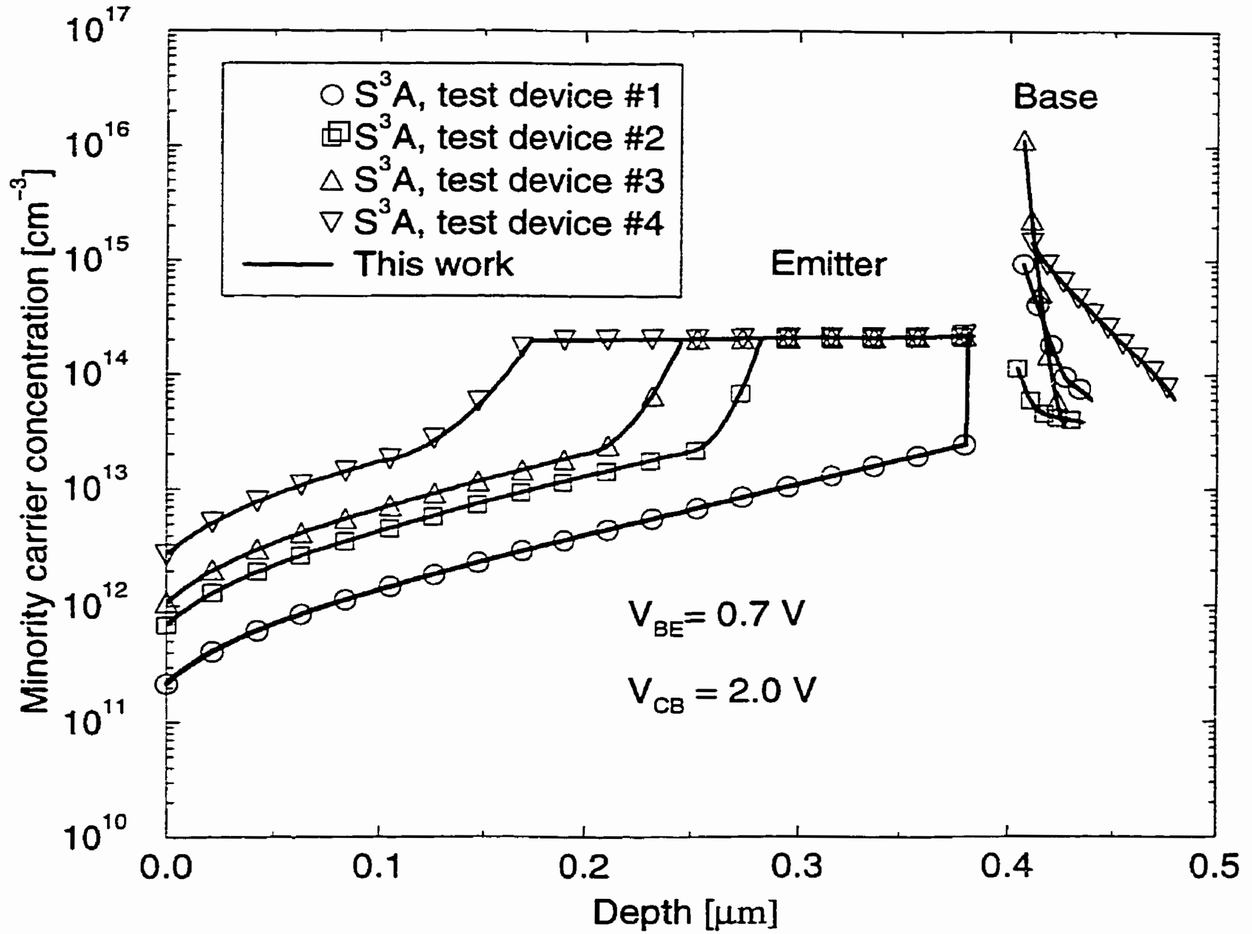

Figure 2.7: D. C. minority carrier profiles in neutral emitter and base regions of SiGe HBTs. Comparison of results generated by analytical expressions, derived in this work, and those from the numerical simulator S^3A ($S_m = 10^6$ cm/s, $S = 10^7$ cm/s, $S_p = 0$ cm/s).

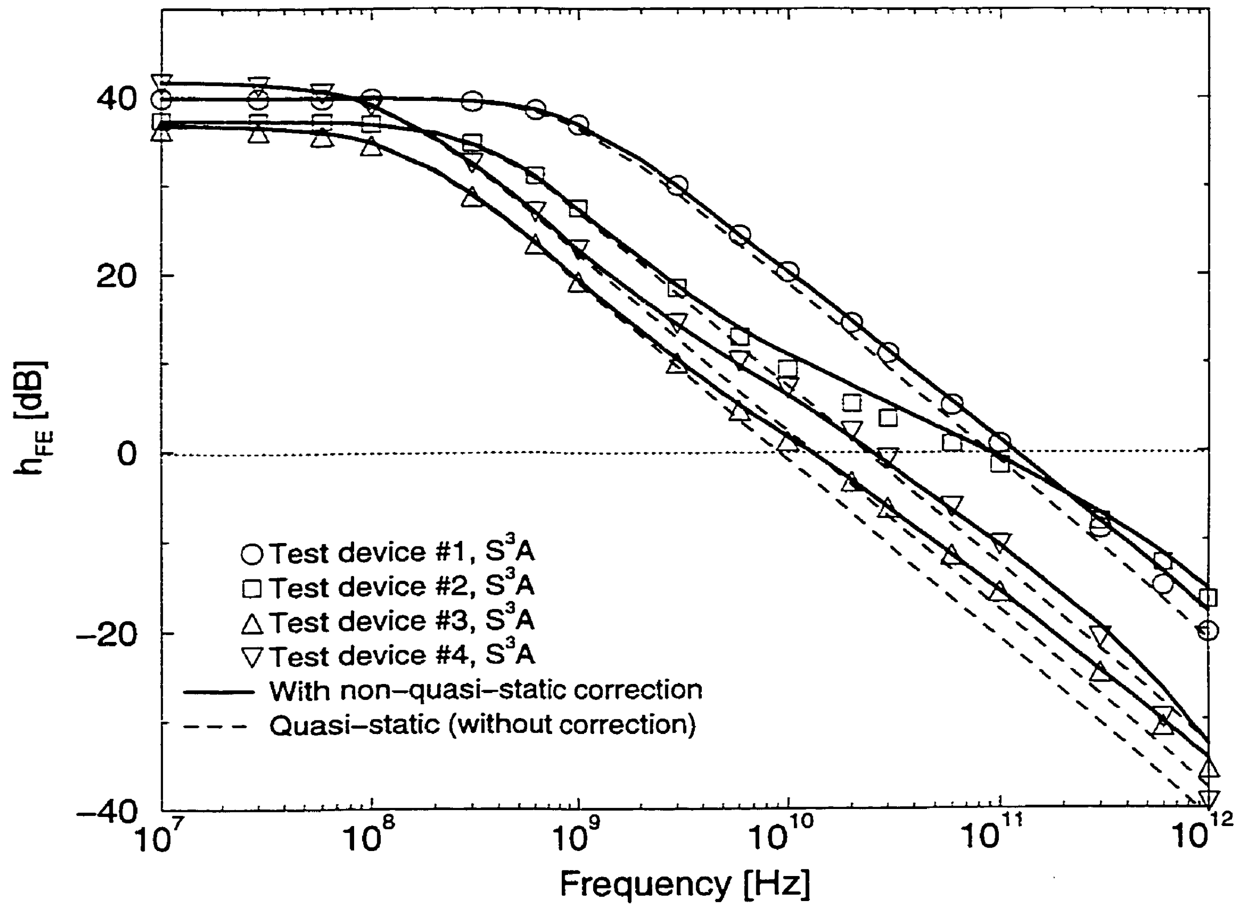

Figure 2.8: A.C. current gain (h_{FE} or β) as a function of the frequency of SiGe test HBTs. Comparing of numerical simulation results from S^3A (symbols) with results generated by analytical expressions, based on the quasi-static assumption (dash lines, (2.68)), and those using non-quasi-static correction (solid lines, (2.106)).

quency increases toward f_T and beyond, only the expression with non-quasi-static correction can predict the A.C. current gain accurately for all four devices. The quasi-static approximation predicts a total delay time of 6.82 ps and 12.1 ps for test devices #2 and #4, respectively. But the actual total delay times predicted by S^3A are only 2.2 ps and 6.0 ps, respectively. It shows that the quasi-static approximation here over-estimates the total delay time and consequently f_T by a factor of 2 to 3. This affirms the necessity of accounting for non-quasi-static effects when calculating the total delay times and also of the accuracy of the analytical expression (2.106). Furthermore, although not shown, devices with uniformly-doped poly-Si and Gaussian-doped mono-Si region of different thicknesses have also been tested and the results have the same degree of accuracy as seen here.

2.4 Results and Discussions

In this section, the emitter and base delay times and the total delay time are calculated with and without non-quasi-static correction. The calculations are performed for Ge profiles of identical Ge dose incorporated into the base of SiGe HBTs with different emitter geometries and base profiles. Profile definitions presented in Section 2.2 are followed. For the Ge profiles, $y_{av} = 0.1$ and $\frac{-0.2}{W_b} < g < \frac{0.2}{W_b}$ are used. For all devices studied in this section, the following assumptions are made: $V_{BE} = 0.7$ V and $V_{CB} = 2$ V, $N_{ep} = 10^{20}\text{cm}^{-3}$, $N_{em} = 10^{18}\text{cm}^{-3}$, $N_{epi} = 10^{17}\text{cm}^{-3}$, $W_{em1} = (X_{em} - w_1)/4$, $\alpha = 1$ (i.e. HF devices), $S_m = 10^6$ cm/s, $S = 10^7$ cm/s (except for devices with uniform base doping where an infinite recombination velocity is assumed). Depletion layer widths are calculated in the same manner as described in Section 2.3.

The quantity S_p is set to zero. As shown in Appendix A, S_p is actually the

Transistor	W_{ep} [μm]	X_{em} [μm]	X_b [nm]	θ	N_{bo} [cm^{-3}]	R_b [$k\Omega/\square$]
1a	0.38	0.02	100	-†	7.5×10^{18}	1.15
1b	0.38	0.02	100	0.2	2×10^{19}	1.22
1c	0.38	0.02	100	0.5	2×10^{19}	1.15
1d	0.38	0.02	100	0.8	2×10^{19}	1.08

Table 2.2: Profile details of transistor set #1 (†indicates a uniform base profile).

recombination velocity at the oxide/mono-Si interface (called S_{is} by Rinaldi [109]). Referencing Patton *et al.* 's work [167], Suzuki set S_p to 10^4–5×10^4 cm/s for HF devices. However, in Patton *et al.* 's work, the surface saturation current density (J_{os}) and not S_p was reported. It appears that Suzuki extracted S_p by incorrectly treating it as the *effective* recombination velocity at the poly/mono-Si interface (i.e. S_p' in the definition adopted here) and then calculating S_p' from J_{os} and the equilibrium hole concentration at the poly/mono-Si interface. The equilibrium hole concentration can be estimated from Slotboom and de Graaff's bandgap narrowing model [161] given the doping density in the poly-Si region. This likely over-estimates S_p. Based on the values of the hole capture cross section, the thermal velocity, and the total density of interface traps per unit area, Yu *et al.* [106] estimated S_p to be around 1.6×10^3 cm/s. However, as noted by Ashburn *et al.* [107], there are no reliable experimental values for S_p yet. As a result of this, and that only HF devices are studied, S_p is consistently set to zero in this section. Even when a value of 1.6×10^3 cm/s is used for S_p, the effective recombination velocity S_p' varies by less than 5% when compared with the case where a zero value of S_p is used since S_{poly} is in the order of at least 3.8×10^4 cm/s (for 0.1 μm $< W_{ep} <$ 0.38 μm and $N_{ep} = 10^{20}$cm^{-3}, see (2.24)).

First, a set of transistors with the same emitter structure and base width but

different base profiles are studied (see Table 2.2). Fig 2.9 plots the emitter delay time for transistor set #1 as a function of the Ge grading. It shows that the minimum τ_e point for the uniform doping profile (transistor # 1a) occurs at zero Ge grading (i.e. a Ge box profile). The same result was obtained by Roulston and McGregor [97]. However, for non-uniform base profiles (transistors #1b, 1c, 1d), the minimum τ_e points vary with the location of the profile doping peak. The closer the peak is to the emitter-base junction, a less positive Ge grading is required to reach the minimum τ_e. For example, for $\theta = 0.8$ (transistor #1d), the peak base doping is located at 80 nm away from the emitter-base junction and the minimum τ_e occurs at a Ge grading of around 10^4/cm. Whereas for $\theta = 0.2$ (transistor #1b), the peak base doping is located at 20 nm away from the emitter-base junction, the minimum τ_e point occurs at a *negative* Ge grading of around 0.5×10^4/cm. For $\theta = 0.5$, the minimum τ_e point is located at a Ge grading between the two. To explain this, the D.C. common-emitter current gain (β_{dc}) is plotted in Fig. 2.10. It shows that the locations of maximum β_{dc} points for different base profiles coincide with their respective minimum τ_e points. This is because τ_e is inversely proportional to β_{dc} in (2.52). However, why do the maximum β_{dc} points follow the peak location of the base doping? Since the current gain is inversely proportional to the effective base Gummel number G_b, from (2.66) the current gain can be minimized if $n_o(x) = n_{ie}^2(x)/N_b(x)$ is maximized over the neutral base. Therefore, if the peak doping location is near the base-emitter junction, a negative Ge grading will give a high Ge fraction near the doping peak to maximize $n_o(x)$ and hence minimize the base Gummel number and maximize β_{dc}.

Fig. 2.11 shows the base delay time as a function of the Ge grading for transistor set #1. As expected, for all four base doping profiles, τ_b decreases with increasing Ge grading. However, τ_b saturates when the Ge grading is very high because the

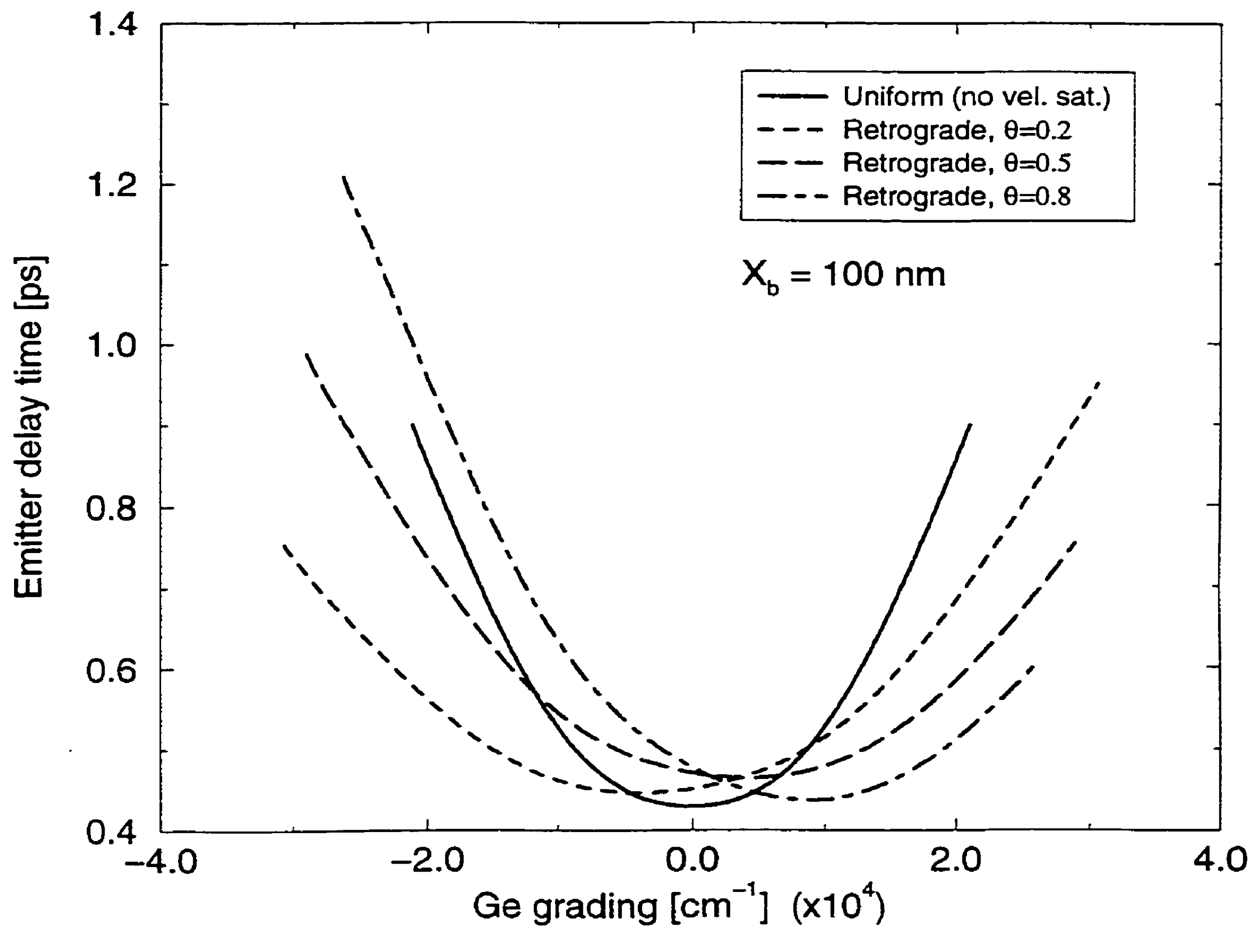

Figure 2.9: Emitter delay time vs. Ge grading for transistor set #1 ($W_{ep} = 0.38$ μm, $X_{em} = 0.02$ μm, $X_b = 100$ nm).

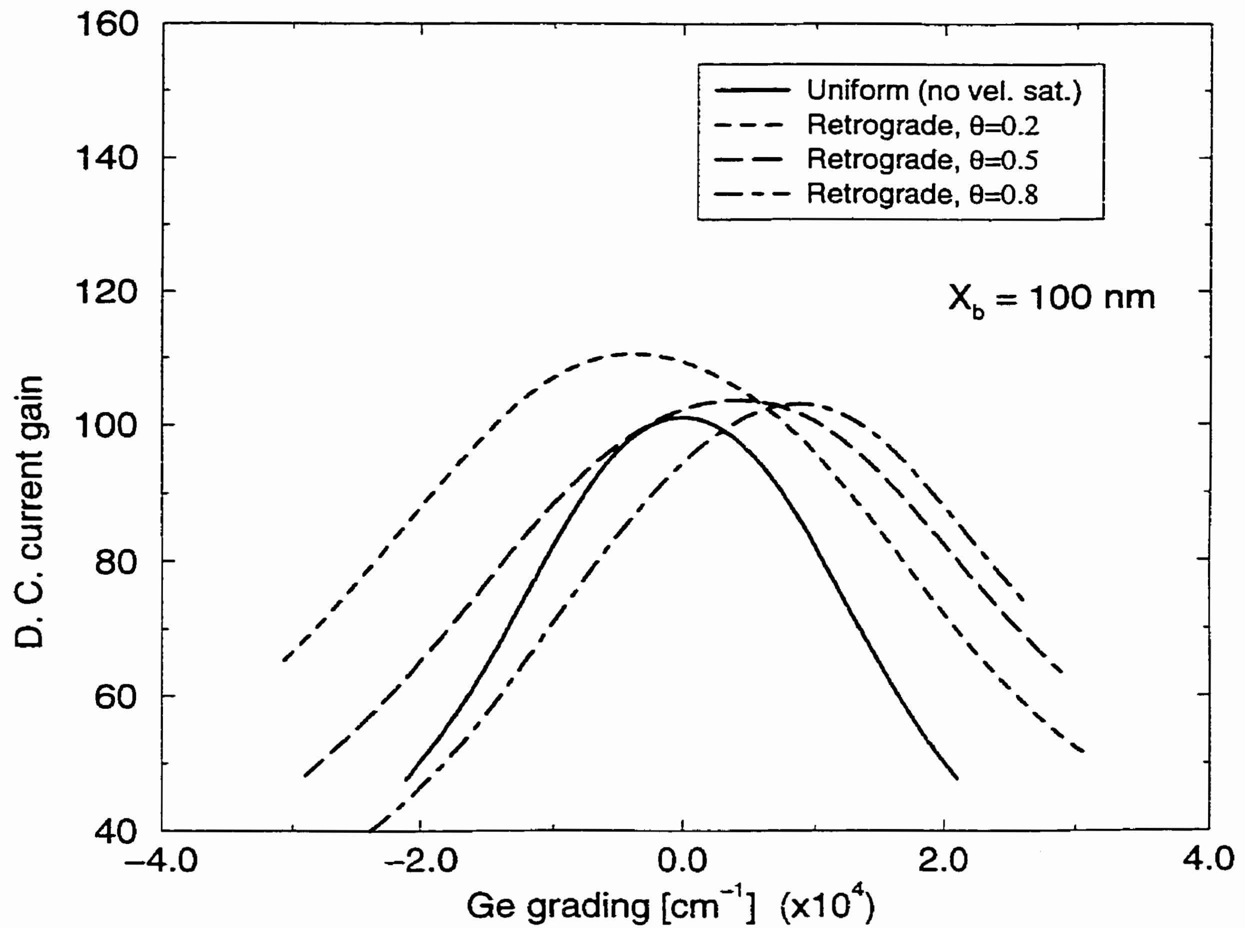

Figure 2.10: D.C. current gain as a function of Ge grading for transistor set #1 ($W_{ep} = 0.38$ μm, $X_{em} = 0.02$ μm, $X_b = 100$ nm).

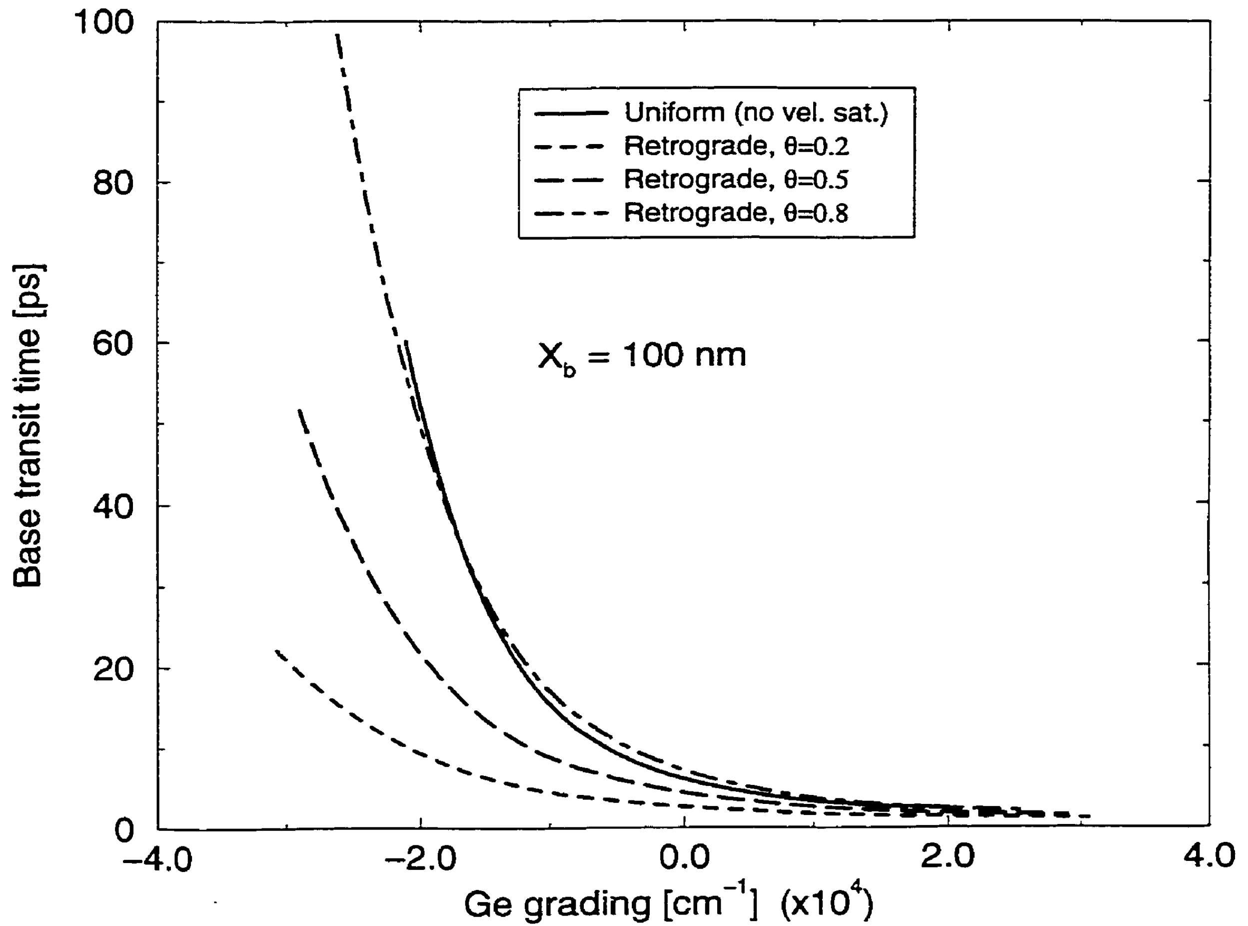

Figure 2.11: Base delay time as a function of Ge grading for transistor set #1 ($W_{ep} = 0.38\ \mu$m, $X_{em} = 0.02\ \mu$m, $X_b = 100$ nm).

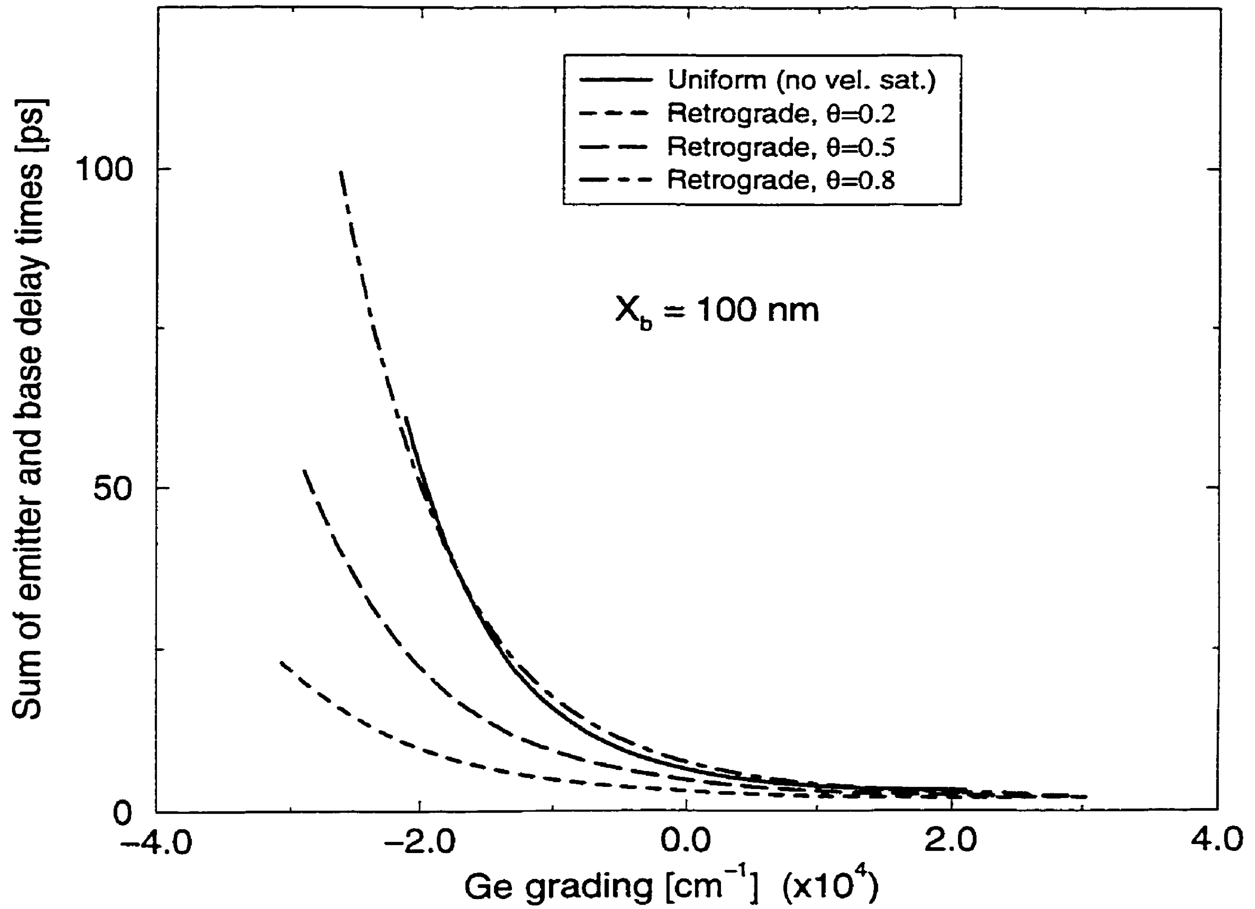

Figure 2.12: Total delay time (without non-quasi-static correction) as a function of Ge grading for transistor set #1 (W_{ep} = 0.38 μm, X_{em} = 0.02 μm, X_b = 100 nm).

effect of the field-dependent mobility begins to set in at such a high Ge grading (i.e. electric field). If the quasi-static approximation is used to estimate f_T, one can simply sum the emitter and base delay times to see the influence of the Ge profile on τ_t ($= 1/f_T$). This is shown in Fig. 2.12. The figure illustrates that the higher the Ge grading, the smaller is the total delay time. In other words, if only base and emitter delay times are concerned, a graded Ge profile with a positive Ge gradient instead of a box Ge profile is preferred. This conclusion is expected because, in

this particular case, τ_b is larger than τ_e by at least an order of magnitude and a higher Ge grading can reduce τ_b as shown in Fig. 2.11. To measure the advantage of adopting a graded Ge profile, with the Ge grading required to minimize τ_t, over the use of a box Ge profile, a graded profile leverage factor (γ) is defined as follows:

$$\gamma = \frac{\tau_t(g=0) - \tau_t(g_{min})}{\tau_t(g=0)} \cdot sgn(g_{min}) \tag{2.107}$$

where τ_t is the either the total delay time (in the case without non-quasi-static correction) or the *effective* total delay time (in the case with non-quasi-static correction), g is the Ge grading, and g_{min} is the Ge grading where τ_t is the minimum and the function $sgn(g_{min})$ is the sign of g_{min} ($= 1$ if $g_{min} \geq 0$ or $= -1$ if $g_{min} < 0$). A factor of $\gamma = 0$ implies that the minimum total delay time or the effective total delay time is attained with a box Ge profile. A positive γ indicates that a graded Ge profile with a positive Ge grading is preferred over the box profile. A negative γ indicates that a graded Ge profile with a negative Ge grading is preferred over the box profile. Quantitatively, the absolute value of γ is the percentage drop, with respect to a box Ge profile, in the effective total delay time when an optimal graded profile is used. From Fig. 2.12, γ is 50% for the uniform base profile and 54% for the retrograde base profile with $\theta = 0.5$.

If non-quasi-static effects are considered, the effective total delay time can be obtained through (2.106) and shown in Fig. 2.13. It is similar to Fig. 2.12 except that the absolute magnitude is generally smaller by less than 1 ps. For uniform base doping, the effective delay time is around 6 ps at zero Ge grading and 2.5 ps at the minimum point where the Ge grading is 2.1×10^4/cm. This translate to a γ of 58%. For the retrograde base profile with $\theta = 0.5$, a γ of 63% can be attained when a graded Ge profile with a grading of 2.9×10^4/cm is used instead of a box

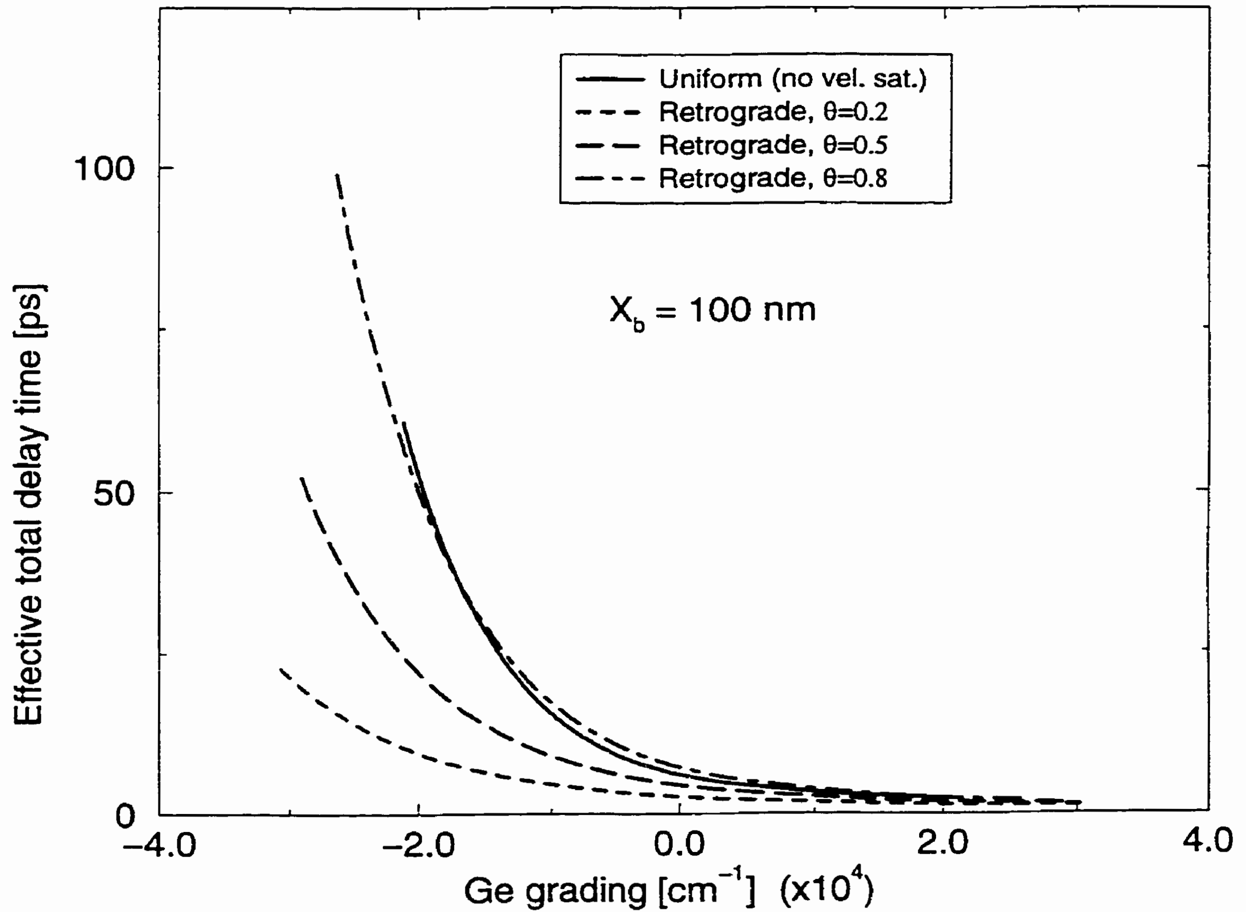

Figure 2.13: Effective total delay time (with non-quasi-static correction) as a function of Ge grading for transistor set #1 ($W_{ep} = 0.38\ \mu$m, $X_{em} = 0.02\ \mu$m, $X_b = 100$ nm).

Transistor	W_{ep} [μm]	X_{em} [μm]	X_b [nm]	θ	N_{bo} [cm^{-3}]	R_b [$k\Omega/\Box$]
2a	0.38	0.02	30	-†	1.8×10^{19}	2.0
2b	0.38	0.02	30	0.2	5×10^{19}	2.16
2c	0.38	0.02	30	0.5	5×10^{19}	2.02
2d	0.38	0.02	30	0.8	5×10^{19}	1.89

Table 2.3: Profile details of transistor set #2 (†indicates a uniform base profile).

profile.

That is, an increase of around 10% in γ (for both base doping profiles) is caused by the non-quasi-static correction. In other words, the conclusion that a graded Ge profile with positive gradient should be used is even more valid with the non-quasi-static effects taken into account, even though the non-quasi-static effects are not very significant in this particular case. The insignificance is due to the fact that the emitter delay time is small relative to the base delay time. From (2.106), one can see that a small τ_e will suppress the denominator term associated with charge sharing in the neutral emitter region. Also, the base-related terms with charge sharing effect are not significant at frequencies much smaller than $1/(1 - \alpha_b)\tau_b$ which is higher than the transition frequency.

But what if the base width is reduced? A second set of transistors with smaller base widths is studied (see Table 2.3). As can be seen from Table 2.3, a higher base concentration is required in order to avoid high base sheet resistances which degrade the maximum oscillation frequency. However, the base concentration cannot be too high, otherwise the depletion layer extending into the emitter will be wider than the mono-Si emitter thickness. Fig 2.14 plots τ_e for the transistors with $X_b = 30$ nm in set #2 as a function of the Ge grading. Compared with transistor set #1 (Fig. 2.9), τ_e is smaller but still in the same order of magnitude. The slight decrease

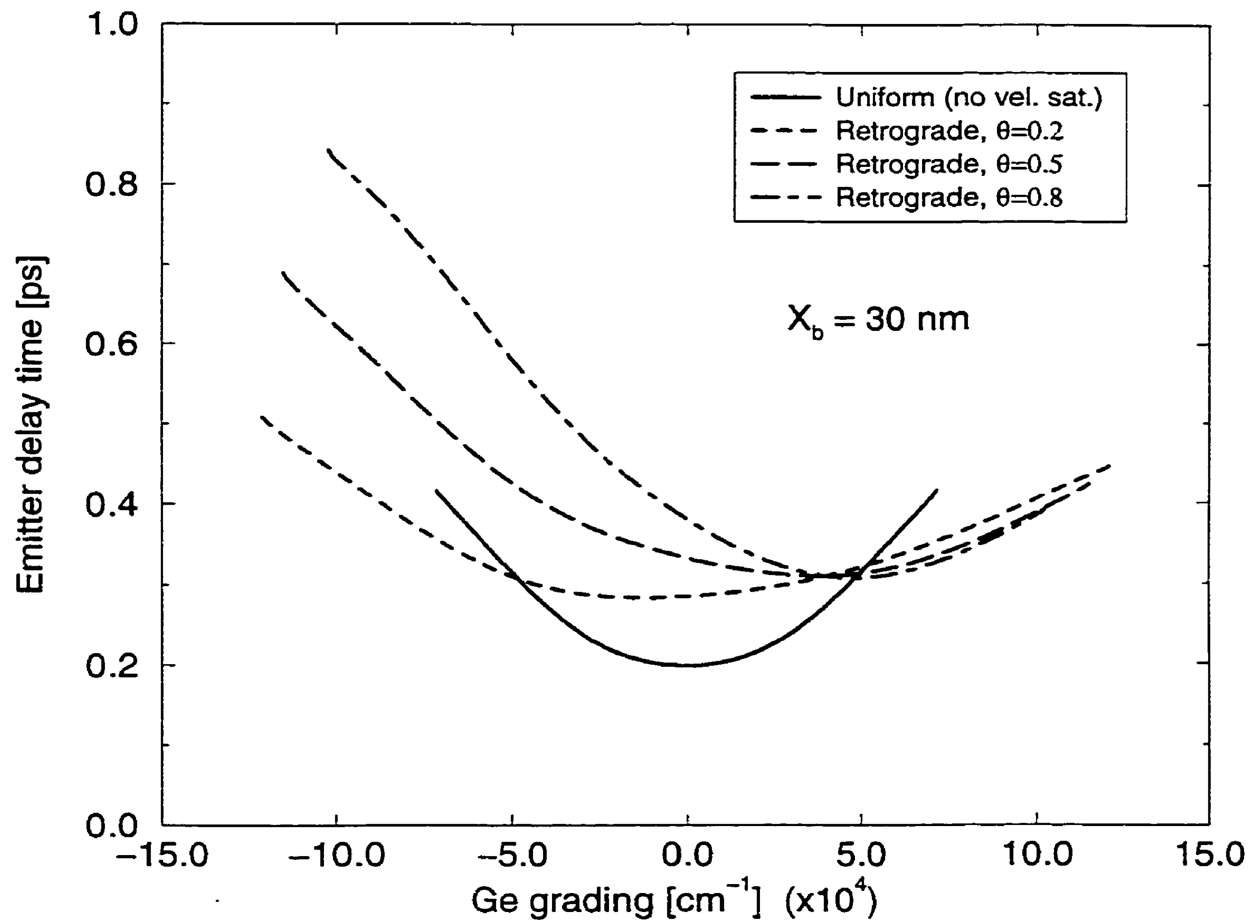

Figure 2.14: Emitter delay time vs. Ge grading for transistor set #2 (W_{ep} = 0.38 μm, X_{em} = 0.02 μm, X_b = 30 nm).

is caused by the base width reduction that decreases G_b and thus increases β_{dc}. The relative locations of the minimum τ_e points on the Ge grading axis are similar to those found in Fig. 2.9. However, the base delay times for transistor set #2 (see Fig. 2.15) are smaller, by an order of magnitude, than those for transistor #1 mainly because of the reduced base width. The total delay time and the effective total delay time are plotted in Fig. 2.16 and Fig. 2.17, respectively.

For the case without non-quasi-static correction (Fig. 2.16), the minimum τ_t point occurs at Ge gradings 3.38×10^4 and 6.71×10^4/cm corresponding to the uniform base and the retrograde base with $\theta = 0.5$, respectively. The graded Ge profile leverage factor γ is 16% and 27% for the uniform base profile and retrograde base profile ($\theta = 0.5$), respectively. After non-quasi-static effects are considered (Fig. 2.17), the minimum effective total delay times occur at higher Ge gradings: 7.2×10^4/cm for the uniform base profile and 11.5×10^4/cm for the retrograde base with $\theta = 0.5$. The leverage factor γ is 45% for the uniform base and 53% for the retrograde base ($\theta = 0.5$).

In comparison with transistor set #1, two observations can be made: i) the graded profile leverage is smaller by 10-13%, i.e., the advantage of using the graded Ge profile over the box Ge profile becomes smaller as the base width is reduced, ii) non-quasi-static effects become more important when the base width is reduced. An increase in γ of 25-30% is obtained when compared with the case without non-quasi-static correction.

The smallest metallurgical base width (X_b) we have studied is 30 nm. For metallurgical base width smaller than 30 nm (i.e., neutral base width (W_b) less than 10 nm for the base and emitter profiles under this study), it will be difficult to estimate τ_t accurately because the drift-diffusion transport theory may not be valid any more. Simulations based on hydrodynamic formulations, Monte Carlo

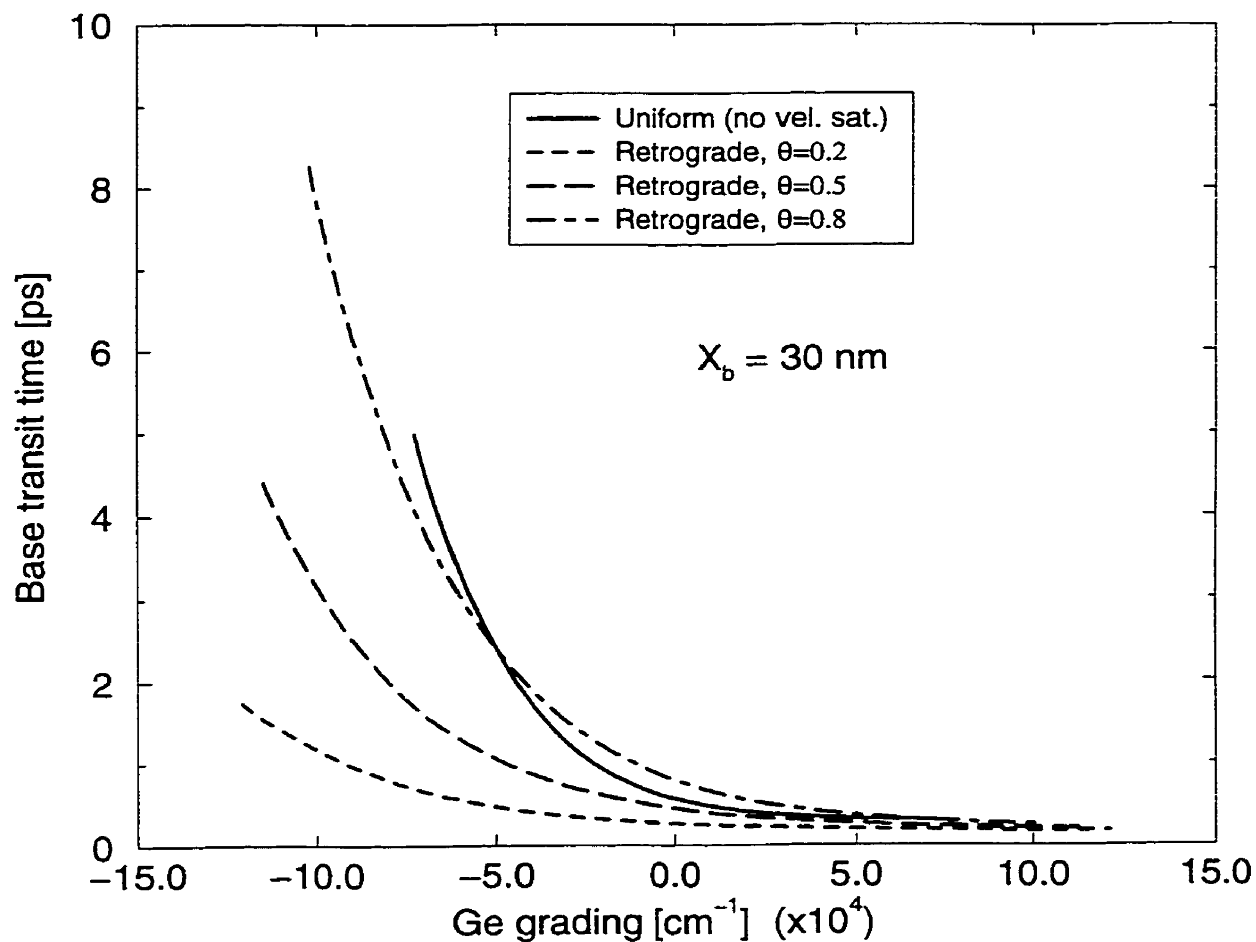

Figure 2.15: Base delay time as a function of Ge grading for transistor set #2 ($W_{ep} = 0.38\ \mu$m, $X_{em} = 0.02\ \mu$m, $X_b = 30$ nm).

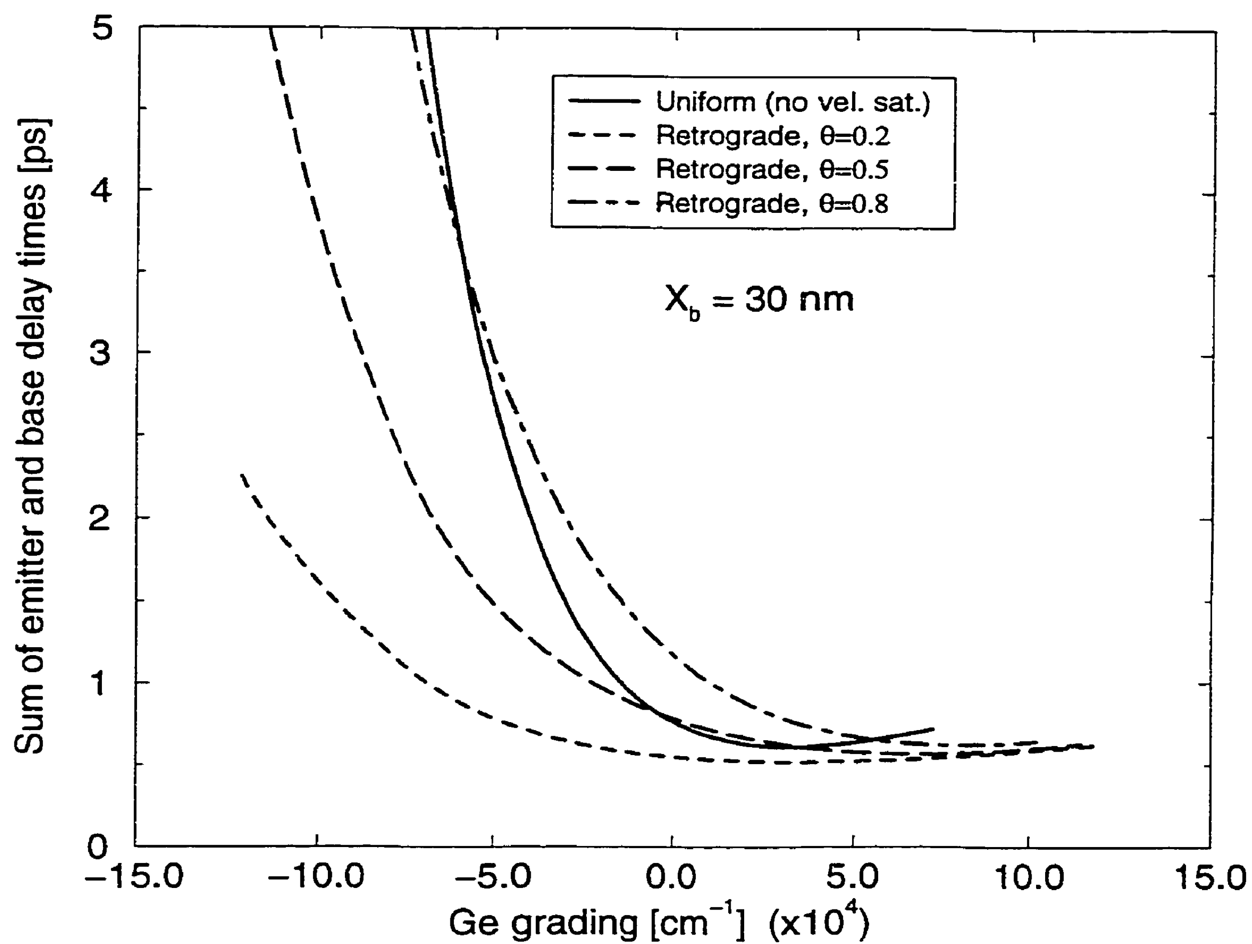

Figure 2.16: Total delay time (without non-quasi-static correction) as a function of Ge grading for transistor set #2 ($W_{ep} = 0.38\ \mu$m, $X_{em} = 0.02\ \mu$m, $X_b = 30$ nm).

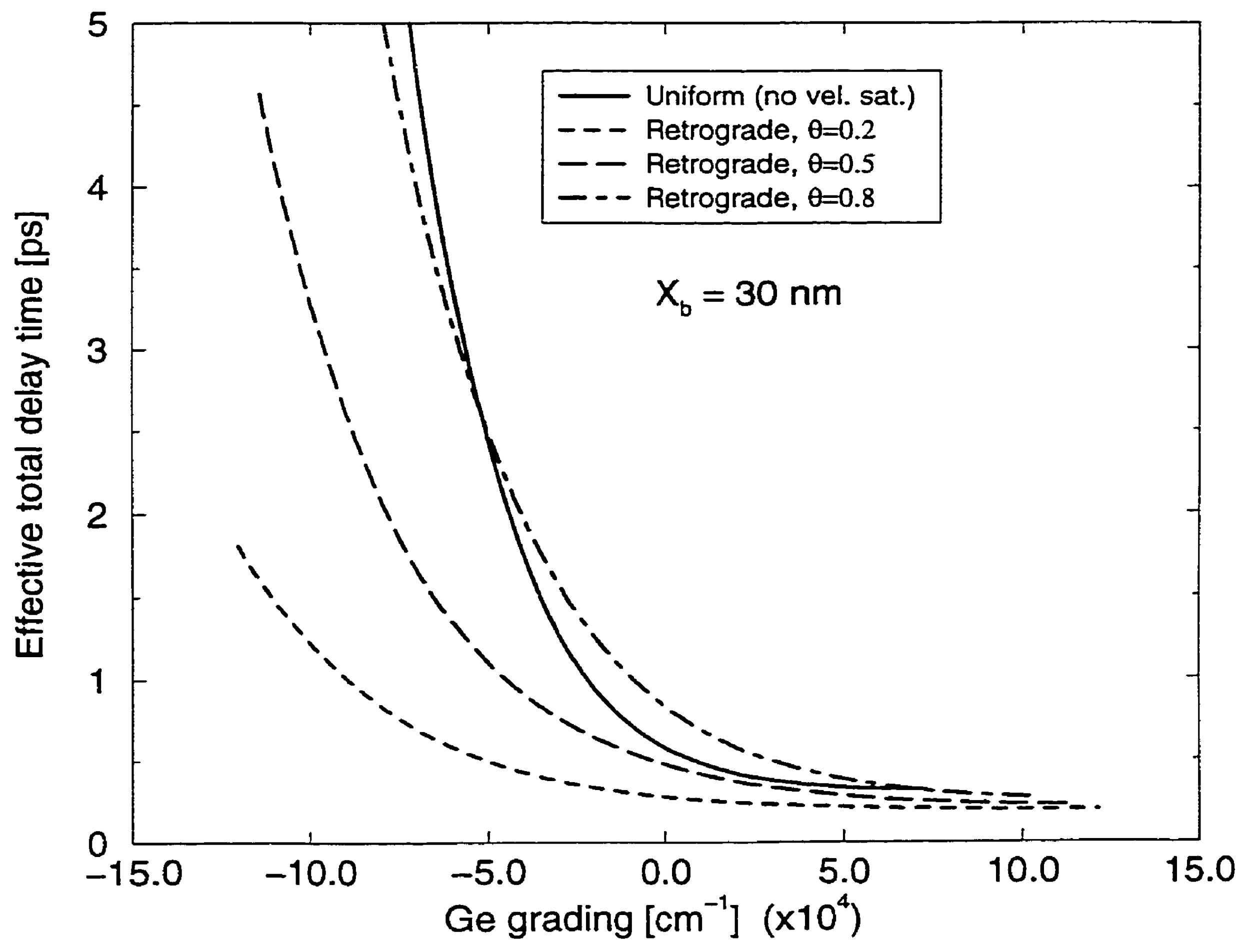

Figure 2.17: Effective total delay time (with non-quasi-static correction) as a function of Ge grading for transistor set #2 ($W_{ep} = 0.38$ μm, $X_{em} = 0.02$ μm, $X_b = 30$ nm).

Transistor	W_{ep} [μm]	X_{em} [μm]	X_b [nm]	θ	N_{bo} [cm^{-3}]	R_b [$k\Omega/\square$]
3a	0.25	0.15	30	-†	1.8×10^{19}	2.0
3b	0.25	0.15	30	0.2	5×10^{19}	2.16
3c	0.25	0.15	30	0.5	5×10^{19}	2.02
3d	0.25	0.15	30	0.8	5×10^{19}	1.89

Table 2.4: Profile details of transistor set #3 (†indicates a uniform base profile).

techniques, etc. will be required, which are beyond the scope of this study.

What if the emitter thickness is large so that τ_e becomes dominant ? To answer this question, another new set of transistors (see Table 2.4) is considered. In transistor set #3, although the *total* emitter widths remain unchanged from those in sets #1 and #2, the thickness of the mono-Si emitter (i.e. emitter junction depth) is larger than those in #1 and #2 by 0.13 μm. This will increase τ_e (see Fig. 2.18) when compared with sets #1 and #2 and it is more effective to increase τ_e through the mono-Si thickness than the poly-Si thickness (for example, compare Fig. 2.4 and Fig. 2.5). It should be noted that since the same base structures are used in set #2 and #3, the base transit times are identical for transistors in both sets. The comparison of Fig. 2.15 with Fig. 2.18 shows that for Ge gradings higher than -5 $\times$ 10^4/cm, τ_e can be even higher than τ_b. Fig. 2.19 and Fig. 2.20 plot the total delay time and the effective total delay time, respectively, for set #3.

In Fig. 2.19, the minimum τ_t point for the uniform base profile and the retrograde base profile ($\theta = 0.5$) is at a Ge grading of 0.9×10^4 and 4.4×10^4/cm, respectively. The factor γ is 2% for the uniform base profile and 11% for the retrograde base profile ($\theta = 0.5$). These numbers, calculated without the non-quasi-static correction, are much smaller than those observed in sets #1 and #2. More surprisingly, for the retrograde base profile with ($\theta = 0.2$), the minimum τ_t is actu-

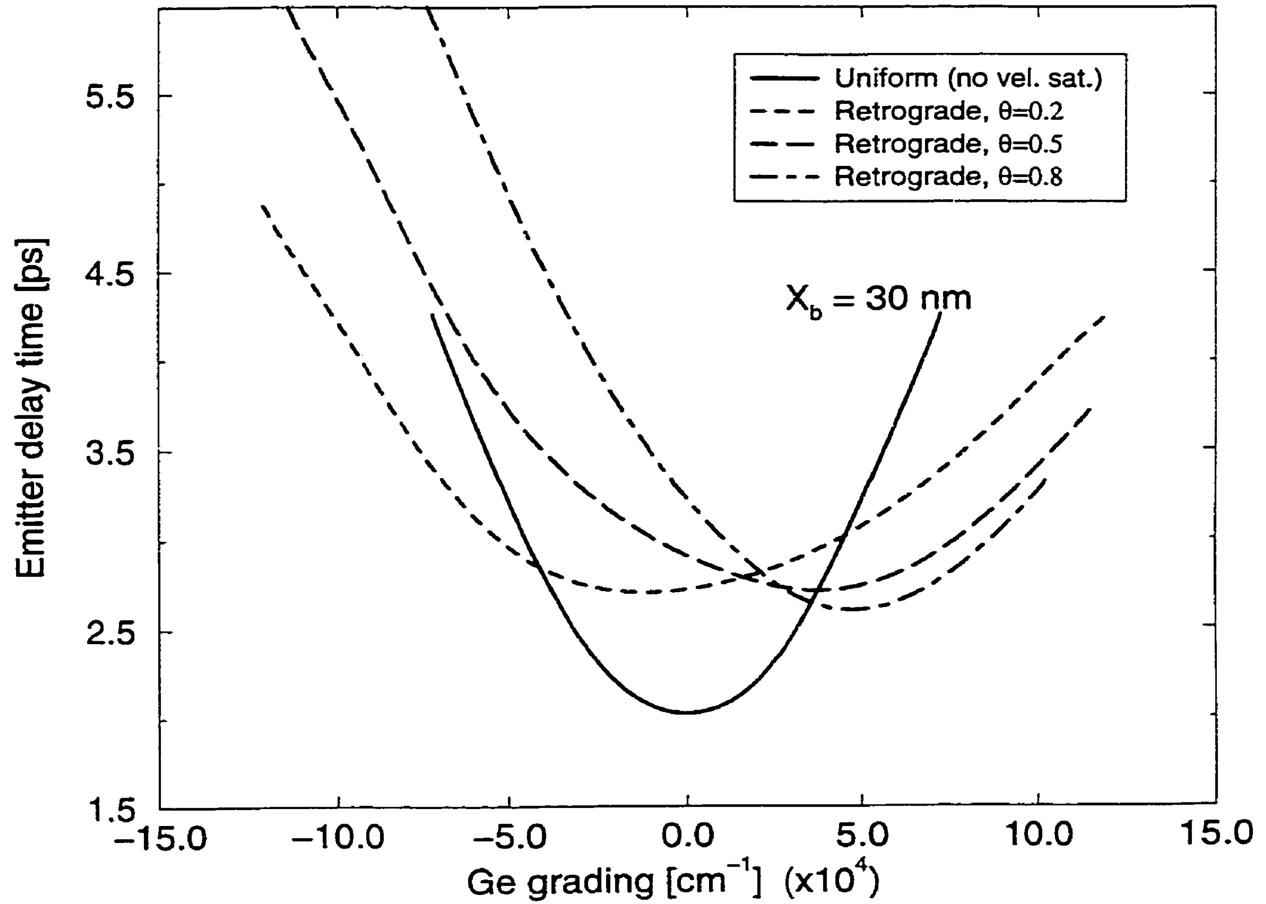

Figure 2.18: Emitter delay time vs. Ge grading for transistor set #3 (W_{ep} = 0.25 μm, X_{em} = 0.15 μm, X_b = 30 nm).

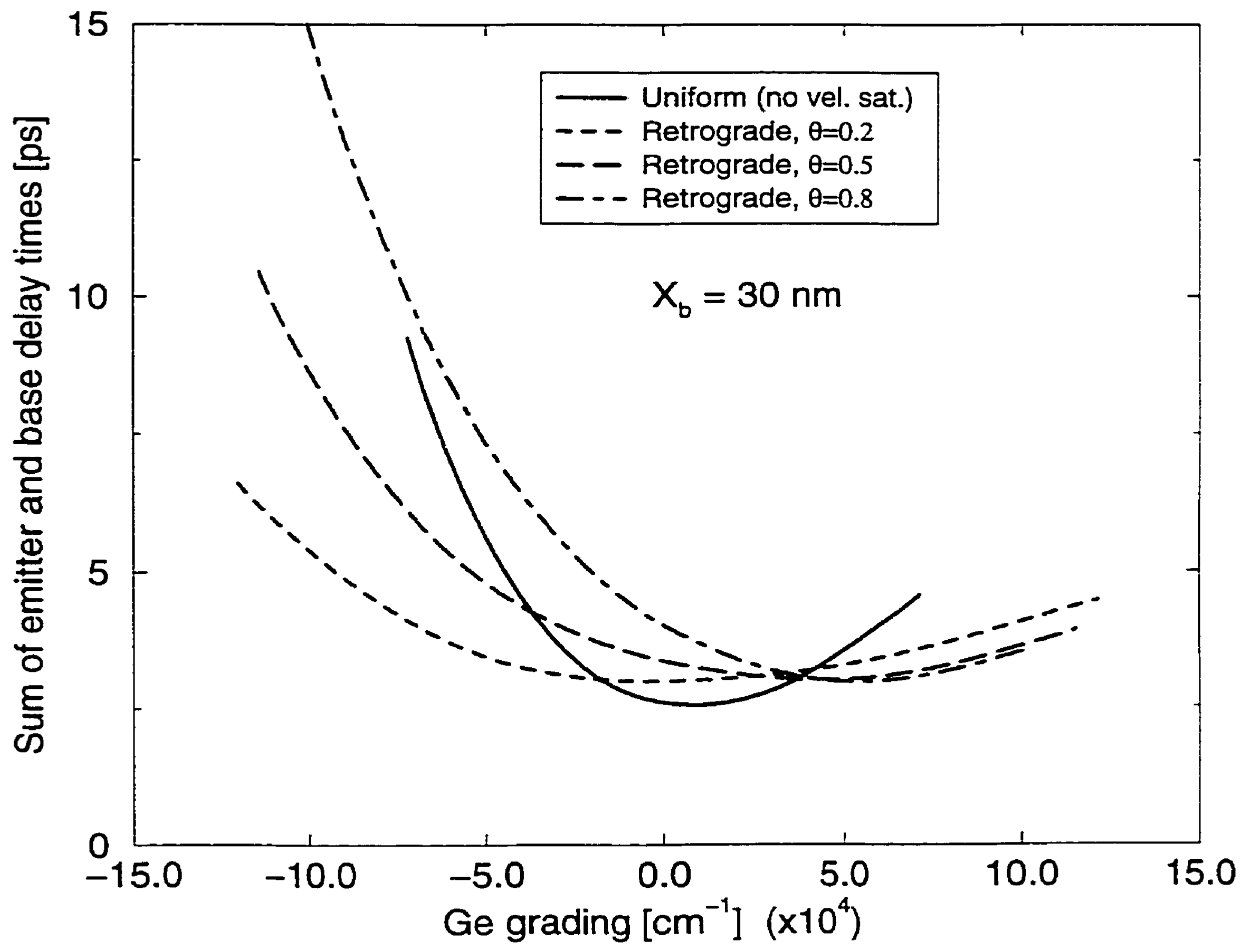

Figure 2.19: Total delay time (without non-quasi-static correction) as a function of Ge grading for transistor set #3 ($W_{ep} = 0.25$ μm, $X_{em} = 0.15$ μm, $X_b = 30$ nm).

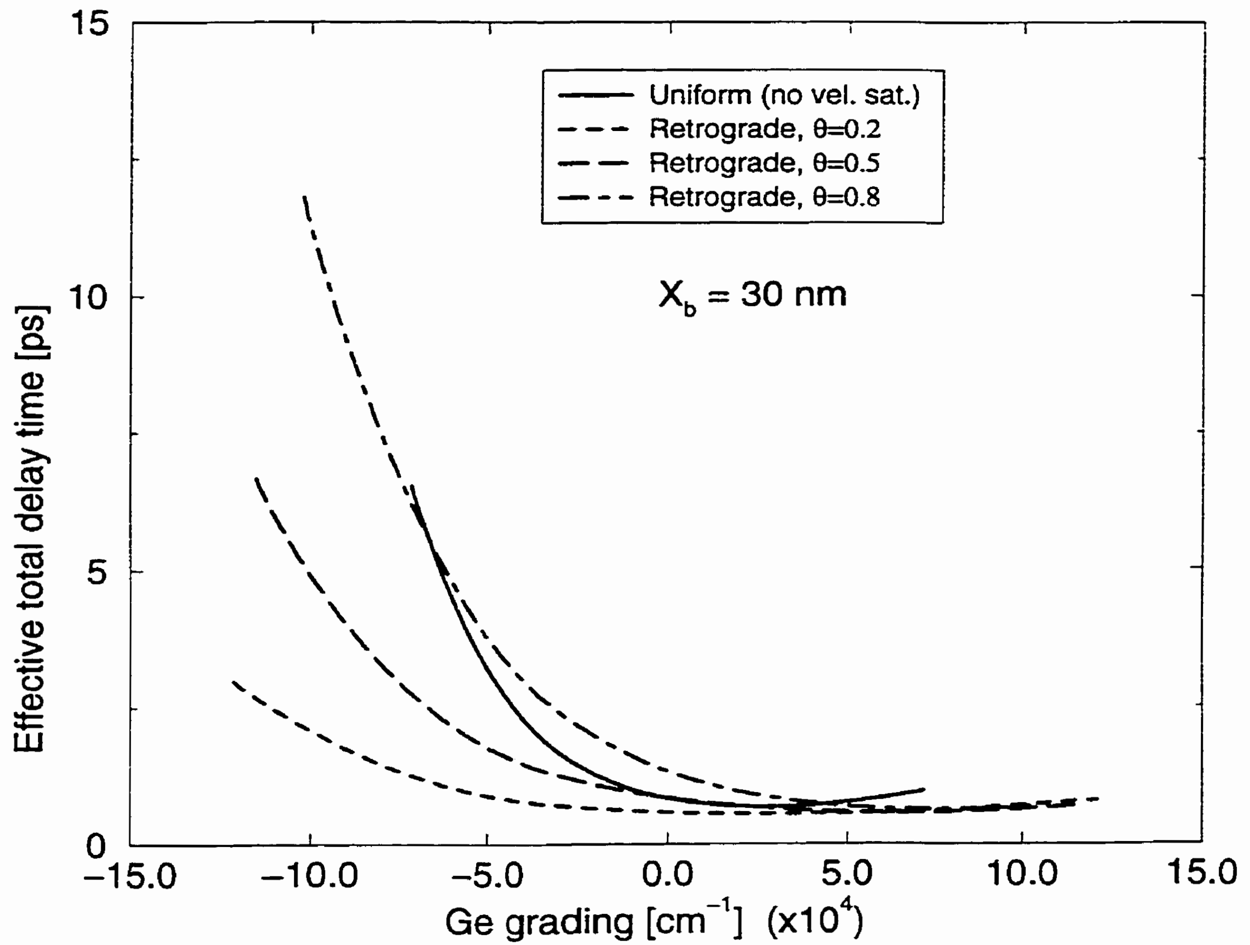

Figure 2.20: Effective total delay time (with non-quasi-static correction) as a function of Ge grading for transistor set #3 (W_{ep} = 0.25 μm, X_{em} = 0.15 μm, X_b = 30 nm).

ally achieved by a negative Ge grading. However, after non-quasi-static effects are taken into account (Fig. 2.20), the minimum τ_t points for all devices are pushed towards positive Ge gradings. The factor γ becomes 19% for the uniform base profile and 32% for the retrograde base profile ($\theta = 0.5$). These numbers are also smaller than their counterparts in sets #1 and #2, implying a smaller advantage for using a graded Ge profile over the box profile than in those two sets. Also, these numbers are around 17-21% higher than the γ's calculated without non-quasi-correction. In other words, the non-quasi-static effects in this case are more significant than in set #1 but less significant than in set #2. This suggests that the significance of non-quasi-static effects on the final estimate of τ_t depends, not only on the relative magnitudes of τ_e and τ_b (as shown by comparing sets #1 and #2), but also on the emitter charge-partitioning factor. The emitter charge partitioning factors α_e vary from 0.51 to 0.6 in set #2 and are around 0.93 in set #3. This explains why the non-quasi-static effects in set #3 are smaller by around 20% than in set #2 despite the larger τ_e achieved by transistors in set #3.

So far it has been observed that the graded Ge profile leverage factor γ tends to decrease when the magnitude of τ_e increases. To further explore this trend, τ_e is increased by using a higher base doping such that the current gain drops. In Table 2.5, transistors are designed with the peak base concentrations two times higher than those in the previous sets. This brings down the base sheet resistance closer to those of transistors with a base width of 100 nm (i.e., set #1). However, it should be noted that for transistor #4a, with a uniform base profile, the depletion layer of the emitter-base junction on the emitter side is larger than 20 nm. It does not pose problems in this particular case (as the emitter junction depth is 0.15 μm), but it will not work for transistors with both small emitter junction depths (e.g. sets #1 and 2) and uniformly doped bases. Although not shown, the base transit

Transistor	W_{ep} [μm]	X_{em} [μm]	X_b [nm]	θ	N_{bo} [cm^{-3}]	R_b [$k\Omega/\square$]
4a	0.25	0.15	30	-†	3.5×10^{19}	1.11
4b	0.25	0.15	30	0.2	10^{20}	1.19
4c	0.25	0.15	30	0.5	10^{20}	1.12
4d	0.25	0.15	30	0.8	10^{20}	1.06

Table 2.5: Profile details of transistor set #4 (†indicates a uniform base profile).

times are in the same order of magnitude as those in sets #2 and #3. The effective total delay time with non-quasi-static correction is plotted in Fig. 2.21. It shows higher total effective delay times than those in Fig. 2.20 because the emitter delay times are higher, as expected. The factor γ is 13 % for the uniform base profile and 28 % for the retrograde base profile (θ = 0.5). These values are the lowest compared with those in sets #1-3, implying that the advantage of the graded Ge profile over the box Ge profile is the smallest with set #4. This confirms the previously observed trend that the a graded Ge profile with positive Ge gradings have less advantage as τ_e continues to rise. Furthermore, for the retrograde base profile with θ = 0.2 (transistor #4b), the effective total delay time is minimum when a negative Ge grading of -1.73 $\times$ 10^4/cm is applied across the neutral base. However, the corresponding τ_t achieved at this Ge grading is practically the same as that at zero Ge gradient (i.e. a box Ge profile).

To make the study complete, transistors with very large emitter junction depth (X_{em} = 0.3μm, W_{ep} = 0.1μm) and transistors with base width of 50 nm (somewhere between the base widths of transistors studied above) are investigated along with the transistors in set #1–3. Fig. 2.22–2.25 summarize all the results by plotting the graded profile leverage factor (calculated with non-quasi-static correction) as a function of the metallurgical base width and the emitter depth for four different

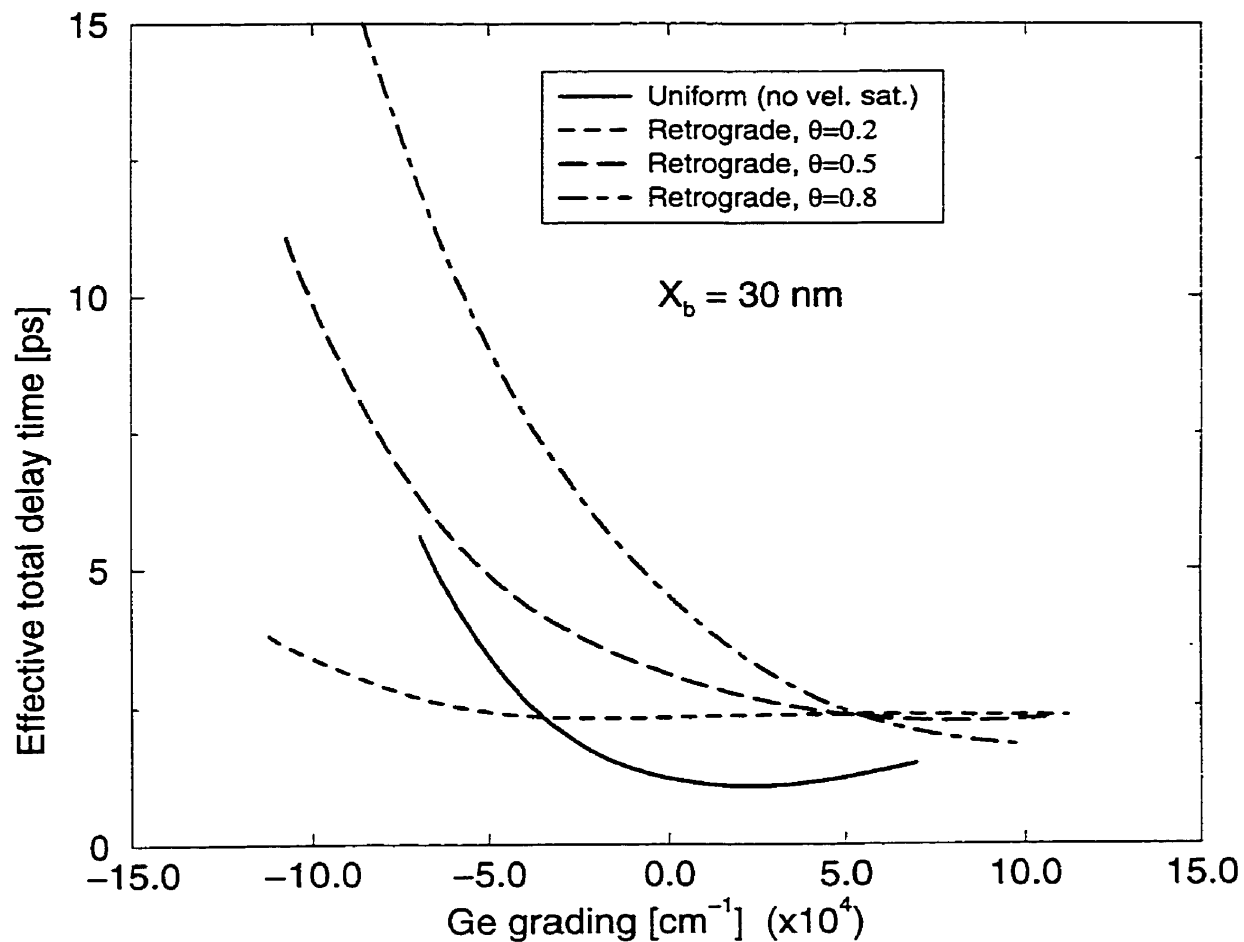

Figure 2.21: Effective total delay time (with non-quasi-static correction) as a function of Ge grading for transistor set #4 ($W_{ep} = 0.25$ μm, $X_{em} = 0.15$ μm, $X_b = 30$ nm, N_{bo} $=10^{20}$cm^{-3}).

shapes of base doping profile.

Fig. 2.22 is plotted for transistors with uniformly doped bases. It should be noted that a recombination velocity, at the collector-base junction, of 10^7 cm/s instead of an infinite recombination velocity (which is assumed earlier for studying the same uniformly doped base profiles) is used. Fig. 2.23 shows results for the retrograde base profiles with $\theta = 0.2$. In this particular case, additional data points obtained from transistors with extra base doping (which can indirectly increase τ_e as discussed earlier) are included in order to see more clearly the point where the leverage of the graded Ge profile over the box profile diminishes. As shown in the figure, for $X_b = 30$ nm, when the emitter junction depth (or the mono-Si emitter thickness) increases further (between 0.15 and 0.3 μm in this case), either a box Ge profile or a graded profile with a *negative* Ge grading is preferred over the graded Ge profile with a positive grading. In fact, the data point for $X_{em} = 0.15$ μm was taken from Fig. 2.21. What is new is that it shows a continuous drop in the leverage factor as X_{em} increases further to 0.3 μm. Furthermore, when comparing Fig. 2.23–2.25 together, one can see that the factor γ tends to decrease as the base doping profile is tilted towards the emitter-base junction (i.e. a smaller θ). This is true for transistors of all three sizes of base width $X_b = 30, 50, 100$ nm and emitter junction depths (0.02, 0.15, 0.3 μm).

2.5 Conclusions

By keeping the Ge dose constant as the Ge grading is varied to include the box Ge profile and graded Ge profiles with both positive and negative gradings, the emitter and base delay times, for SiGe HBTs of different base profile shapes, base concentrations, base widths, and emitter junction depths, have been studied with

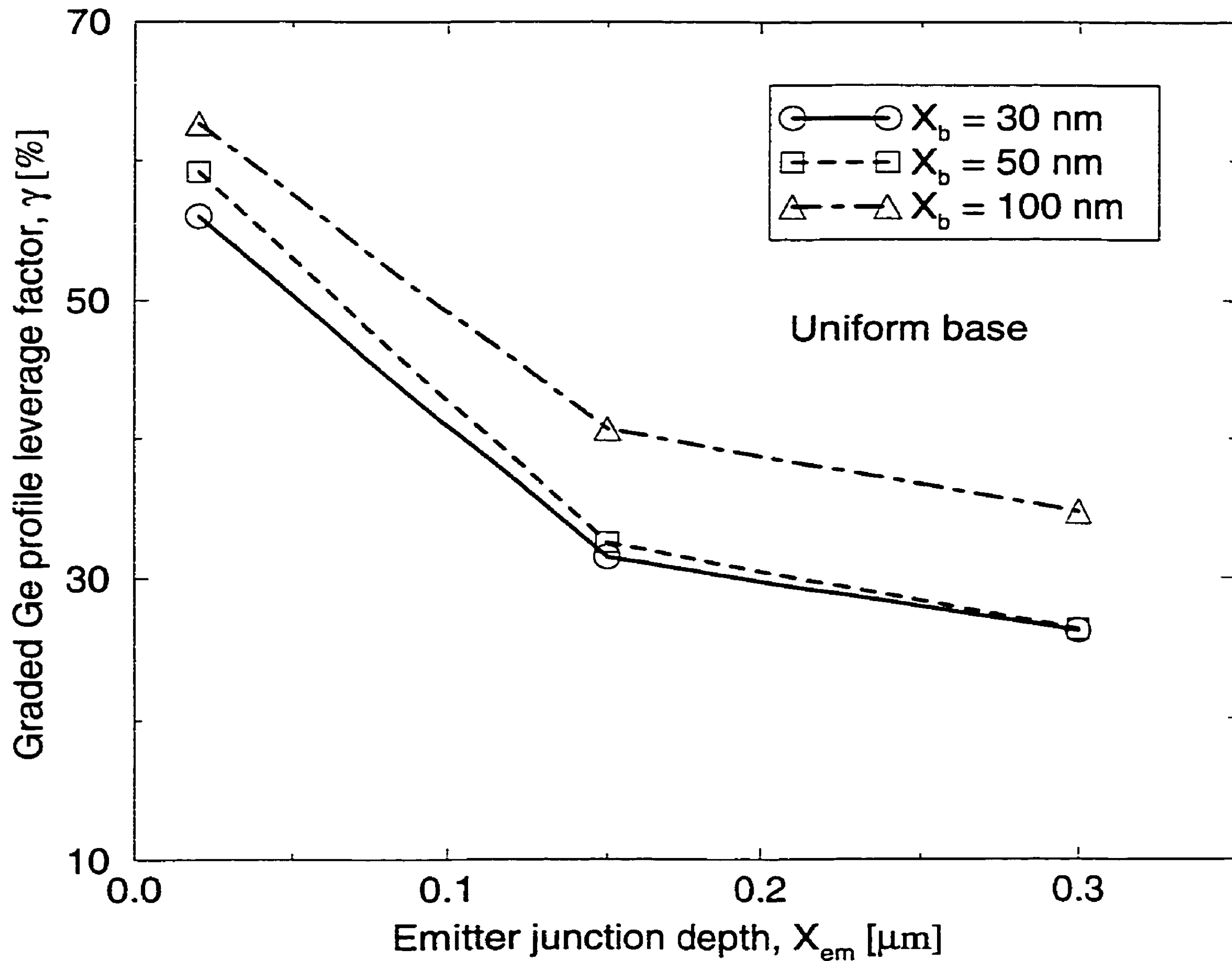

Figure 2.22: Graded Ge profile leverage factor vs. emitter junction depth for uniformly-doped base profiles (Circles: X_b = 30 nm, $N_{bo} = 1.8 \times 10^{19}\text{cm}^{-3}$; squares: X_b = 50 nm, $N_{bo} = 1.8 \times 10^{19}\text{cm}^{-3}$; triangles: X_b = 100 nm, $N_{bo} = 7.5 \times 10^{18}\text{cm}^{-3}$, total emitter width = 0.4 μm, $S = 10^7$cm/s).

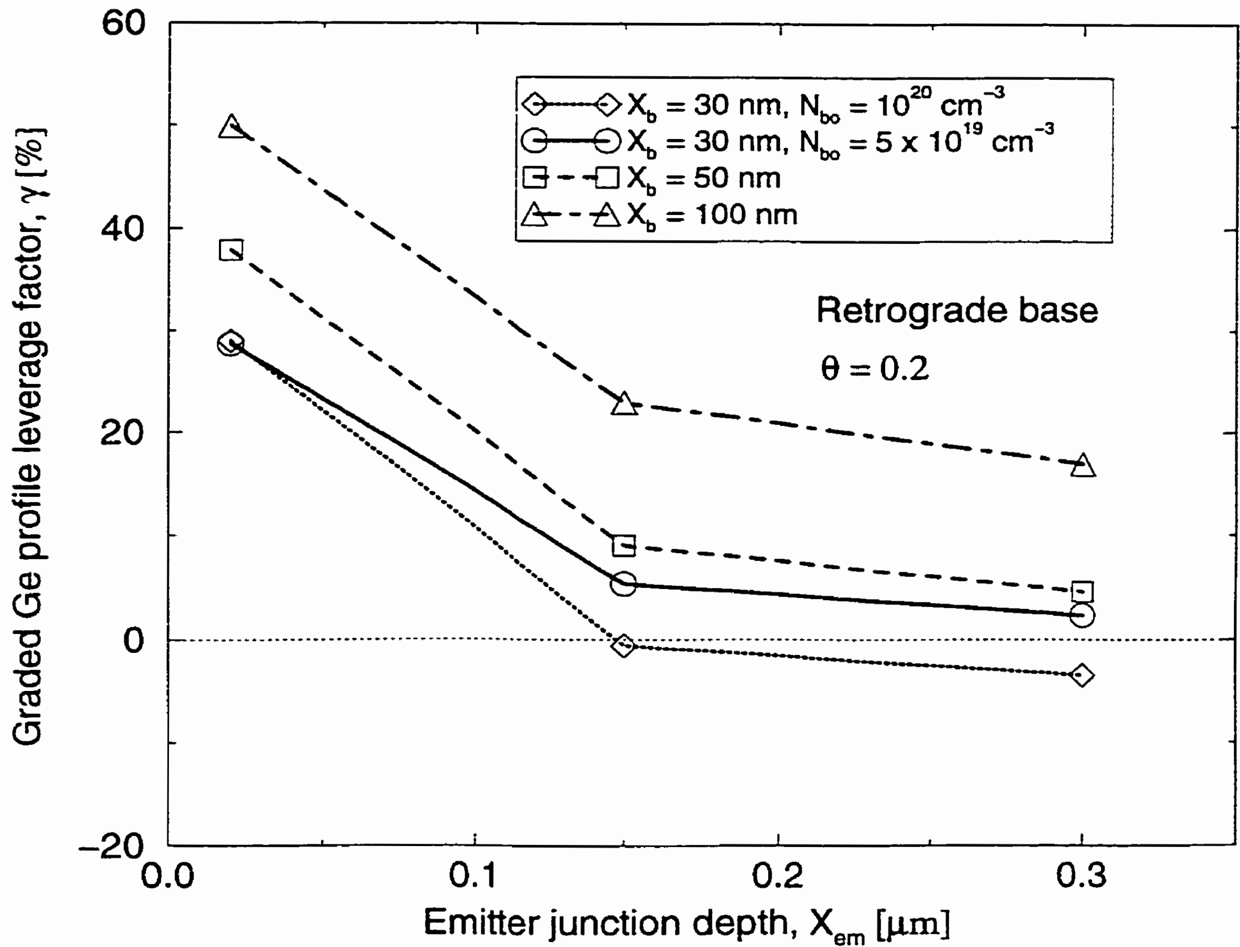

Figure 2.23: Graded Ge profile leverage factor vs. emitter junction depth for retrograde base profile with $\theta = 0.2$ (Diamonds: $X_b = 30$ nm, $N_{bo} = 10^{20}\text{cm}^{-3}$; circles: $X_b = 30$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; squares: $X_b = 50$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; triangles: $X_b = 100$ nm, $N_{bo} = 2 \times 10^{19}\text{cm}^{-3}$, total emitter width = 0.4 μm, $S = 10^7$cm/s).

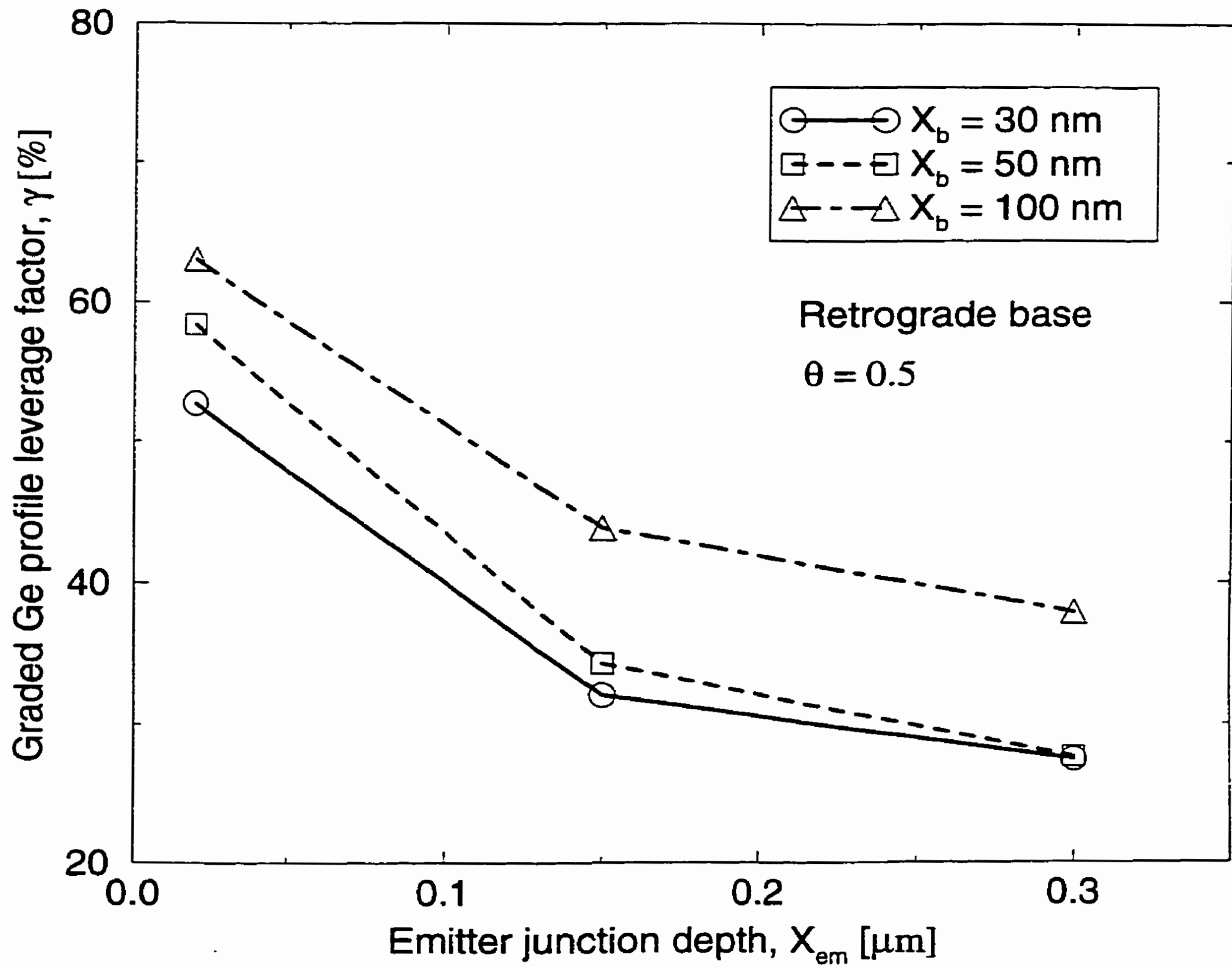

Figure 2.24: Graded Ge profile leverage factor vs. emitter junction depth for retrograde base profile with $\theta = 0.5$ (Circles: X_b = 30 nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; squares: X_b = 50 nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; triangles: X_b = 100 nm, $N_{bo} = 2 \times 10^{19}\text{cm}^{-3}$, total emitter width = 0.4 μm, $S = 10^7$cm/s).

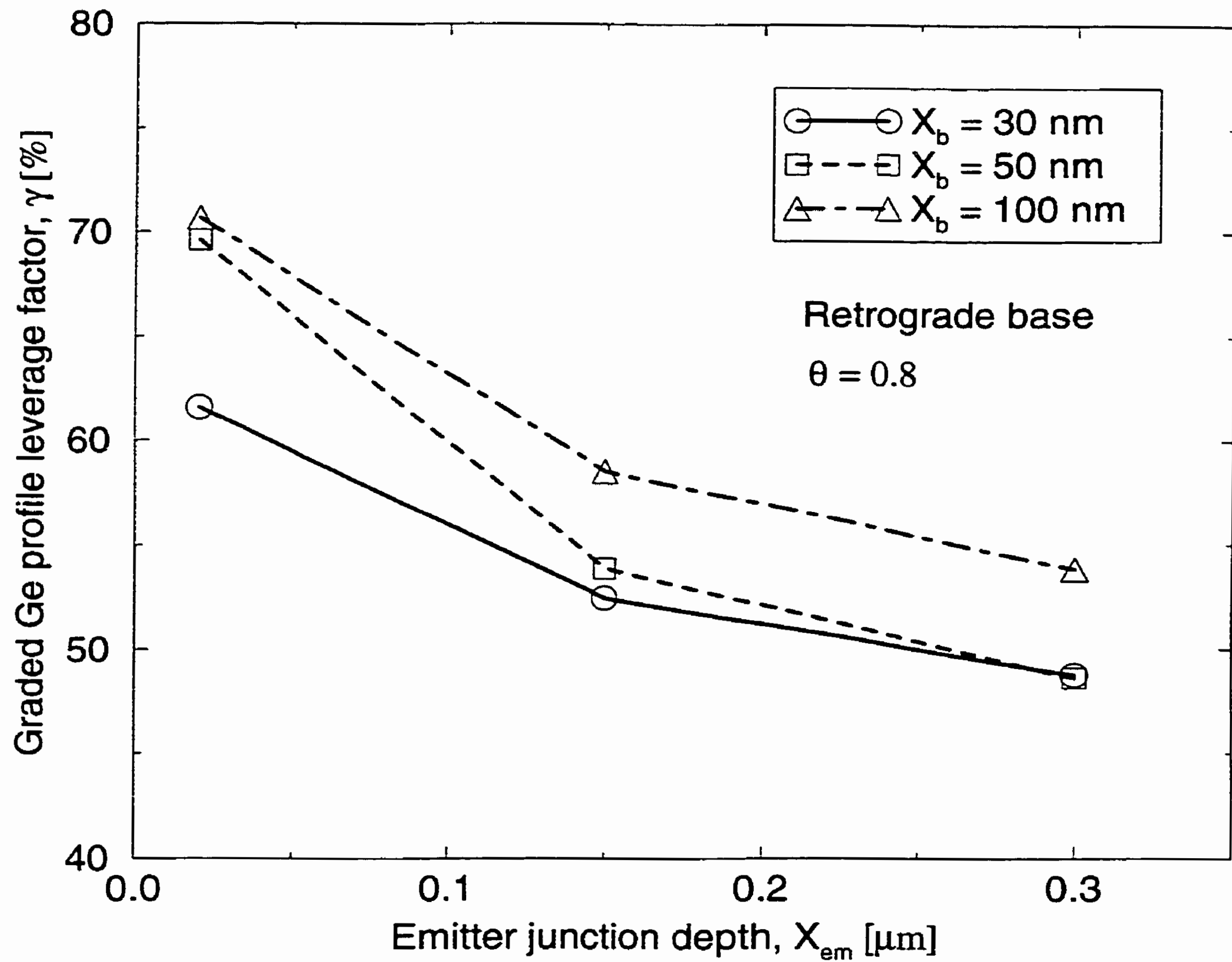

Figure 2.25: Graded Ge profile leverage factor vs. emitter junction depth for retrograde base profile with $\theta = 0.8$ (Circles: $X_b = 30$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; squares: $X_b = 50$ nm, $N_{bo} = 5 \times 10^{19}\text{cm}^{-3}$; triangles: $X_b = 100$ nm, $N_{bo} = 2 \times 10^{19}\text{cm}^{-3}$, total emitter width $= 0.4$ μm, $S = 10^7$cm/s).

non-quasi-static effects taken into account. Considering the total emitter and base delay times, the following trends are observed:

i) The advantage of the graded Ge profile with a positive Ge grading for minimizing τ_t varies considerably with the detailed shape of the base doping profile.

ii) The advantage of the graded Ge profile with a positive Ge grading for minimizing τ_t is reduced when the emitter junction depth increases.

In particular, the advantage of the graded Ge profile with a positive Ge grading is considerably reduced in the case of a retrograde base doping profile tilted toward the base-emitter junction (e.g. $\theta = 0.2$). The graded Ge leverage factors γ for retrograde doping profiles with $\theta = 0.2$ and with $\theta = 0.8$ differ by around 20-50%, dependent on the emitter and base dimensions.

However, for retrograde doping profiles ($\theta = 0.5$ or 0.8) and the uniform doping profile studied here, the Ge graded profile with a positive Ge grading is still preferred, as a reduction in the effective total delay time by at least 25%, with respect to a Ge box profile, can be achieved, for transistors with emitter depth smaller than 0.3 μm and metallurgical base width larger than 30 nm. In particular, for transistors with an emitter depth of 20 nm and metallurgical base width of 30 nm ($\theta = 0.5$), a reduction of around 50% is obtained, with respect to a box Ge profile, when a Ge profile graded from 0 near the emitter to 0.2 near the collector is used. Even when the emitter depth is increased to 150 nm, a reduction of 30% can be still obtained with a Ge profile graded from 0.06 near the emitter to 0.14 near the collector.

For transistors with a retrograde doping profile tilted toward the emitter-base junction (i.e., $\theta = 0.2$), a peak doping of 10^{20}cm^{-3}, and a base sheet resistance of

around 1 $k\Omega/\Box$, a box or a graded Ge profile with a small negative Ge grading is more effective than the graded Ge profile with a positive Ge grading if the metallurgical base width is 30 nm or smaller and the emitter junction depth is larger than 0.15 μm. In theory, this might be possible in future generations of SiGe HBTs as the base width continues to shrink such that the base concentration has to be increased to keep the the base sheet resistance at a reasonable level. In this case, the mono-Si emitter thickness (or the emitter junction depth) must be large enough to accommodate a wider depletion layer that extends into the emitter. However, for today's SiGe HBTs with typical dimensions, i.e., $X_b > 30$ nm and $X_{em} < 0.15$ μm, as far as τ_e and τ_b are concerned, the advantage of a positively graded Ge profile remains.

Chapter 3

Analytical expressions of base delay time for retrograde doping profiles

3.1 Introduction

Advances in epitaxial growth techniques in the past two decades have allowed integration of device-quality SiGe films as base layers into bipolar devices. This has led to the emergence of SiGe HBTs. One of the important figures of merit for SiGe HBTs is the transition frequency, f_T. As discussed in Chapter 1, the base transit (or delay) time is one of the significant delay times contributing to the f_T of a bipolar transistor. Thus, developing an accurate analytical model for the base transit time is useful. An analytical expression is preferred to a numerical approach since it reveals explicit dependencies of the base transit time on different parameters and provides valuable *engineering insights* for device design and optimization.

The use of the analytical expression allows us to evaluate the impact of the retrograde portion of a retrograde base doping profile on τ_b. The importance of the retrograde base profile lies in its ability to reduce the capacitance and the tunneling leakage current of the emitter-base junction of a SiGe HBT. As high base doping, i.e., small intrinsic base resistance, is often used to improve the maximum oscillation frequency of SiGe HBTs, the issues of high emitter-base capacitance and tunneling leakage current become more important. As shown in Chapter 2, the graded Ge profile is more effective than the box Ge profile in minimizing the emitter and base delay time contribution to the transition frequency. This result holds for transistors with base widths larger than 30 nm and emitter junction depths smaller than 150 nm. Therefore, we assume a graded Ge profile in our derivation of the τ_b analytical expression, even though the derivation in the case of a box Ge profile is more straightforward. A brief survey of pertinent works on deriving analytical expressions for τ_b follows.

Since the works of Moll and Ross [168] and Lindmayer and Wrigley [169] in the 1960's, different models and analytical expressions of τ_b have been derived for bipolar transistors. Kroemer [135] in 1985 generalized Moll and Ross' integral expression to account for the impact of the non-uniform energy bandgap caused by the composition variation in a uniformly-doped base of a HBT. Szeto and Reif [170] considered the effect of the non-uniform bandgap narrowing caused by heavy doping effects on τ_b in an exponentially-doped base. Suzuki [171] proposed a more accurate τ_b model by taking into account the electric field due to the base doping profile, heavy doping effects, and the electric-field and concentration dependencies of the carrier diffusivity. Later, Suzuki and Nakayama [134] also considered the effect of velocity saturation at the collector-base junction. Then, Lu and Kuo [172] derived an analytical τ_b expression for an exponentially-doped retrograde base, taking

into account the effects of heavy doping and doping-dependent diffusivity. Jahan and Anwar [173] presented an analytical τ_b expression for an exponentially-doped base, which considers these two effects together with the velocity saturation at the collector-base junction. However, it was later pointed out by Rinaldi [174] that Jahan and Anwar's results were inaccurate because the effective doping concentration, instead of just the doping concentration, was used to calculate the low-field electron mobility.

For SiGe HBTs, Gao and Morkoc [175] presented a set of analytical expressions for an exponentially-doped *retrograde* base profile (see Fig. 3.1), which reduces the junction capacitance and the tunneling current at the emitter-base junction. However, their expressions were incorrect since Kroemer's integral relation for τ_b (equation (12) in [135]) was improperly applied to the three subregions of the neutral base. They also neglected important effects like electric-field-dependent diffusivity, heavy doping effects, etc. Chen *et al.* [85] included heavy doping effects alongside the effect of the non-uniform bandgap narrowing due to the Ge gradient in the base, and presented an analytical τ_b expression for an exponentially-doped base. In addition to these effects, Lu *et al.* [114] considered the effect of doping-dependent diffusivity in deriving a closed-form analytical τ_b expression for an exponentially-doped retrograde base. Recently, Rinaldi [174] reported a closed-form analytical expression of τ_b for an exponentially-doped base with all effects, including the electric-field-dependent diffusivity and the velocity saturation at the collection-base junction, taken into account. However, thus far, no correct analytical τ_b expression, which considers all the effects mentioned above, for each subregion of an exponentially-doped *retrograde* base has been reported.

This chapter proposes a regional model and derives a set of closed-form analytical τ_b expressions for the subregions in the same base profile considered by Gao

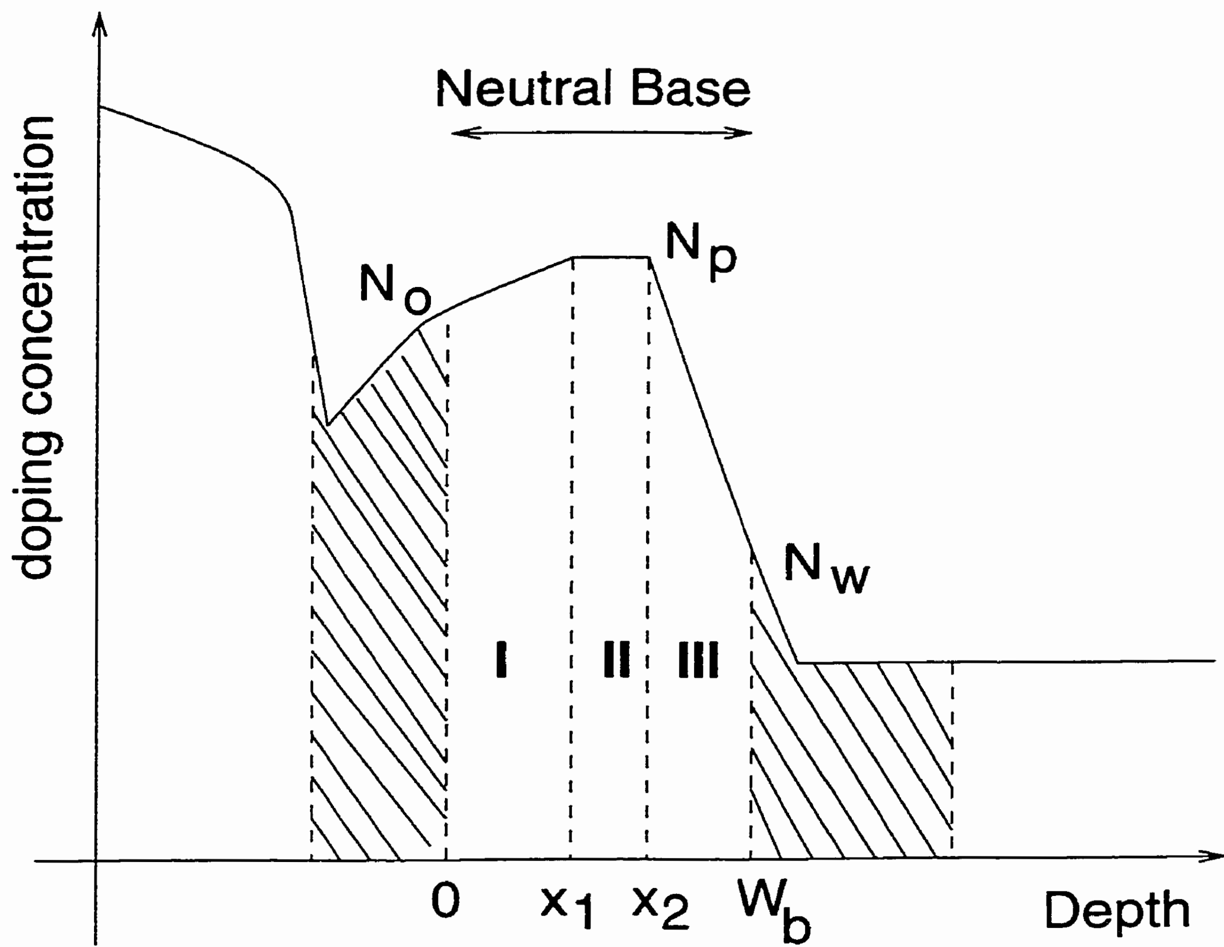

Figure 3.1: Doping profile for an exponentially-doped retrograde base.

and Morkoc [175], i.e., an exponentially-doped retrograde SiGe base (see Fig. 3.1). Our derivation considers the following effects: (1) the built-in electric field caused by the non-uniform base doping profile, (2) the electric field due to the non-uniform apparent bandgap narrowing associated with heavy doping effects in the base, (3) the electric field induced by the non-uniform bandgap narrowing due to the Ge concentration gradient in the base, (4) the doping dependency of the electron diffusivity, (5) the electric-field dependency of the electron diffusivity, and (6) the velocity saturation at the collector-base junction. To simplify the derivation, we only consider low-level injection and neglect the neutral base recombination[1] and the effects of ballistic transport and velocity overshoot. We assume that material parameters (except the bandgap energy) for Si and those for SiGe are identical and that the transistors operate at 300 K. The work in this chapter has been published in [89].

Section 3.2 derives the τ_b expressions for a regional model. Results calculated from the regional model are presented in Section 3.3 for *assessing the relative importance* of each of these effects and also the impact of the retrograde portion of the base doping profile on τ_b. Section 3.4 verifies the results by numerical integration and compares them with published results in the literature. Conclusions are drawn in Section 3.5

[1] Please see Chapter 2 for the justification of this assumption.

3.2 Regional model

3.2.1 Base profile approximation

Using the same definitions as in [175], we approximate and divide the base profile of a n-p-n SiGe HBT into three subregions (Fig. 3.1) as follows:

$$N_B(x) = \begin{cases} N_o\ e^{b_1 x} & 0 \leq x \leq x_1 \text{ (Region I)}, \\ N_p & x_1 \leq x \leq x_2 \text{ (Region II)}, \\ N_p\ e^{-b_2(x-x_2)} & x_2 \leq x \leq W_b \text{ (Region III)} \end{cases} \tag{3.1}$$

$$b_1 = ln\left(\frac{N_p}{N_o}\right)/x_1, \qquad b_2 = ln\left(\frac{N_p}{N_w}\right)/(W_b - x_2)$$

where N_o, N_p, N_w, x_1, x_2, and W_b are indicated in Fig. 3.1.

3.2.2 Apparent bandgap narrowing due to heavy doping effects

As in [114, 170, 172, 173], we approximate Klaassen *et al.*'s [132] expression of apparent bandgap narrowing due to heavy doping effects as:

$$\Delta E_{g,HD}(x) \sim 2V_{go}\ ln\left[\frac{N_B(x)}{N_2}\right] \qquad \text{for } N_B(x) > N_2 \tag{3.2}$$

where $V_{go} = 6.92$ meV and $N_2 = 1.3 \times 10^{17}\text{cm}^{-3}$. Thus, the effective intrinsic carrier concentration becomes:

$$\begin{aligned} n_{ie}^2(x) &= n_{io}^2(x) e^{\Delta E_{g,HD}(x)/kT} \\ &= n_{io}^2(x)\left(\frac{N_B(x)}{N_2}\right)^{2V_{go}/kT} \end{aligned} \tag{3.3}$$

where n_{io} is the intrinsic carrier concentration in Si.

3.2.3 Bandgap narrowing due to Ge concentration gradient

According to People and Bean [70], the bandgap narrowing $\Delta E_{g,Ge}$ in a coherently strained $Si_{1-y}Ge_y$ on <100> Si is related to the Ge fraction, y, as $\Delta E_{g,Ge} = 0.74\ y$ (eV) for $y \leq 0.3$. For a linearly-graded Ge profile in the base without Ge at $x = 0$, $\Delta E_{g,Ge}(x) = 0.74\ x\ y_{Ge}/W_b$, where y_{Ge} is the Ge fraction at $x = W_b$, indicating the Ge grading across the base. Thus the effective intrinsic carrier concentration given in (3.3) needs to be modified as:

$$n_{ie}^2(x) = n_{io}^2(x) \left[\frac{N_B(x)}{N_2}\right]^{2V_{go}/kT} e^{ax} \tag{3.4}$$

where $a = 0.74\ y_{Ge}/(W_b kT)$. Hence, the minority electron concentration at thermal equilibrium becomes:

$$n_o(x) = \frac{n_{ie}^2(x)}{N_B(x)} \tag{3.5}$$

Using (3.1), (3.4), and (3.5), we can express the electron concentration at thermal equilibrium in each subregion of the base as:

$$n_o(x) = \begin{cases} n_{o1}(x) = C_1\ e^{(a-b_1')x} & 0 \leq x \leq x_1 \text{ (Region I)}, \\ n_{o2}(x) = C_2\ e^{ax} & x_1 \leq x \leq x_2 \text{ (Region II)}, \\ n_{o3}(x) = C_2\ e^{(a+b_2')x - b_2' x_2} & x_2 \leq x \leq W_b \text{ (Region III)} \end{cases} \tag{3.6}$$

$$C_1 = \frac{n_{io}^2}{N_o}\left(\frac{N_o}{N_2}\right)^{2V_{go}/kT}, \qquad C_2 = \frac{n_{io}^2}{N_p}\left(\frac{N_p}{N_2}\right)^{2V_{go}/kT}$$

$$b_1' = b_1(1 - 2V_{go}/kT), \qquad b_2' = b_2(1 - 2V_{go}/kT)$$

3.2.4 Doping-dependent diffusivity

Klaassen [121, 125] recently reported a more accurate low-field electron mobility model which results in a single function of the local donor and acceptor concentrations and takes into account different scattering mechanisms, and screening and clustering effects. To facilitate our derivation, we approximate Klaassen's results as follows [174]:

$$\mu_n^l(x) = \frac{\mu_{no}}{1 + (N_B(x)/N_m)^m} \tag{3.7}$$

where μ_n^l is the low-field mobility; μ_{no}, N_m, m are fitting parameters. As shown in Fig. 3.2, a good fit is obtained for $N_B(x) = 10^{17} \rightarrow 10^{19}\text{cm}^{-3}$ by using the least square method with $\mu_{no} = 4\times10^6$ cm^2 V^{-1}s^{-1}, N_m =$1.46\times10^4\text{cm}^{-3}$, and $m = 0.295$ (Note. A local donor concentration of 10^{16}cm^{-3} is assumed in generating Klaassen's curve). Using Einstein's relation, the low-field electron diffusivity can be obtained from (3.7):

$$D_n^l(x) = \frac{D_{no}}{1 + (N_B(x)/N_m)^m} \tag{3.8}$$

where $D_{no} = \mu_{no}V_T$ and V_T is the thermal voltage kT/q.

3.2.5 Velocity saturation at collector-base junction

From conventional drift-diffusion transport theory, J_n can be expressed as:

$$J_n = q\mu_n(x)n(x)\mathcal{E}(x) + qD_n(x)\frac{dn(x)}{dx} \tag{3.9}$$

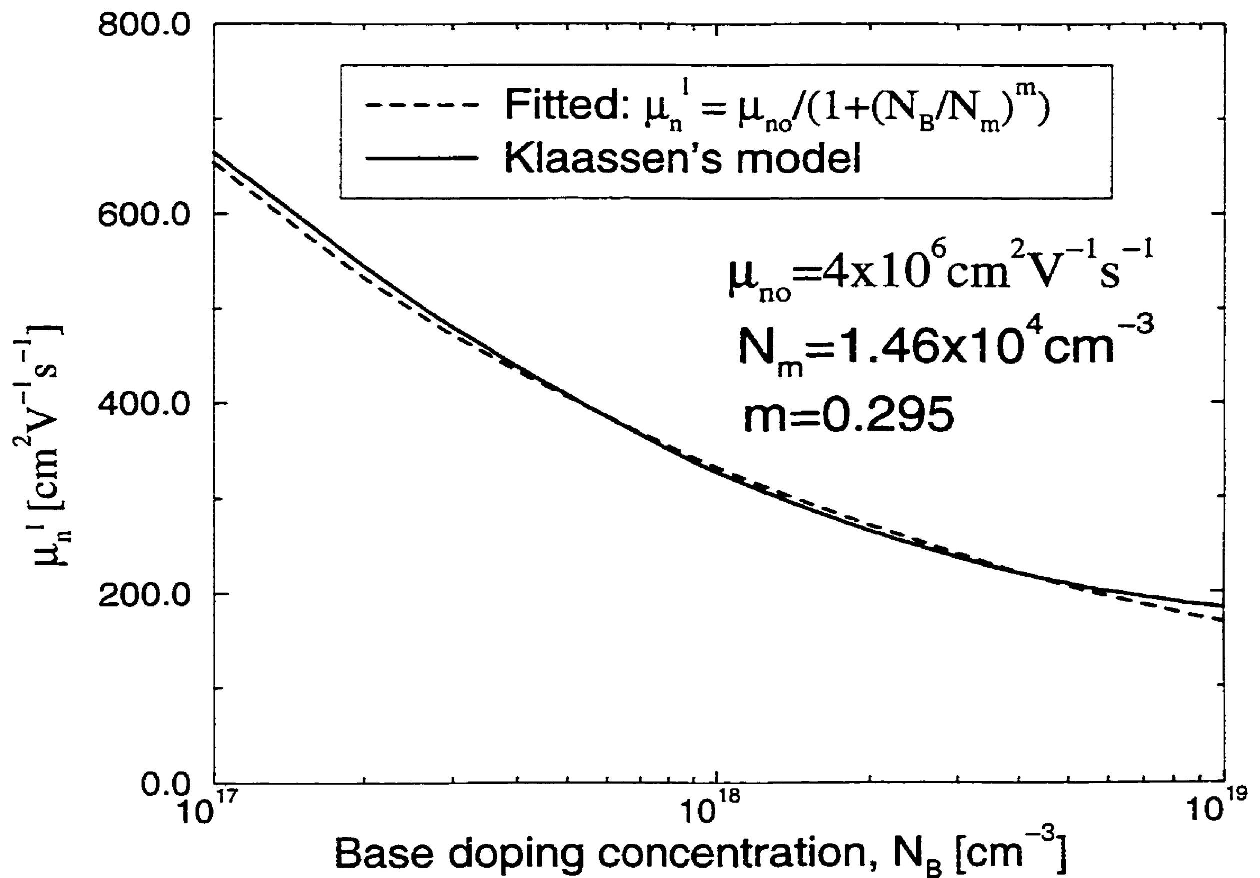

Figure 3.2: Minority electron low-field mobility as a function of base concentration. Solid line: Klaassen's results [121, 125] with local $N_d = 10^{16}\text{cm}^{-3}$. Dashed line: fitted by (3.7).

where μ_n and D_n are the electron mobility and diffusion coefficient, respectively. The electric field in the neutral base $\mathcal{E}$ is the sum of the doping-induced electric field ($\mathcal{E}_d$) and the electric field caused by the non-uniform bandgap narrowing ($\mathcal{E}_g$). Neglecting the hole current [135], we can express the doping-induced electric field as:

$$\mathcal{E}_d = V_T \frac{d[lnN_B(x)]}{dx} \tag{3.10}$$

where V_T is the thermal voltage and N_b is the acceptor concentration in the neutral base. Considering the non-uniform bandgap narrowing in the base, the induced electric field $\mathcal{E}_g$ can be written as:

$$\mathcal{E}_g = -V_T \frac{dln(n_{ie}^2)}{dx} \tag{3.11}$$

where n_{ie} is the effective intrinsic carrier concentration and can be written as [131]:

$$n_{ie}^2 = n_{io}^2 exp\left[\frac{\Delta E_g(x)}{kT}\right] \tag{3.12}$$

where n_{io} as the intrinsic carrier concentration and $\Delta E_g(x)$ the apparent bandgap narrowing. From (3.10) and (3.11), the total induced electric field in the neutral base can be written as:

$$\mathcal{E} = -V_T \frac{d}{dx}[ln(n_o(x))] \tag{3.13}$$

where $n_o = \frac{n_{ie}^2}{N_B}$ is the equilibrium electron concentration obtained in (3.6). Substituting (3.13) into (3.9), we obtain:

$$J_n = qD_n n_o \frac{d}{dx}\left(\frac{n}{n_o}\right) \tag{3.14}$$

Integrating (3.14) from x to W_b with the boundary condition at $z = W_b$ (i.e., $J_n = -qSn_3(W_b)$, where S is the interface velocity or B-C junction velocity [176] and n_3 is the electron concentration in subregion III), we obtain:

$$\begin{aligned} \frac{J_n}{q}\int_x^{W_b} \frac{dz}{D_n(z)n_o(z)} &= \frac{n_3(W_b)}{n_{o3}(W_b)} - \frac{n(x)}{n_o(x)} \\ &= -\frac{J_n}{qSn_{o3}(W_b)} - \frac{n(x)}{n_o(x)} \end{aligned} \tag{3.15}$$

3.2.6 Regional base transit times

The base transit times in the three different subregions are defined as follows:

$$\tau_{b1} = q\int_0^{x_1} \frac{n_1(x)}{|J_n|}dx \tag{3.16}$$

$$\tau_{b2} = q\int_{x_1}^{x_2} \frac{n_2(x)}{|J_n|}dx \tag{3.17}$$

$$\tau_{b3} = q\int_{x_2}^{W_b} \frac{n_3(x)}{|J_n|}dx \tag{3.18}$$

where $n_1(x), n_2(x)$, and $n_3(x)$ are the electron concentrations in subregions I, II, and III, respectively. It is noteworthy that J_n is constant throughout all three

subregions. From (3.15), we can express the ratio $n(x)/J_n$ as:

$$\frac{n(x)}{J_n} = -\frac{n_o(x)}{q}\left[\frac{1}{Sn_{o3}(W_b)} + \int_x^{W_b} \frac{dz}{D_n(z)n_o(z)}\right] \tag{3.19}$$

Please note that (3.19) is similar to the expressions presented by Jahan and Anwar [173] and by Suzuki and Nakayama [134] if n_o is replaced by $n_{ie}^2/N_B(x)$. However, it differs from Suzuki and Nakayama [134]'s expression where S is assumed to be the thermal saturation velocity, v_s.

Substituting (3.19) into (3.16)–(3.18) gives:

$$\tau_{b1} = \int_0^{x_1} n_{o1}(x) \int_x^{W_b} \frac{dz}{D_n(z)n_o(z)} dx + \frac{\int_0^{x_1} n_{o1}(x)dx}{Sn_{o3}(W_b)} \tag{3.20}$$

$$\tau_{b2} = \int_{x_1}^{x_2} n_{o2}(x) \int_x^{W_b} \frac{dz}{D_n(z)n_o(z)} dx + \frac{\int_{x_1}^{x_2} n_{o2}(x)dx}{Sn_{o3}(W_b)} \tag{3.21}$$

$$\tau_{b3} = \int_{x_2}^{W_b} n_{o3}(x) \int_x^{W_b} \frac{dz}{D_n(z)n_o(z)} dx + \frac{\int_{x_2}^{W_b} n_{o3}(x)dx}{Sn_{o3}(W_b)} \tag{3.22}$$

It should be noted that the upper limits of the inner integrals in the first terms of (3.20)–(3.22) should be the same, i.e., W_b, instead of x_1, x_2, and W_b, as incorrectly assumed by Gao and Morkoc [175].

3.2.7 Electric-field-dependent diffusivity

Since the electric field in the non-uniform base influences the electron diffusivity, D_n in (3.20)–(3.22) is different from the low-field diffusivity obtained from (3.8). Using the Caughey-Thomas mobility formula [126] and Einstein relation, we relate

the electron diffusivity to the low-field diffusivity by the electric field across the base as follows:

$$D_n(x) = \frac{D_n^l(x)}{[1 + (|E(x)|\ D_n^l(x)/(V_T v_s))^{\beta}]^{1/\beta}} \tag{3.23}$$

where $\beta = 1$ [177, 178] or 2 [126]. The electric field, E, is either negative or zero in subregions II and III (i.e., $|E| = -E = V_T \frac{d}{dx}[ln(n_{o(2,3)}(x))]$, from (3.13)). While in Region I, $E = -V_T \frac{d}{dx}[ln(n_{o1}(x))] = -V_T(a - b_1')$. Thus, the direction of the electric field depends on the strength of the electric field associated with the doping profile and the heavy doping effects, relative to the electric field due to the Ge grading:

$$|E|_{\text{in Region I}} = \begin{cases} 0 & \text{if } a = b_1', \\ -E = V_T \frac{d}{dx}[ln(n_{o1}(x))] & \text{if } a > b_1', \\ E = -V_T \frac{d}{dx}[ln(n_{o1}(x))] & \text{if } a < b_1' \end{cases} \tag{3.24}$$

Therefore, for $\beta = 1$, by substituting (3.23) and (3.24) into (3.20)–(3.22), we can show that (see Appendix B):

$$\begin{aligned} \tau_{b1} = & \int_0^{x_1} n_{o1}(x) \int_x^{W_b} \frac{dz}{D_n^l(z) n_o(z)} dx + \frac{x_1 |a - b_1'|}{v_s(a - b_1')} + \\ & \left\{ \frac{1}{S n_{o3}(W_b)} - \frac{1}{v_s} \left[\frac{|a - b_1'|/(a - b_1') - 1}{n_{o1}(x_1)} + \frac{1}{n_{o3}(W_b)} \right] \right\} \int_0^{x_1} n_{o1}(x) dx \end{aligned} \tag{3.25}$$

$$\tau_{b2} = \int_{x_1}^{x_2} n_{o2}(x) \int_x^{W_b} \frac{dz}{D_n^l(z) n_o(z)} dx + \frac{x_2 - x_1}{v_s} + \left(\frac{1}{S} - \frac{1}{v_s} \right) \frac{\int_{x_1}^{x_2} n_{o2}(x) dx}{n_{o3}(W_b)} \tag{3.26}$$

$$\tau_{b3} = \int_{x2}^{W_b} n_{o3}(x) \int_{x}^{W_b} \frac{dz}{D_n^l(z) n_o(z)} dx + \frac{W_b - x_2}{v_s} + \left(\frac{1}{S} - \frac{1}{v_s}\right) \frac{\int_{x_2}^{W_b} n_{o3}(x) dx}{n_{o3}(W_b)} \tag{3.27}$$

3.2.8 Analytical expressions for base delay time

In order to obtain closed-form analytical expressions, the inner integrals in the first terms of (3.25)–(3.27) need to be expanded. So, (3.25)–(3.27) are re-written as:

$$\begin{aligned} \tau_{b1} = & \int_0^{x_1} n_{o1}(x) \left[\int_x^{x_1} \frac{dz}{D_{n1}^l(z) n_{o1}(z)} + \int_{x_1}^{x_2} \frac{dz}{D_{n2}^l(z) n_{o2}(z)} + \int_{x_2}^{W_b} \frac{dz}{D_{n3}^l(z) n_{o3}(z)} \right] dx \\ & + \frac{x_1 |a - b_1'|}{v_s (a - b_1')} + \left\{ \frac{1}{S n_{o3}(W_b)} - \frac{1}{v_s} \left[\frac{|a - b_1'|/(a - b_1') - 1}{n_{o1}(x_1)} + \frac{1}{n_{o3}(W_b)} \right] \right\} \\ \times & \int_0^{x_1} n_{o1}(x) dx \end{aligned} \tag{3.28}$$

$$\begin{aligned} \tau_{b2} = & \int_{x_1}^{x_2} n_{o2}(x) \left[\int_x^{x_2} \frac{dz}{D_{n2}^l(z) n_{o2}(z)} + \int_{x_2}^{W_b} \frac{dz}{D_{n3}^l(z) n_{o3}(z)} \right] dx \\ & + \frac{x_2 - x_1}{v_s} + \left(\frac{1}{S} - \frac{1}{v_s}\right) \frac{\int_{x_1}^{x_2} n_{o2}(x) dx}{n_{o3}(W_b)} \end{aligned} \tag{3.29}$$

$$\tau_{b3} = \int_{x2}^{W_b} n_{o3}(x) \int_{x}^{W_b} \frac{dz}{D_{n3}^l(z) n_{o3}(z)} dx + \frac{W_b - x_2}{v_s} + \left(\frac{1}{S} - \frac{1}{v_s}\right) \frac{\int_{x_2}^{W_b} n_{o3}(x) dx}{n_{o3}(W_b)} \tag{3.30}$$

Taking into consideration all the effects mentioned in Sections 3.2.2-3.2.5 and 3.2.7, we can obtain the analytical expression for τ_b in each subregion by substituting (3.1),

(3.6), (3.8) into (3.28)–(3.30):

$$\tau_{b1} = \frac{1}{D_{no}}\left\{\frac{e^{(b_1'-a)x_1}-1}{(a-b_1')^2} + \frac{x_1}{a-b_1'} + \left(\frac{N_o}{N_m}\right)^m\left[\frac{e^{b_1 m x_1} - e^{(b_1 m - a + b_1')x_1}}{(a-b_1')(b_1 m - a + b_1')}\right.\right.$$
$$\left. - \frac{e^{b_1 m x_1}-1}{b_1 m(b_1 m - a + b_1')}\right] + C_1\left[\frac{1+\left(\frac{N_p}{N_m}\right)^m}{C_2 a}(e^{-ax_1} - e^{-ax_2}) + A\right]\frac{e^{(a-b_1')x_1}-1}{a-b_1'}\Bigg\}$$
$$+\frac{x_1|a-b_1'|}{v_s(a-b_1')} + \left\{\frac{C_1}{SC_2 e^{(a+b_2')W_b - b_2' x_2}} - \frac{1}{v_s}\left[\frac{|a-b_1'|/(a-b_1')-1}{e^{(a-b_1')x_1}}\right.\right.$$
$$\left.\left. + \frac{C_1}{C_2 e^{(a+b_2')W_b - b_2' x_2}}\right]\right\}\frac{e^{(a-b_1')x_1}-1}{a-b_1'} \tag{3.31}$$

$$\tau_{b2} = B + \frac{x_2 - x_1}{v_s} + \left(\frac{1}{S} - \frac{1}{v_s}\right)\frac{e^{ax_2} - e^{ax_1}}{a e^{W_b(a+b_2') - b_2' x_2}} \tag{3.32}$$

$$\tau_{b3} = \frac{1}{D_{no}}\left\{\frac{e^{(a+b_2')(x_2-W_b)}-1}{(a+b_2')^2} + \frac{W_b - x_2}{a+b_2'} + \left(\frac{N_p}{N_m}\right)^m\left[\frac{e^{(a+b_2'+b_2 m)(x_2-W_b)} - e^{b_2 m(x_2-W_b)}}{(a+b_2')(b_2 m + a + b_2')}\right.\right.$$
$$\left.\left. + \frac{1-e^{b_2 m(x_2-W_b)}}{b_2 m(b_2 m + a + b_2')}\right]\right\} + \frac{W_b - x_2}{v_s} + \left(\frac{1}{S} - \frac{1}{v_s}\right)\frac{1-e^{(a+b_2')(x_2-W_b)}}{a+b_2'} \tag{3.33}$$

where

$$A = \frac{e^{-ax_2} - e^{b_2'(x_2-W_b)-aW_b}}{C_2(a+b_2')} + \frac{\left(\frac{N_p}{N_m}\right)^m\left[e^{-ax_2} - e^{(b_2 m + b_2')(x_2-W_b)-aW_b}\right]}{C_2(b_2 m + a + b_2')}$$
$$B = \begin{cases} \frac{1+\left(\frac{N_p}{N_m}\right)^m}{a^2 D_{no}}\left[e^{a(x_1-x_2)}-1\right] + \frac{1+\left(\frac{N_p}{N_m}\right)^m}{aD_{no}}(x_2-x_1) + \frac{C_2 A}{aD_{no}}(e^{ax_2} - e^{ax_1}) & \text{if } a \neq 0 \\ \frac{1+\left(\frac{N_p}{N_m}\right)^m}{2D_{no}}(x_2-x_1)^2 + \frac{C_2 A}{D_{no}}(x_2-x_1) & \text{if } a = 0 \end{cases} \tag{3.34}$$

It should be noted that (3.31) is invalid when $a = b_1'$ or $b_1 m = a - b_1'$. When $a = b_1'$, the following equation, which is directly derived from (3.20) with $D_n = D_n^l$, should be used:

$$\tau_{b1} = \frac{1}{D_{no}}\left\{\frac{x_1^2}{2} + \left(\frac{N_o}{N_m}\right)^m \left[\frac{x_1 e^{b_1 m x_1}}{b_1 m} - \frac{e^{b_1 m x_1} - 1}{(b_1 m)^2}\right] + C_1 x_1 \left[\frac{1 + \left(\frac{N_p}{N_m}\right)^m}{C_2 a}(e^{-a x_1} - e^{-a x_2}) + A\right]\right\} + \frac{C_1 x_1}{S\ C_2 e^{(a+b_2')W_b - b_2' x_2}} \tag{3.35}$$

Furthermore, when $b_1 m = a - b_1'$, the following expression, which is derived from (3.28) directly, should be used:

$$\begin{aligned}\tau_{b1} = {} & \frac{1}{D_{no}}\left\{\frac{e^{(b_1'-a)x_1} - 1}{(a-b_1')^2} + \frac{x_1}{a-b_1'} + \left(\frac{N_o}{N_m}\right)^m \left[-\frac{x_1}{a-b_1'} + \frac{e^{-x_1(a-b_1')}}{(a-b_1')^2}\right.\right. \\ & \left.\left. - \frac{1}{(a-b_1')^2}\right] + C_1\left[\frac{1+\left(\frac{N_p}{N_m}\right)^m}{C_2 a}(e^{-a x_1} - e^{-a x_2}) + A\right]\frac{e^{(a-b_1')x_1} - 1}{a-b_1'}\right\} + \frac{x_1|a-b_1'|}{v_s(a-b_1')} \\ & + \left\{\frac{C_1}{S C_2 e^{(a+b_2')W_b - b_2' x_2}} - \frac{1}{v_s}\left[\frac{|a-b_1'|/(a-b_1') - 1}{e^{(a-b_1')x_1}} + \frac{C_1}{C_2 e^{(a+b_2')W_b - b_2' x_2}}\right]\right\} \\ & \times \frac{e^{(a-b_1')x_1} - 1}{a-b_1'}\end{aligned} \tag{3.36}$$

If no bandgap narrowing due to heavy doping effects is considered, b_1' and b_2' become b_1 and b_2, respectively. If velocity saturation at the collector-base junction is ignored and $n(W_b)$ is assumed to be zero, S approaches infinity in (3.31)–(3.36). Similarly, set $v_s \to \infty$ if the electric-field dependency of the diffusivity is ignored. If the dependency of diffusivity on doping concentration is neglected, both $(\frac{N_o}{N_m})^m$ and $(\frac{N_p}{N_m})^m$ should be set to zero in (3.31)–(3.36). If no Ge exists in the base, set $v_s \to \infty$ in (3.32) and the limits of (3.31)–(3.36), when $a \to 0$, should be taken instead.

3.3 Results

Fig. 3.3 shows the sensitivity of τ_{b1}, τ_{b2}, and τ_{b3} to the base concentration at the edge of the emitter-base depletion layer, N_o, for the Si BJT and the SiGe HBT studied by Gao *et al.* in [175]. This result is obtained by assuming $S = v_s$ =10^7cm/s and a =7.74 × 10^6/cm (i.e. a Ge grading of 13.5%), and that the retrograde region occupies 30% of the neutral base.

Both τ_{b2} and τ_{b3} are insensitive to the variation of N_o because they depend only on parameters in regions II and III (see (3.29) and (3.30)). Whereas, τ_{b1} decreases significantly with N_o since a higher N_o lessens the retarding electric field induced by the retrograde profile in region I. However, this variation in τ_{b1} diminishes for SiGe HBTs (denoted by the dashed line) because the electric field induced by the Ge concentration gradient may dominate the retarding field in the retrograde region.

Fig. 3.4 demonstrates how the Ge grading across the base affects the sensitivity of the τ_{b1}/τ_b ratio to N_o for SiGe HBTs of different fractions of the retrograde region I (10% and 50%) and neutral base widths (50 nm and 200 nm). Again, we assume that $S = v_s$ =10^7cm/s.

For both base widths, the variation of the τ_{b1}/τ_b ratio is less than 15% as N_o is increased from 5×10^{17}cm^{-3} to 9×10^{18}cm^{-3}. Furthermore, this variation reduces with a higher Ge grading. This result is consistent with Fig. 3.3 in that the retarding electric field in the retrograde region I does not affect τ_b significantly. In fact, τ_b is more sensitive to the ratio of the retrograde region width to the neutral base width. As shown in Fig. 3.4, the τ_{b1}/τ_b ratio increases by around 60% when x_1/W_b changes from 0.1 to 0.5.

Fig. 3.5 compares Gao *et al.*'s results (using (7)–(10) in [175]) with our results for a Si BJT and a SiGe HBT. In obtaining our results, all effects are considered

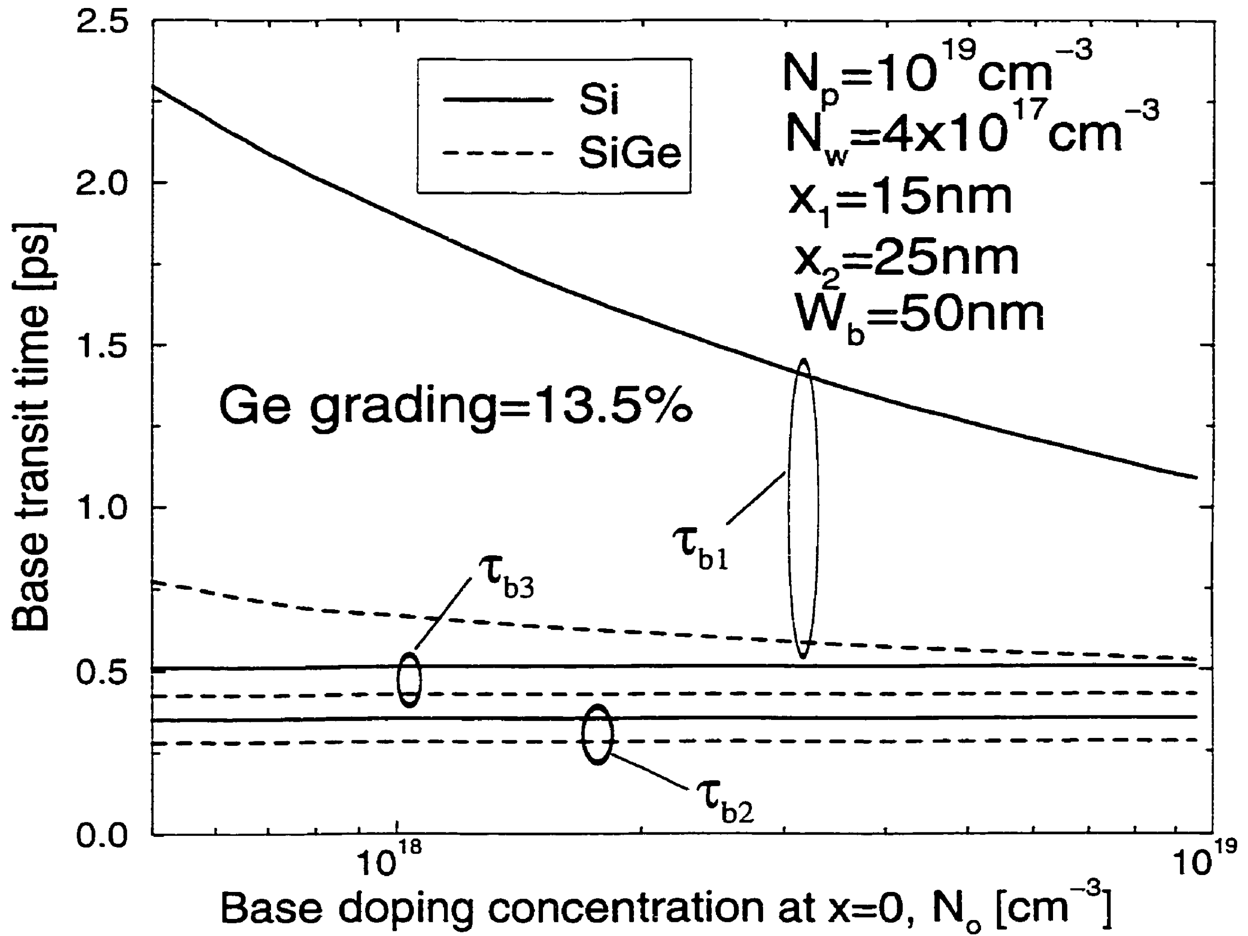

Figure 3.3: Base transit times in different subregions as a function of the base concentration at the edge of the depletion layer (with all effects considered and $S = v_s = 10^7$cm/s assumed).

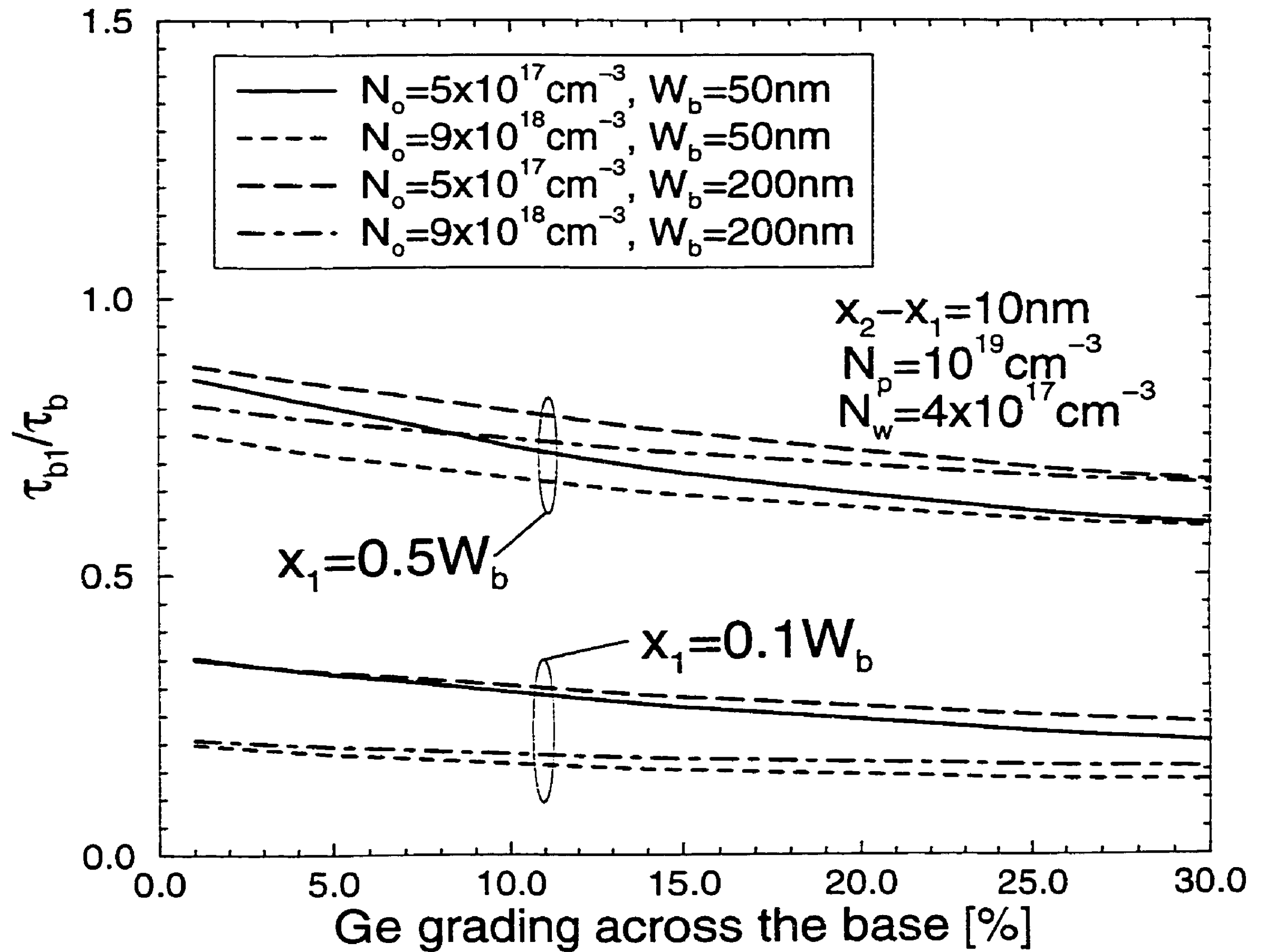

Figure 3.4: Fraction of the base transit time of the retrograde region as a function of the Ge grading across the base (with all effects considered and $S = v_s = 10^7$cm/s assumed).

and $S = v_s = 10^7$cm/s is assumed. Other parameters used for both sets of results are $N_p = 10^{19}$cm^{-3}, $N_w = 4 \times 10^{17}$cm^{-3}, $x_1 = 15$ nm, $x_2 = 25$ nm, and $W_b = 50$ nm.

In Fig. 3.5, both τ_b and τ_{b1} are shown. According to Gao *et al.*'s results (denoted by dashed lines) for Si and SiGe transistors, the τ_{b1}/τ_b ratio increases by ~0.30 and ~0.26, respectively, as N_o drops from 9.5×10^{18} to 5×10^{17}cm^{-3}. However, our results indicate a less significant increase: ~0.17 for Si and ~0.09 for SiGe. In terms of the total base transit time of SiGe transistors, for the same N_o drop, our results show an increase of 20%, whereas Gao *et al.* 's results show an increase of 71%. The effect of the retrograde region is smaller for the SiGe transistor because the electric field induced by the Ge concentration gradient dominates the retarding field in the retrograde region. This implies that the retarding electric field does not influence τ_b as much as Gao *et al.* have reported. The discrepancy between Gao *et al.* 's results and our results is mainly due to the incorrect derivation for Gao *et al.* 's analytical expressions.

Fig. 3.6 assesses the relative importance of each effect, mentioned above, on τ_b as the Ge grading varies across the base of the two SiGe HBTs with different x_1/W_b ratios. Since the y-axis is the ratio of the τ_b modelled with one of the effects neglected to the τ_b modelled with all effects considered, the closer to unity the curve is, the less significant is the neglected effect. In the case where the effect of the doping dependency of electron diffusivity is ignored, an averaged diffusivity, D_n^*, calculated from the averaged base concentration and Klaassen's low-field mobility model [121], is used.

For both transistors with different x_1/W_b ratios, the curves associated with the effect of the doping dependency of the electron diffusivity ($D_n(N)$) are close to unity and does not vary much with the Ge grading. This implies that the influence of $D_n(N)$ effect on τ_b is neither significant nor altered much by the Ge grading.

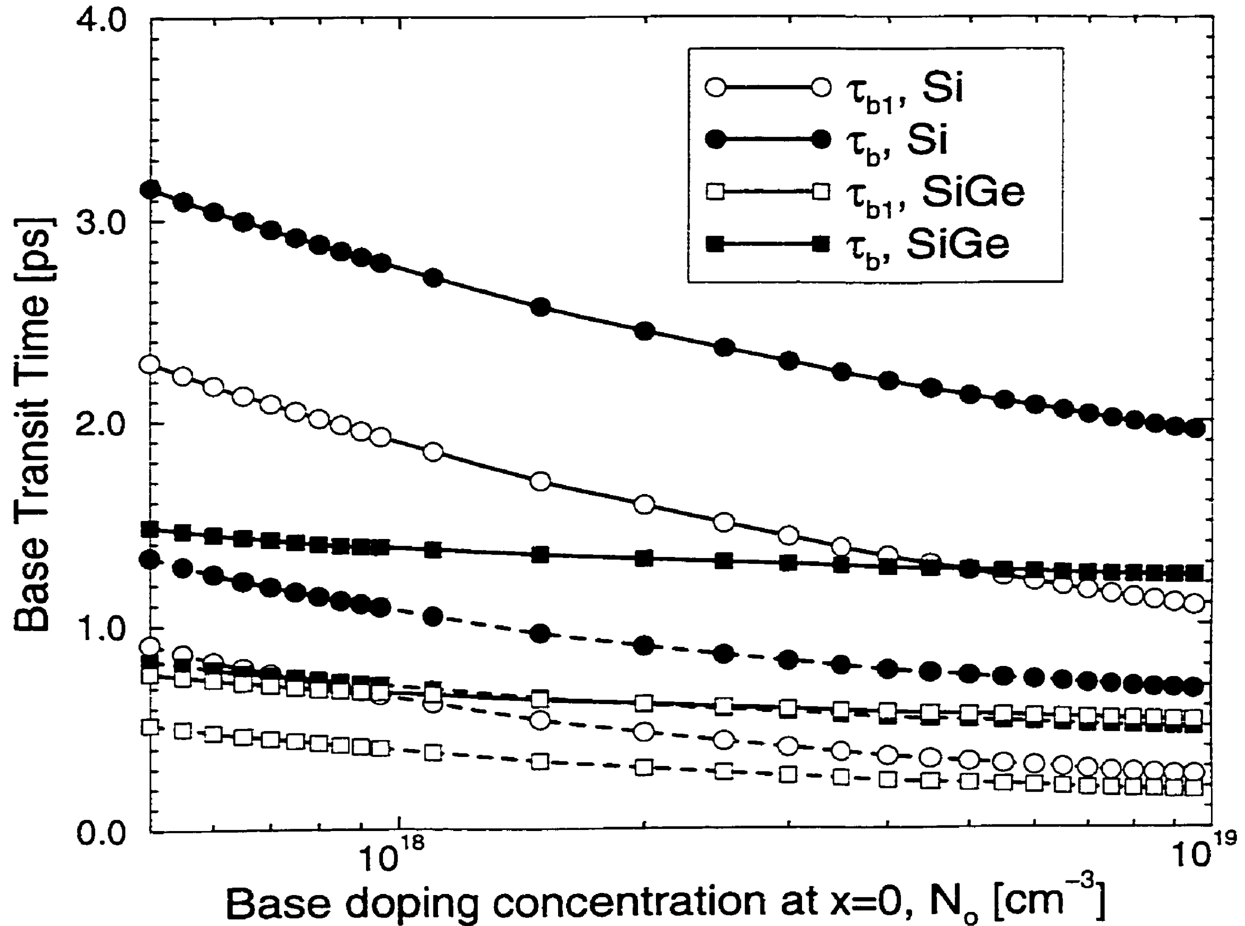

Figure 3.5: Base transit time of the retrograde region and the total base transit time as a function of the base concentration at the edge of the depletion layer. Solid lines: our results with all effects considered and $S = v_s = 10^7$cm/s assumed. Dashed lines: Gao *et al.*'s results using (7)–(10) in [175] with $D_n^* = 4.4$ cm^2/s and F = 20 kV/cm (i.e., a Ge grading of 13.5%). Other parameters: $x_1 = 15$ nm, $x_2 = 25$ nm, $W_b = 50$ nm, $N_p = 10^{19}\text{cm}^{-3}$, $N_w = 4 \times 10^{17}\text{cm}^{-3}$.

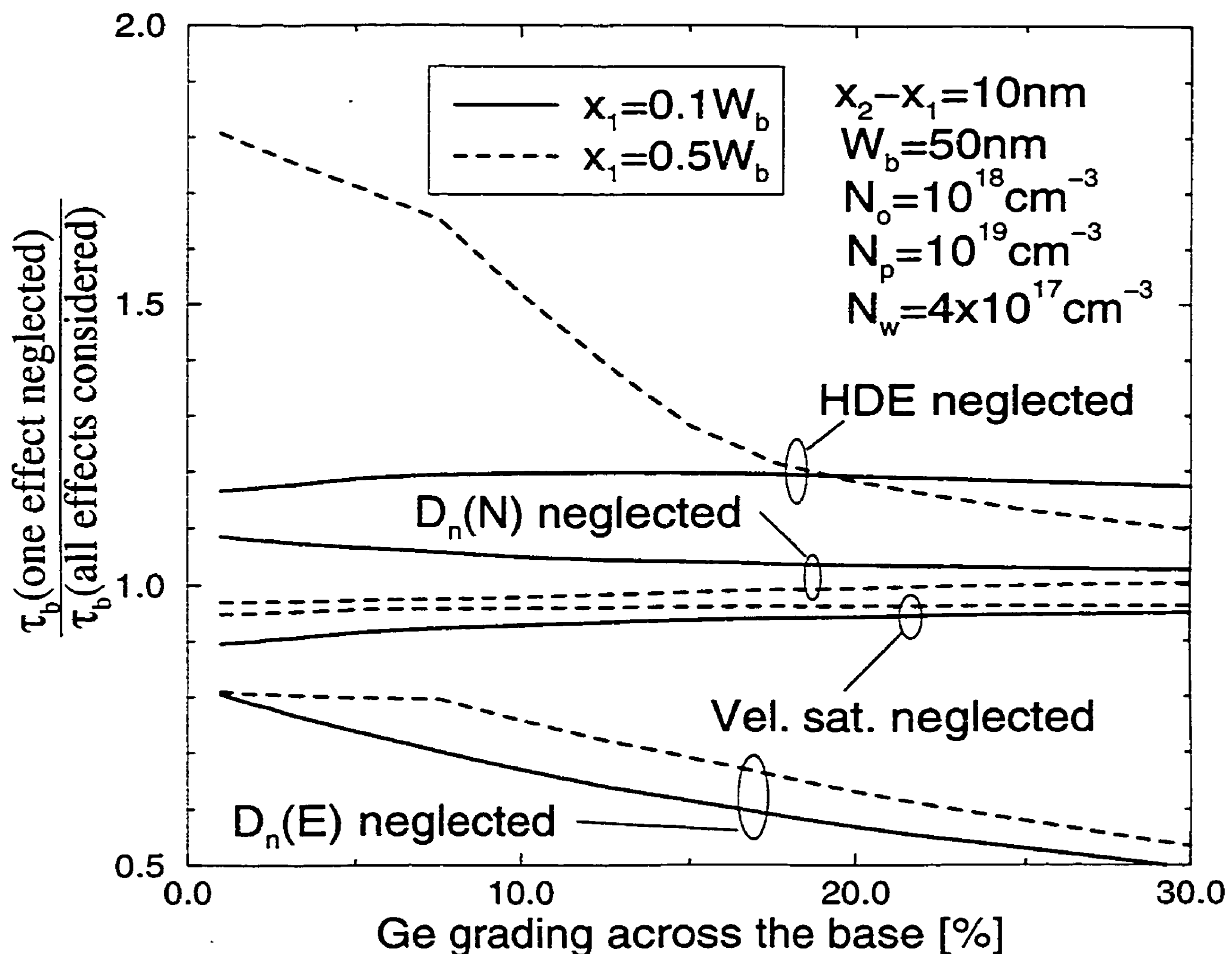

Figure 3.6: Ratio of the base transit time with one effect neglected to the base transit time with all effects considered as a function of the Ge grading across the base ($S = v_s = 10^7$cm/s assumed).

The curves corresponding to the effect of velocity saturation at the collector-base junction are very close to unity, regardless of the Ge grading. However, it should not be concluded that the effect of velocity saturation is insignificant because we have assumed $S = v_s = 10^7$cm/s in our calculation, which is a relatively high value. In fact, we have found that when S is reduced to 10^6cm/s, the corresponding curves will move farther away from unity (Note. The effect of S can also be seen in Fig. 3.7).

The effect of the electric-field dependency of the diffusivity ($D_n(E)$) becomes increasingly significant as the Ge grading increases. This is expected because a higher Ge grading across the base induces a higher electric field. The higher the electric field, the more pronounced this effect will be.

The impact of heavy doping effects (HDE) on τ_b is weakly affected (within 5%) by the Ge grading, except when the retrograde region occupies a significant fraction of the neutral base width. For $x_1/W_b = 0.5$, at low Ge gradings, the effect of HDE becomes important. But its effect diminishes with increasing Ge grading.

3.4 Verification and comparison

Results obtained from our analytical expressions, numerical integration, and Gao *et al.* [175] are compared with the published simulation data of Patton *et al.* [93] for their 75-GHz SiGe HBT. The base region of Patton *et al.*'s SiGe HBT is approximated as follows:

$$\begin{aligned}
N_{B1}(x) &= 4.14 \times 10^{18} e^{5.532\times10^5 x} \quad , 0 \le x \le 14.9 \text{ nm} \\
N_{B2}(x) &= 9.44 \times 10^{18} \quad , 14.9 \text{ nm} \le x \le 20 \text{ nm} \\
N_{B3}(x) &= 9.44 \times 10^{18} e^{-1.29\times10^6 (x-20\times10^{-7})} \quad , 20 \text{ nm} \le x \le 46.7 \text{ nm}
\end{aligned} \tag{3.37}$$

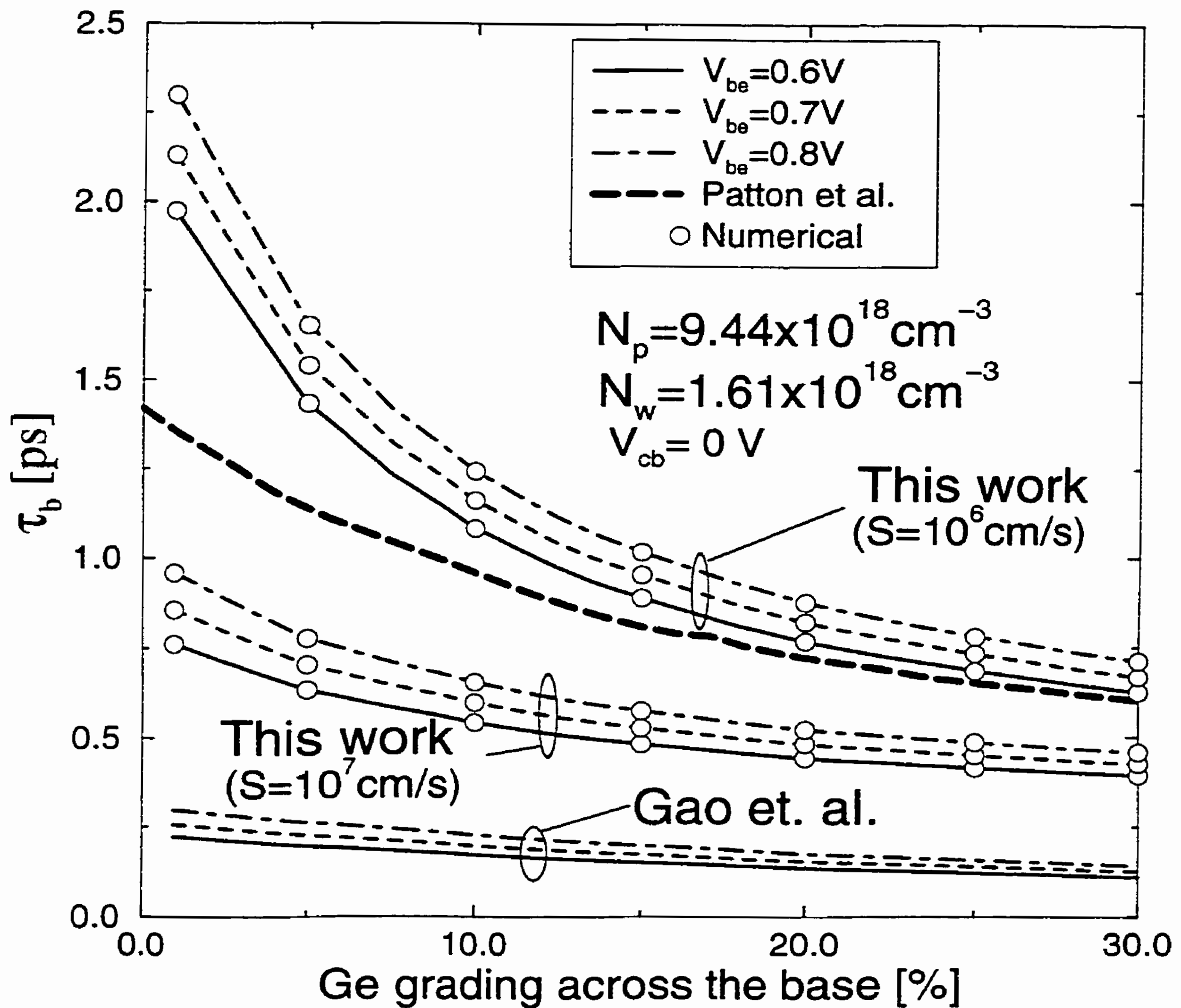

Figure 3.7: Base transit time as a function of the Ge grading across the base. Darker dashed line: Patton *et al.*'s simulation results [93]. Solid line: $x_1 = 7.4$ nm, $x_2 = 12.5$ nm, $W_b = 26.2$ nm, $N_o = 6.27 \times 10^{18}\mathrm{cm}^{-3}$. Dashed line: $x_1 = 8.9$ nm, $x_2 = 14$ nm, $W_b = 27.7$ nm, $N_o = 5.77 \times 10^{18}\mathrm{cm}^{-3}$. Dot-dashed line: $x_1 = 10.4$ nm, $x_2 = 15.5$ nm, $W_b = 29.2$ nm, $N_o = 5.31 \times 10^{18}\mathrm{cm}^{-3}$. Gao *et al.*'s results are generated using (7)–(10) in [175] with $D_n^* = 4.4$ cm^2/s.

where x is in cm. Since the emitter-base voltage (V_{be}) was not specified in [93], three values of V_{be} (0.6, 0.7, and 0.8 V) are considered here. The locations of the neutral base region (or the depletion layer widths of the emitter-base junction and the base-collector junction on the base side) are determined by MEDICI numerical simulations [179] at the specified base-emitter voltages. It is found that the depletion layer widths of the emitter-base junction on the base side at $V_{be} = 0.6, 0.7, 0.8$ V are 7.5, 6, and 4.5 nm, respectively. The depletion layer width of the collector-base junction does not vary much with the base-emitter voltage and is $\sim$ 13 nm under a zero collector-base voltage. For each value of V_{be} and the corresponding depletion layer width, we then approximate the resulting neutral base profile in each subregion using (3.37).

Fig. 3.7 shows how τ_b varies with the Ge grading across the base under the three bias conditions. Assuming that S is in the range of 10^6cm/s $\rightarrow$ 10^7cm/s is reasonable because the Ge grading can reduce the formation of energy spikes at the collector-base junction and can cause most of the band offset in the valence band in SiGe [93]. Gao *et al.*'s results are obtained from analytical expressions ((7)–(10) published in [175]) with $D_n^* = 4.4\text{cm}^2/\text{s}$. Identical structural parameters are used for generating our results and Gao *et al.*'s. As shown in the figure, our results agree well with Patton *et al.*'s simulations, while Gao *et al.*'s results deviate significantly from both Patton *et al.*'s and our results. Furthermore, the results obtained from our analytical expressions are in agreement with those obtained by numerical integration.

To further validate our analytical expressions, we consider the special case of a uniformly-doped neutral base with a linearly graded Ge profile (i.e., $b_1 = b_1' = b_2 = b_2' = 0, C_1 = C_2, N_o = N_p = N_w$). When the effects of doping dependency and electric-field dependency of the electron diffusivity are ignored (as what Kroemer

did in [135]), (3.31)–(3.34) become:

$$\begin{aligned}
\tau_{b1} =& \frac{1}{D_{no}}\left\{\frac{e^{-ax_1}-1}{a^2}+\frac{x_1}{a}+C_1\left[\frac{e^{ax_1}-e^{ax_2}}{C_2 a}+A\right]\frac{e^{ax_1}-1}{a}\right\} \\
\tau_{b2} =& B \\
\tau_{b3} =& \frac{1}{D_{no}}\left\{\frac{e^{a(x_2-W_b)}-1}{a^2}+\frac{W_b-x_2}{a}\right\} \\
A =& \frac{e^{-ax_2}-e^{-aW_b}}{C_2 a} \\
B =& \frac{e^{a(x_1-x_2)}-1}{a^2 D_{no}}+\frac{x_2-x_1}{aD_{no}}+\frac{C_2 A}{aD_{no}}(e^{ax_2}-e^{ax_1})
\end{aligned} \tag{3.38}$$

Therefore,

$$\begin{aligned}
\tau_b &= \tau_{b1}+\tau_{b2}+\tau_{b3} \\
&= \frac{W_b}{aD_{no}}\left(1-\frac{1-e^{-aW_b}}{aW_b}\right)
\end{aligned} \tag{3.39}$$

This is the same as Kroemer's expression (equation (16) in [135] with $a = F/V_T$ and $e^{-aW_b} = (n_{io}/n_{iw})^2$).

3.5 Conclusions

We have presented a regional base transit time (τ_b) model which results in a set of closed-form analytical expressions for an exponentially-doped retrograde base of a SiGe HBT with a linearly graded Ge profile. This model considers the retarding built-in electric field due to the retrograde region, heavy doping effects, the effect of velocity saturation at the collector-base junction, the effects of the doping dependency and the electric-field dependency of the electron diffusivity, and the electric field induced by the Ge concentration gradient. The impact of these different effects

on τ_b has been assessed. Our model results show that the retarding field due to the retrograde region in the base does not increase τ_b as much as Gao *et al.* reported, especially when a high Ge grading exists and when the retrograde region is small compared with the total base width. For a Ge grading of 13.5 at.%, the base transit time increases by $\sim$ 20% (vs. 71% predicted by Gao *et al.*) when the base concentration near the emitter-base junction drops from $9.5 \times 10^{18}\text{cm}^{-3}$ to $5 \times 10^{17}\text{cm}^{-3}$. The impact of the retrograde region on τ_b reduces when a high Ge grading (e.g. 13.5 %) exists in a base with the retrograde region occupying only a small fraction of the total base width (e.g. 10%). For a high Ge grading across the base, the effect of the electric-field dependency of the electron diffusivity on τ_b becomes very pronounced and must be considered in the model. We also show that this model results in Kroemer's analytical expression for the special case of a uniformly-doped base. The results have been verified by numerical integration. They agree well with the published simulation data of Patton *et al.* for their 75-GHz SiGe HBT.

Chapter 4

Doping and Ge profile optimization for minimum base delay time before onset of Kirk effect

4.1 Introduction

To fully utilize the high-frequency performance of HBTs, the base doping and germanium profiles must be carefully designed. Since the base delay (or transit) time (τ_b) is one of the important delay time contributors to f_T, a few base profile optimization studies for minimizing τ_b have been published [180–185]. Broadly speaking, two approaches have been taken in these studies. The first approach adopts numerical methods [180] or optimization theory [181] to obtain the optimal Ge profile for minimizing τ_b. However, such approach has two limitations: i) a

uniformly doped base and a low-injection condition are often assumed in order to make the problem more tractable, and ii) with the constraint of a fixed Ge dose in the base, the resulting optimal Ge profile is often difficult to realize in fabrication.

The second approach is to derive closed-form analytical expressions to compare τ_b for different combinations of more realistic Ge and doping profiles in the base. Under this paradigm, two regimes of operation can be considered: low injection [182–185] and high injection before the onset of the Kirk effect. Unlike other works in the low-injection regime, Patri and Kumar's study [185] provided a more meaningful comparison by considering Ge profiles under the important criterion of identical film stability. However, they did not account for the electric-field dependency of the electron diffusion coefficient, which has been shown to be significant in a HBT with Ge grading [89]. Also, in their comparison, the peak base concentration, instead of the base concentration near the emitter, was kept constant. A more useful comparison is to keep the latter the same as it is a stronger factor than the former in determining the emitter-base junction capacitance. This was first suggested by Varnerin [81] and recently emphasized by Hamel [186].

In the high-injection regime, to our knowledge, no profile comparison study for SiGe HBTs is available, although modern transistors are often biased at high currents to reduce charging times. Even for Si bipolar transistors, most studies in this regime focus only on how to accurately and efficiently solve for the minority carrier concentration, the current density, and subsequently τ_b [115, 187–192]. Although Suzuki [193] and Ma *et al.* [194] have compared different base doping profiles in Si bipolar transistors for minimizing τ_b, the doping profiles considered in both studies correspond to different base resistances. This makes the comparison less meaningful.

Unlike Suzuki and Ma *et al.*, Yuan [195] compared a uniform doping profile and

an exponential one with the same base resistance. The electric field due to the variation of bandgap narrowing in the base and the modulation of injected carriers under high injection was considered in determining the diffusion coefficient. But in calculating the minority carrier concentration for the case of the exponentially-doped base, this electric field was neglected. This compromises the accuracy of the comparison.

Apart from the aforementioned limitations, three important points are seldom addressed in the base transit time studies for high injection: i) the definition of base-emitter voltage, and ii) the impact of plasma-induced energy bandgap narrowing (BGN), and iii) the assumption of identical transport parameters for Si and SiGe. Although τ_b is often calculated for a given V_{BE}, most studies did not give a clear definition for V_{BE}. One exception is the study by Suzuki [193], who correctly distinguished between $V_{BE,t}$, the base-emitter *terminal* voltage, and V_{BE}, the built-in voltage minus the base-emitter *junction* voltage. Under high injection, the electric field in the base becomes significant and can cause a substantial difference between $V_{BE,t}$ and V_{BE}.

Regarding the second point, Wu and Lindholm [196] have pointed out the significance of the phenomenon of bandgap narrowing in the presence of high concentrations of mobile carriers, which is likely under high injection. However, to our knowledge, this has not been considered when studying the base profile design for minimizing τ_b in the high-injection regime. Finally, except for the Ge-induced bandgap narrowing, most profile comparison studies simply assume Si values for SiGe parameters.

The objective of this chapter is to study and compare the τ_b performance for different combinations of Ge profile and doping profile in SiGe HBTs operating in both low-injection and high-injection regimes before the onset of the Kirk effect.

To provide a meaningful comparison, we fix the intrinsic base resistance (under low-injection condition), the base concentration near the emitter, and the Ge dose in the base while comparing Ge profiles of different shapes and doping profiles of different base widths. In calculating τ_b for a given collector current density J_n (instead of V_{BE}), we adopt a consistent set of SiGe transport parameters and include important effects such as the electric-field dependency of diffusion coefficient and plasma-induced bandgap narrowing. Results in this chapter have been reported in [90].

The organization of the chapter is as follows. Section 4.2 presents the theoretical framework for calculating τ_b, which will be verified in Section 4.3. Sections 4.4 and 4.5 present and discuss the profile comparison results for the regimes of low injection and high injection, respectively. Conclusions are drawn in Section 4.6.

4.2 Theory

4.2.1 Assumptions

The following assumptions are made in our study:

1. The transistor operates under conditions before the onset of high current effects such as emitter crowding [197], Kirk effect [198] and van der Ziel-Agouridis effect [199]. Base conductivity modulation caused by the rise of the majority carrier concentration in the high-injection regime takes place only in the intrinsic device region. We consider a SiGe base HBT with a collector of $9 \times 10^{16}\text{cm}^{-3}$ and 0.5 μm, operating with a collector-base terminal voltage of 2 V, for which the Kirk onset current density is $\sim 1.4 \times 10^5$ A/cm^2. The transistor operates at 300 K.

2. All base doping and Ge profiles considered here are located within the neutral base region. The neutral base width is constant before the onset of Kirk effect.

3. The drift-diffusion equations are still valid for the base widths considered in this study ($\sim 25-100$ nm). Stettler and Lundstrom [200] have shown that the carrier travels ballistically only when the base width is *much smaller* than the carrier's collision-free path length ($\sim 10-20$ nm for heavily-doped silicon).

4. The base current density can be approximated to be zero. Yue *et al.* [192] have conducted an ambipolar study without assuming a zero base current and found that the assumption of zero base current will be more valid when the base width decreases and the base concentration increases. This together with the fact that a thin base is often used in modern bipolar transistors (less than 100 nm in this work) allows us to neglect neutral base recombination (i.e. J_n is constant).

5. Boltzmann statistics can still be used as an approximation. The averaged base concentration considered in this study does not exceed 10^{19}cm^{-3}.

6. Boltzmann quasi-equilibrium (or the non-equilibrium equivalent of Boltzmann relation) is still valid even under high forward base-emitter bias [201]. This allows the use of the law of the junction or the Fletcher boundary conditions.

7. The law of the junction can approximate the Fletcher boundary conditions [202] and be applied across the space charge region of the emitter-base junction as long as the emitter concentration is much higher than the electron concentration at the emitter edge of the base (e.g. $N_{DE} = 10^{19}\text{cm}^{-3}$) [203].

8. For the boundary condition at the base-collector junction, it is assumed that for both low and high injection the electron concentration at the collector edge

of the base is related to the electron current density through the saturated velocity, v_s. This is more accurate than assuming a zero electron concentration (i.e. an infinite carrier velocity) at the edge for a finite electron current density. For high collector current density, the electric field at the junction begins to drop and it is not accurate to assume an infinite carrier velocity at the collector edge of the base [204].

4.2.2 Profile Definitions

Although any arbitrary doping profile can be used in this work, we have chosen four representative profiles (shown in Fig. 4.1) for comparison: uniform, Gaussian, exponential, and retrograde profiles. The four profiles can be analytically described as follows:

$$N_b(x) = \begin{cases} N_p e^{-\eta_e[(R_p - x)/R_p]^\alpha}, & 0 \leq x \leq R_p \\ N_p e^{-\eta_c[(x - R_p)/(W_b - R_p)]^\alpha}, & R_p \leq x \leq W_b \end{cases} \tag{4.1}$$

where $\eta_e = ln(N_p/N_{be})$, $\eta_c = ln(N_p/N_{bc})$, R_p is the peak location of the profile, N_{be} is the base concentration near the emitter, N_{bc} is the base concentration near the collector, N_p is the peak concentration of the profile, and W_b is the neutral base width. For a uniform doping profile, $\eta_e = \eta_c = 0, N_p = N_{be}$. For the exponential and the Gaussian profiles, $\alpha = 1$ and 2, respectively, and $N_p = N_{be}, R_p = 0$. For a retrograde profile, $\alpha = 2$, and N_p must be larger than N_{be}.

Following [185], we describe the Ge profile as follows. Since the Ge dose in the base is kept constant, regardless of the shape of the Ge profile, the Ge fraction $y(x)$

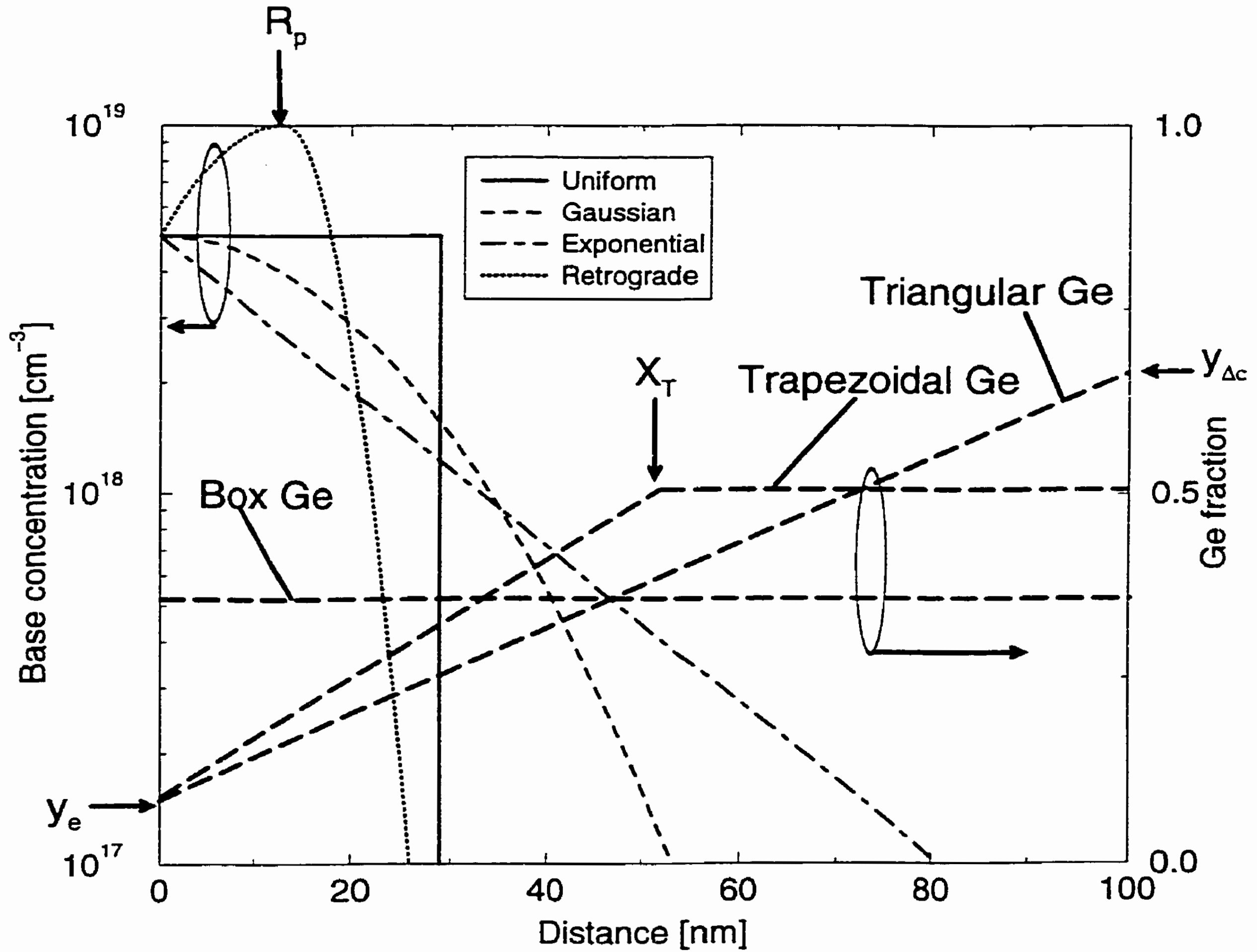

Figure 4.1: Ge profiles (W_b is arbitrarily set to 100 nm for illustrative purpose) and base doping profiles with intrinsic base resistance of $5\text{k}\Omega/\Box$, N_{be} $=5\times10^{18}\text{cm}^{-3}$,$N_{bc}$$=9\times10^{16}\text{cm}^{-3}$: uniform ($W_b = 29\ nm$), Gaussian ($W_b = 54\ nm$), exponential ($W_b = 83\ nm$), retrograde ($W_b = 26\ nm, R_p = 13\ nm, N_p = 10^{19}\text{cm}^{-3}$).

is defined as:

$$y(x) = \begin{cases} [y(X_T) - y_e] \frac{x}{X_T} + y_e, & 0 \leq x \leq X_T \\ y(X_T), & X_T \leq x \leq W_b \end{cases} \tag{4.2}$$

where

$$y(X_T) = \frac{y_{\Delta c} - y_e}{2 - X_T/W_b} + y_e$$

and X_T defines the shape of the Ge profile. This is, $X_T = 0$ for a box Ge profile, $X_T = W_b$ for a triangular Ge profile, and $0 < X_T < W_b$ for a trapezoidal Ge profile (see Fig. 4.1). The quantity y_e is the Ge fraction near the emitter, while $y_{\Delta c}$ is the Ge fraction near the collector when the Ge profile becomes triangular.

4.2.3 Solution method

Considering an n-p-n transistor, the base transit time is defined as the ratio of the injected charge to the electron or collector current density[1]:

$$\tau_b = \frac{q \int_0^{W_b} n(x)dx}{J_n} \tag{4.3}$$

where $n(x)$ is the electron concentration; q, the electronic charge; J_n, the conventional collector current density in the base. Please note that we define J_n as the absolute value of the actual collector current density. To calculate τ_b, one needs to know $n(x)$ and J_n.

Unlike other similar studies [115, 187–191], where researchers calculate J_n and

[1]As neutral base recombination is neglected, J_n and J_c are used interchangeably.

$n(x)$, and subsequently τ_b for a given value of V_{BE}, we assume that J_n is given and use it to calculate V_{BE} and $n(x)$, and subsequently τ_b. There are two advantages of this approach. First, it naturally leads to the τ_b-J_n plot, which is more useful than the τ_b-V_{BE} plot since it is more common for circuit designers to refer the operating or biasing point of a device to J_n than V_{BE}. Second, quoting τ_b by J_n instead of V_{BE} avoids the possibility of confusing the definition of V_{BE} as mentioned.

In the following, we shall derive expressions relating V_{BE}, J_n, and $n(x)$. A key parameter is the electron diffusion coefficient, $D_n(x)$, which is a function of the electric field, $\mathcal{E}(x)$, in the base. Therefore, we shall first derive the expression for $E(x)$.

With $J_p \sim 0$, we can write the electric field in the base as:

$$\mathcal{E}(x) = V_T \frac{d}{dx} [ln(p(x))] - \frac{1}{q} \frac{d\Delta E_g(x)}{dx} \tag{4.4}$$

where V_T is the thermal voltage; p is the hole concentration in the base; $\Delta E_g(x)$ is the apparent electrical bandgap narrowing in the base, which can be expressed in terms of the effective intrinsic carrier concentration, n_{ie}, as follows:

$$\Delta E_g(x) = kT\ ln \left(\frac{n_{ie}^2(x)}{n_i^2} \right) \tag{4.5}$$

where n_i is the intrinsic carrier concentration; k, the Boltzmann constant; T, the temperature. Substituting (4.5) into (4.4) gives:

$$\mathcal{E}(x) = V_T \frac{d}{dx} \left[ln \left(\frac{p(x)}{n_{ie}^2(x)} \right) \right] \tag{4.6}$$

Now, we are ready to relate V_{BE}, J_n, and $n(x)$ together. Substituting (4.6) into

the electron current transport equation

$$-J_n = q\frac{D_n(x)n(x)\mathcal{E}(x)}{V_T} + qD_n(x)\frac{dn(x)}{dx},$$

we obtain:

$$-\frac{J_n p}{qD_n(x)n_{ie}^2(x)} = \frac{d}{dx}\left[\frac{p(x)n(x)}{n_{ie}^2(x)}\right] \tag{4.7}$$

Using the principle of charge neutrality ($p(x) = n(x) + N_b(x)$), the law of the junction at $x = 0$, and the boundary condition at $x = W_b$ (i.e., $J_n = qv_s n(W_b)$), (4.7) is integrated from 0 to W_b:

$$J_n = \frac{qe^{V_{BE}/V_T}}{\int_0^{W_b} \frac{n(x)+N_b(x)}{D_n(x)n_{ie}^2(x)}dx + \frac{n(W_b)+N_B(W_b)}{v_s n_{ie}^2(W_b)}} \tag{4.8}$$

Observing that $J_n = qv_s n(W_b)$ at $x = W_b$, we can write (4.8) as,

$$J_n = \frac{qe^{V_{BE}/V_T}}{\int_0^{W_b} \frac{n(x)+N_b(x)}{D_n(x)n_{ie}^2(x)}dx + \frac{N_B(W_b)}{v_s n_{ie}^2(W_b)} + \frac{J_n}{qv_s^2 n_{ie}^2(W_b)}} \tag{4.9}$$

Therefore, the base-emitter voltage (defined as the built-in voltage at the base-emitter junction minus the base-emitter junction voltage after a terminal V_{BE} voltage is applied) can be expressed as:

$$V_{BE} = V_T\, ln\left\{\frac{J_n}{q}\left[\int_0^{W_b} \frac{n(x)+N_b(x)}{D_n(x)n_{ie}^2(x)}dx + \frac{N_b(W_b)}{v_s n_{ie}^2(W_b)} + \frac{J_n}{qv_s^2 n_{ie}^2(W_b)}\right]\right\} \tag{4.10}$$

To determine $n(x)$, we integrate (4.7) from 0 to x and apply the principle of

charge neutrality and the law of the junction at $x = 0$:

$$-\frac{J_n}{q}\int_0^x \frac{n(z)+N_b(z)}{D_n(z)n_{ie}^2(z)}dz = \frac{[n(x)+N_b(x)]n(x)}{n_{ie}^2(x)} - e^{V_{BE}/V_T} \tag{4.11}$$

A quadratic equation of $n(x)$ is then obtained:

$$n(x)^2 + N_b(x)n(x) - n_{ie}^2(x)e^{V_{BE}/V_T} + \frac{J_n n_{ie}^2(x)}{q}\int_0^x \frac{n(z)+N_b(z)}{D_n(z)n_{ie}^2(z)}dz = 0 \tag{4.12}$$

with the solution[2]:

$$n(x) = -\frac{N_b(x)}{2} + \sqrt{\left[\frac{N_b(x)}{2}\right]^2 + n_{ie}^2(x)e^{V_{BE}/V_T} - \frac{J_n n_{ie}^2(x)}{q}\int_0^x \frac{n(z)+N_b(z)}{D_n(z)n_{ie}^2(z)}dz} \tag{4.13}$$

Two points should be noted here. First, $n_{ie}(x)$ in (4.10) and (4.13) also depends on $n(x)$ because of plasma-induced bandgap narrowing in the case of high injection. This becomes clear when considering the different terms in ΔE_g. To account for the bandgap narrowing due to the Ge presence, the heavy doping effects, the high concentrations of carriers, the difference between $N_{c,Si}N_{v,Si}$ and $N_{c,SiGe}N_{v,SiGe}$, and the Fermi-level shift due to the difference between Fermi-Dirac statistics and Boltzmann statistics, we write ΔE_g in (4.5) as follows [205]:

$$\begin{aligned}\Delta E_g(x) =& \Delta E_{g,HDE} + \Delta E_{g,Ge} + \Delta E_{g,plasma} + kT\ ln\left(\frac{N_{c,SiGe}N_{v,SiGe}}{N_{c,Si}N_{v,Si}}\right) \\ &+ kT\left[ln\left(\frac{N_b}{N_{v,SiGe}}\right) - \mathcal{F}_{\frac{1}{2}}^{-1}\left(\frac{N_b}{N_{v,SiGe}}\right)\right]\end{aligned} \tag{4.14}$$

[2]The solution corresponding to the positive square root term is taken since $J_n\int_0^x \frac{n(x)+N_b(x)}{D_n(x)n_{ie}^2(x)}dx < q\ e^{V_{BE}/V_T}$ according to (4.9).

where $\mathcal{F}_{\frac{1}{2}}^{-1}$ is the inverse function of $\frac{2}{\sqrt{\pi}}\times$ the Fermi-Dirac integral of $\frac{1}{2}$ order (Please note that the details of the physical models used in this section to describe the electron transport in the SiGe base are given in Appendix C).

Based on Klaassen's new mobility model [121,125], the apparent bandgap narrowing, due to the heavy doping effects in Si after correcting for the Fermi-level shift associated with the use of Fermi-Dirac statistics, is extracted by Sokolic and Amon [206]:

$$\Delta E_{g,HDE}(x) = \left[(aN_b^c(x))^{-4} + (bN_b^d(x))^{-4}\right]^{-1/4} \quad [eV] \tag{4.15}$$

where a =6.76 $\times$ 10^{-11}, b =3.58 $\times$ 10^{-7}, $c = 0.5$, and $d = 0.28$.

Using the $\Delta E_{g,HDE}$ above and taking into account the Fermi-level shift associated with the use of Fermi-Dirac statistics, and the difference between the Si and SiGe N_cN_v products, Sokolic and Amon [206] extracted the bandgap narrowing due to the Ge presence as:

$$\Delta E_{g,Ge}(x) = 0.937y(x) - 0.5y^2(x) \quad [eV] \tag{4.16}$$

where $y(x)$ is the Ge fraction defined by (4.2).

For plasma-induced bandgap narrowing, we have fitted the experimental results of Neugroschel *et al.* [207] as follows:

$$\Delta E_{g,plasma}(x) = \begin{cases} 0, & n(x) < 2 \times 10^{16}\text{cm}^{-3} \\ a_2[ln(n(x))]^2 + a_1 ln(n(x)) + a_0, & 2 \times 10^{16} \leq n(x) \leq 10^{19}\text{cm}^{-3} \end{cases} \tag{4.17}$$

where $a_2 = 1.4898\ meV/cm^6, a_1 = -1.0346 \times 10^2\ meV/cm^3, a_0$ =1.7833 $\times$ $10^3\ meV$.

The ratio of the $N_c N_v$ product between Si and SiGe for $N_b < 10^{19}\text{cm}^{-3}$ calculated by Sokolic and Amon [205] can be fitted as a function of the Ge fraction:

$$\frac{N_{c,SiGe}N_{v,SiGe}(x)}{N_{c,Si}N_{v,Si}} = 0.07768 + \frac{0.9223}{1 + 78.02(y(x))^{1.213}} \tag{4.18}$$

for $y < 0.35$.

To calculate $N_{v,SiGe}$ for the Fermi-level shift, we assume $N_{c,SiGe} = \frac{2}{3}N_{c,Si}$ [206] and the values of $N_{c,Si}$ and $N_{v,Si}$ are taken from [71].

Second, $D_n(x)$ in (4.10) and (4.13) also depends on $n(x)$ and $n_{ie}(x)$ through the electric field, the carrier-carrier scattering, and the carrier-impurity scattering in the base. According to the Caughey-Thomas model [126][3], $D_n(x)$ is related to $\mathcal{E}(x)$ as follows:

$$D_n(x) = \frac{D_n^l(x)}{[1 + (|\mathcal{E}(x)|\ D_n^l(x)/(V_T v_s))^{\beta}]^{1/\beta}} \tag{4.19}$$

where $\beta = 2$. The electron diffusion coefficient at low electric field, $D_n^l(x)$, is a function of the base doping, carrier concentration, and the Ge fraction. It can be written as:

$$D_n^l(x) = D_{n,Si}^l(x) D_{n,rel}^l(x) \tag{4.20}$$

where $D_{n,rel}^l$ is the ratio of the electron diffusion coefficient in SiGe to that in Si, which is modelled by Decoutere *et al.* [209] as a function of the Ge fraction. Taking the carrier-carrier scattering and carrier-impurity scattering into account, Klaassen [121,125] has recently modelled $D_{n,Si}^l(x)$ as a function of $N_b(x), n(x), p(x)(= n(x) + N_b(x))$, and the donor concentration (which is set to the background collector con-

[3] In our implementation, Thornber's model [208] can also be chosen.

centration in our study)[4]. Because of the quasi-neutrality in the base, $\mathcal{E}(x)$ in (4.19) can be modified from (4.6) as:

$$\mathcal{E}(x) = V_T \frac{d}{dx}\left[ln\left(\frac{n(x)+N_b(x)}{n_{ie}^2(x)}\right)\right] \tag{4.21}$$

From (4.10) and (4.13), we can see that V_{BE} and $n(x)$ are interdependent, and $D_n(x)$ and $n_{ie}(x)$ must be known in order to obtain V_{BE} and $n(x)$. This, together with the fact that both $D_n(x)$ and $n_{ie}(x)$ depend on $n(x)$, as just noted, implies that an iterative method is needed to solve (4.10) and (4.13) for V_{BE} and $n(x)$. We have chosen to use the low-injection values as the initial solutions for the iteration. For low injection, $n(x) \ll N_b(x)$, i.e., $p(x) \sim N_b(x)$, (4.21), (4.10) and (4.13) become:

$$\mathcal{E}_L(x) = V_T \frac{d}{dx}\left[ln\left(\frac{N_b(x)}{n_{ie,L}^2(x)}\right)\right] \tag{4.22}$$

$$V_{BE,L} = V_T\ ln\left\{\frac{J_n}{q}\left[\int_0^{W_b}\frac{N_b(x)}{D_{n,L}(x)n_{ie,L}^2(x)}dx + \frac{N_b(W_b)}{v_s n_{ie,L}^2(W_b)} + \frac{J_n}{qv_s^2 n_{ie,L}^2(W_b)}\right]\right\} \tag{4.23}$$

$$n_L(x) = -\frac{N_b(x)}{2} + \sqrt{\left[\frac{N_b(x)}{2}\right]^2 + n_{ie,L}^2(x)e^{V_{BE,L}/V_T} - \frac{J_n n_{ie,L}^2(x)}{q}\int_0^x \frac{N_b(z)}{D_{n,L}(z)n_{ie,L}^2(z)}dz} \tag{4.24}$$

To determine $D_{n,L}(x)$ and $n_{ie,L}(x)$ in (4.23) and (4.24), we set $\Delta E_{g,plasma} = 0$

[4]We consider Klaassen's low-field mobility model most suitable for simulating bipolar transistors operating under high injection because it accounts for the carrier-related scattering. However, other low-field mobility models can also be chosen within our implementation.

in (4.14), substitute (4.22) in (4.19) and set $p(x) = N_b(x)$ in Klaassen's low-field mobility model.

In short, we can determine $n(x)$ iteratively for a given J_n as follows:

$$\begin{aligned}
n_{ie,1}(x) &\leftarrow F_{n_{ie}}[n_L(x)] \\
D_{n,1}(x) &\leftarrow F_D[n_L(x), n_{ie,1}(x)] \\
V_{BE,1} &\leftarrow F_V[n_L(x), D_{n,1}(x), n_{ie,1}(x)] \\
n_1(x) &\leftarrow F_n[n_L(x), D_{n,1}(x), V_{BE,1}, n_{ie,1}(x)] \\
n_{ie,2}(x) &\leftarrow F_{n_{ie}}[n_1(x)] \\
D_{n,2}(x) &\leftarrow F_D[n_1(x), n_{ie,2}(x)] \\
V_{BE,2} &\leftarrow F_V[n_1(x), D_{n,2}(x), n_{ie,2}(x)] \\
n_2(x) &\leftarrow F_n[n_1(x), D_{n,2}(x), V_{BE,2}, n_{ie,2}(x)] \\
&\cdots\cdots \\
n_{ie,j}(x) &\leftarrow F_{n_{ie}}[n_{j-1}(x)] \\
D_{n,j}(x) &\leftarrow F_D[n_{j-1}(x), n_{ie,j}(x)] \\
V_{BE,j} &\leftarrow F_V[n_{j-1}(x), D_{n,j}(x), n_{ie,j}(x)] \\
n_j(x) &\leftarrow F_n[n_{j-1}(x), D_{n,j}(x), V_{BE,j}, n_{ie,j}(x)]
\end{aligned} \tag{4.25}$$

where $F_{n_{ie}}, F_D, F_V, F_n$ are functions determined by (4.5), (4.10), (4.13)–(4.21) and the subscript j denotes the number of iteration.

4.3 Verification

The results generated by the iteration scheme and expressions described in Section 4.2 are compared with those reported in [89,185,193,210], which cover a wide range

of base region characteristics (base width, doping level, Ge concentration, profile shape) and injection level.

Fig. 4.2 and Fig. 4.3 show that the solutions, obtained from the expressions and iteration scheme described in Section 4.2, converge quickly. For the profiles used for Fig. 4.2–4.4, solutions converge at the 5th iteration. In Fig. 4.2, for the Si case (Suzuki's profiles), the V_{BE} solutions at the 3rd and the 5th iteration are identical over a wide range of J_n. For SiGe HBT (Rinaldi's profile), for $J_n < 10^5 A/cm^2$, solutions of V_{BE} at the 3rd and the 5th iteration are identical. For $J_n > 10^5 A/cm^2$, the error at 3rd iteration relative to the 5th iteration is less than 1%. In Fig. 4.3, no difference can be seen between the solution of the normalized electron concentration at the 3rd iteration and that at the 5th iteration. For the base transit time, the error at the 3rd iteration relative to the 5th iteration is less than 1% even at very high V_{BE} biases.

As shown in Figs. 4.2–4.4, our results agree well with those reported. Discrepancies are possibly due to the use of slightly different parameters for modelling heavy doping effects, the diffusion coefficient, and different $N_c N_v$ (in the case of SiGe) or N_{bc} values which are not clearly specified in the compared works. Whenever possible, the same models and parameters, and profile definitions specified in [89, 185, 193, 210] are used in the comparison. The effect of plasma-induced bandgap narrowing is not considered in the comparison as it was neglected in those works. In Fig. 4.4, for low injection, our result agrees well with Rinaldi's analytical expression. For high injection, our result (generated with a J_n of $10^8 A/cm^2$) approaches the limit set by Rinaldi's analytical expression [210] for *strong* high injection.

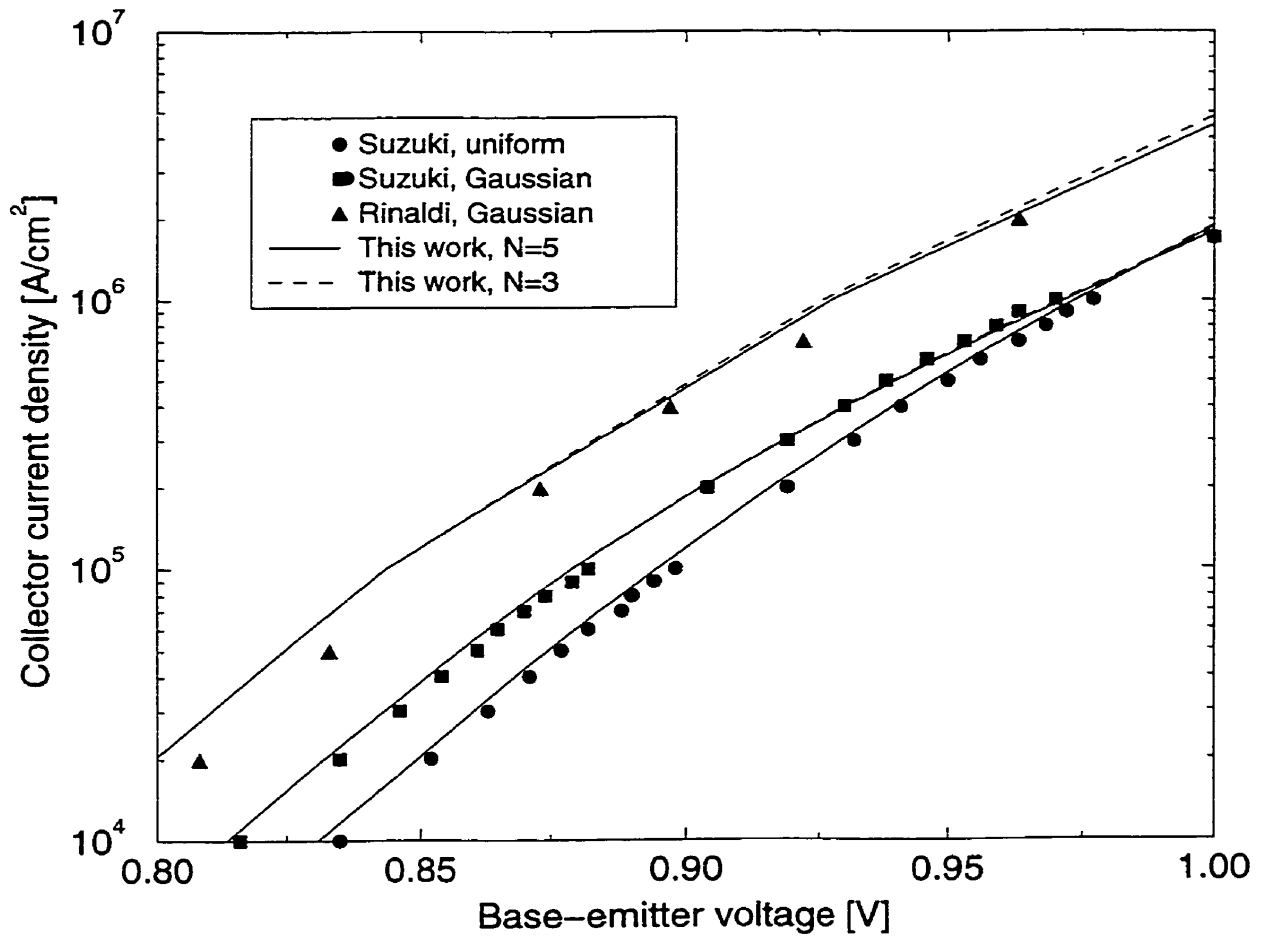

Figure 4.2: Collector current density vs. base-emitter voltage. Suzuki's uniform and Gaussian base: $W_b = 100\,nm, N_{be} = 2\times10^{18}\text{cm}^{-3}$. $N_{bc} = 2\times10^{16}\text{cm}^{-3}$ is used for Suzuki's Gaussian base [193]. Rinaldi's Gaussian base [210]: $W_b = 100\,nm, y(W_b) = 0.15$, triangular Ge profile, $N_{be} = 10^{18}\text{cm}^{-3}$, $N_{bc} = 5 \times 10^{16}\text{cm}^{-3}$. The variable N is the number of iteration.

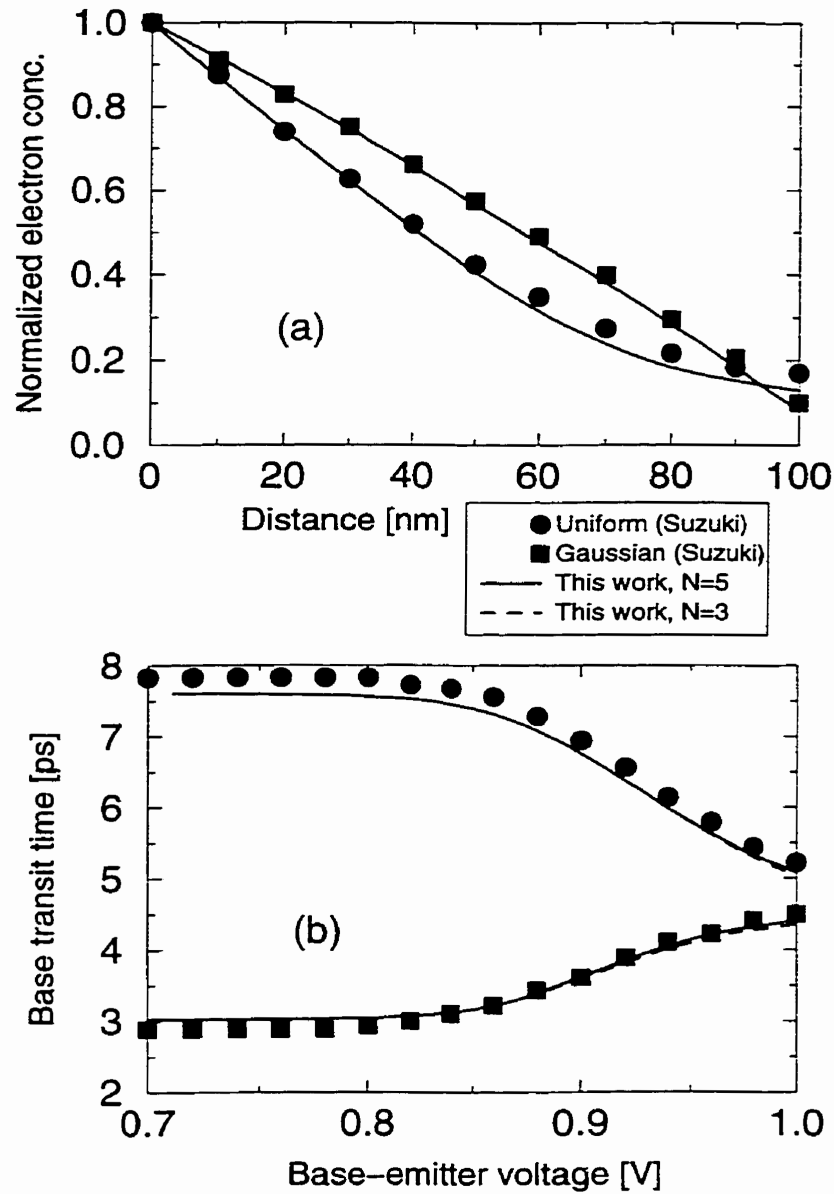

Figure 4.3: (a) Normalized injected electron concentration profiles at $V_{BE} = 0.9\ V$, (b) Base transit time vs. base-emitter voltage. For both (a) and (b), Suzuki's uniform and Gaussian bases [193]: $W_b = 100\ nm, N_{be} = 2 \times 10^{18}\text{cm}^{-3}$. $N_{bc} = 2 \times 10^{16}\text{cm}^{-3}$ is used for the Gaussian base.

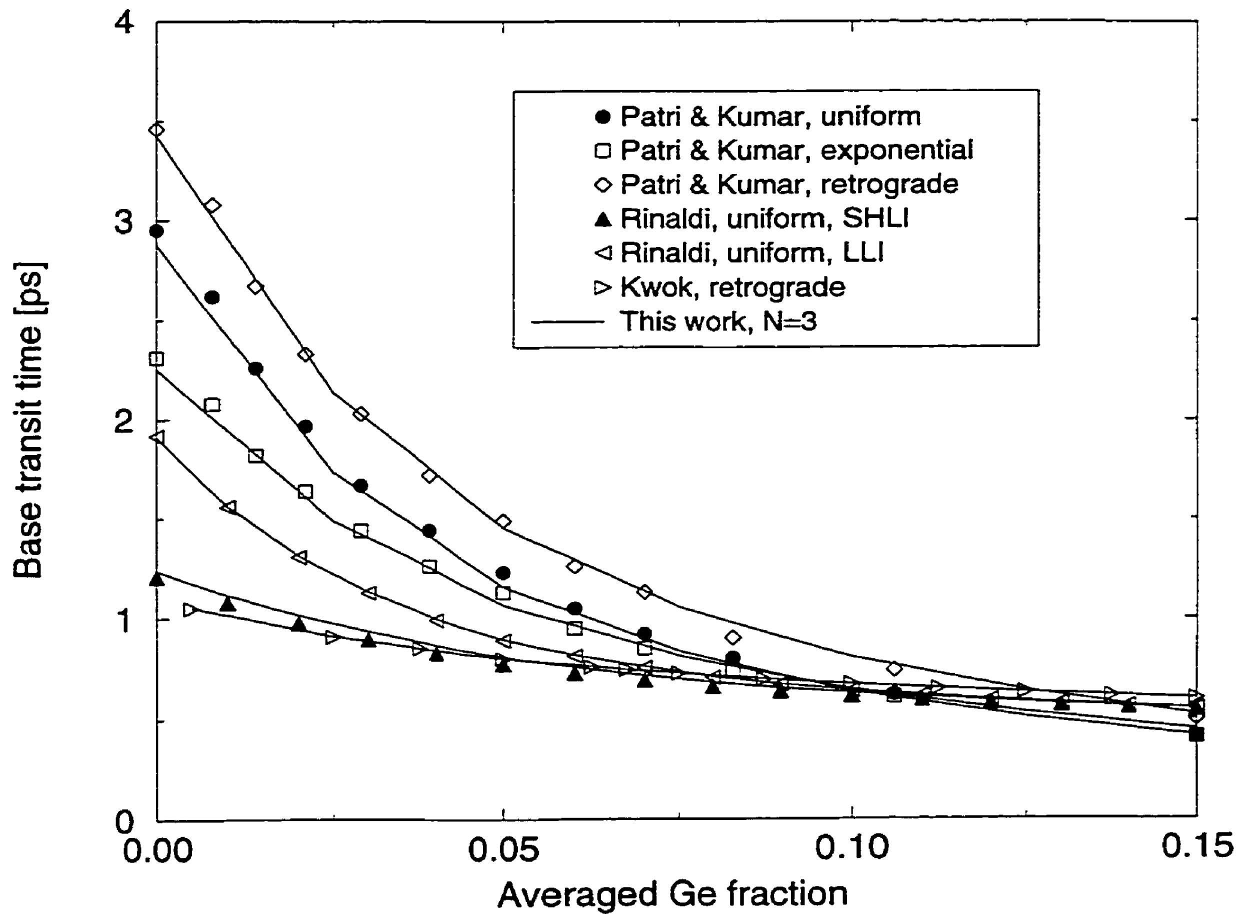

Figure 4.4: Base transit time vs. averaged Ge fraction (defined as $y_{\Delta c}/2$ for triangular Ge profile). Patri & Kumar's profiles [185]: $W_b = 60\ nm, R_b = 5k\Omega/\square$. Their results are generated under low injection without considering electric-field dependency of the diffusion coefficient. Rinaldi's uniform base [210]: $N_b = 10^{18}\text{cm}^{-3}$. The triangular symbols for Rinaldi's base are calculated using equations (3) and (5) in [210], which correspond to the low injection and strong high injection, respectively. Kwok's retrograde base [89]: $W_b = 38.5\ nm, x_1 = 6.5\ nm, x_2 = 11.5\ nm, N_{be} = 6.55 \times 10^{18}\text{cm}^{-3}, N_{bc} = 3 \times 10^{17}\text{cm}^{-3}$, $N_p = 9.4 \times 10^{18}\text{cm}^{-3}$ where x_1 and x_2 define the neutral base region with a constant base concentration of N_p. Kwok's results are produced under low injection.

4.4 Low injection

The following points can be noted from comparing the four doping profiles shown in Fig. 4.1. The doping profiles have the same intrinsic base sheet resistance[5] and base concentration near the emitter. The base transit time and the intrinsic base sheet resistance for each doping profile are plotted against different shapes of the Ge profile (i.e., different X_T/W_b values) in Fig. 4.5. The Ge dose remains unchanged for all doping profiles and shapes of the Ge profile. Results are generated at $J_n = 10^3$ A/cm^2 (for low injection) with the following effects considered: concentration and electric-field dependencies of the electron diffusion coefficient, non-uniform bandgap narrowing due to the Ge presence and the heavy doping effects, and velocity saturation near the collector. Hereafter, all results are obtained at the 5th iteration. Although not shown, we have found that the same results are obtained regardless of whether or not plasma-induced bandgap narrowing is considered. This is because the electron concentration at low injection is small relative to the base doping level.

Fig. 4.5 shows that the intrinsic base resistances for all doping profiles are around 5 $k\Omega/\square$ For all shapes of Ge profile considered (i.e. $0 \leqslant X_T/W_b \leqslant 1$), both the Gaussian and the exponential doping profiles result in higher τ_b than the uniform and the retrograde profiles due to their larger base widths. A larger base width simply offsets the enhancement from the aiding electric field caused by the doping gradient. The retrograde doping profile gives a smaller τ_b than the uniform profile because of its slightly smaller base width and the aiding field associated with

[5]The intrinsic base sheet resistance R_b is calculated as

$$R_b = 1/\int_0^{W_b} q\mu_p(x)[N_b(x) + n(x)]dx$$

where $\mu_p(x)$ is a function of $N_b(x)$ and $n(x)$ and can be obtained by Klaassen's unified mobility model [121]. The electron concentration $n(x)$ is set to zero in the case of low injection.

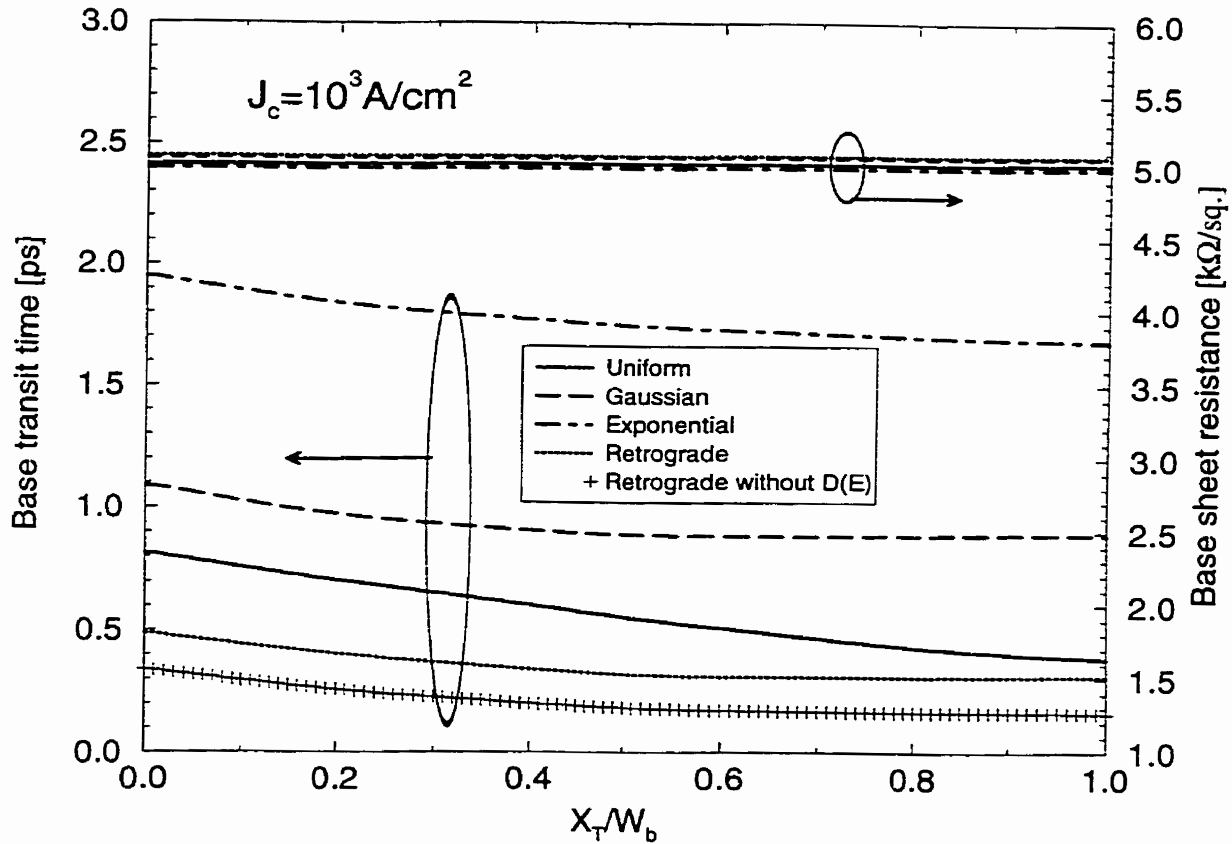

Figure 4.5: Base transit time and intrinsic base sheet resistance vs. X_T/W_b. J_n $=10^3 A/cm^2, R_b = 5k\Omega/\square$, $y_e = 0, N_{be} = 5 \times 10^{18}\text{cm}^{-3}$, $N_{bc} = 9 \times 10^{16}\text{cm}^{-3}$. Uniform doping profile (W_b = 29 $nm, y_{\Delta c}$ =0.1724), Gaussian (W_b = 54 $nm, y_{\Delta c}$ =0.0926), exponential (W_b = 83 $nm, y_{\Delta c}$ =0.0602), retrograde (W_b = 26 $nm, y_{\Delta c}$ = 0.1923, R_p = 13 nm, N_p =10^{19}cm^{-3}).

its tail region.

Regarding the optimal Ge profile, Fig. 4.5 suggests that for the retrograde doping profile, graded Ge profiles with $X_T/W_b \geqslant 0.5$ (i.e. trapezoidal or triangular) achieve the minimum or near minimum τ_b. To be precise, the X_T/W_b value for the minimum τ_b is 0.68. Although not shown, it is found that the optimum X_T/W_b value is always less than one for the retrograde profile considered here when R_p is less than W_b. However, if the electric-field dependency of the diffusion coefficient, $D(E)$, is ignored (denoted by the plus symbols in Fig. 4.5), a false optimum point will appear at $X_T/W_b = 1$ regardless of the location of the peak concentration. Therefore, the electric field dependency of the diffusion coefficient should be considered when determining the optimum point.

In Fig. 4.5, we set $y_{\triangle c}W_b/2$ (a measure of the Ge dose) to 2.5 (i.e., the Ge dose is equivalent to a triangular Ge profile with W_b of 50 nm and a Ge fraction of 0.1 near the collector). Whereas in Fig. 4.6, the base transit time is plotted against the Ge dose of a trapezoidal Ge profile with $X_T/W_b = 0.68$ for different doping profiles. It shows that the retrograde doping profile always gives a smaller τ_b than the uniform profile for all Ge doses. Also, a Ge dose ($y_{\triangle c}W_b/2$) of 2.5 is sufficient to reach the near minimum τ_b. Additional Ge doses will not reduce τ_b effectively.

4.5 High injection

The four doping profiles in Fig. 4.1 are again compared at high-injection levels. Fig. 4.7 (Fig. 4.8) plots the base transit time and base sheet resistance of a triangular Ge HBT against the collector current density for the four doping profiles without (with) plasma-induced BGN. The Ge doses are identical for all doping profiles. It should be noted that our discussions are limited up to the Kirk onset current

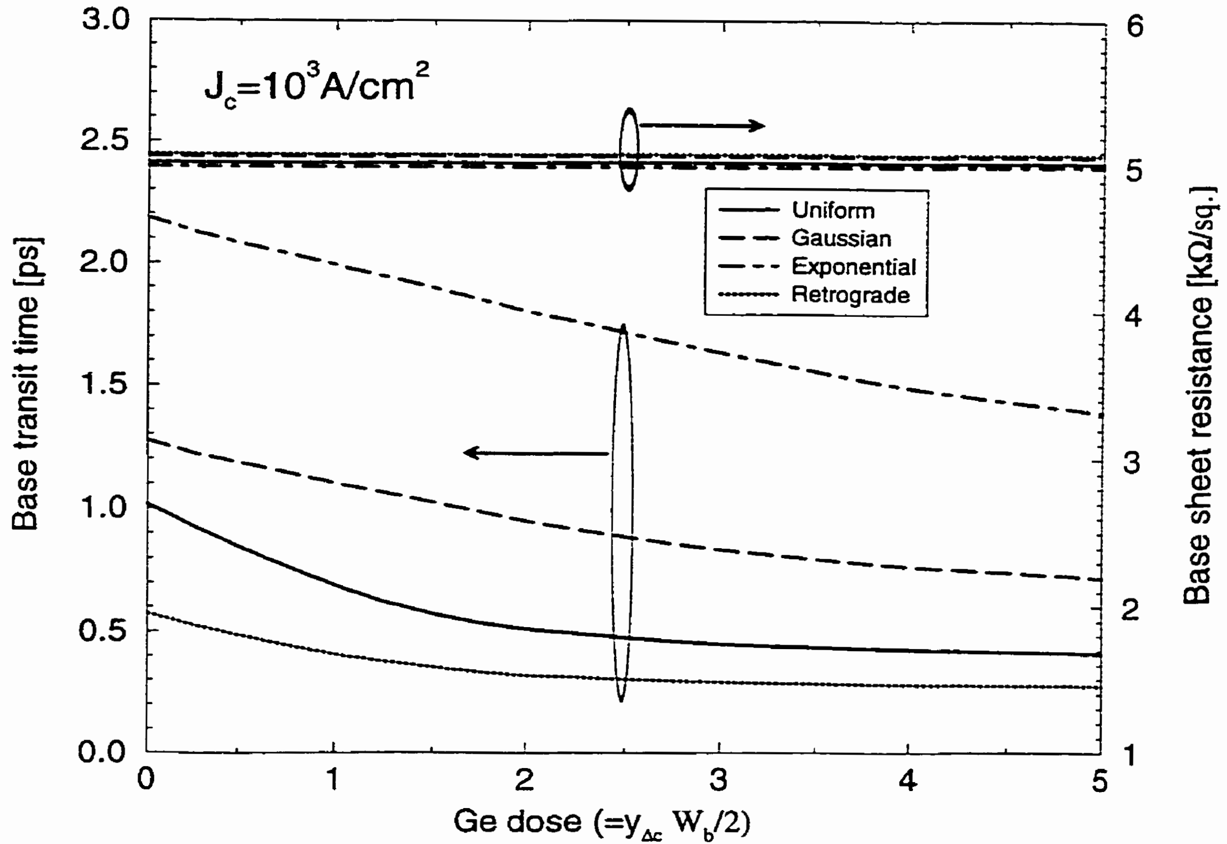

Figure 4.6: Base transit time and intrinsic base sheet resistance vs. Ge dose of a trapezoidal Ge profile with $X_T/W_b = 0.68$ for different doping profiles. $J_n = 10^3 A/cm^2, R_b = 5k\Omega/\square$, $y_e = 0, N_{be} = 5 \times 10^{18}\text{cm}^{-3}$, $N_{bc} = 9 \times 10^{16}\text{cm}^{-3}$. Uniform ($W_b = 29\ nm$), Gaussian ($W_b = 54\ nm$), exponential ($W_b = 83\ nm$), retrograde ($W_b = 26\ nm, R_p = 13\ nm, N_p = 10^{19}\text{cm}^{-3}$).

density (around $1.4 \times 10^5\ A/cm^2$).

A comparison of Fig. 4.7 and Fig. 4.8 shows that the effect of plasma-induced BGN is not significant for the triangular Ge profile considered. The base transit times for all doping profiles are not strong functions of the collector current density (at least within the range under discussion). The base transit time for the uniform doping profile is slightly higher than that for the retrograde profile. Both are only around 40% and 20% of those of the Gaussian and the exponential profiles, respectively. R_b begins to drop for $J_n > 10^4 A/cm^2$ (as the excess carrier concentration increases with J_n). At $J_n = 1.4 \times 10^5 A/cm^2$, R_b of the retrograde and the uniform profiles are around 5% and 20% larger than those of the Gaussian and exponential profiles, respectively. Therefore, the retrograde profile is still preferred to the Gaussian doping profile. For applications where the maximum oscillation frequency (which is affected by R_b) is a more important figure of merit, an in-depth trade-off study is required to compare the effects of the retrograde and the exponential doping profiles.

Fig. 4.9 examines the effect of the Ge profile shape on the base transit time and base resistance for different doping profiles at $J_n = 10^5 A/cm^2$, without considering plasma-induced BGN. The base transit time for the retrograde doping profile is smaller than that for the uniform profile for all Ge profile shapes considered. Since the base resistance for each doping profile is sufficiently constant over different X_T/W_b values, it is still meaningful to compare different Ge profile shapes. For the retrograde profile, graded Ge profiles with $X_T/W_b \geqslant 0.5$ give the minimum (or near minimum) τ_b. To be precise, the X_T/W_b value for the minimum τ_b is 0.7.

With plasma-induced bandgap narrowing taken into account (see Fig. 4.10), the base transit time for the retrograde doping profile is again the lowest for all Ge profile shapes considered. Graded Ge profiles with $X_T/W_b \geqslant 0.5$ are found to

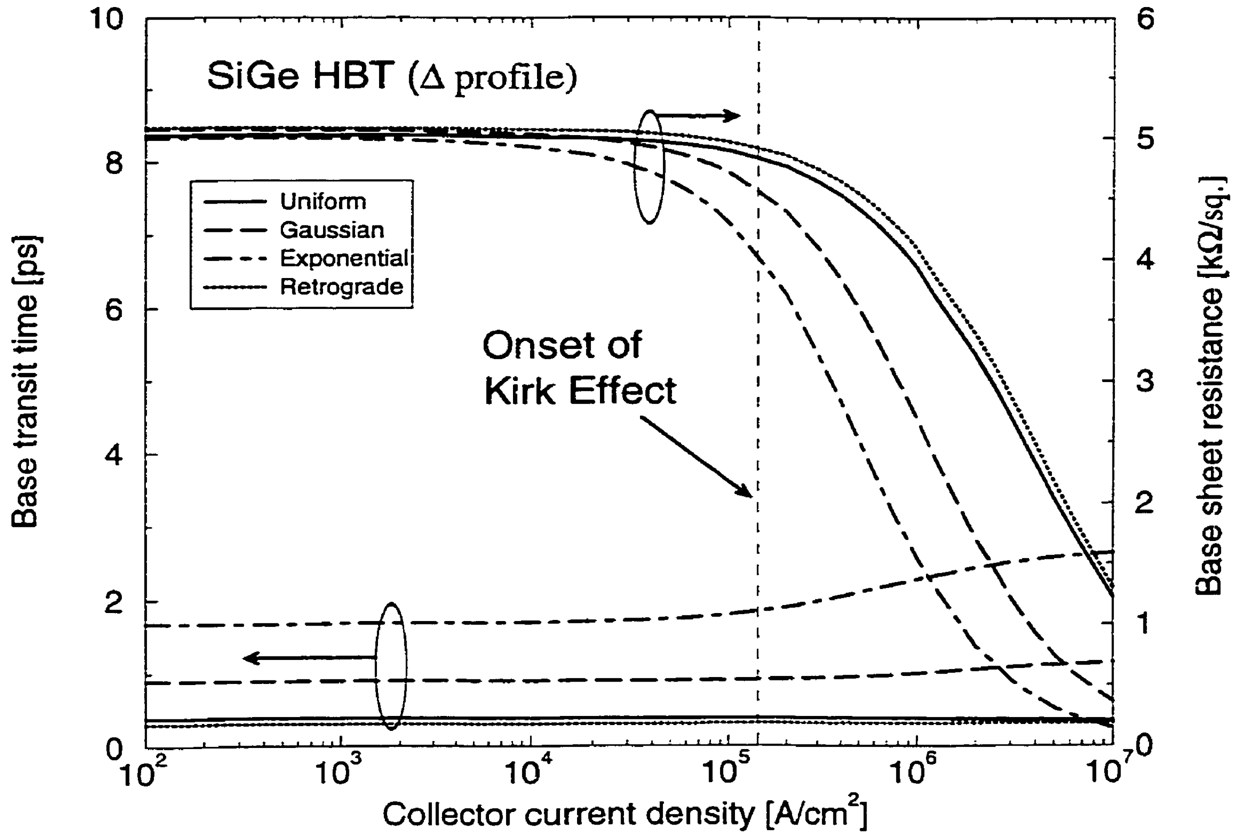

Figure 4.7: Base transit time and intrinsic base sheet resistance vs. collector current density for a triangular Ge profile with different doping profiles. $R_b = 5k\Omega/\Box$, $y_e = 0, N_{be} = 5 \times 10^{18}\text{cm}^{-3}$, $N_{bc} = 9 \times 10^{16}\text{cm}^{-3}$. Uniform doping profile ($W_b = 29\ nm, y_{\Delta c} = 0.1724$), Gaussian ($W_b = 54\ nm, y_{\Delta c} = 0.0926$), exponential ($W_b = 83\ nm, y_{\Delta c} = 0.0602$), retrograde ($W_b = 26\ nm, y_{\Delta c} = 0.1923, R_p = 13\ nm, N_p = 10^{19}\text{cm}^{-3}$).

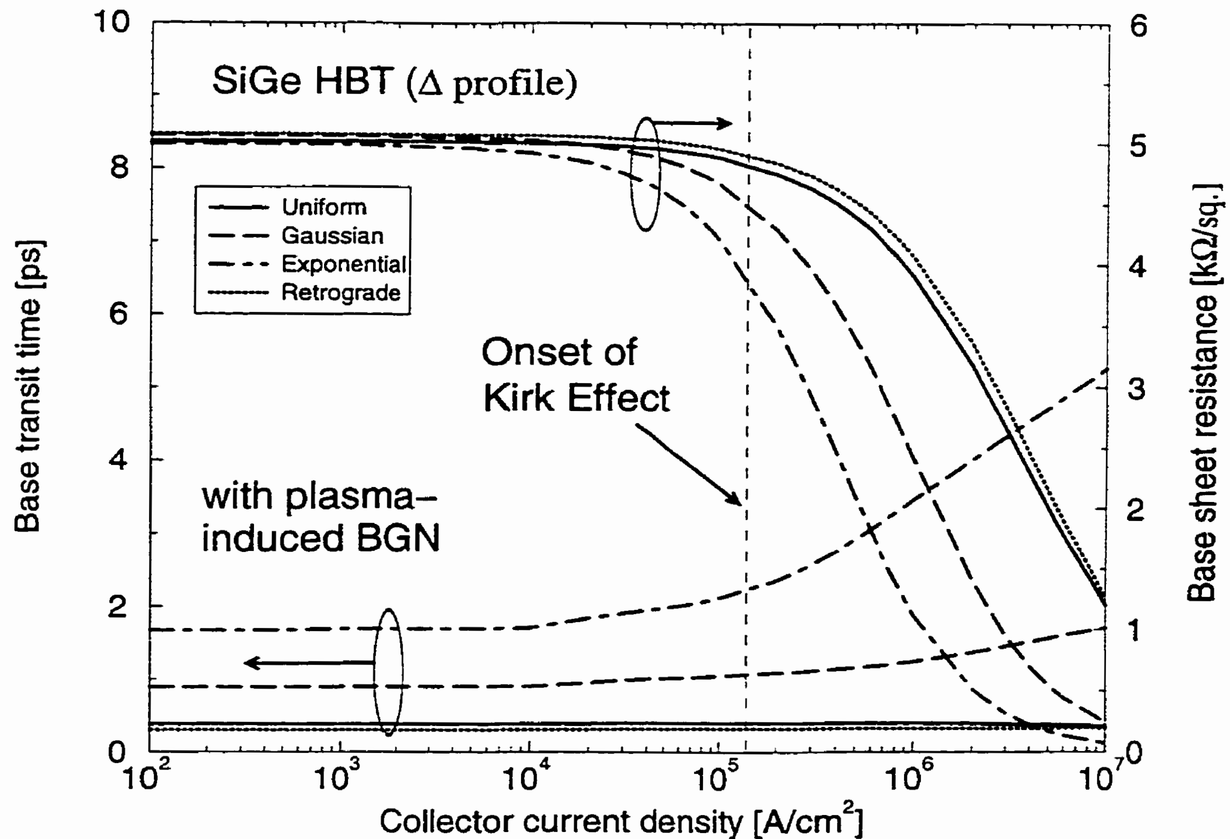

Figure 4.8: Base transit time and intrinsic base sheet resistance vs. collector current density for a triangular Ge profile with different doping profiles. Plasma-induced bandgap narrowing is considered. $R_b = 5k\Omega/\Box$, $y_e = 0$, N_{be} =5×10^{18}cm^{-3}, N_{bc} =9×10^{16}cm^{-3}. Uniform doping profile ($W_b = 29\ nm, y_{\Delta c}$ =0.1724), Gaussian ($W_b = 54\ nm, y_{\Delta c}$ =0.0926), exponential ($W_b = 83\ nm, y_{\Delta c}$ =0.0602), retrograde ($W_b = 26\ nm, y_{\Delta c} = 0.1923, R_p = 13\ nm, N_p$ =10^{19}cm^{-3}).

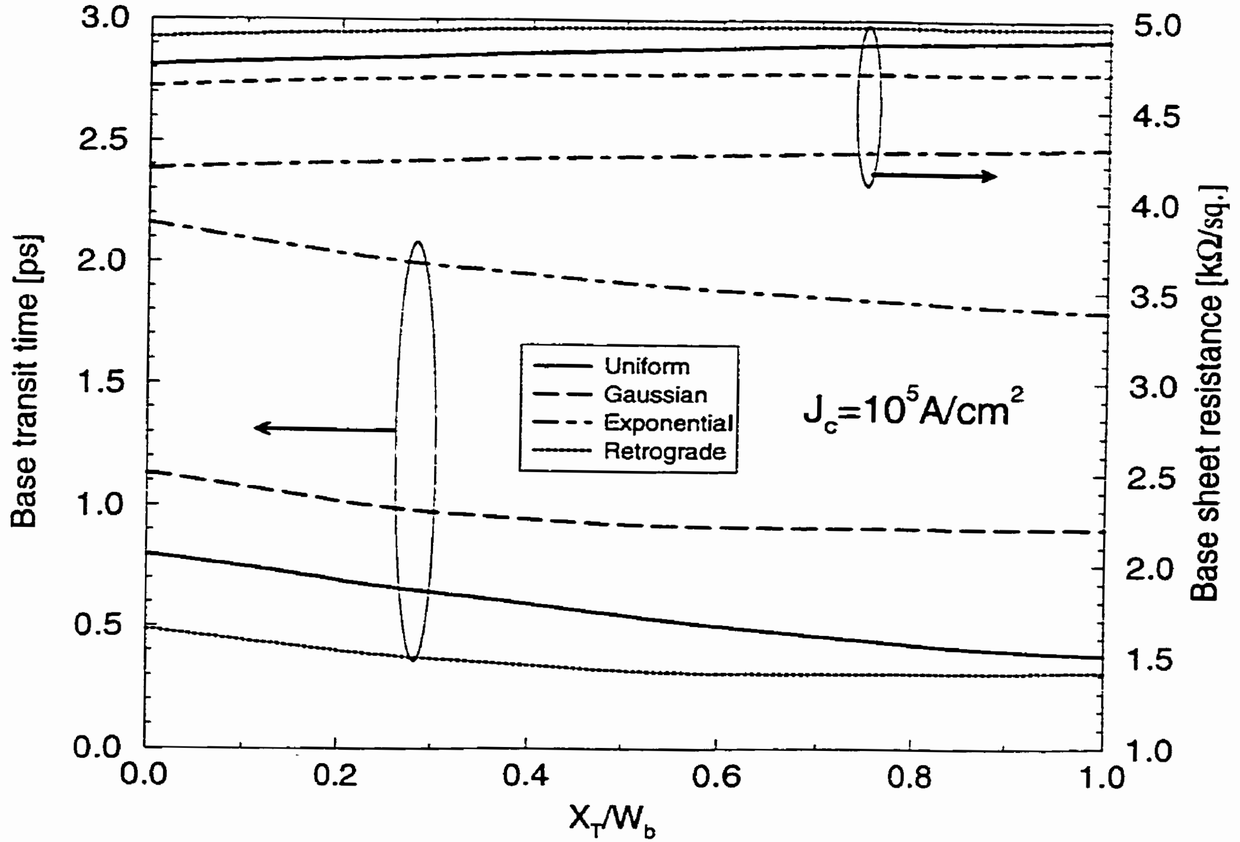

Figure 4.9: Base transit time and intrinsic base sheet resistance vs. X_T/W_b. J_n $=10^5 A/cm^2, R_b = 5k\Omega/\square$, $y_e = 0, N_{be} = 5 \times 10^{18}$cm^{-3}, $N_{bc} = 9 \times 10^{16}$cm^{-3}. Uniform doping profile ($W_b = 29\ nm, y_{\Delta c} = 0.1724$), Gaussian ($W_b = 54\ nm, y_{\Delta c} = 0.0926$), exponential ($W_b = 83\ nm, y_{\Delta c} = 0.0602$), retrograde ($W_b = 26\ nm, y_{\Delta c} = 0.1923, R_p = 13\ nm, N_p = 10^{19}$cm^{-3}).

be optimal for the retrograde doping profile. The exact minimum τ_b point occurs at $X_T/W_b = 0.67$. Compared with Fig. 4.9, the base transit times are higher for all doping profiles. The increase is more significant for Ge profiles with a box shape (i.e. small X_T/W_b values) than for those with a triangular shape (i.e. large X_T/W_b values). This can be explained by the fact that the large Ge-induced aiding electric field associated with the triangular Ge profile dominates the effect of the retarding electric field caused by plasma-induced bandgap narrowing. It should be noted that a higher current density associated with an increase in the base-emitter voltage results in a larger negative electron concentration gradient in the neutral base. As predicted by (4.4), (4.14), and (4.17), a larger negative electron concentration gradient creates a higher positive (i.e. retarding) electric field related to plasma-induced bandgap narrowing.

Figs. 4.11 and 4.12 assess the relative importance of different effects in determining τ_b as a function of the Ge profile shape and the Ge dose, respectively, for the retrograde doping profile. The relative importance of the effect 'X' is measured by the τ_b ratio $\frac{\tau_b(\text{effect 'X' is neglected})}{\tau_b(\text{all effects considered})}$. The closer this ratio to unity, the less important the effect 'X' is. Fig. 4.11 shows that the effect of plasma-induced BGN is more significant when the shape of the Ge profile is box-like rather than triangle-like, which is consistent with the difference between Fig. 4.9 and Fig. 4.10. Fig. 4.12 indicates that the effect of plasma-induced BGN is very significant in the case of a Si BJT (i.e. zero Ge dose) and becomes less prominent when the Ge fraction near the collector of a triangular Ge HBT increases. A Ge-induced electric field associated with a Ge fraction of 0.2 near the collector can reduce the adverse effect of plasma-induced BGN to a τ_b difference of 5%. Furthermore, both Figs. 4.11 and 4.12 confirm the importance of the electric-field dependency of D_n in the presence of the Ge-induced electric field.

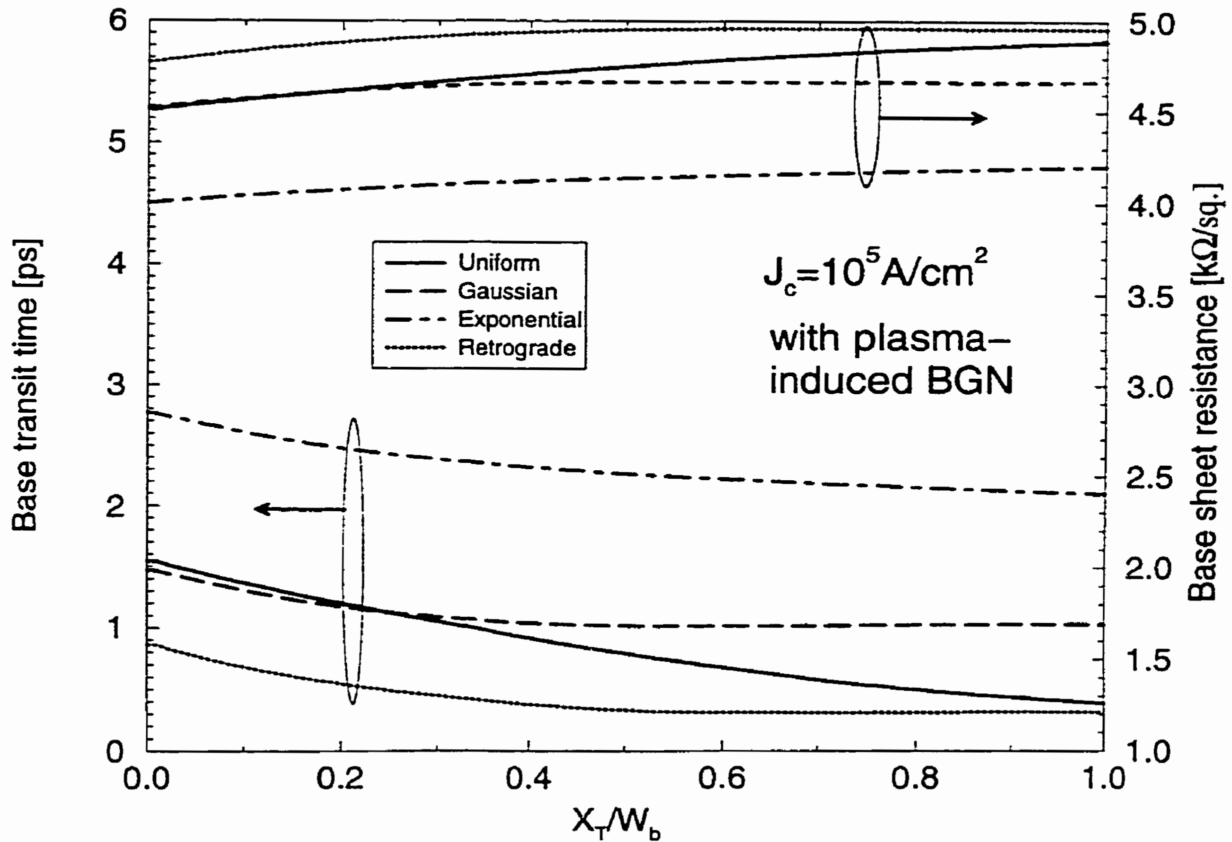

Figure 4.10: Base transit time and intrinsic base sheet resistance vs. X_T/W_b. Plasma-induced bandgap narrowing is considered. J_n =$10^5 A/cm^2, R_b = 5k\Omega/\Box$, $y_e = 0, N_{be}$ =5 × 10^{18}cm^{-3}, N_{bc} =9 × 10^{16}cm^{-3}. Uniform doping profile (W_b = 29 $nm, y_{\Delta c}$ =0.1724), Gaussian (W_b = 54 $nm, y_{\Delta c}$ =0.0926), exponential (W_b = 83 $nm, y_{\Delta c}$ =0.0602), retrograde (W_b = 26 $nm, y_{\Delta c}$ = 0.1923, R_p = 13 nm, N_p =10^{19}cm^{-3}).

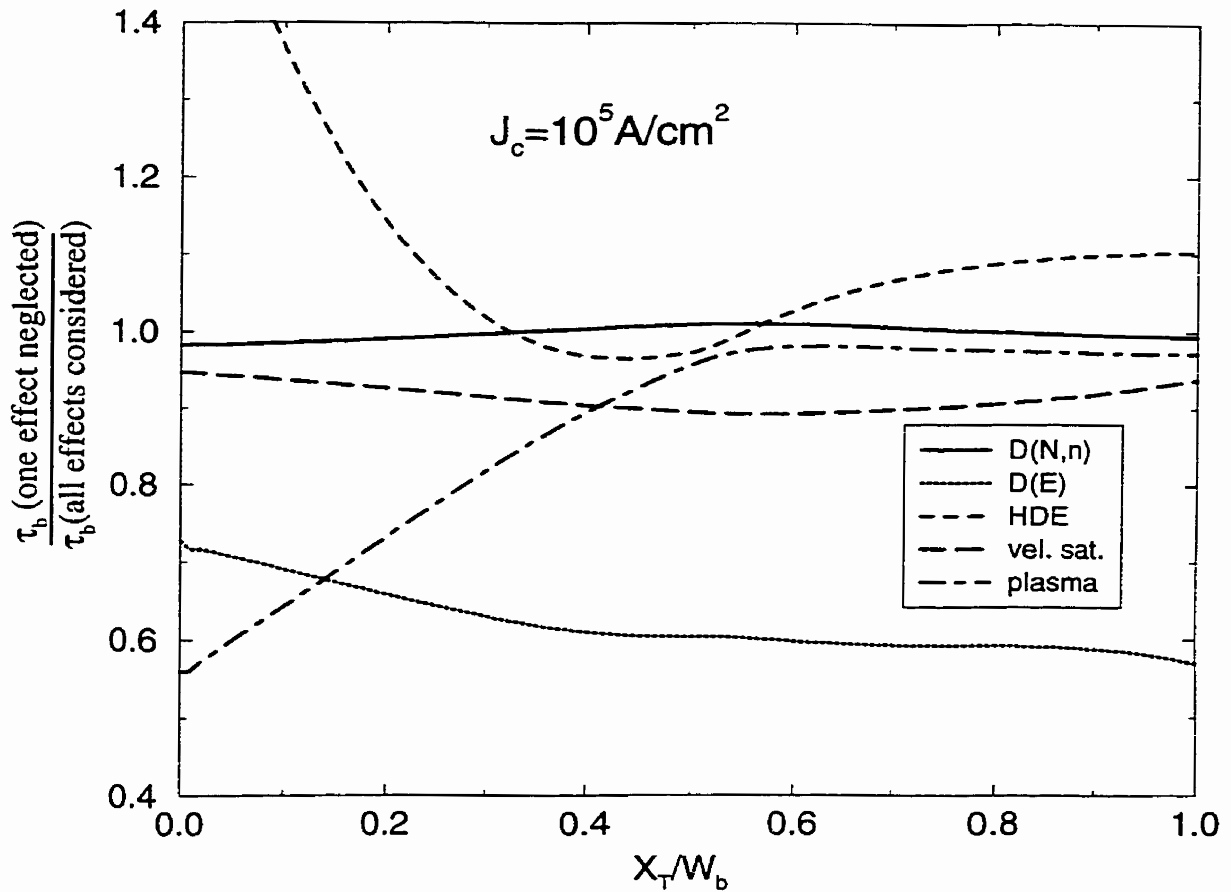

Figure 4.11: Ratio of base transit time with one effect neglected to base transit time with all effects considered vs. X_T/W_b for a retrograde doping profile in a SiGe base. $J_n = 10^5 A/cm^2$, $y_e = 0, y_{\Delta c} = 0.1923$, $N_{be} = 5 \times 10^{18}\text{cm}^{-3}$, $N_{bc} = 9 \times 10^{16}\text{cm}^{-3}$, $N_p = 10^{19}\text{cm}^{-3}$, $W_b = 26\ nm$, $R_p = 13\ nm$.

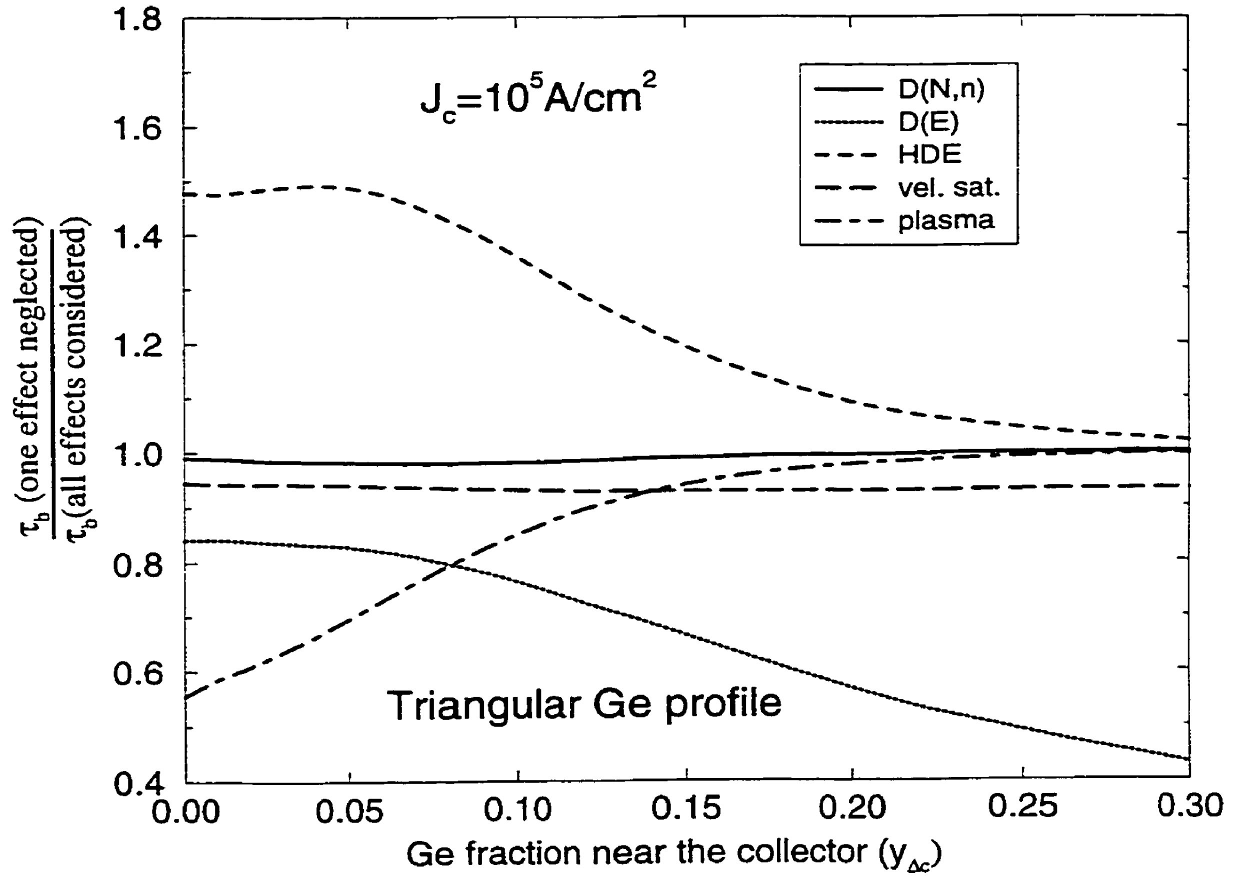

Figure 4.12: Ratio of base transit time with one effect neglected to base transit time with all effects considered vs. Ge fraction at the collector edge of a triangular Ge profile for a retrograde doping profile. J_n =$10^5 A/cm^2$, $y_e = 0$, N_{be} =5×10^{18}cm^{-3}, N_{bc} =9×10^{16}cm^{-3}, N_p =10^{19}cm^{-3},$W_b = 26\ nm$, $R_p = 13\ nm$.

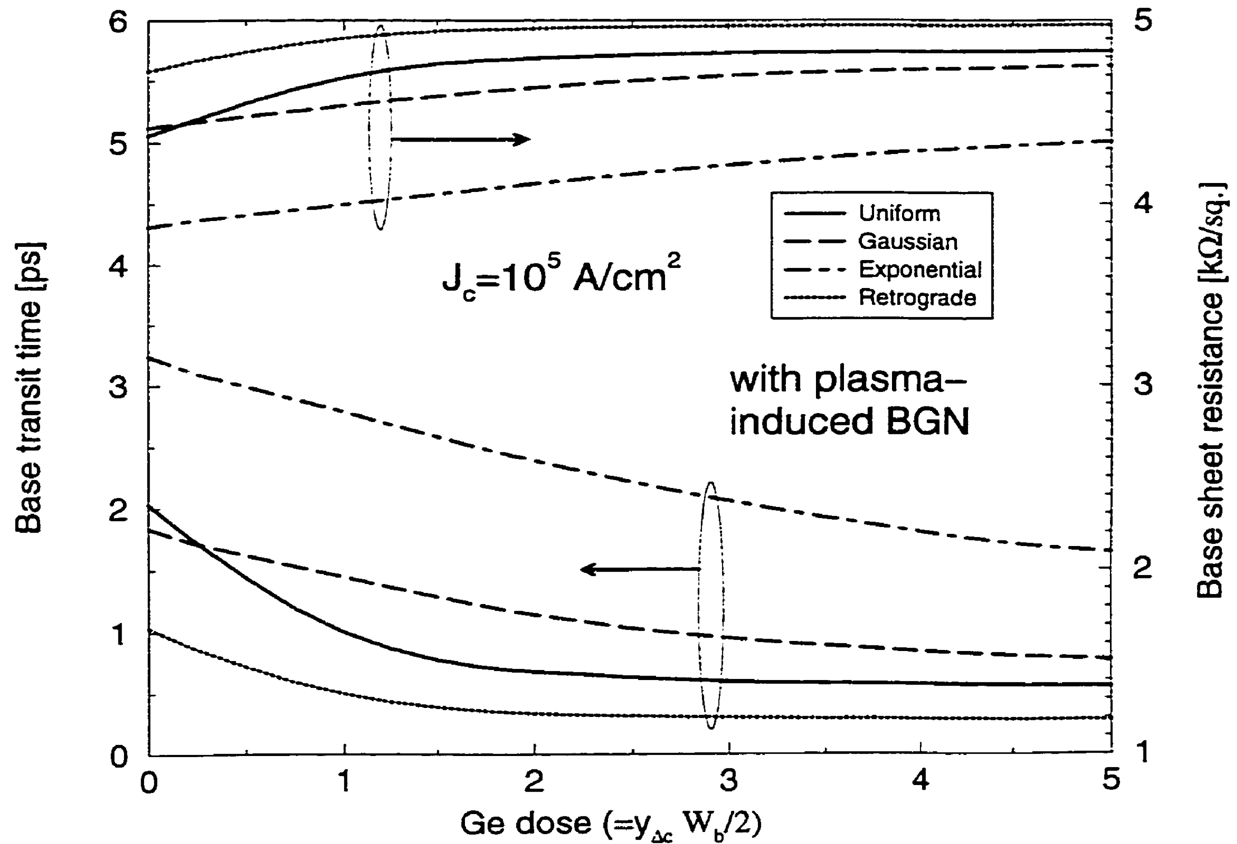

Figure 4.13: Base transit time and intrinsic base sheet resistance vs. Ge dose of a trapezoidal Ge profile with X_T/W_b = 0.67 for different doping profiles. Plasma-induced bandgap narrowing is considered. $J_n = 10^5 A/cm^2, R_b = 5k\Omega/\Box$, $y_e = 0, N_{be} = 5 \times 10^{18}\text{cm}^{-3}$, $N_{bc} = 9 \times 10^{16}\text{cm}^{-3}$. Uniform ($W_b$ = 29 nm), Gaussian (W_b = 54 nm), exponential (W_b = 83 nm), retrograde ($W_b = 26\ nm, R_p = 13\ nm, N_p = 10^{19}\text{cm}^{-3}$).

Figs. 4.13 shows how the base transit times of different doping profiles for a trapezoidal Ge profile ($X_T/W_b = 0.67$) vary with the Ge dose. The effect of plasma-induced BGN is considered. Just as in the case of low injection, a Ge dose ($y_{\Delta c}W_b/2$) of 2.5 is sufficient to reach the near minimum τ_b.

4.6 Conclusions

An iteration scheme to calculate τ_b for a given collector current density instead of a base-emitter voltage has been developed in order to determine the optimal doping profile and Ge profile in the neutral base for minimizing τ_b of SiGe HBTs under all levels of injection before the onset of the Kirk effect. We have adopted a consistent set of SiGe physical models with parameters tuned to measurement data, and included important effects such as the electric-field dependency of the diffusion coefficient and plasma-induced bandgap narrowing in our study for the first time. The iteration scheme has been verified by comparing its results with simulation results reported in the literature. Using this iteration scheme in both low- and high-injection regimes, we have compared, for the first time, base transit times for a range of doping and Ge profiles with identical Ge doses, intrinsic base sheet resistances, and base concentrations near the emitter. Based on the results in Sections 4.4 and 4.5, we draw the following conclusions:

1. For the triangular Ge profile and the retrograde or uniform doping profile considered, the base transit times in the low-injection regime do not differ much from those in the high-injection regime before the onset of Kirk effect.

2. For the given Ge dose, Ge profile shape, N_{be}, and R_b, the retrograde rather than the uniform doping profile gives the minimum τ_b.

3. For the given Ge dose, N_{be}, and R_b, graded Ge profiles with $X_T/W_b \geqslant 0.5$ achieve the minimum (or near minimum) τ_b for a retrograde doping profile.

4. For the retrograde doping profile and the trapezoidal Ge profile with $X_T/W_b \sim 0.67$ considered, Ge doses higher than $y_{\triangle c}W_b/2 = 2.5$ will not reduce τ_b effectively.

5. The effect of the electric-field dependency of D_n is significant in the presence of a Ge-induced electric field and must be considered when determining the optimal Ge profile.

6. In the high-injection regime, the effect of plasma-induced BGN tends to increase the base transit time. For a Ge dose $y_{\triangle c}W_b/2 = 2.5$, plasma-induced BGN can increase the base transit time by $\sim 70\%$ for the box Ge profile with the retrograde doping profile.

7. Whether one should take into consideration the effect of plasma-induced BGN depends on the magnitude of the aiding Ge-induced electric field, which acts against the retarding field associated with plasma-induced BGN. For the retrograde doping profile in a HBT with the triangular Ge profile considered, the effect should be accounted for if the Ge fraction is less than 0.15 near the collector. The effect should also be considered if the Ge profile is box-shaped.

Chapter 5

Fabrication of SiGe HBTs by high-dose Ge implantation and solid-phase epitaxy with Si amorphization

5.1 Introduction

The emerging SiGe technology has been shown to greatly improve the performance of Si devices in high-performance circuits [211]. However, the success of this technology will depend on its ability to be integrated into the existing Si production line. At present, device-quality SiGe growth is mainly realized by epitaxial deposition techniques such as molecular beam epitaxy (MBE) [35], rapid thermal chemical vapor deposition (RTCVD) or limited reaction processing (LRP) [45], atmospheric pressure chemical vapor deposition (APCVD) [46], low pressure chemical

vapor deposition (LPCVD) [212], and ultra-high vacuum chemical vapor deposition (UHV/CVD) [42]. Most of these techniques involve the introduction of a new processing equipment in the production line. Some have low throughput (e.g. MBE). Issues relating to integration, reliability, and downscaling are yet to be fully explored.

Alternatively, the technique of high-dose[1] Ge implantation with solid-phase epitaxy offers many attractive advantages.[2] This technique easily allows multiple selective growth of SiGe regions and is fully compatible with the existing Si facilities. Several attempts [214–222] have been made to fabricate SiGe-base HBTs by this new technique. Fukami *et al.* [214] managed to show transistor action but no heterojunction characteristics could be demonstrated. Both surface defects caused by the high-dose Ge implantation and end-of-range extended defects at the original amorphous/crystalline interface were observed. Based on the work of Fukami *et al.*, Gupta *et al.* [215] improved the current gain of the HBT up to 110 (versus 100 for the Si BJT). However, the current gain enhancement was caused more by the narrower base width (80 nm versus 130 nm in Si) than by the 7 at.% Ge presence, i.e., no heterojunction characteristics could be confirmed. Just like the case of Fukami *et al.*, surface defects and end-of-range extended defects were observed. Ma [216] successfully demonstrated the heterojunction effect of the base-collector junction but did not observe any transistor action. Dopant profile obtained by Secondary Ion Mass Spectrometry (SIMS) showed that the arsenic profile overdiffused into the base, resulting in a large emitter junction depth too close to the end-of-range extended defects near the original amorphous/crystalline interface.

---

[1] In this work, the phrase "high-dose" refers to an implantation resulting in a peak concentration of at least 5 atomic % of the Si substrate, i.e. $2.5 \times 10^{21}\mathrm{cm}^{-3}$.

[2] A review of this technique can be found in [213].

Better electrical characteristics were then demonstrated by Lombardo *et al.* [217–219]. Their SiGe HBT achieved a current gain larger than the Si BJT by a factor of 2.67. However, it is still questionable as to the origin of the enhancement because it was also shown that the base width of the SiGe HBT is smaller than that of the Si BJT by a factor of two. Even if the current gain did arise from the heterojunction effect, it is clear that the Ge leverage was not fully exploited. More recently, the enhancement of the collector current was reported by the same group [220, 221]. Again, since base width reduction due to the Ge presence was also shown, it became difficult to confirm the heterojunction characteristics. Although a bandgap narrowing of 27 meV was estimated by a temperature-dependent current measurement, both the temperature dependency of the electron mobility and the effect of dopant ionization were not considered when interpreting the measured data. This renders the claim of the heterojunction effect questionable. Using the technique of Ge implantation, Mitchell *et al.* [222] fabricated SiGe HBTs of both n-p-n and p-n-p types. Transistor action was shown only by the p-n-p SiGe HBT and no heterojunction characteristics shown by either type. In fact, a collector-emitter leakage was observed in the n-p-n SiGe HBT due to enhanced arsenic diffusion. Also, hairpin dislocations extending from the original amorphous/crystalline interface to the surface were found in the regrown SiGe region.

Since it has been shown that extended defects near the device junctions can adversely affect the device performance [223, 224], it is beneficial to use a high-energy Si amorphization step to push the original amorphous/crystalline interface (created during the high-dose Ge implantation) and the associated end-of-range (EOR) extended defects farther away from the base-collector junction [225, 226]. However, no studies, using the high-dose Ge implantation followed by the solid-phase epitaxy *and* a high-energy Si amorphization to realize SiGe HBTs, have been

reported.

In this chapter, a brief report of the electrical characteristics of Si BJTs, SiGe HBTs, and SiGe HBTs with Si amorphization fabricated by the high-dose Ge implantation and solid-phase epitaxy is presented. Due to limited resources, this aspect of the research could not be pursued to its logical end; however, certain original experimental results are presented here. In particular, transistor action of the SiGe HBT with Si amorphization is reported for the first time.

5.2 Goal of the experiment

The goal of this experiment was threefold. First, to demonstrate the feasibility of fabricating a SiGe HBT without introducing any new processing equipment. Second, to demonstrate heterojunction effects of the SiGe HBTs fabricated by high-dose Ge implantation and solid-phase epitaxy, with or without Si amorphization. Third, to see if the Si amorphization step can improve the D.C. characteristics of SiGe HBTs fabricated by high-dose Ge implantation and solid-phase epitaxy only.

In order to fairly compare the three types of transistors (Si BJTs, SiGe HBTs, SiGe HBTs with Si amorphization), they were fabricated side-by-side in the same dice to minimize the effects of possible processing deviations. Figure 5.1 shows the schematic of a cross sectional view of the three types of transistors. Figure 5.2 shows how the Ge implantation, Si amorphization, and solid-phase epitaxy steps are incorporated into a polyemitter Si BJT fabrication process.

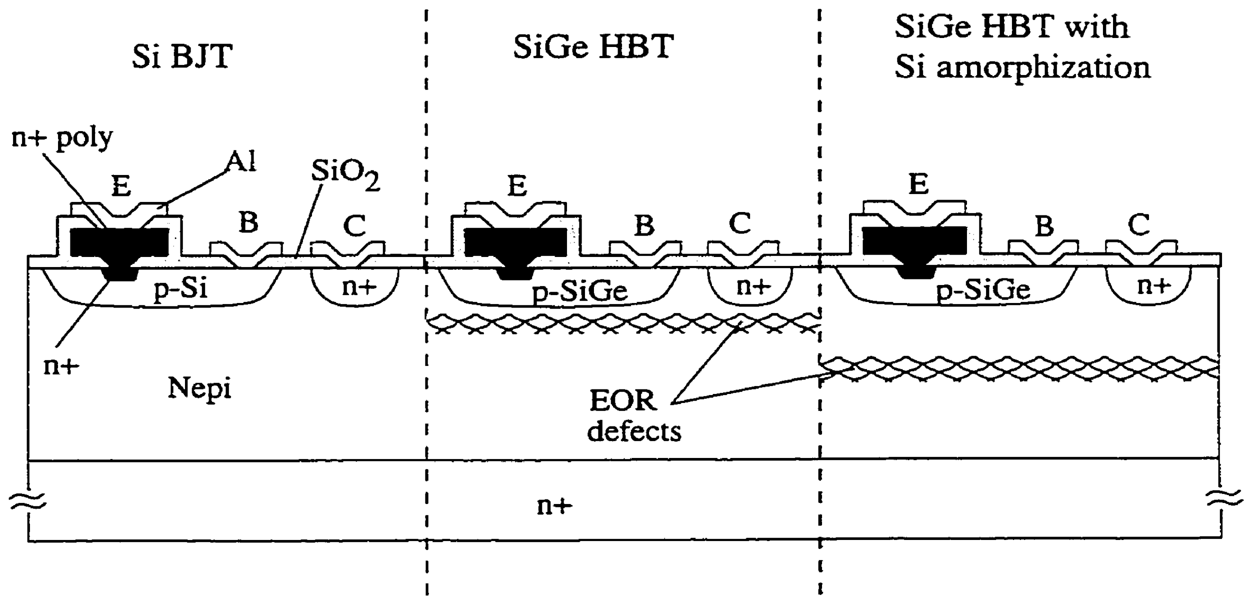

Figure 5.1: Schematic of a cross-sectional view of Si BJT, SiGe HBT, and SiGe HBT with Si amorphization.

Wafer	Temperature	Time
1	700°C	30 min
2	800°C	30 min
3	900°C	30 min
4	1000°C	30 sec
5	1050°C	30 sec
6	1100°C	30 sec

Table 5.1: Conditions used in the first annealing step (Note: all annealing steps were performed in a N_2 ambient).

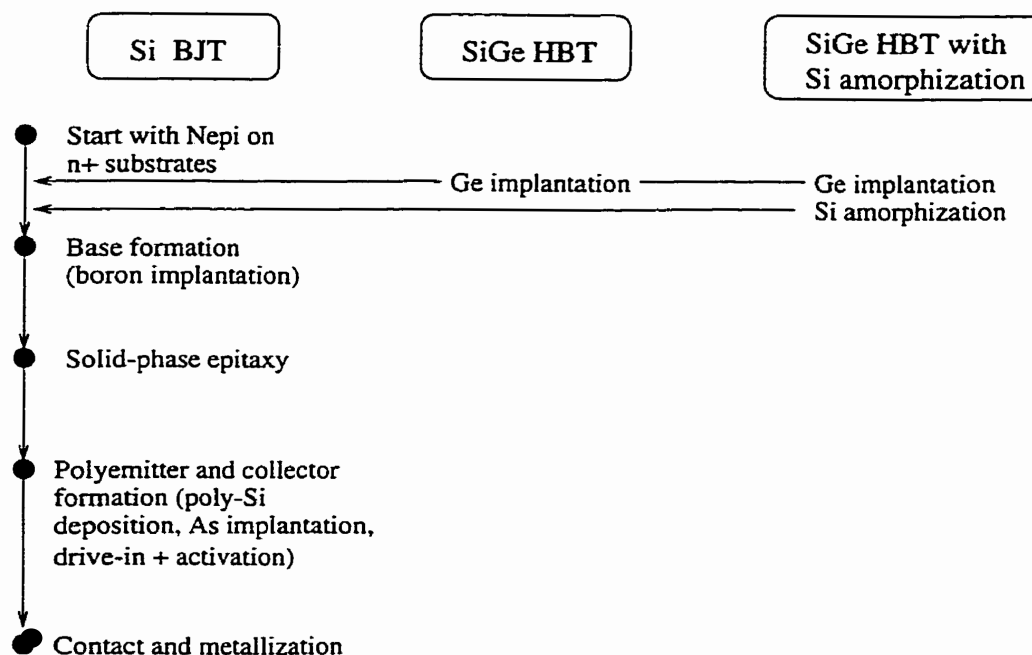

Figure 5.2: Fabrication process flow.

5.3 Device fabrication

Six 3-inch (100) Nepi on N+ wafers were used as the starting materials. The Nepi layer is 1 μm thick with a doping density of $7 \times 10^{16}\text{cm}^{-3}$. Except implantation, all fabrication steps were performed in the SiDIC Lab at the University of Waterloo. Only n-p-n transistors were fabricated. All transistors have the same lateral dimensions (an emitter area of 15×150 μm^2). Two annealing steps are required in the fabrication process: solid-phase epitaxial growth, and dopant drive-in and activation. Wafers were subjected to different conditions of the first annealing step. A brief summary of the process steps and details of the Ge implantation are given in the following. Details of other process steps can be found in Appendix D.

First, a steam oxide was grown and windows for Ge implantation were opened. Wafers covered with photoresist were then sent to Implant Sciences for Ge implan-

tation (^{74}Ge, 190 keV, 3.5×10^{16}cm^{-2}). An implant current density of 1.25 μA/cm^2 was used. The backsides of the wafers were coated with APD Cry-Con thermal grease to improve the thermal conductivity to the sample holder and the temperature of the wafers was monitored and kept below 125 K at the specified implant current density[3]. It has been found that lowering the substrate temperature can reduce end-of-range defects[4] [227, 228]. The thermal grease at the back was then removed. Photoresist on the front was stripped off using a "piranha" etch (7:3 H_2SO_4:H_2O_2) followed by a hot stripper. Wafers were cleaned using the conventional RCA recipe. Then, a low-temperature oxide was deposited in a LPCVD system at 410°C in the SiDIC Lab at the University of Waterloo. The low deposition temperature ensures that no solid-phase epitaxy can take place until the first high-temperature annealing step. Windows were then opened for Si amorphization. For the Si implant, an energy of 200 keV and a dose of 10^{16}cm^{-2} were used. Again, the same low-temperature setup used for the Ge implantation was employed for the Si implantation. Photoresist stripping and wafer cleaning then followed. The high-energy Si implantation will locate the initial amorphous/crystalline interface around 0.38 μm from the silicon surface [229]. In other words, the end-of-range extended defects remaining after the solid-phase epitaxial growth can be kept farther away from the depletion layer of the base-collector junction. Fig. 5.3 shows the Ge profile generated from the profile simulator PROFILE CODE developed by

[3]Although it is possible for the Ge implantation to be performed at room temperature as it is argued that the subsequent low-temperature high-dose Si implantation will re-amorphize the substrate anyway, there has been no comparison studies on the differences between results obtained from the two different wafer temperatures during Ge implantation.

[4]This is expected because as the substrate temperature is lowered, ion-beam-induced recrystallization is suppressed and a thicker amorphous layer can be formed. A thicker amorphous layer implies a deeper original a/c interface, leading to a smaller number of excess Si interstitials that form the end-of-range defects.

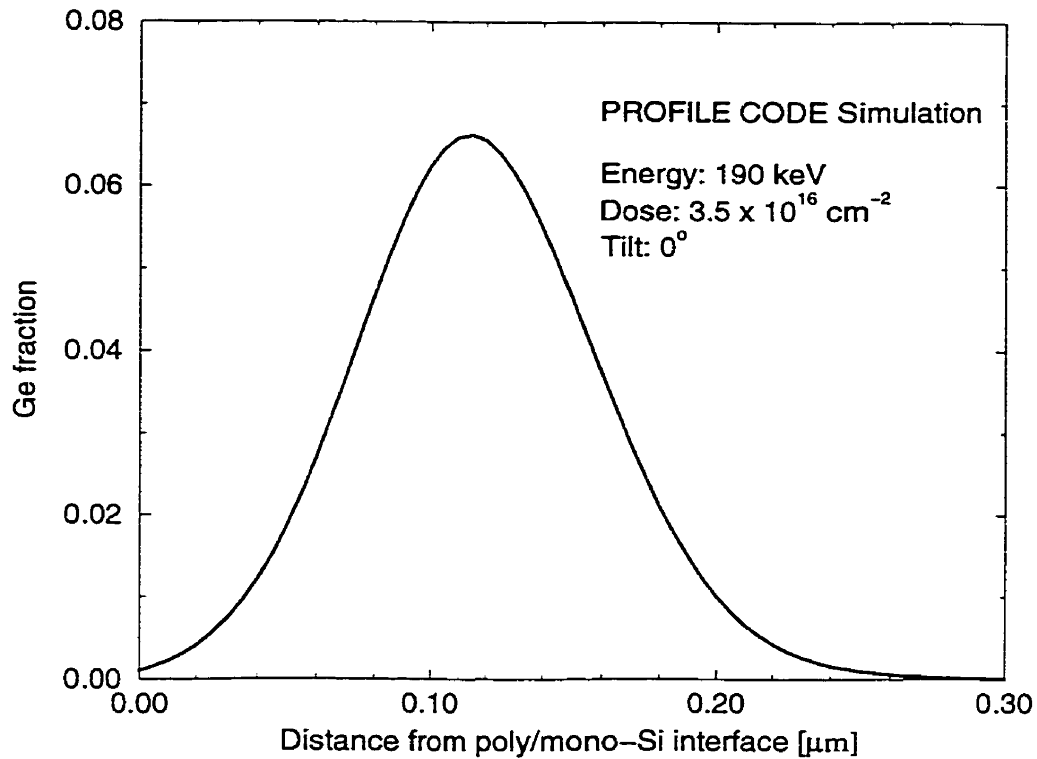

Figure 5.3: Simulated as-implanted Ge profile (190 keV and $3.5 \times 10^{16}\text{cm}^{-2}$) from PROFILE CODE.

Implant Sciences. The corresponding peak at.% is estimated around 6.7. Fig. 5.4 shows the net doping profiles from TSUPREM simulations for Si BJTs on wafer #3.

The remaining steps include the base, collector and polyemitter formations, and the Al sputtering for metallization. After boron (20 keV, $2 \times 10^{13}\text{cm}^{-2}$, tilt=7°, room temperature) was implanted for the base formation, the wafers were subject to the first annealing step for the solid-phase epitaxial growth (see Table 5.1 for details of annealing conditions). Then, emitter windows were opened and polysilicon was deposited at 585°C in a LPCVD system. Collector windows were opened and arsenic was implanted (120 keV, $2 \times 10^{16}\text{cm}^{-2}$, tilt=7°, room temperature) to form

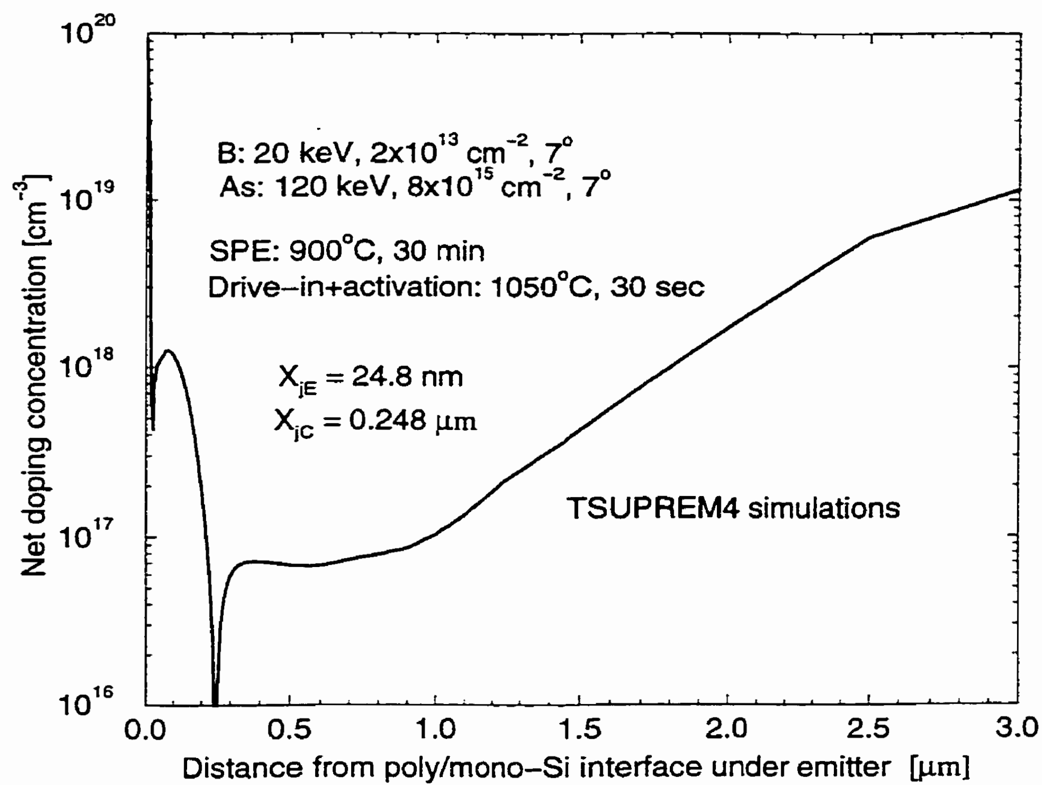

Figure 5.4: Simulated net vertical doping profile under the emitter of Si BJTs on wafer #3.

the collector and to dope the polysilicon. Polysilicon was patterned and low temperature oxide was grown. In order to drive in and activate the arsenic for emitter formation and break up the oxide at the polysilicon/mono-silicon interface, a rapid thermal annealing step at 1050°C for 30 seconds was performed in a N_2 ambient. Aluminum was then sputtered and patterned on the front side of the wafers, followed by sputtering on the back side for backside collector contacts. Since devices will be probed one at a time, no isolation between devices is required.

5.4 Experimental results

Gummel plots and common-emitter output characteristics were measured by two Keithley 236 Source Measure Units at room temperature. Most of the transistors showed poor characteristics and hence are not reported here, although considerable effort was expended in studying them including failure mode analysis. Transistors on wafer #3 show relatively better characteristics and their electrical characteristics are summarized as follows. Figure 5.5 shows the Si BJT Gummel plot and Fig. 5.6 shows the output characteristics of Si BJT. Figs. 5.7 and 5.8 show the I-V characteristics of the emitter-base and collector-base junctions of SiGe HBTs. Fig. 5.9 shows the output characteristics of SiGe HBTs with Si amorphization. While SiGe HBT with Si amorphization shows transistor action for the first time, its performance is not very satisfactory. Due to resource limitations, these studies could not be taken to logical conclusions. Further experimentation and refinements are needed which would probably be left for future work.

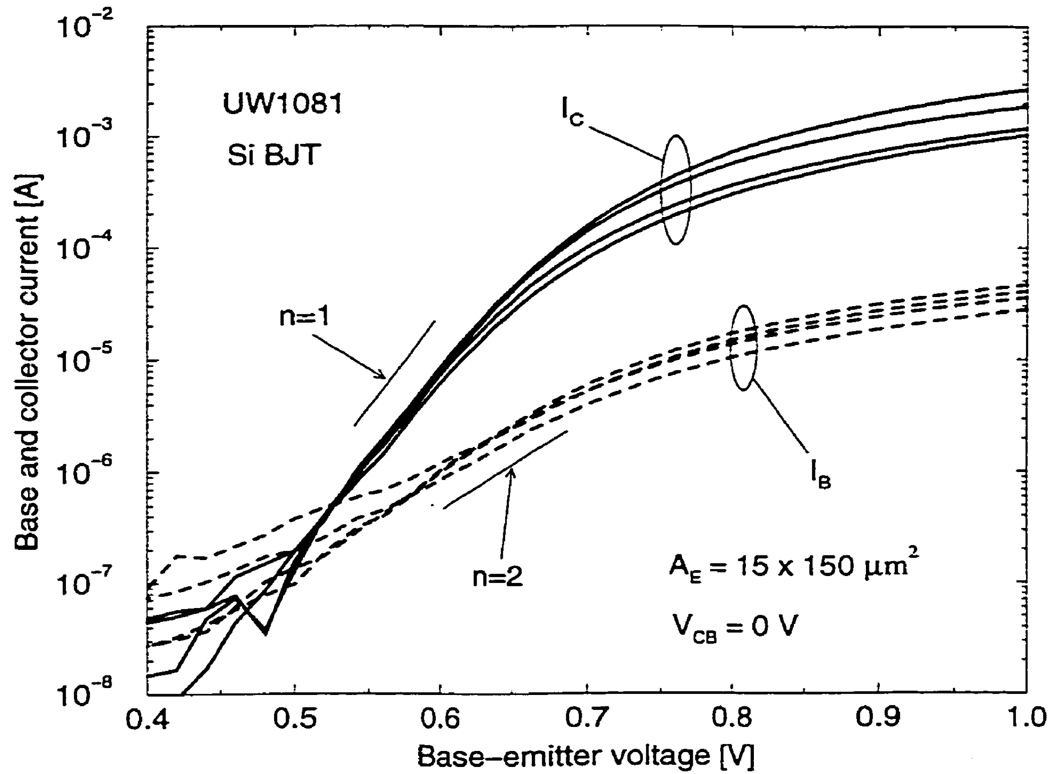

Figure 5.5: Typical measured Gummel plots of Si BJTs.

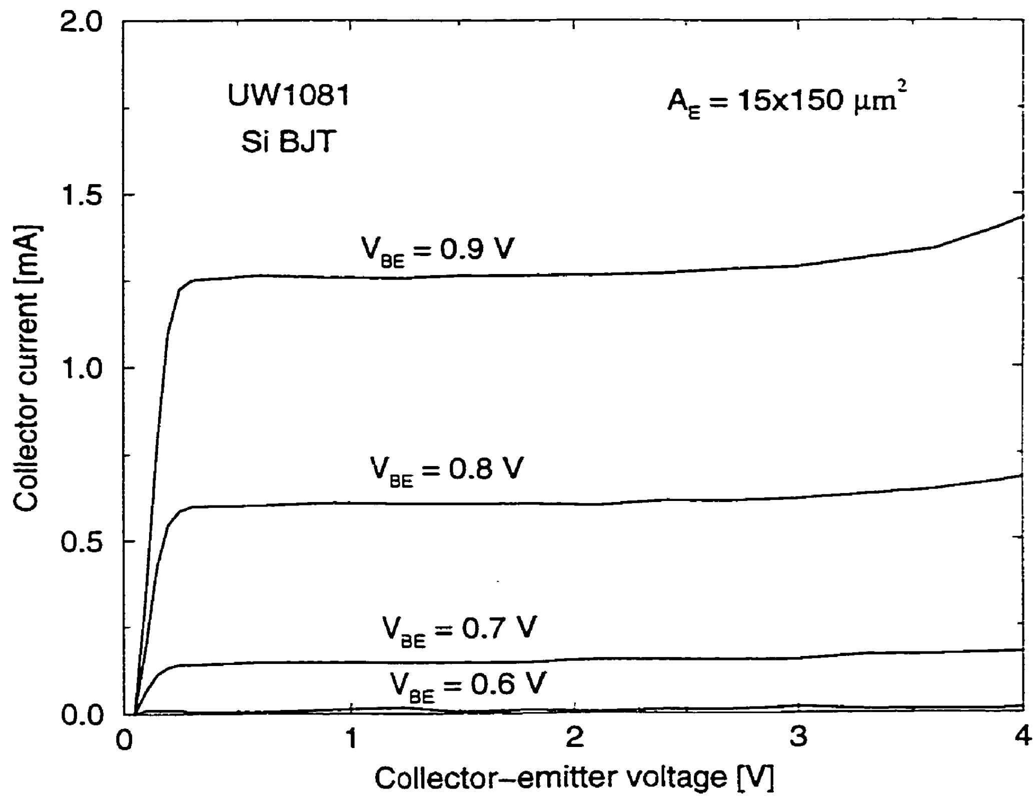

Figure 5.6: Typical measured output characteristics of Si BJT.

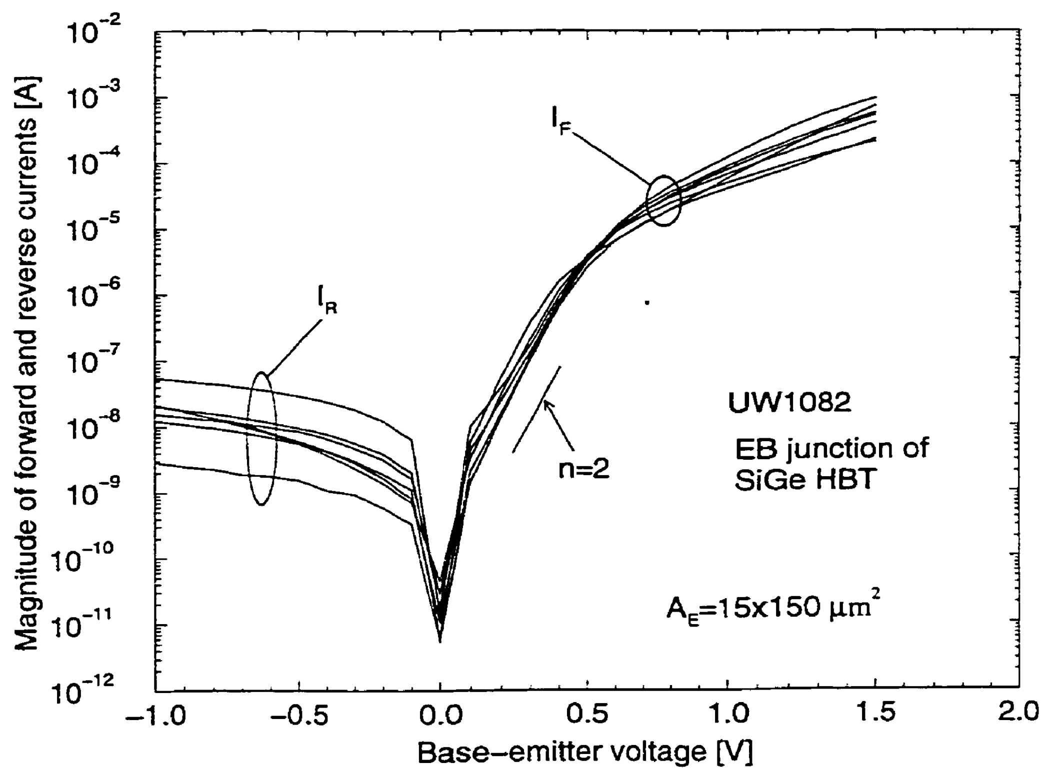

Figure 5.7: Typical emitter-base junction characteristics of SiGe HBTs.

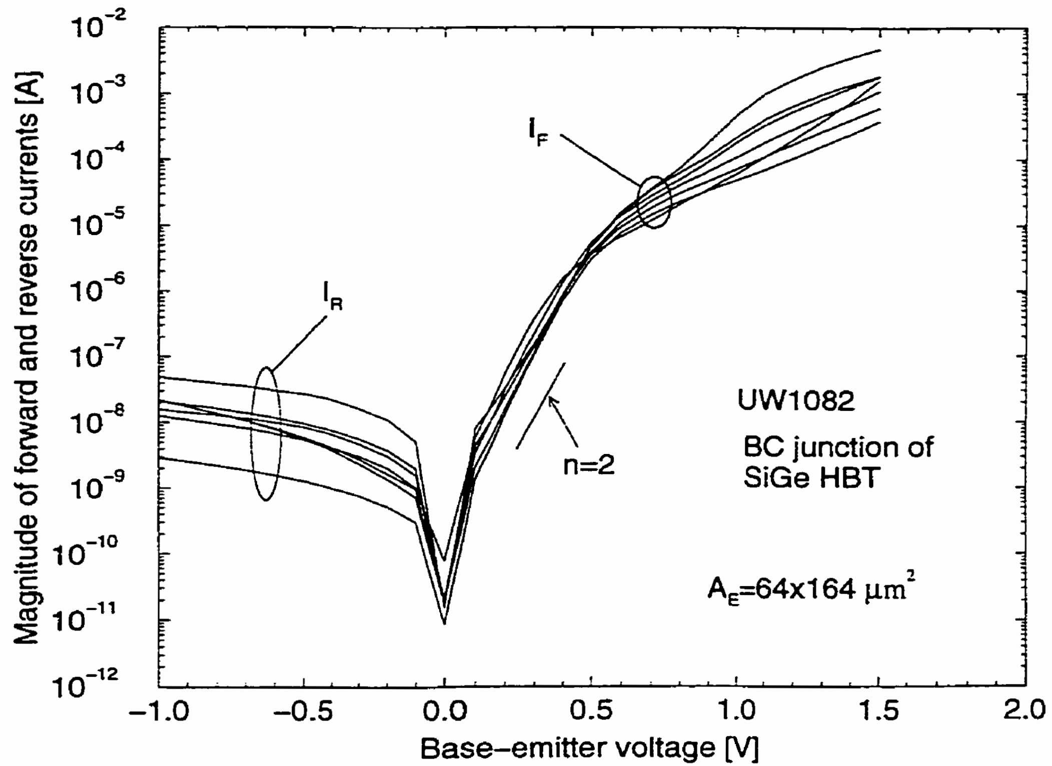

Figure 5.8: Typical base-collector junction characteristics of SiGe HBTs.

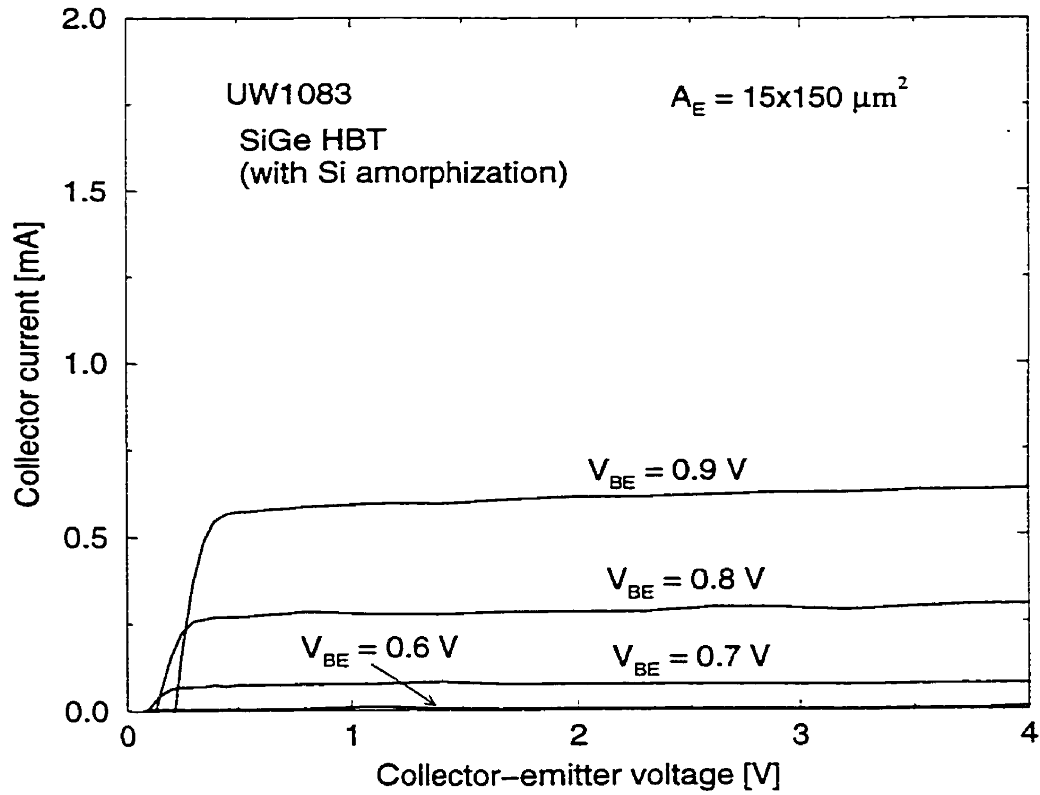

Figure 5.9: Typical measured output characteristics of SiGe HBT with Si amorphization.

5.5 Conclusions

SiGe HBTs, with and without Si amorphization, have been fabricated side-by-side with Si BJTs. Transistor action is shown by Si BJTs. SiGe HBTs without Si amorphization show no transistor action nor heterojunction effects. Transistor action of SiGe HBTs with Si amorphization is reported for the first time even though no heterojunction effect can be confirmed.

Further experimentation is required to improve the characteristics. Studies of dopant diffusion in implantation-formed SiGe and defect characterization of implanted SiGe would be helpful in further refinement of process architecture.

Chapter 6

Conclusions

Three topics related to SiGe HBTs have been studied. The first topic involved two vertical base profile optimization studies for improving the frequency performance of SiGe HBTs. In the first study, we mapped out important factors that impact the performance of the Ge profile in minimizing the contribution of the emitter and base delay times to the transition frequency in the low-injection regime. A new emitter delay time expression that allows a finite effective recombination velocity at the metal/poly-Si interface was derived. For the first time, a fixed Ge dose was adopted as the constraint, with non-quasi-static effects taken into account, in an optimization study.

We found that the detailed shape of a retrograde doping profile can have great influence on determining the optimal Ge profile in minimizing the total emitter and base delay time contribution to f_T. In particular, a retrograde base doping profile tilted toward the emitter-base junction will favor the choice of a less graded Ge profile. We also confirmed that the advantage of a graded Ge profile over a box Ge profile is greater for shallower emitter junctions, especially when non-quasi-static effects are considered. The main result of the study was: a graded Ge profile is still

more effective than a box Ge profile in minimizing the emitter and base delay time contribution to f_T for SiGe HBTs with typical emitter and base dimensions, i.e., an emitter junction depth smaller than 0.15 μm and a metallurgical base width larger than 30 nm.

In the second base profile optimization study, we compared the base delay time for different combinations of the base doping and Ge profiles in both low- and high-injection regimes before the onset of Kirk effect. We have presented a new iteration scheme that calculates the base delay time for a wide range of collector current densities instead of base-emitter biases. This is the first study which adopts identical Ge doses, intrinsic base resistances, and base concentrations, as optimization constraints, together with the effect of plasma-induced bandgap narrowing taken into consideration.

Three main results were obtained in the second optimization study. First, for the triangular Ge profile with the retrograde or uniform base doping profile considered, the base delay times in low-injection (e.g. $J_c = 10^3$ A/cm^2) and high-injection (e.g. $J_c = 10^5$ A/cm^2) regimes do not differ much before the onset of Kirk effect. Second, for the given Ge dose (equivalent to the total content of a uniform Ge profile of 19 at.% over a neutral base of 26 nm), intrinsic base resistance ($\sim 5k\Omega/\square$), and base concentration near the emitter ($N_{be} = 5 \times 10^{18}$cm^{-3}), graded Ge profiles ($X_T/W_b \geqslant 0.5$) together with a retrograde doping profile achieve the minimum (or near minimum) base delay time. Third, for the graded Ge profile (trapezoidal profile with $X_T/W_b = 0.675$) considered, Ge doses higher than the equivalent Ge content of a uniform profile at 10 at.% across a base of 25 nm will not effectively reduce the base delay time.

The results from the two base profile optimization studies are of particular importance to a SiGe BiCMOS process where only a limited amount of Ge can be

incorporated into the base for maintaining the thermal stability of the SiGe layer. These results can serve as starting points for choosing the appropriate base doping and Ge profiles to improve the high-frequency performance of future generations of SiGe HBTs.

The second topic in this thesis involved the analysis of the retrograde portion of a retrograde doping profile in a SiGe base with a graded Ge profile in the low-injection regime. We discovered a longstanding error in the derivation for a set of analytical base delay time expressions in the literature. We corrected the mistake and developed a new set of closed-form analytical expressions for the base delay time. Our expressions take into consideration important physical effects such as the built-in electric field caused by the non-uniform base doping profile, heavy doping effects, Ge-induced bandgap narrowing, and both the doping and the field dependencies of the electron diffusivity. One of the contributions arising from this analysis is the assessment of the relative importance of these physical effects in determining the base delay time.

We found that the adverse effect of the retrograde portion of the retrograde base profile considered was over-estimated by the incorrect analytical expressions. For a Ge grading of 13.5 at.% over a base of 50 nm, a peak doping concentration of 10^{19}cm^{-3}, and a retrograde region of 15 nm wide, the base delay time increases by ~ 20 % (versus 71 % predicted by the incorrect expressions) when the base concentration near the emitter drops from 9.5×10^{18} to $5 \times 10^{17}\text{cm}^{-3}$. This result will be of interest to researchers who adopt the retrograde base doping profile to reduce the capacitance and tunneling leakage current of the emitter-base junction in a SiGe HBT with a high base peak concentration.

The results from assessing the relative importance of the physical effects mentioned above showed that the field dependency of electron diffusivity has great

impact on determining the base delay time. This is especially true in the case of a high Ge grading across the SiGe base. For a Ge grading of 10% across a retrograde base of 50 nm, the effect of the field dependency of electron diffusivity can make a difference of $\sim$ 30% in the base delay time. This confirms the similar finding obtained in the second optimization study for minimizing the base delay time in both low- and high-injection regimes. In addition, we found from the second optimization study that the effect of plasma-induced bandgap narrowing should be considered for a SiGe base graded from 0 near the emitter to less than 0.15 near the collector in the high-injection regime. These findings are useful for modelling purposes.

The third topic of this thesis involved the fabrication of SiGe HBTs by high-dose Ge implantation, Si amorphization, and subsequent solid phase epitaxy. Transistor action of SiGe HBTs fabrication by high-dose Ge implantation and solid-phase epitaxy, with Si amorphization, is reported for the first time.

As a final remark, the work presented in this thesis has only dealt with a few of many important issues relating to SiGe HBTs. For example, the problem of optimizing emitter and base profiles in the emitter-base depletion layer was not considered. The base profile optimization study for minimizing the emitter and base delay time contribution to the transition frequency, with non-quasi-static effects considered in the high-injection regime, was not discussed. In addition, more novel Ge profiles, for instance, Ge profiles with different Ge gradings in more than two base subregions, can be considered in the optimization studies. Regarding the SiGe HBT fabrication, structural characterization for dopant distributions and extended defects can be attempted. After all, the research area of SiGe HBTs is vast. It is hoped that the studies in this thesis have made a meaningful contribution to the rapidly growing field of SiGe HBT technology.

Appendix A

General formulation for S'_p

Recently, Rinaldi [109] proposed a general formulation of the effective recombination velocity at the poly/mono-Si interface (S'_p) to clarify the similarities and differences between various poly-emitter models. The key to his formulation is the re-consideration of the current continuity at *two* interfaces: one between the poly-Si and the interfacial oxide ($x = W_{em}$) and the other one between the interfacial oxide and the mono-Si region ($x = W'_{em}$) (see Fig. A.1). This Appendix shows how equations (2.10), (2.19) and (2.27) arise from Rinaldi's general formulation.

For the current continuity at $x = W_{em}$,

$$J_i = J_{p2}(W_{em}) + J_{pr}(W_{em}) \tag{A.1}$$

where $J_{p2}(W_{em})$ is the hole current injected into the poly-Si region; $J_{pr}(W_{em})$ is the hole recombination current due to the traps at the poly-Si/oxide interface; J_i is the tunneling and/or thermionic-emission current across the interfacial oxide.

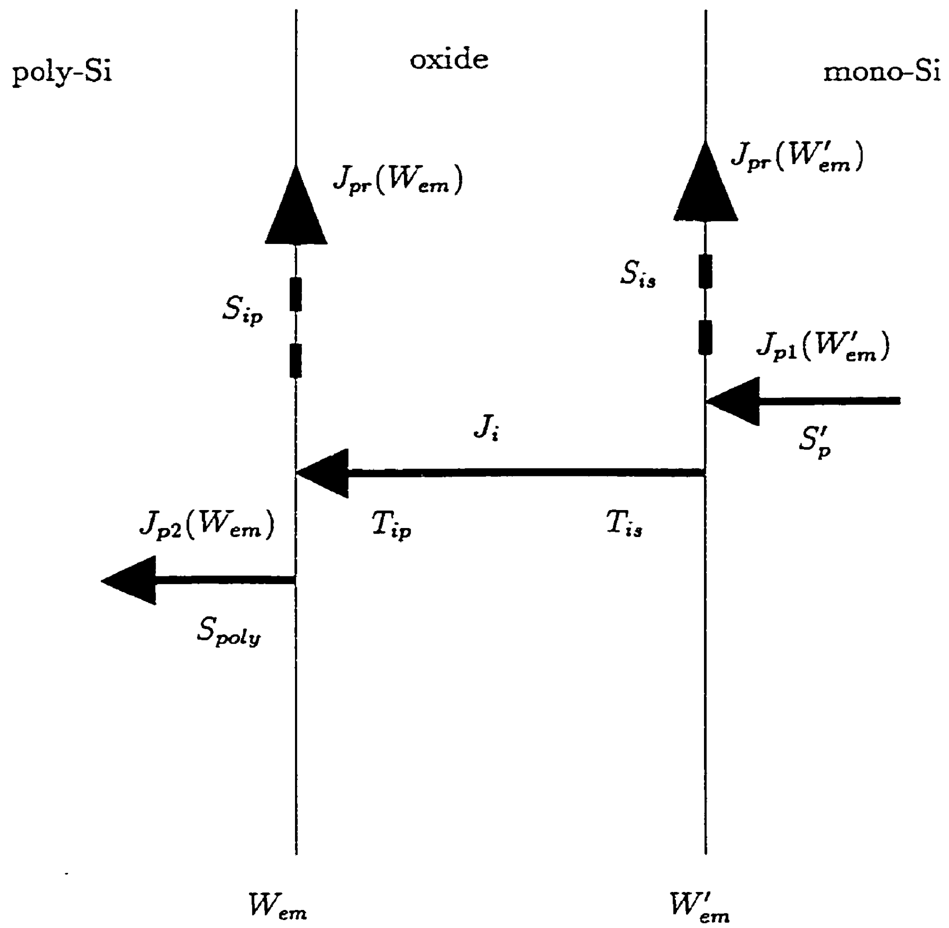

Figure A.1: A schematic of current flows at the poly-Si/oxide/mono-Si interfaces.

Parameters	Definition
S_{poly}	Effective recombination velocity with respect to the poly-Si bulk region
S'_p	Effective recombination velocity at the poly-Si/mono-Si interface
S_{is}	Interface recombination velocity at the oxide/mono-Si interface
S_{ip}	Interface recombination velocity at the poly-Si/oxide interface
T_{is}	Interface blocking factor at the oxide/mono-Si interface
T_{ip}	Interface blocking factor at the poly-Si/oxide interface

Table A.1: Definitions of poly-emitter model parameters

For the current continuity at $x = W'_{em}$,

$$J_{p1}(W'_{em}) = J_i + J_{pr}(W'_{em}) \tag{A.2}$$

where $J_{pr}(W'_{em})$ is the hole recombination current due to the traps at the oxide/mono-Si interface; $J_{p1}(W'_{em})$ is the hole current injected from the mono-Si region into the interfacial oxide layer.

The five current components in (A.2) and (A.1) can be expressed as follows:

$$\begin{aligned} J_{p2}(W_{em}) &= qS_{poly}p_2(W_{em}) \\ J_{pr}(W_{em}) &= qS_{ip}p_2(W_{em}) \\ J_i &= q[T_{is}p_1(W'_{em}) - T_{ip}p_2(W_{em})] \\ J_{pr}(W'_{em}) &= qS_{is}p_1(W'_{em}) \\ J_{p1}(W'_{em}) &= qS'_p p_1(W'_{em}) \end{aligned} \tag{A.3}$$

where p_1 and p_2 are the hole concentrations in the mono-Si and poly-Si region, respectively, and the definitions of S_{poly}, S_{ip}, S_{is}, S'_p, T_{is}, and T_{ip} are listed in Table A.1.

By making different assumptions on the relative importance of the each possible current gain mechanism mentioned in Chapter 2, various polyemitter models assign

different expressions to the model parameters. The parameter S_{poly} encapsulates the transport property in the poly-Si bulk. S_{is} and S_{ip} model recombination at the two interfaces. T_{is} and T_{ip} model the tunneling and/or barrier effect due to the interfacial oxide. Finally, the parameter S_p' encapsulates the overall effect due to the presence of the poly-Si region and the interfacial oxide layer. Substituting (A.3) into (A.2) and (A.1) gives:

$$S_p' = S_{is} + \frac{T_{is}(S_{poly} + S_{ip})}{S_{poly} + S_{ip} + T_{ip}} \tag{A.4}$$

$$\frac{p_2(W_{em})}{p_1(W_{em}')} = \frac{T_{is}}{T_{ip} + S_{poly} + S_{ip}} \tag{A.5}$$

Substituting (A.5) into (A.4) yields:

$$S_p' = S_{is} + \frac{p_2(W_{em})}{p_1(W_{em}')}(S_{poly} + S_{ip}) \tag{A.6}$$

If $T_{is} = T_{ip}$ and $S_{is} = S_{ip}$, then (A.4) reduces to:

$$S_p' = S_{is} + \left(\frac{1}{T_{is}} + \frac{1}{S_{poly} + S_{is}}\right)^{-1} \tag{A.7}$$

which is equation (18) in Yu *et al.* 's model [106]. Yu *et al.* also set T_{is} and T_{ip} to $v_c\alpha$ where v_c is the collection velocity and α is the tunneling probability $\frac{e^{-b_h}}{1-C_hkT}$ [120]. As observed by Post *et al.* [98], T_{is} and T_{ip} take on different expressions in various polyemitter models.

For Suzuki's model [108] (on which our derivation is based), S_{ip} is set to zero, T_{ip} and T_{is} are assumed to be much greater than S_{poly}, and T_{is}/T_{ip} is equated to

the tunneling probability [120]. Equations (A.5) and (A.6) then become:

$$\frac{p_2(W_{em})}{p_1(W'_{em})} = \frac{T_{is}}{T_{ip}} = \frac{e^{-b_h}}{1 - C_h kT} \tag{A.8}$$

$$S'_p = S_{is} + \frac{p_2(W_{em})}{p_1(W'_{em})} S_{poly} \tag{A.9}$$

Recognizing $\frac{e^{-b_h}}{1-C_h kT}$ as α, (A.8) and (A.9) can be written as:

$$\alpha = \frac{p_2(W_{em})}{p_1(W'_{em})} \tag{A.10}$$

and

$$S'_p = S_{is} + \alpha S_{poly} \tag{A.11}$$

which are (2.10) and (2.27), respectively, in Chapter 2 with $S_{is} = S_p$. Since $S_{ip} = 0$, the hole recombination current at the poly-Si/oxide interface ($J_{pr}(W_{em})$) becomes zero, therefore (A.2) and (A.1) can be combined into:

$$J_{p1}(W'_{em}) = J_{pr}(W'_{em}) + J_{p2}(W_{em}) \tag{A.12}$$

which is (2.19) in Chapter 2 if $J_{pr}(W'_{em})$ is simply written as J_{pr}.

Appendix B

Derivation for equation (3.26)

Take (3.26) as an example. Substituting (3.23) and $|E| = V_T \frac{d}{dx}[ln(n_o(x))]$ (Note. $E < 0$ in regions II & III for $a > 0$) into (3.21) gives:

$$\begin{aligned}
\tau_{b2} &= \int_{x_1}^{x_2} n_{o2}(x) \int_x^{W_b} \frac{1 + \frac{D_n^l(z)}{n_o(z)v_s}\frac{dn_o(z)}{dz}}{D_n^l(z)n_o(z)} dz\, dx + \frac{\int_{x_1}^{x_2} n_{o2}(x)dx}{Sn_{o3}(W_b)} \\
&= \int_{x_1}^{x_2} n_{o2}(x) \int_x^{W_b} \frac{dz}{D_n^l(z)n_o(z)} dx + \frac{1}{v_s}\int_{x_1}^{x_2} n_{o2}(x) \int_x^{W_b} \frac{dn_o(z)}{n_o^2(z)} dx + \frac{\int_{x_1}^{x_2} n_{o2}(x)dx}{Sn_{o3}(W_b)} \\
&= \int_{x_1}^{x_2} n_{o2}(x) \int_x^{W_b} \frac{dz}{D_n^l(z)n_o(z)} dx - \frac{1}{v_s}\int_{x_1}^{x_2} n_{o2}(x) \int_x^{W_b} dn_o^{-1}(z)dx + \frac{\int_{x_1}^{x_2} n_{o2}(x)dx}{Sn_{o3}(W_b)} \\
&= \int_{x_1}^{x_2} n_{o2}(x) \int_x^{W_b} \frac{dz}{D_n^l(z)n_o(z)} dx - \frac{1}{v_s}\int_{x_1}^{x_2} n_{o2}(x)[n_{o3}^{-1}(W_b) - n_{o2}^{-1}(x)]dx + \frac{\int_{x_1}^{x_2} n_{o2}(x)dx}{Sn_{o3}(W_b)} \\
&= \int_{x_1}^{x_2} n_{o2}(x) \int_x^{W_b} \frac{dz}{D_n^l(z)n_o(z)} dx - \frac{\int_{x_1}^{x_2} n_{o2}(x)dx}{v_s n_{o3}(W_b)} + \frac{x_2 - x_1}{v_s} + \frac{\int_{x_1}^{x_2} n_{o2}(x)dx}{Sn_{o3}(W_b)}
\end{aligned} \tag{B.1}$$

Similarly, (3.25) and (3.27) can be obtained.

Appendix C

Physical models for electron transport in SiGe

This Appendix reviews the most recent and consistent physical models for describing the electron transport in SiGe, with the focus on how these models were developed in a consistent manner from both measured data and theoretical studies. In particular, Sokolic and Amon's framework of interpreting the collector current data will be adopted [205] since it accounts for the lower effective densities of states in the energy bands of SiGe due to the strain-induced energetic splitting and the change in the hole effective mass [230], and the Fermi-level shift associated with the use of the Fermi-Dirac statistics in the case of a heavily doped base [231].

From Chapter 2, the collector current density of a SiGe HBT in the low-injection regime is written as:

$$J_{c,SiGe} = \frac{q\ e^{V_{BE}/V_T}}{\int_0^{W_b} \frac{dx}{D_{n,SiGe} n_o} + \frac{1}{S n_o(W_b)}} \tag{C.1}$$

Assuming negligible velocity saturation at the collector junction (i.e. $S \to \infty$),

$J_{c,SiGe}$ becomes:

$$J_{c,SiGe} = \frac{q\ e^{V_{BE}/V_T}}{\int_0^{W_b} \frac{N_b}{D_{n,SiGe} n_{ie}^2} dx} \tag{C.2}$$

because $n_o = n_{ie}^2/N_b$. To relate the profile information to $J_{c,SiGe}$, the effective intrinsic electron concentration n_{ie} is expanded as follows [205]:

$$n_{ie}^2 = n_i^2\ exp\left(\frac{\Delta E_g}{kT}\right) \tag{C.3}$$

with the apparent bandgap narrowing

$$\begin{aligned} \Delta E_g =& \Delta E_{g,HDE} + \Delta E_{g,Ge} + kT\ ln\left(\frac{N_{c,SiGe} N_{v,SiGe}}{N_{c,Si} N_{v,Si}}\right) \\ &+ kT\left[ln\left(\frac{N_b}{N_{v,SiGe}}\right) - \mathcal{F}_{\frac{1}{2}}^{-1}\left(\frac{N_b}{N_{v,SiGe}}\right)\right] \end{aligned} \tag{C.4}$$

The notation $\mathcal{F}_{\frac{1}{2}}^{-1}$ denotes the inverse function of Fermi-Dirac integral of order 1/2. The quantity $N_{c,SiGe}(N_{v,SiGe})$ is the density of states in the conduction (valence) band of SiGe. The first two terms of (C.4), $\Delta E_{g,HDE}$ and $\Delta E_{g,Ge}$, denote the bandgap narrowing due to the doping and the Ge presence, respectively. The last two terms account for the reduced effective densities of states due to the Ge, and the Fermi-level shift associated with Fermi-Dirac statistics, respectively.

Under the compressive strain caused by the lattice mismatch between Si and Ge, energy band splitting occurs in both the conduction and the valence bands of SiGe [232]. The effective density of states in the conduction band of SiGe ($N_{c,SiGe}$) can be approximated as $\frac{2}{3}N_{c,Si}$ since the lower fourfold degenerate states after splitting are Si-like as long as the Ge fraction is less than 0.85 [233] (as the ellipsoidal constant energy surfaces are unchanged to first order under strain [232]). However,

the situation for the effective density of states in the valence band is more complicated and will be discussed later. Regarding the heavy-doping induced bandgap narrowing ($\Delta E_{g,HDE}$) in SiGe, Poortmans *et al.* [234] have predicted a difference of $\sim 5-10$ meV larger than the Si value for base doping less than 10^{19}cm^{-3}and Ge fraction less than 0.2. Their prediction is in agreement with the experimental data obtained by photoluminescence [235]. Since the accuracy of experimental procedure is ~ 10 meV, it is reasonable to assume that the doping-induced bandgap narrowing in Si and SiGe are practically the same [205].

In short, in order to model $J_{c,SiGe}$ accurately for different base doping and Ge profiles and operating temperatures, one has to first obtain the physical models for $D_{n,SiGe}$, $\Delta E_{g,Ge}$, $N_{v,SiGe}$, and $\Delta E_{g,HDE}$ over a wide range of Ge fraction, doping concentration, and temperature. In fact, many experimental studies have attempted to extract $\Delta E_{g,Ge}$ as a function of the Ge fraction by DC measurements on SiGe HBTs with uniform base doping and Ge profiles. Assuming uniform doping and Ge profiles in the base (i.e., $D_{n,SiGe}$ and n_{ie} are position-independent), $J_{c,SiGe}$ from (C.2) becomes:

$$J_{c,SiGe} = \frac{qn_{ie}^2 D_{n,SiGe}}{N_b W_b} exp\left(\frac{V_{BE}}{V_T}\right) \tag{C.5}$$

However, similar to the observation made by del Alamo *et al.* [236] for Si transistors, DC measurements can only give information on the product $n_{ie}^2 D_{n,SiGe}$ in (C.5) but not the individual terms n_{ie}^2 and $D_{n,SiGe}$. Even if n_{ie}^2 could be extracted, it still requires the knowledge of the $N_{v,SiGe}$ in order to obtain $\Delta E_{g,Ge}$ accurately from (C.4).

Therefore, a better method to consistently obtain the physical models $\Delta E_{g,HDE}$, $N_{v,SiGe}$, $D_{n,SiGe}$, and $\Delta E_{g,Ge}$ is the following. First of all, $\Delta E_{g,HDE}$ is determined

from DC measurements on Si BJTs. The collector current density of a Si BJT with uniformly-doped base can be written as:

$$J_{c,Si} = \frac{qn_i^2 D_{n,Si}}{N_b W_b} exp\left(\frac{V_{BE}}{V_T}\right) exp\left(\frac{\Delta E_g}{kT}\right) \tag{C.6}$$

where n_i can be obtained from the recent Green's model [71] and

$$\Delta E_g = \Delta E_{g,HDE} + kT\left[ln\left(\frac{N_b}{N_{v,Si}}\right) - \mathcal{F}_{\frac{1}{2}}^{-1}\left(\frac{N_b}{N_{v,Si}}\right)\right] \tag{C.7}$$

with $N_{v,Si}$ also taken from Green's model. Based on Klaassen's latest mobility model [121, 125] a new $\Delta E_{g,HDE}$ model which is independent of the Fermi-level shift associated with the use of Fermi-Dirac statistics can be obtained from (C.6) and (C.7) [206]. It should be noted the unified bandgap narrowing model obtained consistently by Klaassen *et al.* [132] from Klaassen's latest mobility model does consider Fermi-Dirac statistics for heavily doped bases.

Secondly, $N_{v,SiGe}$ is determined from photoluminescence measurement at different temperatures and used to extract $D_{n,SiGe}$ from a set of temperature-dependent measurements over a wide range of Ge fraction. Decoutere *et al.* [209] assumed identical temperature and doping dependences of the the electron diffusivities in Si and SiGe. Using Klaassen's mobility model and $N_{c,SiGe}N_{v,SiGe}$ experimental values from Poortmans *et al.* [237], and recognizing the limitation of Boltzmann Statistics at high base concentrations, Decoutere *et al.* expressed $D_{n,SiGe}$ as a product of Klaassen's dopant-dependent mobility value and a Ge-dependent multiplying factor. They have shown that their $D_{n,SiGe}$ model agrees well with the equivalent mobility values obtained by Manku and Nathan's calculation [238] and Monte-Carlo simulations of Hinckley *et al.* [239]. In Decoutere's model, the electron mobility for SiGe is always higher than that in Si for a Ge fraction less than 0.2. This is con-

sistent with the physical picture because a higher out-of-plane (i.e. perpendicular to the growth plane) electron mobility is caused by the population of the lowered fourfold degeneracy of the conduction band valleys and the reduced inter-valley scattering. As the Ge fraction is higher, the mobility will drop as a result of alloy scattering [240].

Thirdly, taking into account the Fermi-level shift associated with Fermi-Dirac statistics as shown in (C.4), using Decoutere *et al.* 's $D_{n,SiGe}$ model, the re-calculated $\Delta E_{g,HDE}$ model, and their new $N_{v,SiGe}$ model, Sokolic and Amon re-calculated $\Delta E_{g,Ge}$ from a set of of DC measurements over a range of Ge fraction. They [205] found that the extracted values of $\Delta E_{g,Ge}$ agree well with those obtained from optical measurement [69]. It should be noted that Green's n_i^2 model was consistently adopted in interpreting the measured data.

Finally, a few words on Sokolic and Amon's $N_{v,SiGe}$ model are in order. The compressive strain in SiGe not only causes the energetic splitting between the heavy-hole and light-hole valence bands and down shifting of the split-off band, but also distorts the electronic structure of the valence band. This change in the valence band results in a smaller hole effective mass m_p^* and consequently a smaller $N_{v,SiGe}$. Therefore, one needs to first determine $m_p^*(T, N_b, y)$ in order to model $N_{v,SiGe}$ accurately over a wide range of Ge fraction (y), doping level (N_b), and temperature (T). Building upon the theoretical works on SiGe valence band structures by Manku and Nathan [241] and Fu *et al.* [242], and incorporating the temperature dependence of m_p^* from Green [71] and doping dependence of m_p^* from Fu *et al.*, Sokolic and Amon managed to derive a $m_p^*(T, N_b, y)$ model [230] over a wide range of temperature, base doping, and Ge fraction. This new $m_p^*(T, N_b, y)$ model then leads to the new $N_{v,SiGe}$ model presented in [205], which agrees reasonably well with Poortmans *et al.* 's $N_{c,SiGe}N_{v,SiGe}$ values [237] used previously in obtaining $D_{n,SiGe}$ by Decoutere

et al. . Such agreement strengthens the consistency of Sokolic and Amon's set of physical models for describing the electron transport in a strained SiGe layer.

Appendix D

Ge-implanted base and polysilicon-emitter HBT process

This appendix includes the details of the fabrication process of a SiGe HBT formed by high-dose Ge implantation, solid-phase epitaxy, and Si amorphization.

The starting materials are 3-inch Nepi on N+ Si (100) wafers. The N+ substrate is 380 μm thick with a resistivity of 0.001-0.004 Ω-cm. The n-doped epitaxial (Nepi) layer is 1 μm thick with a doping density of $7 \times 10^{16}\text{cm}^{-3}$. Both the Nepi layer and the N+ substrate are doped with arsenic. All steps except the implantation steps were performed in the SiDIC lab at the University of Waterloo. The mask set, named UW-108, consists of eight masks: germanium, silicon-post-amorph, base, emitter, collector, poly, contact, and metal (numbered in this order). Details of the process steps [243] are given in Tables D.1 and D.2. It should be noted that wafers sent to Implant Sciences for implantation were covered with photoresist.

Step	Parameters	Comments
1	RCA I & II, HF dip	Clean wafer surface
2	1100°C, 7 min, steam	Oxide for patterning, 0.18 μm
3	8:1 BHF, mask #1 (Ge)	Open window for Ge implant
4	Ge, 190 keV, 3.5×10^{16}cm^{-2}, tilt=0, LN_2	Implant Sciences
5	7:3 H_2SO_4:H_2O_2, hot stripper	Strip PR by piranha etch
6	RCA I & II, HF dip	Clean wafer surface
7	410°C, 40 min, 0.32 torr	Oxide for patterning, 0.35 μm
8	8:1 BHF, mask #2 (Si)	Open window for Si amorphization
9	Si, 200 keV, 10^{16}cm^{-2}, tilt=0, LN_2	Push EOR defects to 0.38 μm
10	7:3 H_2SO_4:H_2O_2, hot stripper	Strip PR by piranha etch
11	RCA I & II, HF dip	Clean wafer surface
12	410°C, 40 min, 0.32 torr	Oxide for patterning, 0.175 μm
13	8:1 BHF, mask #3 (base)	Open base window
14	B, 20 keV, 2×10^{13}cm^{-2}, RT, tilt=7°	Implant Sciences
15	hot stripper	Strip PR
16	RCA I & II, HF dip	Clean wafer surface
17	410°C, 20 min, 0.37 torr	Oxide for patterning and annealing, 0.47 μm
18	see Table 5.1	Solid-phase epitaxy
19	8:1 BHF, mask #4 (emitter)	Open emitter window
20	RCA I & II, HF dip	Clean emitter surface

Table D.1: Fabrication process steps.

Step	**Parameters**	**Comments**
21	585°C, 40 min, 0.32 torr	Polysilicon deposition, 0.36 μm
22	410°C, 20 min, 0.37 torr	Oxide for patterning, 0.47 μm
23	8:1 BHF, mask #5 (collector)	Open collector window
24	KOH, 55°C	Etch polysilicon in collector
25	8:1 BHF	Remove oxide in collector
26	As, 120 keV, 8×10^{15}cm^{-2}, tilt=7°, RT	Implant Sciences
27	RCA I & II, HF dip	Clean wafer surface
28	410°C, 20 min, 0.38 torr	Oxide for patterning, 0.45 μm
29	8:1 BHF, mask #6 (poly)	Pattern polysilicon
30	KOH, 55°C	Etch polysilicon
31	RCA I & II, HF dip	Clean wafer surface
32	410°C, 30 min, 0.39 torr	Oxide for patterning, 0.67 μm
33	RTA, 1050°C, 30 sec, N_2	Drive-in, break up interfacial oxide, dopant activation
34	8:1 BHF, mask #7 (contact)	Open contact window
35	RCA I & II, HF dip	Clean wafer surface
36	150W, 31.3% Ar, 5 mtorr	Sputter Al, 1.8 μm thick
37	mask #8 (metal), 37°C, $H_3PO_4/HNO_3/H_4C_2O_2$	Pattern metal
38	hot wash	Clean wafer surface
39	wipe backside with HF	Deglaze backside
40	150W, 32.1% Ar, 5 mtorr	Sputter Al back contact, 0.66 μm

Table D.2: Fabrication process steps (cont'd).

Bibliography

[1] J.D. Cressler, "SiGe HBT technology: a new contender for Si-based RF and microwave circuit applications", *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, no. 5, pp. 572–589, 1998.

[2] J. Sitch, "HBTs in telecommunications", *Solid-St. Electron.*, vol. 41, no. 10, pp. 1397–1405, 1997.

[3] T. Baumheinrich and U. Langmann, "Design of high speed bipolar Si/SiGe ICs for optical wide band communications", in *ISCAS'99. Proceedings of the 1999 IEEE International Symposium on Circuits and Systems. VLSI*, 1999, vol. 2, pp. 496–499.

[4] H.-M. Rein, R. Schmid, P. Weger, T. Smith, T. Herzog, and R. Lachner, "A versatile Si-bipolar driver circuit with high output voltage swing for external and direct laser modulation in 10 Gb/s optical-fiber links", *IEEE Journal of Solid-State Circuits*, vol. 29, no. 9, pp. 1014–1021, 1994.

[5] A. Felder, M. Moller, J. Popp, J. Bock, and H.-M. Rein, "46 Gb/s DEMUX, 50 Gb/s MUX, and 30 GHz static frequency divider in silicon bipolar technology", *IEEE Journal of Solid-State Circuits*, vol. 31, no. 4, pp. 481–486, 1996.

[6] W. Bogner, U. Fischer, E. Gottwald, and E. Mullner, "20 Gbit/s TDM nonrepeatered transmission over 198 km DSF using Si-bipolar IC for demultiplexing and clock recovery", in *Proceedings of European Conference on Optical Communication*, 1996, vol. 2, pp. 203–206.

[7] M. Wurzer, J. Bock, W. Zirwas, H. Knapp, F. Schumann, A. Felder, and L. Treitinger, "40 Gb/s integrated clock and data recovery circuit in a silicon bipolar technology", in *Proceedings of the 1998 Bipolar/BiCMOS Circuits and Technology Meeting*, 1998, pp. 136–139.

[8] H. M. Rein and M. Moller, "Design considerations for very-high-speed Si-bipolar IC's operating up to 50 Gb/s", *IEEE Journal of Solid-State Circuits*, vol. 31, no. 8, pp. 1076–1090, 1996.

[9] A. Edman and B. Rudberg, "SDH 10 Gb/s regenerator framer in 0.6 μm CMOS", in *1997 IEEE International Solids-State Circuits Conference. Digest of Technical Papers*, 1997, pp. 156–157.

[10] E. Sickinger and W.C. Fischer, "A 3-GHz 32-dB CMOS limiting amplifier for SONET OC-48 receivers", *IEEE Journal of Solid-State Circuits*, vol. 35, no. 12, pp. 1884–1888, 2000.

[11] J. Savoj and B. Razavi, "A 10-Gb/s CMOS clock and data recovery circuit", in *2000 Symposium on VLSI Circuits. Digest of Technical Papers*, 2000, pp. 136–139.

[12] A. Tanabe, M. Umetani, I. Fujiwara, T. Ogura, K. Kataoka, M. Okihara, H. Sakuraba, T. Endoh, and F. Masuoka, "A 10 Gb/s demultiplexer IC in 0.18 μm CMOS using current mode logic with tolerance to the threshold voltage

fluctuation", in *2000 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, 2000, pp. 62–63.

[13] K. Murata, T. Otsuji, M. Yoneyama, and M. Tokumitsu, "A 40-Gbit/s superdynamic decision IC fabricated with 0.12-μm GaAs MESFET's", *IEEE Journal of Solid-State Circuits*, vol. 33, no. 10, pp. 1527–1535, 1998.

[14] Y. Kuriyama, M. Asaka, T. Sugiyama, N. Iizuka, and M. Obara, "Over 40 Gbit/s ultrahigh-speed multiplexer IC implemented with high f_{max} AlGaAs/GaAs HBTs", *Electronics Letters*, vol. 30, no. 5, pp. 401–402, 1994.

[15] Y. Suzuki, H. Shimawaki, Y. Amamiya, N. Nagano, T. Niwa, H. Yano, and K. Honjo, "50-Ghz bandwidth baseband amplifiers using GaAs-based HBT's", *IEEE Journal of Solid-State Circuits*, vol. 33, no. 9, pp. 1336–1340, 1998.

[16] H. Suzuki, K. Watanabe, K. Ishikawa, H. Masuda, K. Ouchi, T. Tanoue, and R. Takeyari, "Very-high-speed InP/InGaAs HBT IC's for optical transmission systems", *IEEE Journal of Solid-State Circuits*, vol. 33, no. 9, pp. 1313–1319, 1998.

[17] T. Otsuji, M. Yoneyama, Y. Imai, T. Enoki, and Y. Umeda, "A 64-Gbit/s multiplexer IC using InAlAs/InGaAs/InP HEMT's", *Electronics Letters*, vol. 33, no. 17, pp. 1488–1489, 1997.

[18] T. Otsuji, K. Murata, T. Enoki, and Y. Umeda, "An 80-Gbit/s multiplexer IC using InAlAs/InGaAs/InP HEMT's", *IEEE Journal of Solid-State Circuits*, vol. 33, no. 9, pp. 1321–1326, 1998.

[19] J.R. Long and M.A. Copeland, "A 1.9 GHz low-voltage silicon bipolar receiver front-end for wireless personal communications systems", *IEEE Journal of Solid-State Circuits*, vol. 30, no. 12, pp. 1438–1448, 1995.

[20] J. Durec, "An integrated silicon bipolar receiver subsystem for 900 MHz ISM band applications", in *Proceedings of the 1997 Bipolar/BiCMOS Circuits and Technology Meeting*, 1997, pp. 57–60.

[21] H.D. Wohlmuth, W. Simburger, H. Knapp, and A.L. Scholtz, "2 GHz Meissner VCO in Si bipolar technology", in *Proceedings of 29th European Microwave Conference*, 1999, vol. 1, pp. 190–193.

[22] G.L. Puma, K. Hadjizada, S. van Waasen, C. Grewing, P. Schrader, W. Geppert, A. Hanke, M. Seth, and S. Heinen, "An RF transceiver for digital wireless communication in a 25 GHz Si bipolar technology", in *2000 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, 2000, pp. 144–145.

[23] A.N. Karanicolas, "A 2.7-V 900-MHz CMOS LNA and mixer", *IEEE Journal of Solid-State Circuits*, vol. 31, no. 12, pp. 1939–1944, 1996.

[24] P. Litmanen, P. Ikalainen, and K. Halonen, "A 2.0-GHz submicron CMOS LNA and a downconversion mixer", in *ISCAS '98 Proceedings of the 1998 IEEE International Symposium on Circuits and Systems*, 1998, vol. 4, pp. 357–359.

[25] F. Behbahani, J.C. Leete, Y. Kishigami, A. Roithmeier, K. Hoshino, and A.A. Abidi, "A 2.4-GHz low-IF receiver for wideband WLAN in 0.6-μm CMOS architecture and front-end", *IEEE jounral of Solid-State Circuits*, vol. 35, no. 12, December 2000.

[26] E.A. McShane and K. Shenai, "Technologies and design of low-power RF microsystems", in *2000 22nd International Conference on Microelectronics. Proceedings*, vol.1, 2000, pp. 107–115.

[27] Ting-Ping Liu and E. Westerwick, "5-GHz CMOS radio transceiver front-end chipset", *IEEE Journal of Solid-State Circuits*, vol. 35, no. 12, pp. 1927–1933, 2000.

[28] T. Wakimoto, T. Hatano, C. Yamaguchi, H. Morimura, and S. Konaka, "Sub 1-V 5-GHz-band up- and down-conversion mixer cores in 0.35-μm CMOS", in *2000 Symposium on VLSI Circuits. Digest of Technical Papers*, 2000, pp. 98–9.

[29] H.R. Rategh, H. Samavati, and T.H. Lee, "A CMOS frequency synthesizer with an injection-locked frequency divider for a 5-GHz wireless LAN receiver", *IEEE Journal of Solid-State Circuits*, vol. 35, no. 5, pp. 780–787, 2000.

[30] H. Samavati, H.R. Rategh, and T.H. Lee, "A 5-GHz CMOS wireless LAN receiver front end", *IEEE Journal of Solid-State Circuits*, vol. 35, no. 5, pp. 765–772, 2000.

[31] C. Kermarrec, T. Tewsbury, G. Dawe, R. Baines, B. Meyerson, D. Harame, and M. Gilbert, "SiGe HBTs reach the microwave and millimeter-wave frontier", in *BCTM Tech. Dig.* IEEE, 1994, pp. 155–162.

[32] D. J. Roulston, *Bipolar Semiconductor Devices*, p. 8, McGraw-Hill, 1990.

[33] J. C. Bean, "Silicon-based semiconductor heterostructures: column IV bandgap engineering", *Proc. IEEE*, vol. 80, no. 4, pp. 571–587, 1992.

[34] J. C. Bean, T. T. Sheng, L. C. Feldman, A. T. Fiory, and R. T. Lynch, "Pseudomorphic growth of Ge_xSi_{1-x} on silicon by molecular beam epitaxy", *Appl. Phys. Lett.*, vol. 44, no. 1, pp. 102–104, January 1984.

[35] S.S. Iyer, G.L. Patton, S.L. Delage, S. Tiwari, and J.M.C. Stork, "Silicon-germanium base heterojunction bipolar transistors by molecular beam epitaxy", in *Proceedings of the Second International Symposium on Silicon Molecular Beam Epitaxy*, J.C. Bean and L.J Schowalter, Eds., Honolulu, October 1987, pp. 114–125.

[36] H. Temkin, "Ge_xSi_{1-x} optoelectronic devices", in *Proceedings of the Second International Symposium on Silicon Molecular Beam Epitaxy*, J.C. Bean and L.J Schowalter, Eds., Honolulu, October 1987, pp. 28–35.

[37] S. S. Iyer, G. L. Patton, S. L. Delage, S. Tiwari, and J. M. C. Stork, "Silicon-germanium heterojunction bipolar transistors by molecular beam epitaxy", in *IEDM Technical Digest*. IEEE, 1987, pp. 874–876.

[38] T. Tatsumi, H. Hirayama, and N. Aizaki, "Si/$Ge_{0.3}Si_{0.7}$/Si heterojunction bipolar transistor made with Si molecular beam epitaxy", *Appl. Phys. Lett.*, vol. 52, no. 11, pp. 895–897, March 1988.

[39] H. Temkin, J. C. Bean, A. Antreasyan, and R. Leibenguth, "Ge_xSi_{1-x} strained-layer heterostructure bipolar transistors", *Appl. Phys. Lett.*, vol. 52, no. 13, pp. 1089–1091, March 1988.

[40] G. L. Patton et al., "Silicon-Germanium-Base heterojunction bipolar transistors by molecular beam epitaxy", *IEEE Electron Device Letters*, vol. 9, no. 4, pp. 165–167, April 1988.

[41] B. S. Meyerson, K. J. Uram, and F. K. Legoues, "Cooperative growth phenomena in silicon/germanium low-temperature epitaxy", *Appl. Phys. Lett.*, vol. 53, pp. 2555–2557, December 1988.

[42] G. L. Patton, D. J. Harame, J. M. C. Stork, B. S. Meyerson, G. J. Scilla, and E. Ganin, "SiGe-base, poly-emitter heterojunction bipolar transistors", in *Digest of technical papers / Symposium on VLSI technology.* IEEE, May 1989, pp. 95–96.

[43] G. L. Patton, D. Harame, J. M. C. Stork, B. S. Meyerson, G. J. Scilla, and E. Ganin, "Graded-SiGe-Base, poly-emitter heterojunction bipolar transistors", *IEEE Electron Device Lett.*, vol. 10, no. 12, pp. 534–536, December 1989.

[44] S.E. Fischer, R.K. Cook, R.W. Knepper, R.C. Lange, K. Nummy, D.C. Ahlgren, M. Revitz, and B.S. Meyerson, "A 45 GHz strained-layer SiGe heterojunction bipolar transister fabricated with low temperature epitaxy", in *International Electron Devices Meeting 1989. Technical Digest*, 1989, pp. 890–892.

[45] C. A. King, J. L. Hoyt, C. M. Gronet, J. F. Gibbons, M. P. Scott, and J. Turner, "$Si/Si_{1-x}Ge_x$ heterojunction bipolar transistors produced by limited reaction processing", *IEEE Electron Device Letters*, vol. 10, no. 2, pp. 52–54, February 1989.

[46] A. Pruijmboom, C. E. Timmering, J. Van Rooij-Mulder, D. Gravesteijn, W. de Boer, et al., "Heterojunction bipolar transistors with $Si_{1-x}Ge_x$ base", *Microelectronic Engineering*, vol. 19, pp. 427–434, 1992.

[47] G. L. Patton, J. H. Comfort, B. S. Meyerson, E. F. Crabbe, G. J. Scilla, E. de Fresart, J. M. C. Stork, J. Y.-C. Sun, D. L. Harame, and J. N. Burghartz, "75-GHz f_T SiGe-Base heterojunction bipolar transistors", *IEEE Electron Device Letters*, vol. 11, no. 4, pp. 171–173, April 1990.

[48] E. F. Crabbe, B. S. Meyerson, J. M. C. Stork, and D. L. Harame, "Vertical profile optimization of very high frequency epitaxial Si- and SiGe-base bipolar transistors", in *IEDM Technical Digest.* IEEE, 1993, pp. 83–86.

[49] D. L. Harame, E. F. Crabbe, J. D. Cressler, J. H. Comfort, J. Y. C. Sun, S. R. Stiffer, E. Kobeda, J. N. Burghartz, M. M. Gilbert, J. C. Malinowski, and A. J. Dally, "A high performance epitaxial SiGe-base ECL BiCMOS technology", in *IEDM Tech. Dig.*, 1992, pp. 19–22.

[50] D. L. Harame, K. Schonenberg, M. Gilbert, D. Nguyen-Ngoc, J. Malinowski, S. J. Jeng, B. Meyerson, J. D. Cressler, R. Groves, G. Berg, K. Tallman, K. Stein, G. Hueckel, C. Kermarrec, T. Tice, G. Fitzgibbons, K. Walter, D. Colavito, T. Houghton, N. Greco, T. Kebede, B. Cunningham, S. Subbanna, J. H. Comfort, and E. F. Crabbe, "A 200 mm SiGe-HBT technology for wireless and mixed-signal applications", in *IEDM Tech. Dig.*, 1994, pp. 437–440.

[51] D. C. Ahlgren, M. Gilbert, D. Greenberg, S. J. Jeng, J. Malinowski, D. Nguyen-Ngoc, K. Schonenberg, K. Stein, K. Walter, G. Hueckel, D. Colavito, G. Freeman, D. Sunderland, D. L. Harame, and B. Meyerson, "Manufacturability demonstration of an integrated SiGe HBT technology for the analog and wireless marketplace", in *IEDM Tech. Dig.* IEEE, 1996, pp. 859–862.

[52] J. H. Comfort et al., "Profile leverage in a self-aligned epitaxial Si or SiGe base bipolar transistors", in *IEDM Tech. Dig.* IEEE, 1990, pp. 21–24.

[53] A. Pruijmboom, D. Terpstra, C. E. Timmering, et al., "Selective-epitaxial base technology with 14 ps ECL-gate delay, for low power wide-band communication systems", in *IEDM Tech. Dig.* IEEE, 1995, pp. 747–750.

[54] D. L. Harame, J. M. C. Stork, K. Y.-J. Hsu, J. Cotte, K. A. Jenkins, J. D. Cressler, P. Restle, E. F. Crabbe, S. Subbanna, T. E. Tice, B. W. Scharf, and J. A. Yasaitis, "Optimization of SiGe HBT technology for high speed analog and mixed-signal applications", in *IEDM Tech. Dig.*, 1993, pp. 71–74.

[55] M. Muller, H.-M. Rein, A. Felder, and T.F. Meister, "60 Gbit/s time-division multiplexer in SiGe-bipolar technology with special regard to mounting and measuring technique", *Electronics Letters*, vol. 33, no. 8, pp. 679–680, 1997.

[56] A. Felder, M. Moller, M. Wurzer, M. Rest, T.F. Meister, and H.-M. Rein, "60 Gbit/s regenerating demultiplexer in SiGe bipolar technology", *Electronics Letters*, vol. 33, no. 23, pp. 1984–1986, 1997.

[57] J.-O. Plouchart, B.-U. Klepser, H. Ainspan, and M. Soyuer, "Fully-monolithic 3 V SiGe differential voltage-controlled oscillators for 5 GHz and 17 GHz wireless applications", in *ESSCIRC '98. Proceedings of the 24th European Solid-State Circuits Conference*, 1998, pp. 332–335.

[58] J.-O. Plouchart, H. Ainspan, and M. Soyuer, "A 5.2 GHz 3.3V SiGe RF transmitter", in *Proceedings of 29th European Microwave Conference*, vol.3, 1999, pp. 279–282.

[59] B. Foley, P. Murphy, and A. Murphy, "A monolithic SiGe 5 GHz low noise amplifier and tuneable image-reject filter for wireless LAN applications", in *2000 High Frequency Postgraduate Student Colloquium*, 2000, pp. 26–31.

[60] J.-O. Plouchart, H. Ainspan, M. Soyuer, and A. Ruehli, "A fully-monolithic SiGe differential voltage-controlled oscillator for 5 GHz wireless applications", in *2000 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium. Digest of Papers*, 2000, pp. 57–60.

[61] S.P. Voinigescu, D. Marchesan, and M.A. Copeland, "A family of monolithic inductor-varactor SiGe-HBT VCOs for 20 GHz to 30 GHz LMDS and fiber-optic receiver applications", in *2000 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium. Digest of Papers*, 2000, pp. 173–177.

[62] M. Rossberg, J. Sachs, P. Rauschenbach, P. Peyerl, K. Pressel, W. Winkler, and D. Knoll, "11 GHz SiGe circuits for ultra wideband radar", in *Proceedings of the 2000 Bipolar/BiCMOS Circuits and Technology Meeting 2000*, 2000, pp. 70–73.

[63] T. Hashimoto, F. Sato, T. Aoyama, H. Suzuki, H. Yoshida, H. Fujii, and T. Yamazaki, "A 73 GHz f_T 0.18 μm RF-SiGe BiCMOS technology considering thermal budget trade-off and with reduced boron-spike effect on HBT characteristics", in *IEDM Tech. Dig.*, 2000, pp. 149–152.

[64] S. Decoutere, F. Vleugels, R. Kuhn, R. Loo, M. Caymax, S. Jenei, J. Croon, S. Van Huylenbroeck, M. Da Rold, E. Rosseel, P. Chevalier, and P. Coppens, "A 0.35 μm SiGe BiCMOS process featuring a 80 GHz f_{max} HBT and integrated high-Q RF passive components", in *Proceedings of the 2000 Bipolar/BiCMOS Circuits and Technology Meeting 2000*, 2000, pp. 106–109.

[65] G. Freeman, D. Ahlgren, D.R. Greenberg, R. Groves, F. Huang, G. Hugo, B. Jagannathan, S.J. Jeng, J. Johnson, K. Schonenberg, K. Stein, R. Volant, and S. Subbanna, "A 0.18 μm 90 GHz f_T SiGe HBT BiCMOS, ASIC-compatible, copper interconnect technology for RF and microwave applications", in *IEDM Tech. Dig.*, 1999, pp. 569–572.

[66] M. Carroll, T. Ivanov, S. Kuehne, J. Chu, C. King, M. Frei, M. Mastrapasqua, R. Johnson, K. Ng, S. Moinian, S. Martin, C. Huang, T. Hsu, D. Nguyen, R. Singh, L. Fritzinger, T. Esry, W. Moller, B. Kane, G. Abeln, D. Hwang, D. Orphee, S. Lytle, M. Roby, D. Vitkavage, D. Chesire, R. Ashton, D. Shuttleworth, M. Thoma, S. Choi, S. Lewellen, P. Mason, T. Lai, H. Hsieh, D. Dennis, E. Harris, S. Thomas, R. Gregor, P. Sana, and W. Wu, "COM2 SiGe modular BiCMOS technology for digital, mixed-signal, and RF applications", in *IEDM Tech. Dig.*, 2000, pp. 145–148.

[67] K. Washio, E. Ohue, H. Shimamoto, K. Oda, R. Hayami, Y. Kiyota, M. Tanabe, M. Kondo, T. Hashimoto, and T. Harada, "A 0.2-μm 180-GHz-f_{max} 6.7-ps-ECL SOI/HRS self-aligned SEG SiGe HBT/CMOS technology for microwave and high-speed digital applications", in *IEDM Tech. Dig.*, 2000, pp. 741–744.

[68] A. Monroy, W. Laurens, M. Marty, D. Dutartre, D. Gloria, J.L. Carbonero, A. Perrotin, M. Roche, and A. Chantre, "BiCMOS6G: a high performance 0.35 μm SiGe BiCMOS technology for wireless applications", in *Proceedings of the 1999 Bipolar/BiCMOS Circuits and Technology Meeting*, 1999, pp. 121–124.

[69] D. V. Lang et al., "Measurement of the bandgap of Ge_xSi_{1-x}/Si strained-

layer heterostructures", *Appl. Phys. Lett.*, vol. 47, no. 12, pp. 1333–1335, December 1985.

[70] R. People and J. C. Bean, "Band alignments of coherently strained Ge_xSi_{1-x}/Si heterostructures on <001> Ge_ySi_{1-y} substrates", *Appl. Phys. Lett.*, vol. 48, no. 8, pp. 538–540, February 1986.

[71] M. A. Green, "Intrinsic concentration, effective densities of states, and effective mass in silicon", *J. Appl. Phys.*, vol. 67, no. 6, pp. 2944–2954, 1990.

[72] S. Subbanna, G. Freeman, D. Ahlgren, D. Greenberg, D. Harame, J. Dunn, D. Herman, B. Meyerson, Y. Greshishchev, P. Schvan, D. Thornberry, G. Sakamoto, and R. Tayrani, "Integration and design issues in combining very-high-speed silicon-germanium bipolar transistors and ULSI CMOS for system-on-a-chip applications", in *IEDM Tech. Dig.*, 1999, pp. 845–848.

[73] T. H. Lee and S. S. Wong, "CMOS RF integrated circuits at 5 GHz and beyond", *Proceedings of the IEEE*, vol. 88, no. 10, pp. 1560–1571, 2000.

[74] L.E. Larson, "High-speed Si/SiGe technology for next generation wireless system applications", *Journal of Vacuum Science & Technology B*, vol. 16, no. 3, pp. 1541–1548, 1998.

[75] L.E. Larson, "Silicon bipolar transistor design and modeling for microwave integrated circuit applications", in *Proceedings of the 1996 Bipolar/BiCMOS Circuits and Technology Meeting*, 1996, pp. 142–148.

[76] L.E. Larson, "Integrated circuit technology options for RFICs-present status and future directions", *IEEE Journal of Solid-State Circuits*, vol. 33, no. 3, pp. 387–399, 1998.

[77] U. Konig, "SiGe and GaAs as competitive technologies for RF-applications", in *Proceedings of the 1998 Bipolar/BiCMOS Circuits and Technology Meeting*, 1998, pp. 87–92.

[78] A. Bandyopadhyay, S. Subramanian, S. Chandrasekhar, A. G. Dentai, and S. M. Goodnick, "Degradation of DC characteristics of InGaAs/InP single heterojunction bipolar transistors under electron irradiation", *IEEE Trans. Electron Devices*, vol. 46, no. 5, pp. 840–849, 1999.

[79] D. J. Roulston, *Bipolar Semiconductor Devices*, p. 176, McGraw-Hill, 1990.

[80] D. J. Roulston, *Bipolar Semiconductor Devices*, p. 241, McGraw-Hill, 1990.

[81] L. J. Varnerin, "Stored charge method of transistor base transit analysis", *Proc. IRE*, vol. 47, pp. 523–527, 1959.

[82] R. G. Meyer and R. S. Muller, "Charge-control analysis of the collector-base space-charge-region contribution to bipolar-transistor time constant τ_T", *IEEE Trans. Electron Devices*, vol. 34, no. 2, pp. 450–452, 1987.

[83] K. J. Negus and D. J. Roulston, "Simplified modelling of delays in the emitter-base junction", *Solid-St. Electron.*, vol. 31, no. 9, pp. 1464–1466, 1988.

[84] D. J. Roulston and F. Hebert, "Study of delay times contributing to the f_t of bipolar transistors", *IEEE Electron Device Lett.*, vol. 7, no. 8, pp. 461–462, 1986.

[85] J. Chen, G. B. Gao, and H. Morkoc, "Comparative analysis of the high-frequency performance of $Si/Si_{1-x}Ge_x$ heterojunction bipolar and Si bipolar transistors", *Solid-St. Electron.*, vol. 35, no. 8, pp. 1037–1044, 1992.

[86] B. Heinemann, K. E. Ehwald, P. Schley, G. Fischer, F. Herzel, D. Knoll, T. Morgenstern, W. Ropke, and W. Winkler, "Collector profile design of SiGe HBTs for optimized static and high-frequency performance", in *ESSDERC*, 1995, pp. 615–618.

[87] A. Joseph, J. Cressler, D. Richey, and G. Niu, "Optimization of SiGe HBT's for operation at high current densities", *IEEE Trans. Electron Devices*, vol. 46, no. 7, pp. 1347–1353, 1999.

[88] G. Zhang, J. Cressler, G. Niu, and A. Pinto, "A comparison of npn and pnp profile design tradeoffs for complementary SiGe HBT technology", *Solid-St. Electron.*, vol. 44, pp. 1949–1954, 2000.

[89] K. H. Kwok, "Analytical expressions of base transit time for SiGe HBTs with retrograde base profiles", *Solid-St. Electron.*, vol. 43, no. 2, pp. 275–283, 1999.

[90] K. H. Kwok and C. R. Selvakumar, "Profile design considerations for minimizing base transit time in SiGe HBTs for all levels of injection before onset of Kirk effect", *IEEE Trans. Electron Devices*, 2001 (accepted for publication).

[91] D. Terpstra, W. B. de Boer, and J. W. Slotboom, "High-performance Si-SiGe HBTs SiGe-Technology developement in ESPRIT project 8001 TIBIA: An overview", *Solid-St. Electron.*, vol. 41, no. 10, pp. 1493–1502, 1997.

[92] A. Schuppen, "SiGe-HBTs for mobile communications", *Solid-St. Electron.*, vol. 43, pp. 1373–1381, 1999.

[93] G. L. Patton, J. M. C. Stork, J. H. Comfort, E. F. Crabbe, B. S. Meyerson, D. L. Harame, and J. Y.-C. Sun, "SiGe-base heterojunction bipolar tran-

sistors: physics and design issues", in *IEDM Tech. Dig.* IEEE, 1990, pp. 13–16.

[94] D. L. Harame, J. H. Comfort, J. D. Cressler, et al., "Si/SiGe epitaxial-base transistors–Part I: materials, physics, and circuits", *IEEE Trans. Electron Devices*, vol. 42, no. 3, pp. 455–468, March 1995.

[95] R. J. E. Hueting, J. W. Slotboom, A. Pruijmboom, W. B. de Boer, C. E. Timmering, and N. E. B. Cowern, "On the optimization of SiGe-base bipolar transistors", *IEEE Trans. Electron Devices*, vol. 43, no. 9, pp. 1518–1523, September 1996.

[96] D. M. Richey, J. D. Cressler, and A. J. Joseph, "Scaling issues and Ge profile optimization in advanced UHV/CVD SiGe HBT's", *IEEE Trans. Electron Devices*, vol. 44, no. 3, pp. 431–440, March 1997.

[97] D. J. Roulston and J. M. McGregor, "Effect of bandgap gradient in the base region of SiGe heterojunction bipolar transistors", *Solid-St. Electron.*, vol. 35, no. 7, pp. 1019–1020, 1992.

[98] I. R. C. Post, P. Ashburn, and G. R. Wolstenholme, "Polysilicon emitters for bipolar transistors: A review and re-evaluation of theory and experiment", *IEEE Trans. Electron Devices*, vol. 39, no. 7, pp. 1717–1730, July 1992.

[99] C. R. Selvakumar, "Theoretical and experimental aspects of polysilicon emitter bipolar transistors", in *Polysilicon emitter bipolar transistors*, A. K. Kapoor and D. J. Roulston, Eds., pp. 3–16. IEEE Press, 1989.

[100] H. C. DeGraaff and J. G. DeGroot, "The SIS tunnel emitter: a theory for

emitters with thin interface layers", *IEEE Trans. Electron Devices*, vol. 26, no. 11, pp. 1771–1776, November 1979.

[101] G. W. Bakker, D. J. Roulston, and A. A. Eltoukhy, "Effective recombination velocity of polysilicon contacts for bipolar transistors", *Electron. Lett.*, pp. 622–624, 1984.

[102] C. C. Ng and E. S. Yang, "A thermionic-diffusion model of polysilicon emitter", in *IEDM Tech. Dig.* IEEE, 1986, pp. 32–35.

[103] T. H. Ning and R. D. Issac, "Effect of emitter contact on current gain of silicon bipolar devices", *IEEE Trans. Electron Devices*, vol. 27, no. 11, pp. 2051–2055, November 1980.

[104] A. Neugroschel, M. Arienzo, Y. Komem, and R. D. Issac, "Experimental study of the minority-carrier transport at the polysilicon monosilicon interface", *IEEE Trans. Electron Devices*, vol. 32, pp. 807–816, April 1985.

[105] A. A. Elthoukhy and D. J. Roulston, "The role of the interfacial layer in polysilicon emitter bipolar transistors", *IEEE Trans. Electron Devices*, vol. 29, no. 12, pp. 1862–1869, December 1982.

[106] Z. Yu, B. Ricco, and R. W. Dutton, "A comprehensive analytical and numerical model of polysilicon emitter contacts in bipolar transistors", *IEEE Trans. Electron Devices*, vol. 31, no. 6, pp. 773–784, June 1984.

[107] P. Ashburn, D. J. Roulston, and C. R. Selvakumar, "Comparison of experimental and computed results on arsenic- and phosphorus-doped polysilicon emitter bipolar transistors", *IEEE Trans. Electron Devices*, vol. 34, no. 6, pp. 1346–1353, June 1987.

[108] K. Suzuki, "Unified minority-carrier transport equation for polysilicon or heteromaterial emitter contact bipolar transistors", *IEEE Trans. Electron Devices*, vol. 38, no. 8, pp. 1868–1877, August 1991.

[109] N. F. Rinaldi, "On the modeling of polysilicon emitter bipolar transistors", *IEEE Trans. Electron Devices*, vol. 44, no. 3, pp. 395–403, March 1997.

[110] K. Suzuki, "Emitter and base transit time of polycrystalline silicon emitter contact bipolar transistors", *IEEE Trans. Electron Devices*, vol. 38, no. 11, pp. 2512–2518, November 1991.

[111] L. M. Castaner, S. Sureda, D. Bardes, and R. Alcubilla, "A compact charge ratio expression for the emitter delay of polysilicon emitter bipolar transistors", *IEEE Trans. Electron Devices*, vol. 41, no. 3, pp. 454–455, March 1994.

[112] S. Basu, R. N. Mitra, and A. N. Daw, "On the transit time of polysilicon emitter transistors", *Solid-St. Electron.*, vol. 43, pp. 189–191, 1999.

[113] C. C. Chang, T. C. Lu, and J. B. Kuo, "A closed-form analytical poly-emitter BJT forward transit time model considering bandgap-narrowing effects and concentration-dependent diffusion coefficients for poly-emitter BJT devices", *Solid-St. Electron.*, vol. 35, no. 10, pp. 1556–1560, 1992.

[114] T. C. Lu, H. P. Chen, and J. B. Kuo, "SiGe-base heterojunction bipolar transistors:an analytical current gain and forward transit time model", *Solid-St. Electron.*, vol. 36, no. 9, pp. 1313–1320, 1993.

[115] Y. Chyan, C. Chang, S. Sze, M. Lin, K. Liao, and R. Reif, "Analytical

modeling of polycrystalline silicon emitter bipolar transistors under high-level injection", *Solid-St. Electron.*, vol. 37, no. 8, pp. 1521–1529, 1994.

[116] P. Ma, L. Zhang, and Y. Wang, "Analytical model for high injection in Si/SiGe heterojunction bipolar transistors (HBT)", *Solid-St. Electron.*, vol. 41, no. 6, pp. 913–916, 1997.

[117] D. J. Roulston, *Bipolar Semiconductor Devices*, p. 196, McGraw-Hill, 1990.

[118] D. J. Roulston, *Bipolar Semiconductor Devices*, p. 201, McGraw-Hill, 1990.

[119] E. L. Heasell, "Recombination beneath ohmic contacts and adjacent oxide covered regions", *Solid-St. Electron.*, vol. 22, pp. 89–93, 1979.

[120] P. Ashburn and B. Soerowirdjo, "Comparison of experimental and theoretical results on polysilicon emitter bipolar transistors", *IEEE Trans. Electron Devices*, vol. 31, no. 7, pp. 853–860, July 1984.

[121] D. B. M. Klaassen, "A unified mobility model for device simulation–I. Model equations and concentration dependence", *Solid-St. Electron.*, vol. 35, no. 7, pp. 953–959, 1992.

[122] J. Y. W. Seto, "The electrical properties of polycrystalline silicon films", *J. Appl. Phys.*, vol. 46, no. 12, pp. 5247–5254, December 1975.

[123] D. M. Kim, F. Qian, C. U. Bickford, and S. Yu, "Resistivity and carrier mobility in heavily doped polycrystalline silicon thin films", in *Mat. Res. Soc. Symp. Proc.* MRS, 1988, vol. 106, pp. 261–266.

[124] D-L. Chen, D. W. Greve, and A. M. Guzman, "Minority-carrier hole diffusion length in heavily-doped polysilicon and its influence on polysilicon-emitter

transistors", *IEEE Trans. Electron Devices*, vol. 35, no. 7, pp. 1045–1053, July 1988.

[125] D. B. M. Klaassen, "A unified mobility model for device simulation–II. Temperature dependence of carrier mobility and lifetime", *Solid-St. Electron.*, vol. 35, no. 7, pp. 961–967, 1992.

[126] D. M. Caughey and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field", *Proc. IRE*, vol. 55, pp. 2192–2193, 1967.

[127] H. J. J. DeMan, "The influence of heavy doping on the emitter efficiency of a bipolar transistor", *IEEE Trans. Electron Devices*, vol. 18, no. 10, pp. 833–835, 1971.

[128] H. C. DeGraaff, J. W. Slotboom, and A. Schmitz, "The emitter efficiency of bipolar transistors", *Solid-St. Electron.*, vol. 20, pp. 515–521, 1977.

[129] C. R. Selvakumar, "Simple general analytical solution to the minority carrier transport in heavily doped semiconductors", *J. Appl. Phys.*, vol. 56, no. 12, pp. 3476–3478, 1984.

[130] J. A. del Alamo and R. M. Swanson, "The physics and modeling of heavily doped emitters", *IEEE Trans. Electron Devices*, vol. 31, no. 12, pp. 1878–1888, 1984.

[131] R. P. Mertens, R. J. van Overstraeten, and J. J. de Man, "Heavy doping effects in silicon", *Advances in electronics and electron physics*, vol. 55, pp. 77–118, 1981.

[132] D. B. M. Klaassen, J. W. Slotboom, and H. C. DeGraaff, "Unified apparent

bandgap narrowing in n- and p-type silicon", *Solid-St. Electron.*, vol. 35, no. 2, pp. 125–129, 1992.

[133] M. A. Shibib, F. A. Lindholm, and F. Therez, "Heavily doped transparent-emitter regions in junction cells, diodes, and transistors", *IEEE Trans. Electron Devices*, vol. 26, no. 6, pp. 959–965, June 1979.

[134] K. Suzuki and N. Nakayama, "Base transit time of shallow-base bipolar transistors considering velocity saturation at base-collector junction", *IEEE Trans. Electron Devices*, vol. 39, no. 3, pp. 623–628, March 1992.

[135] H. Kroemer, "Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region", *Solid-St. Electron.*, vol. 28, no. 11, pp. 1101–1103, 1985.

[136] J. S. Hamel, "An Accurate Charge Control Approach for Modeling Excess Phase Shift in the Base Region of Bipolar Transistors", *IEEE Trans. Electron Devices*, vol. 43, no. 7, pp. 1092–1098, July 1996.

[137] J. S. Hamel, "Compact Modeling of the Influence of Emitter Stored Charge on the High Frequency Small Signal A.C. Response of Bipolar Transistors Using Quasi-static Parameters", *IEEE J. Solid-State Circuits*, vol. 31, no. 1, pp. 106–113, January 1996.

[138] J. K. Gummel, "A charge control relation for bipolar transistors", *Bell Syst. Tech. J.*, pp. 115–120, January 1970.

[139] R. Beaufoy and J. Sparkes, "The junction transistor as a charge-controlled device", *ATE J.*, vol. 13, pp. 310–327, 1957.

[140] H. K. Gummel and J. C. Poon, "An integral charge-control model of bipolar transistors", *Bell Syst. Tech. J.*, vol. 49, pp. 827–852, 1970.

[141] I. E. Getreu, *Modeling the bipolar transistor*, pp. 69–126, Elsevier Scientific Publishing Company, 1978.

[142] J. G. Fossum and S. Veeraraghavan, "Partitioned-Charge-Based Modeling of Bipolar Transistors for Non-Quasi-Static Circuit Simulation", *IEEE Electron Device Lett.*, vol. 7, no. 12, pp. 652–654, December 1986.

[143] J. A. Seitchik, A. Chatterjee, and P. Yang, "An Accurate Bipolar Model for Large Signal Transient and AC Applications", in *IEDM Tech. Dig.*, 1987, pp. 244–247.

[144] M. K. Chen, F. A. Lindholm, and B. S. Wu, "Comparison and Extension of Recent One Dimensional Bipolar Transistor Models", *IEEE Trans. Electron Devices*, vol. 35, no. 5, pp. 1096–1106, July 1988.

[145] J. Lindmayer and C. Y. Wrigley, *Fundamentals of semiconductor devices*, pp. 42–50, Affiliated East-West Press Pvt. Ltd., East-West edition, 1965.

[146] H. Klose and A. W. Wieder, "The Transient Integral Charge Control Relation - A Novel formulation of Currents in a Bipolar Transistor", *IEEE Trans. Electron Devices*, vol. 34, no. 5, pp. 1090–1099, 1987.

[147] R. J. McDonald, "Generalized Partitioned-Charge-Based Bipolar Transistor Modeling Methodology", *Electron. Lett.*, vol. 24, no. 21, pp. 1302–1303, 1988.

[148] J. S. Hamel and C. R. Selvakumar, "The General Transient Charge Control Relation: A New Charge Control Relation for Semiconductor Devices", *IEEE Trans. Electron Devices*, vol. 38, no. 6, pp. 1467–1476, 1991.

[149] G. A. M. Hurkx, "A New Approach to A. C. Characterization of Bipolar Transistors", *Solid-St. Electron.*, vol. 31, no. 8, pp. 1269–1275, 1988.

[150] B. S. Wu and F. A. Lindholm, "One-Dimensional Non-Quasi-Static Models for Arbitrarily and Heavily Doped Quasi-Neutral Layers in Bipolar Transistors", *IEEE Trans. Electron Devices*, vol. 36, no. 4, pp. 727–737, April 1989.

[151] J. S. Hamel, "Integral relations for determining non-quasi-static charge partitioning in bipolar devices from static charge distribution", *IEEE Trans. Electron Devices*, vol. 40, no. 9, pp. 1713–1716, September 1993.

[152] N. F. Rinaldi, "Modeling of small-signal minority-carrier transport in bipolar devices at arbitrary injection levels", *IEEE Trans. Electron Devices*, vol. 45, no. 7, pp. 1501–1510, July 1998.

[153] R. F. Rinaldi, H. C. DeGraaff, and J .L. Tauritz, "Analysis and Modeling of Small-Signal Bipolar Transistor Operation at Arbitrary Injection Levels", *IEEE Trans. Electron Devices*, vol. 45, no. 8, pp. 1817–1825, August 1998.

[154] D. E. Thomas and J. L. Moll, "Junction transistor short-circuit current gain and phase determination", *Proc. IRE*, vol. 46, pp. 1177–1184, 1958.

[155] J. te Winkel, "Extended charge-control model for bipolar transistors", *IEEE Trans. Electron Devices*, vol. 20, no. 4, pp. 389–394, 1973.

[156] J. S. Hamel, "An accurate quasi-static method for determining the excess phase shift in the base region of bipolar transistors", in *IEEE Bipolar Circuits and Technology Meeting*. IEEE, 1992, pp. 208–211.

[157] J. S. Hamel and D. J. Roulston, "Impact of the emitter and base diffusion

capacitances on the a.c. behavior of bipolar transistors", *Solid-St. Electron.*, vol. 36, no. 8, pp. 1175–1182, 1993.

[158] J. S. Hamel, "A quasi-static approach for modeling the influence of emitter stored charge on the high frequency small signal A.C. response of bipolar transistors", in *IEEE Bipolar Circuits and Technology Meeting.* IEEE, 1994, pp. 44–47.

[159] J. S. Hamel, "Techniques for non-quasi-static analysis of bipolar transistors", Master's thesis, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1, 1989.

[160] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors", *Bell Syst. Tech. J.*, vol. 28, pp. 435–489, July 1949.

[161] J. W. Slotboom and H. C. DeGraaff, "Measurements of bandgap narrowing in Si bipolar transistors", *Solid-St. Electron.*, vol. 19, pp. 857–862, 1976.

[162] J. del Alamo, S. E. Swirhun, and R. M. Swanson, "Simultaneous measurement of hole lifetime, hole mobility and bandgap narrowing in heavily doped n-type silicon", in *IEDM Technical Digest.* IEEE, 1985, pp. 290–293.

[163] S. E. Swirhun, Y.-H. Kwark, and R. M. Swanson, "Measurement of electron lifetime, electron mobility and band-gap narrowing in heavily doped p-type silicon", in *IEDM Technical Digest.* IEEE, 1986, pp. 24–27.

[164] J. S. Hamel, "Non-quasi-static modeling of neutral region A.C. photo currents in integrated photodiode detectors", *Solid-St. Electron.*, vol. 37, no. 3, pp. 473–479, 1994.

[165] C. R. Selvakumar, K. H. Kwok, and John. S. Hamel, "A simple general analytical solution to the 1D A.C. continuity equation and its use in non-quasi-static analysis of bipolar transistors and HBTs", *IEEE Trans. Electron Devices*, December 2000 (submitted).

[166] H. C. Lin, "Depletion-layer calculatons of a double-diffused junction", *IEEE Trans. Electron Devices*, vol. 27, no. 9, pp. 1839–1841, September 1980.

[167] G. L. Patton, J. C. Bravman, and J. D. Plummer, "Physics, technology and modeling of polysilicon emitter contacts for VLSI bipolar transistors", *IEEE Trans. Electron Devices*, vol. 33, no. 11, pp. 1754–1768, November 1986.

[168] J. L. Moll and I. M. Ross, "The dependence of transistor parameters on the distribution of base layer resistivity", *Proc. IRE*, vol. 44, pp. 72–78, January 1956.

[169] J. Lindmayer and C. Wrigley, "The high-injection-level operation of drift transistors", *Solid-St. Electron.*, vol. 2, pp. 79–84, 1961.

[170] S. Szeto and R. Reif, "Reduction of f_t by nonuniform base bandgap narrowing", *IEEE Trans. Electron Devices*, vol. 10, no. 8, pp. 341–343, August 1989.

[171] K. Suzuki, "Optimum base doping profile for minimum base transit time", *IEEE Trans. Electron Devices*, vol. 38, no. 9, pp. 2128–2133, September 1991.

[172] T. C. Lu and J. B. Kuo, "A closed-form analytical BJT forward transit time model considering bandgap-narrowing effects and concentration-dependent diffusion coefficients", *Solid-St. Electron.*, vol. 35, no. 9, pp. 1374–1377, 1992.

[173] M. M. Jahan and A. F. M. Anwar, "An analytical expression for base transit time in an exoponentially doped base bipolar transistor", *Solid-St. Electron.*, vol. 39, no. 1, pp. 133–136, 1996.

[174] N. Rinaldi, "Analytical relations for the base transit time and collector current in BJTs and HBTs", *Solid-St. Electron.*, vol. 41, no. 8, pp. 1153–1158, 1997.

[175] G.-B. Gao and H. Morkoc, "Base Transit Time for SiGe-Base Heterojunction Bipolar Transistors", *Electron. Lett.*, vol. 27, no. 16, pp. 1408–1410, August 1991.

[176] M. S. Lundstrom, "An Ebers-Moll model for the heterostructure bipolar transistors", *Solid-St. Electron.*, vol. 29, no. 11, pp. 1173–1179, 1986.

[177] R. Zuleeg and K. Lehovec, "Temperature dependence of the saturation current of MOST's", *IEEE Trans. Electron Devices*, vol. 15, no. 12, pp. 987–989, 1968.

[178] L. M. Dang and M. Konaka, "A two-dimensional computer analysis of triode-like characteristics of short-channel MOSFETs", *IEEE Trans. Electron Devices*, vol. 27, no. 8, pp. 1533–1539, 1980.

[179] TMA, *Medici user's manual*, Avant, 1997.

[180] S. Winterton, C. Peters, and N. Tarr, "Composition grading for base transit time minimization in SiGe-baseheterojunction bipolar transistors", *Solid-St. Electron.*, vol. 36, no. 8, pp. 1161–1164, 1993.

[181] P. Rinaldi and N. Rinaldi, "Composition grading for base transit time minimization in HBTs: An analytical approach", *Solid-St. Electron.*, vol. 41, no. 1, pp. 59–66, 1997.

[182] J. Cressler, E. Crabbe, J. Comfort, J. Stork, and J. Sun, "On the profile design and optimization of epitaxial Si and SiGe-base bipolar transistor technology for 77K applications – Part II: Circuit performance issues", *IEEE Trans. Electron Devices*, vol. 40, pp. 542–556, 1993.

[183] J. Song and J. Yuan, "Comments on "On the base profile design and optimization of epitaxial Si- and SiGe-base bipolar technology for 77K applications – Part II: Circuit performance issues"", *IEEE Trans. Electron Devices*, vol. 44, no. 5, pp. 915–917, May 1997.

[184] J. Song and J. Yuan, "Optimum Ge profile for base transit time minimization of SiGe HBT", *Solid-St. Electron.*, vol. 41, no. 12, pp. 1957–1959, 1997.

[185] V. Patri and M. Kumar, "Profile design considerations for minimizing base transit time in SiGe HBT's", *IEEE Trans. Electron Devices*, vol. 45, no. 8, pp. 1725–1730, 1998.

[186] J. S. Hamel, "Comments on "A new approach to optimizing the base profile for high-speed bipolar transistors", *IEEE Trans. Electron Devices*, vol. 12, no. 2, pp. 81–82, February 1991.

[187] K. Suzuki, "Analytical base transit time model of uniformly-doped-base bipolar transistor for high-injection regions", *Solid-St. Electron.*, vol. 36, no. 1, pp. 109–110, 1993.

[188] P. Ma, L. Zhang, and Y. Wang, "Analytical relations pertaining to collector current density and base transit time in bipolar transistors", *Solid-St. Electron.*, vol. 39, no. 1, pp. 173–175, 1996.

[189] N. F. Rinaldi, "Modeling of minority-carrier transport in semiconductor regions with position-dependent material parameters at arbitrary injection levels", *IEEE Trans. Electron Devices*, vol. 43, no. 8, pp. 1256–1263, August 1996.

[190] D. Bardes and R. Alcubilla, "Analytical modelling of BJT neutral base region under variable injection conditions", *Solid-St. Electron.*, vol. 41, no. 8, pp. 1177–1180, 1997.

[191] P. Ma, S. Zhang, and M. Ostling, "A new set of initial conditions for fast and accurate calculation of base transit time and collector current density in bipolar transistors", *Solid-St. Electron.*, vol. 42, no. 11, pp. 2023–2036, 1998.

[192] Y. Yue, J. Liou, A. Ortiz-Conde, and F. Sanchez, "Effects of high-level free-carrier injection on the base transit time of bipolar junction transistors", *Solid-St. Electron.*, vol. 39, no. 1, pp. 27–31, 1996.

[193] K. Suzuki, "Analytical base transit model for high-injection regions", *Solid-St. Electron.*, vol. 37, no. 3, pp. 487–493, 1994.

[194] P. Ma, L. Zhang, and Y. Wang, "Analytical model of collector current density and base transit time based on iteration method", *Solid-St. Electron.*, vol. 39, no. 11, pp. 1683–1686, 1996.

[195] J. Yuan, "Effect of base profile on the base transit time of the bipolar tran-

sistor for all levels of injection", *IEEE Trans. Electron Devices*, vol. 41, no. 2, pp. 212–216, 1994.

[196] B. S. Wu and F. A. Lindholm, "One-dimensional all injection nonquasic-static models for arbitrarily doped quasi-neutral layers in bipolar junction transistors including plasma-induced energy-gap narrowing", *IEEE Trans. Electron Devices*, vol. 37, no. 1, pp. 250–261, January 1990.

[197] J. R. Hauser, "The effects of distributed base potential on emitter-current injection density and effective base resistance for stripe transistor geometries", *IEEE Trans. Electron Devices*, vol. 11, pp. 238–242, May 1964.

[198] C. T. Kirk, "A theory of transistor cutoff frequency (f_T) falloff at high current densities", *IEEE Trans. Electron Devices*, vol. 11, pp. 164–174, March 1962.

[199] A. van der Ziel and D. Agouridis, "The cutoff frequency falloff in UHF transistors at high currents", *IEEE Proc. Letts.*, vol. 54, pp. 411–412, 1966.

[200] M. A. Stettler and M. S. Lundstrom, "A microscopic study of transport in thin base silicon bipolar transistors", *IEEE Trans. Electron Devices*, vol. 41, no. 6, pp. 1027–1033, June 1994.

[201] A. Van Der Ziel, "Boundary conditions for forward biased p-n junctions", *Solid-State Electron.*, vol. 16, pp. 1509–1511, 1973.

[202] N. H. Fletcher, "General Semiconductor junction relations", *J. Electron*, vol. 2, pp. 609–610, May 1957.

[203] B. L. Grung and R. M. Warner, "An analytical model for the epitaxial bipolar transistor", *Solid-State Electron.*, vol. 20, pp. 753–771, 1977.

[204] D. J. Roulston, "Early voltage in very-narrow-base bipolar transistor", *IEEE Electron Device Lett.*, pp. 88–89, February 1990.

[205] S. Sokolic and S. Amon, "Physics and modelling carrier transport in the base of npn SiGe HBTs", in *Fourth International Symp. Low temperature Electronics and high temperature superconductors*. Electrochemical Society, 1997, pp. 254–269.

[206] S. Sokolic and S. Amon, "Characterization of minority carrier concentration in the base of npn SiGe HBT", in *ESSDERC*, 1996, pp. 657–660.

[207] A. Neugroschel, J. S. Wang, and F. A. Lindholm, "Evidence for excess carrier storage in electron-hole plasma in silicon transistors", *IEEE Electron Device Letters*, vol. 6, no. 5, pp. 253–255, 1985.

[208] K. K. Thornber, "Relation of drift velocity to low-field mobility and high-field saturation velocity", *J. Appl. Phys.*, vol. 51, no. 4, pp. 2127–2136, 1980.

[209] S. Decoutere, J. Poortmans, L. Deferm, and J. Nijs, "Investigation of the high frequency noise figure reduction of SiGe heterojunction bipolar transistors using actualised physical models", *Solid-St. Electron.*, vol. 38, no. 1, pp. 157–162, 1995.

[210] N. Rinaldi, "All-injection modeling of collector current and transit time in SiGe bipolar transistors", in *ESSDERC*, 1995, pp. 785–788.

[211] A. Gruhle, A. Schuppen, U. Konig, U. Erben, and H. Schumacher, "Monolithic 26 GHz and 40 GHz VCOs with SiGe heterojunction bipolar transistors", in *IEDM Tech. Dig.* IEEE, 1995, pp. 725–728.

[212] G. Parker, P. Ashburn, J. Bonar, H. Gregory, G. Kennedy, and J. Hamel, "LPCVD SiGe for heterojunction bipolar transistors", in *IEE Colloquium on Advanced MOS and Bi-polar Devices.* IEE, 1995, pp. 1–5.

[213] P. Hemment, F. Cristiano, A. Nejim, S. Lombardo, K. Larssena, F. Priolo, and R. Barklie, "Ge^+ ion implantation – a competing technology?", *J. Crystal Growth*, vol. 157, pp. 147–160, 1995.

[214] A. Fukami, K. Shoji, T. Nagano, T. Tokuyama, and C. Yang, "Graded-bandgap SiGe bipolar transistor fabricated with germanium ion implantation", *Microelectronic Engineering*, vol. 15, pp. 15–18, 1991.

[215] A. Gupta et al., "Characterization of germanium implanted $Si_{1-x}Ge_x$ layer", *J. Elec. Mater.*, vol. 22, no. 1, pp. 125–128, 1993.

[216] T. Ma, "SiGe base heterojunction bipolar transistor", Master's thesis, Univeristy of Waterloo, Waterloo, Ontario, Canada N2L 3G1, 1993.

[217] S. Lombardo, A. Pinto, V. Raineri, P. Ward, and S. U. Campisano, "Si/Ge_xSi_{1-x} HBTs with Ge_xSi_{1-x} base formed by high dose Ge implantation in Si", in *IEDM Technical Digest.* IEEE, 1995, pp. 1019–1021.

[218] S. Lombardo, V. Raineri, R. Portoghese, S. U. Campisano, A. Pinto, G. Rosa, and P. Ward, "Characterization of Si/Ge_xSi_{1-x} heterojunction bipolar transistors formed by Ge ion implantation in Si", *Nuclear Instruments and Methods in Physics Research B*, vol. 120, pp. 169–172, 1996.

[219] S. Lombardo, A. Pinto, V. Raineri, P. Ward, G. La Rosa, G. Privitera, and S. U. Campisano, "Si/Ge_xSi_{1-x} heterojunction bipolar transistors with the

Ge_xSi_{1-x} base formed by Ge ion implantation in Si", *IEEE Electron Device Letters*, vol. 17, no. 10, pp. 485–487, October 1996.

[220] S. Lombardo, V. Privitera, A. Pinto, P. Ward, G. Rosa, and S. Campisano, "Band-gap narrowing and high-frequency characteristics of Si/Ge_xSi_{1-x} heterojunction bipolar transistors formed by Ge ion implantation in Si", *IEEE Trans. Electron Devices*, vol. 45, no. 7, pp. 1531–1537, 1998.

[221] S. Lombardo, C. Spinella, S. Campisano, A. Pinto, and P. Ward, "Si/Ge_xSi_{1-x} heterojunction bipolar transistors formed by Ge ion implantation in Si. Narrowing of band gap and base width", *Nuclear Instruments and Methods in Physics Research B*, vol. 147, pp. 56–61, 1999.

[222] M. Mitchell, S. Nigrin, F. Cristiano, P. Ashburn, and P. Hemment, "Characterisation of NPN and PNP SiGe heterojunction bipolar transistors fomed by Ge^+ implantation", in *High performance electron devices for microwave and optoelectronics applications*. IEEE, 1999, pp. 254–259.

[223] M. E. Lunnon and J. T. Chen, "Furnace and rapid thermal annealing of p^+/n junctions in BF_2-implanted silicon", *J. Electrochem. Soc.*, vol. 132, no. 10, pp. 2473–2475, October 1985.

[224] S. Lombardo, K. Kyllesbech Larsen, V. Raineri, F. La Via, S. U. Campisano, S. Lagomarsino, and A. Kazimirov, "Characterization of C coimplanted Ge_xSi_{1-x} epitaxial layers formed by high dose Ge ion implantation in (100) Si", *J. Appl. Phys.*, vol. 79, no. 7, pp. 3456–3463, 1996.

[225] F. Cristiano, A. Nejim, D. Hope, M. Houlton, and P. Hemment, "Structural studies of ion beam synthesised SiGe/Si heterostructures for HBT applica-

tions", *Nuclear Instruments and methods in Physics Research B*, vol. 112, pp. 311–315, 1996.

[226] Z. Xia, E. Ristolainen, and P. Holloway, "SiGe/Si heterostructures produced by double-energy Si^+ and Ge^+, and Ge^+ and Ge^{2+} ion implantations", *J. Vac. Sci. Technol. B*, vol. 15, no. 5, pp. 2629–2632, 1998.

[227] K. S. Jones, S. Prussin, and E. R. Weber, "A systematic analysis of defects in ion-implanted silicon", *Appl. Phys. A*, vol. 45, pp. 1–34, 1988.

[228] K. Shoji, A. Fukami, T. Nagano, T. Tokuyama, and C. Yang, "Improved crystalline quality of $Si_{1-x}Ge_x$ formed by low- temperature germanium ion implantation", *Appl. Phys. Lett.*, vol. 60, no. 4, pp. 451–453, 1992.

[229] D. C. Paine, D. J. Howard, and N. G. Stoffel, "Strain relief in compositionally graded $Si_{1-x}Ge_x$ formed by high dose ion implantation", *J. Elec. Mater.*, vol. 20, no. 10, pp. 735–746, 1991.

[230] S. Sokolic and S. Amon, "Temperature dependent model for hole effective mass in heavily doped p-type SiGe", *J. de Phys. IV*, vol. 6, pp. C3–137–142, 1996.

[231] S. Sokolic and S. Amon, "Influence of Fermi-Dirac Statistics on Collector current of npn SiGe HBT", *J. de Phys. IV*, vol. 6, pp. C3–131–136, 1996.

[232] R. People, "Indirect band gap of coherently strained Ge_xSi_{1-x} bulk alloys on <001> silicon substrates", *Phys. Rev. B*, vol. 32, no. 2, pp. 1405–1408, July 1985.

[233] R. Braunstein, A. Moore, and F. Herman, "Intrinsic optical absorption in germanium-silicon alloys", *Phys. Rev.*, vol. 109, no. 3, pp. 695–710, 1958.

[234] J. Poortmans, S. C. Jain, D. H. J. Totterdell, M. Caymax, J. F. Nijs, R. P. Mertens, and R. Van Overstaeten, "Theoretical calculation and experimental evidence of the real and apparent bandgap narrowing due to heavy doping in p-type Si and strained $Si_{1-x}Ge_x$ layers", *Solid-St. Electron.*, vol. 36, no. 12, pp. 1763–1771, 1993.

[235] A. Souifi, G. Bremond, T. Benyattou, G. Guillot, D. Dutartre, and P. Warren, "Band-gap narrowing determination by photoluminescence on strained B-doped $Si_{0.82}Ge_{0.18}$ layers grown on Si", *Appl. Phys. Lett.*, vol. 62, no. 23, pp. 2986–2988, 1993.

[236] J. del Alamo, S. Swirhun, and R. Swanson, "Measuring and modeling minority carrier transport in heavily doped silicon", *Solid-State Electronics*, vol. 28, pp. 47–54, 1985.

[237] J. Poortmans, M. Caymax, A. Van Ammel, M. Libezny, K. Werner, S. C. Jain, J. Nijs, and R. Mertens, "On the electron minority carrier mobility and the effective bandgap in heterojunction bipolar transistors with strained $Si_{1-x}Ge_x$-base", in *Proc. ESSDERC*, 1993, pp. 317–320.

[238] T. Manku and A. Nathan, "Electron drift mobility model for devices based on unstrained and coherently strained $Si_{1-x}Ge_x$ grown on $< 001 >$ silicon substrate", *IEEE Trans. Electron Devices*, vol. 39, no. 9, pp. 2082–2089, 1992.

[239] J. M. Hinckley, V. Sankaran, and J. Singh, "Charged carrier transport in $Si_{1-x}Ge_x$ pseudomorphic alloys matched to Si-strain-related transport improvements", *Appl. Phys. Lett.*, vol. 55, no. 19, pp. 2008–2010, 1989.

[240] J. Poortmans, S. C. Jain, J. Nijs, and R. Van Overstraeten, "Materials properties of (strained) SiGe layers", in *Advanced silicon and semiconducting silicon-alloy based materials and devices*, J. Nijs, Ed., pp. 185–214. Institute of Physics Pub., 1994.

[241] T. Manku and A. Nathan, "Effective mass for strained p-type $Si_{1-x}Ge_x$", *J. Appl. Phys.*, vol. 69, no. 12, pp. 8414–8416, 1991.

[242] Y. Fu, S. C. Jain, M. Willander, and J. J. Loferski, "Valence band structures of heavily doped strained Ge_xSi_{1-x} layers", *J. Appl. Phys.*, vol. 74, no. 1, pp. 402–407, 1993.

[243] K. H. Kwok and C. R. Selvakumar, "Fabrication process for polyemitter, Ge-implanted base heterojunction bipolar transistors", unpublished report, 2000.