ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACULTY OF TECHNOLOGY

ONLINE SIMULATION OF SILICON-GERMANIUM HETEROJUNCTION BIPOLAR TRANSITOR USING NANOHUB SIMULATOR

A thesis submitted to the School of Graduate Studies of Addis Ababa University in partial fulfillment of the Degree of Masters of Science in Electrical Engineering

By

Damot Tesfaye

Advisor

Prof. Dr. Gerald Higelin

ADDIS ABABA

Ethiopia

February, 2008

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACULTY OF TECHNOLOGY

SIMULATION OF SILICON-GERMANIUM HETEROJUNCTION BIPOLAR TRANSITOR

BY

Damot Tesfaye

Approval by Board of Examiners

Dr. Mengesh Mamo_____ _____

Chairman (Department of Graduate Committee) Signature

Prof. Dr. Gerald Higelin _____

Advisor Signature

Dr. Ing. Mohammed Abdo _____

Internal Examiner Signature

Dr. Genene Tessema _____

External Examiner Signature

DECLARATION

I, the undersigned, hereby declare that this thesis is my original work performed under the supervision of Prof. Dr. Gerald Higelin has not been presented as a thesis for a degree program in any other university and all sources of materials used for the thesis are duly acknowledged.

Name: **Damot Tesfaye**

Signature: ----------

Place: Addis Ababa

Date of submission: ----------

This thesis has been submitted for examination with our approval as university advisor.

Prof. Dr. Gerald Higelin ----------

Signature

Addis Ababa

Ethiopia

February, 2008

Acknowledgments

First of all I would like to thank my research advisor, Professor Dr.Gerald Higelin, for his guidance and support in writing this thesis. I would also like to thank the late advisor and Senior Instructor Prof. Santanam Agore who he had given me valuable suggestions and materials.

I am grateful to the National Science Foundation (NFS) and to the Network for Computational Nanotechnology (NCN) for funding the Nanohub webpage and allowing it freely for researchers, with out which it would have been impossible for me to get the device simulator *PADRE* and finish my thesis work. I thank my friend G/Meskel Teklemariyam for encouraging me to start Postgraduate study program.

I am also very grateful to L/Col Hailemelkot Ambaw for his positive attitude to human beings and for his help to me to reach at this level in giving the valuable time and freedom. Also I am very grateful to L/Col Maharu Hekiso for being a father and understanding person who was there from the start to the end and made everything comfortable. Without their support and the collaboration of colleagues at the office it would have been more challenging, or even impossible to finish the thesis. I also acknowledge Yerksew for her valuable support.

I am very thankful to my friend Tesfaye Tamirat; words can't express the inspiration he did for the completion of this thesis work. My admiration goes to all my friends, Teshome, Debela, Samuel, Meron and especially to Abraham for giving me valuable comments in the thesis writing and suggesting relevant materials and Kassaye Berhanu who helped me to meet the dead-line.

Finally, much thanks to my mother whose wishes and prays got me here and lots of thanks to sisters and brother for their support, understanding and Inspiration. Above all I thank GOD and Saint Merry for helping me to finish this work.

CONTENTS

ACRONYMS

MOSFET ……...Metal Oxide Semiconductor Field Effect Transistor

HBT ……Heterojunction Bipolar Transitor

MOS ……...Metal Oxide Semiconductor

BJT…….. Bipolar Junction Transistor

SiGe ……Silicon-Germanium

GaAs ……...Gallium Arsenide

IC …….Integrated Circuit

Ge …….Germanium

Si ……Silicon

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Evolution of Bipolar Technology

The design and study of a new semiconductor device structure hold promise at both the device level, where the transistor's electrical behavior may lead to novel effects, and circuit level, where the device characteristics may be exploited to enhance functional performance. Since the revolutionary invention of the point-contact transistor at Bell Laboratories in 1947, numerous new transistor structures have been proposed and demonstrated. Of the many transistors demonstrated in the last fifty years, however, the IC market is dominated by just two devices: the BJT with a market share of about 20 %, and the metal-oxide semiconductor field-effect transistor (MOSFET) with 75 %. BJTs and MOSFETS are the dominant high performance devices in silicon technology. The historical advantage of the bipolar device is the fact that its vertical dimensions are easier to control than the lateral MOS structure. Current gain in a homojunction npn bipolar transistor is mainly determined by the ratio of densities of electrons injected from the emitter into the base and densities of holes reinjected from the base into the emitter, and results in a finite dc current gain.

1.2 Heterojunction Bipolar Transistor

The idea of varying the bandgap in a bipolar transistor structure to increase the emitter efficiency is almost as old as the bipolar junction transistor. Shockely described the idea in his application for a patent on the junction bipolar transistor [1]. The inherent performance advantages of HBTs over conventional bipolar junction transistors have been recognized and Kromer [2] first explained the underlying principle of the heterojunctions. The heterojunction offers a larger set of device configurations and has become the basis for the so-called field of bandgap enginnering.

The principle operation of an HBT is identical to that of the BJT, except that the bandgap of the emitter region exceeds that of the base region ΔE_g , typically of the order of 0.10–0.2 eV.

The resultant $e\Delta E_g / kT$ exponential increase in current gain permits scaling of the base region to smaller thickness and higher doping levels.

1.3 Development of SiGe HBT technology.

As Silicon BJTs, reach their fundamental limits on speed because of the physical properties of the semiconductor material, advanced high-speed devices require heterojunction technology, as has been demonstrated in the pervious section. Although Ge had made its mark as point-contact electrode on the first transistor, Si eventually became the semiconductor of choice for its material properties. In 1957, Kromer patented the first heterojunction Si bipolar transistor and eighteen year later, Erich kasper at Daimler-Benz (now Daimler-Chrysler) made the first SiGe strained layer [63].

SiGe HBTs are particularly exciting because of their ability to take immediate advantage of highly developed Silicon processing techniques. Impressive improvements in high-speed SiGe bipolar technology have been made through the growth of device quality strained–$Si_{1-x}Ge_x$ layers. This strain, which occurs because of a ~4 % difference in the lattice constants of Si and Ge, is used to vary the bandgap energy, band discontinuity and other properties of the material.

Although the introduction of Ge in the base increases process integration complexity, it offers an additional degree of freedom which relaxes a series of trade-offs affecting device design.

The design of SiGe HBT, for a particular technology generation is optimized by appropriate scaling of the emitter, base and collector regions and their associated doping profiles. A SiGe HBT offers additional design flexibility in that the bandgap of the base may be tailored by grading the Ge concentration. Reducing the width of the base region reduces the base transit time with associated improvement in cut-off frequency, but inevitably increase overall base resistance with possible reduction in f_{max}. For effective design, it is thus essential to use an appropriate Simulation tool. Devices like SiGe HBTs designed with such techniques do have better performance.

Thus, SiGe HBTs are superior to Si bipolar junction transistors (BJTs) and comparable to the best GaAs transistors, in that they are ideally suitable for low-voltage, low-power wireless communication applications. Promising research results, combined with recent commercialization announcements, have generated considerable optimism. Silicon has been pushed to the 1-2 GHz frequency domain. However, many new RF applications, such as handled and personal communication systems (PCS), direct broadcast TV, local multipoint distribution systems and wireless LANs, require circuit operation at frequencies up to 30 GHz.

High-speed digital communications (up to 40 Gbps) such as synchronous optical network (SONET) applications also require high-speed devices-typically with a maximum oscillation frequency, f_{max} in excess of 100 GHz. It is now believed that, in many of these markets, SiGe will provide direct competition for GaAs on the grounds of cost and design flexibility. Indeed, it is possible that SiGe technology may eventually be applicable in the frequency range above 30 GHz.

The revolution in wireless communication has been brought about by a combination of advances in digital integrated circuit technology, RF communications, digital communications and networking techniques [64]. RF communication systems can be broadly categorized in two market sectors, namely, 'low end' such as pagers, cordless phones etc, and 'high end' such as personal communication services (PCS), GSM, IS-136 etc. SiGe HBTs are suitable for applications in the high-end applications where the best performance is essential, while CMOS technology will dominate the low-end applications.

1.4. Motivation

The motivation of this research work is the inability of homojunction bipolar transistor (BJT) to provide applicable high current gain, high speed and high frequency ICs. This problem is overcome by the use of hetero-junctions at the emitter-base and collector-base junction of the transistor (i.e. by using a hetero-structure material at the base of a transistor.)

1.5. Objective of the thesis

1.5.1. General objective

The prime objective of this thesis is studying the DC Characteristics of the device SiGe HBT via simulation of the device behavior by the approximate numerical solution of the (approximate) physical transport and field equations governing charge flow in the device using PADRE.

1.5.2. Specific objective

The specific objective of this thesis is simulating the energy band of Si bipolar junction transistor, the energy band of SiGe HBT, I-V characteristics of Si BJT and the I-V characteristics of SiGe HBT. Calculating the current gain of Si BJT and SiGe HBT is also a job taken as specific objective.

1.6. Thesis Organization

Organization of the thesis begins with presenting the background theory of SiGe HBTs in chapter 2, which includes examining the underlying physics of the npn SiGe HBT, with a particular emphasis on the fundamental differences between the operation of the SiGe HBT and Si BJT. Chapter 3 describes issues relating to optimal design of SiGe HBTs and considers how device simulation can be used to determine key indictors of device performance. Chapter 4 introduces the device simulator padre and includes explanations of padre syntax, choice of numerical methods and techniques for solution. Chapter 5 discusses results of the simulations and finally chapter 6 summarizes the conclusions of this research and offers suggestions for future work.

Chapter 2

Theory of SiGe HBT

In this chapter the underlying physics of npn SiGe HBT is examined with a particular emphasis on the fundamental difference between the operation of the SiGe HBT and the Si BJT.

The concept of a bipolar transistor in which the emitter has a greater bandwidth than the base dates back to the time of Shockley's original patent on the junction bipolar transistor [1]. A detailed theoretical analysis of the potential performance advantage of this type of device, commonly known as a heterojunction bipolar transistor, was presented by Kroemer in 1957 [2]. However, it was not until 1987 that the functional HBT employing a base layer was demonstrated. The introduction of Ge into the base of an npn Si BJT reduces the bandgap of the SiGe alloy in the P-doped base, relative to Si in the n-doped emitter and collector regions. This bandgap discontinuity creates the heterojunctions needed for the enhanced performance of SiGe HBT.

Before discussing heterojunction action in a bipolar transistor, it is better to run through well-established Si BJT fundamentals [3]. If the effect of carrier recombination is initially ignored, the electron and hole injection currents in a forward biased pn junction can be expressed as

$$I_n = \frac{qAD_{nb}}{L_{nb}} n_{po}\left(\exp\left\{\frac{qv_{be}}{kT}\right\}-1\right) \qquad (2.1)$$

$$I_p = \frac{qAD_{pe}}{L_{pe}} p_{no}\left(\exp\left\{\frac{qv_{be}}{kT}\right\}-1\right) \qquad (2.2)$$

where v_{be} is the applied bias, A is the area of junction, D_{nb} and D_{pe} are the minority carrier diffusion constants, L_{nb} and L_{pe} are minority carrier diffusion lengths, and n_{po} and p_{no} are the equilibrium minority carrier concentration in the neutral base and emitter, respectively.

In conventional homojunction transistors, the doping concentration in the emitter is considerably higher than in the base, in order to obtain a high injection efficiency .For a typical gain of 100, the emitter must be doped 100 times more heavily than base. As the doping

concentration increases more than $10^{18} cm^{-3}$, bandgap narrowing due to heavy doping becomes more significant [4].

The following substitution can be made in equations (2.1) and (2.2)

$$n_{po} = \frac{n_{io}^{\ 2}}{N_b} \tag{2.3}$$

$$p_{no} = \frac{n_{io}^{\ 2}}{N_e} \tag{2.4}$$

$$n^2_{ie} = n^2_{io} \exp \frac{\Delta E_{bgn}}{kT} \tag{2.5}$$

where n_{io} is the intrinsic carrier concentration and ΔE_{bgn} represents the bandgap reduction in the emitter due to heavy doping.

When bandgap narrowing is included, the current gain β becomes

$$\beta_{si} = \frac{N_e L_{pe} D_{nb}}{N_b L_{nb} D_{pe}} \exp\left(\frac{-\Delta E_{bgn}}{kT}\right) \tag{2.6}$$

In an HBT with a narrow bandgap base, the bandgap of the emitter is larger than the base and therefore the injection effieciency can be very high, even if the base is doped more heavily than the emitter [2, 5]. In a SiGe HBT, a narrow bandgap SiGe base is used and the bandgap difference between the emitter and base is $\Delta E_g(x) = E_{g,Si} - E_{g,SiGe}(x)$. Due to its smaller bandgap, the intrinsic carrier concentration in the SiGe base increases.

The difference between the HBT and BJT is that the concentration of the injected electrons is much higher (several orders of magnitude) into the base due to lower conduction band barrier .The current gain for a SiGe HBT becomes

$$\beta_{SiGe} = \beta_{Si} \exp\left(\frac{\Delta E_g(x)}{kT}\right). \qquad (2.6)$$

This means that the collector current will be much larger than in a similar doped BJT, by a factor of $\exp\left(\Delta E_g(x) \big/ kT\right)$, while the base current is not affected . In a SiGe HBT, the bandgap difference $\Delta E_g(x)$ can be made much larger than kT. For example, a Ge fraction x=0.2 in the base yields a bandgap difference of more than 170 meV. Therefore, the current gain of the HBT can be made large, irrespective of the doping ratio in the emitter and the base. However, the real advantage of the HBT is not to achieve a very high current gain, but to trade it against a high base doping, necessary to reduce the base resistance.

High value of maximum oscillation frequency and low values of gate delay τ_d (for digital switching applications) can be obtained in HBTs [6, 7]. Base resistance is an important parameter in determining $f_{\max}$. In a well designed HBT, a value of 50 for β is usually sufficient, so emitter injection efficiency can be traded for increased doping in the base. Increased base doping give rise to reduced base resistance which is also desirable in helping to avoiding punch-through as the base-collector voltage is increased.

High base doping may contribute to the onset of tunneling current at the emitter-base junction .This can be avoided by deliberately reducing the doping concentration in the emitter. Indeed, in the HBT, it is in principle feasible to consider the possibility of interchanging collector and emitter, providing additional advantage in some digital circuits. Many of the specific issues involved in transistor design are more fully covered in chapter 3. For the reminder of this chapter, focus is given to device physics of the base region and the emitter-base and base-collector junction.

2.1. ENERGY BAND

The first step in understanding how a heterostructure device will operate is to consider the energy band diagram. For homostructre, the electron affinity and bandgap are position independent, and there is no need to worry about the reference value. But for heterostructures, a reference level is essential, normally taken to be the field-free vacuum level. To draw energy band diagram for devices with a position-dependent alloy composition junctions, it is essential to know how the bandgap varies with position and also the band line up at composition junctions.

Figure 2.1 shows the band diagram of an npn bipolar transistor. In forward active mode, the emitter-base junction is forward biased by the input voltage v_{be} , and the base-collector junction is reverse biased by the output voltage v_{bc} [2] .The collector current Ic consists of electrons which are injected from the n-emitter into the thin p-base, move through the base by drift diffusion, and are collected in the n-collector layer (a drift field in the base can be caused by either a doping or a bandgap gradient). The number of electrons injected into the emitter side of the base is determined by the height of the potential barrier, ΔV_n, in the conduction band between the emitter and the base , which can be controlled by the input voltage $v_{be.}$ The dominant component of the base current I_b consists of holes which are injected from the p-base into the n-emitter (no holes are injected into the n-collector in forward active mode because the base-collector junction is reverse biased). The number of holes injected into the emitter is determined

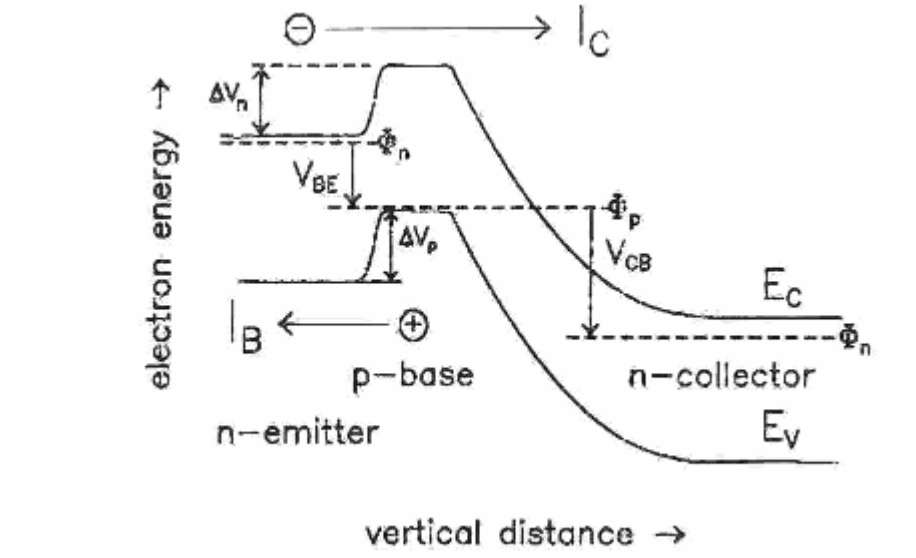

Figure 2.1 Principal Energy band diagram of an npn bipolar transistor

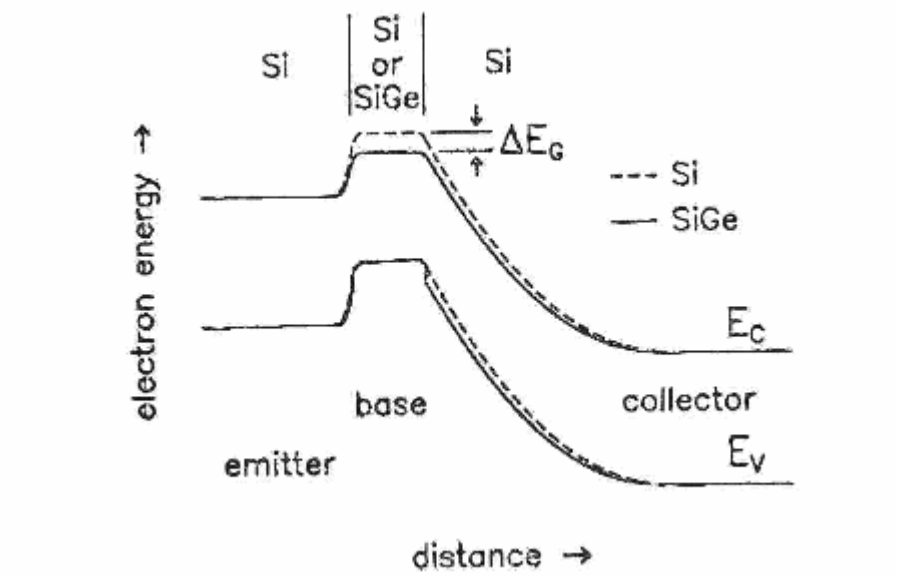

Figure 2.2 Band diagram of a narrow bandgap npn heterojunction bipolar transistor [4].

by the potential barrier $\Delta V_{p\ in}$ the valence band between base and emitter, which is also controlled by the input voltage v_{be}. The key idea of an HBT is to lower the potential barrier seen by the carrier responsible for the output current (electrons in npn devices) compared with the one seen by the carriers constituting the input current (holes in npn devices), there by increasing the ratio of output to input current ,the common emitter current gain of the HBT [5]. This is done by fabricating the emitter and the base using materials having different bandgaps. Depending on the layer in which the bandgap is changed compared to a homojunction device, two HBT configurations can be distinguished:

(i) in a narrow bandgap base HBT , the bandgap in the base is lowered thereby increasing the collector current, whereas

(ii) in a wide bandgap emitter HBT, the bandgap in the emitter is increased

compared to the base resulting in a lower base current.

In both cases, the common emitter current gain is increased by a factor proportional to $\exp\left(\Delta E_g / kT\right)$ if spike and notch effects at the heterojunction are neglected. Note that in an HBT, where the emitter bandgap is larger than in the base, the current gain β should increase when the temperature is lowered, making it possible to operate the transistor more effectively at cryogenic temperature.

2.2 Terminal currents in a SiGe HBT

In this section SiGe HBT will be compared with the equivalent Si BJT. For the purpose of comparison, it is assumed that both the silicon bipolar and the SiGe HBT are identical, other than the fact that germanium is present in the SiGe HBT. Figure 2.3 shows how the base bandgap changes are brought about by the incorporation of Ge into the base [8].

In thermal equilibrium, the Fermi level, $E_{F,}$ is constant across the junction. Therefore, for an abrupt Si/SiGe interface, the difference in bandgap between the emitter and base causes discontinuity to exist at the conduction and valence bands, shown in Figure 2.3 as ΔE_c and ΔE_v, respectively .Also, the total discontinuity, $\Delta E_c + \Delta E_v$, is equal to the base bandgap difference between the silicon emitter and SiGe base , $\Delta E_g^{\,c-b}$. In SiGe, the valence band discontinuity, ΔE_v, tends to be considerably larger than the conduction band discontinuity, ΔE_c.

Figure 2.4 shows the band diagram in forward active mode, where in this more general case, the germanium concentration is graded linearly across the base, increasing from emitter toward the collector [8]. With the presence of germanium, the electron injection barrier from emitter to base, ψ_n, is reduced and there will be greater electron injection from emitter to base. This means an increase in the collector current. However, the hole injection barrier from base to

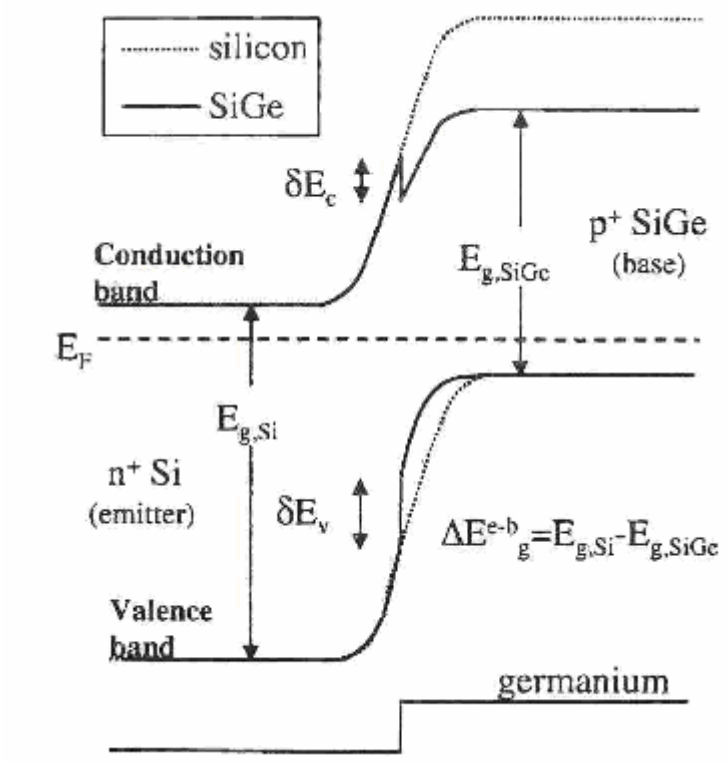

Figure 2.3 Effective of strained-SiGe layer n the bandgap of emitter-base junction for an abrupt Si/SiGe interface.

emitter, ψ_p, remains the same as in a silicon bipolar transistor .Therefore, the hole current from base to emitter, which is the main contributor to base current, remains the same. Hence, silicon bipolar transistor and SiGe HBTs tend to have approximately the same base current.

The following derivations [8], is used to show enhancements resulting from Ge incorporation in the base, closely follow derivation contained in [9]. We consider the most general case of germanium grading and show how constant grading may be treated as a particular case for which the theoretical treatment is still valid. The collector current of a graded SiGe HBT can be obtained by altering the collector current equation of a silicon bipolar transistor.

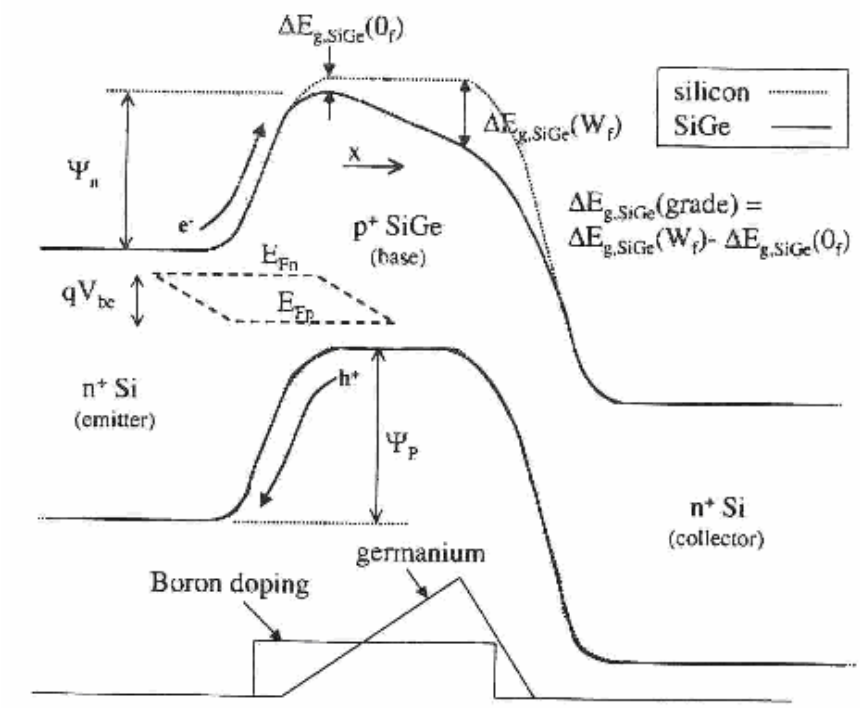

Figure 2.4 Bandgap energy diagram across a graded SiGe HBT in forward active mode of operation.

Assuming uniform base doping for the device, the silicon bipolar collector current density, $J_{c,Si}$, for uniformly-doped base can be written using the Moll-Ross relation [10]

$$J_{c,Si}=q(\exp(qV_{be}/kT)-1)\left[\int_{O_f}^{W_f}\frac{N_b(x)dx}{D_{nb}(x)n_{ie}^{\ 2}(x)}\right]^{-1} \qquad (2.8)$$

$$=\frac{qD_{nb}n^{2}_{ie}}{W_bN_b}\exp(\Delta E_{gb}^{app}/kT)[\exp(qV_{be}/kT)-1] \qquad (2.9)$$

where q is the charge on an electron ,V_{be} is the forward biased emitter-base voltage, k is the boltzmann constant, T is the temperature, O_f and W_f are the base electrical junction position at

the emitter and collector side of the neutral base, in forward active mode, W_b is the neutral base width, $N_b(x)$ is the position-dependent base doping concentration $D_{nb}(x)$ and $n_{ie}(x)$ are positional-dependent base electron diffusion coefficient and effective intrinsic carrier concentration, respectively, n_{io} is the intrinsic carrier concentration in the absence of heavy doping effects, N_b is the base doping, and ΔE_{gb}^{app} is the base apparent bandgap narrowing due to the heavy doping effect.

In equation (2.8), $n_{ie}(x)$ accounts for the effective intrinsic concentration across the base and is a function of the bandgap. For a graded SiGe HBT, bandgap changes across the base, as depicted in figure 2.4 can be accounted for [9]

$$n_{ie}^2(x) = \gamma n_{ie}^2 \exp\left(\frac{\Delta E_{gb}^{app}}{kT} + \frac{\Delta E_{g,SiGe}(graded)(x/W_b)}{kT} + \frac{\Delta E_{g,SiGe}(O_f)}{kT}\right) \qquad 2.10$$

Where [11]

$$\gamma = \frac{(N_c N_v)_{SiGe}}{(N_c N_v)_{Si}} \approx 0.4 \qquad 2.11$$

and neutral base width, $W_b = W_f - O_f$. The term $\Delta E_{g,SiGe}(grade)$ represents the bandgap difference across the neutral base. The term $\Delta E_{g,SiGe}(O_f)$ represents the bandgap difference at the emitter side of the neutral base, N_c and N_v are the density of states in the conduction and valence bands respectively.

Putting equations (2.8) and (2.10) together and integrating, the most general form for the SiGe HBT collector current density, $J_{c,SiGe}$ incorporating both bandgap offset and grading, can be written as [9]

$$J_{c,SiGe} = \bar{\varsigma}\,\bar{\gamma}\frac{qD_{nb}n_{io}^2}{W_b N_b}\left[\exp\left(\frac{\Delta E_{gb}^{app}}{kT}+\frac{qV_{be}}{kT}\right)-1\right]\times\frac{\Delta E_{g,SiGe}(graded)}{kT}\frac{\exp\left(\Delta E_{g,SiGe}(O_f)/kT\right)}{1-\exp(-\Delta E_{g,SiGe}(grade)/kT)} \tag{2.12}$$

where

$$\bar{\varsigma} = \frac{(D_{nb})_{SiGe}}{(D_{nb})_{Si}} > 1 \tag{2.13}$$

where the symbol '_' refers to a position averaged quantity. The ratio of $(D_{nb})_{SiGe}$ to $(D_{nb})_{Si}$ accounts for the strain enhancement of the minority carrier electron mobility with increasing germanium content [12].

Taking the ration of $J_{c,SiGe}$ to $J_{c,Si}$, the collector current enhancement due to bandgap engineering can be estimated by,

$$\frac{J_{c,SiGe}}{J_{c,Si}} = \bar{\varsigma}\,\bar{\gamma}\frac{\Delta E_{g,SiGe}(graded)}{kT}\frac{\exp\left(\Delta E_{g,SiGe}(O_f)/kT\right)}{1-\exp(-\Delta E_{g,SiGe}(grade)/kT)} \tag{2.14}$$

where important conclusion can be drawn by considering the magnitudes of the terms in the above equation in giving rise to the collector current enhancement, i.e.,

- $\bar{\varsigma} > 1$ defines the effect of the difference in diffusivity/mobility between SiGe and Si;
- $\exp\left(\frac{\Delta E_{g,SiGe}(O_f)}{kT}\right) > 1$ defines the effect of basic heterojunction action due to bandgap shrinkage in the base; and
- $\frac{\left(\Delta E_{g,SiGe}(grade)/kT\right)}{1-\exp(-\Delta E_{g,SiGe}(grade)/kT)} > 1$ defines the effect of bandgap grading across the base.

It should be pointed out that equation (2.12) applies in the general case. In the limiting case, where there is no gradient, the latter term tends to unity as $\Delta E_{g,SiGe}(grade)$ tends to zero, and the overall expression for collector current is still valid in a much simplified form.

Even though $\bar{\gamma} < 1$ [11], exp $(\Delta E_{g,SiGe}(O_f)/kT$ increases. The SiGe HBT having a germanium concentration varying from 4% at the emittere-base junction to 12% with a trapezoidal shape across the base (see figure 2.5), a collector current enhancement by a factor of 4.5 has been reported [9].

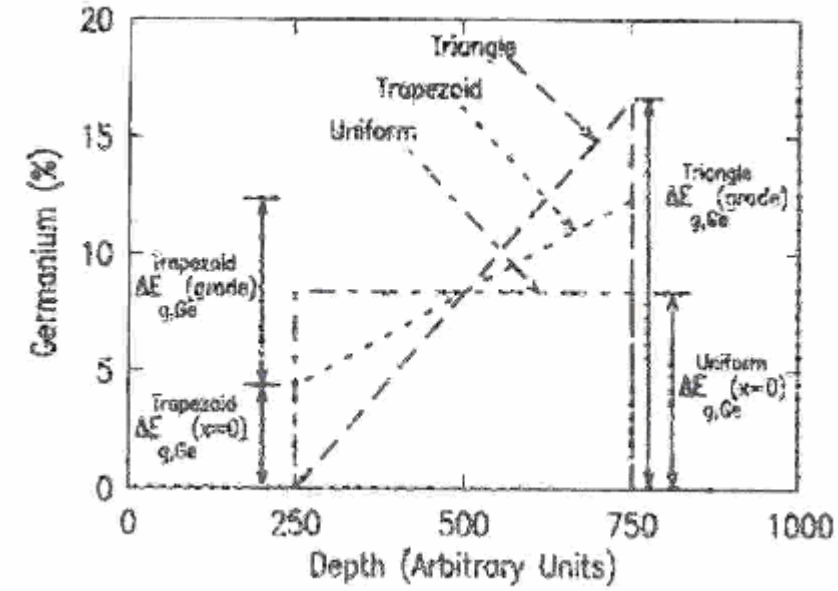

Figure2.5 Uniform (flat), triangle, and trapezoidal Ge profile in the base of a SiGe HBT.

The base current in a bipolar transistor, consists of several components. In the emitter, holes can recombine with electrons at the emitter surface, in the neutral emitter, or in the wide bandgap emitter-base space charge region, or neutral base. In the narrow bandgap base, electrons can recombine with holes in the narrow bandgap part of the emitter-base space charge region, or in the neutral base. An additional source of collector and base current consists of electron-hole pairs created by avalanche multiplication or thermal generation in the base-collector space charge region. The various base current components can be distinguished by their dependence on emitter-base voltage, base-collector voltage and temperature. If both base and emitter material

have a high minority carrier lifetime, which is usually the case in SiGe HBTs, the base current is dominated by emitter surface recombination current or the current in the neutral emitter.

Since the boundary condition for the injected minority carriers into the emitter remains the same as in the homojunction, the reverse injected hole current can be written as

$$J_p = \frac{qD_{pe}n_{ie,Si}^2}{N_{de}W_e}\left(e^{qV_{be}/kT} - 1\right) \qquad (2.15)$$

Equation (2.15) assumes a short, uniformity-doped emitter. For emitters with a short minority carrier lifetime, W_e is replaced by the diffusion length $L_{pe} = \sqrt{D_{pe}\tau_p}$ where D_{pe} and τ_p are the respective minority carrier diffusivity and lifetime in the emitter region, giving

$$J_p = \frac{qD_{pe}n_{ie}^2}{N_{de}L_{pe}}\left(e^{qV_{be}/kT}\right) \qquad (2.16)$$

where N_{de} and L_{pe} are the emitter doping density and hole diffusion length, respective. Equation (2.16) implies that J_p has an ideality factor of unity.

The potential barrier for hole injection into the emitter is the same for both the homojunction and the narrow bandgap heterojunction device, which implies that this component of the base current should be identical in the two devices, if they have similar emitters.

Auger recombination deals with the heavy doping effects. This band-to-band recombination mechanism occurs at dopant concentrations beyond 10^{19} cm^{-3} [13]. One of the main objectives in SiGe HBT design is to lower the base resistance by increasing the base doping concentration. The lower base resistance improves high-frequency performance. In the highly doped emitter of a BJT, the net effect of Auger recombination is a lower effective lifetime in the emitter, leading to a shortened diffusion length and increased base current. In a device simulator this effect is easily included as an extra term in the current continuity equations.

2.3 TRANSIT TIME

Bandgap grading across the base creates a drift electric field that accelerates the electron minority carrier through the base. The graded electric field reduces the amount of base stored charge per unit collector current. This reduces the energy and time required to move charge in and out of the base during transients. As a result, the base transit time, τ_b, decreases.

In any bipolar transistor, the base transit time for constant base doping can be written as [10].

$$\tau_b = \frac{Q_b}{I_c} = \int_{O_f}^{W_f} \frac{n_{ie}^2(z)}{N_b(z)} \left[\int_z^{W_f} \frac{N_b(y)dy}{D_{nb}(y) n_{ie}^2(y)} \right] dz \qquad (2.17)$$

where Q_b is the total base stored charge and I_c is the collector current. Putting equation (2.10) into (2.17) and integrating, $\tau_{b,Si}$ [13, 14] and $\tau_{b,siGe}$ [9] becomes:

$$\tau_{b,Si} = \frac{W_b^2}{2D_{nb}} \qquad (2.18)$$

$$\tau_{b,SiGe} = \frac{W_b^2 kT}{\overline{\varsigma} D_{nb} \Delta E_{g,SiGe}(grade)} \times \left[1 - \frac{kT}{\Delta E_{g,SiGe}(grade)} \left(1 - \exp\left(\frac{-\Delta E_{g,SiGe}(grade)}{kT} \right) \right) \right] \qquad (2.19)$$

Taking the ration of $\tau_{b,SiGe} / \tau_{b,Si}$ gives:

$$\frac{\tau_{b,SiGe}}{\tau_{b,Si}} = \frac{2}{\overline{\varsigma}} \frac{kT}{\Delta E_{g,SiGe}(grade)} \times \left[1 - \frac{kT}{\Delta E_{g,SiGe}(grade)} \left(1 - \exp\left(\frac{-\Delta E_{g,SiGe}(grade)}{kT} \right) \right) \right] \qquad (2.20)$$

For a finite germanium grading of more than 1% at room temperature, $\tau_{b,SiGe} / \tau_{b,si}$ will be less than 1 and therefore the SiGe HBT base transit time will be shorter than silicon bipolar. The cut-off frequency, f_T of a bipolar device, as explained in section 2.7, is a function of base transit time implying that bandgap grading will also increase the usable frequency of operation of the device.

An additional benefit of incorporating Ge into the base is a reduction in emitter transit time τ_e, compared to silicon BJT. Since τ_e is inversely proportional to the collector current, for a given base doping profile, the enhancement in τ_e, is obtained from the inverse of (2.14)

$$\frac{\tau_{e,SiGe}}{\tau_{e,Si}} \approx \frac{J_{c,Si}}{J_{c,SiGe}} = \frac{1 - e - \Delta E_{g,SiGe}\ (grade\ ) / kT}{\overline{\varsigma}\ \overline{\gamma}\ \frac{\Delta E_{g,SiGe}\ (graded\ )}{kT} e^{\Delta E_{g,SIGe}\ (0)/kT}} \quad (2.11)$$

The emitter transit time can potentially be a limiting factor in HBTs which include a low-doped emitter region to avoid tunneling current from the base to emitter.

2.4 EARLY VOLTAGE

For analogue circuit applications, a high value of the product of current gain and Early voltage (βV_A) is desirable. There are several physical effects which cause the collector current to increase with collector-emitter voltage for a constant base current. The most important of these is the increase of collector current caused by a decreased of the width of the neutral base with base-collector reverse bias [15].

Reduction of the neutral base width leads to an increase in the gradient of the injected electron distribution in the p-type base. Since the electron diffusion current across the base is directly proportional to this gradient, the collector current will increase.

The Early voltage (ignoring recombination in the base) is given by

$$V_A \approx J_C \frac{\partial V_{ce}}{\partial J_C} = \Big/ \left(\frac{\partial V_{ce}}{\partial W_b} \frac{\partial W_b}{\partial J_C} \right) \quad (2.22)$$

The rate of change of the neutral base width W_b width respect to the base-collector voltage, for constant emitter-base voltage, is given by

$$\frac{\partial W_b}{\partial V_{bc}} = -\frac{C_{jc}}{qN_b(W_b)} \qquad (2.23)$$

and

the change of collector current density with respect to the base width is

$$\frac{\partial J_C}{\partial W_b} = -J_C \frac{N_b(W_b) \Big/ (n_{ie}^2(W_b) D_{nb}(W_b))}{\int_0^{W_b} (N_b(x) \Big/ (n_{ie}^2(x) D_{nb}(x)))dx} \qquad (2.24)$$

For a constant base profile, combining equations (2.23) and (2.24) one gets

$$\mathrm{V_A} = \frac{q n_{ie}^2(W_b) D_{nb}(W_b)}{C_{jc}} \int_0^{W_b} (N_b(x) \Big/ (n_{ie}^2(x) D_{nb}(x)))dx \qquad (2.25)$$

where $n_{ie}^2(W_b)$ denotes the intrinsic carrier density at the end of the neutral base on the collector side. Combining equation (2.25) with the standard equation for bipolar current gain

$$\beta = \frac{q}{J_{bo}} \left[\int_0^{W_b} p(x) \Big/ (n_{ie}^2(x) D_n(x))dx \right]^{-1} \qquad (2.26)$$

and

assuming p(x) = $\mathrm{N_b}$(x) yields an important figure-of-merit for bipolar transistors, βV_A, given by

$$\beta V_A = \frac{q^2}{J_{bo} C_{jc}} \left(n_{ie}^2(W_b) D_n(W_b) \right). \qquad (2.27)$$

The following three points are significant:

- βV_A is a strong function of Ge concentration at the end of the neutral base (base-collector junction);
- βV_A is larger in SiGe than in silicon due to the larger $n_{io}^2(w_b)$ value in SiGe; and
- To maximize βV_A, the base-collector junction capacitance should be as low as possible.

Harame et al [9] showed that Early voltage enhancement in of a graded SiGe HBT can be expressed as

$$\frac{V_{A,SiGe}}{V_{A,Si}} \approx \exp\left(\frac{\Delta E_{g,SiGe}}{kT}\right)\left[\frac{1-\exp(-\Delta E_{g,SiGe}(grade)/kT}{\Delta E_{g,SiGe}(grade)/kT}\right] \tag{2.28}$$

Combining equations (2.28) and (2.14), the enhancement in βV_A at constant emitter-base voltage can be shown as

$$\frac{\beta V_{A,SiGe}}{\beta V_{A,Si}} \approx \gamma\varsigma e^{\Delta E_{g,SiGe}(o_f)/kT} e^{\Delta E_{g,SiGe}(grade)/kT} \tag{2.29}$$

which is significant greater than unity for a profile with finite Ge content. For finite germanium grading, $\Delta E_{g,SiGe}(grade),$ of more than 1% across the base, $\tau_{b,SiGe}/\tau_{b,Si}$, ratio will be larger than 1. Therefore, grading Ge across the neutral base improves not only base transit time, but also Early voltage. Furthermore, since current gain is essentially enhanced by the difference in bandgap at the emitter-base junction and Early voltage by Ge grading across the base, respectively, the composition product βV_A is significantly enhanced by up two orders of magnitude.

2.5 Heterojunction barrier effects

The computed conduction band offset in the silicon to strained-$Si_{1-x}Ge_x$ heterojunction is small (typical 20 meV) [16]. If a significant conduction band offset exists, reduction in the gain may result. In a heterostructure, compositional grading across the heterojunction may be used to eliminate the conduction band spike. In the case of an Si/SiGe/Si system, the conduction band spike is not a severe problem if the emitter dopant concentration is larger than the base doping concentration, as the band bending appears on the side with lower doping. In an npn transistor any small conduction band spike is disregarded. However, it is not true for the pnp transistor, as

the spike will be larger in this case because valence band offsets are much larger than the conduction band offsets. At high current densities or high forward bias, the transport of carrier is strongly influenced by the potential barrier that develops due to alloy grading potential of the heterojunction. A retrograde Ge profile near the collector junction also creates a barrier to the flow of majority carriers [17].

2.6 HIGH LEVEL INJECTION

In a bipolar transistor, two difference type of high level injection (HLI) can occur. The first occurs in the base region from the large number of electrons injected at high emitter-base voltage. The effect was analysed for Si BJTs by Webster [19]. Since the reverse injected base current retains an $e^{qV_{be}/kT}$ dependence, the current gain falls off inversely proportional to I_c [3]. In general, this effect does not appear in HBTs if the base dopant concentration is high.

The other HLI effect occurring in the collector region is the kirk effect [18] which arises as the base-collector depletion width spreads into the collector at high current levels due to electron velocity saturation. The effect of velocity saturation at large collector current density depends on the relative base and collector doping concentrations. Forward bias of the internal base-collector junction increases the base current due to hole injection into the collector and results in a rapid drop in dc current gain. In SiGe HBT, the valence band offset prevents the injection of holes into the collector and subsequent the collector current saturates at densities less than the classical kirk effect. In addition, excess charge is stored in the base, which results in decreased current gain and $f_{T.}$

Cottrell and Yu [20] and Yu et al [21] attempted to model the valence band barrier effects at high collector current densities for a SiGe HBT. The authors noted that the valence band barrier effect appears at high current densities for npn and at all current densities for pnp devices. Other researchers [22, 23] examined the effect of two dimensional lateral carriers diffusion on the gain.

2.7 HIGH-FRQUENCY FIGURES-OF- MERIT

For high frequency ac operation, bipolar transistors are often assessed according to two figures-of-merit. The first is known as the unity gain cut-off or transition frequency, f_T . The second is known as maximum frequency oscillation frequency. While both figures-of-merit may not necessarily be suitable for all application of SiGe HBTs, both are still widely quoted particularly in device research publications.

2.7.1. Unity gain cut-off frequency, f_T

f_T is defined as the frequency at which the common emitter short circuits ac current gain is unity [22]. It is related physically to the bipolar device, as the total delay for the minority carrier across the device from emitter to collector, τ_{ec} [3]. The total delay consists of the minority carrier stored charge delay and the junction capacitance charging delay, is often related to f_T through the equation :

$$f_T = \frac{1}{2\pi\tau_{ec}} \qquad 2.30$$

where the total transit time τ_{ec} compromise of a number of components:

$$\tau_{ec} = \tau_e + \tau_{eb} + \tau_b + \tau_{bc} + \tau_{je} + \tau_c \qquad (2.31)$$

The major components, due to minority carrier stored charge, are τ_e for the neutral emitter and τ_b for the neutral base region (as previously discussed in section 2.3). The term τ_{eb} represents minority carrier transit time in the emitter-base depletion region, and is often small enough to be included in the emitter transit time term. The transit time τ_b, the delay due to the excess minority

carrier storage in the base, is generally the most significant term in equation (2.31) and the relevant expressions for a SiGe HBT and the effect of Ge grading have been given in equation (2.19)-(2.20).

The delay term τ_{bc} is known as the collector depletion layer transit time. It can be approximated as [13, 24]

$$\tau_{bc} = \frac{W_{jc}}{2v_{scl}} \tag{2.32}$$

where W_{jc} is the base-collector depletion layer width, υ_{scl} is the carrier scattering limited velocity which is approximately equal to 1×10^7 cm s^{-1} at room temperature for silicon [25]. For high-speed devices, as the base width is consistently scaled down, τ_b reduces, and τ_e and τ_{bc} become progressively more significant.

The delay term τ_{je} is the total charging time associated with emitter base and base-collector depletion layers and is given by [3]

$$\tau_{je} = \frac{kT}{qI_c}\left(C_{je} + C_{jc}\right) \tag{2.33}$$

where C_{je} and C_{jc} are the emitter-base and the base collector depletion capacitances. As the collector current increases, it is often assumed that this transit time component becomes negligible. However, for low power devices, the effect of low I_c on τ_{je} becomes more significant, emphasizing very clearly the importance of minimizing the junction capacitances C_{je} and C_{jc}.

The delay term τ_c is the collector charging time [3]

$$\tau_c = R_c C_{jc}. \tag{2.34}$$

In a well-designed transistor, R_c is usually quite small and therefore τ_c is usually not very significant. By combining all equations, f_T can be conveniently formulated as

$$f_T = \frac{1}{2\pi}\left(\frac{kT}{qI_c}(C_{je} + C_{jc}) + \frac{W_b^2}{\alpha D_{nb}} + \tau_e + \tau_{eb} + \frac{W_{jc}}{2\upsilon_{scl}} + R_c C_{jc}\right)^{-1} \tag{2.35}$$

From equation (2.33) it is clear that τ_{je} is dominant at low collector current, and therefore f_T tends to increase with increase in I_c. However, the influence of τ_{je} reduces drastically as the collector current continues to increase.

2.7.2 Maximum oscillation frequency, f_{max}

The unity gain cut-off frequency provides a good indication of the intrinsic delay associated with a bipolar transistor. However, it is not a realistic parameter for a circuit environment, as it assumes that the output is short circuited. In addition, it is independent of base resistance and hence does not take the base resistance base-collector depletion capacitance time constant into account. These are important parameters for determining the transient behavior of bipolar circuits. Therefore, another more practical and widely accepted figure-of-merit, f_{max}, is commonly used, which characterizes the power transfer in and out of the bipolar device. f_{max} is defined as the frequency at which the unilateral current gain becomes unity. Here the output is essentially isolated from the input by an appropriate external matching circuit compromising reactive and resistive components. The load that it drives is also assumed to be conjugately matched to the transistor output impedance. It can be shown [26] that:

$$f_{\max} = \sqrt{\frac{f_T}{8\pi C_{jc} R_b}} \tag{2.36}$$

where R_b is the base resistance. Equation (2.36) shows that it is not sufficient to obtain a high value f_T, by decreasing base width, but that base resistance and base-collector capacitance must be minimized. However, as base width decreases rapidly to achieve high f_T, R_b will increase unless the doping is increased. To counter that effect, the base needs to be more highly doped, which means that emitter doping has to be lowered to prevent emitter-base junction tunneling for

very high base doping levels. The increased current gain capability of a SiGe base enables lowering of emitter doping without jeopardizing sufficient current gain.

2.8. BREAKDOWN VOLTAGE, BV_{CEO}

Although several breakdown voltage are defined for a bipolar transistor, the most important is the collector-emitter breakdown voltage, BV_{ceo} , as it determines the maximum supply voltage that can be applied. The collector-emitter breakdown is limited by two different reverse bias junction breakdown mechanisms: Zener and avalanche. Zener breakdown occurs when both sides of a junction have high dopant concentrations. Avalanche breakdown occurs when a large electric field appears across the depletion region causing impact ionization and generation of electron-hole pairs. BV_{ceo}, limited by avalanche breakdown, occurs when the product of the avalanche multiplication factor and dc current gain approaches unity. For design purpose it is often approximated by

$$BV_{ceo} \cong \frac{BV_{cbo}}{\sqrt[m]{\beta}} \tag{2.37}$$

where BV_{cbo} is the base-collector breakdown voltage with emitter open circuited and m ranges from 2-3 for silicon [27].

In general, the optimization of breakdown voltages for a homojunction transistor and an HBT does not differ. However, extension of the Ge profile into the collector region to avoid the parasitic heterojunction barrier may lead to increase impact ionization. But simulation of carrier energy seem to indicate that impact ionization is more likely to occur deeper into the collector than originally thought [28]. Therefore, a narrow bandgap Si_{1-x} Ge_x base may not affect the breakdown voltage. A trade-off exists between the breakdown voltage and the collector velocity saturation effects. Increases in breakdown voltage for both emitter-base and base-collector junctions have been obtained by placing lightly-doped spacers on both slides of the heavily doped base without incurring collector velocity saturation effect [29].

CHAPTER 3

DESIGN OF SIGE HBTS

As semiconductor technology continues to evolve, numerical modeling of the electrical behavior of advanced devices has become vital. Numerical device modeling based on the self-consistent solution of the fundamental semiconductor equations dates back to the famous work Gummel in 1964 [30]. In Gummel's one-dimensional (1D) discretization, the Poisson equation and the continuity equation are decoupled and solved sequentially until convergence. Gummel's approach was later extended by de Mari [31] and applied to transient simulations of 1D p-n junction. A very important breakthrough in the discretization of the current transport equation was reported by Scharefetter and Gummel in 1969 [32]. The Scharefetter-Gummel discretization scheme has since been used by all important device simulation programs.

During the 1970s and 1980s, several 1D and 2D programs were developed, and made freely available to the research community. Examples include SEDAN [33] for 1D simulations, MINIMOS [34] for 2D MOS transistor simulations, BAMBI [35] for arbitrary semiconductor structures and PISCES [36], A 2D finite-element simulator, which rapidly become an industry standard and formed the basis of future commercial products such as Silvaco-Atlas [37], Avant-Medici [38] and PISCES-2ET [39]. But in this thesis PADRE, a 2D/3D TCAD Device Simulator is used which will be discussed in detail in chapter 4.

Although the drift-diffusion (DD) model is the most widely used and understood tool for semiconductor device simulation, unfortunately fails to predict non stationary transport effects. As a derivative of the Boltzmann transport equation (BTE), it also fails to reflect the quantum mechanical nature of carrier transport. The continuous push towards a smaller device has led to a need to address these shortcomings, and to the development of sophisticated physical models, such as the hydrodynamic and energy balance models [40, 41], the spherical harmonic expansion method [42] and the Monte Carlo technique [43]. Unfortunately, since the Monte Carlo technique involves keeping statistics on large number carriers undergoing random collisions, it is very expensive in terms of computer time. The simulation of a complete transistor requires tracking a prohibitive number of carriers in order to attain statistical significance. This typical limits the Monte Carlo technique to use an aid in studying only part of the transistor, for instance the emitter-base junction.

In the hydrodynamic or energy transport model, the first three moments of the BTE are taken, yielding the particle, momentum and energy conservation equation [40]. To solve these equations, it is generally necessary to make many assumptions (for instance invocation of the relaxation time approximation). As the drift-diffusion model is pushed to its limits, more people are trying the hydrodynamic method of solution. A complete hierarchy of approaches and analyses has been reviewed by Ravaioli [44]. However, the increased rigour of such models comes at the expense of increased CPU time; so far the simulations reported in different publication are confined to discussion almost exclusively to the drift-diffusion model. Regardless of the modeling methodology used, the ultimate responsibility will always rest on the users of the simulator to intelligently interpret the result results and know when the assumptions inherent to the method are being violated. Otherwise, as was pointed out by Tang and Laux [45], …computational sophisticated 2D or even 3D devices simulation are rendered merely expensive, and perhaps misleading, curve fitting program'.

The aim of this chapter is to give some insight into the formulation of a physical device model for a SiGe HBT and to show how it can be applied for HBT transistor design. The model equations account for the position-dependent variation of energy bandgap, the dependence of mobility on different scattering mechanisms, carrier velocity saturation, doping-dependent carrier lifetime and heavy doping effects. The resulting HBT model corresponds closely to that implemented in the Silvaco-ATLAS device simulator [37], which has been in a number of the examples considered. A number of studies are presented where model prediction is compared to measured data.

3.1. DEVICE MODELING

Physically based device simulation predicts the electrical characteristics associated with a specified physical structure and bias conditions. This is achieved by mapping the structure on to a two-dimensional or three-dimensional grid consisting of a number of grid points called nodes. By applying a set of partial differential equations, derived from Maxwell's equation to this grid, transport of carriers can be simulated. By specification of appropriate boundary conditions, dc, ac and transient models of operation can be modeled. Physical simulation has two important characteristics. It is much quicker and cheaper than performing experiments. In addition it provides information that is difficult or impossible to measure. The main drawback is that all the relevant physics must be incorporated into the simulator. The user must specify the problem to be solved by defining:

- the physical structure
- the physical models; and
- the bias conditions for which electrical characteristics are required.

A basic requirement for a successful physical simulation of a semiconductor device is a mathematical model describing its operation. The model is characterized by a set of fundamental equation which links the electrostatic potential and the carrier densities within some predefined simulation domain. These equations are derived from Maxwell's laws and consists of Poisson's equation relates variations in electrostatic potential to the space-charge density and is given by,

$$\nabla \bullet (\varepsilon \nabla \psi) = -q(p - n + N_D^+ - N_A^-) - \rho_s \qquad (3.1)$$

where ψ is the electrostatic potential, ε is the local dielectric permittivity, q is the charge of an electron, p and n are the hole and electron concentrations, N_D and N_A are the ionized donor and acceptor impurity concentrations and ρ_s is the surface charge density.

The continuity equations, which describe the way that electron and hole carrier densities evolve as a result of transport process, generation and recombination processes, are given by,

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \bullet \vec{J}_n + (G - R) \tag{3.2}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot \vec{J}_p + (G - R) \tag{3.3}$$

where J_n and J_p are the electron and hole current densities, and G and R are the generation and the recombination rates, respectively. The above equations provide the general framework for device simulation. However, further secondary equations are necessary to specify, generation recombination rates. The current density equations are usually obtained by applying approximations and simplification to the BTEs. These assumptions can result in a number of possible transport models such as the drift-diffusion model [46], the energy balance and the hydrodynamic models [40]. The choice of transport model can impact on the choice of generation and recombination model. By far the simplest and most commonly used model in device simulation is the drift-diffusion model. Until recently this model was adequate for nearly all semiconductor devices but it tends to become less accurate for small features sizes [44].

In the drift-diffusion model, the current densities are expressed in terms of quasi-Fermi levels ϕ_n and ϕ_p as

$$\vec{J}_n = -q\mu n\nabla\phi_n \tag{3.4}$$

$$\vec{J}_p = -q\mu_p\nabla\phi_p \tag{3.5}$$

where μ_n and μ_p are the electron and hole mobilities. Using Boltzmann approximations, the quasi-Fermi levels may be related to the carrier concentration and the potential as given by

$$n = n_{ie}\exp\left[\frac{q(\psi - \phi_n)}{kT_L}\right] \tag{3.6}$$

$$p=n_{ie}\exp\left[\frac{-q(\psi-\phi_p)}{kT_L}\right] \tag{3.7}$$

where n_{ie} is the effect intrinsic carrier concentration and T_L is the lattice temperature. These two equations may then be rewritten as

$$\phi_n = \psi - \frac{kT_L}{q}\ln\frac{n}{n_{ie}} \tag{3.8}$$

$$\phi_p = \psi + \frac{kT_L}{q}\ln\frac{p}{n_{ie}} \tag{3.9}$$

By substituting these equations into the current density expressions, one obtains

$$\vec{J}_n = qD_n\nabla n - qn\mu_n\nabla\psi - \mu_n nkT_L\nabla(\ln(n_{ie})) \tag{3.10}$$

$$\vec{J}_p = -qD_p\nabla p - qp\mu_p\nabla\psi + \mu_p pkT_L\nabla(\ln(n_{ie})) \tag{3.11}$$

where the last term accounts for the gradient in the effect intrinsic carrier concentration, taking into account bandgap narrowing effects. Effective electric field are given by

$$\vec{E}_n = -\nabla\left(\psi + \frac{kT_L}{q}\ln n_{ie}\right) \tag{3.12}$$

$$\vec{E}_p = -\nabla\left(\psi - \frac{kT_L}{q}\ln n_{ie}\right) \tag{3.13}$$

From the above and using Einstein relationships, the familiar drift-diffusion expressions are as follows:

$$\vec{J}_n = q\mu_n\vec{E}_n + qD_n\nabla n \tag{3.14}$$

$$\vec{J}_p = q\mu_p \vec{E}_p - qD_p \nabla p \tag{3.15}$$

In the case of Boltzmann statistics, D_n and D_p are given by

$$D_n = \frac{kT_L}{q}\mu_n \tag{3.16}$$

$$D_p = \frac{kT_L}{q}\mu_p \tag{3.17}$$

In the case of the energy balance (EB) model, a higher-order solution to the generalized BTE is necessary to include an additional coupling of the current density to the carrier temperature (energy). Then the current density and energy flux densities are expressed as

$$\vec{J}_n = qD_n \nabla n - \mu_n n \nabla \psi + qnD_n^T \nabla T_n \tag{3.18}$$

$$\vec{S}_n = -K_n \nabla T_n - \left(\frac{k\delta_n}{q}\right)\vec{J}_n T_n \tag{3.19}$$

$$\vec{J}_p = qD_p \nabla p - \mu_p p \nabla \psi + qpD_p^T \nabla T_p \tag{3.20}$$

$$\vec{S}_p = -K_p \nabla T_p - \left(\frac{k\delta_p}{q}\right)\vec{J}_p T_p \tag{3.21}$$

Where $K_{n,p}$ and $\delta_{n,p}$ are respective transport coefficient for electrons and holes that depend on the corresponding carrier temperatures. T_n and T_p. S_n and S_p are the flux of energy (or heat) from the carrier to the lattice. Full details of the formation are given in [47].

3.2. Numerical Methods

3.2.1. Numerical Solution Techniques

Several different numerical methods can be used for calculating the solutions to semiconductor device problems. Different solution methods are optimum in different situations and some guidelines will be given here. Numerical methods are given on the METHOD statements of the input file.

Different combinations of models will require PADRE to solve up to six equations. For each of the model types there are basically three types of solution techniques: (a) de-coupled (GUMMEL), (b) fully coupled (NEWTON) and (c) BLOCK. In simple terms, the de-coupled technique like the Gummel method will solve for each unknown in turn keeping the other variables constant, repeating the process until a stable solution is achieved. Fully coupled techniques such as the Newton method solve the total system of unknowns together. The combined or block methods will solve some equations fully coupled, while others are de-coupled. In general, the Gummel method is useful where the system of equations is weakly coupled, but has only linear convergence. The Newton method is useful when the system of equations is strongly coupled and has quadratic convergence. The Newton method may however spend extra time solving for quantities which are essentially constant or weakly coupled. Newton also requires a more accurate initial guess to the problem to obtain convergence. Thus, a block method can provide for faster simulations times in these cases over Newton. Gummel can often provide better initial guesses to problems. It can be useful to start a solution with a few Gummel iterations to generate a better guess and then switch to Newton to complete the solution. Specification of the solution method is carried out as follows:

METHOD GUMMEL BLOCK NEWTON

The exact meaning of the statement depends upon the particular models it is applied to. This will be discussed in the following sections.

3.2.1. Basic Drift Diffusion Calculations

The isothermal drift diffusion model requires the solution of three equations for potential, electron concentration and hole concentration. Specifying GUMMEL or NEWTON alone will produce simple Gummel or Newton solutions as detailed above. For almost all cases the Newton method is preferred and it is the default.

Specifying:

METHOD GUMMEL NEWTON

will cause the solver to start with Gummel iterations and then switch to Newton, if convergence is not achieved. This is a very robust, although more time consuming way of obtaining solutions for any device. However this method is highly recommended for all simulations with floating regions such as SOI transistors. A floating region is defined as an area of doping which is separated from all electrodes by a pn junction. BLOCK is equivalent to NEWTON for all isothermal drift-diffusion simulations.

3.2.2. Drift Diffusion Calculations with Lattice Heating

When the lattice heating model is added to drift diffusion an extra equation is added. The BLOCK algorithm solves the three drift diffusion equations as a Newton solution and follows this with a decoupled solution of the heat flow equation. The NEWTON algorithm solves all four equations in a coupled manner. NEWTON is preferred once the temperature is high; however BLOCK is quicker for low temperature gradients. Typically the combination used is:

METHOD BLOCK NEWTON

3.2.3. Energy Balance Calculations

The energy balance model requires the solution of up to 5 coupled equations. GUMMEL and NEWTON have the same meanings as with the drift diffusion model (i.e. GUMMEL specifies a decoupled solution and NEWTON specifies a fully coupled solution).
However, BLOCK performs a coupled solution of potential, carrier continuity equations followed by a coupled solution of carrier energy balance, and carrier continuity equations. It is possible to switch from BLOCK to NEWTON by specifying multiple solution methods on the same line.

For example:

```
METHOD BLOCK NEWTON
```

will begin with BLOCK iterations then switch to NEWTON if convergence is still not achieved. This is the most robust approach for many energy balance applications.
The points at which the algorithms switch is predetermined, but can also be changed on the METHOD statement.

3.2.4. Energy Balance Calculations with Lattice Heating

When non-isothermal solutions are performed in conjunction with energy balance models, a system of up to six equations must be solved. GUMMEL or NEWTON solve the equations iteratively or fully coupled respectively. BLOCK initially performs the same function as with energy balance calculations, then solves the lattice heating equation in a de-coupled manner.

Setting the number of carriers can be made using the parameter CARRIERS.

For example,

```
METHOD CARRIERS=2
```

specifies a solution for both carriers is required. This is the default. With one carrier the parameter ELEC or HOLE is needed. For example, for hole solutions only:

METHOD CARRIERS=1 HOLE

To select a solution for potential only specify:

METHOD CARRIERS=0

3.3. MATERIAL PARAMETER FOR SIMULATION

Electron and holes in a device are accelerated by electric fields but lose momentum as a result of various scattering process. These scattering mechanisms include lattice vibration, impurity ions, other carriers, interface and material imperfection. To simplify these mechanisms for modeling purpose, mobility is usually defined as a function of lattice temperature, local electric field and doping concentration. In a device simulator, a mobility model is further subdivided into

- low-field behavior,
- high-field behavior,
- bulk semiconductor regions, and
- inversion layers.

In the low-field region, mobility is principally dependent on phonon and impurity scattering, both of which tend to decrease the low-field mobility. High-field behavior shows that carrier mobility decreases with electric field. The mean drift velocity no longer increases linearly with increasing electric field, but rises more slowly. Eventually the velocity saturates at a constant velocity commonly denoted by υ_{sat} which is principally a function of lattice temperature.

Modeling mobility in bulk material involves characterizing μ_{no} and μ_{po} as a function of doping and lattice temperature and describing the transition between low-field and high-field regions. Modeling carrier mobility in inversion layers presents additional complication due to surface scattering and quantum mechanical effects. These effects are important for accurate simulation of MOS devices. The transverse electric field is often used to characterize mobility variation within inversion layers.

Different silicon mobility models are available in the ATLAS manual [37]. The low-field mobility can be characterized in five different ways: user defined; look up table as a function of doping; an analytic function of doping and temperature [48]; a carrier scattering model relating mobility to carrier-carrier scattering

As carrier is accelerated in an electric field, their velocity will begin to saturate at high electric field. This effect has to be accounted for by a reduction of effect mobility, since the drift velocity is the product of the mobility and electric field in the direction of current flow. The following expression [50] is used to implement a field-dependent mobility for both holes and electrons, that provides a smooth transition between low-field and high behavior,

$$\mu(E) = \mu_0 \left[1/1 + \left(\frac{\mu_0 E}{\upsilon_{sat}} \right)^{-1} \right]^{1/\beta} \tag{3.27}$$

where μ_0 is the low-field mobility, E is the electric field parallel to the direction of current flow, β is a constant, and υ_{sat} is the saturation velocity. The coefficient β is one for holes and two for electrons. The saturation velocity υ_{sat} is calculated by default from the temperature-dependent model,

$$\upsilon_{sat}(T) = \frac{2.4 \times 10^7}{1 + 0.8 \exp(T/600)} \tag{3.28}$$

but specific values for holes and electrons can be specified, if required.

The incorporation of germanium significantly changes the properties of the base region and the emitter-base and base-collector junction a SiGe HBT. While silicon has been well characterized over the past 40 years, still not nearly as much is known about strained-SiGe. Many simplifying assumption are made in the SiGe material parameters. The addition of Ge reduces the bandgap Si, leading to the narrow bandgap SiGe base of the HBT, as discussed in chapter 2. The lattice difference of $Si_{1-x}Ge_x$ alloy differs considerably from that of Si. The incorporation of Ge also modifies the energy band structure, and density of states in the conduction and valence bands. In addition, carrier mobilities and diffusivities change owing to changes in the effective masses and alloy scattering. Finally, the dielectric constant, built in potential and depletion widths in the p-n heterojunction depends on the Ge concentration.

3.3.1. SiGe: hole mobility

There have been few reports on the measurement of mobility in strained-$Si_{1-x}Ge_x$ alloys. Mansevit et al [51] reported enhanced electron mobilities at room temperature, but the Ge mole fraction of the samples was not accurately known. Monte Carlo simulations of electron mobility heavily-doped SiGe at room temperature indicate that μ_n will be almost 50% higher than for silicon due to the smaller effective mass in SiGe [52]. Enhanced low-temperature have been also observed for both holes and electrons [53]. In addition to phono, impurity and alloy scattering mechanisms, strain is expected to play a major role in determining carrier mobility. Due to strain effects, mobilities in SiGe are different for carriers traveling parallel and perpendicular to the direction of growth.

The composition, temperature and doping dependent hole mobility is given by:

(i) for majority carriers

$$\mu_p = 49.0\left(\frac{T}{300}\right)^{-1} + \frac{480.0(T/300)^{-2.2} - 49.0(T/300)^{-0.45}}{\left(1.0 + (T/300)^{-2.4}\right)\left(N_{tot}/1.7\times10^{17}\right)^{0.74}} \qquad (3.29)$$

(ii) for minority carrier

$$\mu_p = \left(\left[122.3\left(\frac{T}{300}\right)^{-0.45} + \frac{480.0(T/300)^{-2.2} - 122.3(T/300)^{-0.45}}{(1.0 + (T/300))^{-2.4})(N_{tot}/1.4\times10^{17})^{0.7}}\right]\rho\right) \times\left(1.0 + \frac{1.0}{0.5 + (7.2\times10^{20}/N_{tot})^2}\right)^{-1} \qquad (3.30)$$

where

$$\rho = \left(\mu_{\min}(x) + \frac{(\mu_{\max}(x) - \mu_{\min}(x))}{1 + (N_{tot}/2.35\times10^{17})^{0.88}}\right)\times\left(\mu_{\min}(0) + \frac{\mu_{\max}(0) - \mu_{\min}(0)}{1 + \left(N_{tot}/2.35\times10^{17}\right)^{0.88}}\right)^{-1} \qquad (3.31)$$

where

$$\mu_{\min}(x) = 68.7\exp(51.2x^3 - 34.2x^2 + 8.7x) \qquad (3.32)$$

and

$$\mu_{\max}(x) = 461.9\exp(32.5x^3 - 22.2x^2 + 6.4x) \qquad (3.33)$$

3.3.2. SiGe: electron mobility

The alloy scattering limited electron mobility components for coherently strained $Si_{1-x}Ge_x$, along directions perpendicular and parallel to the growth direction are given by [43]

$$\mu_{\perp}^{alloy} = \frac{5.5\times10^{18}T}{22.0N_c x(1-x)m_t^2} \tag{3.34}$$

$$\mu_{//}^{alloy} = \frac{5.5\times10^{18}T}{4.0N_c x(1-x)m_1^2} \tag{3.35}$$

where N_c is the effective density of states for silicon.

It may be noted that the alloy mobility decreases with increasing Ge content. At low doping levels, alloy scattering and phonon scattering predominate, both of which have an $E^{1/2}$ dependence. Since the conduction band of SiGe for x<0.3 is similar to that of silicon, and all the predominant scattering rates have an $E^{1/2}$ dependence, the individual parallel and perpendicular components may be defined.

The parallel component of electron mobility in SiGe can thus be obtained by using Mathiessen's rule

$$\frac{1}{\mu_{//}^{SiGe}} = \frac{1}{\mu_{//}^{Si}} + \frac{1}{\mu_{//}^{alloy}} \tag{3.36}$$

and the corresponding perpendicular component becomes

$$\frac{1}{\mu_{\perp}^{SiGe}} = \frac{1}{\mu_{\perp}^{SiGe}} + \frac{1}{\mu_{\perp}^{alloy}} \tag{3.37}$$

where the mobility of silicon for parallel and perpendicular to the growth plane is expressed as [43].

$$\mu_{\perp}^{Si} = \frac{3.0\mu_{Si}}{(m_t/m_1 + 2.0)} \tag{3.38}$$

$$\mu_{//}^{Si} = \frac{3.0\mu_{Si}}{2.0(m_t / m_1) + 1.0} \qquad (3.39)$$ where

m_1 and m_t are longitudinal and transverse density of state masses in silicon

At very high concentrations, the Caughey-Thomas relationship [38] no longer suffices to describe the carrier mobility. The effect of ultrahigh concentrations on mobility have been analysed by Klaassen [36], and the modified expression for majority and minority mobility for electron in silicon is given by:

(i) for majority carriers

$$\mu^{Si} = \left[74.5\left(\frac{T}{300}\right)^{-0.45} + \frac{1430.(T/300)^{-2.3} - 74.5(T/300)^{-0.45}}{(1.0 + (T/300)^{-2.6}(N_{tot}/8.6\times10^{16})^{0.77})} \right] / Z \qquad (3.40)$$

$$\mu^{Si} = \left[74.5\left(\frac{T}{300}\right)^{-.45} + \frac{1430.(T/300)^{-2.3} - 74.5(T/300)^{-0.45}}{(1.0 + (T/300)^{-2.6}(N_{tot}/8.6\times10^{16})^{0.77})} \right] \qquad (3.41)$$ where

$$Z = 1.0 + \frac{1.0}{0.21 + (4.0\times10^{20}/N_{tot})^2} \qquad (3.42)$$ where

N_{tot} is the total doping and the clustering function Z(N) is fitted analytically.

To evaluate the mobility of strained-SiGe, alloy scattering as well as energy shifts in the conduction band have to be included. The shift are taken into account through the electron concentration, since the total mobility is given by a weighted average of the unstrained electron concentration of the i^{th} conduction band, with the corresponding strained electron concentration.

3.3.3. SiGe: bandgap

The most significant material parameter to be specified in the simulation of SiGe HBTs is the bandgap narrowing induced by incorporation of a Ge fraction x. A number of different models have been put forward. Polynomial fits by Bludau et al [55] describe the temperature dependence of the energy bandgap of pure silicon at or below room temperature. The high-temperature model from Sze [3] is slightly modified to match the room temperature value and is given by

$$E_g(T) = 1.170 + 1.059\times10^{-5}T - 6.05\times10^{-7}T^2 \qquad 0 \le T \le 170K \tag{3.46}$$

$$E_g(T) = 1.1785 - 9.025\times10^{-5}T - 3.05\times10^{-7}T^2 \qquad 170 \le T \le 300K \tag{3.47}$$

$$E_g(T) = 1.170 - \frac{4.73\times10^{-4}T^2}{T + 624.93} \qquad T \ge 300K \tag{3.48}$$

An empirical a fit to the data provided by People [46] for the bandgap of strained- $Si_{1-x}Ge_x$ alloys on Si(100) substrate is given by

$$E_g(x) = 1.124 - 1.22x + 0.88x^2 \qquad x \le 0.6. \tag{3.49}$$

A linear fit is used for $0.6 < x < 1.0$, which assumes that the bandgap of strained pure Ge on (100) Si is 0.6 eV. Note that the bandgap of strained-SiGe is considerable smaller that of bulk-SiGe.

In ATLAS, to give increased accuracy, the SiGe bandgap is modeled by a complex piecewise linear function of x, as defined in full in the ATLAS manual. For values of x likely to be encountered in SiGe HBT ($x<0.245$), the following equation applies

$$E_g(x) = 1.08 + x(0.945 - 1.08)/0.245. \tag{3.50}$$

In ATLAS, an alternative temperature dependent of the bandgap $E_g(T)$ for SiGe as

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{T+\beta} = E_g(300) + \alpha\left[\frac{300^2}{300+\beta} - \frac{T^2}{T+\beta}\right] \tag{3.51}$$

where the composition dependence of α and β are given by:

$$\alpha = (4.73(1-x) + 4.77x)10^{-4}$$

$$\beta = 636.0(1-x) + 235.0x$$

The electron affinity of SiGe is assumed to be independent of the composition x and equal to 4.07 eV, identical to that Si.

In a BJT model, the intrinsic carrier concentration n_{io}, which depends on the effective density of states in the conduction and valence bands and the bandgap, plays an important role. The effective conduction and valence band density of states in silicon are given by the well-known expressions:

$$N_c = 2\left(\frac{2\pi m_n^* kT}{h^2}\right)^{3/2} \qquad N_v = 2\left(\frac{2\pi m_p^* kT}{h^2}\right)^{3/2} \qquad (3.52)$$

where h is Plank's constant, and m_n^* and m_p^* are the effective masses of the electron and hole density of states.

The effective density of states decrease with increasing Ge content, because the amount of degeneracy in both the valence and conduction band decrease of densities of states for SiGe is given by:

$$N_c = 2.80\times10^{19} + x(1.04\times10^{19} - 2.8\times10^{19}) \qquad (3.53)$$

$$N_v = 1.04\times10^{19} + x(6.0\times10^{18} - 1.04\times10^{19}) \qquad (3.54)$$

By using equation (3.53) and (3.54), one can calculate the intrinsic carrier concentration as a function of the Ge content.

$$n_{io}^2(x) = N_c N_v \exp\left(-\frac{E_g(x,T)}{kT}\right) \qquad (3.55)$$

In addition to the Ge-indued bandgap narrowing, the high doping in the base induces additional bandgap narrowing, similar to that observed in silicon. Although several bandgap narrowing and mobility and models have been proposed for silicon [57-59], little information is available in the

literature for $Si_{1-x}Ge_x$. The default model in ATLAS assumes that the bandgap narrowing due to heavy doping is the same as that in silicon. This approach has the advantage that any differences in the simulation of Si BJT and SiGe HBTs can then be unambiguously attributed to heterojunction action (due to Ge incorporation), rather than differences in model parameters. This assumption of equal values of doping induced bandgap narrowing in silicon and $Si_{1-x}Ge_x$ is lower than that in Silicon.

Bandgap narrowing effects due to heavy doping are modeled by replacing the intrinsic carrier concentration n_{io} with an effective carrier concentration $n_{ie}(x,y)$ where

$$n_{ie}(x,y) = n_{io} \exp\left[\frac{qa_1}{2kT}\left\{\ln\frac{N(x,y)}{a_2} + \left(\left(\ln\frac{N(x,y)}{a_2}\right)^2 + a_3\right)^{1/2}\right\}\right] \qquad (3.56)$$ where

a_1=0.00692, $a_2=1.3\times10^{17}$ cm^{-3} and a_3=0.5 are model parameters. In ATlAS, the dielectric constant of SiGe as a function of composition is given by

$$\varepsilon = 11.9 + 4.1x.$$

3.3.4. Recombination and carrier lifetime

The dominant recombination processes in bulk-Si are Shockely-Read-Hall (SRH) and Auger recombination. Radiative recombination is negligible since silicon is an indirect bandgap semiconductor, and recombination involving excitons and shallow-level traps is only important at low temperature. The total recombination rate due to Auger and SRH recombination can be written as:

$$R = \left[A_n n + A_p p + \frac{1}{\tau_n(p+p_1) + \tau_p(n+n_1)}\right]\left(np - n_{ie}^2\right). \qquad (3.58)$$

In equation (3.58), A_n and A_p are the electron and hole Auger recombination coefficients and n_{ie} is the effective intrinsic carrier concentration including bandgap narrowing effects. τ_n and τ_p are the minority carrier SRH lifetimes and n_1 and p_2 are constants which depend on the energy of the deep-level traps. Commonly used (default) values for the radiative and Auger recombination coefficients are $A_n=5.0\times10^{-32}$ and $A_p=9.9\times10^{-32}$ for silicon [index]. Since strained-SiGe is similar to silicon in band structure, exactly the same recombination model is assumed for SiGe.

The minority carrier lifetimes in silicon are doping-dependent. For doping concentration up to 10^{19} cm^{-3}, an empirical fit to experimental fit to experimental data gives

$$\tau(N) = \frac{\tau(0)}{1 + N/N_0} \qquad (3.59)$$

for both electron and hole. $\tau(0)$ is the minority carrier lifetime in lightly doped silicon and N_0 is the reference doping.

3.4. DEVICE DESIGN ISSUES

In the following sections, important parameters of SiGe HBTs (f_T, f_{max}, and V_A) will be considered in detail and attempts are made to illustrate how simulation has been used to optimize the device design for circuit applications. Base, emitter and collector profile design issues at room temperature will be discussed.

3.4.1. Base design

Let's consider a uniform (flat or box) Ge profile (x=0.12) in the base. In the conventional Si BJT, β is inversely proportional to the integrated base charge. Since base doping cannot be increased indefinitely while maintaining adequate β, the flat Ge profile is particularly useful in realizing a transistor with either a very high β, or a moderate β with a lower intrinsic base resistance.

However, any significant enhancement in peak f_T of a SiGe HBT over a Si BJT depends principally on the utilization of Ge grading across the base.

The Ge grading (0-12%), is effective for reducing $\tau_{b,}$ and thus increasing f_T. In this type of Ge profile design, there is no Ge-induced bandgap reduction at the emitter-base junction; an enhancement in β of approximately 5 has been simulated. In high-speed analogue applications, which require a high βV_A product, the triangular Ge profile would appear to offer a superior design [60]. Because β is still enhanced for the triangular Ge profile, it is still possible to trade β for lower base resistance. Using this approach, both f_T and base resistance can be tailored to significantly increase f_{max}.

A trapezoidal profile would appear to be a logical compromise between the two previous Ge profiles. This type of profile was used successfully to realize the first 1.0 Gb s^{-1} 12-bit digital-to-analogue converter [60].

3.4.2 Emitter design

An ideal emitter should provide low emitter saturation current density low emitter resistance, low charge storage, low emitter-base depletion capacitance, and good passivation at the perimeter of the emitter. The polysilicon emitter contact used in conventional Si technology meets most of these requirements. The polysilicon-silicon interface also provides a barrier-to-hole injection into the emitter. An alternative approach to the polysilicon emitter contact is to use single-crystal emitter. Such a structure is ideal to decouple the base the base from the emitter, thereby allowing arbitrarily high base dopant concentrations. Furthermore, it allows a reduction in emitter-base capacitance, leading to higher f_T at lower collector current density, as long as the delay associated with minority carrier charge storage in the quasi-neutral emitter can be minimized by maintaining sufficient current gain. A high-low emitter profile, consisting of heavily-doped polysilicon contact on top of a thin epitaxial emitter cap addresses both requirements [61]. The emitter cap thickness should be small to minimize charge storage and is typically 200-300Å. The highly-doped polysilicon contact ensures low total emitter resistance.

3.4.3 COLLECTOR DESIGN

The design of the collector is dictated by conflicting requirement to simultaneous achieve high breakdown voltage BV_{ceo}, low base-collector capacitance, low base-collector signal delay τ_{bc}, and a high value of the knee current density at which f_T decreases. The collector doping profile determines two critical performance parameters of the transistor: the base-collector delay time τ_{bc}, which is a significant component of the total intrinsic delay τ_{ec}, and the intrinsic base-collector capacitance which governs circuit performance. A convectional approach to suppress base widening is simply to utilization a thin highly-doped epitaxial collector layer. Consequently, base widening is suppressed at the expense of BV_{ceo} degradation.

In determining HBT performance, it should be recalled that the collector-emitter breakdown voltage BV_{ceo} is directly related to the cut-off frequency, according to the theoretical `Johnson limit, and falls monotonically for increasing value of f_T [62]. A 50 GHz transistor corresponds to a breakdown voltage of 3.3 V. In general, therefore, some degree of optimization is always required to yield the appropriate higher f_T for lower BV_{ceo}.

Increasing the peak collector doping density (N_{coll}) above 1 x 10^{17} cm-3 improves the frequency performance in two ways:

(i) a reduction in transit time τ_{bc} giving increase in f_T; and

(ii) a delay onset of Kirk effect permitting operation at higher collector current density since the Kirk (knee) current density (J_k) is proportional to the collector doping.

CHAPTER- 4

THE Device Simulator PADRE

PADRE is a modular and extensible framework for one, two and three dimensional semiconductor device simulation. It is implemented using modern engineering practices that promote reliability, maintainability and extensibility. Products that use the padre framework meet the device simulation needs of many semiconductor applications.

PADRE simulates the electrical behavior of semiconductor devices under steady state, transient conditions or AC small-signal analysis. Multiple devices can be treated, along with lumped element circuit networks. PADRE can simulate physical structures of arbitrary geometry (including heterostructures)-with arbitrary doping profiles that can be obtained using analytical functions or directly from multidimensional process simulators such as PROPHET or BICEPS.

PADRE was developed in 1994 at Bell Laboratories by Mark Pinto, R. Kent Smith, and Ashraful Alam. In this chapter it is pointed out to the most critical issues one faces when creating the input file for particular device structures.

4.1. The PADRE SYNTAX

A padre command file is a list of commands for padre to execute. This list is stored as an ASCII text file using any text editor. The input file contains a sequence of statements. Each statement consists of a keyword that identifies the statement and a set of parameters. The general format is:

<STATEMENT> <PARAMETER>=<VALUE>

Some hints on the proper structure of the statements are given below:

1. The statement keyword must come first, but after this the order of parameters within a statement is not important.
2. It is only necessary to use enough letters of any parameter to distinguish it from any other parameter on the same statement.

3. Logicals can be explicitly set to false by preceding them with the ^ symbol.

4. Any line beginning with # is ignored. These lines are used as comments. Note that the “#” can be put on any padre input line; all information to the left of the character is processed, and that to the right is ignored.

5. Padre can read up to 256 characters on one line.

However, it is best to spread long input statements over several lines to make the input file more readable. The character + at the beginning of a line indicates continuation. There are four groups of statements, and these must occur in the correct order. These groups are indicated in Figure 1.

Each input file must contain these four groups in order. Failure to do this will usually cause an error message and termination of the program, but it could lead to incorrect operation of the program. For example, material parameters or models set in the wrong order may not be used in the calculations. The order of statements within the mesh definition, structural definition, and solution groups is also important.

Group		Statement
1. **Structure Specification**	_____	**MESH** **REGION** **ELECTRODE** **DOPING**
2. **Material Model Specification**	_____	**MATERIAL** **MODELS** **CONTACT** **INTERFACE**
3. **Numerical Method Selection**	_____	METHOD
4. **Solution Specification**	_____	**LOG** **SOLVE** **LOAD** **SAVE**

Figure 4.1 Padre Command groups with the primary statements in each group.

4.2. Choice of the numerical method

Several different numerical methods can be used for calculating the solutions to semiconductor device problems. Different solution methods are optimum in different situations and some guidelines will be given here. Different combinations of models will require padre to solve up to six equations. For each of the model types there are basically three types of solution techniques: (a) de-coupled (GUMMEL), (b) fully coupled (NEWTON) and (c) BLOCK that are specified on the SYSTEM line. In simple terms, the de-coupled technique like the Gummel method will solve for each unknown in turn keeping the other variables constant, repeating the process until a stable solution is achieved. Fully coupled techniques such as the Newton method solve the total system of unknowns together. The coupled or block methods will solve some equations fully coupled, while others are decoupled.

In general, the Gummel method is useful where the system of equations is weakly coupled, but has only linear convergence. The Newton's method is useful when the system of equations is strongly coupled and has quadratic convergence. The Newton method may however spend extra time solving for quantities which are essentially constant or weakly coupled. Newton also requires a more accurate initial guess to the problem to obtain convergence. Thus, a coupled method can provide for faster simulation times in these cases over Newton. Gummel can often provide better initial guesses to problems. It can be useful to start a solution with a few Gummel iterations to generate a better guess and then switch to Newton to complete the solution. The

following specifies a PDE system for a simulation with only holes and using the Gummel method:

System Gummel Carr=1 holes

The Poisson equation is always solved, and optionally one can specify that continuity and/or energy balance partial differential equation (PDEs) be solved for the carriers. The continuity equations can be selected by setting carriers to the number of carriers (0, 1, 2) with the single carrier defined through the electrons and/or holes parameter; the default is electrons for carriers = 1. In general, one can specify under the system line a general nonlinear construct (with arbitrary groupings of PDEs) using the COUPLING parameter. In this case the PDEs are selected using an integer code as follows:

1. Poisson
2. Electron continuity equation
3. Hole continuity equation
4. Electron energy balance equation
5. Hole energy balance equation

4.3. Solutions obtained

PADRE can calculate DC, AC small signal, and transient solutions. Obtaining solutions is rather analogous to setting up parametric test equipment for device tests. One usually defines the voltages on each of the electrodes in the device. PADRE then calculates the current through each electrode. PADRE also calculates internal quantities, such as carrier concentrations and electric fields throughout the device. This is information that is difficult or impossible to measure.

In all simulations the device starts with zero bias on all electrodes. Solutions are obtained by stepping the biases on electrodes from this initial equilibrium condition. Results are saved using the Log or Plot statements.

The following example performs an initial bias point, saving the solution to the data file OUT0:

```
Solve init outf=out0
```

In the next example, bias stepping is illustrated. The two solve lines produce the following bias conditions:

Bias point #	V1	V2	V3
1	0.0	0.5	-0.5
2	1.0	0.5	0.0
3	2.0	0.5	0.0
4	3.0	0.5	0.0
5	4.0	0.5	0.0
6	5.0	0.5	0.0

The solutions for these bias points will be saved to the files OUT1, OUTA, OUTB, OUTC, OUTD and OUTE. Note that the initial guess for the first bias point is obtained directly from the preceding solution because the PREVIOUS option was specified. The initial guesses for bias points 2 and 3 will also be obtained as if PREVIOUS had been specified since two electrodes (numbers 1 and 3) had their biases changed on bias point 2. However, for bias points 4, 5 and 6, PADRE will use a projection to obtain an initial guess since starting with bias point 4, both of its preceding solutions (bias points 2 and 3) only had the same electrode bias (number 1) altered.

```
Solve Prev V1 = 0 V2 = 0.5 V3 = -0.5   OUTF = OUT1
Solve Prev V1 = 1 V2= 0.5 V3 = 0   Vstep=1  Nstep=4
+ Electr =1 OUTF = OUTA
```

Here is a case where two electrodes are stepped (2 and 3). The bias points solved for will be (0,0,1), (0,0.5,1.5), (0,1,2) and (0,2,3). *PADRE* will use the *PROJECT* option to predict an initial guess for the third and fourth bias points since the bias voltages on both electrodes 2 and 3 have been altered by the same amount between each point.

```
Solve Proj V1=0 V2=0 V3=1Vstep = 0.5 Nsteps= 2 Elect = 23
Solve Proj V2=2 V3=3
```

If no new voltages are specified and a VSTEP is included, the first bias point solved for is the preceding one incremented appropriately by VSTEP. This is illustrated by repeating the above example as a three line sequence:

```
Solve   V1 = 0    V2 = 0      V3 = 1
Solve   proj      Vstep = 0.5   Nsteps = 2   ELECT=23
Solve   proj      Vstep= 1      Nsteps = 1   ELECT=23
```

The following sequence is an example of a time-dependent solution. The *METHOD line* specifies the second-order discretization and automatic time-step selection option, along with Newton-Richardson. The first *SOLVE* line then computes the solution for a device with 1 volt on V1 and 0 on V2 in steady-state. The second SOLVE line specifies that V1 is to be ramped to 2 volts over a period of 10ns and is left on until 25 ns. Each solution is written to a file; the name of the file is incremented in a manner similar to that described above for a dc simulation (UP1, UP2, etc.).
Note that an initial time step had to be specified on this line. The third SOLVE line ramps V1 down from 2 volts to -1 volts in 20 ns (end of ramp is at t = 45ns). The device is then solved at this bias for another 55 ns (out to 100 ns). Note that again each solution is saved in a separate file

(DOWN1, DOWN2, etc.) and that no initial time-step was required since one had been estimated from the last transient solution for the previous *SOLVE line.* Finally, the fourth *SOLVE* line performs the steady-state solution at V1=-1 and V2=0.

```
Method 2nd  TautoETHOD 2ND TAUTO AUTONR
Solve V1=1 V2=0
Solve V1=2 Tstart =1E-12 TSTOP=25E-9 RAMPTIME=10E-9
+ OUTF = UP1
Solve V1 = -1 Tstop = 100E-9 Ramptime = 20E-9 OUTF=DOWN1
Solve V1 = -1 V2 = 0
```

In all of the above examples, to obtain convergence for the equations used it is necessary to supply a good initial guess for the variables to be evaluated at each bias point. The PADRE solver uses this initial guess and iterates to a converged solution. Provided a reasonable grid is used, almost all convergence problems in PADRE are caused by a poor initial guess to the solution. During a bias ramp, the initial guess for any bias point is provided by a projection of the two previous results. Problems tend to appear near the beginning of the ramp when two previous results are not available. If one previous bias is available, it is used alone.

The LINALG line sets parameters associated with the numerical methods used for the solution of linear algebraic systems. There can be more than one LINALG line in a single simulation, so that parameters can be altered. Subsequent LINALG lines only alter those coefficients specified. The METHOD, PRECONDITION and ACCELERATION vectors determine the basic method, the preconditioned and the iterative acceleration technique used to perform the linear algebraic solutions. Each of these consists of a set of single digits corresponding exactly to the PDE groupings on the SYSTEM line, defined as follows:

```
Method =          1     Direct sparse LU decomposition
```

	2	Point iterative
	3	ABF
Precondition =	1.	LDU
	2.	ILU
	3.	Block-ILU
Acceleration =	1	Conjugate gradients
	2	Orthomin
	3	Biconjugate gradients
	4	CGS
	5	Generalized conjugate residuals
	6	GMRES
	7	CGNE
	8	LSQR
	9.	USQR
	10.	BICR
	11.	BICG/CGS
	12.	BICG1
	13.	BICG2

4.4. Advanced solution techniques

4.4.1. Obtaining Solutions around Breakdown Voltage

Obtaining solutions around the breakdown voltage can be difficult using the standard PADRE approach. It requires special care in the choice of voltage steps and also in interpreting the results. The curve tracer is the most effective method in many cases. The following statement traces an IV curve using terminal 4. Extrapolation is used to obtain initial guesses, and the simulation will

terminate when V4 is 5 volts or I4 is 1E-4 Amps/mum. The initial step will be <= .05*5 = .25v on V4 and <= 5E-6A in I4.

CONTIN ELECT=4 SIGMA=.05 VMAX=5 IMAX=1E-4 TOL.SIG=5E-3 EXTRA

4.4.2. Using Current Boundary Conditions

In all of the examples considered in the basic description of the SOLVE statement it was assumed that voltages were being forced and currents were being measured. PADRE also supports the reverse case through current boundary conditions. The current through the electrode is specified in the SOLVE statement and the voltage at the contact is calculated. Current boundary conditions are set using the CONTACT statement. The syntax of the SOLVE statement is altered once current boundary conditions are specified. Often it is best to ramp the voltage until the current is above 1pA/mm and then switch to current forcing. When interpreting the results, it is important to remember the calculated voltage on the electrode with current boundary conditions is stored as the ‘internal biases.

Pseudo continue

TRAp	=	logical	(default is true)
DGmin	=	real	(default is 3.0e-9)
A.Trap	=	real	(default is 0.5)
N.Trap	=	real	(default is 2)
M.Trap	=	real	(default is 1)
I.Trap	=	real	(default is 10)
MAxneg	=	integer	(default is 10)

DI.Trap	=	real	(default is 1.0e-5)
DV.Trap	=	real	(default is 1.0e-5)
OUT.Trap	=	logical	(default is true)
IGN.Inner	=	logical	(default is false)
STop	=	logical	(default is true)

TRAP specifies the initiation of a "pseudo-continuation" method which attempts to detect if a solution process starts to diverge. If it does, the electrode bias steps taken from the initial guess are reduced by the multiplicative factor A.TRAP (up to a maximum of I.TRAP cutbacks or until DI.TRAP or DV.TRAP are violated - see below). One criterion for determining divergence is that the maximum number of Newton iterations is exceeded with any group of PDEs. Also, PADRE can monitor internal quantities - e.g. the norm of the residuals (RHSNORM), the norm of the updates (XNORM) and the size of any damping coefficient - during the Newton iterations if it is clear that convergence is slow (this feature can be disabled by specifying IGN.INNER).
Divergence tendencies are ignored if the residual norms are below a small threshold (to account for round-off error), DGMIN. N.TRAP and M.TRAP specify the number of Newton iterations to be ignored before checking and the consecutive number of occurrences on which the norm must go up before the TRAP feature is enabled. A trap can also be initiated by reaching a maximum number of consecutive iterations which produce negative concentrations (MAXNEG). DI.TRAP and DV.TRAP specify the minimum size of the allowed bias steps after application of A.TRAP;DV.TRAP is given in volts, while DI.TRAP is defined in a relative sense. OUT.TRAP controls whether solution output files are written for any points computed due to a trap; the file names used are derived from the name given to the originally attempted point, with the string "Xn" (n is a digit, starting from 0) appended. If the TRAP feature is not used, setting the STOP parameter will force the simulator to suspend execution if convergence is not reached.

Damping

DAMPEd	=	character (default is "single") [Expert]

ITDamp	=	integer (default is 1) [Expert]
Delta	=	real (default is 1.0e-6) [Expert]
DAMP Loop	=	integer (default is 10) [Expert]
DFactor	=	real (default is 10.0) [Expert]
DPower	=	real (default is 2) [Expert]
DVLimit	=	real (default is 10.0) [Expert]
TRUncate	=	logical (default is false) [Expert]
VMargin	=	real (default is 0.01*kT/q) [Expert]

The *DAMPED* parameter indicates the use of a more sophisticated damping scheme proposed by Bank and Rose (this is the recommended option, particularly for large bias steps). The method is recommended for single PDE loops (DAMPED="single"), but it can also be performed for coupled groups of PDEs (DAMPED="all"). To turn this damping method off, use DAMPED="none". The Bank/Rose damping scheme is controlled by the parameters ITDAMP, DELTA, DAMPLOOP, DFACTOR and DPOWER which specify the iteration on which to start to damp, the reduction threshold, maximum number of damping loops, update reduction factor and update reduction power respectively. Other damping schemes that can be performed either separately or in conjunction with Bank/Rose include a simple limit on any potential updates (DVLIMIT) and truncation of all potentials so that they lie between the minimum and maximum possible voltages (TRUNCATE) with a margin on either end of VMARGIN.

4.5. Plot and log statement

4.5.1. Plot Statement

The PLOT.1D line plots a specific quantity along a line segment through the device (mode A), or plots an I-V curve of data (mode B). These data are then used as inputs in Rappture for efficient graphical presentation.

X.Start or A.X	=	real
Y.Start or A.Y	=	real
X.End or B.X	=	real
Y.End or B.Y	=	real
Z.pos	=	real (default is 0)
COordinate	=	character

The above parameters define the Cartesian coordinates of the start (A.X, A.Y) and the end (B.X, B.Y) of the line segment along which the specified quantity is to be plotted. By default, the data is plotted as a function of distance from the start (A) on the z-plane closest to the z coordinate given by Z.POS; the data can alternatively be plotted against a particular coordinate ("x", "y" or "z") along the same segment. The line segment may not be defaulted, and it is required in mode A.

function is one of:

POTential	=	logical Mid-gap potential
QFN	=	logical Electron quasi-fermi level
QFP	=	logical Hole quasi-fermi level
N.temp	=	logical Electron temperature
P.temp	=	logical Hole temperature
Doping	=	logical Total net impurity concentration
IOn.imp	=	logical Net ionized impurity concentration
Electrons	=	logical Electron concentration
Holes	=	logical Hole concentration
NET.Charge	=	logical Net charge concentration
NET.CArrier	=	logical Net carrier concentration

J.Conduc	=	logical Conduction current
J.Electr	=	logical Electron current
V.Electr	=	logical Electron velocity
J.Hole	=	logical Hole current
V.Hole	=	logical Hole velocity
J.Displa	=	logical Displacement current
J.Total	=	logical Total current
E.Field	=	logical Electric field
REcomb	=	logical Net recombination
BAND.Val	=	logical Valence band potential
BAND.Con	=	logical Conduction band potential
or		
X.Axis	=	character
Y.Axis	=	character
Freq	=	real
INFile	=	character

The above parameters specify the quantity to be plotted. There is no default. In mode A, one of the solution variables is plotted versus distance into the device. For vector quantities, the magnitude is plotted. In mode B, terminal characteristics can be plotted against each other by choosing the value to be plotted on each axis (XAXIS =, YAXIS =). Quantities available for plotting include applied device biases

(XAXIS / YAXIS= VA1, VA2, ... , VA9, VA0), actual contact bias which may differ from applied bias in the case of lumped element boundary conditions (V1,V2, etc.), terminal current (I1,I2, etc.), AC capacitances (C11,C12,C21, etc.), AC conductance (G11,G12,G21, etc.) and AC admittance (Y11,Y12,Y21, etc.). Also, voltages at circuit nodes can be plotted as "Vnode_name" where "node_name" is the name of the node, and currents for labeled lumped elements can be plotted by name as "Iname" where "name" is the lumped element name (see LUMP.ELEMENT

line). Additionally, any of the voltages or currents can be plotted versus time for transient simulations, and any AC quantity can be plotted versus frequency. For plotting AC parameters versus bias at a particular frequency, use the FREQUENCY parameter. The values plotted are the I-V or AC data of the present run, provided a log is being kept (see the LOG line). Alternatively, a different log file can be loaded with INFILE.

Data-control

ABsolute	=	logical (default is false)
Logarithm	=	logical (default is false)
X.Log	=	logical (default is false)
DEcibels	=	logical (default is false)
INTegral	=	logical (default is false)
NEGative	=	logical (default is false)
INVerse	=	logical (default is false)
D.order	=	real (default is 0)
X.Component	=	logical (default is false)
Y.Component	=	logical (default is false)
MIx.mater	=	logical (default is false) [Expert]
SPline	=	logical (default is false)
NSpline	=	logical (default is 100)

ABSOLUTE specifies that the absolute value of the variable be taken. For rapidly varying quantities, the LOGARITHM (X.LOG) is often more revealing. To get the true logarithm of a quantity, specify ABSOLUTE and LOGARITHM - the absolute is taken first and there is no danger of negative arguments. DECIBELS convert the function to a measure of gain in decibels defined by INTEGRAL plots the integral of the specified ordinate, and NEGATIVE negates the ordinate values; INVERSE plots the value of the function raised to the (-1) power. D.ORDER0

specifies that the derivative of the function (order=D.ORDER) with respect to the ordinate be plotted.

X.COMPONENT and Y.COMPONENT force the x and y components respectively of any vector quantities to be plotted as opposed to the default total magnitude. MIX.MATER specifies that local vector averaging should be done over all materials to which the node belongs as opposed to just the hierarchical choice (see the PRINT line). The SPLINE option indicates that spline-smoothing should be performed on the data using NSPLINE interpolated points (maximum is 500).

Examples

The following plots a graph of potential along a straight line from (0.0,0.0) to (5.0,0.0):

```
PLOT.1D POTEN A.X=0 A.Y=0 B.X=5 B.Y=0
```

In the next example, the log of the electron concentration is plotted from (1.0,-0.5) to (1.0, 8.0) with bounds on the plotted electron concentration of 1.0e10 and 1.0e20. A spline interpolation is performed with 300 interpolated points. The non-spline-interpolated points are marked.

```
PLOT.1D ELECT LOG A.X=1 A.Y=-.5 B.X=1 B.Y=8
+ MIN=10 MAX=20 SPLINE NSPL=300 POINTS
```

In the following example, the current in contact 1 is plotted as a function of contact 2 voltage, then the curve is compared with a previous run.

```
PLOT.1D X.AXIS=V2 Y.AXIS=I1
PLOT.1D X.AXIS=V2 Y.AXIS=I1 INF=logf0 UNCH
```

The following plots the actual contact voltage on a contact versus the applied voltage.

```
PLOT.1D X.AXIS=V3 Y.AXIS=VA3
```

Finally, the following shows a plot of two capacitance components versus the log of frequency. A different line type is chosen for the second component.

```
PLOT.1D X.AXIS=FREQ Y.AXIS=C21 X.LOG
PLOT.1D X.AXIS=FREQ Y.AXIS=C31 X.LOG UNCH LINE=4
```

4.5.2. Log Statement

The LOG line allows the I-V and/or AC characteristics of a run to be logged to disk. Then, they are ready for plotting using Rappture. Any I-V or AC data subsequent to the line is saved. If a log file is already open, it is closed and a new file opened. The following example specifies the command: Save the I-V data in a file called IV1 and AC data in AC1.

```
LOG OUTF=IV1 ACFILE=AC1
```

Chapter 5

Simulation Results and Discussion

To investigate and understand the theoretical performances of the DC and transient characteristics of the proposed SiGe heterojunction bipolar transistor, PADRE, the 2D/3D

dimensional simulator, is used for simulation purpose. Based on the physical structures created on Structure Specification line, Padre calculates the electrical characteristics, which are associated with the specified physical structure and bias conditions. It provides the internal device mechanism or device behavior by solving the well-established drift-diffusion transport equation. Drift Diffusion calculations are carried out using appropriate physical models. To perform this, physical models such as concentration dependent mobility, field dependent mobility, and Klaassen mobility are used and the band gap narrowing effect is taken into account. Carrier statistics are performed by defining Fermi-Dirac distribution and minority carrier lifetime including the effect of Shockely-Read-Hall and Auger recombination mechanism.

5.1 **DC characteristics**

DC characteristic simulation is performed to illustrate some of the advantages seen in a SiGe transistor over that of a Si transistor. The schematic cross section of the common emitter SiGe HBT is indicated in figure 5.1 below. It is this physical structure, generated on the PADRE Device Simulator at Structure Specification line, which typically generated by the comprehensive results of Mesh line, Region line, Doping line and Electrode line.

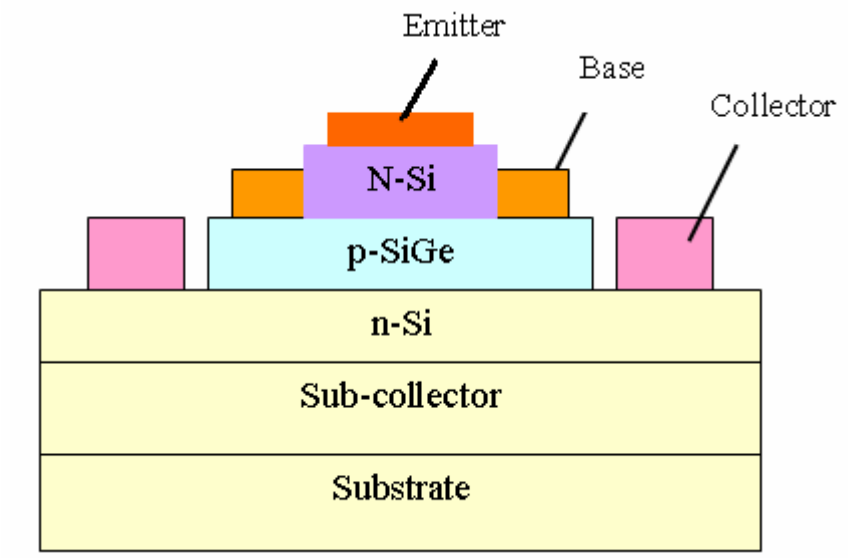

Figure 5.1 Schematic diagram of the lateral cross section of an npn SiGe HBT

Figure 5.2 and figure 5.3 show a plot of simulated conduction energy band diagram of Si BJT and SiGe HBT respectively. To fully characterize the energy band diagram of the transistors, valence band of the transistors have also to be simulated. Thus, figure 5.4 and figure 5.5 show a plot of the simulated valence band diagram of Si and SiGe HBT transistors respectively.

The combination of fig. 5.2 and fig. 5.4 give a full energy band diagram of a Si BJT which helps for the comparison of physical device parameters with the full energy band diagram of a SiGe HBT given by the combination of fig.5.3 and fig.5.5. Introducing an alloy of SiGe in the base of a bipolar transistor provides a valence band offsets in the base (fig.5.5) that creates a drift field for the enhanced transport of electrons. The valence band offset in the HBT represents itself as a narrowing of the energy bandgap and it is 22.5 meV.

The simulated energy band diagram and various distribution plots under thermal equilibrium condition at room temperature for an npn SiGe HBT and Si BJT have neutral contacts. All plots were extracted vertically along the center of Base-collector junction of the device, beginning beneath the emitter contact and ending midway through the substrate.

The parameters used in device simulation of the transistors are given in the table below.

Parameters	Symbol	Si BJT	SiGe HBT
Active device area	A	$10\,\mu m \times 1\mu m$	$10\,\mu m \times 1\mu m$
Surface emitter doping	N_{E1}	5×10^{19}	5×10^{19}
Emitter doping	N_{E2}	10^{20} cm^{-3}	1×10^{19} cm^{-3}
Base doping	N_B	10^{18} cm^{-3}	2×10^{21} cm^{-3}
Surface Emitter thickness	W_{E1}	$2\,\mu m$	$2\,\mu m$
Emitter thickness	W_{E2}	$2\ \mu m$	$2\ \mu m$
Base thickness	W_B	$1\ \mu m$	$1\ \mu m$
Collector thickness	W_C	$2\ \mu m$	$2\ \mu m$
Epitaxial layer thickness (W_E+W_B+W_C)	W_{epi}	$4\ \mu m$	$4\ \mu m$
Collector doping	N_C	10^{15} cm^{-3}	1×10^{17} cm^{-3}
Sub-Collector doping	N_{SUB}	10^{19} cm^{-3}	10^{19} cm^{-3}
Sub-Collector thickness	W_{SUB}	$1\ \mu m$	$1\ \mu m$
Germanium composition	Ge	0	0.2

Table 5.1 Design constraints used in SiGe HBT and Si BJT

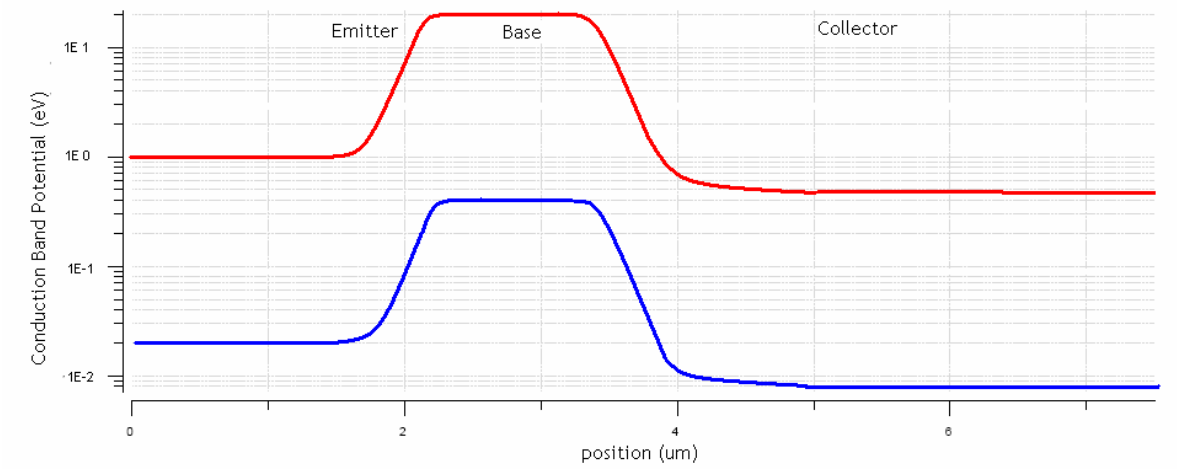

Figure 5.2 Simulated Conduction band of Si bipolar transistor (Si BJT)

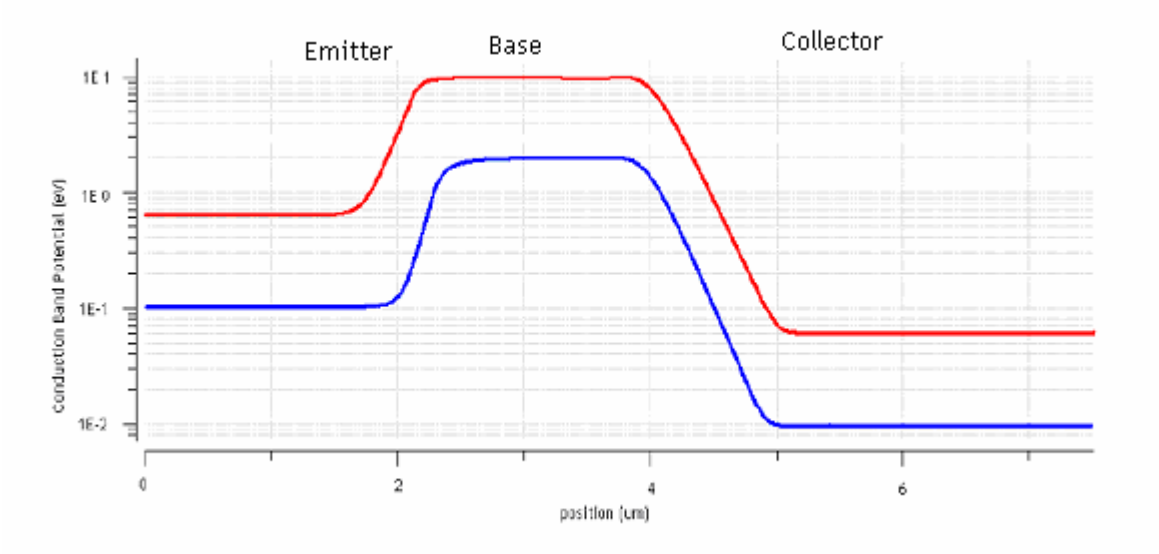

Figure 5.3 Simulated conduction band of SiGe Heterojunction Bipolar Transistor

In examining the energy band diagram, a forward-bias applied to the emitter-base junction has simultaneously lowered the potential barrier for conduction band electrons in the emitter to flow into the base and valence band holes in the base to flow in the emitter.

The curve with red color at figure 5.2 indicates conduction band of Si BJT with emitter-base is forward biased shows a reduction in height when compared with blue color curve in fig.5.2 which is at no bias condition. The barrier potential for holes coming from the base to emitter side decrease through bias which is shown at the valence band diagram at fig.5.2

As shown from figure 5.3. Conduction energy band of SiGe is lower than that of Si BJT which is a result of Ge incorporation in the base of SiGe HBT. The figure shows a small slope in the conduction energy band in the base for a SiGe transistor with 20 % mole fraction, whereas a flat conduction energy band is observed for a Si transistor.

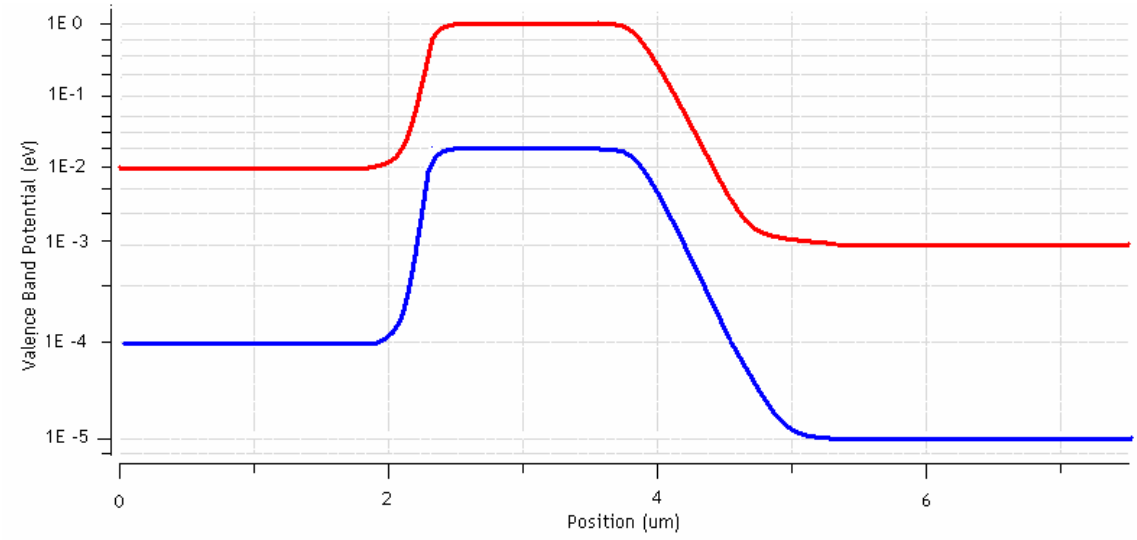

Figure 5.4 Simulated valence band diagram of Si Bipolar transistor

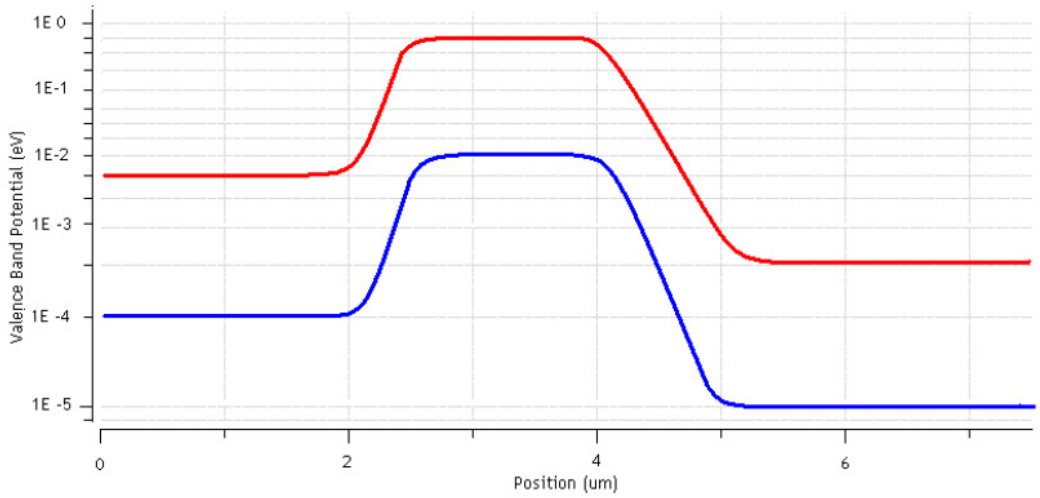

Figure 5.5 Simulated valence band diagram of SiGe HBT

The simulated common emitter configuration of the lateral SiGe HBT and Si BJT are illustrated in fig. 5.3 and fig. 5.4. Fig 5.5 and fig 5.6 shows the Gummel plot of base current of SiGe HBT and Si BJT which are results of simulation for a fixed
Collector-emitter base bias voltage (V_{CE}=3V) respectively.

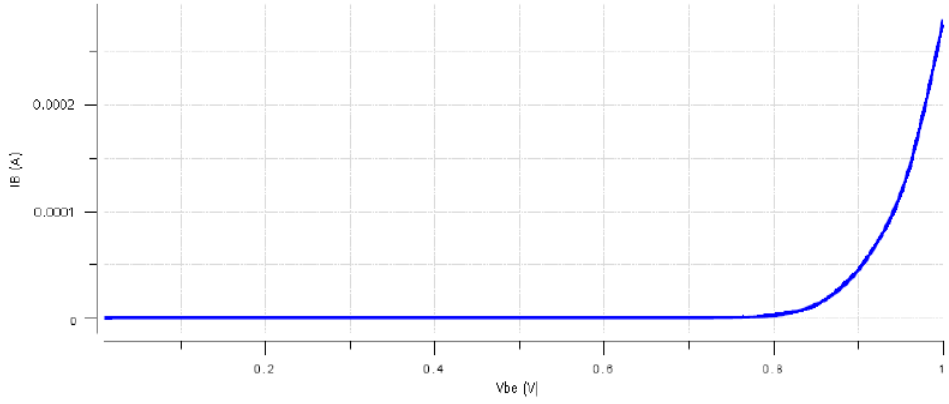

Figure 5.6 Simulated base current of a common emitter of Si BJT

It is observed that Si BJT exhibits a lower collector current than that of the npn HBT. The SiGe material properties used for the simulation purpose are taken from the theoretical works.

The collector current of a 20 % Ge in a SiGe transistor is about 1.5 times larger compared to the collector current of the Si transistor. The 1.5 increase in the collector current (and the current gain) of the HBT compared to the homojunction transistor is due to the narrower bandgap in the base. Since the base current of silicon and SiGe HBTs are virtually identical, the current gain enhancement due to germanium incorporation is similar to the collector current enhancement. One reason for the amount of collector current for the SiGe transistor is not as large because of the small amount of the Ge mole fraction used.

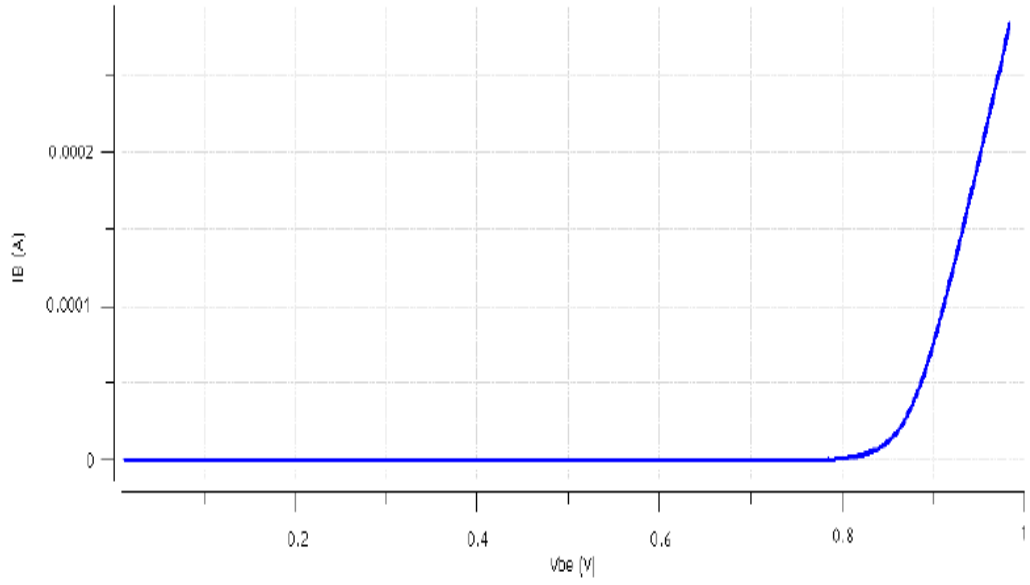

Figure 5.6 Simulated Base current Vs Base-Emitter biase voltage of SiGe HBT

The Current-voltage characteristics of the SiGe base npn transistor are compared in fig. 5.7 and fig. 5.8 with that of the lateral Si npn BJT.

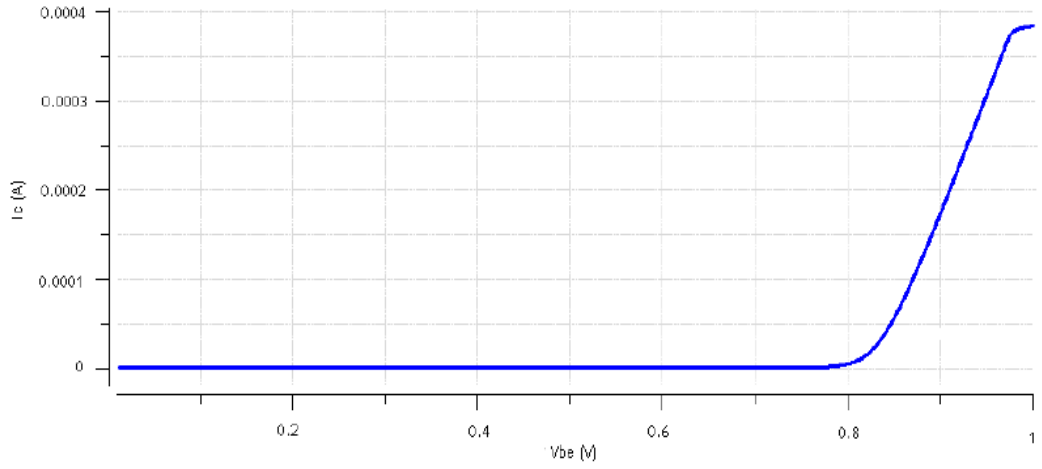

Figure 5.7 Simulated collector current Vs Base-Emitter voltage of SiGe HBT.

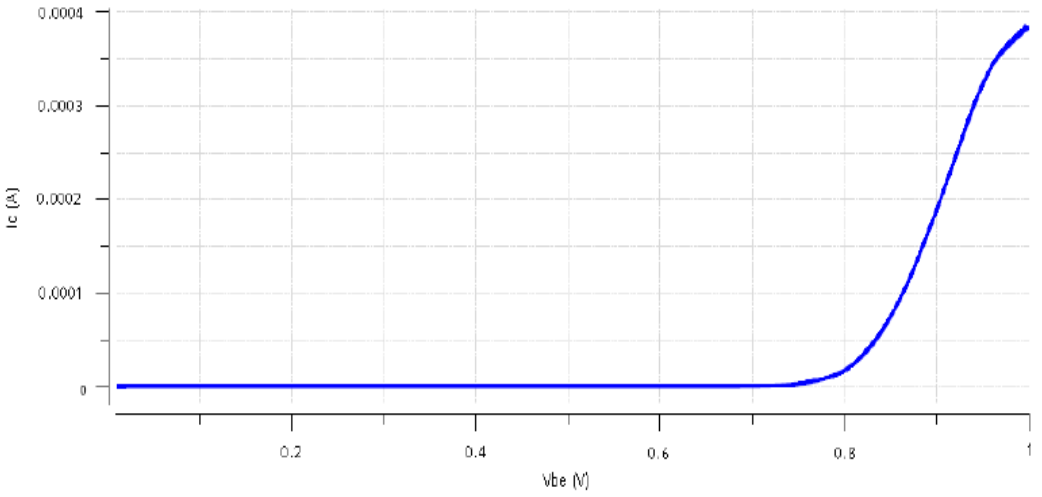

Figure 5.8 Simulated collector current Vs Base-Emitter voltage of Si BJT

The current gain performance of the npn SiGe HBT is compared with npn Si BJT for identical device dimensions and bias conditions but for different doping densities. The physical parameters and doping profiles in the different regions of these structures are shown in fig.5.1. The surface emitter doping of 5×10^{19} cm^{-3} and its thickness w_{E1} of 2 μm is chosen to provide ohmic contact. The emitter doping of 1×10^{19} cm^{-3} and its thickness w_{E2} of 2μm is selected to lower

emitter-base junction capacitance. This two step emitter profile is typical of the heterostructure devices, provided the heterostructure

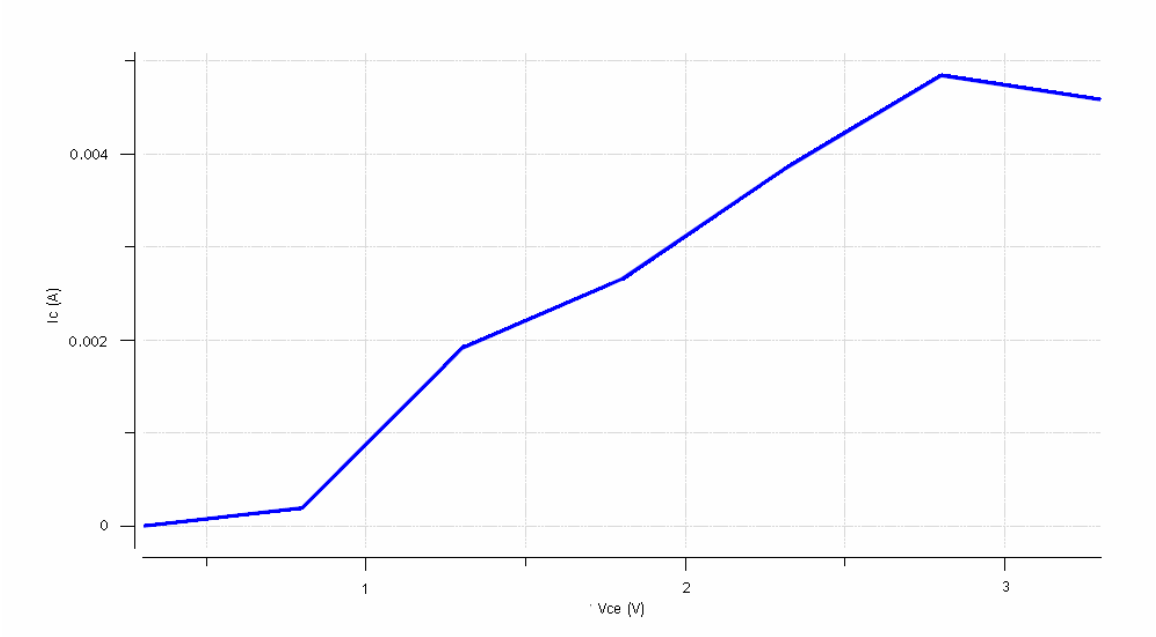

Figure 5.9 Simulated Collector current Vs Collector-Emitter voltages of SiGe HBT

device has an inherent high current gain which could be traded-off to achieve lower internal emitter doping of 8×10^{8} cm^{-3} is chosen in both the structure.

V_{be} (V)	0.8	0.85	0.9	0.95	1.0
I_C (mA)	10e-3	50e-3	175e-3	0.3	0.375
I_B (mA)	10e-3	20e-3	75e-3	0.2	0.3
$\beta^{Si} = \frac{I_C}{I_B}$	1	2.5	2.33	1.5	1.25

Table 5.2 Base Current, Collector Current and Current Gain of Si BJT at Base-Emitter bias Voltage.

V_{be} (V)	0.8	0.85	0.9	0.95	1.0
I_C (mA)	25e-3	75e-3	180e-3	325e-3	375e-3
I_B (mA)	15e-3	125e-3	50e-3	125e-3	300e-3
$\beta^{SiGe} = \frac{I_C}{I_B}$	1.66	6	3.6	2.6	1.25

Table 5.3 Base Current, Collector Current and Current Gain of SiGe HBT at base-emitter bias Voltage.

In the design of a BJT for DC operation in the common emitter configuration, a high current gain is desirable neglecting space charge recombination; referring equation 5.1, the DC current gain is directly proportional to the emitter-to-base doping and thickness ratio, and inversely proportional to the minority carrier diffusivity. But in this thesis, the current gain is not high because thickness of the base is about 1um which makes the gain to be less than that of transistors with base thickness of 300 nm, 325 nm and 500 nm [62]. The reason is that the

emitter-to-base thickness ratio decreases as the base thickness increases further which decreases current gain.

$$\beta \cong \frac{W_E D_n N_{d,E}^{+}}{W_B D_p N_{a,E}^{-}} \tag{5.1}$$

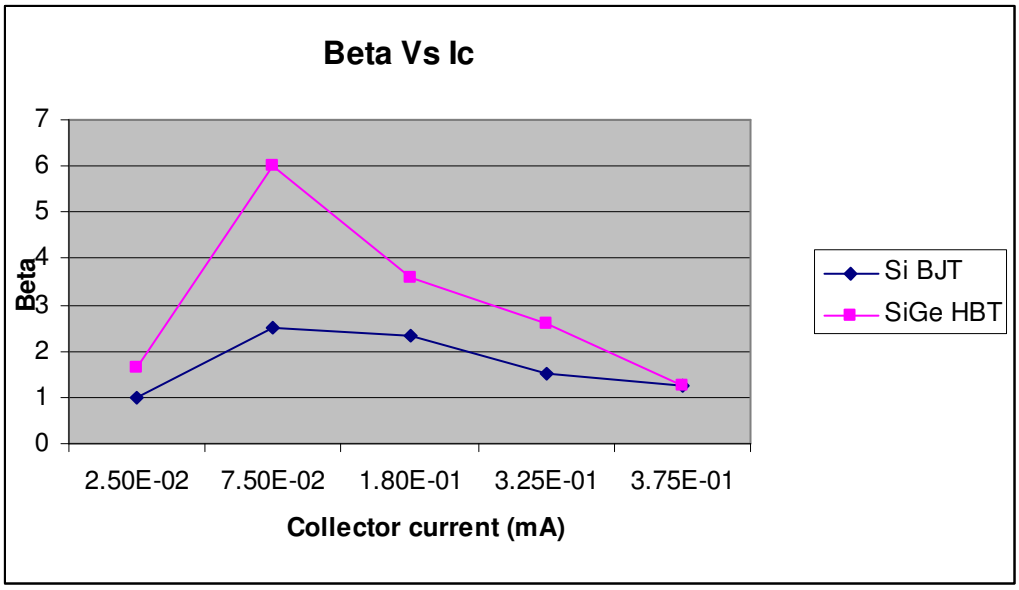

Figure 5.10 Comparison of Current Gain of Si BJT and SiGe HBT

As can be seen in figure 5.10, a 20 % uniform Ge box profile incorporation to the base of a Si transistor produces the 2.4 fold increase in β for at 300 K, since the enhancement depends exponentially on the bandgap reduction at the emitter base junction. In the conventional Si BJT, β is inversely proportional to the integrated base charge. Since base doping cannot be increased indefinitely while maintaining adequate β, the flat Ge profile is particularly useful in realizing a transistor with lower intrinsic base resistance. The calculated result of β of SiGe HBT using equation 5.1 is about 11.5 and comparing it with 6 of the simulated β has a deviation of 47 %.

Chapter 6
Conclusion

6.1. Summary

Motivated by the limitation in trade-off between doping profile and DC and AC characteristics in the conventional design of Si BJT, the device structure with a modified doping profile of the base was proposed. The DC characteristics simulations of the device structure are performed and the results of the simulation support and maintain the basic principles of semiconductor devices. This is realized by the plots of energy band diagram which show the bandgap of SiGe HBT is narrower than Si BJT, due to the incorporation of Ge into the base of the transistor which lowers barrier height to electrons and reduces base resistance to the flow of currents. Hence, the performance of the SiGe HBT is better than Si BJT in terms of DC characteristics. This is recognized by

- Increase in collector current and
- Increase in current Gain.

.

But due to the operational inability of the command "extract" in the proposed device simulator ***PADRE***, AC characteristics of device has not been studied through simulation.

6.2 Future work

Future work should include simulation of the AC characteristics of the SiGe HBT for high speed and high frequency application. In order to have better values of collector current and current gain, meshing of the device has to be fine enough and dimensions of region has to be minimized which goes with power of device simulator version and type of simulator. So in order to have better performance of what has been simulated, one has to define the device with fine meshes and has to create minimized regions with well known device simulator like Industrial PADRE Version Simulator, Commercialized Medici device simulator and Silvaco-Atlas device simulator.

References

[1] Shockley W, ‘’US patent, Specification 2569347’’, 1951

[2] Kroemer H, ‘’Theory of a wide-gap emitter for transistor’’, Proc. IRE 45, 1535-7, 1957

[3] Sze S M Physics of semiconductor Devices, 2nd, edn, John Wiley, 1981

[4] Slotboom J W and de Graaff H C, Measurement of bandgap narrowing in Si bipolar transistor, Solid-State Electron **19,** 857-62, 1976

[5] Kromer H, Heterojunction bipolar transistor and integrated circuits, Proc. IEEE **70,** 13-25, 1982

[6] Vaidyanathn M and Roulston D J, “Effective base-collector time constants for calculating he maximum oscillation frequency of bipolar transistor” Solid-State Electron, **38,** 509-16, 1995

[7] Vaidyanathn M and Pulfrey DL, “Extrapolated f_{max} of heterojunction bipolar transistor”, IEEE Trans. Electron Devices, **46,** 301-9, 1999

[8] Tang Y T “Advantage characteristics and modeling of SiGe HBTs”, PhD Thesis, University of Southampton, 2000

[9] Harame D L, Comfort J H, Cresseler J D, Crabble E F, SunJ Y-C, Meyerson B S and Tice T “Si/SiGe epitaxial-base transistor”, part I: materials, physics and circuits, IEEE Tran. Devices, **46,** 301-9, 1995

[10] Kroemer H “Two integral relations pertaining to the electron transport through a bipolar with a non-uniform energy gap in the base region”, Solid-state Electron, **28,** 1101-3,1985

[11] Slotboom J W, Streuker G, Pruijmboom A and Gravesteijin D J “Parasitic energy barriers in SiGe HBTs”, IEEE Electron Device Lett., **12,**486-8, 1991

[12] Kay L E and Tang T-W, “Monte Carlo calculation of strained and unstrained electron Mobilities in $Si_{1-x}Ge_x$ using an improved ionized impurity model”, J.Appl. Phys., **70,** 1483-1488, 1991.

[13] Ashburn P Design and Realization of bipolar transistor (chichester: Wiley), 1988

[14] Lindmayer J and Wrigley C, “The high injection level operation of drift transistors”, Solid-State Electron, **2,** 79-84, 1961

[15] Early J M, "Effects of space-charge layer widening in junction transistors", Proc. IRE, **40,** 1401-6, 1952

[16] Pepole R , "Physics and application of Ge_xSi_{1-x}/Si strained layer heterostructure ", IEEE J. Quantum Electron, **22,** 1696-710, 1986

[17] Harame D L, strok J M C, Meyerson B S, Crabbe E F, Scilla G J, de Fresart E, Megadanis A C, Stanis C L, patton G L, comfort J H, Bright A A, Johnson J B and Furkay S S, "30 GHz polysilicon-emitter and Single-Crystal –emitter graded SiGe-base pnp transistors", IEEE IEDM Tech. Dig. 33-6, 1900

[18] Kirk C T, "A theory of transistor cut-off frequency f_T falloff at high current densities" IRE Trans. Electron Devices, **9,** 164-74, 1962

[19] Webster W M, "On the variation of junction-transistor current amplification with emitter Current", Proc. IRE, **42**, 914-20, 1954

[20] Cottrel P and Yu Z, "velocity of Saturation in the collector of Si/Ge_xSi_{1-x}/Si HBTs" IEEE Electron Device Lett., **11,** 431-3, 1990

[21] Yu Z, Cottrell P E and Dutton R, "Modeling and Simulation of high level injection behavior in double heterojunction bipolar transistors", IEEE BCTM, Proc.pp 192-4, 1990

[22] Gao G-B, Fan Z-F and Morko H 1991 Analysis of cut-off frequency roll-off at high currents in SiGe double-heterojunction bipolar transistor Appl. Phys. Lett. 58 2951-3

[23] Mazhari B and Morko H 1991 Effect of collector-base valence discontinuity on kirk effect in double-heterojunction bipolar transistor Appl. Phys. Lett. 59 2962-4

[24] Meyer RG and Muller R S 1987 charge-control analysis of the collector base space charge region contribution to bipolar transistor time constant τ_t IEEE Trans. Electron Devices **34** 450-2

[25] Smith P, Inoue M and Frey J 1980 Electron velocity in Si and GaAs at very high electric fields Appl. Phys. Lett. **37** 797-8

[26] Pritchard R L 1955 High-frequency power gain of junction transistor Proc. IRE **43** 1075-85

[27] Roulston D J 1990 Bipolar Semiconductor Devices (Singapore: McGraw-Hill)

[28] Patton G L, strok J M C, comfort J H, Crabble E F, Meyerson B S Harame D L and

Sun J Y-C 1990 SiGe-base heterojunction bipolar transistor: physics and design issues IEEE IEDM Tech. Dig. Pp 13-16

[29] Comfort J H, Patton G L, cresseler J D, Lee W, Crabbe E F, Meyerson B S, Sun Canceled

[30] Gummel H K 1964 A self-consistent iterative scheme for one dimensional steady-state transistor calculations IEEE Trans. Electron Devices 11 455-65

[31] DeMari A 1968 An accurate numerical steady-state one-dimensional solution of the P-N junction solid-state Electron. **11** 33-58

[32] Scharefetter D L and Gummel H K 1969 Large-signal analysis of a silicon read diode Oscillator IEEE Trans. Electron Devices 16 64-77

[33] D'Avanzo D C, Vanzi M and Dutton R W 1979 one-dimensional semiconductor device analysis (SEDAN) Report G-201-5 Stanford University

[34] Selberherr S, Schutz A and Potzl H W 1980 MINIMOS-a Two-dimensional semiconductor MOST Transistor Analyzer IEEE Trans. Electron Devices **27** 1540-50

[35] Franz A F and Franz G A 1985 BAMBI-a design model for power MOSFETs IEEE Trans. Comput.-Aided Des. **4** 177-89

[36] Pinto M R 1985 PISCES-IIB Manual (Stanford,CA: Stanford University)

[37] Silvaco International 1997 Silvaco-ATLAS Manual, Ver 4.0

[38] Technology Modelling Associated 1997 MEDICI, 2D Device Simulator

[39] Stanford University 1994 PISCES-2ET 2D Device simulator

[40] Blotekjaer K 1970 transport equations for electrons in two-valley semiconductors IEEE Trans. Electron Devices **17** 38-47

[41] Baccarani G and Wordeman M R 1985 An investigation of steady-state velocity overshoot in silicon Solid-state Electron. **28** 407 -16

[42] Lin H, Goldsman N and Mayergorz I D 1992 An efficient deterministic solution of the space dependent Boltzmann transport equation for Silicon Solid-State Electron. **35** 33-42

[43] Fischetti M V and Laux S E 1988 Monte Carlo analysis of electron transport in small semiconductor device including bad-structure and space-charge effects phys. Rev. B **138** 9721-45

[44] Ravaioli R 1998 Hierarchy of simulation approaches for hot carrier transport.

[45] Tang J Y F and Laux S E 1986 MONTE: a program to simulate the heterojunction devices in two dimensions IEEE Trans. Comput.-Aided Des. **5** 645-52

[46] Selberherr S 1984 Analysis and Simulation of Semiconductor Devices (Vienna: Springer-Verlag)

[47] A panovich Y, Lyumkis E, polsky B, Shur A and Blakey P 1994 Steady state and transient analysis of submicron devices using energy balance and simplified hydrodynamic models. IEEE Trans. Comput.- Aided Des. **13** 702-7

[48] Arora N D, Hauser J R and Roulston D J 1982 Electron and hole mobility in silicon as a function of concentration and temperature IEEE Trans. Electron Devices **29** 292-5.

[49] Klaassen D B M 1992 A unified mobility model for device simulation-I Model equation and concentration dependence solid-state Electron. **35** 953-9

[50] Caughey D M and Thomas R E 1967 Carrier mobility in silicon empirical related doping and field Proc. IEEE **55** 2192-3

[51] Manasevit H M, Gergis I S and Jones A B 1982 Electron mobility enhancement in epitaxial multilayer Si-$Si_{1-x}Ge_x$ alloys film on (100) Si Appl. Phys. Lett **41** 464-6

[52] Smith C and Welbourn A D 1987 Prospects for a hetero-structure bipolar transistor using a silicon- germanium alloy IEEE BCTM Proc. Pp 57-64

[53] Jorke H and Herzog H J 1985 Mobility enhancement in modulation-doped Si-$Si_{1-x}Ge_x$ superlattice grown by molecular beam epitaxy Proc. 1st Int. symp. Silicon Molecular beam Epitaxy ed J C Bean (Pennington, NJ: Electrochemical Society) pp 352-9

[54] Manku T and Nathan A 1992 Electron drift mobility model for device based on unstrained and coherently strained $Si_{1-x}Ge_x$ grown on (001) Silicon substrate IEEE Trans. Electron Devices **39** 2082-9

[55] Bludau W, Onton A and Heinkle W 1974 Temperature dependence of the bnadgap of Silicon J. Appl. Phys. **45** 1846-8

[56] People R 1986 physics and application of Ge_xSi_{1-x}/Si strained layer

[57] del Alamo J A, Swirhun S and Swanson R M 1985 Simultaneous measurement of hole life time, hole mobility and bandgap narrowing in heavily doped n-type silicon IEDM Tech. Dig. pp 290-3

[58] Swirhun S E, Kwark Y H and Swanson R M 1986 Measurement of electron lifetime, electron mobility and bandgap narrowing in heavily-doped p-type silicon IEEE IEDM Tech. Dig pp 24-7

[59] Klaassen D B M, Slotboom J W and de Graaff H C 1992 Unified apparent bandgap narrowing in n- and p-type silicon Solid-State Electron. **35** 125-9

[60] Harme D L, Strok J M C, Meyerson B S, Hsu K Y J, Cotte J, Jenkins K A, Cresseler J D, Restle P, Crabbe E F, Subbans S, Tice T, Scharf B W and Yasaitis J A 1993 Optimization of SiGe HBT technology for high speed analog and mixed-signal application IEEE IEDM Tech. Dig. pp 874-6

[61] Crabbe E F, comfort J H, cresseler J D, Sun J Y-C and Stork J M C1993 High-low polysilicon-emitter SiGe-base bipolar transistors IEEE electron Device Lett. 14 478-80

[62] Johnson E O, "Physical limitation on frequency and power parameters of transistors" RCA Rev. pp 163-77, 1965

[63] Kasper E, Herzog H J and kibble H A, "One-dimensional SiGe Superlattices grown by UHV epitax", ppl. Phys., **8**199-205, 1975

[64] Larson LE, "Integrated Circuit technology options for RFICs-present status and future directions", IEEE , J.solid-state circuits, **33,** 387-99, 1998

APPENDIX A: **Mobility vs Impurity Concentration for Silicon and Gallium Arsenide (T=300 K)**

Concentration (cm^{-3})	Mobility in Silicon (cm^2/V-s)		Mobility in GaAs (cm^2/V-s)	
	Electrons	Holes	Electrons	Holes
1.0E14	1350.0	495.0	8000.0	390.0
2.0E14	1345.0	495.0	7718.0	380.0
4.0E14	1335.0	495.0	7445.0	375.0
6.0E14	1320.0	495.0	7290.0	360.0
8.0E14	1310.0	495.0	7182.0	350.0
1.0E15	1300.0	491.1	7300.0	340.0
2.0E15	1248.0	487.3	6847.0	335.0
4.0E15	1200.0	480.1	6422.0	320.0
6.0E15	1156.0	473.3	6185.0	315.0
8.0E15	1115.0	466.9	6023.0	305.0
1.0E16	1076.0	460.9	5900.0	302.0
2.0E16	960.0	434.8	5474.0	300.0
4.0E16	845.0	396.5	5079.0	285.0
6.0E16	760.0	369.2	4861.0	270.0
8.0E16	720.0	348.3	4712.0	245.0
1.0E17	675.0	331.5	4600.0	240.0
2.0E17	524.0	279.0	3874.0	210.0
4.0E17	385.0	229.8	3263.0	205.0
6.0E17	321.0	203.8	2950.0	200.0
8.0E17	279.0	186.9	2747.0	186.9
1.0E18	252.0	178.0	2600.0	170.0
2.0E18	182.5	130.0	2060.0	130.0
4.0E18	140.6	90.0	1632.0	90.0
6.0E18	113.6	74.5	1424.0	74.5
8.0E18	99.5	66.6	1293.0	66.6
1.0E19	90.5	61.0	1200.0	61.0
2.0E19	86.9	55.0		
4.0E19	83.4	53.7		
6.0E19	78.8	52.9		
8.0E19	71.6	52.4		
1.0E20	67.8	52.0		
2.0E20	52.0	50.8		
4.0E20	35.5	49.6		
6.0E20	23.6	48.9		
8.0E20	19.0	48.4		
1.0E21	17.8	48.0		