TCAD Based SiGe HBT Advanced Architecture Exploration

Mahmoud Al-Sa'Di

Numéro d'ordre: 4239

THESE

présentée à

L'UNIVERSITE BORDEAUX I

ECOLE DOCTORALE DE SCIENCES PHYSIQUES ET DE L'INGENIEUR

par **Mahmoud AL-SA'DI**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPECIALITE : **ELECTRONIQUE**

TCAD Based SiGe HBT Advanced Architecture Exploration

Soutenue le: 25 Mars 2011

Après avis de :

M. Mikael ÖSTLING	Professeur	KTH Royal Institute of Technology, Stockholm	Rapporteurs
M. Alain CHANTRE	H.D.R	STMicroelectronics, Crolles	Rapporteurs

Devant la commission d'examen formée de:

M. Mikael ÖSTLING	Professeur	KTH Royal Institute of Technology	Rapporteurs
M. Alain CHANTRE	H.D.R	STMicroelectronics	Rapporteurs
M. Sébastien FREGONESE	Chargé de recherche	Université Bordeaux 1	Co-Directeur de thèse
M. Thomas ZIMMER	Professeur	Université Bordeaux 1	Directeur de thèse

Membre invité:

Mme Cristell MANEUX	Maître de Conférences\ H.D.R	Université Bordeaux 1

Thèse préparée au Laboratoire IMS, 351 Cours de la Libération, 33405 Talence Cedex, France.

"IN THE NAME OF ALLAH,

THE BENEFICENT, THE MERCIFUL"

To the soul of my father

To my mother

To Christine

Acknowledgment

I would like to express my honest appreciation to all the people who ever gave me help and support during my PhD study period. Particularly, I would like to thank my research advisor Prof. Thomas Zimmer for his continual instruction and encouragement. I have greatly benefited from his profound knowledge and charming personality. I would like to thank my co-advisor Dr. Sebastien Fregonese for his guidance and valuable suggestions and always keeping his door open whenever I need help and advice. It has been always fruitful and joyful to discuss with him. I would also like to warmly thank Prof. Cristell Manuex for her continuous support and help during my Ph.D study period. Also I thank my committee members Prof. Mikael Östling from KTH, Royal Institute of Technology and Dr. Alain Chantre from STMicroelectronics for being on my thesis committee.

I would like to thank my former and present lab members, especially the Compact Modeling team members: Brice Grandchamp, Mario Weisz, Montassar Najari, Jad Bazzi, Gilles Kone, Si-Yu Liao, Arkaprava Bhattacharyya, Sudip Ghosh, Amit Kumar and my former colleague Johnny Goguet. Their help and support during these years was invaluable. It was fun to be part of this research group and share the time in the laboratory.

Finally, I wish to thank my mother, my brothers and sisters for their love, for supporting my decision to pursue my PhD far from them, and for providing invaluable assistance and encouragement. I am forever indebted to my family.

Table of Contents

Abstract

The Impact of strain engineering technology applied on NPN-Si-BJT/NPN-SiGe-HBT devices on the electrical properties and frequency response has been investigated. Strain engineering technology can be used as an additional degree of freedom to enhance the carriers transport properties due to band structure changes and mobility enhancement. The mobility of charge carriers in bipolar devices can be enhanced by creating mechanical tensile strain in the direction of electrons flow to improve electron's mobility, and by creating mechanical compressive strain in the direction of holes flow to improve hole's mobility. Consequently, new devices concepts and novel device architectures that are based on strain engineering technology have been explored using TCAD modeling. The physical models and parameters used in TCAD simulations have been calibrated in collaboration with Bundeswehr University-Munich using Monte Carlo simulation. Specific models for SiGe bandgap, bandgap-narrowing, effective mass, energy relaxation, mobility for hydrodynamic and drift-diffusion simulations have been calculated and implemented in the house simulator using tabulated models compiled in C code. Two approaches have been used in this study to generate the proper mechanical strain inside the device. The first approach was through introducing strain engineering technology principle at the device's base region using SiGe extrinsic stress layer. The second approach was through introducing strain engineering technology principle at the device's collector region using strain layers. The study examined not only the transistor DC performance but also the RF performance through multiple optimizations for the explored vertical transistors. Simulation results showed that the strained silicon BJT/HBT devices exhibited better DC performance and high frequency characteristics in comparison with equivalent standard conventional BJT/HBT devices. An approximately 42% improvement in f_T and 13% improvement in f_{MAX} have been achieved for BJT device employing strain at the base region. As well, an enhancement of the collector current by nearly three times in strained silicon BJT device has been attained. The obtained results for applying the same

technique on NPN-SiGe-HBT device have shown that applying strain on the base region of the HBT device is less efficient in comparison with the BJT device, as the SiGe base is already stressed due to the existence of Ge at the base. Moreover, utilizing a strain layer at the device's collector region will result in an approximately 9%-14% improvement in f_T and 7%-12% improvement in f_{MAX} in comparison with an equivalent standard conventional NPN-SiGe-HBT device. Despite of the very small decrease in the breakdown voltage BV_{CE0} value (1%─4%), the $f_T \times B_{VCE0}$ product enhancement is about 12% by means of strain engineering at the collector region.

Résumé

Dans le but d'améliorer les transistors bipolaires TBH SiGe, nous proposons d'étudier l'impact de la contrainte mécanique sur leurs performances. En effet, cette contrainte permet de libérer un degré de liberté supplémentaire pour améliorer les propriétés du transport grâce à un changement de la structure de bande d'énergie du semiconducteur. Ainsi, nous avons proposé de nouvelles architectures de composants basées sur l'ingénierie de la contrainte mécanique dans les semiconducteurs. Deux approches ont été utilisées dans cette étude pour générer la tension mécanique adéquate à l'intérieur du dispositif. La première approche consiste à appliquer une contrainte mécanique sur la base du transistor en utilisant une couche de SiGe extrinsèque. La seconde approche vise à appliquer une contrainte dans la région du collecteur en utilisant une couche contrainte. Les résultats obtenus montrent que cette méthode peut être une approche prometteuse pour améliorer les performances des TBH.

CHAPTER 1

Introduction & Background

1. Motivation

For the last two decades BiCMOS technology has been the dominant technology for analog RF high speed applications, due to the high linearity, high current drive capability, the lower 1/f noise, high output resistance, and voltage capability of the SiGe HBT. HBT has been a key device in this branch of semiconductor industry. However, due to the continuous demand for devices having higher frequency response, it becomes imperative to develop new bipolar architectures suitable for high frequency and power applications. Various techniques and efforts have been proposed to improve the performance of bipolar devices: introduction of a grading germanium into the base [1], introduction of carbon to improve 1D doping profile [2] and reduction of emitter width [3]. An additional way to improve the device performance is to enhance the carrier transport by changing the material properties. In advanced semiconductor devices strain can be used as an additional degree of freedom to enhance transport properties due to band structure changes [4]. Mobility of charge carriers in a bipolar device can be increased by creating mechanical tensile strain in the device in the direction of electrons flow to improve electron's mobility, and by creating mechanical compressive strain in the device in the direction of holes flow to improve hole's mobility [5].

The main motivation of this thesis is to clarify the impact of strain engineering technology on NPN-SiGe-HBT device's electrical properties and frequency response, through exploring new device concepts and proposing new device's architectures that are based on strain engineering

technology, and investigating the feasibility of such concepts in an industrial environment. These new concepts have been studied using finite element simulation TCAD tools.

The work methodology consists of the following steps:

- Virtually fabricate the device using process simulations.
- Study the sensitivity of the device's different zones to strain.
- Propose new methods to incorporate strain in the process and evaluate the strain level that can be obtained inside the device.
- Define simulation parameters and physical models (the model parameters have been calibrated in collaboration with Universität der Bundeswehr München).
- Perform numerical (device) simulations to analyze the device electrical performance.

2. Thesis Organization

The thesis is organized as follows:

Chapter 1: Introduction & Background; provides a brief discussion of the SiGe HBT technology, theoretical background, state-of-the-art and the motivation for carrying out this research.

Chapter 2: Strain Technology; gives an overview of the elasticity theory of solids, physics behind strain, different strain types, and application techniques of strain.

Chapter 3: TCAD Simulation & Modeling; provides an overview of the Technology Computer Aided Design (TCAD), process simulation, device simulation and the physical models used in this study.

Chapter 4; TCAD Simulation Results; gives a detailed discussion of the approaches used to create the desired mechanical strain inside the device, the devices architectures and the obtained results.

3. Principles of SiGe-HBT

The fundamental advantage of SiGe HBT over conventional Si BJT arises from the bandgap engineering technology principle. Bandgap engineering technology can be used to improve several key figure-of-merits for a bipolar transistor. The first step in understanding how a heterostructure device operates is to consider the energy band diagram. The energy band diagram for a forward biased ideal graded-base SiGe HBT with constant doping in the emitter, the base and the collector regions, and a comparable Si BJT are shown in Fig.1. The Ge content is linearly graded from 0% near the metallurgical emitter–base (EB) junction to some maximum value of Ge content near the metallurgical collector–base (CB) junction, and then rapidly ramped back down to 0% Ge [6].

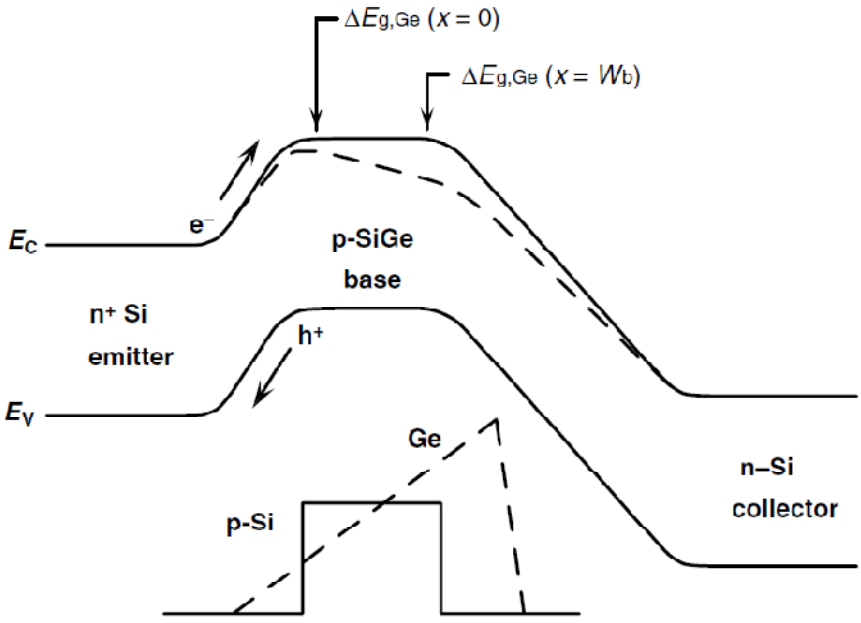

Fig.1**:** Energy band diagram for a graded-base SiGe HBT and a Si BJT.

The effect of introducing a graded Ge profile into the base region is apparent in the band structure changes shown above, where $\Delta E_{g,Ge}(x = 0)$ and $\Delta E_{g,Ge}(x = W_b)$ represent the germanium-induced bandgap narrowing at the emitter end of the base (x = 0) and the germanium-induced bandgap narrowing at the collector end of the base ($x = W_b$), respectively. As shown in the figure, a slight reduction in the base bandgap at the EB junction and a much larger reduction at the CB junction are observed with respect to the Si BJT device.

To illustrate the heterojunction action in a bipolar transistor, the operation of a conventional Si BJT will be considered first. When V_{BE} is applied to forward bias the EB junction, electrons are injected from the electron-rich emitter into the base across the EB potential barrier. The injected electrons diffuse across the base, and are swept into the electric field of the CB junction, yielding a useful collector current. At the same time, the applied forward bias on the EB junction produces a back-injection of holes from the base into the emitter. If the emitter region is heavily doped with respect to the base, the density of back-injected holes will be small compared to the forward-injected electron density, and hence a finite current gain results.

The introduction of Ge into the base region will affect the *dc* and *ac* operations of the Si BJT device. The reduction of the potential barrier to injection of electrons from emitter into the base will yield exponentially more electron injection for the same applied V_{BE}, translating into higher collector current, and hence higher current gain. In addition to that, the presence of a finite Ge content at the CB junction will positively influence the output conductance of the transistor, yielding higher Early voltage. For *ac* operation, the Ge-gradient-induced drift field across the neutral base will accelerate the injected minority electrons across the base, and thereby decreasing the base transit time. In what follows, the underlying physics of the NPN-SiGe-HBT, with particular emphasis on the fundamental differences between the operations of the SiGe HBT and the Si BJT will be introduced. For analysis purposes, a SiGe HBT and a Si BJT are taken to be of identical geometry, with identical emitter, base and collector-doping profiles, apart from the Ge in the base of the SiGe HBT. The Ge profile is assumed to be linearly graded from the EB to the CB junction as depicted in Fig.1. This analysis assumes standard low injection conditions, negligible bulk and surface recombination and Boltzmann statistics.

3.1 Current Gain

Due to the introduction of Ge into the base region, the potential energy barrier in the conduction band at the EB junction is lowered allowing more electrons to be injected into the base and thereby leads to an increase in the collector current. Furthermore, hole back injection is also reduced by the large valence band discontinuity reducing the base current. Overall, this increase in the collector current dramatically improves the current gain. For a constant base doping profile, the collector current density (J_C) can be derived from the generalized Moll-Ross relation as

$$J_C = \frac{q(e^{qV_{BE}/kT} - 1)}{\int_0^{W_b} \frac{P_b(x)dx}{D_{nb}(x)n_{ib}^2(x)}} \quad 1.1$$

where W_b is the neutral base width for the applied bias V_{BE}, $P_b(x)$ is the base doping, D_{nb} is the minority electron diffusivity in the base and n_{ib} is the intrinsic carrier density.

For SiGe HBT the intrinsic carrier density can be written as

$$n_{ib}^2(x) = \gamma\, n_{i0}^2\, e^{\Delta E_{gb}^{app}/kT}\, e^{[\Delta E_{g,Ge}(W_b) - \Delta E_{g,Ge}(0)]x/(W_b kT)}\, e^{\Delta E_{g,Ge}(0)/kT} \quad 1.2$$

Where ΔE_{gb}^{app} is the heavy-doping induced apparent bandgap narrowing in the base region, $\Delta E_{g,Ge}(0)$ is the Ge-induced band offset at x=0, $\Delta E_{g,Ge}(W_b)$ is the Ge-induced band offset at x=W_b, n_{io} is the low-doping intrinsic carrier density for Si, and γ is the effective density-of-states ratio between SiGe and Si

$$\gamma = \frac{(N_C N_V)_{SiGe}}{(N_C N_V)_{Si}} < 1 \quad 1.3$$

where $(N_C N_V)_{SiGe}$ represents the position-dependent Ge-induced changes associated with both the conduction and valence band effective density-of-states. The collector current density for SiGe HBT can be written as

$$J_{C,SiGe} = \frac{qD_{nb}}{N_{ab}^- W_b}\left(e^{qV_{BE}/kT} - 1\right) n_{i0}^2\, e^{\Delta E_{gb}^{app}/kT} \left[\frac{\tilde{\gamma}\tilde{\eta}\Delta E_{g,Ge}(grade)/kT e^{\Delta E_{g,Ge}(0)/kT}}{1 - e^{-\Delta E_{g,Ge}(grade)/kT}}\right] \quad 1.4$$

where, $\tilde{\eta}$ is the minority electron diffusivity ratio between SiGe and Si, and is given by

$$\tilde{\eta} = \frac{\left(\tilde{D}_{nb}\right)_{SiGe}}{(D_{nb})_{Si}} \quad 1.5$$

The first term in Equation 1.4 corresponds to the Si BJT ,and the second term represents the modification of J_C due to the introduction of Ge into the base region. As shown in the equation, J_C in a SiGe HBT depends exponentially on the EB boundary value of the Ge-induced band offset, and is linearly proportional to the Ge-induced bandgap grading factor. Consequently, The SiGe HBT exhibits higher collector current and approximately the same base current as the Si BJT. This increase in J_C for the SiGe HBT will lead to an increase in current gain. For

identically constructed devices, the ratio of the current gain (β) between a linearly graded germanium profile SiGe HBT and a Si BJT is given by

$$\frac{\beta_{SiGe}}{\beta_{Si}} \cong \frac{J_{C,SiGe}}{J_{C,Si}} = \frac{\tilde{\gamma}\tilde{\eta}\,\Delta E_{g,Ge}(grade)/kT e^{\Delta E_{g,Ge}(0)/kT}}{1 - e^{-\Delta E_{g,Ge}(grade)/kT}} \quad 1.6$$

Equation 1.6 represents a ratio which is larger than unity for a finite Ge content, and indicates that the smaller base bandgap in a SiGe HBT exponentially increases the number of minority carriers injected into the base, causing an increase in the collector current for the same forward-bias, and hence higher current gain value.

3.2 Early Voltage

The most important physical effect which causes the collector current to increase with the collector-emitter voltage for a constant base current is the increase of collector current caused by a decreased neutral base width with base-collector reverse bias (V_{CB}). The output conductance is a measure of collector current variation with V_{CB}. Therefore, low output conductance is desirable to achieve invariant output current in low-frequency analogue applications. The Early voltage (V_A) is a measure of the change in output conductance with changing V_{CB}. For analogue circuit applications, a high value of the product of current gain and Early voltage is desirable, which is conventionally defined as the figure-of-merit for analog circuit design.

An additional benefit of using a graded-Ge profile in a SiGe HBT is the enhancement of the Early voltage. The enhancement of the Early voltage of a graded SiGe HBT can be expressed as

$$\frac{V_{\mathrm{A,SiGe}}}{V_{\mathrm{A,Si}}}\Big|_{\mathrm{V_{BE}}} \cong \exp\left(\frac{\Delta E_{g,Ge}(grade)}{\mathrm{kT}}\right)\left[\frac{1 - e^{-\Delta E_{g,Ge}(grade)/kT}}{\Delta E_{g,Ge}(grade)/kT}\right] \quad 1.7$$

This ratio is larger than unity, and it indicates that for a finite grading across the quasi-neutral base, the exponential dominates the functional dependence on Ge content, yielding an improvement for a SiGe HBT compared to a Si BJT.

3.3 Transit Time

As a result of the quasi-electric field due to the grading of the Ge content in the base, the electrons injected from the emitter to the collector will be accelerated; thereby effectively reducing the base transit time. For bipolar transistors, the base transit time for a constant base doping is given by

$$\tau_b = \int_0^{W_b} \frac{n_{ib}^2(x)}{P_b(x)} \left[\int_0^{W_b} \frac{P_b(y)dy}{D_{nb}(y) n_{ib}^2(y)} \right] dx \qquad 1.8$$

By substituting equation 1.2 into equation 1.8 and integrating, the SiGe HBT transit time becomes

$$\tau_{b,SiGe} = \frac{W_b^2}{\widetilde{D}_{nb}} \frac{kT}{\Delta E_{g,Ge}(grade)} \left(1 - \frac{kT}{\Delta E_{g,Ge}(grade)} \left[1 - e^{-\Delta E_{g,Ge}(grade)/kT} \right] \right) \qquad 1.9$$

For Si BJT, the base transit time is given by

$$\tau_{b,Si} = \frac{W_b^2}{2D_{nb}} \qquad 1.10$$

Hence, the enhancement of the base transit time due to the existence of a graded Ge profile into the base region is given by ratio of the base transit time for SiGe ($\tau_{b,SiGe}$), and Si BJT ($\tau_{b,Si}$) as follows

$$\frac{\tau_{b,SiGe}}{\tau_{b,Si}} = \frac{2}{\tilde{\eta}} \frac{kT}{\Delta E_{g,Ge}(grad)} \left(1 - \frac{kT}{\Delta E_{g,Ge}(grad)} \left[1 - e^{-\frac{\Delta E_{g,Ge}(grad)}{kT}} \right] \right) \qquad 1.11$$

For a finite Ge grading, the ratio ($\tau_{b,SiGe}/\tau_{b,Si}$) will be less than unity, and therefore the SiGe HBT base transit time will be shorter than Si BJT.

Moreover, the forward bias in EB junction creates a back-injection of holes from the base to the emitter, and gives rise to emitter charge storage delay time (τ_e). The emitter charging time will be reduced due to the lower barrier which carriers must overcome before being injected into the base due to the existence of the graded Ge profile at the base region.

3.4 High Frequency Figures-of-Merit

The SiGe base gives new degrees of freedom for the design of SiGe HBTs and allows much higher values of the figures of merit; the cut-off frequency (f_T) and the maximum oscillation frequency (f_{MAX}), to be achieved in comparison with the conventional silicon BJTs. The unity-gain cut-off frequency and the maximum oscillation frequency parameters are positively impacted by the Ge content in the base. As mentioned previously τ_b and τ_e decrease in SiGe HBT.

The unity-gain cutoff frequency (f_T) is given by the relation

$$f_T = \frac{1}{2\pi}\left[\frac{kT}{qI_C}(C_{EB}+C_{CB}) + \tau_b + \tau_e + \frac{W_{CB}}{2v_{sat}} + r_C C_{CB}\right]^{-1} \quad 1.12$$

Where I_c is the collector current, C_{EB} and C_{CB} are the EB and CB depletion capacitances, W_{CB} is the CB space-charge region width, v_{sat} is the saturation velocity, and r_c is the dynamic collector resistance. Therefore, improvements in τ_b and τ_e due to the presence of SiGe will directly translate into an enhanced f_T of the transistor at fixed bias current. The maximum oscillation frequency (f_{MAX}) figure of merit is defined as

$$f_{MAX} = \sqrt{\frac{f_T}{8\pi C_{CB} r_B}} \quad 1.13$$

Where r_B is the base resistance. Therefore, the increase in f_T and the reduction in the base resistance aids in improving the f_{MAX} of the SiGe HBT [7] [8].

4. State-of-the-Art

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 the 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, the progress made in the field of semiconductor technology have a constant increase in performance according to Moore' law which states that the number of transistors per chip doubles every 18-24

months. However, 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. The idea of varying the bandgap in a bipolar transistor structure to increase the emitter injection efficiency is almost as old as the bipolar junction transistor itself. Shockley described the idea in his application for a patent on the junction bipolar transistor [9][10]. The inherent performance advantages of heterojunction bipolar transistors (HBTs) over conventional bipolar junction transistors have been recognized, and Kroemer first explained the underlying principle of the heterojunctions in 1957 [11].

The work by Kroemer on the heterojunction opened a new door for achieving higher speeds than anticipated with silicon bipolar transistors. Although the performance advantages of HBTs over BJTs were well understood, no fabrication technologies were available to produce high quality heterojunctions until the 1970s, where epitaxial growth of SiGe heterostructures demonstrated by Erich Kasper using molecular beam epitaxy (MBE) [12]. With the advent of heteroepitaxy, the concept of strained layers has been extended to include other elemental semiconductors. These developments set the stage for IBM's development of SiGe HBTs in 1987 using MBE. The successful demonstrations of SiGe HBT technology, in both high-performance digital and analogue circuit applications, are the results of over 15 years of steady research progress from initial material preparations in 1984, through device demonstrations from 1987–1992 to large scale circuit fabrication in 1994 and commercial products in 1998. Since then, the interest in SiGe HBTs from both academia and industry rose progressively through the years causing the figures of merit to increase and reach current state-of-the-art levels. Today's state-of-the-art SiGe pre-production technologies have shown cut-off frequency close to 300 GHz, and maximum frequency of oscillation close to 400 GHz as reported recently by our partners of the European joint research project DOTFIVE [13] [14].

The rapid improvement in SiGe HBT performance is mainly due to careful tailoring of Ge and doping profiles, and also due to the continuing downscaling of device dimensions. Material growth development is also an important factor for the success of SiGe technology and the most recent example is the incorporation of carbon in the SiGe epitaxial layers, which further improves performance. The developments of SiGe HBTs technology are summarized in Table.1 [7].

Historical Event	Year	Ref.
First demonstration of transistor action	1947	[15]
Basic HBT concept	1948	[9]
First demonstration of a bipolar junction transistor	1951	[10]
First demonstration of a silicon bipolar transistor	1956	[16]
Drift-base HBT concept	1954	[17]
Fundamental HBT theory	1957	[18][11]
First epitaxial silicon transistors	1960	[19]
First SiGe HBT	1987	[20]
First ideal SiGe HBT grown by CVD	1989	[21]
First SiGe HBT grown by UHV/CVD	1989	[22]
First high-performance SiGe HBT	1990	[23][24]
First self-aligned SiGe HBT	1990	[25]
First SiGe HBT ECL ring oscillator	1990	[25]
First pnp SiGe HBT	1990	[26]
First operation of SiGe HBTs at cryogenic temperatures	1990	[27]
First SiGe HBT BiCMOS technology	1992	[28]
First LSI SiGe HBT integrated circuit	1993	[29]
First SiGe HBT with peak f_T above 100 GHz	1993	[30][31]
First SiGe HBT technology in 200-mm manufacturing	1994	[32]
First SiGe HBT technology optimized for 77K	1994	[33]
First radiation tolerance investigation of SiGe HBTs	1995	[34]
First report of low-frequency noise in SiGe HBTs	1995	[35]
First SiGe:C HBT	1996	[36]
First high-power SiGe HBTs	1996	[37][38]
First sub-10psec SiGe HBT ECL circuits	1997	[39]
First high-performance SiGe:C HBT technology	1999	[40]
First SiGe HBT with peak f_T above 200 GHz	2001	[41]
First SiGe HBT with peak f_T above 300 GHz	2002	[42]
First complementary (npn+pnp) SiGe HBT technology	2003	[43]
First C-SiGe technology with npn and pnp f_T above 100 GHz	2003	[44]
First vertical SiGe HBT on thin film (CMOS compatible) SOI	2003	[45]
First SiGe HBT with both f_T and f_{MAX} above 300 GHz	2004	[46]
First Conventional Double-Polysilicon FSA-SEG Si/SiGe:C HBT with f_{MAX} 423 GHz and f_T 273 GHz	2009	[14]
First Fully Self-Aligned SiGe:C HBT Architecture with f_{MAX} 400 GHz	2009	[13]
First SiGe HBT with f_{MAX} 500 GHz (IHP: to be published in IEDM 2010)	2010	

Table 1: Milestones in the development of SiGe HBTs [7].

4.1 Strained SiGe HBT

Due to the continuous demand for devices having higher frequency response, it becomes imperative to develop new bipolar architectures suitable for high frequency and power applications. Therefore, various efforts and approaches have been proposed to improve the bipolar devices performance. In advanced semiconductor devices technology, strain engineering technology is considered as an additional degree of freedom to enhance transport properties due to band structure changes. Recent experimental and theoretical work point out uniaxially and biaxially stressed Si and SiGe substrates as a possible technique for the new generation HBT devices. Experimentally, two major methods have been used; wafer bending [47], and virtual substrates [48]. Wafer bending allows inducing uniaxial stress while virtual substrates create biaxial stress. Both methods induce a uniform stress over the device. In what follows, a brief description of the latest work done on strained SiGe HBTs using wafer bending and virtual substrate approaches is presented.

Wang et al. investigated the impact of mechanical uniaxial stress on the characteristics of SiGe heterojunction bipolar transistors uing a four-point bending apparatus to apply a uniaxial stress in the range of −200 MPa to + 200 MPa. The SiGe HBTs used in their study were fabricated using 0.18 μm self-aligned SiGe BiCMOS technology, with an emitter area of 0.2x10.16 μm^2. Their results show that the performances of SiGe HBTs are varied with the stress level. The changes in the collector current, base current, current gain, and the breakdown voltage were found to be linearly dependent upon the mechanical uniaxial stress level, for the range of −200 MPa to + 200 MPa. The strain-polarity dependence of the collector current, base current, and current gain was positive under uniaxial compressive stress, whereas that of the breakdown voltage was negative [47].

Yuan et al. in their work reported the performance of Si–SiGe HBT under the biaxial compressive and tensile mechanical stress with the comparison of BJTs. An externally uniform mechanical displacement at the center with the diameter of 13 mm on 100 mm wafers for both SiGe HBT and Si BJT devices has been applied as shown in Fig.2. The average biaxial strain used in this study is 0.028%. The current gain variations of the mechanically strained Si–SiGe heterojunction bipolar transistor (HBT) and Si bipolar junction transistor (BJT) devices were investigated. The current gain change for HBT is found to be 4.2% and 7.8% under the biaxial compressive and tensile mechanical strain of 0.028%, respectively. The change for BJT is found

to be 4.9% and 5.0% under the biaxial compressive and tensile mechanical strain of 0.028%, respectively. Their results are shown in Fig.3 . Moreover, their results show that the current gain changes show a good linear dependence on external biaxial mechanical stress as shown in Fig.4 [49].

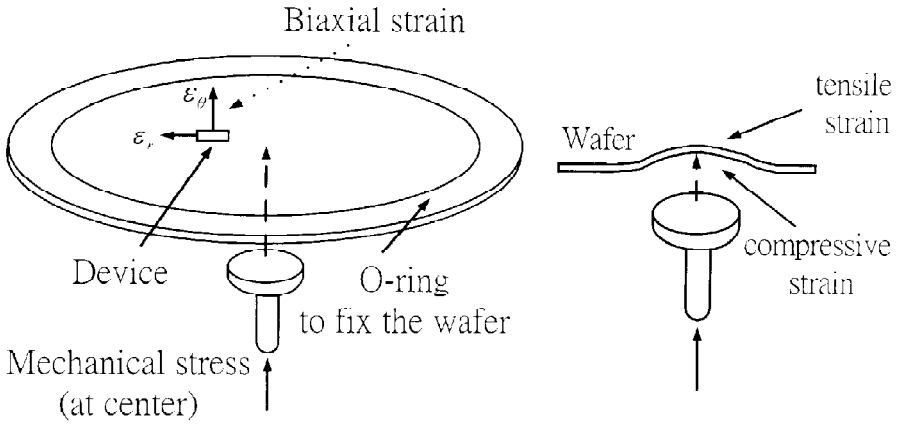

Fig.2 : Schematic diagram of the externally applied mechanical stress on the wafer.

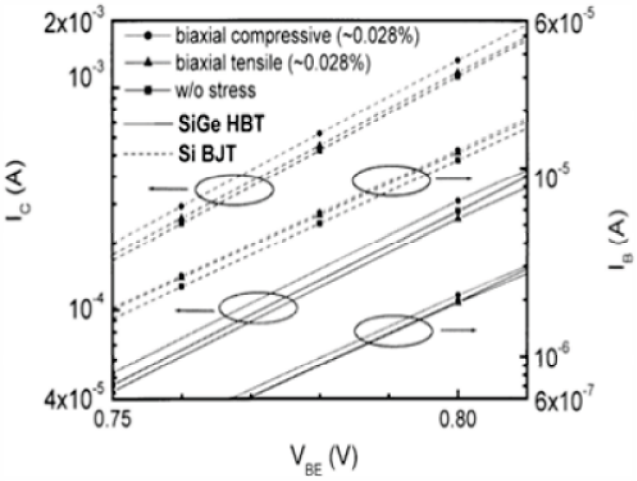

Fig.3: Gummel plot of Si BJT and SiGe HBT devices without and with mechanical stress (V_{BC}=0 V).

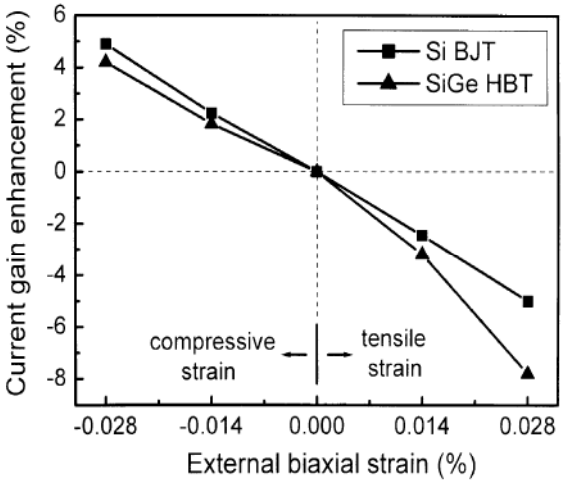

Fig.4: Current gain changes of SiGe HBT and Si BJT device as a function of stress level.

Persson et al. [48] reported in their work (Fabrication and characterization of strained Si heterojunction bipolar transistors on virtual substrates), a strained Si HBT with a maximum current gain of 3700 using a relaxed $Si_{0.85}Ge_{0.15}$ virtual substrate, $Si_{0.7}Ge_{0.3}$ base and strained Si emitter. The schematic of the complete structure used in their study is shown in Fig.5. Their

results demonstrate major improvements in current gain compared with co-processed pseudomorphic SiGe HBTs and Si BJTs as shown in Fig.6. In addition, strained Si HBTs exhibit larger collector current than SiGe HBT and Si BJTs as shown in Fig.7.

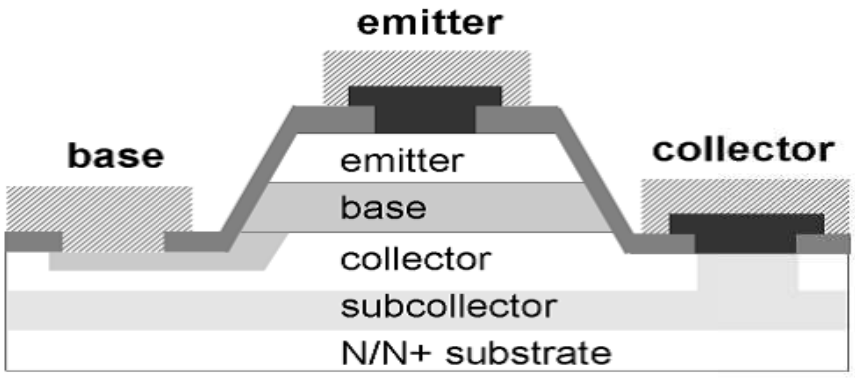

Fig.5: A schematic of the complete strained Si HBT structure used by Persson et al.

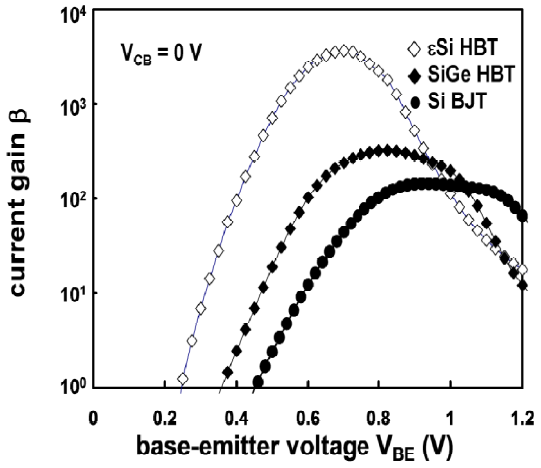

Fig.6: Collector current (I_C) vs. collector-emitter voltage (V_{CE}) characteristics for strained Si HBT, SiGe HBT, and Si BJT at I_B = 3 μA.

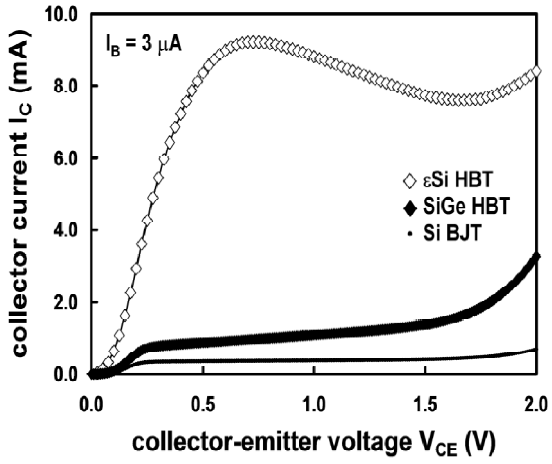

Fig.7: Current gain β vs. base-emitter voltage V_{BE}. β is increased by almost one order of magnitude in the strained Si HBT.

Haugerud et al. studied the effects of mechanical planar biaxial tensile strain applied, post-fabrication, to Si/SiGe HBT BiCMOS technology. Planar biaxial tensile strain was applied to the samples, which included standard Si CMOS, SiGe HBTs, and an epitaxial-base Si BJT control. Their results show that at a strain level of 0.123%, the Si BJT/SiGe HBTs showed a consistent decrease in collector current and hence current gain after strain as illustrated in Fig.8

and Fig.9. This decrease in the collector current is attributed to the compressive strain in the orthogonal plane which degrades the electron transport [50].

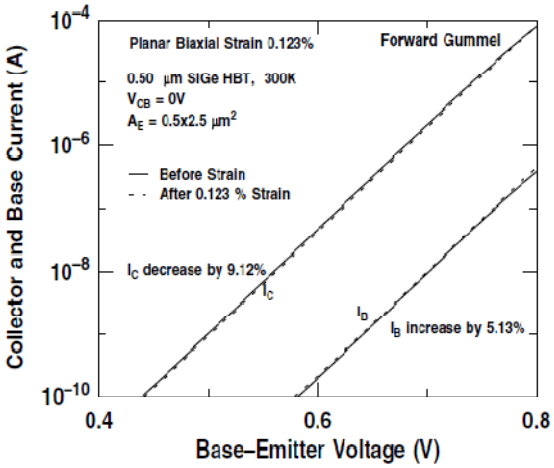

Fig.8: Forward Gummel characteristics of a first generation SiGe HBT for both pre-strain and post 0.123% biaxial strain.

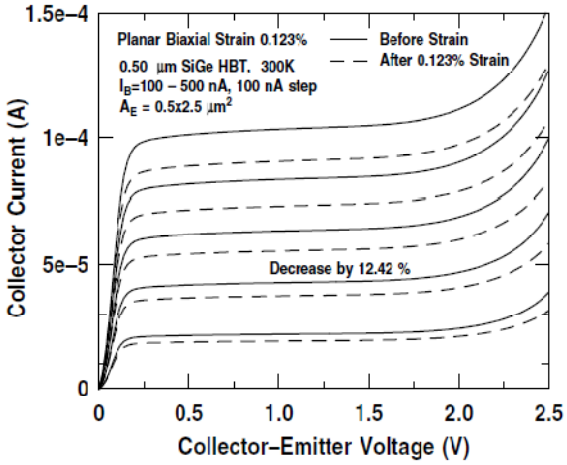

Fig.9: Output characteristics of a first generation SiGe HBT for both pre-strain and post 0.123% biaxial strain

Moreover, two patents have been proposed to introduce stress into bipolar transistors: The first patent by Chidambarrao et al. proposes to create stress at the base region of the device as shown in Fig.10. In their structure, the compressive and tensile strains are created by forming a stress layer in close proximity to the intrinsic base of the device resulting in an enhancement of the mobility of the charge carriers [51]. The second patent by Dunn et al. proposes a method of forming a semiconductor device having two different strains inside the device [52]. This proposal is more complicated since the stress is applied on the emitter, the base and the collector regions as shown in Fig.11.

To the best of our knowledge, there haven't been any simulation or measurement results available in the literature to evaluate the effect of the stress layers in the above structures, thus we have analyzed the impact of strain engineering on SiGe-HBTs via modified ideas from the above patents using a specific structure provided by IMEC Microelectronic-Belgium as a reference device [53].

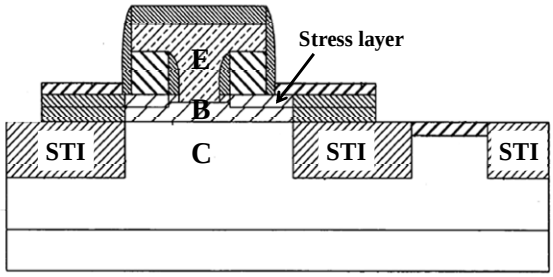

Fig.10: A cross- sectional view of a complete BJT device formed according to Chidambarrao et al.

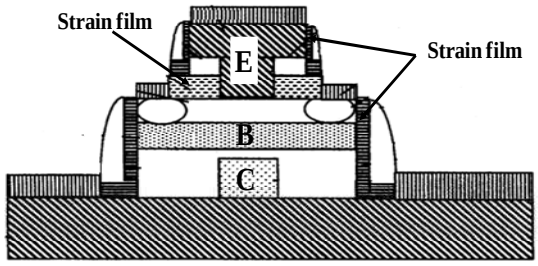

Fig.11: A cross- sectional view of a complete BJT device formed according to Dunn et al.

Regarding the theoretical work done in this field, Jankovic et al. investigated the influence of strained-Si cap layers on n–p–n heterojunction bipolar transistors fabricated on virtual substrates as shown in Fig.12. Using an approximate theoretical model, they found that the presence of a strained-Si/SiGe (relaxed) heterojunction barrier in the emitter can substantially improve the HBT's current gain as shown in Fig.13. Furthermore, two-dimensional numerical simulations of a virtual substrate HBT with a realistic geometry demonstrate that, besides the current gain enhancement, a three times improvement in f_T and f_{MAX} were realized when a strained-Si/SiGe emitter is incorporated as shown in Fig.14 and Fig.15.

Moreover, Jankovic et al. presented a computational study by commercial TCAD of the potential electrical and thermal properties of n–p–n HBTs fabricated on relaxed $Si_{1-y}Ge_y$ virtual substrates. The dependences of dc, ac and self-heating characteristics of virtual substrate HBTs (VS HBTs) on alloy composition were investigated in details. It is found that symmetrical VS HBTs generally exhibit higher current drive capabilities compared with equivalent HBTs formed pseudomorphically on Si substrates, but at the expense of a lower f_T and a decreased f_{MAX}. In addition, simulated results show that self-heating effects become increasingly significant for VS HBTs, substantially degrading device electrical parameters such as the early voltage [54] [55].

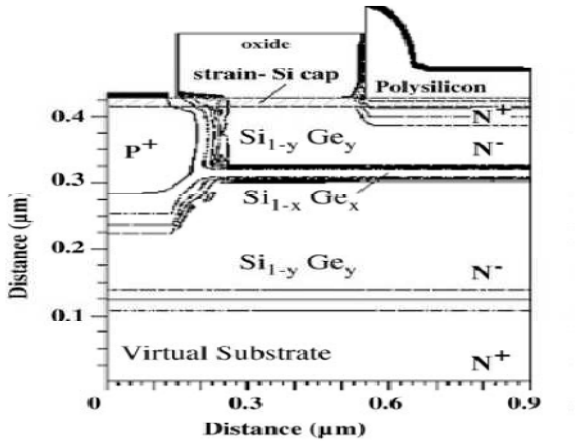

Fig.12: 2D cross-section of the simulated virtual substrate HBT device with strained-Si/SiGe heterointerface emitter

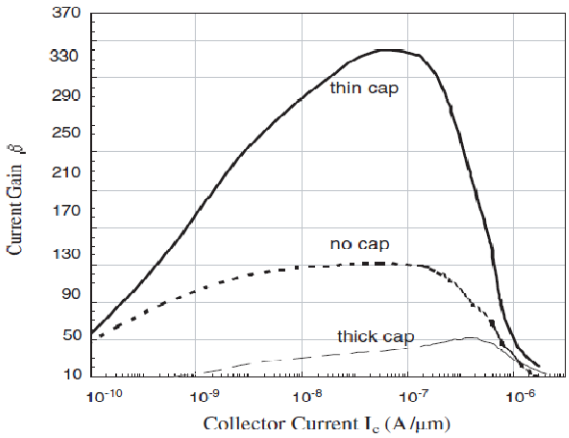

Fig.13: the β(Ic) characteristics of virtual substrate HBTs with thin (10 nm), thick (50 nm) and without strained-Si layers in the emitter

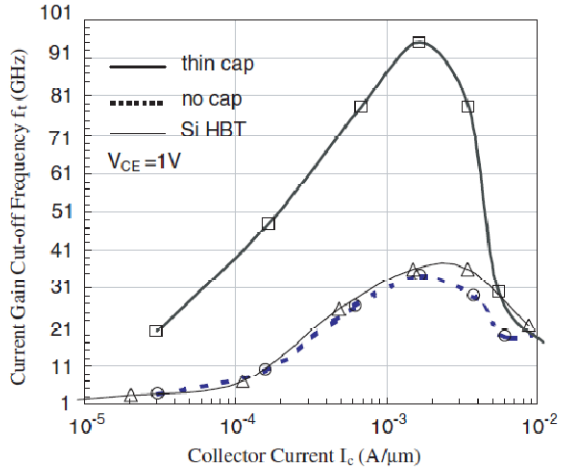

Fig.14: f_T versus collector current Ic extracted for the n-p-n HBTs with and without strained-Si cap layer and for conventional silicon-based HBT.

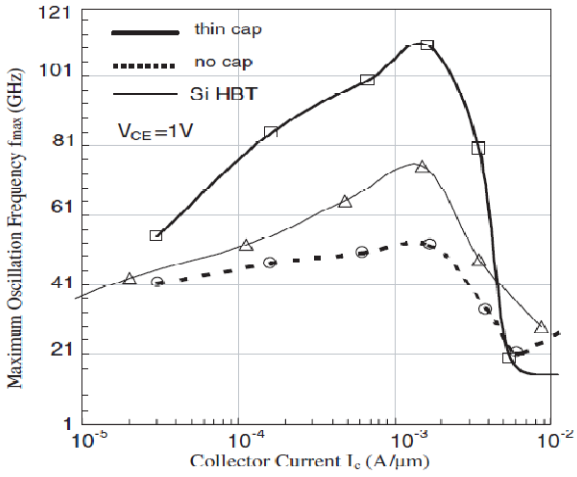

Fig.15: f_{MAX} versus collector current Ic extracted for the n-p-n HBTs with and without strained-Si cap layer and for conventional silicon-based HBT.

Simulation results reported by Jankovic et al. show optimistic improvements of the HBT deice performance by means of strain engineering technology. As they are using an approximate theoretical model in their simulations, this approximate theoretical model might need to be calibrated and/or taking more effects in consideration to achieve more precise results.

4.2 Current State-of-the-Art

As mentioned previously, the European joint research project DOTFIVE partners have achieved SiGe:C HBTs with maximum frequency of oscillation (f_{MAX}) 400 GHz which are today's state-of-the-art SiGe:C HBTs. These partners achieve this milestone through completely different architectures. The first partners (IMEC & IHP Microelectronics) approache 400 GHz f_{MAX} through fully self-aligned (FSA) SiGe:C architecture. While the second partner (ST Microelectronics) approaches 400 GHz f_{MAX} through a conventional double-polysilicon FSA selective epitaxial growth (SEG) Si/SiGe:C HBT. In what follows a brief description of each approach is given.

IMEC and IHP have developed two novel device architectures for half terahertz RF performance to reduce further the device parasitic elements compared to the reference quasi self-aligned (QSA) architecture (f_T = 205 GHz, and f_{MAX}= 275 GHz). The first architecture named G1G architecture is the novel FSA IMEC architecture. The second architecture is the novel FSA IHP architecture. The novel device architecture that IMEC explores, attempts to reduce the device parasitics significantly. The key element to achieve this, is that the emitter, base and collector regions are self-aligned to each other. This is accomplished by growing the collector/base and capping layer non-selectively, then etching the extrinsic device region away using a sacrificial emitter, and then using the sacrificial emitter to self-align a reconstructed external base to the emitter region. Initial results with this approach yielded marginal devices with poor performance. But the concept to fabricate such a device was demonstrated, and as the reasons for the poor performance and process marginalities were identified, IMEC continued to optimize this architecture in the frame of the DOTFIVE project. A schematic cross-section of the novel FSA IMEC device architecture is shown in Fig.16.

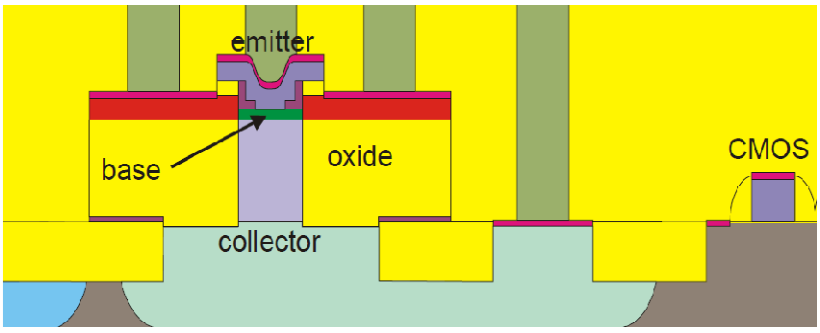

Fig.16: Schematic cross-section of novel device architecture of IMEC.

IMEC started from an initial process flow, which yielded marginally functional devices with f_{MAX} below 200 GHz. This initial process flow has been completely reviewed, and the unit process steps have been significantly improved. A SEM cross-section showing the device architecture after the unit process step improvements, compared to the original structure is shown in Fig.17. The major deficiencies of the original structure have been alleviated, such as the marginal thickness of the external base connection, the marginal overlap of the polyemitter over the L-shaped spacer, the marginal L-shaped spacer formation, and the sloped profile of the pedestal etch. The resulting device has an effective emitter width of 80nm.

This fully self-aligned SiGe:C HBT architecture featuring a single step epitaxial collector/base process, removal of the extrinsic part of the device using sacrificial emitter, external base reconstruction, L-shaped spacer formation after removal of the sacrificial emitter, and in-situ As doped polyemitter demonstrated 400 GHz f_{MAX}. In addition, introducing carbon in the collector suppressed segregation of P in the collector to the Si/SiGe interface, which resulted in a strong increase of the f_{MAX} because of the reduction of C_{BC}. The resulting f_T and f_{MAX} curves are shown in Fig.18, and the base-collector reverse diode current is shown in Fig.19. The influence of the addition of 0.2% carbon to the undoped collector part is demonstrated. This improvement is the accumulated result of a further lateral scaling of the HBT device, a significant decrease of the base resistance and of the base-collector capacitance compared with the reference HBT device.

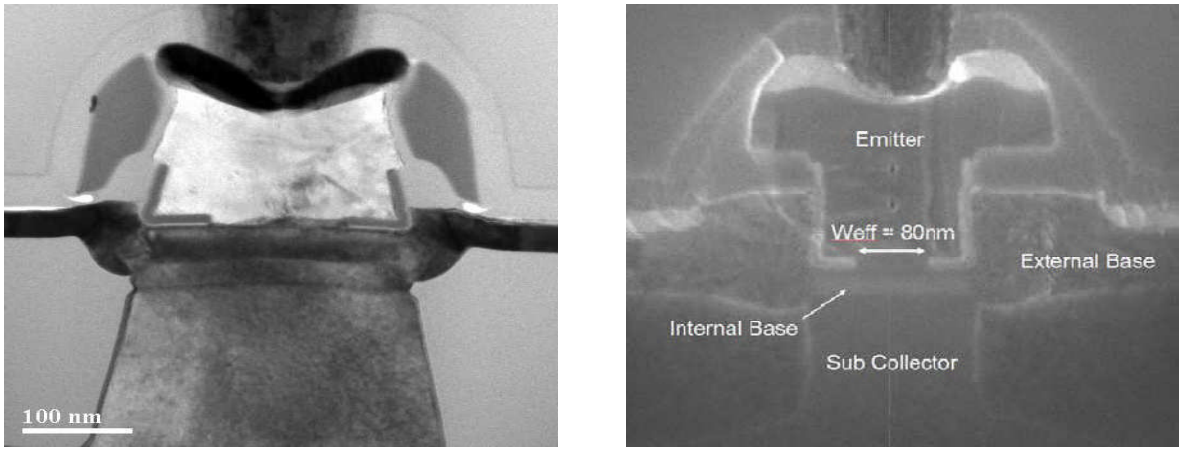

Fig.17:Cross-section SEM picture of the original (left) and improved (right) emitter/base structure.

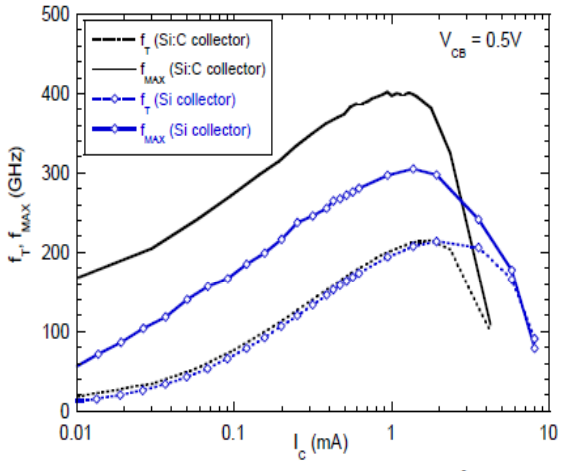

Fig.18: Base-collector reverse diode current for a 0.15x1.0μm^2 HBT device

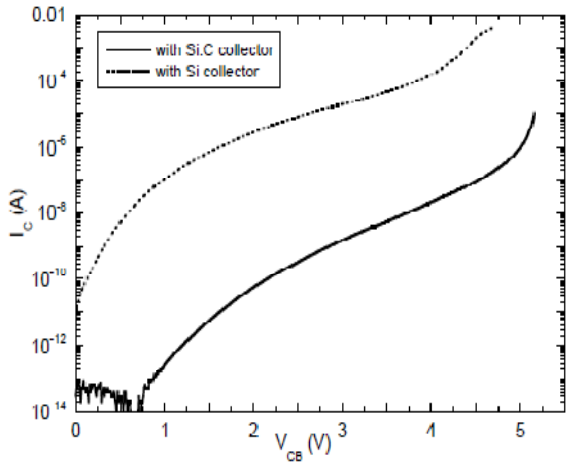

Fig.19 :f_T and f_{MAX} versus Ic for a 0.15x1.0μm^2 HBT.

Regarding the IHP novel architecture, a new collector construction for high-speed SiGe:C HBTs that substantially reduces the parasitic base-collector capacitance by selectively under etching the collector region is presented. A schematic cross-section of the novel IHP device architecture is shown in Fig.20. The IHP novel architecture provides f_T values that are higher compared to the G1G architecture and presents less variation with decreasing emitter width. The C_{BE} and R_E values of the IHP architecture are lower than the G1G architecture helping to obtain higher f_T values. The f_{MAX} values for both architectures show a similar increase with decreasing the emitter width. The IHP architecture however provides slightly lower f_{MAX} values at wider emitter widths while the situation reverses for the smallest simulated dimensions [13] [56].

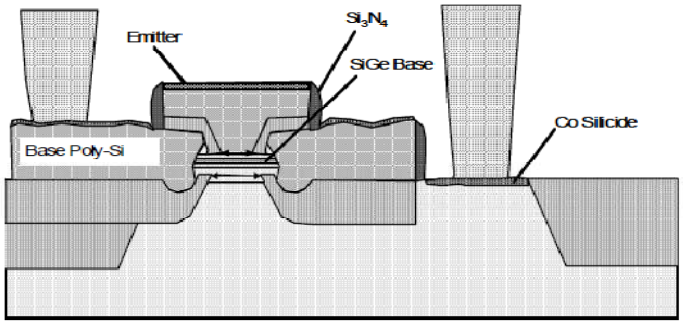

Fig.20: Schematic cross-section of the IHP novel device architecture.

On the other hand, ST Microelectronics approaches 400 GHz f_{MAX} through a conventional double-polysilicon FSA selective epitaxial growth (SEG) Si/SiGe:C HBT. Starting from the high speed SiGe BiCMOS technology BiCMOS9MW (Fig.21) which features a SA selective epitaxial SiGe HBT with 230 GHz / 290 GHz f_T / f_{MAX} , two shrinking phases (B3T and B4T) have been performed by STMicroelectronics. The path followed to move from a 300 GHz f_{MAX} HBT (BiCMOS9MW) to a 400 GHz f_{MAX} HBT (B4T) is shown in Fig.22.

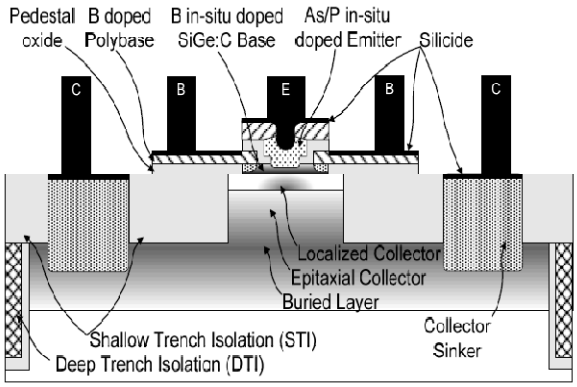

Fig.21: Sketch of the FSA-SEG SiGe HBT architecture

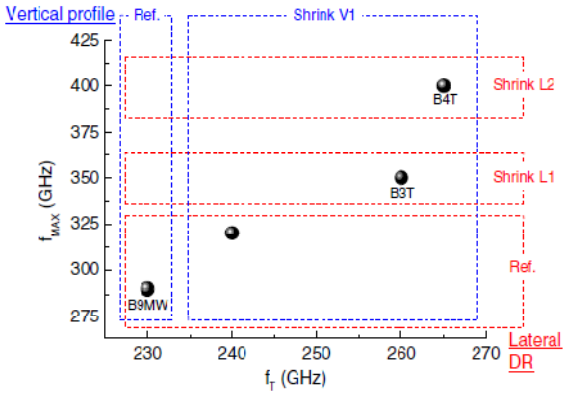

Fig.22 : From BiCMOS9MW to B4T: Splits of the vertical and lateral scaling contributions.

The vertical shrink V1 corresponds to a slight reduction of the spike annealing temperature, a different base profile and a reduction of the collector doping. Lateral shrink L1 corresponds to a reduction of the collector area, the polyemitter and emitter inside spacer widths, the final emitter width being unchanged (0.13µm). The result of this first shrinking phase is a technology called B3T, featuring f_T = 260 GHz and f_{MAX} = 350 GHz. Using still a conventional SA selective epitaxial base HBT, a second shrinking phase resulted in the B4T technology providing a maximum oscillation frequency of 400 GHz together with a transit frequency of 265 GHz (wafer averages). These outstanding performance data have been obtained for a collector base breakdown voltage of 6.0 V and a collector emitter breakdown voltage of 1.5 V (W_E is reduced to 0.11 µm for the lateral shrink L2). The f_T and f_{MAX} characteristics versus collector current of a B4T transistor, compared to those of B3T and BiCMOS9MW HBTs having the same drawn emitter window length are shown in Fig.23. Moreover the main electrical parameters of these devices are summarized in Table 2 [14] [56].

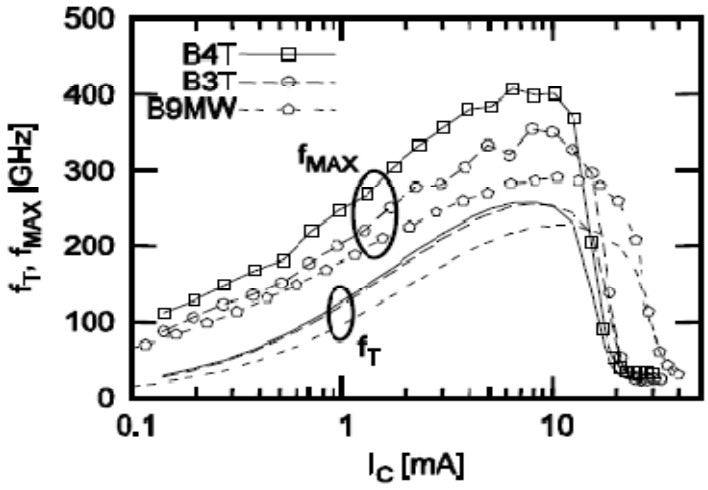

Fig.23: f_T & f_{MAX} vs. I_C for BiCMOS9MW, B3T and B4T technologies at V_{CB} = 0.5 V (L_E ~ 5 μ m).

Parameter	Measurements conditions	BiCMOS9MW	B3T	B4T	Unit
f_T	V_{CB}=0.5 V	230	260	265	GHz
f_{MAX}	V_{CB}=0.5 V	290	350	400	GHz
W_E	TEM	0.13	0.13	0.11	µm
J_C	Peak f_T	15.0	11.8	13.0	mA/µm^2
β	V_{BE}=0.75V	950	1595	1750	-
V_{AF}		>200	>200	>200	V
BV_{EBO}		2.2	2.1	2.1	V
BV_{CEO}	V_{BE}=0.7V	1.5	1.55	1.50	V
BV_{CBO}		5.5	6.0	6.0	V
R_{Bi}		2.7	2.5	2.5	kΩ/sq

Table 2: HBT parameters comparison (wafer averages) FOR BICMOS9MW, B3T and B4T technologies (L_E~5 µm).

However, IHP Microelectronics reached a SiGe HBT device with f_T/f_{MAX} of 300 GHz/500 GHz, their results will be published in IEDM 2010 proceedings. The speed-improvement compared to previous SiGe HBT technologies originates from the reduced specific collector-base capacitance and base resistance and scaling of the device dimensions.

5. DOTFIVE Project

THz technology is an emerging field which has demonstrated a wide-ranging potential. Extensive research in the last years has identified many attractive application areas and has paved the technological path towards broadly usable THz systems. THz technology is currently in a pivotal phase and will soon be in a position to radically expand our analytic capabilities via its intrinsic benefits. In this context, DOTFIVE is planned to establish the basis for fully integrated cost efficient electronic THz solutions. An illustration of some exemplary applications of Terahertz radiation is shown in Fig.24.

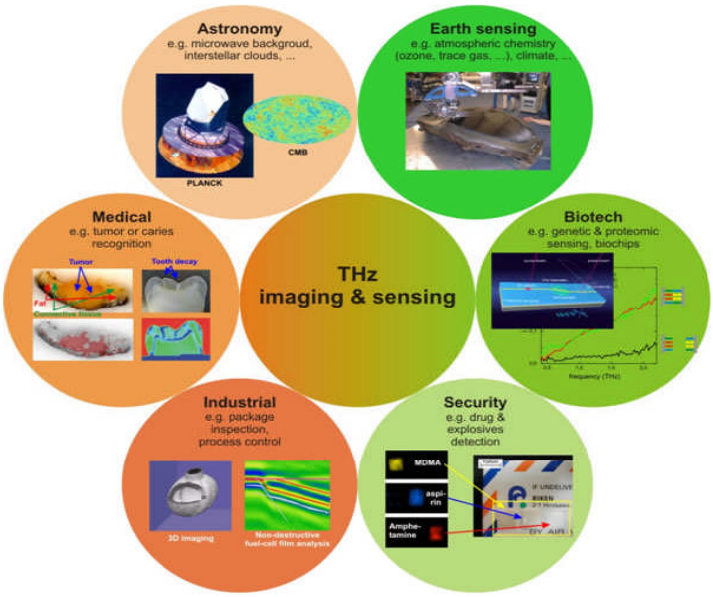

Fig.24: Illustration of some exemplary applications of Terahertz radiation.

DOTFIVE is an ambitious three-year European project supported by the European Commission through the Seventh Framework Program for Research and Technological Development, focused on advanced Research, Technology, and Development activities necessary to move the SiGe-HBT into the operating frequency range of 0.5 THz (500 GHz). This high frequency performance is currently only possible with more expensive technology based on III-V semiconductors, making high integration and functionality for large volume consumer applications difficult. The new transistors developed by DOTFIVE will be used for designing circuits enabling power efficient millimeter-wave applications such as automotive radar (77 GHz)

or WLAN communications systems (60 GHz –Wireless Local Area Network). In addition to these already evolving markets, DOTFIVE technology sets out to be a key enabler for silicon based millimeter-wave circuits with applications in the security, medical and scientific areas. A higher operating speed can open up new application areas at very high frequencies, or can be traded for lower power dissipation, or can help to reduce the impact of process, voltage and temperature. The project involves 15 partners from industry and academia in five countries teaming up for research and development work on silicon-based transistor architectures, device modeling, and circuit design. The scientific aspects of the DOTFIVE project are tackled by five work packages during a period of 36 months as illustrated in the schematic shown in Fig.25.

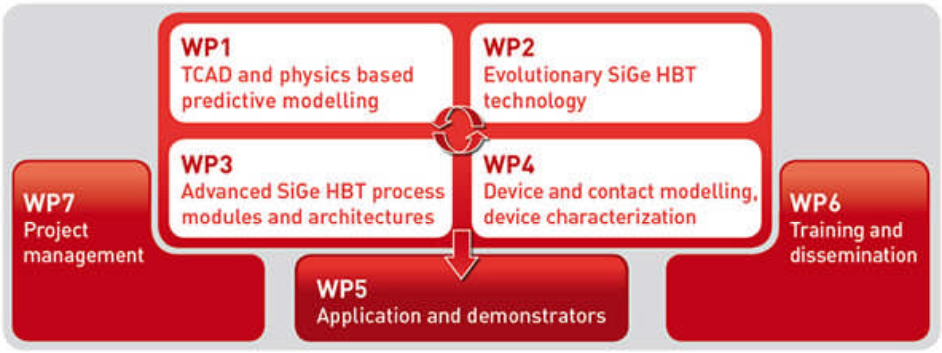

Fig.25: DOTFIVE project work packages

The work of this thesis is a part of the work package1 (WP1) which is dedicated to "physics-based predictive modeling" using Technology Computer Aided Design (TCAD) tools, that allow the simulation of processing steps and electrical characteristics of devices. Due to the complexity of transport phenomena in nano-scale transistors, advanced device simulation tools (e.g., solution of Boltzmann transport, Schrödinger-Possion solver) from DOTFIVE partners are used. Based on such advanced TCAD platform, it will be possible to achieve a deep understanding of the electrical behavior of near-terahertz devices and to develop guidelines for doping and architecture optimization. In particular, WP1 will support continuously the technology development in WP2 and WP3 by, e.g., assessing the achievable performance limits, identifying the critical limitations, and exploring new device concepts and architectures. To make computationally more efficient drift-diffusion and energy-balance based simulators predictive for high performance devices, their physical models for, e.g., carrier transport are obtained from first principles solutions of the Boltzmann transport equation (BTE). Furthermore, WP1 will investigate the ultimate limits of SiGe HBT technology in terms of device performance, transport

limits, quantum effects, and safe operation area limitations [56]. The partners involved in this package are: University of Naples-Italy, ST Microelectronics-France, IMEC Microelectronics-Belgium, IMS-University of Bordeaux 1-France and Bundeswehr University Munich-Germany.

References

[1] S.L. Salmon, J. D. Cressler, R. C. Jaeger, D. L. Harame, "The Influence of Ge Grading on the Bias and Temperature Characteristics of SiGe HBT's for Precision Analog Circuits" vol. 47, 2000.

[2] B. Tillack, B. Heinemann, D. Knoll, H. Rücker, Y. Yamamoto, "Base doping and dopant profile control of SiGe npn and pnp HBTs" vol. 254, 2008.

[3] F. Ellinger, L. C. Rodoni, G. Sialm, C. Kromer, G. von Büren, M. Schmatz, C. Menolfi, T. Toifl, T. Morf, M. Kossel, H. Jäckel, "Effects of emitter scaling and device biasing on millimeter-wave VCO performance in 200 GHz SiGe HBT technology" vol. 52, 2004.

[4] P.R. Chidambaram, C. Bowen, S. Chakravarthi,C. Machala, R. Wise, "Fundamentals of silicon material properties for successful exploitation of strain engineering in modern CMOS manufacturing" vol. 53, 2006.

[5] J.L. Egley, D. Chidambarrao, "Strain effects on device characteristics: Implementation in drift-diffusion simulators" vol. 36, pp. 1653-1664, 1993.

[6] J. D. Cressler, *Silicon heterostructure handbook*. Taylor & Francis, 2006.

[7] John D. Cressler, *Silicon Heterostructure Devices*. CRC Press, 2007.

[8] P. Ashburn, *SiGe Heterojunction Bipolar Transistors*. Wiley, 2003.

[9] "Circuit element utilizing semiconductive material" U.S. Patent 2569347.

[10] W. Shockley, M. Sparks, G.K. Teal, "p-n Junction Transistors" *Physical Review*, vol. 83, 1951.

[11] H. Kroemer, "Theory of a wide-gap emitter for transistor" *Proceedings of the IRE*, vol. 42, 1957.

[12] E. Kasper, H. J. Herzog and H. Kibbel, "A one-dimensional SiGe superlattice grown by UHV epitaxy" *Applied Physics A: Materials Science & Processing*, vol. 8, 1975.

[13] S. Van Huylenbroeck et al., "A 400GHz f_{MAX} Fully Self-Aligned SiGe:C HBT Architecture" *Proceedings IEEE BCTM*, p. 5, 2009.

[14] P. Chevalier et al., "A Conventional Double-Polysilicon FSA-SEG Si/SiGe:C HBT Reaching 400 GHz f_{MAX}" *Proceedings IEEE BCTM*, p. 1, 2009.

[15] J. Bardeen, W.H. Brattain, "The transistor, a semi-conductor triode" *Physical Review*, vol. 71, 1947.

[16] M. Tanenbaum, D.E. Thomas, "Diffused emitter and base silicon transistors" *Bell System Technical Journal*, vol. 35, 1956.

[17] H. Kroemer, "Zur theorie des diffusions und des drift transistors. Part III." *Archiv der Elektrischen Ubertragungstechnik*, vol. 8, 1954.

[18] H. Kroemer, "Quasielectric and quasimagnetic fields in nonuniform semiconductors" *RCA Review*, vol. 18, 1957.

[19] H.C. Theuerer, J.J. Kleimack, H.H. Loar, and H. Christensen, "Epitaxial diffused transistors" *Proceedings of the IRE*, vol. 48, 1960.

[20] S.S. Iyer, G.L. Patton, S.L. Delage, S. Tiwari, and J.M.C. Stork, "Silicon–germanium base heterojunction\line bipolar transistors by molecular beam epitaxy" *Technical Digest of the IEEE International Electron Devices Meeting, San Francisco*, 1987.

[21] C.A .King, J.L. Hoyt, C.M. Gronet, J.F. Gibbons, M.P. Scott, and J. Turner, "$Si/Si_{1-x}/Ge_x$ heterojunction\line bipolar transistors produced by limited reaction processing" *IEEE Electron Device Letters*, vol. 10, 1989.

[22] G.L. Patton, D.L. Harame, J.M.C. Stork, et al., "Graded-SiGe-base, polyemitter heterojunction bipolar transistors" *IEEE Electron Device Letters*, vol. 10, 1989.

[23] G.L Patton, J.H .Comfort, B.S. Meyerson, et al., "63-75 GHz f_T SiGe-base heterojunction-bipolar technology" *Technical Digest IEEE Symposium on VLSI Technology, Honolulu,*, 1990.

[24] G.L. Patton, J.H. Comfort, B.S.Meyerson, et al., "75 GHz f_T SiGe base heterojunction bipolar transistors" *IEEE Electron Device Letters*, vol. 11, 1990.

[25] J.H. Comfort, G.L. Patton, J.D. Cressler,et al., "Profile leverage in a selfaligned epitaxial Si or SiGe-base bipolar technology" *Technical Digest IEEE International Electron Devices Meeting, Washington*, 1990.

[26] D.L. Harame, J.M.C. Stork, B.S. Meyerson, et al., "SiGe-base PNP transistors fabrication with n-type UHV/CVD LTE in a ''NO DT'' process" *Technical Digest IEEE Symposium on VLSI Technology, Honolulu*, 1990.

[27] E.F. Crabbee´, G.L.Patton, J.M.C Stork, B.S. Meyerson, and J.Y-C Sun, "Low temperature operation of Si and SiGe bipolar transistors" *Technical Digest IEEE International Electron Devices Meeting, Washington*, 1990.

[28] D.L.Harame, E.F. Crabbe´, J.D. Cressler, et al., "A high-performance epitaxial SiGe-base ECL BiCMOS technology" *Technical Digest IEEE International Electron Devices Meeting, Washington*, 1992.

[29] D.L. Harame, J.M.C. Stork, B.S. Meyerson, et al., "Optimization of SiGe HBT technology

for high speed analog and mixed-signal applications" *Technical Digest IEEE International Electron Devices Meeting, San Francisco*, 1993.

[30] E. Kasper, A. Gruhle, H. Kibbel, "High speed SiGe-HBT with very low base sheet resistivity" *Technical Digest IEEE International Electron Devices Meeting, San Francisco*, 1993.

[31] E.F Crabbe´, B.S. Meyerson, J.M.C. Stork, and D.L. Harame, "Vertical profile optimization of very high frequency epitaxial Si- and SiGe-base bipolar transistors" *Technical Digest IEEE International Electron Devices Meeting, Washington*, 1993.

[32] D.L Harame, K. Schonenberg, M. Gilbert,, "A 200mm SiGe-HBT technology for wireless and mixed-signal applications" *Technical Digest IEEE International Electron Devices Meeting, Washington*, 1994.

[33] J.D. Cressler, E.F. Crabbe´, J.H. Comfort, .JY-C Sun, and J.M.C Stork, "An epitaxial emitter cap SiGe base bipolar technology for liquid nitrogen temperature operation" *IEEE Electron Device Letters*, vol. 15, 1994.

[34] J.A. Babcock, J.D. Cressler, L.S. Vempati, et al., "Ionizing radiation tolerance of high performance SiGe HBTs grown by UHV/CVD" *IEEE Transactions on Nuclear Science*, vol. 42, 1995.

[35] L.S. Vempati, J.D. Cressler, R.C. Jaeger, and D.L. Harame, "Low-frequency noise in UHV/CVD Si- and SiGe-base bipolar transistors" *Proceedings of the IEEE BCTM, Minnneapolis*, 1995.

[36] L. Lanzerotti, A. St Amour, C.W. Liu, et al., "Si/$Si_{1-x-y}Ge_xC_y$ /Si heterojunction bipolar transistors" *IEEE Electron Device Letters*, vol. 17, 1996.

[37] A. Schuppen, S. Gerlach, H. Dietrich, et al, "1-W SiGe power HBTs for mobile communications" *IEEE Microwave and Guided Wave Letters*, 1996.

[38] P.A. Potyraj, K.J. Petrosky, K.D.Hobart, et al, "A 230-Watt S-band SiGe heterojunction junction bipolar transistor" *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, 1996.

[39] K. Washio, E. Ohue, K. Oda, et al., "A selective-epitaxial SiGe HBT with SMI electrodes featuring 9.3-ps ECL-Gate Delay" *Technical Digest IEEE International Electron Devices Meeting, San Francisco*, 1997.

[40] H.J. Osten, D. Knoll, B. Heinemann, et al., "Carbon doped SiGe heterojunction bipolar transistors for high frequency applications" *Proceedings of the IEEE BCTM, Minnneapolis*, 1999.

[41] S.J. Jeng, B. Jagannathan, J-S.Rieh, et al., "A 210-GHz f_T SiGe HBT with nonself- aligned structure" *IEEE Electron Device Letters*, vol. 22, 2001.

[42] J.S. Rieh, B. Jagannathan, H. Chen, et al., "SiGe HBTs with cut-off frequency of 350 GHz" *Technical Digest of the IEEE International Electron Devices Meeting, San Francisco*, p. 2002.

[43] B. El-Kareh, S. Balster, W. Leitz, et al., "A 5V complementary SiGe BiCMOS technology for high-speed precision analog circuits" *Proceedings of the IEEE BCTM, Toulouse*, 2003.

[44] B. Heinemann, R. Barth, D. Bolze, et al., "A complementary BiCMOS technology with high speed npn and pnp SiGe:C HBTs" *Technical Digest of the IEEE International Electron Devices Meeting, Washington*, 2003.

[45] J. Cai, M. Kumar, M. Steigerwalt, et al., "Vertical SiGe-base bipolar transistors on CMOS-compatible SOI substrate" *Proceedings of the IEEE BCTM, Toulouse*, 2003.

[46] J.S. Rieh, D. Greenberg, M. Khater, et al., "SiGe HBTs for millimeterwave applications with simultaneously optimized f_T and f_{MAX}" *Proceedings of the IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, Fort Worth*, p. 2004.

[47] T.J. Wang, H.W. Chen, P.C. Yeh, et al., "Effects of Mechanical Uniaxial Stress on SiGe HBT Characteristics" *Journal of The Electrochemical Society*, vol. 154, 2007.

[48] S. Persson, M. Fjer, E. Escobedo-Cousin, et al., "Fabrication and characterization of strained Si heterojunction bipolar transistors on virtual substrates" *Proceedings IEDM 2008*, 2008.

[49] F. Yuan, S.-R. Jan, S. Maikap,, "Mechanically Strained Si–SiGe HBTs" *IEEE Electron Device Letters*, vol. 25, 2004.

[50] B. M. Haugerud , M. B. Nayeem, R. Krithivasan, et al., "The effects of mechanical planar biaxial strain in Si/SiGe HBT BiCMOS technology" *Solid-State Electronics*, vol. 49, 2005.

[51] D. Chidambarrao, G. G. Freeman, M. H. Khater, "Bipolar Transistor with Extrinsic Stress Layer" U.S. Patent US 7,102,205 B2Sep-2006.

[52] J.S. Dunn, D.L.Harame, J.B. Johnson, A.B. Joseph, "Structure and Method for Performance Improvement in Vertical Bipolar Transistors" U.S. Patent US 7,262,484 B2Aug-2007.

[53] S. Decoutere, A. Sibaja-Hernandez, "IMEC Microelectronics private communication" 2008.

[54] N. D. Jankovic, and A. O'Neill, "Enhanced performance virtual substrate heterojunction bipolar transistor using strained-Si/SiGe emitter" *Semicond. Sci. Technol.*, vol. 18, 2003.

[55] N.D. Jankovic, A. O'Neill, "Performance evaluation of SiGe heterojunction bipolar transistors on virtual substrates" *Solid-State Electronics*, vol. 48, 2004.

[56] "http://www.dotfive.eu/".

CHAPTER 2

Strain Technology

With the continuing reduction of silicon integrated circuits, new engineering solutions and innovative techniques are required to improve bipolar transistors performance, and to overcome the physical limitations of the device scaling. Therefore, strained-silicon technology has become a strong competitor in search for alternatives to transistor scaling and new materials for improved devices and circuits performances. Strained-Si technology enables improvements in electronic devices performance and functionality via replacement of the bulk crystalline-Si substrate with a strained-Si substrate. The improved performance comes from the properties of strained-Si itself through changing the nature of the wafer by stretching and/or compressing the placement of the atoms. This chapter gives an overview of the elasticity theory of solids, physics behind strain, different strain types and strain application techniques.

1. Theory of Elasticity

Elasticity is the ability of a solid body to recover its shape when the deforming forces are removed. The deformation of an elastic material obeys Hooke's law, which states that deformation is proportional to the applied stress up to a certain point. This point is called the elastic limit. Beyond this point additional stresses will cause permanent deformation [1]. The main law governing elasticity of materials is the theory of stress, strain, and their interdependence will be discussed.

1.1 The Stress Tensor

Stress is defined as the force per unit area. When a deforming force is applied to a body, the stress is defined as the ratio of the force to the area over which it is applied. There are two basic types of stress; if the force is perpendicular (normal) to the surface over which it is acting, then the stress is termed as normal stress, and if it is tangential to the surface, it is called a shear stress. Usually, the force is neither entirely normal nor tangential, but it is at some arbitrary intermediate angle. In this case it can be resolved into components which are both normal and tangential to the surface; so the stress is composed of both normal and shearing components. The sign convention is that tensional stresses are positive and compression stresses are negative.

Let's take an arbitrary solid body oriented in a Cartesian coordinate system, with a number of forces acting on it in different directions, such that the net force (the vector sum of the forces) on the body is zero. Conceptually, we slice the body on a plane normal to the X - direction (parallel to the YZ-plane) as shown in Fig.26.

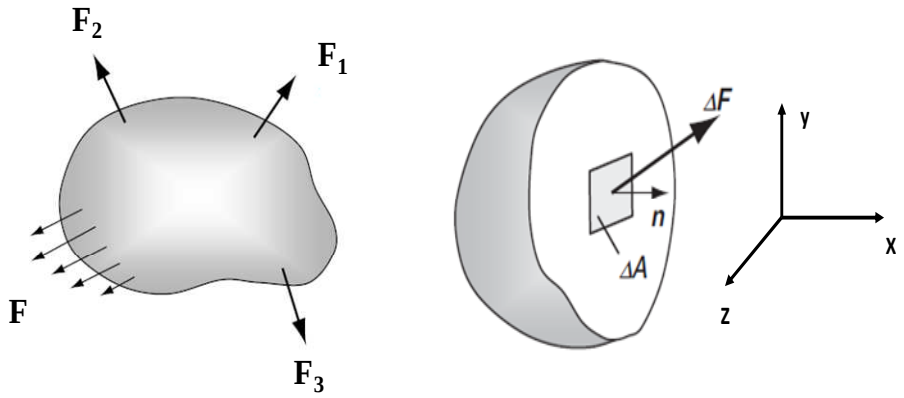

Fig.26: Arbitrary solid body under external forces (left) and a section of the solid body under external force (right).

A small area on this plane can be defined as

$$\Delta A_x = \Delta A\, \hat{\imath} \qquad 2.1$$

The total force acting on this small area is given by

$$\Delta F = \Delta F_x\, \hat{\imath} + \Delta F_y\, \hat{\jmath} + \Delta F_z\, \hat{k} \qquad 2.2$$

We can define three scalar quantities:

$$\left.\begin{aligned}\sigma_{xx} &= \lim_{\Delta A_{x\to 0}} \frac{\Delta F_x}{\Delta A_x} \\ \sigma_{xy} &= \lim_{\Delta A_{x\to 0}} \frac{\Delta F_y}{\Delta A_x} \\ \sigma_{xz} &= \lim_{\Delta A_{x\to 0}} \frac{\Delta F_z}{\Delta A_x}\end{aligned}\right\} \quad 2.3$$

The first subscript refers to the plane and the second refers to the force direction. Similarly considering slices orthogonal to the Y and Z -directions, we obtain

$$\left.\begin{aligned}\sigma_{yy} &= \lim_{\Delta A_{y\to 0}} \frac{\Delta F_y}{\Delta A_y} \\ \sigma_{yx} &= \lim_{\Delta A_{y\to 0}} \frac{\Delta F_x}{\Delta A_y} \\ \sigma_{yz} &= \lim_{\Delta A_{y\to 0}} \frac{\Delta F_z}{\Delta A_y}\end{aligned}\right\} \quad 2.4$$

$$\left.\begin{aligned}\sigma_{zz} &= \lim_{\Delta A_{z\to 0}} \frac{\Delta F_z}{\Delta A_z} \\ \sigma_{zx} &= \lim_{\Delta A_{z\to 0}} \frac{\Delta F_x}{\Delta A_z} \\ \sigma_{zy} &= \lim_{\Delta A_{z\to 0}} \frac{\Delta F_y}{\Delta A_z}\end{aligned}\right\} \quad 2.5$$

For static equilibrium, the shear stress components across the diagonal are identical ($\sigma_{xy} = \sigma_{yx}, \sigma_{xz} = \sigma_{zx}$, and $\sigma_{yz} = \sigma_{zy}$), resulting in six independent scalar quantities. These scalar quantities can be arranged in a matrix form to yield the stress tensor [2]:

$$\boldsymbol{\sigma} = \sigma_{ij} = \begin{pmatrix}\sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{zy} & \sigma_{zz}\end{pmatrix} \quad 2.6$$

1.2 The Strain Tensor

Strain is defined as the change of the object length in a given direction divided by the object initial length in that direction. If a force is applied to a solid object, that may simultaneously translating, rotating, and deforming the object [3]. If we consider the two arbitrary neighboring points P and Q are marked at initial position x and $x + \mathrm{d}x$ respectively. After

deformation these points move to position $x + \mathrm{u}(x)$, and $x + \mathrm{d}x + \mathrm{u}(x + \mathrm{d}x)$ respectively. The absolute squared distance between the deformed points can be written as

$$\grave{l}^2 = \sum_i [dx_i + u_i(x + dx) - u_i(x)]^2 \qquad 2.7$$

For small displacements dx, a Taylor expansion about the point x gives the absolute squared distance as

$$\grave{l}^2 = \sum_i \left[dx_i + \frac{\partial u_i}{\partial x_j} dx_j\right]^2$$

$$= \sum_i {dx_i}^2 + 2\sum_{i,j} \frac{\partial u_i}{\partial x_j} dx_i dx_j + \sum_{i,j,k} \frac{\partial u_i}{\partial x_j} dx_j \frac{\partial u_i}{\partial x_k} dx_k \qquad 2.8$$

The squared distance between the original points can be written as

$$l^2 = \sum_i dx_i^2 \qquad 2.9$$

The change in the squared distance can be written as

$$\grave{l}^2 - l^2 = 2\sum_{i,j} \frac{\partial u_i}{\partial x_j} dx_i dx_j + \sum_{i,j,k} \frac{\partial u_i}{\partial x_j} dx_j \frac{\partial u_i}{\partial x_k} dx_k$$

$$= \sum_{i,j} dx_i \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_i}{\partial x_i}\right) dx_j + \sum_k \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} dx_i\, dx_j$$

$$= \sum_{i,j} dx_i \left[\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_i}{\partial x_i}\right) dx_j + \sum_k \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} dx_i \right] dx_j$$

$$= 2\sum_{i,j} dx_i \varepsilon_{ij}\, dx_j \qquad 2.10$$

Where ε_{ij} are the strain tensor components, and are defined as

$$\varepsilon_{ij} = \frac{1}{2}\left[\frac{\partial u_i}{\partial x_j} + \frac{\partial u_i}{\partial x_i} + \sum_k \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j}\right] \qquad 2.11$$

For $\frac{\partial u_k}{\partial x_i} \ll 1$, the second term in equation (2.11) can be neglected, and the resultant tensor is

$$\varepsilon_{ij} = \frac{1}{2}\left[\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right] \qquad 2.12$$

Therefore, the strain tensor is analogous to the stress tensor and can be written as

$$\varepsilon = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} \qquad 2.13$$

The diagonal terms are the normal strains in the directions X, Y, and Z respectively. While the off-diagonal terms are equal to one half of the engineering shear strain.

The strain components in three dimensions can be written as

$$\left.\begin{aligned} \varepsilon_{xx} &= \frac{\partial u}{\partial x} \quad , \qquad \varepsilon_{xy} = \varepsilon_{yx} = \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) \\ \varepsilon_{yy} &= \frac{\partial v}{\partial y} \quad , \qquad \varepsilon_{yz} = \varepsilon_{zy} = \frac{1}{2}\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right) \\ \varepsilon_{yy} &= \frac{\partial w}{\partial z} \quad , \qquad \varepsilon_{zx} = \varepsilon_{xz} = \frac{1}{2}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right) \end{aligned}\right\} \qquad 2.14$$

Where *u*, *v* and *w* are the displacements in the X, Y and Z directions, respectively [4].

1.3 Stress-Strain Relationship

Stress and strain are linked in elastic media by a stress-strain or constitutive relationship. This relation between stress and strain was first identified by Robert Hook. For Hookean elastic solid, the stress tensor is linearly proportional to the strain tensor over a specific range of deformation. The most general linear relationship between the stress and strain tensors can be written as

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} \qquad 2.15$$

Where C_{ijkl} is a fourth-order elastic stiffness tensor with 81 (3^4) elements.

However, due to the symmetries involved for the stress and strain tensors under equilibrium, C_{ijkl} is reduced to a tensor of 36 elements. To simplify the notations, the stress and strain tensors can be written as vectors using the contracted notations. First the off-diagonal strain

terms are converted to engineering shear strains (The off-diagonal terms are equal to one-half of the engineering shear strain).

$$\begin{bmatrix} \varepsilon_{xx} & 2\varepsilon_{xy} & 2\varepsilon_{xz} \\ 2\varepsilon_{yx} & \varepsilon_{yy} & 2\varepsilon_{yz} \\ 2\varepsilon_{zx} & 2\varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} = \begin{bmatrix} \varepsilon_{xx} & \gamma_{xy} & \gamma_{xz} \\ \gamma_{yx} & \varepsilon_{yy} & \gamma_{yz} \\ \gamma_{zx} & \gamma_{zy} & \varepsilon_{zz} \end{bmatrix} \quad 2.16$$

Where γ is the engineering shear strain.

The resulting matrix is no longer a tensor because it doesn't follow the coordinate-transformation rules. Then the elements are renumbered as the following

$$\left.\begin{matrix} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} = \begin{bmatrix} \sigma_1 & \sigma_6 & \sigma_5 \\ \sigma_6 & \sigma_2 & \sigma_4 \\ \sigma_5 & \sigma_4 & \sigma_3 \end{bmatrix} \\ \begin{bmatrix} \varepsilon_{xx} & \gamma_{xy} & \gamma_{xz} \\ \gamma_{yx} & \varepsilon_{yy} & \gamma_{yz} \\ \gamma_{zx} & \gamma_{zy} & \varepsilon_{zz} \end{bmatrix} = \begin{bmatrix} \varepsilon_1 & \varepsilon_6 & \varepsilon_5 \\ \varepsilon_6 & \varepsilon_2 & \varepsilon_4 \\ \varepsilon_5 & \varepsilon_4 & \varepsilon_3 \end{bmatrix} \end{matrix}\right\} \quad 2.17$$

The relationship between the stress vector and the strain vector can be written as

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13}\, C_{14} & C_{15} & C_{16} \\ C_{12} & C_{22} & C_{23}\, C_{24} & C_{25} & C_{26} \\ C_{13} & C_{23} & C_{33}\, C_{34} & C_{35} & C_{36} \\ C_{14} & C_{24} & C_{34}\, C_{44} & C_{45} & C_{46} \\ C_{15} & C_{25} & C_{35}\, C_{45} & C_{55} & C_{56} \\ C_{16} & C_{26} & C_{36}\, C_{46} & C_{56} & C_{66} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} \quad 2.18$$

The material property matrix with all of the elastic tensor constants (C's) is known as the stiffness matrix. The inverse of the stiffness matrix is called compliance matrix, S, where $S = C^{-1}$ [5]-[6]. The compliance matrix is written as

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13}\, S_{14} & S_{15} & S_{16} \\ S_{12} & S_{22} & S_{23}\, S_{24} & S_{25} & S_{26} \\ S_{13} & S_{23} & S_{33}\, S_{34} & S_{35} & S_{36} \\ S_{14} & S_{24} & S_{34}\, S_{44} & S_{45} & S_{46} \\ S_{15} & S_{25} & S_{35}\, S_{45} & S_{55} & S_{56} \\ S_{16} & S_{26} & S_{36}\, S_{46} & S_{56} & S_{66} \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix} \quad 2.19$$

for linear elastic isotropic materials where the physical properties are independent of direction.

Therefore, Hooke's law takes on a simple form involving only two independent variables [7]. In stiffness form, Hooke's law for the isotropic medium is

$$\begin{bmatrix}\sigma_{xx}\\ \sigma_{yy}\\ \sigma_{zz}\\ \sigma_{yz}\\ \sigma_{zx}\\ \sigma_{xy}\end{bmatrix} = \frac{E}{(1+v)(1-2v)}\begin{bmatrix}1-v & v & v & 0 & 0 & 0\\ v & 1-v & v & 0 & 0 & 0\\ v & v & 1-v & 0 & 0 & 0\\ 0 & 0 & 0 & \frac{1}{2}-v & 0 & 0\\ 0 & 0 & 0 & 0 & \frac{1}{2}-v & 0\\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}-v\end{bmatrix}\begin{bmatrix}\varepsilon_{xx}\\ \varepsilon_{yy}\\ \varepsilon_{zz}\\ \varepsilon_{yz}\\ \varepsilon_{zx}\\ \varepsilon_{xy}\end{bmatrix} \quad 2.20$$

where *E* is the Young's modulus. *v* is the Poisson's ratio, which is defined as the ratio of transverse to longitudinal strains of a loaded specimen.

For anisotropic materials such as cubic crystals (i.e. Si, and Ge crystals), in which their elastic properties are direction dependent. It is possible to simplify Hook's law by considerations of cubic symmetry. If the X, Y, and Z axes coincide with the [100], [010], and [001] directions in the cubic crystal, respectively, then Hooke's law in stiffness form can be written as

$$\begin{bmatrix}\sigma_{xx}\\ \sigma_{yy}\\ \sigma_{zz}\\ \sigma_{yz}\\ \sigma_{zx}\\ \sigma_{xy}\end{bmatrix} = \begin{bmatrix}C_{11} & C_{12} & C_{12} & 0 & 0 & 0\\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0\\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0\\ 0 & 0 & 0 & C_{44} & 0 & 0\\ 0 & 0 & 0 & 0 & C_{44} & 0\\ 0 & 0 & 0 & 0 & 0 & C_{44}\end{bmatrix}\begin{bmatrix}\varepsilon_{xx}\\ \varepsilon_{yy}\\ \varepsilon_{zz}\\ \varepsilon_{yz}\\ \varepsilon_{zx}\\ \varepsilon_{xy}\end{bmatrix} \quad 2.21$$

For cubic crystals, the compliance-stiffness constants relationships are given by

$$C_{11} = \frac{S_{11}+S_{12}}{(S_{11}-S_{12})(S_{11}+2S_{12})} \quad 2.22$$

$$C_{12} = \frac{-s_{12}}{(S_{11}-S_{12})(S_{11}+2S_{12})} \quad 2.23$$

$$C_{44} = \frac{1}{S_{44}} \quad 2.24$$

$$S_{11} = \frac{C_{11}+C_{12}}{(C_{11}-C_{12})(C_{11}+2C_{12})} \quad 2.25$$

$$S_{12} = \frac{-C_{12}}{(C_{11}-C_{12})(C_{11}+2C_{12})} \quad 2.26$$

$$S_{12} = \frac{1}{C_{44}} \quad 2.27$$

The stiffness and compliance coefficients for Si and Ge are listed in Table 3.

	C_{11}	C_{12}	C_{44}	S_{11}	S_{12}	S_{44}
Si	165.64	63.94	79.51	0.7691	-0.2142	1.2577
Ge	128.7	47.7	66.7	0.9718	-0.2628	1.499

Table 3: The elastic compliance coefficients Cij [GPa], and the elastic stiffness coefficients Sij [10^{-12} $m^2.N^{-1}$] values for Si and Ge.

1.4 Young's Modulus

Young's Modulus, E, is defined as the ratio of elastic stress to strain. It is a measure of the material's resistance to elastic deformation. The value of Young's modulus, E, depends on the direction of the applied force (anisotropic). For an arbitrary crystallographic direction, E can be written as:

$$E^{-1} = S_{11} - 2\left(S_{11} - S_{12} - \frac{1}{2}S_{44}\right)(\alpha^2\beta^2 + \alpha^2\gamma^2 + \beta^2\gamma^2) \qquad 2.28$$

where S_{ij} are the elastic compliance constants. α, β, and γ are the direction cosines of the applied force with respect to the crystallographic axis [8].

The following are the measured values for the modulus E in silicon at room temperature for different directions of the applied force [9]-[10].

$$E_{[100]} = \frac{1}{S_{11}} = 131 \text{ GPa} \qquad 2.29$$

$$E_{[110]} = \frac{4}{(2S_{11} + 2S_{12} + S_{44})} = 169 \text{ GPa} \qquad 2.30$$

$$E_{[111]} = \frac{3}{(S_{11} + 2S_{12} + S_{44})} = 187 \text{ GPa} \qquad 2.31$$

Where, $E_{[100]}$, $E_{[110]}$, and $E_{[111]}$ are the Young's modulus that corresponds to the applied forces along the directions [100], [110] and [111], respectively.

1.5 Miller Indices (hkl)

The orientations and properties of the surface crystal planes are important. Since semiconductor devices are built on or near the semiconductor surface. A convenient method of defining the various planes in a crystal is to use Miller indices [11]. Miller indices are a symbolic vector representation in crystallography for the orientation of an atomic plane in a crystal lattice and are defined as the reciprocals of the fractional intercepts which the plane makes with the crystallographic axes, and denoted as *h*, *k* and *l*. The direction [*hkl*] defines a vector direction normal to surface of a particular plane or facet. Fig.27 shows the Miller indices of three important planes in a cubic crystal [12].

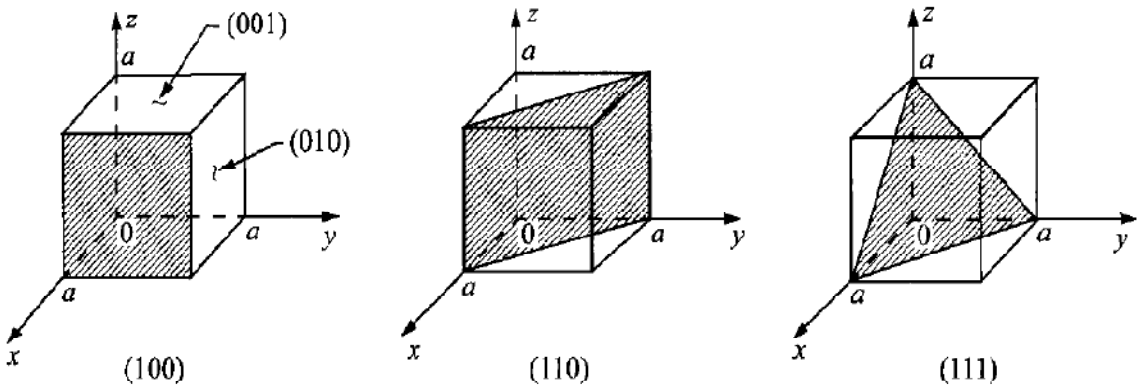

Fig.27: Miller indices of three important planes in a cubic crystal.

1.6 Coordinate Transformation

It is often useful to know the stress tensor in the crystallographic coordinate system for a stress applied along a general direction with respect to the crystallographic coordinate system [13]. A stress applied in a generalized direction $[x', y', z']$ can be transformed to stress in the crystallographic coordinate system $[x, y, z]$ using the following transformation matrix, U

$$U = \begin{bmatrix} cos\theta cos\emptyset & cos\theta sin\emptyset & -sin\theta \\ -sin\emptyset & cos\emptyset & 0 \\ sin\theta cos\emptyset & sin\theta sin\emptyset & cos\theta \end{bmatrix} \quad 2.32$$

Where θ is the polar angle, and $\emptyset$ is the azimuthal angle of the applied stress direction relative to the crystallographic coordinate system as shown in Fig.28.

The stress in the crystallographic coordinate system is given by

$$\sigma = U.\sigma'.U^T \qquad 2.33$$

Where σ' is the stress applied in a generalized coordinate system.

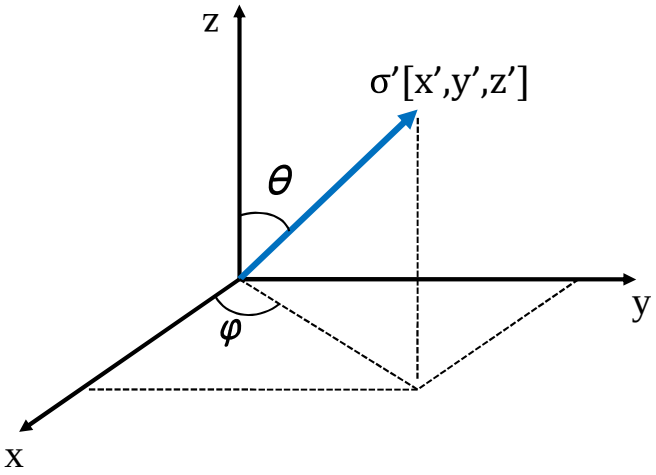

Fig.28: Stress direction [x',y',z'] relative to the crystallographic coordinate system [x,y,z].

2. Piezoresistivity

Piezoresistance is defined as the change in electrical resistance of a solid when subjected to stress. The piezoresistance coefficients (π) that relate the piezoresistivity and stress are defined by

$$\pi = \frac{\Delta R/R}{\sigma} \qquad 2.34$$

Where R is the original resistance that is related to semiconductor sample dimension by $R = \rho\frac{l}{wh}$ where ρ is the resistivity, l, w, and h are the length, the width, and the height of the sample respectively. ΔR signifies the change of resistance, and σ is the applied mechanical stress.

The ratio of ΔR to R can be expressed as the following

$$\frac{\Delta R}{R} = \frac{\Delta l}{l} - \frac{\Delta w}{w} - \frac{\Delta h}{h} + \frac{\Delta \rho}{\rho} \qquad 2.35$$

The first three terms of equation (2.35) represent the geometrical change of the sample under stress, and the last term $\Delta\rho/\rho$ is the resistivity dependence on stress. For most semiconductors, the stress-induced resistivity change is much larger than the geometrical change-induced resistance change, therefore, the resistivity change by stress is the determinant factor of the piezorestivity.

The resistivity change, $\Delta\rho$, is connected to stress by a fourth-rank tensor π_{ijkl} , and is given by

$$\frac{\Delta \rho_{ij}}{\rho} = \sum_{k,l} \pi_{ijkl}\sigma_{kl} \qquad 2.36$$

In the vector form we can rewrite $\Delta\rho_{ij}$ as $\Delta\rho_i$, where i=1,2,…,6. Therefore, equation (2.36) can be written as

$$\frac{\Delta \rho_i}{\rho} = \sum_{\mathrm{k}=1}^{6} \pi_{ik}\sigma_k \qquad 2.37$$

Where π_{ik} is a 6×6 matrix.

For a cubic crystals such as Si, π_{ik} has only three independent elements due to the cubic symmetry.

$$\pi_{ik} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{11} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{12} & \pi_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \pi_{44} \end{bmatrix} \qquad 2.38$$

Where π_{11} describes the piezoresistive effect for stress along the principal crystal axis (longitudinal piezoresistive effect). π_{12} describes the piezoresistive effect for stress directed perpendicular to the principal crystal axis (transverse piezoresistive effect). π_{44} describes the piezoresistive effect on an out-of-plane electric field by the change of the in-plane current induced by in-plane shear stress [14] [15].

3. Element of Bulk Si and Ge

Si and Ge are elements of group IV with four electrons in the outermost shell, and they have diamond lattice structure, where each atom is surrounded by four equidistant nearest neighbors which lie at the corners of a tetrahedron. The unit cell can be considered as two interpenetrating face-centered cubic (fcc) lattices separated by a/4 along each axis of the cell, where, a, is the lattice constant as shown in Fig.29. At 300K, the lattice constants of Si and Ge are 5.431 Å and 5.6575 Å, respectively [9].

The first Brillouin zone represents the central (Wigner-Seitz) cell of the reciprocal lattice. It contains all points nearest to the enclosed reciprocal lattice point. The first Brillouin zone for cubic semiconductors is a truncated octahedron. It has fourteen plane faces; six square faces along the <100> directions and eight hexagonal faces along the <111> directions. The coordinate axes of the Brillouin zone are the wave vectors of the plane waves corresponding to the Bloch states (electrons) or vibration modes (phonons). The points and directions of symmetry are conventionally denoted by Greek letters, as shown in Fig.29. The zone center is called the Γ point (k=0), the directions <100>, <110>, and <111> are called, respectively, Δ, Σ, and Λ directions and their intersections with the zone boundaries are called, X, K and L points respectively [12].

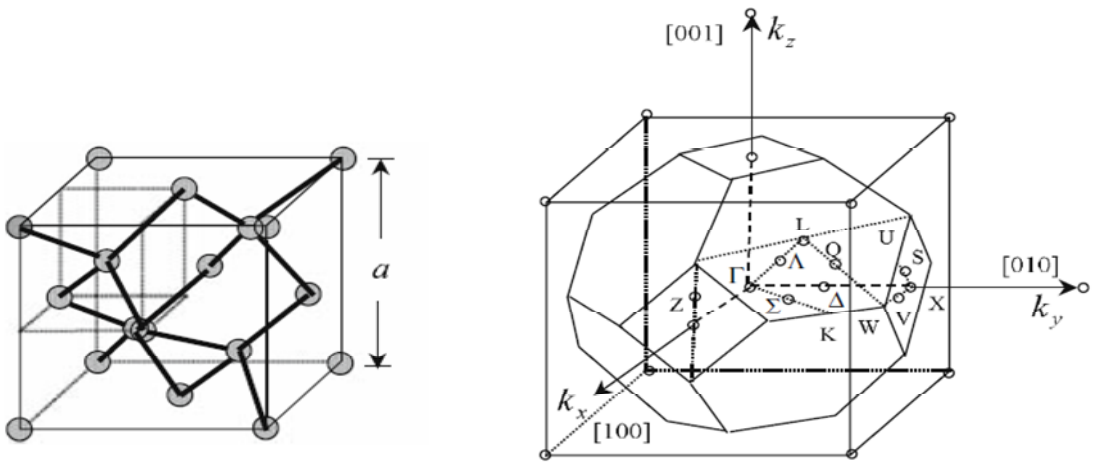

Fig.29 : Structure of the fcc crystal lattice (left), and the first Brillouin zone of the fcc lattice (right).

3.1 Energy Band Structure

Band structure is one of the most important concepts in solid state physics, it describes the variation of energy, E, with the wave vector, k. The band of filled or bonding states is called the valence band. The band of empty or anti-bonding states is called the conduction band. The highest energy occupied states are separated from the lowest energy unoccupied states by an energy region containing no states known as the bandgap. The energy difference between the top of the valence band and the bottom of the conduction band is, Eg, the bandgap energy.

Si and Ge are indirect gap semiconductor materials. The conduction band minima of silicon is a six-fold degenerate, and located close to the X point at $0.85\pi/a$ in the <100> direction. The valence band maximum is located at the G-point and it consists of light hole (LH), heavy hole (HH), and spin-orbit (SO) hole bands. The LH and HH bands are degenerate at the G-point while the SO has 44 meV split from the others bands. In contrast, Germanium has a smaller band gap than Silicon and a higher atomic mass. The Ge conduction band minima is a four-fold degenerate, and located at the L-point along the <111> direction on the first Brillouin zone boundary. The energy band diagram of Si and Ge are shown in Fig.30. At 300 K, the indirect bandgap energy for Si and Ge are 1.12 eV and 0.664 eV, respectively [16] [17].

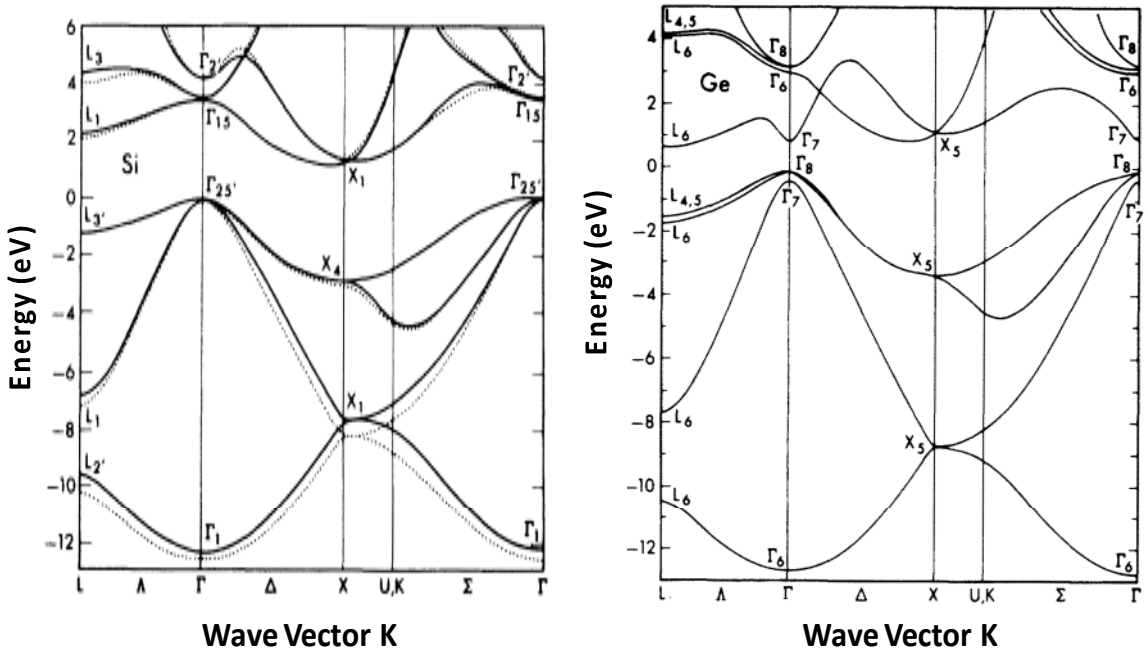

Fig.30 : Electronic band-structure of Si and Ge calculated by Pseudopotential method.

3.2 Calculation of Energy Bands

A wide range of techniques have been employed to calculate the energy band dispersion curves of semiconductor materials. The most frequently used methods are the orthogonalized plane-wave method (OPW), the pseudopotential method and the k.p method.

From quantum mechanics, Schrödinger equation can be solved by expanding the eigenfunction in terms of a complete basis function and developing a matrix eigenvalue equation. A plane wave basis can be used to do so, but it has a difficulty because many plane waves are needed to describe the problem adequately. The OPW method has been proposed by Herring in 1940 [18]. It is an approach to avoid having to deal with a very large number of plane wave states. The basic idea is that the valence and conduction band states are orthogonal to the core states of the crystal, and this fact should be utilized in the selection of the plane wave, resulting in a reduced number of plane wave states used in solving the problem.

The pseudopotential method and the k.p method for calculating the band structure will be discussed in details in the following sections.

3.2.1 The Pseudopotential Method

The pseudopotential method is a technique to solve for band structures of semiconductors. This method makes use of the information that the valence and conduction band states are orthogonal to the core states. In addition to that, this method uses empirical parameters known as pseudopotentials to solve the Schrödinger equation in the one-electron approximation [19]:

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + V(\vec{r})\right]\psi(\vec{r}) = E_n(\vec{k})\psi(\vec{r}) \qquad 2.39$$

Assuming that the electrons wave functions of the core states and their energies are given by φ_j and ε_j respectively. We then have

$$H|\varphi_j\rangle = [H_0 + V(\vec{r})]|\varphi_j\rangle = \varepsilon_j|\varphi_j\rangle \qquad 2.40$$

Where

$$H_0 = -\frac{\hbar^2}{2m_0}\nabla^2 \qquad 2.41$$

The orthogonality condition is given by

$$\langle\varphi_j|\psi\rangle = 0 \qquad 2.42$$

this equation is called the orthogonalized plane wave.

The orthogonality condition is satisfied when we choose the wave function given by

$$|\psi(\vec{k},\vec{r})\rangle = |\chi_n(\vec{k},\vec{r})\rangle - \sum_j \langle\varphi_j|\chi_n\rangle|\varphi_j\rangle \qquad 2.43$$

by substituting this equation into equation (2.39) we have

$$H|\chi_n\rangle - \sum_j \langle\varphi_j|\chi_n\rangle\, H\,|\varphi_j\rangle = E_n(\vec{k})\left[|\chi_n\rangle - \sum_j \langle\varphi_j|\chi_n\rangle|\varphi_j\rangle\right] \qquad 2.44$$

Then, the following relation is obtained

$$H|\chi_n\rangle + \sum_j [E_n(\vec{k}) - \varepsilon_j]|\varphi_j\rangle\langle\varphi_j|\chi_n\rangle = E_n(\vec{k})|\chi_n\rangle \qquad 2.45$$

Equation (2.44) can be written as follows

$$[H + V_{\mathrm{P}}(\vec{r})]|\chi_n\rangle = \mathrm{E}_n(\vec{k})|\chi_n\rangle \qquad 2.46$$

Or

$$[H_0 + V(\vec{r}) + V_{\mathrm{P}}(\vec{r})]|\chi_n\rangle = E_n(\vec{k})|\chi_n\rangle \qquad 2.47$$

Where

$$V_{\mathrm{P}}(\vec{r}) = \sum_j [E_n(\vec{k}) - \varepsilon_j]|\varphi_j\rangle\langle\varphi_j \qquad 2.48$$

There exists an inequality relation between the energies of the core states ε_j and the energies of the valence and conduction bands $E_n(\vec{k})$, which is given by

$$E_n(\vec{k}) > \varepsilon_j \qquad 2.49$$

Thus,

$$V_{\mathrm{P}} \quad > 0 \qquad 2.50$$

Therefore equation (2.45) can be written as

$$[H_0 + V_{\mathrm{Ps}}(\vec{r})]|\chi_n\rangle = E_n(\vec{k})|\chi_n\rangle \qquad 2.51$$

$$V_{\mathrm{Ps}}(\vec{r}) = V(\vec{r}) + V_{\mathrm{P}}(\vec{r}) \qquad 2.52$$

Where $V_{\mathrm{Ps}}(\vec{r})$ is called the pseudopotential, which is periodic and can be expanded by Fourier series as follows

$$V_{\mathrm{Ps}}(\vec{\mathrm{r}}) = \sum_{\mathrm{j}} V_{\mathrm{Ps}}(\vec{G_j})\, \mathrm{e}^{-\mathrm{i}\vec{G_j}.\vec{r}} \qquad 2.53$$

where $V_{\mathrm{Ps}}(\vec{G_j})$ are the Fourier coefficients, and they are given by

$$V_{Ps}(\vec{G_j}) = \frac{1}{\sqrt{\Omega}} \int_{\Omega} V_{Ps}(\vec{r})\, e^{-i\vec{G_j}.\vec{r}}\, d^3r \qquad 2.54$$

$V_{\mathrm{Ps}}(\vec{G_j})$ maybe chosen so that the potential $V_{Ps}(\vec{G_j})$ is expressed with a small number of the Fourier coefficients $V_{\mathrm{Ps}}(\vec{G_j})$, and therefore, the small values of $V_{\mathrm{Ps}}(\vec{G_j})$ can be neglected.

By using the empirical pseudopotential method, the Fourier coefficients of $V_{\mathrm{Ps}}(\vec{r})$ are empirically chosen so that the shape of the critical points and their energies are in good agreement with experimental observation. The energy band calculations based on the empirical pseudopotential method takes into account as few pseudopotentials $V_{\mathrm{Ps}}(\vec{r})$ as possible, and use the Bloch functions of the free-electron bands for the wave functions $|\chi_n\rangle$.

The energy bands are obtained by solving the equation:

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + V_{\mathrm{Ps}}(\vec{r})\right]|\chi_n(\vec{r})\rangle = E_n|\chi_n(\vec{r})\rangle \qquad 2.55$$

Where

$$|\chi_n\rangle = \mathrm{e}^{\mathrm{i}(\vec{k}+\vec{G_j}).\vec{r}} \qquad 2.56$$

Substituting equation (2.53) and equation (2.50) in equation (2.52), then the energy band structures are given by the following eigenvalue equation:

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + \sum_{j'} V_{\mathrm{Ps}}\left(\overrightarrow{G_{j'}}\right) \mathrm{e}^{-\mathrm{i}\overrightarrow{G_{j'}}.\vec{r}}\right] \mathrm{e}^{\mathrm{i}(\vec{k}+\overrightarrow{G_j}).\vec{r}} = E_n\left(\vec{k}\right)\mathrm{e}^{\mathrm{i}(\vec{k}+\overrightarrow{G_j}).\vec{r}} \qquad 2.57$$

To obtain the eigenvalues and eigenfunctions of the above equation, we define the following

$$H_P = -\frac{\hbar^2}{2m_0}\nabla^2 + V_{Ps}(\vec{r}) \qquad 2.58$$

Equation 2.54 becomes:

$$H_P|\vec{k} + \overrightarrow{G_j}\rangle = E_n\left(\vec{k}\right)|\vec{k} + \overrightarrow{G_j}\rangle \qquad 2.59$$

Where

$$|\vec{k} + \overrightarrow{G_j}\rangle = \mathrm{e}^{\mathrm{i}(\vec{k}+\overrightarrow{G_j}).\vec{r}} \qquad 2.60$$

Then the solutions are obtained by solving the determinate

$$\left\|\langle|\vec{k} + \overrightarrow{G_i}\,|H_P|\vec{k} + \overrightarrow{G_j}\rangle - E\left(\vec{k}\right)\delta_{i,j}\right\| = 0 \qquad 2.61$$

3.2.2 The k.p Method

The k.p method starts with the known form of the band structure problem at the edges, and using the perturbation theory to study wave functions according to the crystal symmetry, so that band structures away from the highly symmetry points in k space can be obtained. Additionally, using this method one can obtain analytic expressions for band dispersion and effective masses around high-symmetry points [20] [21].

Assuming that the eigenvalues and Bloch functions are known for a semiconductor with a band edge at k_0 (k_0 is at position Γ-point = [000] in the Brillouin zone). The Schrödinger equation for a one-electron system is given by

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + V(\vec{r})\right]\psi(\vec{k},\vec{r}) = E_n(\vec{k})\psi(\vec{k},\vec{r}) \qquad 2.62$$

Where $V(\vec{r})$ is the potential energy with the lattice periodicity, $\psi(\vec{k},\vec{r})$is the wave function, and $E_n(\vec{k})$ is the total energy.
The solution of equation (2.62) is given by the Bloch function

$$\psi(\vec{k},\vec{r}) = \psi_n(\vec{k}_0,\vec{r})\, e^{i(\vec{k}-\vec{k}_0).\vec{r}} \qquad 2.63$$

Where $\psi_n(\vec{k}_0,\vec{r})$ is a function of the lattice periodicity for band index *n*.

Substituting the Bloch function into equation (2.62), and using the following relations [22]

$$\nabla\psi(\vec{k},\vec{r}) = i(\vec{k}-\vec{k}_0)\,\psi(\vec{k},\vec{r}) + e^{i(\vec{k}-\vec{k}_0).\vec{r}}\nabla\psi_n(\vec{k}_0,\vec{r}) \qquad 2.64$$

$$\begin{aligned}\nabla^2\psi(\vec{k},\vec{r}) &= -(\vec{k}-\vec{k}_0)^2\,\psi(\vec{k},\vec{r}) + 2i(\vec{k}-\vec{k}_0)e^{i(\vec{k}-\vec{k}_0).\vec{r}}\nabla\psi_n(\vec{k}_0,\vec{r}) + e^{i(\vec{k}-\vec{k}_0).\vec{r}}\,\nabla^2\psi_n(\vec{k}_0,\vec{r}) \\ &= e^{i(\vec{k}-\vec{k}_0).\vec{r}}\left(-(\vec{k}-\vec{k}_0)^2 + 2i(\vec{k}-\vec{k}_0).\nabla + \nabla^2\right)\psi_n(\vec{k}_0,\vec{r}) \qquad 2.65\end{aligned}$$

By using the relation $P = -i\hbar\nabla$ for the momentum operator, equation (2.62) becomes

$$\left[\frac{P^2}{2m_0} + V(\vec{r})\right]\psi_n(\vec{k}_0,\vec{r})\, e^{i(\vec{k}-\vec{k}_0).\vec{r}} = \left[\frac{\left(P + (\vec{k}-\vec{k}_0)\right)^2}{2m_0} + V(\vec{r})\right]\psi_n(\vec{k}_0,\vec{r})$$

$$= e^{i(\vec{k}-\vec{k}_0).\vec{r}}\left[\frac{\hbar^2}{2m_0}(\vec{k}-\vec{k}_0)^2 + \frac{\hbar}{m_0}(\vec{k}-\vec{k}_0).P + E_n(\vec{k}_0)\right]\psi_n(\vec{k}_0,\vec{r}) \qquad 2.66$$

The secular equation is represented by

$$\left\|\left\langle\psi_n(\vec{k}_0,\vec{r})\, e^{i(\vec{k}-\vec{k}_0).\vec{r}}\middle|H-E\middle|\psi_n(\vec{k}_0,\vec{r})\, e^{i(\vec{k}-\vec{k}_0).\vec{r}}\right\rangle\right\| = 0 \qquad 2.67$$

Therefore, the eigenvalue determinant becomes

$$\left\|\left\langle\left[\frac{\hbar^2}{2m_0}(\vec{k}-\vec{k}_0)^2 + E_n(\vec{k}_0) - E\right]\delta_{n'n} + \frac{\hbar}{m_0}(\vec{k}-\vec{k}_0).P_{n'n}(\vec{k}_0)\right\rangle\right\| = 0 \qquad 2.68$$

Where $P_{n'n}$ is the momentum matrix element between the different bandedges states, and is given by

$$P_{n'n} = \int \psi^*_{n'}(\vec{k}_0, \vec{r})\, P\, \psi_n(\vec{k}_0, \vec{r})\, d^3r \qquad 2.69$$

$P_{n'n}$ is non zero only for certain symmetries of $\psi^*_{n'}(\vec{k}_0, \vec{r})$ and $\psi_n(\vec{k}_0, \vec{r})$, hence reducing the number of independent parameters.

The k.p description of the non degenerate bands (i.e., conduction bandedge or the split-off band in the valence band for the case of large spin-orbit coupling) can be done using the perturbation theory to obtain the energy wave functions away from k_0.

For k_0=0, the Schrödinger equation for the perturbation Hamiltonian is given by

$$(H_0 + H_1 + H_2)u_{nk} = E_{nk}\, u_{nk} \qquad 2.70$$

Where

$$H_0 = \frac{P^2}{2m_0} + V(\vec{r}) \qquad 2.71$$

$$H_1 = \frac{\hbar}{m_0}\, \vec{k}.\overrightarrow{P} \qquad 2.72$$

$$H_2 = \frac{\hbar^2 k^2}{2m_0} \qquad 2.73$$

And u_{nk} is the central part of the Bloch functions $\psi_n(\vec{k})$.

In the perturbation approach H_0 is a zero order term in $\vec{k}$, H_1 is a first order term in $\vec{k}$, and H_2 is a second order term in $\vec{k}$.

To zero order, we have

$$u_{nk} = u_{n0} \qquad 2.74$$

$$E_{nk} = E_n(0) \qquad 2.75$$

To first order perturbation we have

$$u_{nk} = u_{n0} + \frac{\hbar}{m_0}\sum_{\mathrm{n'\neq n}} \vec{k}.\frac{\langle n0|P|n0\rangle}{E_n(0) - E_{n'}(0)}u_{n0} \quad 2.76$$

$$E_{nk} = E_n(0) + \frac{\hbar}{m_0}\vec{k}.\langle n0|P|n0\rangle \quad 2.77$$

For crystals with inversion symmetry such as Si and Ge, the states $|n0\rangle$ or u_{n0} will have inversion symmetry, therefore, the first order matrix elements vanish because P has an odd parity. For the second order, the energy is given by

$$E_{nk} = E_n(0) + \frac{\hbar^2 \mathrm{k}^2}{2m_0} + \frac{\hbar^2}{{m_0}^2}\sum_{\mathrm{n'\neq n}} \frac{|\,\vec{k}\langle \mathrm{n'}0|\vec{p}|n0\rangle|^2}{E_n(0) - E_{n'}(0)} \quad 2.78$$

Equation 2.78 can be expressed in terms of the effective mass m^* as follows

$$E_{nk} = E_n(0) + \sum_{i,j}\frac{\hbar^2}{m^*_{i,j}}\, k_i.k_j \quad 2.79$$

Where

$$\frac{m_0}{m^*_{i,j}} = \delta_{ij} + \frac{2}{m_0}\sum_{n'\neq n}\frac{\langle n0|P_i|n'0\rangle\langle n'0|P_j|n0\rangle}{E_n(0) - E_{n'}(0)} \quad 2.80$$

This equation is valid for conduction bandedge and the split-off bands. For the valence band, keeping only the valence bandedge bands in the summation, then the energy eigenvalue can be expressed as

$$E_{nk} = E_n(0) + \frac{\hbar^2 k^2}{2m^*_c} \quad 2.81$$

With

$$\frac{1}{m^*_c} = \frac{1}{m_0} + \frac{2P^2_{cv}}{m_0}\,\frac{1}{3}\left(\frac{2}{E_{g\Gamma}} + \frac{1}{E_{g\Gamma} + \Delta}\right) \quad 2.82$$

Where $E_{g\Gamma}$ is the energy gap at the zone center, and Δ is the HH-SO (Split-Orbit) band separation.

For the Split-off band the energy eigenvalue is given by

$$E_{SO} = -\Delta - \frac{\hbar^2 k^2}{2m_{\mathrm{SO}}^*} \quad 2.83$$

Where

$$\frac{1}{m_{\mathrm{SO}}^*} = -\frac{1}{m_0} + \frac{2P_{cv}^2}{3{m_0}^2(E_{g\Gamma} + \Delta)} \quad 2.84$$

The HH valence band structure is that of a free-electron, therefore, the effective mass is the same as free-electron mass, and is given by

$$E_{HH} = -\frac{\hbar^2 k^2}{2m_0} \quad 2.85$$

and the light-hole band structure is given by

$$E_{HH} = -\frac{\hbar^2 k^2}{2m_{\mathrm{LH}}^*} \quad 2.86$$

Where

$$\frac{1}{m_{\mathrm{LH}}^*} = \frac{1}{m_0} + \frac{4P_{cv}^2}{3{m_0}^2 E_{g\Gamma}} \quad 2.87$$

4. Impact of Strain

4.1 Crystal Symmetry

Due to the communication between symmetry operations and the crystal Hamiltonian, crystal symmetry determines the symmetry of the band structure. Therefore, straining the silicon lattice will reduce the crystal symmetry and change the inter-atomic spacing. The breaking of the crystal symmetry also causes band warping from symmetry restrictions. When the band structure of a material is changed, many material properties are altered including band gap, effective mass, carrier scattering, and mobility. Associated modifications in the electronic band structure and density of states contribute to changes in carrier mobility through modulated effective transport masses [23].

4.2 Band Structure and Band Alignment

The impact of biaxial strain on the band structure of a semiconductor can be discussed in two parts; hydrostatic strain (due to the fractional volume change), and uniaxial strain. The hydrostatic component leads to an energy band shift, and a change of the bandgap, while the uniaxial strain component splits the degeneracy of the conduction and valence bands but it has no effects on the average band energy. Under biaxial compressive strain, the six-fold degenerate Si conduction band energy (Δ_6) is splitted. The [001] conduction bands (two-fold degenerate Δ_2 bands) move up in energy, while the [100] and [010] conduction bands (four-fold degenerate Δ_4 bands) move down in energy. On the contrary, under biaxial tensile strain, the [001] conduction bands (two-fold degenerate Δ_2 bands) move down in energy, while the [100] and [010] conduction bands (four-fold degenerate Δ_4 bands) move up in energy. This splitting of degeneracy in the conduction band reduces the conductivity effective mass and suppresses intervalley scattering, hence enhancing the transport properties. The impact of compressive and tensile strains on the conduction band energy is illustrated in Fig.31.

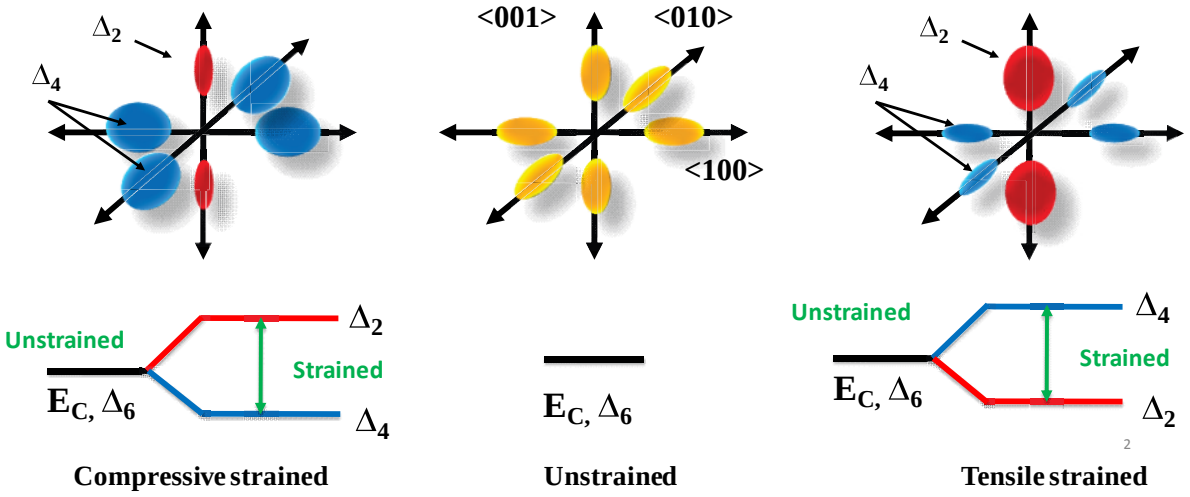

Fig.31 : Schematic representation of the effects of hydrostatic and uniaxial stress on the conduction band energy in Si for tensile and compressive strains.

Regarding the valence band, the biaxial tensile strain splits the top of the valence band, with the heavy hole (HH) moving down in energy and the light hole (LH) being raised in energy. In contrast, biaxial compressive strain splits the top of the valence band, with the light hole (LH)

moving down in energy and the heavy hole (HH) being raised in energy, thereby reducing both intersubband and intrasubband scattering. The impact of compressive and tensile strains on the valence band energy is illustrated in Fig.32.

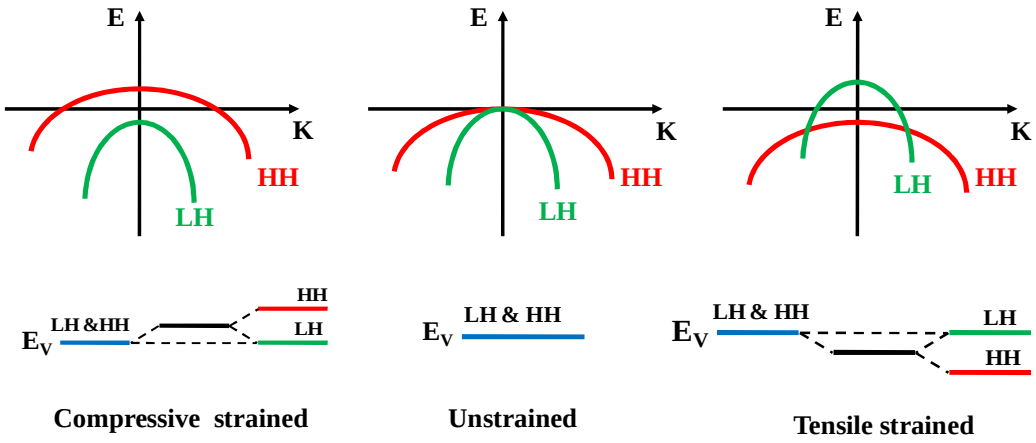

Fig.32: Schematic representation of the effects of hydrostatic and uniaxial stress on the valence band energy in Si for tensile and compressive strains.

The k.p method incorporated with Bir-Pikus strain Hamiltonian is used to calculate strain effect on band structures by introducing an additional perturbation term into the unstrained potential [24]. Therefore, the total Hamiltonian is given by

$$H = H_{k.p} + H_{strain} \tag{2.88}$$

Where

$$H_{k.p} = \frac{\hbar^2(k_l - k_0)^2}{2m_l^*} + \frac{\hbar^2 k_t^2}{2m_t^*} \tag{2.89}$$

and

$$H_{strain} = \Xi_d \, Tr(\varepsilon_{ij}) + \Xi_u \varepsilon_l \tag{2.90}$$

Where Ξ_d and Ξ_u are the dilation and uniaxial deformation potentials at the Si conduction bandedge required for symmetry considerations, $Tr(\varepsilon_{ij})$ is the trace strain tensor, $\varepsilon_l(\varepsilon_t)$ is the longitudinal (transverse) strain component (along [001], $\varepsilon_l = \varepsilon_{zz}$, and $\varepsilon_t = \varepsilon_{xx} + \varepsilon_{yy}$), and m_l^* (m_t^*) is the longitudinal (transverse) effective mass.

The general form of the strain-induced energy change in the energy of carrier bands in silicon is given by

$$\Delta E_{C,i} = \Xi_d(\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}) + \Xi_u \varepsilon_{ii} \quad 2.91$$

$$\Delta E_{V,i} = -a(\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}) \pm \delta E \quad 2.92$$

$$\delta E = \sqrt{\frac{b^2}{2}((\varepsilon_{11} - \varepsilon_{22})^2 + (\varepsilon_{22} - \varepsilon_{33})^2 + (\varepsilon_{11} - \varepsilon_{33})^2) + d^2(\varepsilon_{12}^2 + \varepsilon_{13}^2 + \varepsilon_{23}^2)} \quad 2.93$$

Where, *a*, *b*, and *d* are deformation potentials that correspond to the model, *i* corresponds to the carrier band number, and ε_{ij} are the components of the strain tensor in the crystal coordinate system. The final value of the change in the energy band can be calculated by averaging the energy changes in all the sub-bands. The expression for the change in energy can be summarized as:

$$\frac{\Delta E_C}{kT_{300}} = -ln\left[\frac{1}{n_C}\sum_{i=1}^{n_C} exp\left(\frac{-\Delta E_{C,i}}{kT_{300}}\right)\right] \quad 2.94$$

$$\frac{\Delta E_V}{kT_{300}} = ln\left[\frac{1}{n_V}\sum_{i=1}^{n_V} exp\left(\frac{\Delta E_{V,i}}{kT_{300}}\right)\right] \quad 2.95$$

where n_C and n_V are the number of subvalleys considered in the conduction and valence bands, respectively, and T_{300}=300K [25].

4.3 Mobility Enhancement

To understand the effect of strain on mobility, the simple qualitative Drude model of electrical conduction which explains the transport properties of electrons in materials dictates that

$$\mu = \frac{q\tau}{m^*} \qquad 2.96$$

Where μ is the carrier mobility, τ is the scattering time, and m^* is the conductivity effective mass. Therefore, the mobility improvement in strained silicon takes place mainly due to the reduction of the carrier conductivity effective mass, and the reduction in the intervalley phonon scattering rates.

The conduction band of unstrained bulk silicon has six equivalent valleys along the <100> direction of the Brilloun zone, and the constant energy surface is ellipsoidal with the transverse effective mass, $m_t = 0.19m_0$, and the longitudinal effective mass, $m_l = 0.916m_0$, where m_0 is the free electron mass [26]. If biaxial tensile strain is applied, the degeneracy between the four in-plane valleys (Δ_4) and the two out-of-plane valleys (Δ_2) is broken as shown in Fig.31. As a consequence, the electrons prefer to populate the lower valleys, which are energetically favored. This results in an increased electron's mobility via a reduced in-plane and increased out-of-plane electron conductivity mass. In addition to that, electron scattering is also reduced due to the conduction valleys splitting into two sets of energy levels, which lowers the rate of intervalley phonon scattering. Therefore, if the optimum strain is applied, both reductions in scattering rate and in effective mass will contribute to the electron mobility enhancement. The stress-induced electron mobility enhancement is given by

$$\mu_{ii}^{n} = \mu_n^0 \left[1 + \frac{1 - \frac{m_{nl}}{m_{nt}}}{1 + 2\left(\frac{m_{nl}}{m_{nt}}\right)} \left(\frac{F_n \frac{(F_n - E_C - \Delta E_{C,i})}{kT}}{F_n \frac{(F_n - E_C - \Delta E_C)}{kT}} - 1 \right) \right] \qquad 2.97$$

Where μ_n^0 is electron mobility without the strain, m_{nl} and m_{nt} are the electron longitudinal and transverse masses in the subvalley, respectively, E_C and $\Delta E_{C,i}$ are the change in the energy of the unstrained and the strained carrier sub-valleys, F_n is the quasi-Fermi level of electrons. The index *i* corresponds to a direction (for example, μ_{11}^{n} is the electron mobility in the direction of the x-axis of the crystal system and, therefore, $\Delta E_{C,1}$should correspond to the two-fold subvalley along the x-axis) [27].

For holes, the valence band structure of silicon is more complex than the conduction band. For unstrained silicon at room temperature, holes occupy the top two bands: the heavy and light hole bands. Applying strain, the hole effective mass becomes highly anisotropic due to band warping, and the energy levels become mixtures of the pure heavy, light, and split-off bands.

Thus, the light and heavy hole bands lose their meaning, and holes increasingly occupy the top band at higher strain due to the energy splitting. To achieve high hole's mobility, a low in-plane conductivity mass for the top band is required, in addition to that, a high density of states in the top band and a sufficient band splitting to populate the top band are also required [25].

5. Strain Application Techniques

In the previous section it has been shown that the introduction of a compressive and/or tensile strain in the Si substrate can improve the mobility of both carrier types. Therefore, this provides a very important way to modify and enhance the electrical properties of Si through proper design, implementation, and control of strain in the active layers. Consequently, various methods and approaches have been proposed to induce the desired strain in electronic devices, such as "Global strain" through SiGe epitaxial processes [28] [29], "Local strain" using specially engineered high tensile films [29] and "Mechanical strain" by mechanically bending the wafer post fabrication[30] [31]. The different strain generation methods will be discussed in details.

5.1 Global Strain Approach

Global strain on wafer level is mostly induced by the epitaxial growth of $Si_{1-x}Ge_x$ and Si layers. Because the lattice parameter of $Si_{1-x}Ge_x$ ($0 \leq x \leq 1$) alloys varies between 0.5431 A^0 for Si (x=0) and 0.5657 A^0 for Ge (x=1), tensile strain is induced in a silicon layer epitaxially grown on top of the SiGe layer. And compressive strain is induced in the SiGe layer epitaxially grown on top of a Si layer as shown in Fig.33. In this technology the degree of strain is controlled by changing the content of Ge in the $Si_{1-x}Ge_x$ layer, or by changing the thickness of the strained Si layer. In both cases the strain is in the plane of the layer ($\varepsilon_x = \varepsilon_x = \varepsilon_{||}$), but this strain also produces a perpendicular strain, $\varepsilon_\perp$ resulting in a tetragonal distortion of the lattice.

The strains are connected through the isotropic elasticity theory as

$$\varepsilon_\perp = \frac{-2v}{1-v}\varepsilon_{||} \qquad 2.98$$

Where *v* is the Poisson's ratio.

The tetragonal distortion produced by the perpendicular strain $\varepsilon_{\perp}$ results in a parallel lattice constant, and is given by

$$a_{||} = a_{Si}\left[1 + \frac{f}{\left(1 + \frac{G_{Si}h_{Si}}{G_{SiGe}h_{SiGe}}\right)}\right] \qquad 2.99$$

Where a_{Si} is the Si lattice constant, f is the misfit between the two layers, h_{Si} is the Si layer thickness, h_{SiGe} is the SiGe layer thickness and G_{Si}, G_{SiGe} are the shear moduli of Si and SiGe respectively.

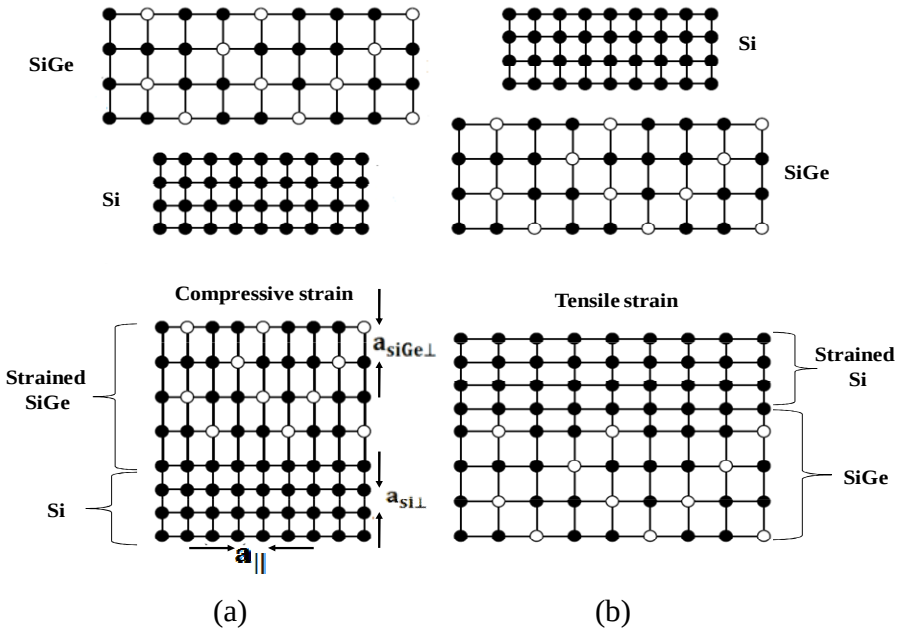

Fig.33: (a) A schematic diagram of the bulk lattice of a thin $Si_{1-x}Ge_x$ film to be grown on top of a thin bulk silicon layer with the top $Si_{1-x}Ge_x$ film being compressively strained. (b) A schematic diagram of the bulk lattice of Si film to be grown on top of a bulk $Si_{1-x}Ge_x$ film with the top Si film being tensile strained.

The misfit f between the two layers is defined as

$$f = \frac{a_{SiGe} - a_{Si}}{a_{Si}} = 4.17\% \qquad 2.100$$

In equilibrium, the in-plane strain in the Si layer and SiGe layer are related together by the relation

$$\varepsilon_{||}^{\mathrm{Si}} = -\left(\frac{G_{siGe}h_{siGe}}{G_{si}h_{si}}\right)\varepsilon_{||}^{\mathrm{SiGe}} \qquad 2.101$$

Under appropriate growth conditions, good quality layers of crystalline $Si_{1-x}Ge_x$ alloys on Si substrates can be grown. If the SiGe thickness remains below a critical thickness (h_C), which depends on the alloy composition and the growth temperature, a pseudomorphic $Si_{1-x}Ge_x$ film can be grown without the introduction of extended defects. If the $Si_{1-x}Ge_x$ thickness exceeds the critical thickness, or the substrate is exposed to sufficiently high temperatures for long period of time, at which the pseudomorphically grown layer is no longer thermodynamically stable, the lattice constant relaxes to its original value. This means that the strain in the $Si_{1-x}Ge_x$ layer will be relaxed and misfit dislocations will generate at the Si/ $Si_{1-x}Ge_x$ interface. Thus, the $Si_{1-x}Ge_x$/Si strained heterostructures are limited in thickness and stability.

Various models have been developed to predict the critical thickness for which the epitaxial strain layer can be grown. Van der Merwe produced a thermodynamic equilibrium model by minimizing the total energy of a system with the generation of a periodic array of dislocations. In his model the critical thickness is when the strain energy equals the interface energy, and is given by

$$h_c \approx \frac{19}{16\pi^2}\left(\frac{1+v}{1-v}\right)\left(\frac{b}{f}\right) \qquad 2.102$$

Where b is the magnitude of the Burger's vector. For a bulk Si substrate b=0.4 nm, and in general $b = a/\sqrt{2}$, where a is the lattice constant of the relaxed substrate.

Matthews and Blackeslee have proposed in their model that the critical thickness is when the misfit stress on an existing threading dislocation equals the line tension of the dislocation, or equivalently, when a dislocation half-loop is stable against the misfit stress. The critical thickness according to Matthews and Blackeslee model is given by

$$h_c = \frac{1}{8\pi}\frac{1}{\varepsilon_{||}}\left(\frac{1-v\,cos^2\theta}{(1+v)\,cos\lambda}\right)ln\left(\frac{4h_c}{b}\right) \qquad 2.103$$

Where θ is the angle between the dislocation line and the Burgers vector, λ is the angle between the Burgers vector and the direction in the interface normal to the dislocation line, v is the Poisson's ratio, and $\varepsilon_{||}$ is the in-plane strain.

However, it was verified that Van der Merwe model and Matthews and Blackeslee model calculations were not consistent with the experimental data of the critical thickness, and if epitaxy conditions are carefully controlled, then a $Si_{1-x}Ge_x$ layer with thickness above (h_C) could be grown. The simplicity of these models, as well, not taking in consideration the nucleation, propagation, and interaction of dislocations in their calculation were the reasons for the models failure. Afterwards, more accurate results were proposed by People and Bean. In their model they tried to explain the metastable critical thickness ($h_{c,MS}$) through a nonequilibrium approach. According to their model, the metastable critical thickness is defined as the film thickness at which its strain energy density becomes greater than the self-energy of an isolated screw dislocation, and is given by

$$h_{c,MS} = \frac{b^2}{16\sqrt{2}} \frac{1}{a_{SiGe}\,\varepsilon_{||}^2}\left(\frac{1-v}{1+v}\right) ln\left(\frac{h_{c,MS}}{b}\right) \qquad 2.104$$

In Fig.34 the equilibrium (stable) and metastable values of critical thicknesses are plotted versus the Ge content, *x*, of a $Si_{1-x}Ge_x$ epitaxial layer grown on a Si substrate. As shown in the figure, increasing the Ge content will increase the strain in the SiGe layer, and thus the critical thickness decreases.

Even though the global strain approach described above has the advantage that it is wafer-level and the transistor fabrication process requires little or no change, it suffers from several process integration issues. The presence of Ge modifies dopants diffusion and changes thermal conductivity of the substrate. The relaxation of SiGe via misfit dislocation formation and thermal processing during the fabrication steps causes degradation of the device performance. Moreover, the growth of a thick SiGe strain-relaxed buffer can be costly [32].

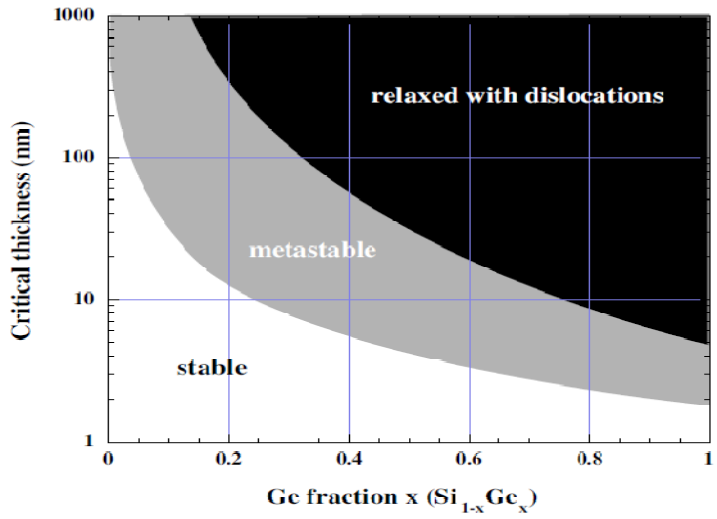

Fig.34 : The equilibrium and metastable critical thickness versus Ge content for pseudomorphic $Si_{1-x}Gex$ layers grown on bulk silicon substrate.

5.2 Local Strain Approach

A second technique for introducing strain in semiconductor devices is the use of a tensile and/or compressive strain layer. In this approach, either uniaxial or biaxial strain is created through the device fabrication process using strain layers such as silicon dioxide (SiO_2), and silicon nitride (Si_3N_4). In this technique, strain develops primarily during the deposition process and consists of two components: the intrinsic strain and the extrinsic strain. The intrinsic strain is the component of strain in the layer caused by the deposition process itself. Processing conditions such as temperature, thickness, pressure, deposition power, reactant and impurity concentrations, are important factors in determining the magnitude and strain type (i.e. compressive or tensile).

The extrinsic strain is the component of strain caused by a change in the external conditions on the wafer. The thermal expansion coefficient of materials like SiO_2, and Si_3N_4 layers are different from the silicon substrate thermal coefficient. Therefore, when the temperature changes, the layer and substrate try to expand or contract by different amounts. Because the substrate and the stress layer are bound together, a stress will develop in both the layer and the substrate. Since layers are typically deposited above room temperature, the process of cooling after deposition will introduce a thermal component of strain. So, after deposition, the film tends to back to its initial state by shrinking if it was stretched earlier, thus creating

compressive intrinsic stress, and similarly tensile intrinsic stress if it was compressed during deposition.

The thermal expansion coefficient, α_T, is defined as the rate of change of strain with temperature, and is given by

$$\alpha_T = \frac{d\varepsilon}{dT} \qquad 2.105$$

Therefore, the thermal strain, ε, induced by a variation in temperature is given by

$$\varepsilon = \alpha_T \, \Delta T \qquad 2.106$$

The intrinsic stress generated can be quantified by Stoney's equation by relating the stress to the substrate curvature as

$$\sigma_f = \frac{E_{Si}}{6(1-v_{Si})R} \frac{h_{Si}}{h_f} \qquad 2.107$$

where E_{Si} and υ_{Si} are Young's modulus and Poisson's ratio of Silicon, h_{Si} and h_f are substrate and film thickness, and R is the radius of curvature of the substrate [33] [34].

The local strain approach through using tensile and/or compressive strain nitride layer has been used to optimize NMOSFET and PMOSFET devices on the same wafer independently by applying different levels of strain as shown in Fig.35 [35]. More than 2 GPa of tensile stress and more than 2.5 GPa of compressive stress have been developed through controlling the growth conditions of Si_3N_4 layers [36].

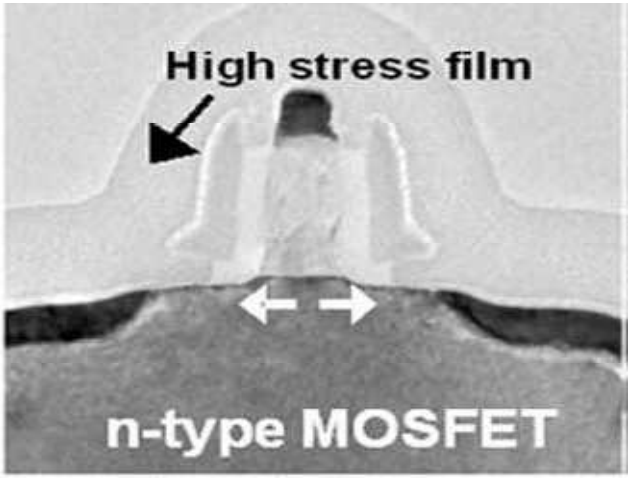

Fig.35 : TEM micrographs of 45-nm n-type MOSFET with nitride-capping film with a large tensile stress [35].

As well, IBM, AMD and Fujitsu [37] [38] have reported a CMOS architecture in which longitudinal uniaxial tensile and compressive stress in the Si channel have been created. In this approach, the process flow consist of a uniform deposition of a highly tensile Si_3N_4 liner post silicidation over the entire wafer, followed by patterning and etching the film over the p-channel transistors. Next, a highly compressive Si_3N_4 layer is deposited, and this film is patterned and etched from n-channel regions. The advantages of this technique over the epitaxial SiGe technique, that the Dual Stress Liner (DSL) approach reduces the process complexity and integration issues. In addition, it simultaneously improves both n- and p-channel transistors. The local strain approach through using SiO_2 and Si_3N_4 strain layers in the collector region will be presented in chapter 4.

5.3 Mechanical Strain Approach

The third technique of introducing strain into the transistors is through external mechanical stress post fabrication. In this approach, the strain is engendered into the Si either through direct mechanical bending of the Si wafer, or by bending a packaged substrate with a Si chip glued firmly onto its surface. One method used to apply external mechanical strain on the Si wafer is shown in Fig.36 [30]-[31]. This technique is an extremely low-cost technique, and it allows the reversible application of either compressive or tensile strain. However, this technique is interesting for experimental study but cannot be used for practical applications.

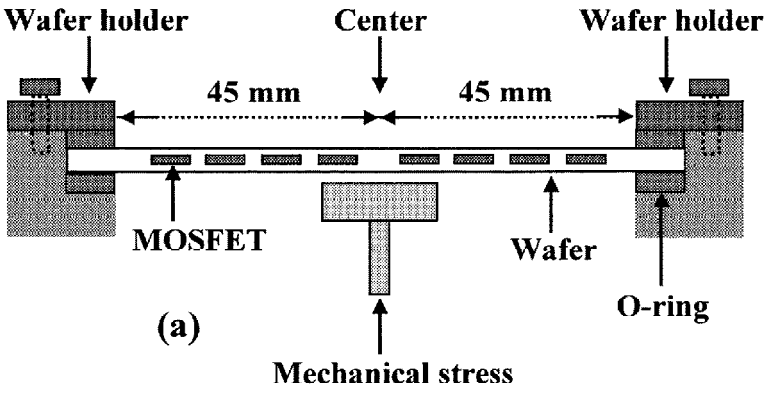

Fig.36 : Schematic diagram of the externally applied mechanical stress on the Si (100) wafer.

6. Conclusion

Strained-Si technology enables improvements in electronic devices performance and functionality via replacement of the bulk crystalline-Si substrate with a strained-Si substrate. Mechanical strain reduces the crystal symmetry and changes the inter-atomic spacing. In addition to that, mechanical strain causes band warping from symmetry restrictions, and induces a change in the band structure. When the band structure of a material is changed, many material properties are altered including band gap, effective mass, carrier scattering, and mobility. Strain can be generated as “global strain” on the wafer level by the growth of SiGe and Si layers, or as “local strain” on the transistor through the device fabrication process by using strain layers or as “mechanical strain” by mechanically bending the wafer post fabrication. Therefore, a profound knowledge and deep understanding of strain physics and strain application techniques are required to achieve the maximum benefit of applying strain technology on standard bipolar devices.

References

[1] J. E. Marsden, T. J. R. Hughes, *Mathematical Foundations of Elasticity*. Dover Publications, Inc., New York., 1983.

[2] M. Filonenko-Borodich, *Theory of elasticity*. Peace Publisher, 1963.

[3] M. H. Sadd, *Elasticity: Theory, Applications, and Numerics*. Elsevier Inc., 2009.

[4] Y.C. Fung, *First Course in Continuum Mechanics (3rd Edition)*. Prentice Hall, 1993.

[5] G. T. Mase and G. E. Mase, *Continuum Mechanics for Engineers, Second Edition*. CRC Press, 1999.

[6] G. Dieter, *Mechanical Metallurgy*. McGraw-Hill Companies, 1986.

[7] J. E. Ayers, *Heteroepitaxy of Semiconductors*. CRC Press, 2007.

[8] P. Howell, J. Ockendon, G. Kozyreff, *Cambridge Texts in Applied Mathematics: Applied Solid Mechanics*. Cambridge University Press, 2009.

[9] S. Adachi, *Properties of Group-IV, III-V and II-VI Semiconductors*. Wiley, 2005.

[10] W.C. O'Mara, R.B. Herring, L.P. Hunt, *Handbook of Semiconductor Silicon Technology*. William Andrew Publishing, 1990.

[11] C. Kittel, *Introduction to Solid State Physics*. Wiley, 1995.

[12] S. M. Sze, K K. Ng, *Physics of Semiconductor Devices*. Wiley-Interscience; 3rd edition, 2006.

[13] F. L. Litvin and A. Fuentes, *Gear Geometry and Applied Theory*. Cambridge University Press, 2004.

[14] R. Hull, *Properties of crystalline silicon*. Institution of Engineering and Technology, 1999.

[15] Y. Sun, S. E. Thompson, T. Nishida, *Strain Effect In Semiconductors: Theory And Device Applications*. Springer, 2009.

[16] J. R. Chelikowsky, and M. L. Cohen, "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors" vol. 14, pp. 556-582, 1976.

[17] C. Junhao, A. Sher, *Physics and Properties of Narrow Gap Semiconductors*. Springer, 2007.

[18] C. Herring, "A New Method for Calculating Wave Functions in Crystals" vol. 57, pp. 1169-1177, 1940.

[19] J. Singh, *Electronic and optoelectronic properties of semiconductor structures*. Cambridge University Press, 2003.

[20] L. C. Lew Yan Voon, M. Willatzen, *The k p Method: Electronic Properties of Semiconductors*. Springer, 2009.

[21] M. Cardona, F. H. Pollak, "Energy-Band Structure of Germanium and Silicon: The k·p Method" vol. 142, pp. 530-543, 1966.

[22] J. R. Barber, *Elasticity Second Edition, Solid Mechanics and Its Applications*. Springer-Verlag New York, LLC, 2002.

[23] Y. Sun, S. E. Thompson,T. Nishida, "Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors" vol. 101, p. 104503, 2007.

[24] G. L. Bir, G. E. Pikus, *Symmetry and Strain-induced Effects in Semiconductors*. Wiley, New York, 1974.

[25] T. Manku, A. Nathan, "Valence energy band structure for strained group IV semiconductors" vol. 73, p. 1205, 1993.

[26] S. Takagi, J. L. Hoyt, J. J. Welser, J. F. Gibbons, "Comparative study of phonon-limited mobility of two dimensional electrons in strained and unstrained Si metal oxide semiconductor field effect transistors" vol. 80, p. 1567, 1996.

[27] J.L. Egley, D. Chidambarrao, "Strain effects on device characteristics: Implementation in drift-diffusion simulators" vol. 36, pp. 1653-1664, 1993.

[28] N. D. Jankovic, A. O'Neill, "Enhanced performance virtual substrate heterojunction bipolar transistor using strained-Si/SiGe emitter" vol. 18, p. 901, 2003.

[29] R. Arghavani, L. Xia, M. M'Saad, M. Balseanu, G. Karunasiri, A. Mascarenhas, S.E. Thompson, "A reliable and manufacturable method to induce a stress of >1 GPa on a P-channel MOSFET in high volume manufacturing" vol. 27, p. 114, 2006.

[30] F. Yuan, S.R. Jan, S. Maikap, Y.H. Liu, C.S. Liang, C.W. Liu, "Mechanically strained Si-SiGe HBTs" vol. 25, p. 483, 2004.

[31] W.Tzu-Juei, C. Hung-Wei, Y. Ping-Chun, K. Chih-Hsin, C. Shoou-Jinn, Y. John, W. San-Lein, L. Chwan-Ying, L. Wen-Chin, T. Denny D, "Effects of mechanical uniaxial stress on SiGe HBT characteristics" vol. 154, p. H105, 2007.

[32] D. J. Paul, "Si/SiGe heterostructures: from material and physics to devices and circuits" vol. 19, p. R75, 2004.

[33] M. Fang, D. Hu, and J. Shao, "Evolution of stress in evaporated silicon dioxide thin films" vol. 8, p. 119, 2010.

[34] A. Szekeres and P. Danesh, “Mechanical stress in SiO2/Si structures formed by thermal oxidation of amorphous and crystalline silicon” vol. 11, p. 1225, 1996.

[35] S.E. Thompson, M. Armstrong, C. Auth, S. Cea, R. Chau, et al., “A Logic Nanotechnology Featuring Strained-Silicon” vol. 25, p. 191, 2004.

[36] G. A. Armstrong, C. K. Maiti, *TCAD for Si, SiGe and GaAs Integrated Circuits (Iet Circuits, Devices and Systems)*. IET, 2007.

[37] S. Pidin, T. Mori, K. Inoue, S. Fukuta, N. Itoh, E. Mutoh, et al., “A novel strain enhanced CMOS architecture using selectively deposited high tensile and high compressive silicon nitride films” p. 213, 2004.

[38] H. S. Yang, R. Malik, S. Narasimha, Y. Li, R. Divakaruni, P. Agnello, et al., “Dual stress liner for high performance sub-45 nm gate length SOI CMOS manufacturing” p. 1075, 2004.

CHAPTER 3

TCAD Simulation & Modeling

1. Introduction

Technology Computer-Aided Design (TCAD) refers to the use of computer simulations to develop and optimize semiconductor processing technologies and devices. TCAD simulation and modeling can be used to predict the device performance and expedite the device development/optimization process for new technology. As well, it greatly enhances the learning process through providing a remarkable physical insight into what real integrated circuits structures look like. The Sentaurus TCAD platform provides a comprehensive capability to simulate detailed and realistic process structures for subsequent electrical analysis by Sentaurus Device and it provides tools for interconnect modeling and extraction that supply critical parasitic information for optimizing chip performance [1].

TCAD modeling is used to design, analyze, and optimize semiconductor technologies and devices with fundamental and accurate models. The use of TCAD in semiconductor manufacturing and development of new technologies is two-fold: Process simulation and Device simulation. Process simulation models the complex flow of semiconductor fabrication steps and ends up with detailed information on geometric shape and doping profile distribution of a semiconductor device. The device simulation uses the information of the process simulation as the input file to calculate the characteristics of semiconductor devices and parameters extraction. This provides a useful way of studying the effects of process parameters on the device performance and both the device structure and the fabrication process can thus be optimized.

However, TCAD process and device simulation tools play a critical role in advanced technology development by giving insight into the relationships between processing choices and nanoscale device performance that cannot be obtained from measurements tools alone. The TCAD process and device simulation tools support a broad range of applications such as CMOS, power, memory, image sensors, solar cells, bipolar, and analog/RF devices. A complete TCAD simulation involves the following steps:

- Virtual fabrication of the device using a process simulator or a device editor.
- Creation of a mesh suitable for device simulation.
- Device simulation that solves the equations describing the device behavior.
- Post processing i.e, generation of figures and plots.

The conventional role of TCAD simulation and modeling in integrated circuit devices processing is illustrated in Fig.37 [1][2]. In what follows, process simulation and device simulation will be discussed in more details.

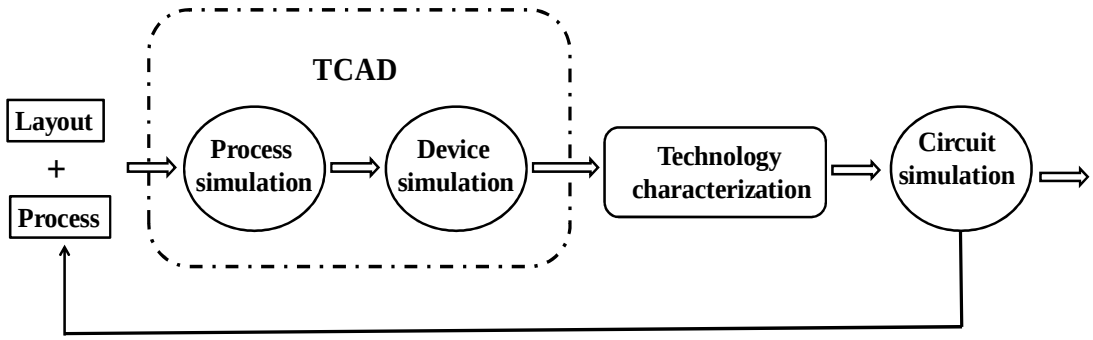

Fig.37: The conventional role of TCAD simulation and modeling in integrated circuit devices processing.

2. Process Simulation

The behavior and properties of all semiconductor devices are defined by their three geometrical dimensions and concentration profile of impurities. The main goal of process simulation is to model a virtual device with geometry and properties identical with the real structure. Sentaurus TCAD process is an advanced, complete, and highly flexible multi-dimensional process simulator for developing and optimizing silicon and compound semiconductor process technologies. It offers unique predictive capabilities for modern silicon and non-silicon technologies, when properly calibrated to a wide range of the latest experimental

data using proven calibration methodology. It uses powerful numerical algorithms that simulate process steps like implantation, diffusion and dopant activation, etching, deposition, oxidation, and epitaxial growth in different semiconductor materials with efficient meshing for robust and stable simulation. The fabrication of integrated circuit devices requires a series of processing steps called a process flow. A process flow is simulated by issuing a sequence of commands that corresponds to the individual process steps. Also process conditions like the ambient chemical composition, temperature, pressure, etc. during individual fabrication steps can be controlled. In addition, several control commands allow selecting physical models and parameters, grid strategies, and graphical output preferences. The final output is a 2D or 3D device structure which can be used for device simulation. The major processing steps involved in the manufacturing of integrated circuit devices are shown in Fig.38. Each of these steps contain numerous possible variations in process controls, and may take several weeks to complete [1][3].

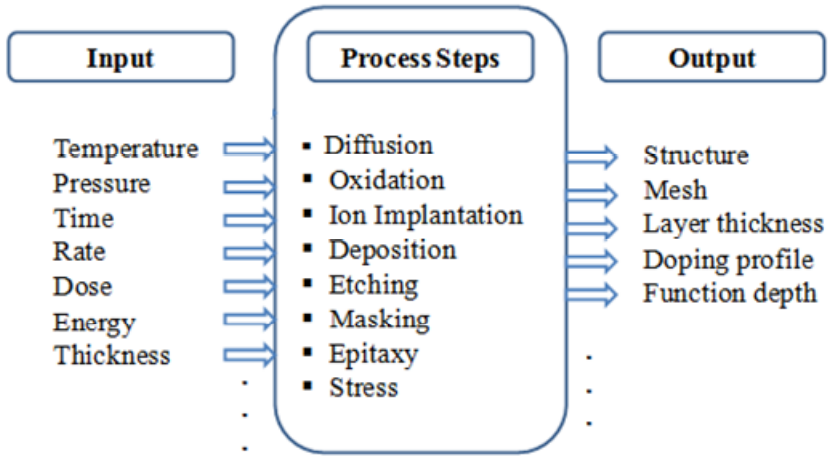

Fig.38: Major process steps involved in the manufacturing of integrated circuit devices.

The TCAD modeling input commands of individual steps make accessible all parameters which characterize the real fabrication processes. In this work TCAD process simulation tools have been used to build the device structures and to calculate the associated mechanical strain generated inside the device due to applying strain engineering technology (i.e. “Global” and “Local” strain techniques), using elastic anisotropic model. The major processing steps in the fabrication process of BJT/HBT devices used in this study will be described, namely deposition, etching, diffusion, oxidation, and mechanical stress computation.

2.1 Deposition

Deposition is the process of growing different layers (insulators, metals, poly Si). The deposition may be isotropic, anisotropic, polygonal (deposition according to a user-supplied polygon), and fill step (fill of the structure with a specific material up to a specific coordinate). The thickness of a deposited layer is defined by the mask, the growth rate, and the deposition time. As the simulated region (volume) is changed, remeshing of the analyzed structure is required. The model parameters and local deposition rates can be specified to construct a geometry which is similar to the one observed in reality. In addition, fields such as stress, pressure, and dopants can be initialized in the deposited layers. With isotropic deposition process, it is possible to specify piecewise linear field values as a function of deposited depth.

2.2 Etching

Etching is the process of removing a material which is in contact with gas. The etching process may be isotropic, anisotropic, directional, Polygonal (Etch according to a user-supplied polygon), Fourier (Angle-dependent etching where etch rate is a cosine expansion of the etching angle), Crystallographic (Angle-dependent etching where etch rate is dependent on the crystallographic direction), and Chemical-mechanical planarization. In general, etching technology consists of both dry and wet etch methods. Dry etching methods include plasma etching, reactive ion etching, sputtering, ion beam etching, and reactive ion beam etching, while wet etching is liquid chemical etching. In Sentaurus TCAD simulator a set of geometry operations is provided which allows defining local etching rates that can be used to approximate the modifications of the structure. During the etching process the thickness of etched layer is defined by the mask, etching rate and etching time. Etch stop and selected material removal can be defined in TCAD simulator.

Sentaurus TCAD MGOALS library can be used to perform etching and deposition operations both in 2D and 3D. The MGOALS library operation can be summarized as follows:

- Analyzing the starting structure for the interfaces that will be changed during the operation.
- Performing the geometry-changing operations.

- Remeshing the entire structure, so that nodes in the silicon region are retained as much as possible in their original locations to minimize the interpolation errors.

2.3 Diffusion

Dopant redistribution is caused by dopant and point defect diffusion as a result of chemical reactions at the interfaces and inside the layers, convective dopant transport due to internal electrical fields and material flow, and moving material interfaces when the substrate is under high temperature process. Diffusion is a high temperature process of diffusion of impurities due to the existing concentration gradient, which depends on temperature, time of diffusion, and boundary conditions characterizing the interface concentration of diffusion species at the Si substrate and gas interface. Sentaurus TCAD simulator includes various physical models with different levels of complexity depending on the type of impurity, point defects and electric field effects. For example, the simplest constant diffusion model neglects the interactions between the dopants and point defects. The pair diffusion model assumes that the gradient of dopant concentration and dopant-defect pairs with the electric field are the driving force of diffusion in active Si regions predefined by the mask. To control the dopant diffusion through various annealing cycles in the fabrication process, the temperature budget should be minimized to ensure very steep and shallow doping profiles for miniaturized structures and devices.

2.4 Oxidation

Oxidation is the process of growth of thermal silicon dioxide (SiO_2) at the silicon surface. This process depends on temperature, time, and oxidation ambient which characterize the diffusion of oxidants from the gas-oxide interface to Si-SiO_2 interface, and its reaction with Si. The ramping up and down temperature cycles with slow temperature changes procedure are used to prevent structure damage due to the induced mechanical stresses and materials motion in the structure caused by the thermal oxidation process, as a result of the thermal expansion coefficients mismatch between SiO_2 layer and silicon substrate, and the growth of oxide on top of silicon substrate. The oxidation process has three steps:

- Diffusion of oxidants (H_2O, O_2) from the gas-oxide interface through the existing oxide to the silicon-oxide interface.
- Reaction of the oxidant with silicon to form new oxide.
- Motion of materials due to the volume expansion, which is caused by the reaction between silicon and oxide.

Various oxidation models are implemented in TCAD simulator. They differ in complexity and coupling of the physical models involved. The oxidant diffusion equation is solved using the generic partial differential equation solver of Sentaurus Process. The model we use in this work is the default model which allows all materials to be simulated as nonlinear viscoelastics.

2.5 Mechanical Stress Computation

Mechanical stress plays an important role in the process simulation and modeling. Mechanical stress modifies the bandgap energy, the carrier's mobility, and can affect the oxidation rates, which can alter the shapes of thermally grown oxide layers. Therefore, accurate computation of mechanical stress is important especially with the continuous trend towards designing process flows that produce the desired kinds of stress in the device to enhance the device performance. In Sentaurus TCAD simulator, stress computation simulations are performed in four distinct steps:

- Define the mechanics equations; the equations used in Sentaurus Process are equations that define force equilibrium in the quasistatic regime.
- Define the boundary conditions for the mechanics equations.
- Define the material properties; in this part the relationship between stresses and strains is defined.
- Define the mechanisms that drive the stresses; this is performed through intrinsic stresses, thermal mismatch, material growth and lattice mismatch, and densification.

Stress is solved in all materials, and the parameters describing materials behavior are included in the simulator's parameters database. In this work, the elastic anisotropy model has been used to calculate the associated mechanical stress with the fabrication process [3][4].

3. Device Simulation

Device simulation and modeling simulates numerically the electrical behavior of semiconductor structures by solving coupled, non-linear, partial differential equations that describe the physics of the device performance. The main purposes of the device simulation are to understand and to describe the physical processes in the interior of a device, and to make reliable predictions of the behavior of the next-generation devices. The electrical characterization includes static, time-dependent, large and small-signal frequency-dependent, electrical behavior, and parameter extraction of the studied structures. The quality of the physical models and the calibrated parameters used in the device simulation is very important for understanding the physical mechanisms in semiconductor devices, and for reliable prediction of the device characteristics. The input device structure typically comes from process simulation steps using Sentaurus process, or through TCAD operations and process emulation steps with the aid of tools like Sentaurus Structure Editor (SSE), or Mesh and Sentaurus Structure Editor. In device simulation, a real semiconductor device, such as a transistor, is approximated by a virtual device as a 2D or 3D structure whose physical properties are discretized onto a nonuniform mesh of nodes. The geometry (grid) of the device contains a description of the various regions, that is, boundaries, material types, the locations of any electrical contacts, and the locations of all the discrete nodes and their connectivity. The data fields contain the properties of the device, such as the doping profiles, in the form of data associated with the discrete nodes. By default, a device simulated in 2D is assumed to have a thickness in the third dimension of 1 μm [5].

Sentaurus TCAD device simulation allows for arbitrary combinations of transport equations and physical models, which allows for the possibility to simulate all spectrums of semiconductor devices, from power devices to deep submicron devices and sophisticated heterostructures [1][4]. In the following, the formulation of physical models and equations used in our structures device simulations will be described.

3.1 Basic Semiconductor Equations

The fundamental semiconductor equations that rule the semiconductor devices are Poisson's equation that solves the relationship between electrostatic potential and charge density,

and charge carrier continuity equation that reflects the fact that sources and sinks of the conduction current are fully compensated by the time variation of the mobile charge [6][7]. The Poisson equation is given by

$$\nabla.\epsilon\nabla\psi = -\rho \quad 3.1$$

Where ϵ where is the electrical permittivity, ψ is the electrostatic potential, and ρ is the charge density.

The charge density ρ is given by

$$\rho = q\,(p - n + N_D - N_A) \quad 3.2$$

Where q is the elementary electronic charge, *n* and *p* are the electron and hole densities, N_D is the concentration of ionized donors, and N_A is the concentration of ionized acceptors.

Therefore, Poisson's equation can be written as

$$\nabla.\epsilon\nabla\psi = -q\,(p - n + N_D - N_A) \quad 3.3$$

The general equations for describing electron and hole transport in a semiconductor under non-equilibrium conditions are the electron and hole continuity equations:

$$\frac{\partial n}{\partial t} = G_n - \mathrm{R_n} + \frac{1}{q}\nabla.\vec{J_n} \quad 3.4$$

$$\frac{\partial p}{\partial t} = G_p - \mathrm{R_p} - \frac{1}{q}\nabla.\vec{J_p} \quad 3.5$$

Where J_n and J_p are the electron and hole current densities respectively, G_n and G_p are the electron and hole generation rates ($m^{-3}s^{-1}$) due to external excitation, and R_n and R_p are the electron and hole recombination rates.

The current densities can be expressed in terms of quasi-Fermi levels and $\emptyset_n$ and $\emptyset_p$ as

$$\vec{J_n} = -q\,\mu_n n\nabla\emptyset_n \quad 3.6$$

$$\vec{J_p} = -q\,\mu_p p\nabla\emptyset_p \quad 3.7$$

Where μ_n and μ_p are the electron and hole mobilities.

The quasi-Fermi levels are linked to the carrier concentrations and the potential through Boltzmann approximations as

$$n = n_{ie} \exp\left(\frac{q(\psi - \emptyset_n)}{kT_L}\right) \quad 3.8$$

$$p = n_{ie} \exp\left(\frac{q(\psi - \emptyset_p)}{kT_L}\right) \quad 3.9$$

Where n_{ie} is the effective intrinsic concentration and T_L is the lattice temperature.

By substituting equations (3.8) and (3.9) into the current density expressions in equations (3.6) and (3.7), we have the following relations

$$\vec{J}_n = qD_n\nabla n - q\,\mu_n n\nabla\psi - \mu_n n\,[kT_L\nabla(ln\, n_{ie})] \quad 3.10$$

$$\vec{J}_p = -qD_p\nabla p - q\,\mu_p p\nabla\psi - \mu_p p\,[kT_L\nabla(ln\, n_{ie})] \quad 3.11$$

The final term accounts for the gradient in the effective carrier concentration, which takes into account the band gap narrowing effect.

The effective electric fields can be defined as

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

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

Now, the conventional formulation of the current as the sum of a diffusion and drift term can be written as

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

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

Here the diffusion current is proportional to the gradient of the carrier concentration, indicating that carriers flow from a region of high concentration to one of low concentration. The constants D_n and D_p are the diffusion coefficients or diffusivities, and are related to the mobilities μ_n and μ_p by Einstein's relations:

$$D_n = \frac{kT_L}{q}\mu_n \quad 3.17$$

$$D_p = \frac{kT_L}{q}\mu_p \quad 3.18$$

The total current density J at any point of the analyzed structure is then calculated as a sum of electron and hole currents

$$\vec{J} = \vec{J}_n + \vec{J}_p \quad 3.19$$

3.2 Transport Models

3.2.1 Drift-Diffusion Model (DD)

The drift-diffusion model is widely used as a starting point for simulating the carrier transport in semiconductors. For isothermal simulation the DD model incorporated in TCAD simulator numerically solves Poisson's equation, and the carrier continuity equation self-consistently to get electron and hole concentrations, and the electrostatic potential at all points within the device structure (equations 3.14 and 3.15), assuming full impurity ionization, non-degenerate statistics, steady state, and constant temperature (the carrier temperatures are assumed to be in equilibrium with the lattice temperature). Even though the DD model is a simple model and not a very precise model for complex semiconductor devices simulation, it is still a good starting point for any device simulation due to the relative simplicity of the model, relative good convergence properties and its ability to provide an initial overview of the device operation in a short time [8].

3.2.2 Hydrodynamic Model (HD)

Due to its limitations, the conventional drift-diffusion model is not applicable for simulations in the submicron regime. As well, it is incapable to describe properly the internal and/or the external characteristics of state of the art semiconductor devices. Mainly, the DD approach cannot reproduce velocity overshoot and often overestimates the impact ionization generation rates. Consequently, the hydrodynamic model (or energy balance model) is preferred

for determining the velocity overshoot and describing properly the characteristics of state of the art semiconductor devices, as it incorporates ballistic effects missing in the DD model. The HD model couples the basic set of semiconductor equations; Poisson equation and continuity equations, with energy balance equations for electrons, holes, and the lattice.

The current densities in the HD case are defined as a sum of four contributions:

$$\vec{J}_n = q\,\mu_n\left(n\nabla E_C + kT_n\nabla n + f_n^{td}kn\nabla T_n - \frac{3}{2}nkT_n\nabla\, ln\, m_n\right) \quad 3.20$$

$$\vec{J}_p = q\,\mu_p\left(p\nabla E_{\mathrm{V}}kT_{\mathrm{p}}\nabla p - f_{\mathrm{p}}^{td}kp\nabla T_{\mathrm{p}} - \frac{3}{2}pkT_p\nabla\, ln\, m_{\mathrm{p}}\right) \quad 3.21$$

Where m_n and m_{p} are the effective masses of electrons and holes, T_n and T_p are the electron and hole carrier temperatures, f_n^{td} and f_{p}^{td} are parameters function of the material. The first term accounts for spatial variations of electrostatic potential, electron affinity, and the bandgap. The three remaining terms take into account the contributions due to the gradient of concentrations and carrier temperature, and the spatial variation of the effective masses, respectively.

The energy fluxes for electrons, holes and the lattice are given by

$$\vec{S}_n = -\frac{5r_n}{2}\left[\frac{k_BT_n}{q}\vec{J}_n + f_n^{hf}\left(\frac{k_B^2}{q}n\mu_nT_n\right)\nabla T_n\right] \quad 3.22$$

$$\vec{S}_p = -\frac{5r_p}{2}\left[-\frac{k_BT_p}{q}\vec{J}_p + f_p^{hf}\left(\frac{k_B^2}{q}p\mu_pT_p\right)\nabla T_p\right] \quad 3.23$$

$$\vec{S}_l = -k_L\nabla T_L \quad 3.24$$

where k_B is the Boltzmann constant, f_n^{hf} and f_p^{hf} are parameters function of the material, and k_L is the lattice thermal conductivity.

The default parameter values of Sentaurus Device are summarized in Table 4. These parameters are accessible in the parameter file of Sentaurus Device. Different values of these parameters can significantly influence the physical results, such as velocity distribution and possible spurious velocity peaks. By changing these parameters, Sentaurus Device can be tuned to a very wide set of hydrodynamic/energy balance models as described in the literature.

	$r_n = r_p$	$f_n^{td} = f_p^{td}$	$f_n^{hf} = f_p^{hf}$
Stratton	0.6	0	1
Blotekjaer	1	1	1

Table 4: HD model parameter values from Stratton and Blotekjaer

In TCAD simulator, the electron, hole, and the lattice temperatures are solved by specifying the keywords in the command file; Electron Temperature, Hole Temperature, and Lattice Temperature, respectively. In spite of the time consuming and convergence problems associated with using the HD model in our study, it has the advantages of good accuracy and proper description of the characteristics of state of the art semiconductor devices [5] [8].

3.3 Generation-Recombination Models

Generation-Recombination are the processes by which mobile charge carriers are created and eliminated. The process by which both carriers annihilate each other is called recombination. In this process the electrons fall in one or multiple steps into the empty state which is associated with the hole, both carriers eventually disappear in the process. The carrier generation is a process where electron-hole pairs are created by exciting an electron from the valence band of the semiconductor to the conduction band, thereby creating a hole in the valence band. The carrier generation-recombination models used in this study are the Shockley-Read-Hall (SRH) recombination model, and Auger recombination model.

3.3.1 SRH-Recombination Model

The SRH model describes the statistics of recombination of holes and electrons in semiconductors occurring through the mechanism of trapping. The net recombination rate for trap-assisted recombination is given by:

$$R_{net}^{SRH} = \frac{np - n_{i,eff}^2}{\tau_p(n + n_1) + \tau_n(p + p_1)} \quad 3.25$$

The variables n_1 and p_1 are defined as

$$n_1 = n_{i,eff}\ exp\left(\frac{E_{trap}}{kT}\right) \quad 3.26$$

$$p_1 = n_{i,eff}\ exp\left(\frac{-E_{trap}}{kT}\right) \quad 3.27$$

Where E_{trap} is the difference between the defect level and intrinsic level, τ_p and τ_p are the electron and hole lifetimes, and $n_{i,eff}$ is the effective intrinsic concentration.

The SRH recombination rate using Fermi statistics is given by

$$R_{net}^{SRH} = \frac{np - \gamma_n\gamma_p n_{i,eff}^2}{\tau_p(n + \gamma_n n_1) + \tau_n(p + \gamma_p p_1)} \quad 3.28$$

With

$$\gamma_n = \frac{n}{N_C} exp\left(\frac{E_C - E_{F,n}}{kT}\right) \quad 3.29$$

$$\gamma_p = \frac{p}{N_V} exp\left(\frac{E_{F,p} - E_V}{kT}\right) \quad 3.30$$

Where N_C and N_V are the effective density-of-states, E_C, E_V are conduction and valence band edges, $E_{F,n}$ and $E_{F,p}$ are the quasi-Fermi energies for electrons and holes respectively.

The doping dependence of the SRH lifetimes is modeled through Scharfetter relation as

$$\tau_{dop}(N_i) = \tau_{min} + \frac{\tau_{max} + \tau_{min}}{1 + \left(\frac{N_i}{N_{ref}}\right)^{\gamma}} \quad 3.31$$

This indicates that the solubility of a fundamental acceptor-type defect is strongly correlated to the doping density.

3.3.2 Auger Recombination Model

Auger recombination is typically important at high carrier densities such as in heavily doped regions or in cases of high injection. The rate of band-to-band Auger recombination is given by

$$R_{net}^{A} = \left(C_n n + C_p p\right)(np + n_{i,eff}^2) \quad 3.32$$

Where C_n and C_p are Auger coefficients.

The temperature-dependant Auger coefficients are given by

$$C_n(T) = \left(A_{A,n} + B_{A,n}\left(\frac{T}{T_0}\right) + C_{A,n}\left(\frac{T}{T_0}\right)^2\right)\left[1 + H_n exp\left(-\frac{n}{N_{0,n}}\right)\right] \quad 3.33$$

$$C_p(T) = \left(A_{A,p} + B_{A,p}\left(\frac{T}{T_0}\right) + C_{A,p}\left(\frac{T}{T_0}\right)^2\right)\left[1 + H_p exp\left(-\frac{p}{N_{0,p}}\right)\right] \quad 3.34$$

Where T_0=300K.

The default coefficients of Auger recombination model for Si are summarized in Table 5. For Ge and SiGe materials, same parameters have been used [5].

Coefficient	A_A [cm^6s^{-1}]	B_A [cm^6s^{-1}]	C_A [cm^6s^{-1}]	H[1]	N_0 [cm^{-3}]
Electrons	6.7×10^{-32}	2.45×10^{-31}	-2.2×10^{-32}	3.46667	1×10^{18}
Holes	7.2×10^{-32}	4.5×10^{-33}	2.63×10^{-32}	8.25688	1×10^{18}

Table 5: Auger Recombination Coefficients of for Silicon.

3.4 Mobility Models

Carrier mobility in semiconductors is determined by a variety of physical mechanisms. Electrons and holes are scattered by thermal lattice variation, ionized and neutral impurities, dislocations, and electrons and holes themselves. Several models for carrier mobility in Si and strained Si have been implemented in TCAD simulator. In this study, Philips unified mobility model [9][10], high field saturation model [6], stress-induced electron mobility model [11], and piezoresistance mobility model have been used .

3.4.1 Philips Unified Mobility Model

The Philips unified mobility model proposed by Klaassen provides unified description of the majority and minority carrier bulk mobilities. In addition to describing the temperature dependence of the mobility, the model includes the effect of electron-hole scattering, screening of ionized impurities by charge carriers, and clustering of impurities. Because the carrier mobility is given as an analytical function of the donor, acceptor, electron and hole concentrations, and the temperature, this model is well suited for simulation purposes, and excellent agreement is obtained with published experimental data on Si.

The bulk carrier mobility for each carrier is given by Mathiessen's rule

$$\frac{1}{\mu_{i,b}} = \frac{1}{\mu_{i,L}} + \frac{1}{\mu_{i,DAeh}} \tag{3.35}$$

The index *i* takes the values "e" for electrons and "h" for holes.

The first term in equation (3.35) represents phonon (lattice) scattering:

$$\mu_{i,L} = \mu_{i,max} \left(\frac{\mathrm{T}}{300\mathrm{K}}\right)^{-\theta_i} \tag{3.36}$$

The second term in equation (3.35) accounts for all other bulk scattering mechanisms (due to free carriers, and ionized donors and acceptors):

$$\mu_{i,DAeh} = \mu_{i,N} \left(\frac{N_{i,sc}}{N_{i,sc,eff}}\right) \left(\frac{N_{i,ref}}{N_{i,sc}}\right)^{\alpha_i} + \mu_{i,c} \left(\frac{n+p}{N_{i,sc,eff}}\right) \tag{3.37}$$

Where

$$\mu_{i,N} = \frac{\mu_{i,max}^2}{\mu_{i,max} - \mu_{i,min}} \left(\frac{T}{300K}\right)^{3\alpha_i - 1.5} \tag{3.38}$$

$$\mu_{i,c} = \frac{\mu_{i,max}\,\mu_{i,min}}{\mu_{i,max} - \mu_{i,min}} \left(\frac{300K}{\mathrm{T}}\right)^{0.5} \tag{3.39}$$

$$N_{e,sc} = N_D^* + N_A^* + p \tag{3.40}$$

$$N_{h,sc} = N_D^* + N_A^* + n \tag{3.41}$$

$$N_{e,sc,eff} = N_D^* + G(P_e)N_A^* + \mathrm{f_e}\,\frac{\mathrm{p}}{\mathrm{F}(P_e)} \qquad 3.42$$

$$N_{h,sc,eff} = N_A^* + G(P_h)N_D^* + \mathrm{f_h}\,\frac{\mathrm{p}}{\mathrm{F}(P_h)} \qquad 3.43$$

Where N_A^* and N_D^* are acceptors and donors concentrations respectively, $G(P_i)$ and $\mathrm{F}(P_i)$ are analytic functions describing minority impurity and electron-hole scattering. θ_i, $N_{i,ref}$, α_i, and f_i are model parameters.

3.4.2 High Field Saturation Model

High-field behavior shows that carrier mobility decreases with electric field. With the increase of the applied electric field, carriers gain energies above the ambient thermal energy and are able to transfer energy to the lattice by optical phonon emission. Consequently, the mean drift velocity is no longer proportional to the electric field, but rises more slowly, and the mobility has to be reduced accordingly. Finally the velocity saturates to a finite velocity known as the saturation velocity (v_{sat}) which is principally a function of lattice temperature.

The high field mobility degradation due to carrier velocity saturation effects is introduced by the Canali model through the relation

$$\mu(F) = \frac{(\alpha+1)\mu_{low}}{\alpha\left[1+\left(\frac{(\alpha+1)\mu_{low}F_{hfs}}{v_{sat}}\right)^{\beta}\right]^{\frac{1}{\beta}}} \qquad 3.44$$

Where

$$v_{sat} = v_{sat,0}\left(\frac{300K}{T}\right)^{v_{sat,exp}} \qquad 3.45$$

Where μ_{low} is the low field mobility, F_{hfs} is the deriving field, β is a temperature dependant model parameter.

3.5 Mobility Models Under Stress

3.5.1 Piezoresistance Mobility Model

This approach focuses on modeling of the variation of the carrier's mobility with the applied stress. The applied stress changes the electrical resistivity (conductivity) of the material, leading to mobility modification according to the relation

$$\tilde{\mu}_\alpha = \mu_\alpha^0 \left(\tilde{1} - \tilde{\Pi}^\alpha . \tilde{\sigma}\right) \quad 3.46$$

and the current density is given by

$$\vec{J}_\alpha = \tilde{\mu}_\alpha . \left(\frac{\vec{J}_\alpha^0}{\mu_\alpha^0}\right) \quad 3.47$$

Where $\alpha = n, p$, $\tilde{\mu}_\alpha$ is the second rank mobility tensor, μ_α^0 is the isotropic mobility without stress, $\tilde{1}$ is the identity tensor, $\tilde{\sigma}$ is the stress tensor, $\tilde{\Pi}^\alpha$ is the tensor of piezoresistance coefficients that depend on the doping concentration and temperature distribution, $\vec{J}_\alpha^0$ is the vector of the carrier current without the stress.

Taking in consideration the change in the effective masses and anisotropic scattering due to the applied stress, the piezoresistance coefficients become

$$\Pi_{ij}^\alpha = \Pi_{ij,con}^\alpha + \Pi_{ij,var}^\alpha P_\alpha(N,T) \quad 3.48$$

The first term is an independent constant represents the change in the effective mass effect, while the second term represents the scattering effect. Where $P_\alpha(N,T)$ is doping-dependent and temperature-dependent factor. Furthermore, the enhancement factors for both electron and hole mobilities due to the applied stress have been calculated using the piezoresistance mobility factor model [5] [12].

3.5.2 Stress-Induced Electron Mobility Model

This approach focuses on the modeling of the mobility changes due to the carrier redistribution between bands in silicon due to the applied stress. The origin of the electron

mobility enhancement has been explained in chapter 2. The stress-induced electron mobility enhancement is given by

$$\mu_{ii}^{n} = \mu_{n}^{0}\left[1 + \frac{1 - \frac{m_{nl}}{m_{nt}}}{1 + 2\left(\frac{m_{nl}}{m_{nt}}\right)}\left(\frac{F_{\mathrm{n}}\frac{(F_n - E_C - \Delta E_{C,i})}{kT}}{F_{\mathrm{n}}\frac{(F_n - E_C - \Delta E_C)}{kT}} - 1\right)\right] \quad 3.49$$

Where μ_n^0 is electron mobility without the strain, m_{nl} and m_{nt} are the electron longitudinal and transverse masses in the subvalley, respectively, E_C and $\Delta E_{C,i}$ are the change in the energy of the unstrained and the strained carrier sub-valleys, F_{n} is the quasi-Fermi level of electrons. The index *i* corresponds to a direction (i.e. μ_{11}^n is the electron mobility in the direction of the x-axis of the crystal system and, therefore, $\Delta E_{C,1}$should correspond to the two-fold subvalley along the x-axis).

The general expression of the mobility along valley *i*, which includes the stress-induced carrier redistribution and change in the intervalley scattering, is given by

$$\mu_{ii}^{n} = \frac{3\mu_n^0}{1 + 2\left(\frac{m_{nl}}{m_{nt}}\right)} \frac{\frac{\tau_i}{\tau^0} F_{1/2}\left(\frac{\eta - \Delta E_{C,i}}{kT}\right) + \frac{\tau_j}{\tau^0}\frac{m_{nl}}{m_{nt}} F_{1/2}\left(\frac{\eta - \Delta E_{C,j}}{kT}\right) + \frac{\tau_l}{\tau^0}\frac{m_{nl}}{m_{nt}} F_{1/2}\left(\frac{\eta - \Delta E_{C,l}}{kT}\right)}{F_{1/2}\left(\frac{\eta - \Delta E_{C,i}}{kT}\right) + F_{1/2}\left(\frac{\eta - \Delta E_{C,j}}{kT}\right) + F_{1/2}\left(\frac{\eta - \Delta E_{C,l}}{kT}\right)} \quad 3.50$$

With

$$\frac{\tau_i}{\tau^0} = \frac{1}{1 + (h_i - 1)\frac{1 - \beta^{-1}}{1 + \left(\frac{N_{tot}}{N_{ref}}\right)^{\alpha}}} \quad 3.51$$

$$\eta = F_n - E_C \quad 3.52$$

Where β, N_{ref}, and α are model fitting parameters, h_i is the ratio between unstrained and strained relaxation times for intervalley scattering in valley *i*, and $\frac{\tau_i}{\tau^0}$ is the ratio between strained and unstrained total relaxation times for the valley *i*.

3.6 Band Structure Models

3.6.1 Bandgap Narrowing

Band-gap narrowing is observed in highly-doped semiconductors. At low impurity concentrations the interaction between electrons and holes and the Coulomb force acting between them doesn't play any role. Thus the impurities do not disturb the property of the crystal and the forbidden energy gap Eg is well defined by sharp band edges. On the other hand, at high impurity concentrations, there is Coulomb interaction between the ionized impurities, and thus overlapping of the wave functions associated with these impurities. This results in a splitting of the impurity energy level, and impurity band is formed. As a result the potential energy of these ionized impurities is reduced resulting in a narrowing of the bandgap.

The Bandgap narrowing for the Slotboom model is given by

$$\Delta E_g^0 = E_{ref}\left[ln\left(\frac{N_{tot}}{N_{ref}}\right) + \sqrt{\left(ln\left(\frac{N_{tot}}{N_{ref}}\right)\right)^2 + 0.5}\right] \quad 3.53$$

Where E_{ref} and N_{ref} are model parameters.

The bandgap narrowing effect has been modeled based on experimental measurements of $\mu_n n^2$ in n-p-n transistor and $\mu_p p^2$ for p-n-p transistor with different base doping concentrations and a 1D model for the collector current [13].

3.6.2 Intrinsic Density

In Sentaurus TCAD device simulation, the intrinsic carrier density for undoped semiconductors as a function of the effective density of states in the conduction and valence band, and the bandgap energy is given by the relation

$$n_i(T) = \sqrt{N_C(T)\, N_V(T)}\; exp\left(-\frac{E_g(T)}{2kT}\right) \quad 3.53$$

The effective intrinsic carrier density is given by

$$n_{i,eff} = n_i\; exp\left(\frac{E_{bgn}}{2kT}\right) \quad 3.54$$

Where E_{bgn} is the doping-dependant bandgap narrowing [5].

3.6.3 Effective Density-of-States

Sentaurus Device computes the effective density-of-states as a function of carrier effective mass. The effective mass may be either independent of temperature or a function of the temperature-dependent band gap. For carriers in silicon the effective mass temperature-dependant is the most appropriate model for calculating the effective density-of-states.

The effective density of states in the conduction band N_C is given by

$$N_C\left(m_{n,}, T_n\right) = 2.540933\; \times 10^{19}\; \left(\frac{m_n}{m_0}\right)^{\frac{3}{2}}\; \left(\frac{T_n}{300K}\right)^{\frac{3}{2}}\; cm^{-3} \quad 3.55$$

Where m_n is the effective mass of electrons.

The effective density of states in the valence band N_V is given by

$$N_V\left(m_{p,}, T_p\right) = 2.540933\; \times 10^{19}\; \left(\frac{m_p}{m_0}\right)^{\frac{3}{2}}\; \left(\frac{T_p}{300K}\right)^{\frac{3}{2}}\; cm^{-3} \quad 3.55$$

Where m_p is the effective mass of holes [5][14].

4. TCAD Calibration

In order to calibrate and to study the accuracy of TCAD simulator, simulations have been performed in accordance with Monte Carlo (MC) simulation performed by Bundeswehr University-Munich [15]. The calibration process has been executed through three approaches; (1) Mobility models calibration through simulating a sample of $Si_{1-x}Ge_x$ substrate parameterized with different Ge content (0% - 28%), different doping concentrations for Arsenic and Boron impurities (10^{17} cm^{-3} - 10^{20} cm^{-3}), and different applied electric fields. (2) Transport models calibration through simulating a reference transistor (f_{Tpeak}=100 GHz) and comparing the obtained results with Monte Carlo simulation results. (3) Simulating IMEC NPN-SiGe-HBT structure and comparing the obtained results with measurement. A description of each approach and the obtained results will be presented.

4.1 Mobility Models Calibration

A sample of $Si_{1-x}Ge_x$ substrate with different homogenous doping concentrations of Boron and Arsenic impurities (10^{17} cm^{-3} to 10^{20} cm^{-3}), and different Germanium content, *x*, (0% - 0.28%) used in the calibration process of Sentaurus TCAD simulator is shown in Fig.39.

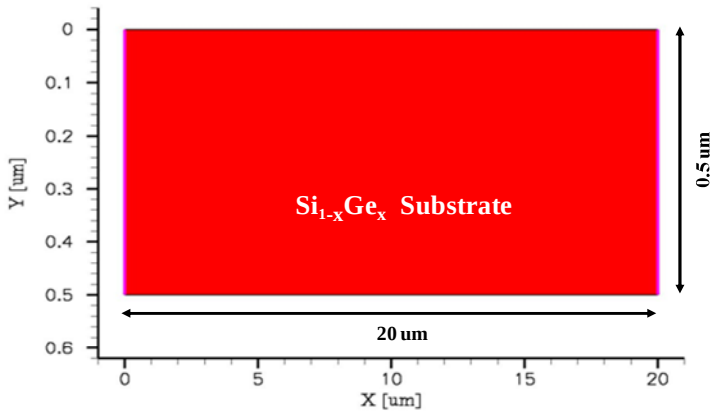

Fig.39: The $Si_{1-x}Ge_x$ substrate used in TCAD simulator calibration process.

The applied bias varied between the electrodes to obtain the desired electric fields (0.2 - 400 Volt). Physical quantities such as mobility and velocity have been computed versus crystal

direction using Hydrodynamic transport model (HD), Philips unified mobility model, and doping dependence mobility model. The device simulation has been run for different temperatures (T=300, 400, and 450 K). The physical models used in the calibration process are summarized in Table 6. The obtained results have been extracted at the middle of the sample and compared with Monte Carlo simulation results from Bundeswehr University-Munich.

Model (I)	Model (II)
Hydrodynamic EffectiveIntrinsicDensity (Slotboom) Mobility • DopingDep. • eHighFieldSaturation • hHighFieldSaturation	Hydrodynamic EffectiveIntrinsicDensity (Slotboom) Mobility • PhuMob. • eHighFieldSaturation • hHighFieldSaturation)

Table 6: Physical models used in the calibration process.

A comparison between TCAD and MC simulation results for electron velocity and hole velocity for both Boron and Arsenic doping impurities are shown in the following figures (Fig.40 to Fig.44) (for Ge_content = 8%, 24%, at T=300K). The complete set of results are shown in the Appendix (i.e, Ge_content = 4%, 12%, 18%, 28%).

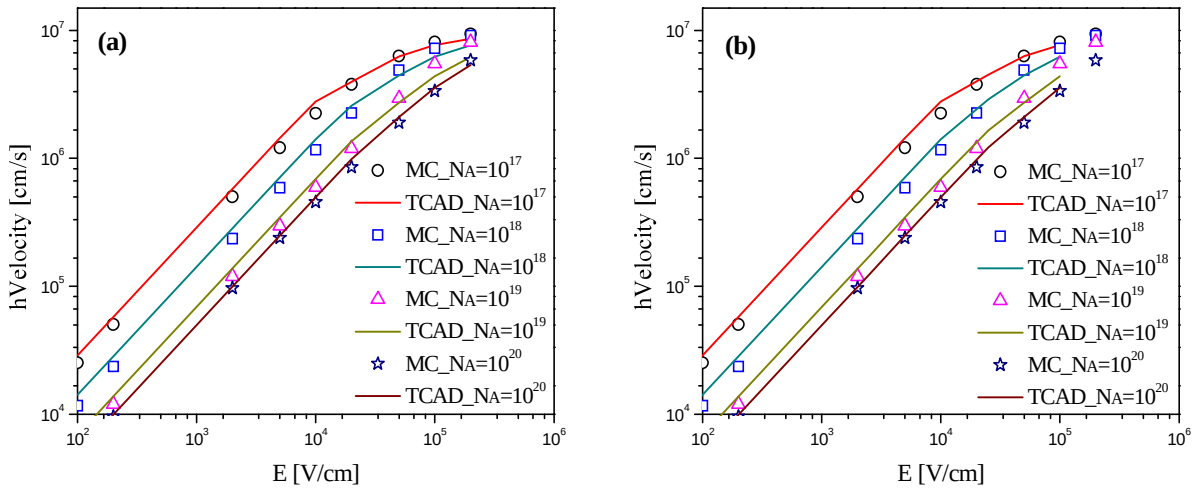

Fig.40: Comparison between MC and TCAD simulation results for hVelocity vs. Electric field using (a) model (I) (b) model (II), (Ge_content = 8%, T=300K).

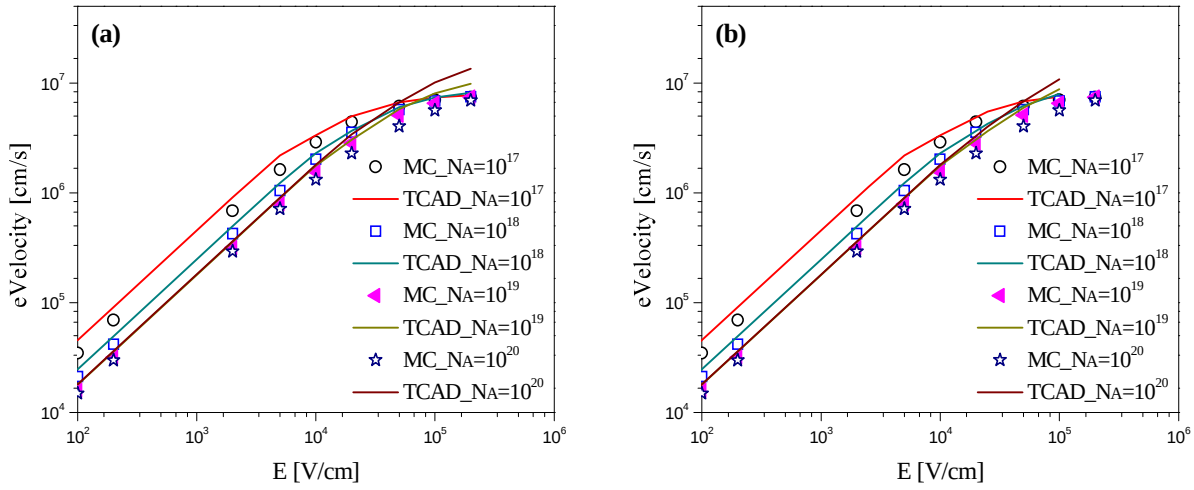

Fig.41: Comparison between MC and TCAD simulation results for eVelocity vs. Electric field using (a) model (I) (b) model (II), (Ge_content = 8%, T=300K).

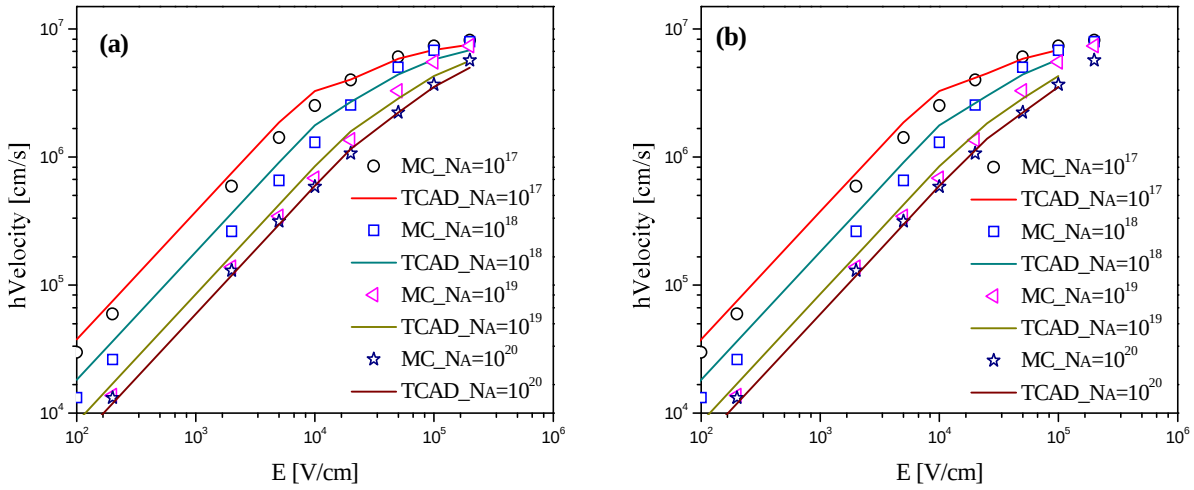

Fig.42: Comparison between MC and TCAD simulation results for hVelocity vs. Electric field using (a) model (I) (b) model (II), (Ge_content = 24%, T=300K).

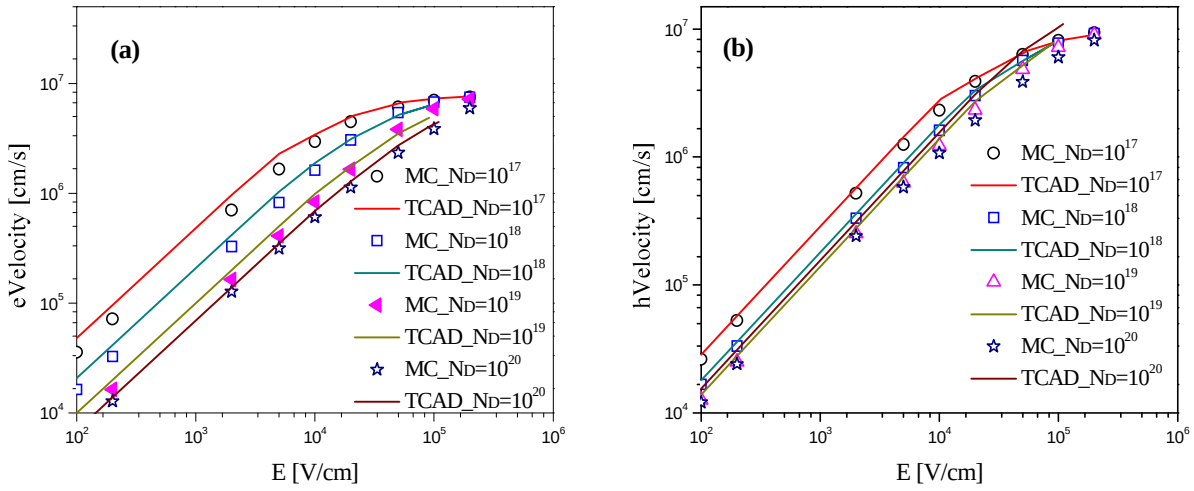

Fig.43: Comparison between MC and TCAD simulation results using model (II) for (a) eVelocity vs. Electric field (b) hVelocity vs. Electric field, (Ge_content = 8%, T=300K).

In addition to that, the minority carriers mobility in $Si_{1-x}Ge_x$ substrate as a function of Germanium content, *x*, at low electric field values has been simulated using TCAD modeling (model (II)) . The TCAD simulation results have been compared with MC simulation results reported by [16] as shown in Fig.44.

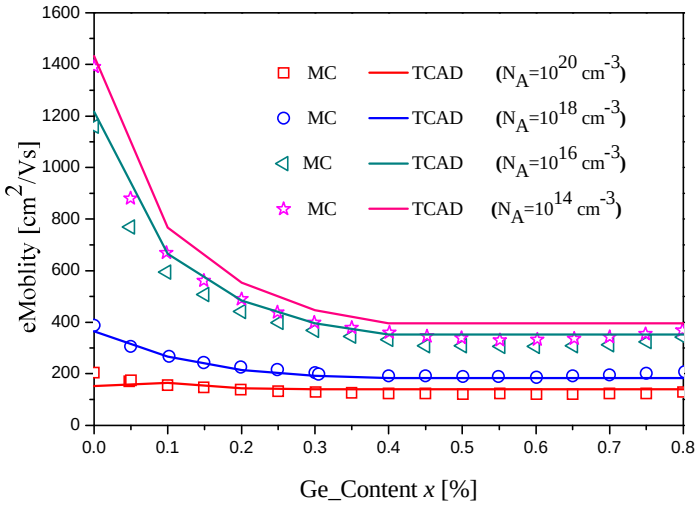

Fig.44: Comparison between TCAD and MC simulation results for the minority carriers mobility in $Si_{1-x}Ge_x$ substrate as a function of Ge content x using model (II).

As shown in the previous figures, good agreement has been achieved between MC and TCAD simulation results. Verifying the validity of the models used in TCAD simulator for mobility and velocity simulations.

4.2 Transport Models Calibration

In order to study the accuracy of TCAD default transport models, a reference NPN-SiGe-HBT device (f_{Tpeak}=100 GHz) simulated by MC simulation reported by [17] has been used. The device structure and the doping profile are shown in Fig.45.

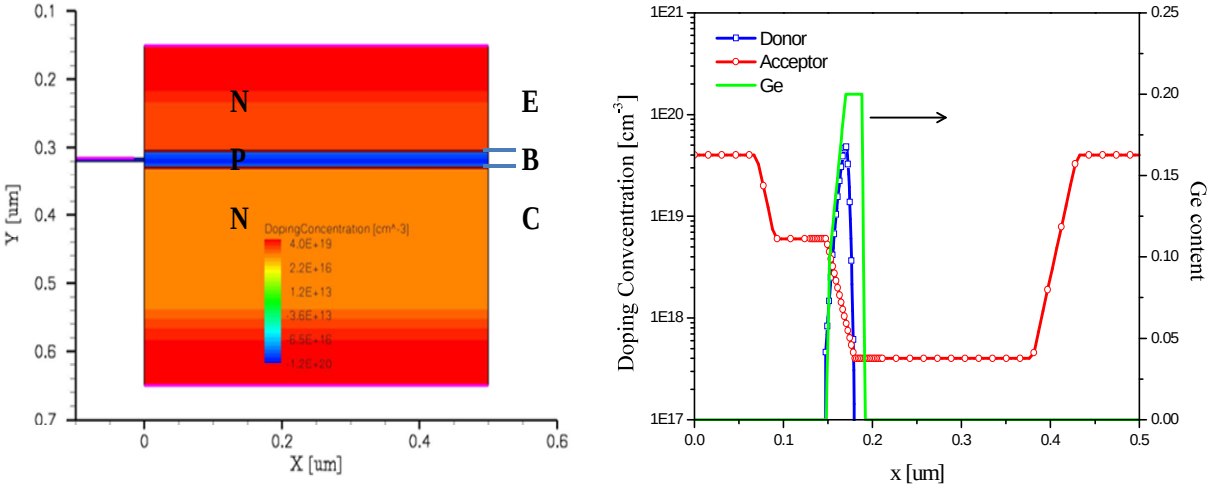

Fig.45: The reference transistor and the doping profile used in the calibration process.

The reference structure has been simulated using TCAD. The simulations have been performed using the default transport models and parameter files. Simulation results are then compared with MC simulation results as shown in Fig.46 and Fig.47. The results indicate that the transport models need to be calibrated (specially the HD model).

To calibrate the HD transport model, TCAD simulations have been run using the default parameters reported by Stratton and Blotekjaer (Table 4). The simulation results are then compared with MC simulation results as shown in Fig.48**.** Good agreement has been observed between TCAD and MC simulation results indicating that Blotekjaer approach provides good

agreement with MC data as shown in Fig.48. In contrast, Blotekjaer approach indicates velocity overshoot and negative slope in the output characteristic as shown in Fig.49.

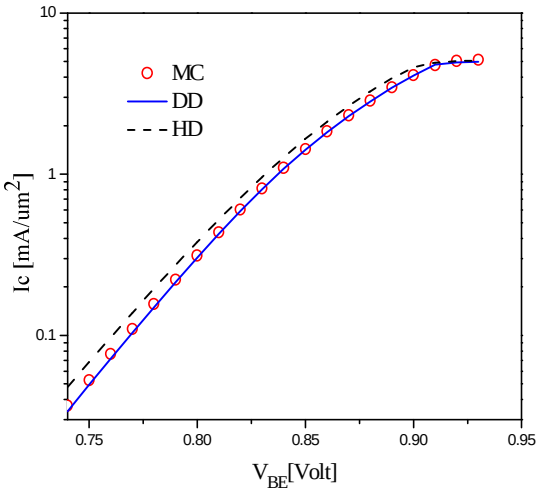

Fig.46 : Ic comparison between TCAD and MC simulation results

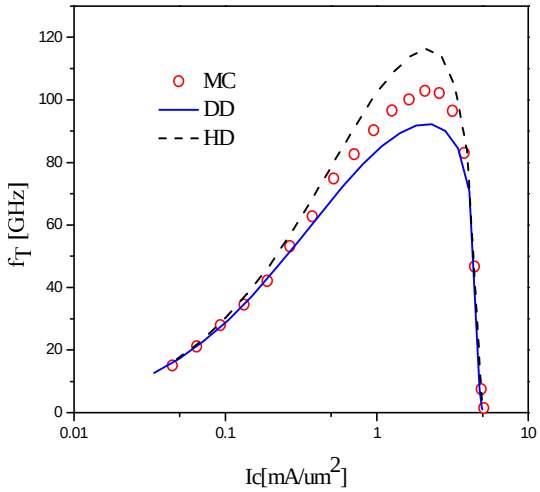

Fig.47 : f_T comparison between TCAD and MC simulation results

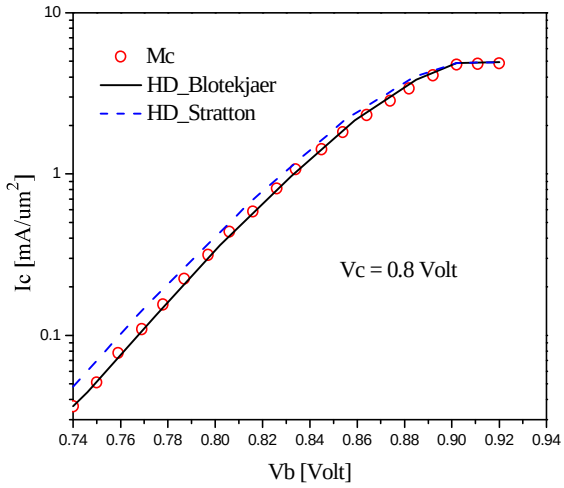

Fig.48 : Ic comparison between MC and HD model simulation results using default parameters

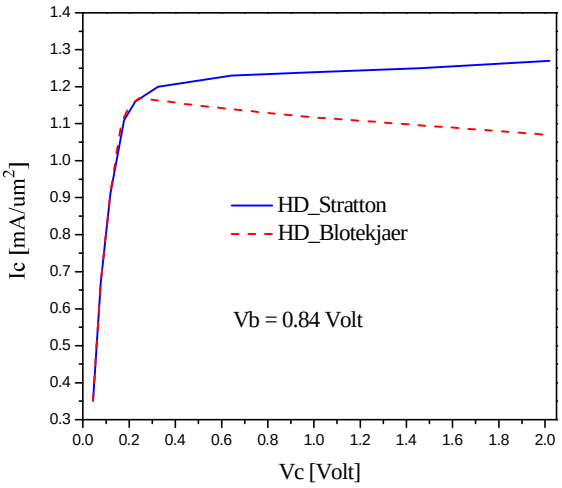

Fig.49: Ic output characteristic for HD model using default parameters

Therefore, in order to avoid current overshoot in the output characteristics, and to achieve good agreement between TCAD and MC simulation results, the parameters f_n^{td} and f_p^{td} have been set to zero, while r_n and r_p values have been optimized to match MC simulation results. The optimized parameter values are shown in Table 7 [18]. Then TCAD device simulations have been performed using the optimized parameters. The obtained results compared with MC simulation results are shown in Fig.50 to Fig.52.

	$r_n = r_p$	$f_n^{td} = f_p^{td}$	$f_n^{hf} = f_p^{hf}$
Stratton	0.6	0	1
Blotekjaer	1	1	1
Calibrated Model	0.2	0	1

Table 7: HD model parameter values from Stratton, Blotekjaer, and calibration

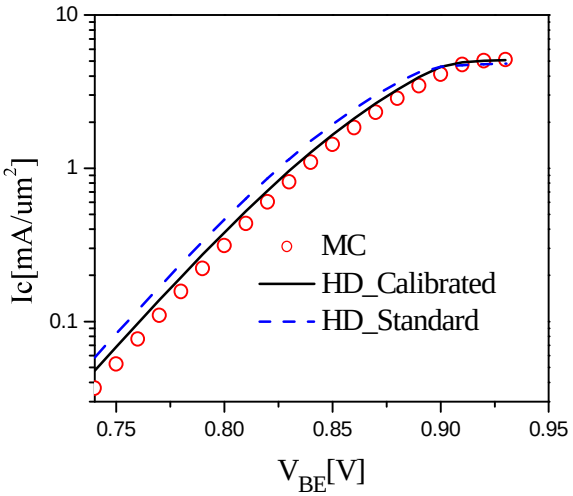

Fig.50: Ic comparison between MC and HD model simulation results

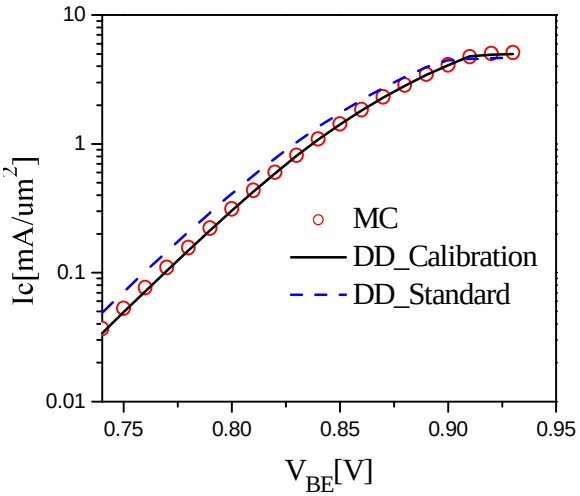

Fig.51: Ic comparison between MC and DD model simulation results

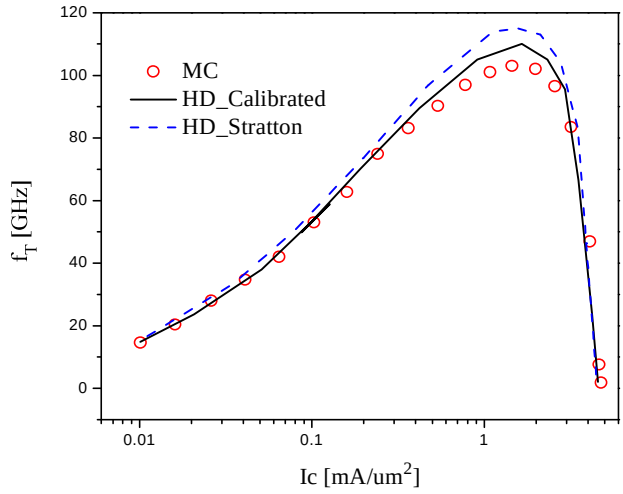

Fig.52 : f_T comparison between MC and HD model simulation results

In addition to that, the electrostatic potential, electron and hole density, electron velocity, electron temperature, collector current, and transit frequency have been simulated using TCAD. Due to the fact that TCAD simulator is not capable to simulate 1D bipolar transistor, different cross-sections at different values (10, 50, 100, 400 nm) have been performed to compare simulation results with the 1D MC simulation results as shown in Fig.53. Fig.54 to Fig.57 show a comparison between the obtained results using TCAD simulation and MC results for electron density for the different cross sections. Fig.58 shows a comparison between TCAD and MC results for hole density, electrostatic potential, electron velocity, and electron temperature respectively, for the cross section 10 nm.

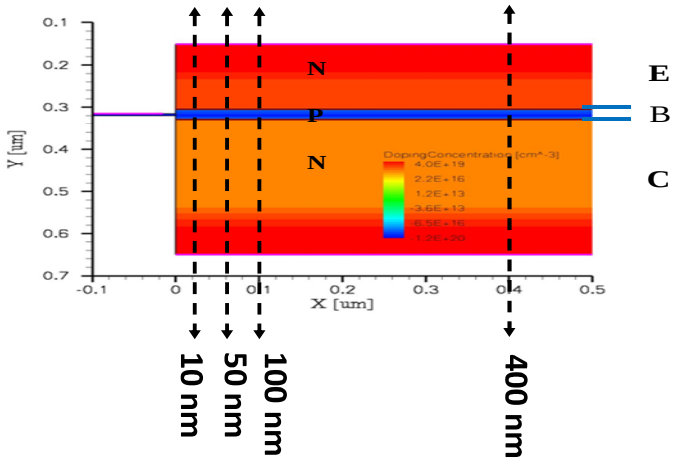

Fig.53 : The reference transistor with different cross sections

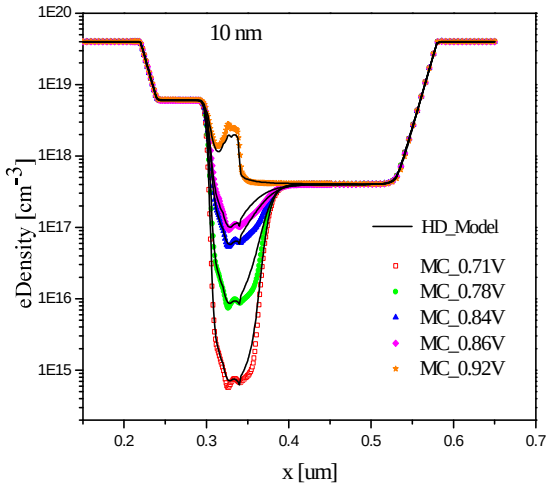

Fig.54 : Electron density comparison between MC and HD model simulation results for the cross-section 10 nm

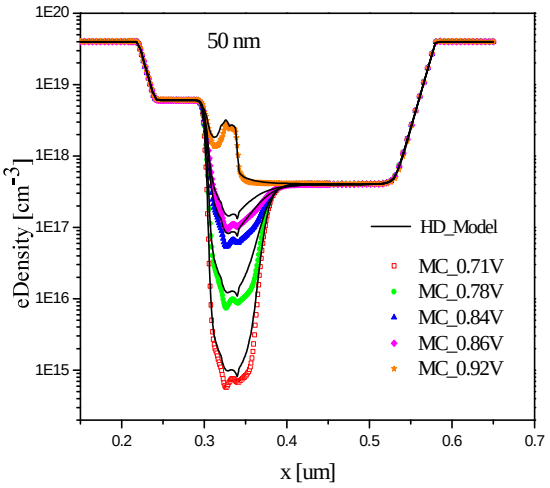

Fig.55 : Electron density comparison between MC and HD model simulation results for the cross-section 50 nm

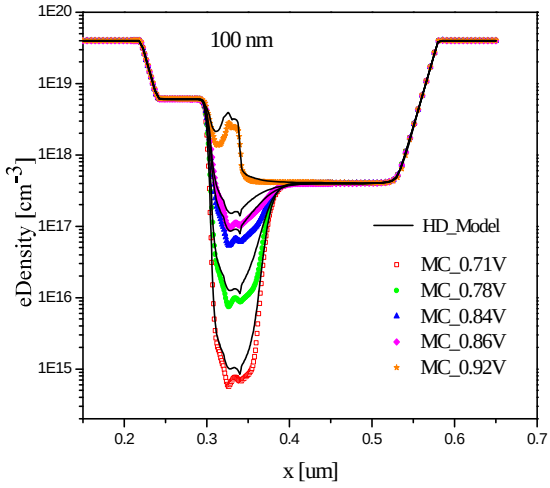

Fig.56: Electron density comparison between MC and HD model simulation results for the cross-section 100 nm

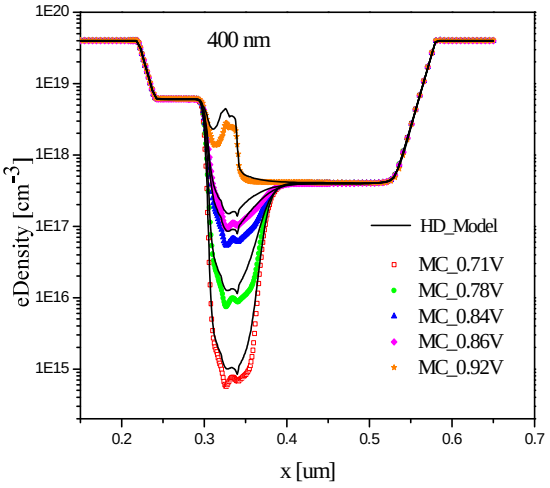

Fig.57 : Electron density comparison between MC and HD model simulation results for the cross-section 400 nm

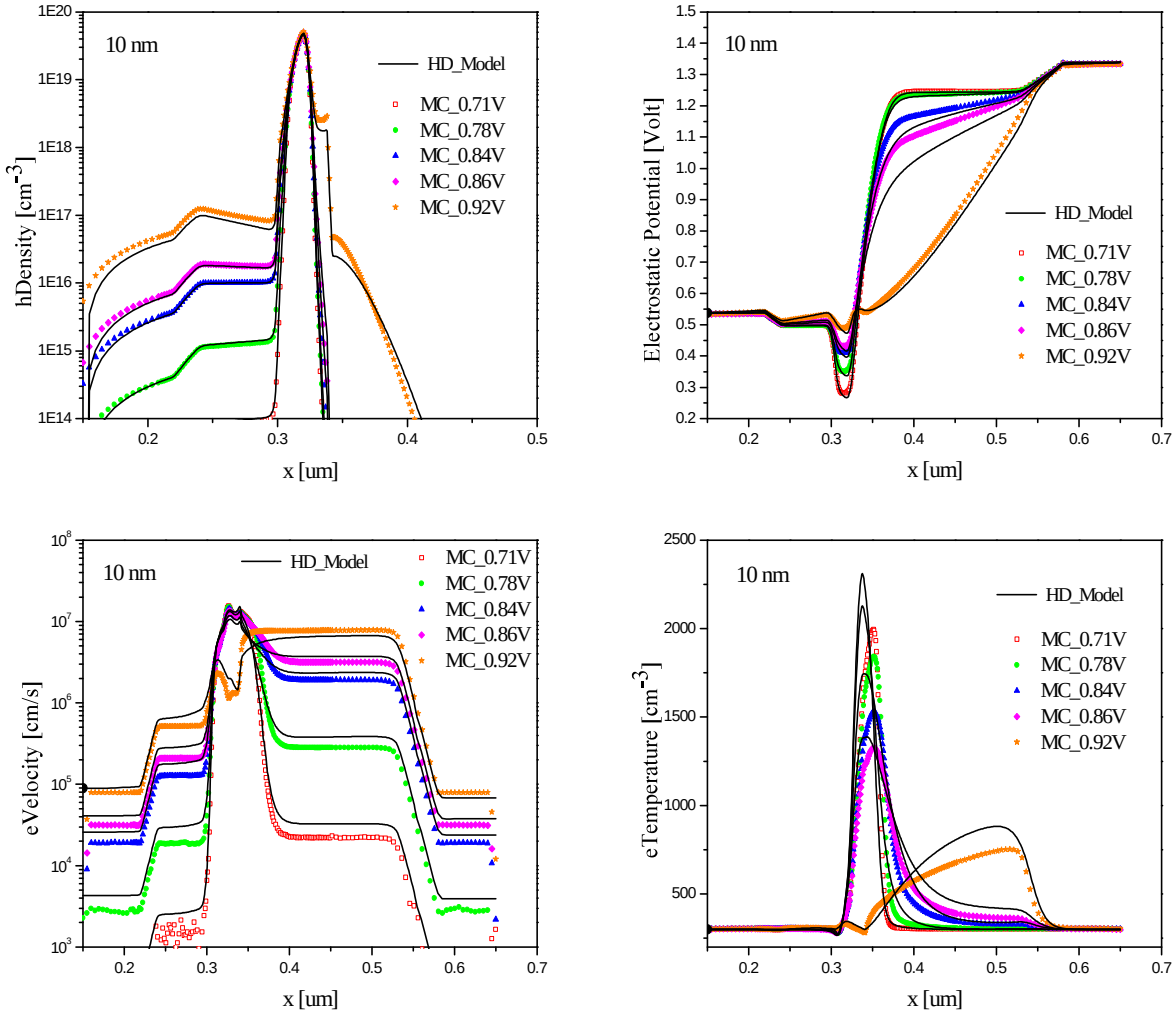

Fig.58 : Comparison between TCAD and MC simulation results (for cross section =10 nm): (a) hole density (b) electrostatic potential (c) electron velocity (d) electron temperature.

Good agreement has been observed between TCAD and MC simulation results for the 100 GHz profile, verifying the validity of the physical models and parameters used in the TCAD simulations. The complete set of results are shown in the Appendix (i.e. TCAD and MC results comparison of the quantities : hole density, electrostatic potential, electron velocity, and electron temperature for cross sections 50 nm, 100 nm, and 400 nm).

4.3 IMEC Structure

After performing step one and step two in the calibration process, a real structure has been simulated using the data from IMEC HBT device structure (f_{Tpeak}=240 GHz). Firstly, process simulation has been performed to build the device structure. The graded Ge profile and the doping profiles in the emitter, base, and collector regions of the device are identical to the ones from IMEC HBT device as shown in Fig.59 [19].

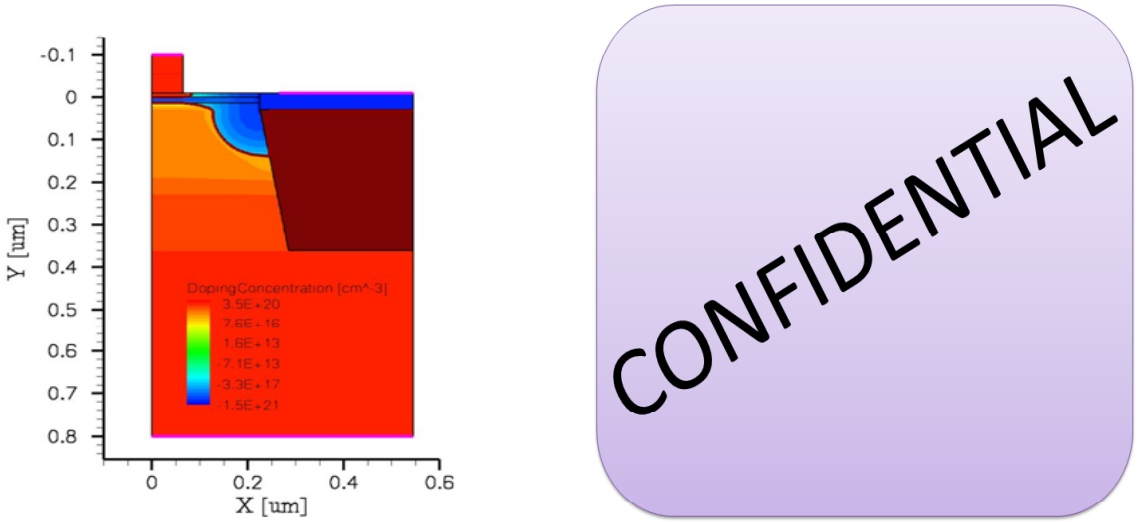

Fig.59 : IMEC HBT device, the graded Ge profile and the doping profiles in the emitter, base, and collector regions.

Then TCAD Sentaurus software tools have been used to perform the 2D device simulations using hydrodynamic model (HD). The carrier temperature equation for the dominant carriers is solved together with the electrostatic Poisson equation and the carrier continuity equations . The model parameters used in TCAD simulations have been calibrated by Universität der Bundeswehr-Munich using Monte Carlo simulations. The carrier mobilities have been calculated using Philips unified mobility model, the high field saturation was calculated through the Canali model by using carrier temperatures as the driving force. The carrier generation-recombination models used are the Shockley–Read–Hall recombination model, and Auger recombination model. As well, doping-induced bandgap narrowing model has been employed. Recombination time and velocity at the polysilicon/silicon interface, and in SiGe have been optimized. Self-heating has also been included by adjusting the thermal resistance (at the contact) and compared to measurement.

A comparison of the forward Gummel plots of IMEC HBT device measurements and TCAD simulation results are shown in Fig.60.

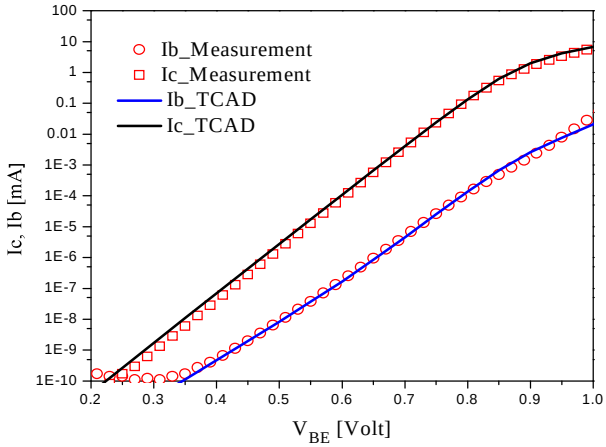

Fig.60 : Forward Gummel plots comparison of measurement and TCAD simulation results.

The f_T curves as a function of the collector current I_C for IMEC NPN-SiGe-HBT device measurements and NPN-SiGe-HBT device simulation results are shown in Fig.61. The base-collector junction capacitance (C_{CB}), and the base-emitter junction capacitance (C_{BE}) are plotted versus base-collector bias (V_{CB}), and base-emitter bias (V_{BE}) respectively, for both measurements and simulation results as shown in Fig.62 and Fig.63. Good agreement between measurements data and simulation results, verifying the validity of the physical models and parameters used in the TCAD simulations.

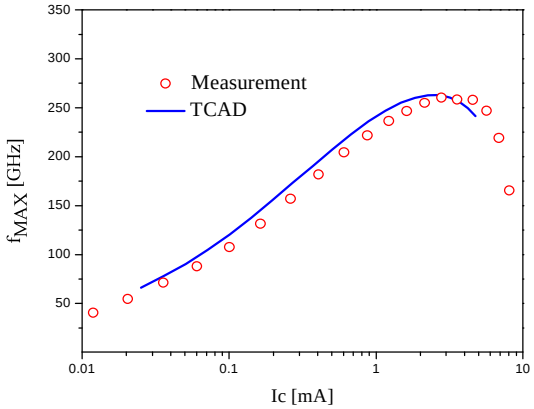

Fig.61 : f_T comparison between measurements and TACD simulation results (V_{BC}=0Volt).

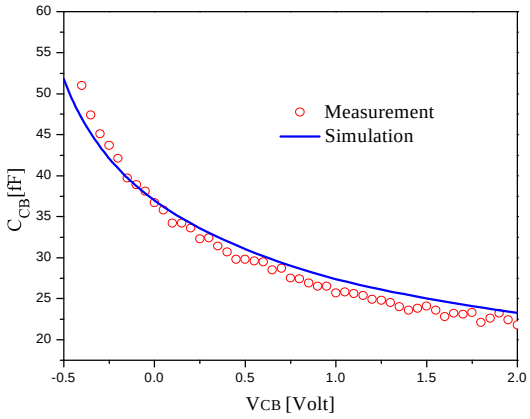

Fig.62 : C_{CB} comparison between measurements and TACD simulation results.

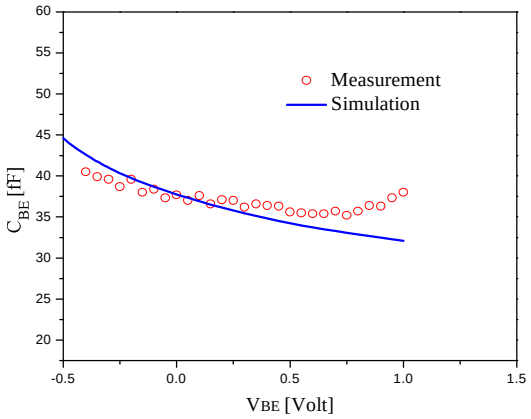

Fig.63 : C_{BE} comparison between measurements and TACD simulation results.

5. Conclusion

TCAD simulation tools are widely used throughout the semiconductor industry to speed up and cut the costs of developing new technologies and devices, since they make it possible to explore new technologies and concepts. TCAD simulations also provide information about the inner operation of devices, thus simplifying improvements on existing technologies. TCAD consists of two main branches: process simulation and device simulation. Process simulation models the complex flow of semiconductor fabrication steps and ends up with detailed information on geometrical shape and doping profile distribution of a semiconductor device. The device simulation uses the information of the process simulation as the input file to calculate the characteristics of semiconductor devices and parameters extraction.

Different physical models have been used in our TCAD simulator, including; HD and DD models, carrier mobility models (Philips unified mobility model, and the high field saturation model). The stress-induced mobility enhancement has been calculated using the Piezoresistivity model, and the stress-induced electron mobility model. The carrier generation-recombination models used are the Shockley– Read–Hall recombination model, and Auger recombination model. As well, the doping induced bandgap narrowing model, the intrinsic density model, and the effective density of states model have been employed. The model parameters used in TCAD

simulations have been calibrated in collaboration with Universität der Bundeswehr München using Monte Carlo simulations.

References

[1] G.A. Armstrong, C.K. Maiti, *Technology Computer Aided Design for Si, SiGe and GaAs Integrated Circuits*. Inspec/Iee, 2008.

[2] J. Vobecky, J. Voves, P. Hazdra, I. Adamcik, "TCAD - A Progressive Tool for Engineers" vol. 2, no. 1, p. 6, 1993.

[3] Synopsys, "Sentaurus Process User Guide" Dec-2007.

[4] Synopsys, "Synopsys Datasheet" 2009.

[5] Synopsys, "Sentaurus Device UserGuide" Dec-2007.

[6] W.C. O'Mara, R.B. Herring, L.P. Hunt, *Handbook of Semiconductor Silicon Technology*. William Andrew Publishing, 1990.

[7] Simon M. Sze, Kwok K. Ng, *Physics of Semiconductor Devices*. Wiley-Interscience; 3 edition, 2006.

[8] Dr. Vasileska, S. M. Goodnick, *Computational Electronics*. Morgan & Claypool, 2006.

[9] D.B.M. Klaassen, "A unified mobility model for device simulation I. Model equations and concentration dependence" vol. 35, p. 935, 1992.

[10] D.B.M. Klaassen, "A unified mobility model for device simulation II. Temperature dependence of carrier mobility and lifetime" vol. 35, p. 961, 1992.

[11] Y. Sun, S. E. Thompson, T. Nishida, *Strain Effect In Semiconductors: Theory And Device Applications*. Springer, 2009.

[12] Y. Sun, S. E. Thompson,T. Nishida, "Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors" vol. 101, p. 104503, 2007.

[13] M. Cardona, F. H. Pollak, "Energy-Band Structure of Germanium and Silicon: The k·p Method" vol. 142, pp. 530-543, 1966.

[14] J.L. Egley, D. Chidambarrao, "Strain effects on device characteristics: Implementation in drift-diffusion simulators" vol. 36, pp. 1653-1664, 1993.

[15] C. Jungemann, Institut fuer Mikroelektronik und Schaltungstechnik (EIT4.1), Bundeswehr University, private communication. .

[16] V. Palankovski, G. Röhrer, T. Grasser, S. Smirnov, H. Kosina, S. Selberherr, "Rigorous modeling approach to numerical simulation of SiGe HBTs" vol. 224, p. 361, 2004.

[17] C. Jungemann, B. Neinhüs, B. Meinerzhagen, "Comparative study of electron transit times

evaluated by DD, HD, and MC device simulation for a SiGe HBT" vol. 48, p. 2216, 2001.

[18] N. Rinaldi, G. Sasso, University of Naples, private communications, University of Naples.

[19] S. Decoutere, A. Sibaja-Hernandez, "IMEC private communication" 2008.

CHAPTER 4

TCAD Simulation Results

In advanced semiconductor devices technology, strain engineering technology can be used as an additional degree of freedom to enhance the carriers transport properties due to band structure changes and mobility enhancement. The mobility of charge carriers in bipolar devices can be enhanced by creating mechanical tensile strain in the direction of electrons flow to improve electrons mobility, and by creating mechanical compressive strain in the direction of holes flow to improve holes mobility. The compressive and tensile strains are created through various methods and techniques such as Global Strain, Local Strain and Mechanical Strain. A detailed description of each technique is presented in chapter 2.

In this work mainly two approaches have been used to create the desired mechanical strain inside the device. The first approach is through introducing strain engineering technology principle at the device base region using SiGe extrinsic stress layer. The second approach is through introducing strain engineering technology principle at the device collector region by means of local strain technique using strain layers. However, another approach to create strain inside the device is through using nitride liners, but this technique was not functional to create the desired strain inside the device, therefore no further work was done on it. In what follows, a detailed description of the main approaches used in this study and the obtained results are presented.

Strain Technology at the Base Region

1. NPN-Si-BJT Device with Extrinsic Stress Layer

The impact of introducing a SiGe stress layer formed over the extrinsic base layer, and adjacent to the intrinsic base of NPN-Si-BJT device on the electrical properties and frequency response has been investigated using TCAD modeling [1][2]. Process simulations are performed using Sentaurus TCAD software tools to build the device structure, and to calculate the associated mechanical stress due to the existence of the extrinsic stress layer using anisotropic elasticity model [3]. In what follows, the major processing flow steps are described. The process simulation starts by the fabrication of shallow (STI) and deep trenches (DTI), then the trenches are filled with silicon oxide to the same level as the surface of the doped silicon substrate as shown in Fig.64. A layer of etch-stop material (silicon oxide) and a thin layer of polysilicon are deposited. The polysilicon layer is then etched using selective etching technique as shown in Fig.65. This is followed by the deposition of a silicon layer to form the device base region as shown in Fig.66. Next, a thin layer of oxide and a layer of nitride are deposited as shown in Fig.67.

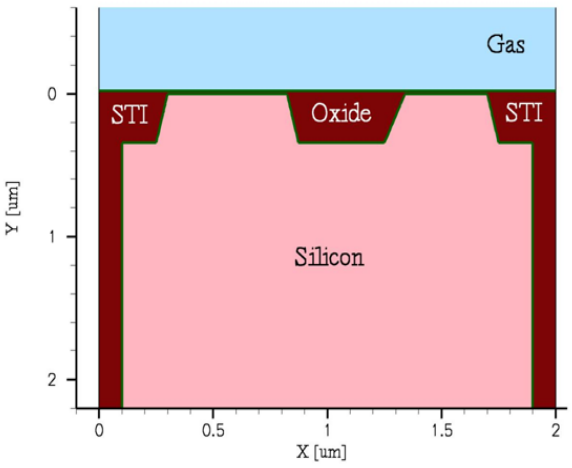

Fig.64: Process simulation: Fabrication of shallow and deep trenches isolation.

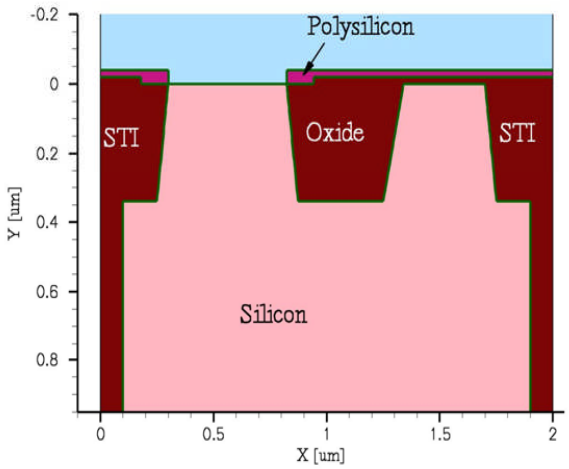

Fig.65: Process simulation: Deposition of etch-stop material and a thin layer of polysilicon.

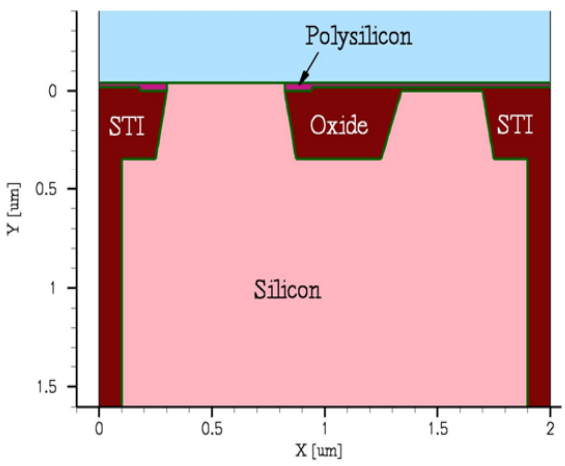

Fig.66: Process simulation: Formation of the device base region.

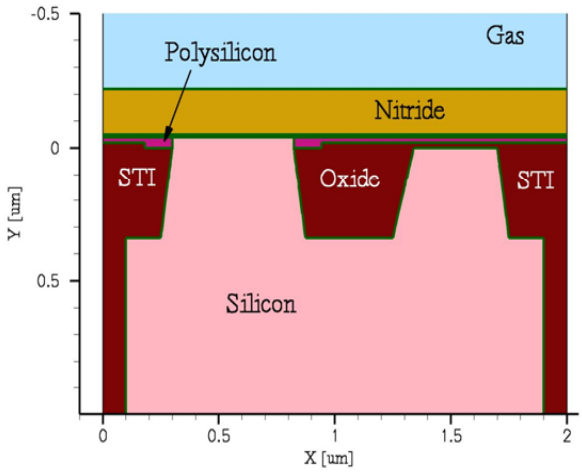

Fig.67: Process simulation: Deposition of oxide and nitride layers.

After that, the nitride and the oxide layers are etched using selective etching technique to form the emitter mandrel structure which is centered over the intrinsic base. The extrinsic base is then etched, resulting in a thinned extrinsic base and recess with a dimension of approximately 10 nm as shown in Fig.68. A SiGe layer is then deposited to form the extrinsic stress layer in the structure. The stress layer is grown up to the same height of the oxide and partially embedded into the intrinsic base as shown in Fig.69.

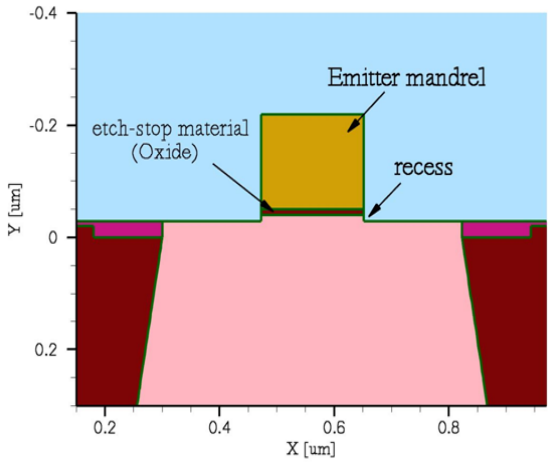

Fig.68: Process simulation: Formation of the emitter mandrel and recesses.

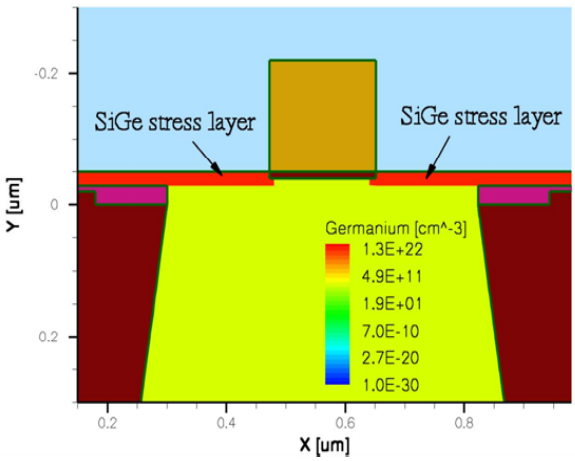

Fig.69: Process simulation: Deposition of the SiGe extrinsic stress layer

An oxide layer is then deposited to the same level of the emitter's mandrel nitride. The nitride of the emitter mandrel is then removed using selective etching technique, and using the underlying oxide layer as an etch-stop material resulting in an emitter opening. The opening extends downward to the oxide, and nitride spacers are formed on the sidewalls of the opening. After that, the underlying oxide is etched also by selective etching technique to expose the base layer in the emitter opening as shown in Fig.70. This is followed by the deposition of polysilicon and a hard mask of nitride. The nitride hard mask is etched and used to etch the polysilicon resulting in a T-shape emitter. Finally, the oxide is etched from all but under the overhanging portion of the emitter structure, and the contacts are formed using a proper technique as shown in Fig.71.

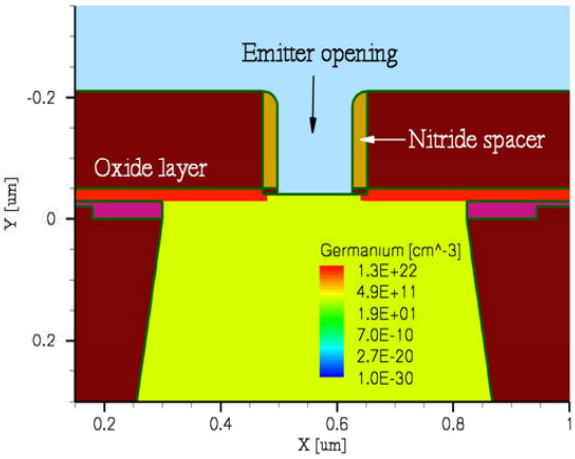

Fig.70: Process simulation: Formation of the emitter opening and the sidewall spacers.

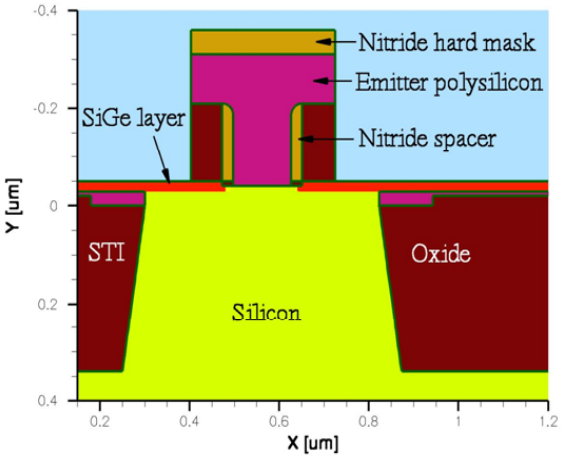

Fig.71: Process simulation: Formation of the T-shape emitter.

For simulation efficiency and saving of simulation resources, only half of the device is used for further device simulations as shown in Fig.72 [4]. The simulated effect of interposing the extrinsic stress layer at the device is shown in Fig.73; the isocontour lines represent the stress values generated in the x-direction (S_{xx}) and y-direction (Syy) of the device due to the existence of the extrinsic SiGe stress layer.

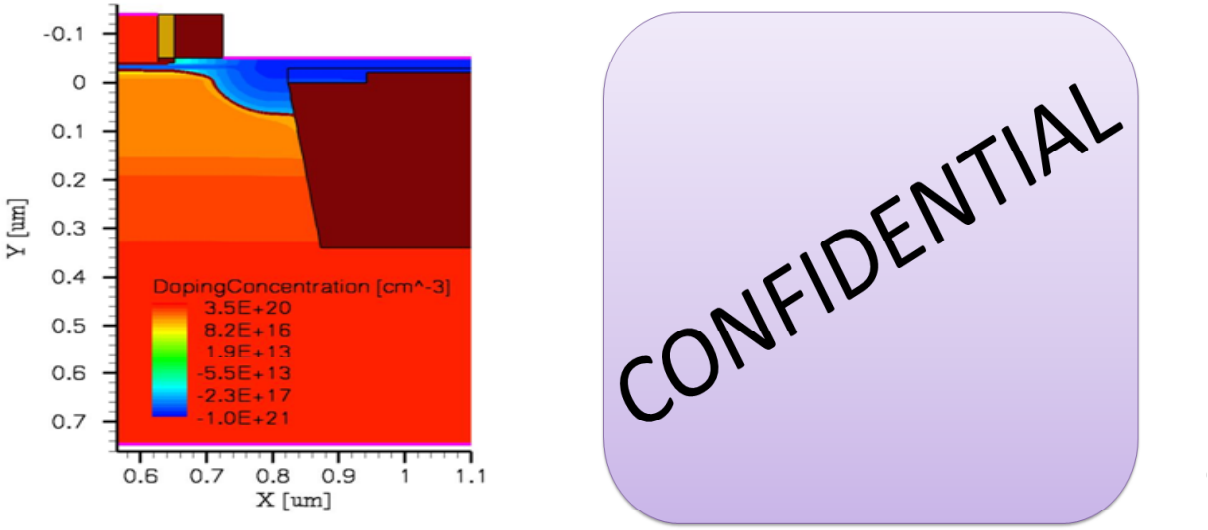

Fig.72: Cross-section of one half of the device (left), and the device doping profile (right) [4].

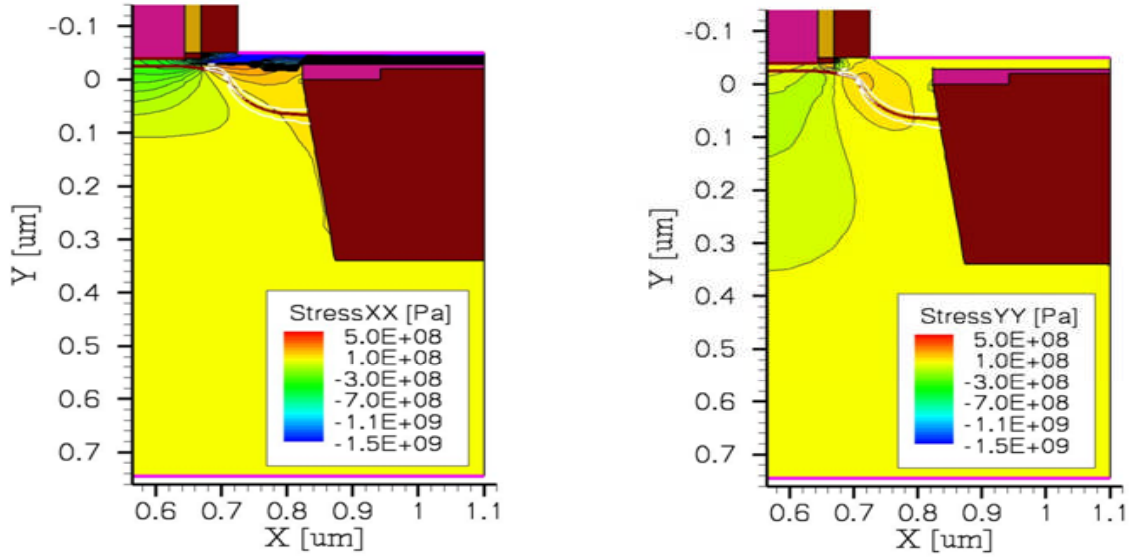

Fig.73 : Cross section of one half the device region of interest, the isocontour lines represent the stress Sxx (left) induced in the x-direction and Syy (right) induced in the y-direction of the structure.

1.1 Impact of Strain

Mechanical stress on semiconductors induces a change in the band structure and this in turn affects the carriers mobility. This effect can be explained by the deformation potential theory [5]. Strain changes the number of carrier sub-valleys and eventually a change in the actual band gap in the material [6]. The carrier redistribution that takes place between the various sub-valleys causes the change in mobility. The mobility enhancement is attributed to the increase in the occupancy of the conduction band valleys [7]. Consequently, incorporating an extrinsic SiGe stress layer at the bipolar device will create tensile strain in the direction of electrons flow to improve electrons mobility, and compressive strain in the direction of holes flow to improve holes mobility [8] . This process will decrease the intrinsic base resistance through the enhanced hole mobility, resulting in an improvement in the maximum oscillation frequency of the device according to the relation

$$f_{MAX} = \sqrt{\frac{f_T}{8\pi R_B C_{CB}}} \tag{4.1}$$

Where f_T is the cut-off frequency, R_B is the base resistance, and C_{CB} is the collector base capacitance.

On the other hand, the vertical tensile strain in the direction of electrons flow under the emitter will enhance the device electrons flow. Furthermore, reduces the electrons transit time through the enhanced electron mobility due to the applied strain, which will improve the cut off frequency according to the relations [9].

$$\tau_{Bf} = \frac{dQ_{nB}}{dI_T} = \frac{w_B^2}{\mu_{nB} V_T} \frac{(\zeta - 1) f_\zeta + 1}{\zeta^2 f_\zeta} + \frac{w_B}{v_c} \frac{f_\zeta - 1}{\zeta f_\zeta} \tag{4.2}$$

$$f_T = \frac{1}{2\pi\, \tau} \tag{4.3}$$

Where τ_{Bf} is the base transit time, Q_{nB} is the minority charge stored in the neutral base region, I_T is the time dependent quasi-static transfer current, w_B is the base width, μ_{nB} is the average electron minority mobility in the neutral base region, V_T is the thermal voltage, ζ is the drift factor, $f_\zeta = \exp(\zeta)$ is the drift function, v_c is the electron velocity at the end of the neutral base region and τ is the total transit time.

Additionally, introducing an extrinsic stress layer at the device base region will decrease the bandgap energy through the reduction of the conduction band energy as shown in Fig.74. This in turn will improve electrons injection efficiency from emitter to collector, and enhance the device electrical performance. As well, the applied strain will induce a change in the band structure and this in turn affects the carrier mobility, resulting in an approximately 27% of mobility improvement in YY direction for electrons at the base region as shown in Fig.75.

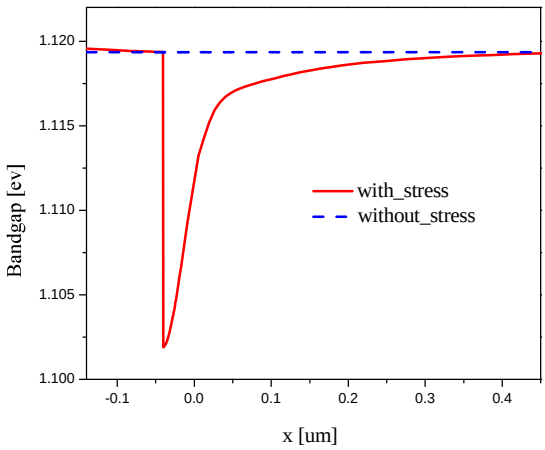

Fig.74: Impact of strain on the bandgap energy.

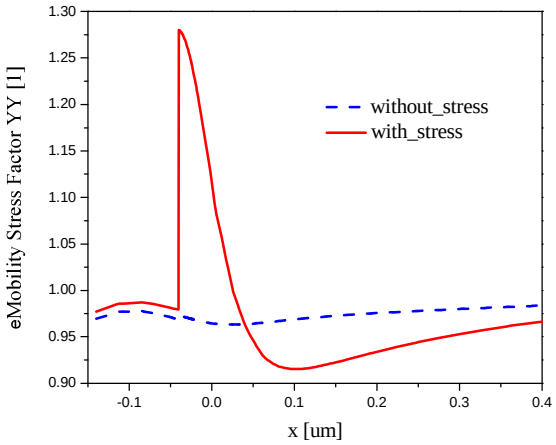

Fig.75: Electron mobility enhancement due to the applied strain.

1.2 Electrical Simulation

Sentaurus TCAD software tools have been used to perform the two-dimensional device simulations using hydrodynamic transport model (HD), where the carrier temperature equation for the dominant carriers is solved together with the electrostatic Poisson's equation and the carrier continuity equations [10]. All the standard silicon models, such as Philips unified mobility model, high field saturation mobility model, Shockley-Read-Hall recombination model, Auger recombination model, piezoresistive model for calculating mobility enhancement due to the applied stress, bandgap narrowing model and default parameter files, are all included in the simulation file. The doping profiles at the emitter, the base and the collector region of the device have been taken from IMEC Microelectronics bipolar device profile [4].

Fig.76 shows the forward Gummel plots obtained by simulating both NPN-Si-BJT incorporating SiGe extrinsic stress layer at the base region, and a standard conventional NPN-Si-BJT device.

Simulation results show that introducing the extrinsic stress layer in the device will increase the collector current by almost three times, resulting in an enhancement of the maximum current gain (β_{max}) in comparison with the standard conventional one. The transit frequency (f_T) and the maximum oscillation frequency (f_{MAX}) are plotted versus the collector current (Ic) for both structures as shown in Fig.77 and Fig.78.

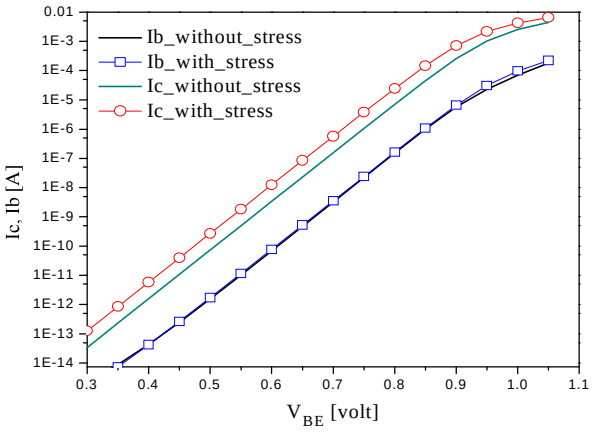

Fig.76: Comparison of forward Gummel plots for both conventional BJT, and BJT with stress layer (Ge=25%, W_E=130 nm).

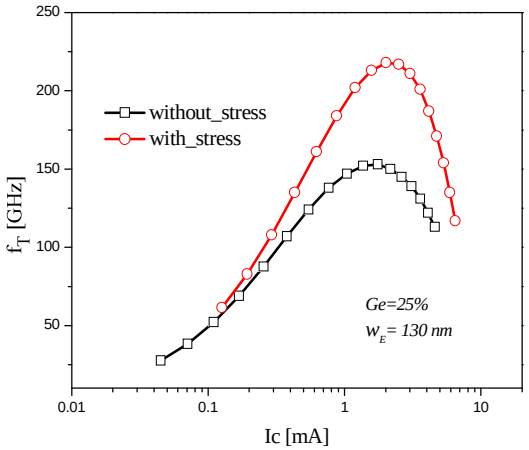

Fig.77: Cut-off frequency as a function of collector current for both devices (Ge=25%, W_E=130 nm).

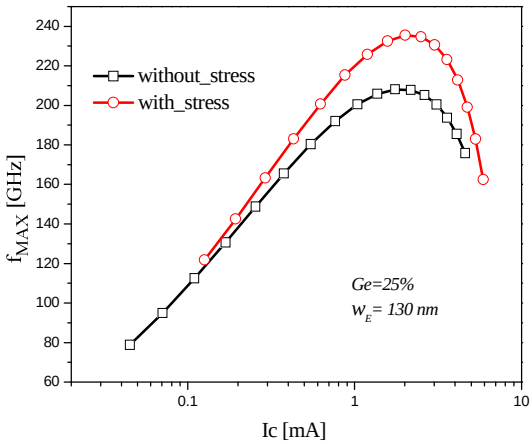

Fig.78: Maximum oscillation frequency as a function of collector current for both devices (Ge=25%, W_E=130 nm).

The results obviously show that bipolar device with extrinsic stress layer exhibit better high frequency characteristics in comparison with an equivalent standard conventional device. An approximately 42% of improvement in f_T, and 13% of improvement in f_{MAX} in NPN-Si-BJT with extrinsic stress layer have been achieved. These improvements are mainly due to the enhanced vertical electron mobility, which can be fully accounted to the impact of interposing the extrinsic stress layer in the device.

The impact of changing the Ge content at the extrinsic stress layer on the stress values generated inside the device (i.e, Sxx, Syy, and Szz), β_{MAX}, f_{MAX} and f_T has been studied, the obtained results are shown in Fig.79 to Fig.81. As shown in the figures, increasing the Ge content at the extrinsic stress layer will increase the stress values generated inside the device, which in turn improves the high frequency characteristics of the device and enhances the current gain. These improvements are related to the increase of the lattice constant difference between the silicon substrate and the SiGe stress layer, which will increase the stress values induced at the base, resulting in a decrease of the conduction band energy and hence the total bandgap energy. This decrease in the bandgap energy will improve the electrons injection efficiency from emitter to collector, consequently enhancing the high frequency characteristics of the device. Unfortunately, increasing the Ge content at the stress layer will also increase the misfit dislocations between the silicon substrate and the stress layer, which may cause a degradation of the device performance. Therefore, the Ge content at the stress layer must be controlled and chosen carefully to avoid such problems.

Moreover, the impact of changing the device's emitter width on the device performance has been studied. The result illustrates that increasing the emitter width will decrease the stress values induced at the base region of the device as shown in Fig.82, causing a degradation of the device performance.

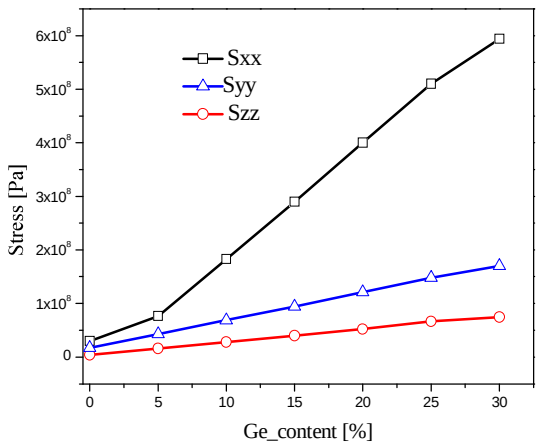

Fig.79: Variation of the stress values generated inside the device with Ge content at the stress layer (W_E=130 nm).

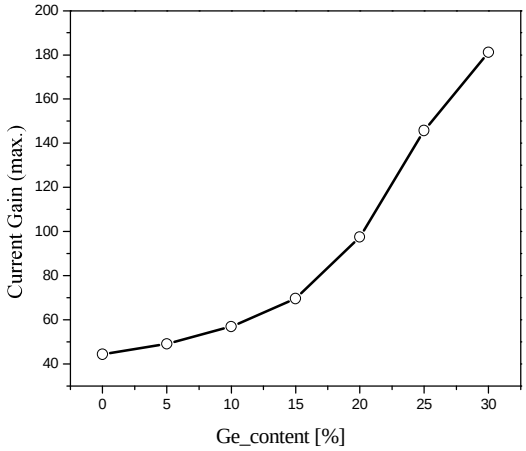

Fig.80: Variation of the maximum current gain with Ge content at the stress layer (W_E=130 nm).

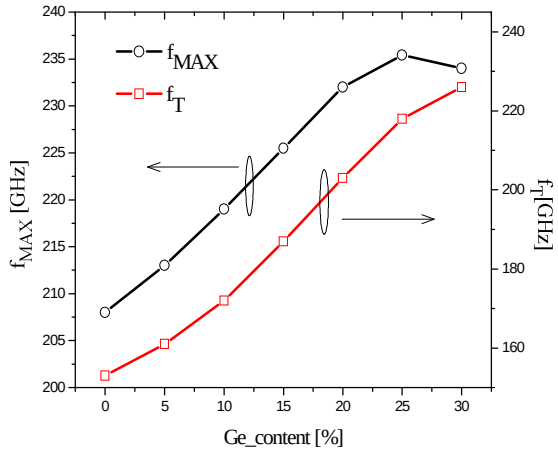

Fig.81:Variation of f_T and f_{MAX} with Ge content at the stress layer (W_E=130 nm).

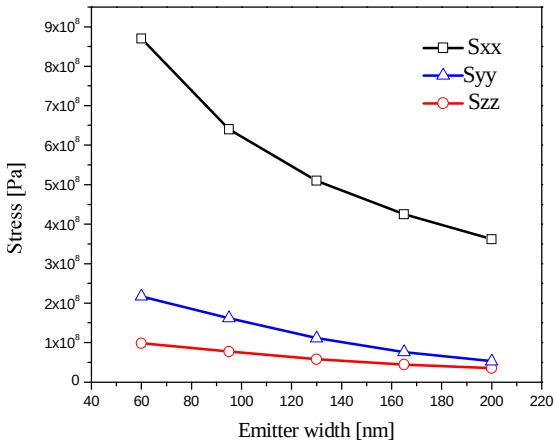

Fig.82: Variation of the stress values generated inside the device with the device's emitter width.

1.3 Conclusion

Simulation results demonstrate that Si bipolar devices with extrinsic stress layer exhibit better high frequency characteristics, and an enhancement of the maximum current gain in comparison with an equivalent standard conventional BJT device. An approximately 42%

improvement in f_T and 13% improvement in f_{MAX} have been achieved in NPN-Si-BJT with extrinsic stress layer at the base region. Furthermore, an enhancement of the collector current by almost three times, and an enhancement of the maximum current gain β_{MAX} in NPN-Si-BJT with extrinsic stress layer have been found. These improvements are related to the electron and hole mobility's enhancement, and to the decrease of the bangap energy. This in turn improves electron injection efficiency from emitter to collector, and improves the whole device electrical performance.

2. NPN-SiGe-HBT Device with Extrinsic Stress Layer

The higher gain, speed and frequency response of the SiGe-HBT make silicon-germanium devices more competitive in areas of technology where high speed and high frequency response are required. However, due to the continuous demand for such devices it becomes imperative to develop new bipolar device architectures suitable for high frequency and power applications. Therefore, various techniques and efforts have been proposed to improve the performance of HBT devices through grading germanium profile at the base [11], introduction of carbon to improve 1D doping profile [12], and reduction of the emitter width [13]. An additional approach to improve the device performance is to enhance the carrier transport by changing the material transport properties by means of strain engineering technology. [14][15]. In what follows the impact of introducing a SiGe extrinsic stress layer formed above the extrinsic base layer, and adjacent to the intrinsic base of NPN-SiGe-HBT device on the electrical properties and frequency response will be presented.

Process simulations have been performed using Sentaurus TCAD software tools to build the device structure and to calculate the associated mechanical stress. The major processing steps of the NPN-SiGe-HBT device architecture are similar to those of NPN-Si-BJT device described previously except that we have a SiGe-base in this case. The complete HBT device structure with SiGe extrinsic stress layer is shown in Fig.83. Likewise, for simulation efficiency and saving simulation resources, only half of the device is used for further device simulations as shown in Fig.84. The graded Ge profile and the doping profiles at the emitter, the base and the collector regions have been taken from IMEC Microelectronics HBT device profile [4].

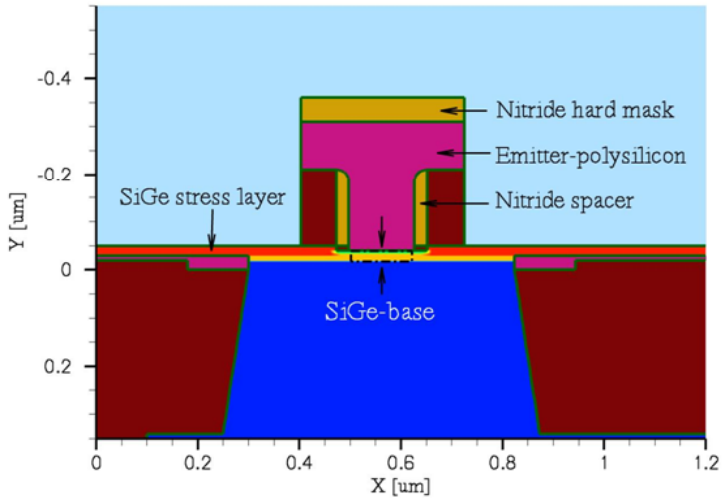

Fig.83: The complete structure of the NPN-SiGe-HBT device with extrinsic SiGe stress layer.

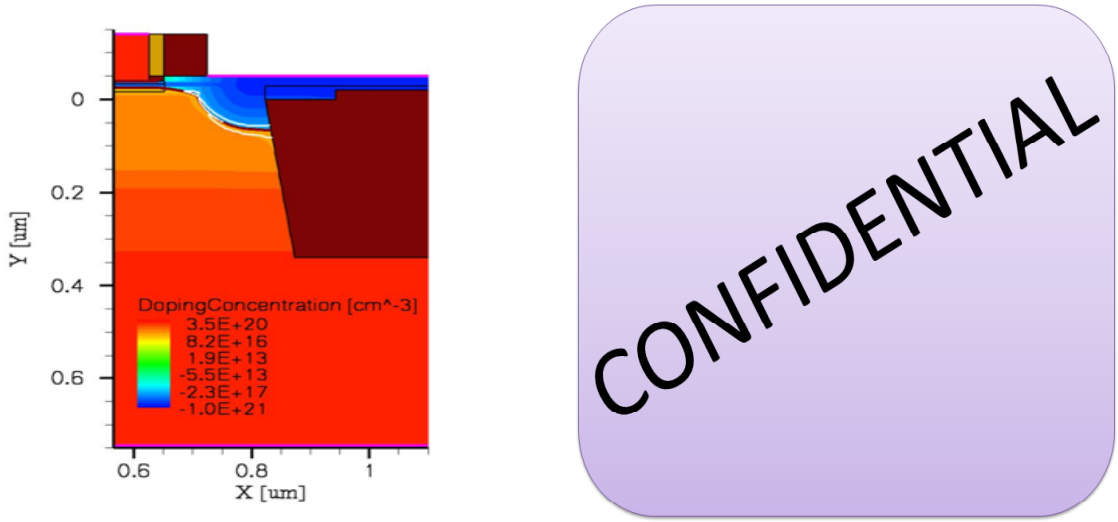

Fig.84: Cross-section of one half of the device (left), and the graded Ge profile, and the doping profile at the emitter, base and collector regions (right) [4].

2.1 Impact of strain

Due to the addition of the extrinsic stress layer, stress is generated inside the device (i.e, Sxx) as shown in Fig.85; these values are extracted at the middle of the base. An approximately 500 MPa of an additive compressive stress (Sxx) is generated at the base region, and 500 MPa of tensile stress (Sxx) is lessen at the collector region as shown in the figure. Chapter.2 provides a detailed discussion of the impact of strain on the bandgap energy and the carrier mobility.

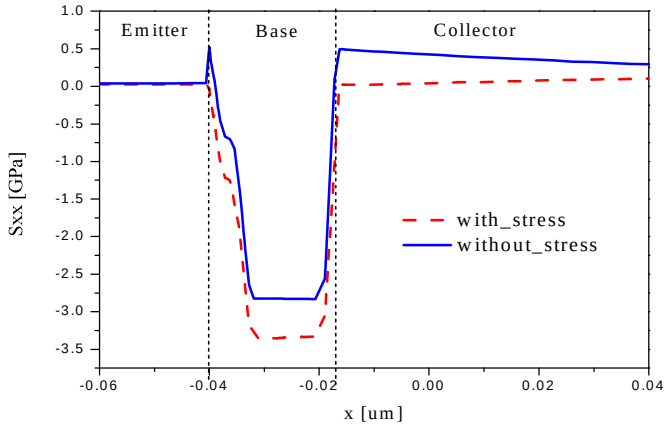

Fig.85: Stress values (Sxx) generated inside the device due to the addition of the extrinsic stress layer.

2.2 Electrical Simulation

Sentaurus TCAD software tools have been used to perform the two-dimensional device simulations using HD transport model. The standard silicon models and parameter files included in TCAD software library cannot be used for the SiGe device with external stress. Therefore, specific parameter files and physical models have been calculated by Universität der Bundeswehr München (BU) using Monte Carlo simulations.

To acquire accurate simulation results, the HBT device has been divided into two regions; The first one is the SiGe-base which is divided into two zones corresponding to different Ge content and doping concentration values as shown in Table 8 and Table 9. The second region is

the remaining part of the device without Ge content (the emitter and the collector regions). The different device regions and zones are shown in Fig.86.

Zone (I)	S_{xx} [Pa]	S_{yy} [Pa]	S_{zz} [Pa]	Boron Active Con. [cm^{-3}]
With_Stress	-1.23444e+9	-1.74456e+7	-1.2265e+9	7.2e+18
Without_Stress	-9.044e+8	585458	-1.1999e+9	7.2e+18

Table 8: The doping concentration and stress values generated inside Zone (I) (corresponds to Ge concentration = 7.6%, and x =0.364 um).

Zone (II)	S_{xx} [Pa]	S_{yy} [Pa]	S_{zz} [Pa]	Boron Active Con. [cm^{-3}]
With_Stress	-3.3417+9	-2.37776e+7	-3.34988e+9	2e+15 - 5.2 e+19
Without_Stress	-3.028e+9	2.32544e+6	-3.32093e+9	2e+15 - 5.2 e+19

Table 9: The doping concentration and stress values generated inside Zone (II) (corresponds to Ge concentration = 20.9%, and *x* =0.0265792 um).

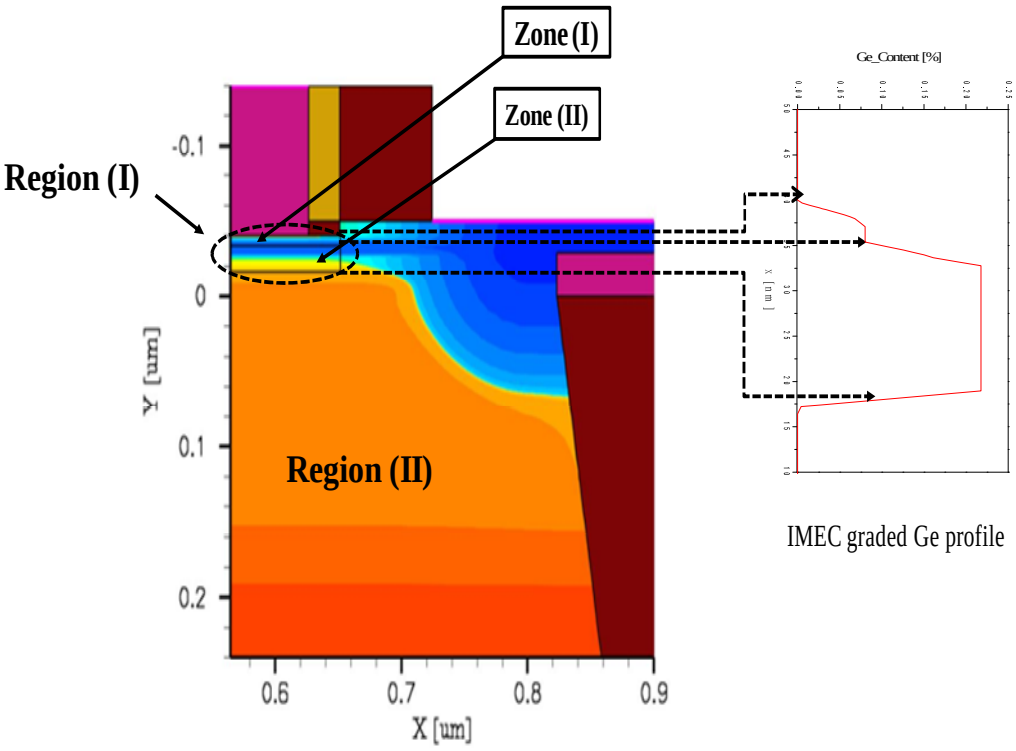

Fig.86: Different regions and zones used for device simulations.

For these regions, specific models for SiGe band-gap, bandgap-narrowing, effective mass, energy relaxation, mobility for hydrodynamic simulation and drift-diffusion simulation have been calculated by BU and implemented in our simulator using tabulated models compiled in C code. Two methods have been used by BU to calculate the impact of stress on the bandgap energy: the first approach is based on Analytical Method, while the second approach uses the Empirical Pseudopotential Method (EPM). The results are summarized in Table 10 and Table 11. However, the results obtained using EPM are more reliable as they are based on a more realistic band structure. Therefore, the models and parameter files calculated using EPM approach are used in further device simulations.

Zone (I)	Eg_Analytical [eV]	Eg_EPM [eV]
With_Stress	1.0251	0.988
Without_Stress	1.0136	0.996

Table 10: Bandgap energy results obtained by BU for zone(I)

Zone (II)	Eg_Analytical [eV]	Eg_EPM [eV]
With_Stress	0.83	0.867
Without_Stress	0.8189	0.8754

Table 11: Bandgap energy results obtained by BU for zone(II)

The cut-off frequency (f_T) and the maximum oscillation frequency (f_{MAX}) are plotted versus the collector current (Ic) using the physical models and parameters provided by Bundeswehr University. Simulation results show that introducing extrinsic stress layer on the HBT device structure will enhance the maximum oscillation frequency f_{MAX}, and the cut-off frequency f_T for both bandgap calculation approaches as shown in the following figures. An approximately 5% of improvement in f_{MAX}, and 3% of improvement in f_T have been achieved in comparison with the standard conventional HBT device (without extrinsic stress layer).

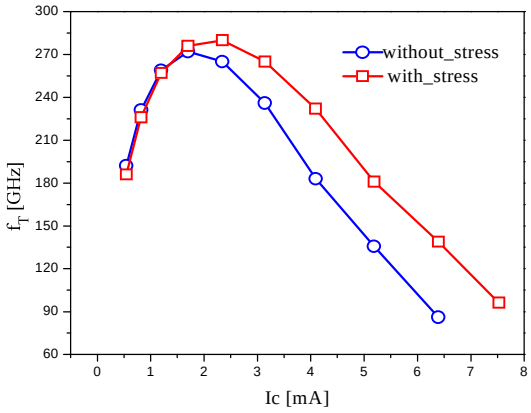

Fig.87: Improvement in f_T due to the addition of the extrinsic stress layer (EPM).

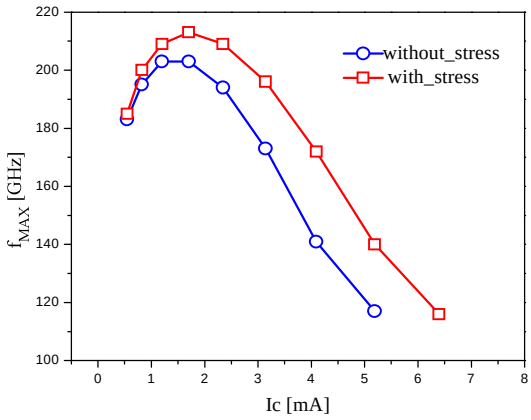

Fig.88 Improvement in f_{MAX} due to the addition of the extrinsic stress layer (EPM).

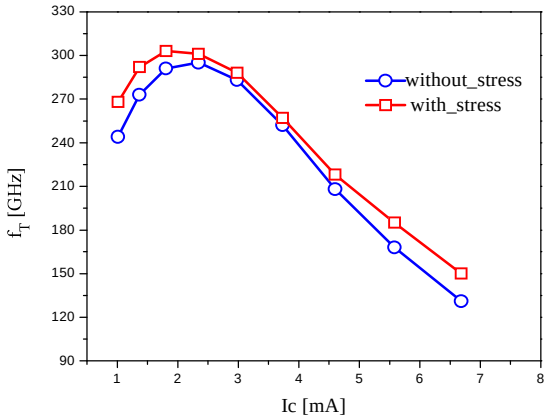

Fig.89: Improvement in f_T due to the addition of the extrinsic stress layer (Analytical Method).

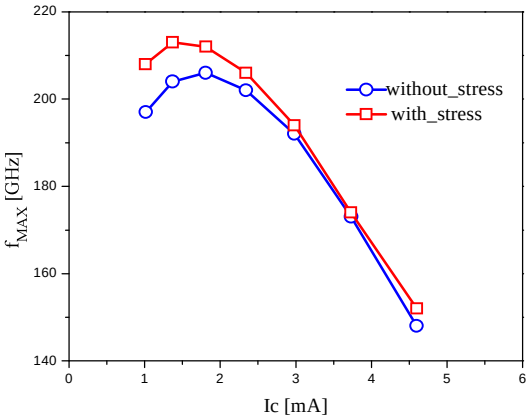

Fig.90: Improvement in f_{MAX} due to the addition of the extrinsic stress layer (Analytical Method).

Furthermore, the components of the transit time have been extracted for both HBT devices as a function of the collector current as shown in Fig.91. Simulation results show that introducing the extrinsic stress layer on the device structure will decrease the total transit time, and hence improve the device performance. This reduction of the transit time is shown in Fig.92 and Fig.93 where the base transit time (t_b), and the collector transit time (t_c) have been plotted versus the collector current separately for illustration.

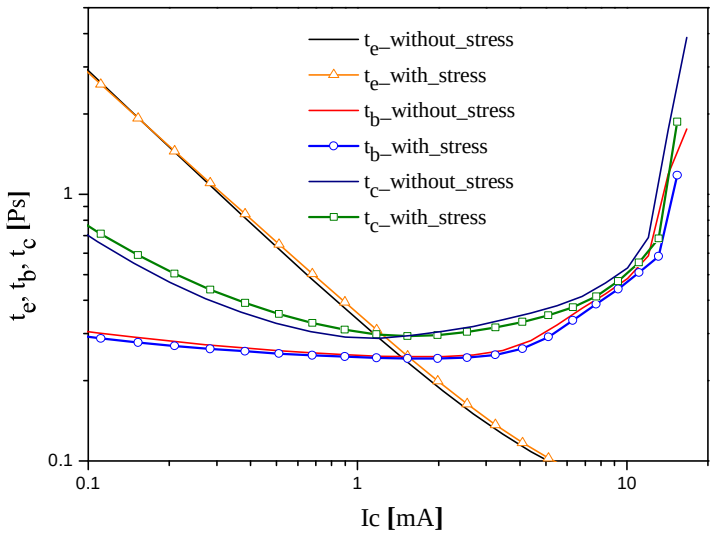

Fig.91:Transit time versus the collector current for HBT device with and without extrinsic stress layer.

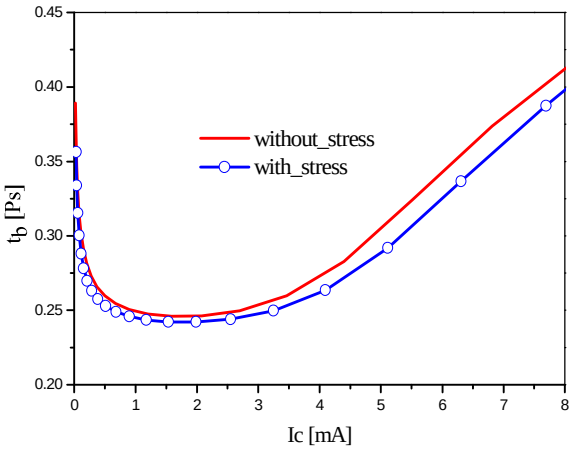

Fig.92: The base transit time versus the collector current for HBT device with and without extrinsic stress layer.

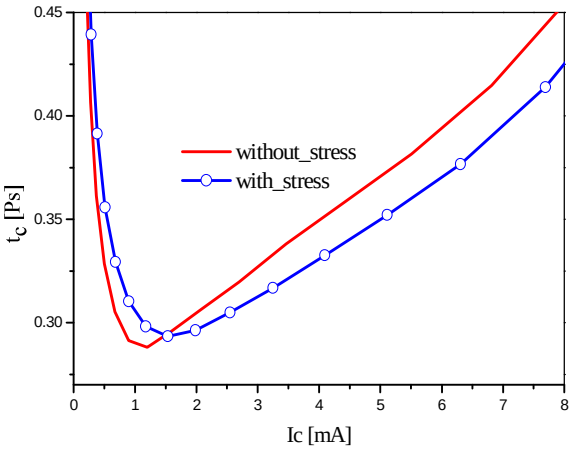

Fig.93: The collector transit time versus the collector current for HBT device with and without extrinsic stress layer.

The impact of changing the Ge content at the extrinsic stress layer on the stress values generated inside the device (Sxx, and Szz) has been studied; the results are shown in Fig.94. As shown, increasing the Ge content at the extrinsic stress layer will increase the stress values generated inside the device. This is related to the increase of the lattice constant difference between the SiGe base layer and the SiGe stress layer, which will increase the stress values induced at the base. Nevertheless, increasing the Ge content at the stress layer will also increase

the misfit dislocations between the two layers, which may cause a degradation of the device performance. Therefore, the Ge content at the stress layer must be controlled and chosen carefully to ensure device performance enhancement.

In addition, the impact of changing the device's emitter width on the device performance has been studied. The result shows that increasing the emitter width will decrease the stress values induced at the base of our device as shown in Fig.95, causing degrade action of the device performance.

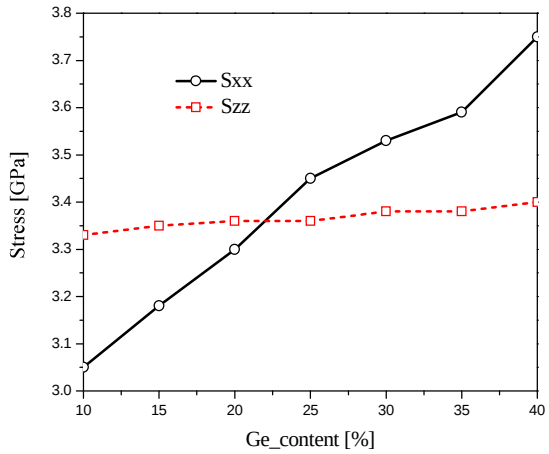

Fig.94: Variation of the stress values generated inside the device with Ge content at the stress layer (W_E=130 nm).

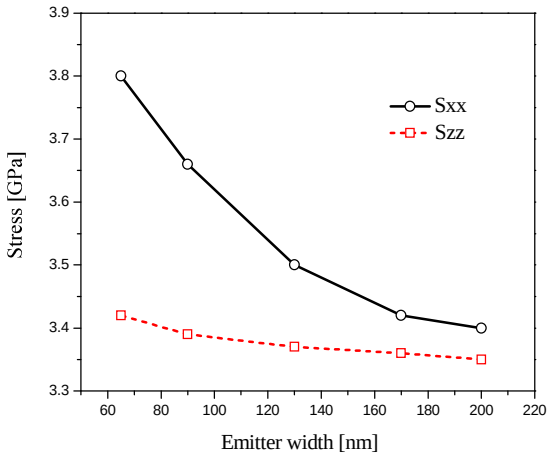

Fig.95: Variation of the stress values generated inside the device with the device emitter width.

2.3 Conclusion

Simulation results show that introducing SiGe extrinsic stress layer on NPN-SiGe-HBT device will enhance the maximum oscillation frequency f_{MAX} (~5%), the cut-off frequency f_T (~3%) and the total transit time in comparison with a standard conventional HBT device (without extrinsic stress layer). However, applying strain engineering technology at the base region of NPN-SiGe-HBT device using the extrinsic SiGe strain layer is less efficient in comparison with the obtained results for the NPN-Si-BJT device. This is related to the high sensitivity of silicon

material to strain in comparison with the SiGe base of NPN-SiGe-HBT device which is already stressed due to the existence of Ge content at the base.

Strain Technology at the Collector Region

Applying strain engineering technology at the base region of NPN-Si-BJT device by means of extrinsic SiGe strain layer can strongly enhance the device's performance due to the sensitivity of silicon material to strain. On the other hand, applying strain engineering technology at the base region of NPN-SiGe-HBT device through using SiGe strain layer is less efficient in comparison with the obtained results for the NPN-Si- BJT device, because the SiGe base is already stressed due to the existence of Ge at the base as presented previously. The intensive study of the transit time in the strained NPN-SiGe-HBT device shows that the main modification of the device's total transit time in comparison with the standard conventional NPN-SiGe-HBT device arises from the reduction of the collector transit time. Verifying that silicon material is more sensitive to strain than the SiGe base region, and the device's performance improvements are mainly due to the impact of the induced strain at the collector region.

Depending on the obtained results reported in [16], [17] and the sensitivity of silicon material to strain, new NPN-SiGe-HBT device's architecture employing strain engineering technology at the collector region will be presented. In this approach the desired strain is introduced during the device fabrication process through a specific device architecture employing silicon nitride (Si_3N_4) and silicon oxide (SiO_2) strain layers at the collector region.

3. NPN-SiGe-HBT Device Employing Si_3N_4 Strain Layer

In this approach the desired strain is generated inside the device through introducing strain engineering technology principle using local strain technique by means of introducing silicon nitride (Si_3N_4) strain layer at the collector region. Nitride films can induce stresses greater

than 1GPa upon thermal treatment, which arises from two sources: Coefficient of thermal expansion mismatch between silicon and nitride film, and intrinsic film stress caused by film shrinkage. The origin of intrinsic stresses comes from the energy configuration of the deposited atoms or ions. Processing conditions such as temperature, pressure, deposition power, reactant and impurity concentrations are important factors in determining the magnitude and strain type (i.e. compressive or tensile) [18] [19].

In what follows the impact of employing a nitride strain layer at the collector region on the device's performance parameters will be presented.

3.1 Process Simulation and Device Structure

Process simulations are performed using Sentaurus TCAD software tools to build the device structure and to calculate the associated mechanical stress. In what follows, the major processing steps will be described. The process simulation starts by the deposition of a silicon layer as shown in Fig.96. Then the silicon substrate is etched using selective etching technique as shown in 97. This is followed by deposition of a nitride layer as shown in Fig.98. Next the nitride layer is etched using selective etching technique, and a silicon layer is deposited, this is followed by selective etching of this layer to form the device intrinsic base region as shown in Fig.99.

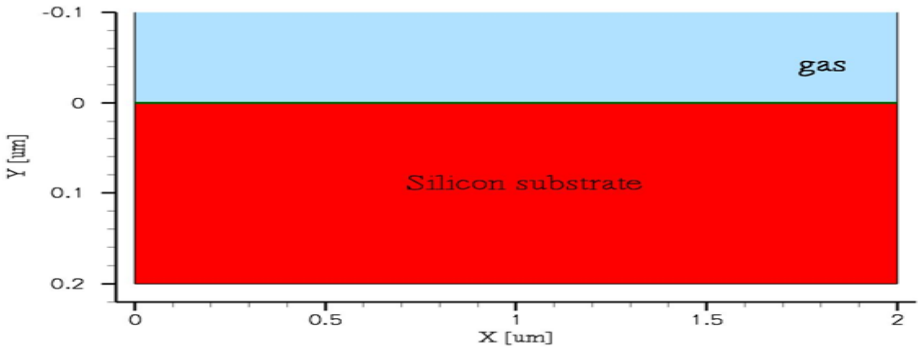

Fig.96: Process simulation: Deposition of silicon substrate.

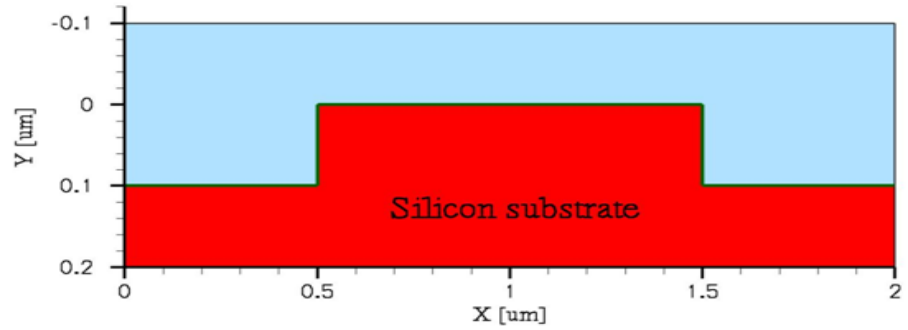

Fig.97**:** Process simulation: Etching of silicon substrate.

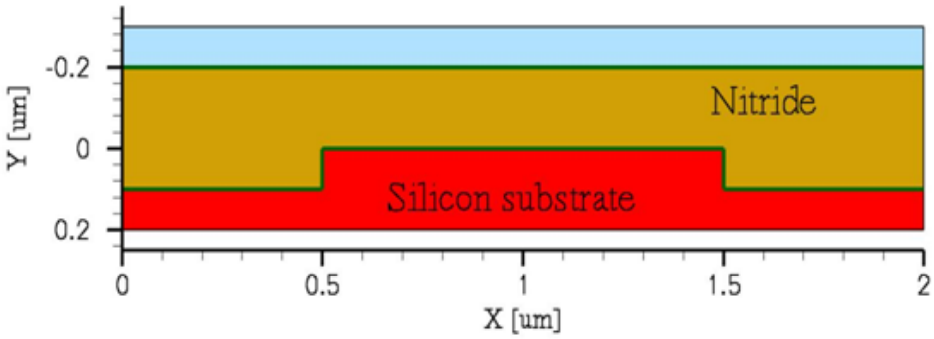

Fig.98**:** Process simulation: Deposition of nitride layer.

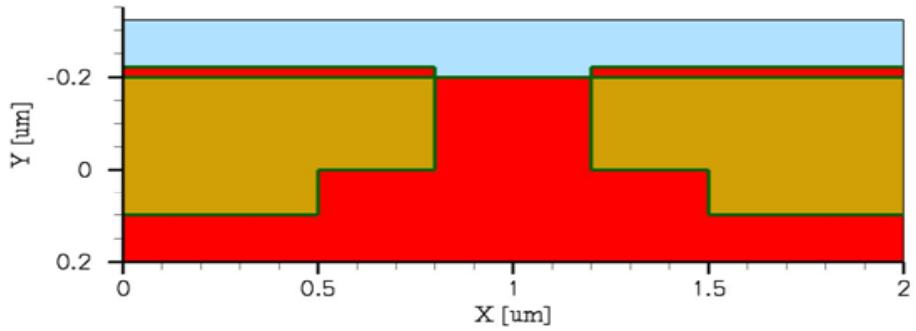

Fig.99: Process simulation: Nitride layer etching, and silicon substrate deposition.

After that, a layer of SiGe alloy is deposited to form the intrinsic base region with a graded Ge profile and doped to have a p-type conductivity as shown in Fig.100. A layer of oxide is deposited, and then etched using selective etching technique to form the emitter opening, followed by deposition of a polysilicon layer as shown in Fig.101.

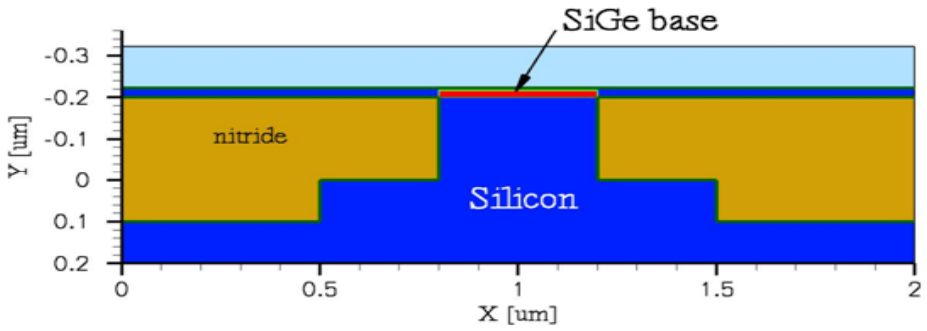

Fig.100: Process simulation: Deposition of the SiGe base.

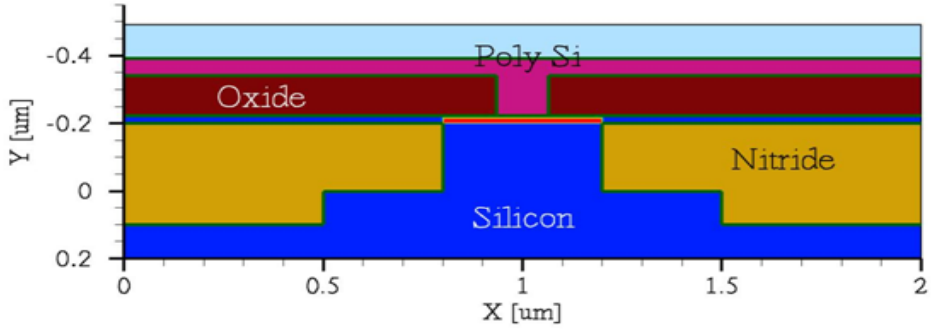

Fig.101: Process simulation: Deposition of oxide layer, formation of emitter opening, and deposition of polysilicon layer.

The oxide and polysilicon layers are then etched resulting in a T-shape emitter as shown in Fig.102. Finally the contacts are formed using a proper technique. The final device structure that illustrates the stress isocontour lines generated inside the device due to the existence of SiGe base, and the nitride strain layer at the collector region in our specific device architecture is shown in Fig.103.

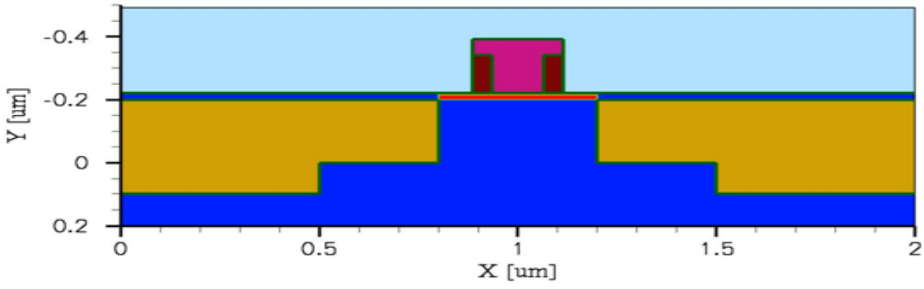

Fig.102: Process simulation: Etching of oxide and nitride layers, and formation of T-shape emitter.

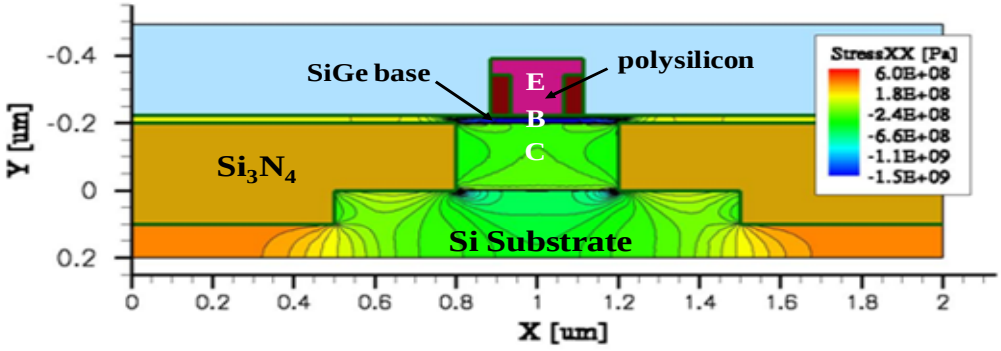

Fig.103: Process simulation: Final device structure showing stress isocontour lines generated inside the device.

For simulation efficiency and saving simulation resources, only half of the device is used for further device simulations due to the device symmetry. Fig.104 shows a comparison between the stress values generated inside the standard conventional HBT device, and the new HBT device architecture employing nitride strain layer at the collector region.

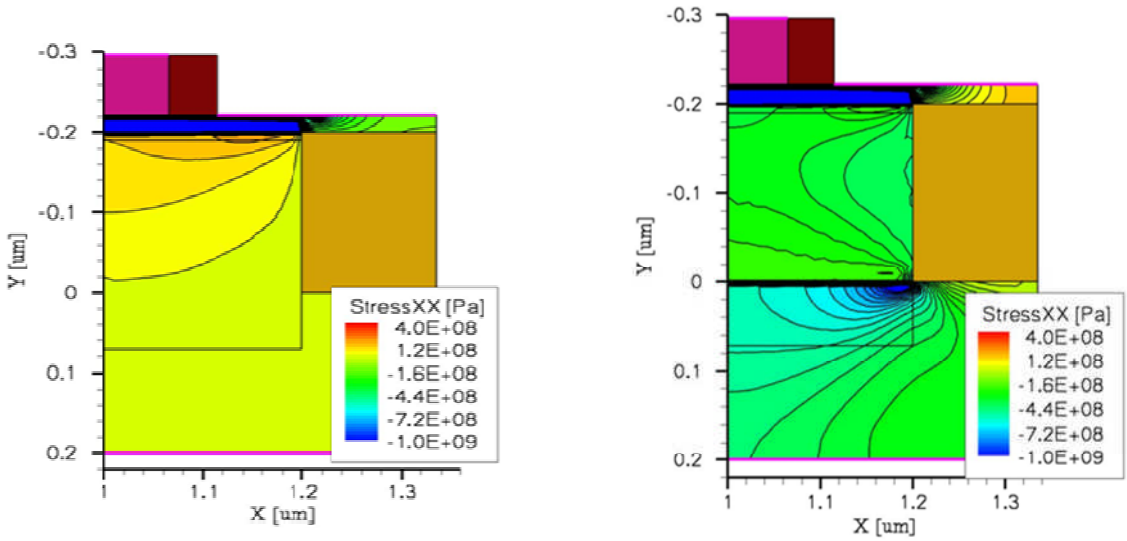

Fig.104: Isocontour lines representing the stress values generated inside the device; standard conventional NPN-SiGe-HBT device (left), and strained silicon NPN-SiGe-HBT (right).

3.2 Impact of Strain

The impact of introducing Si_3N_4 strain layer in the HBT device's architecture on the bandgap energy, and the stress values generated inside the device (Sxx, Syy, and Szz) are shown in Fig.105 to Fig.108. Moreover, the impact of Si_3N_4 strain layer on the carrier mobility is shown in Fig.109.

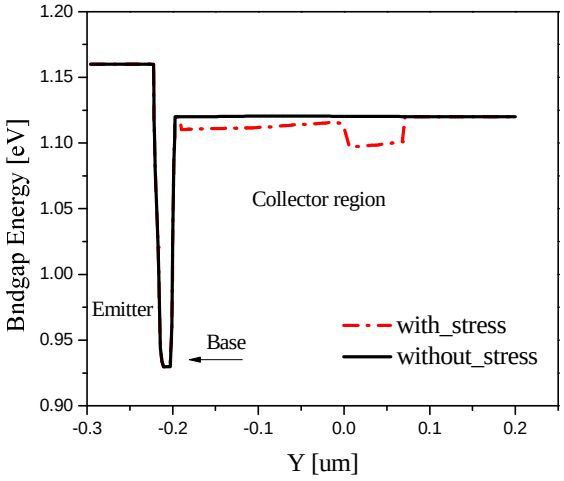

Fig.105: Impact of stress on the bandgap energy.

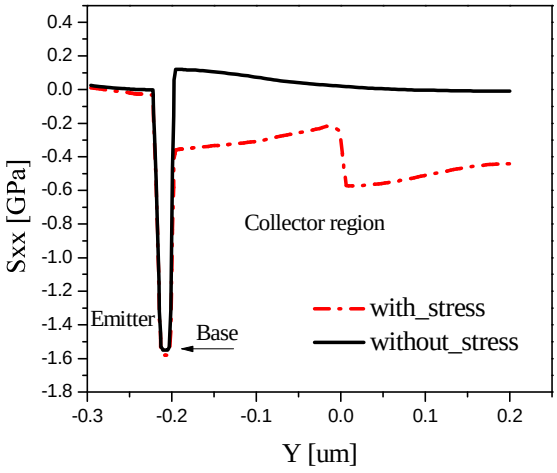

Fig.106: Impact of strain on Sxx.

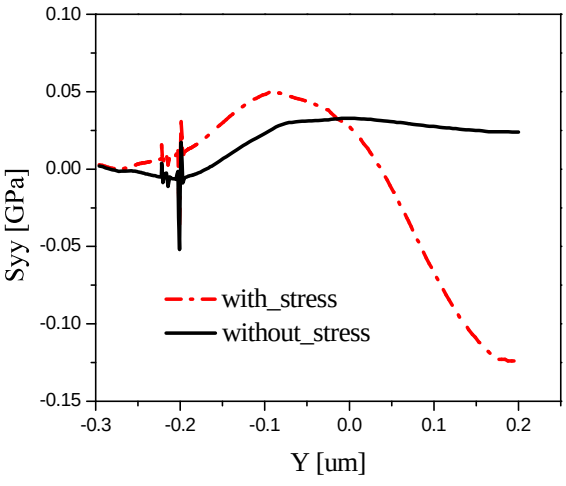

Fig.107: Impact of strain on Syy.

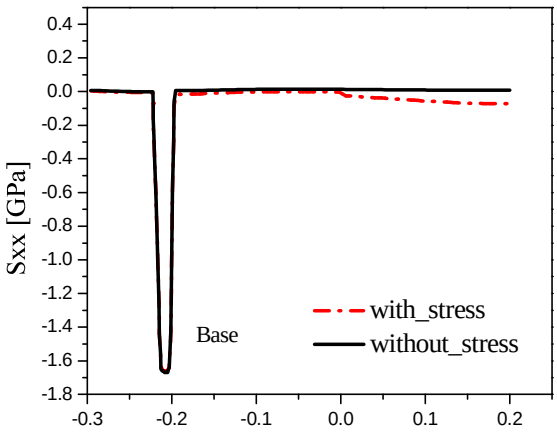

Fig.108: Impact of strain on Szz.

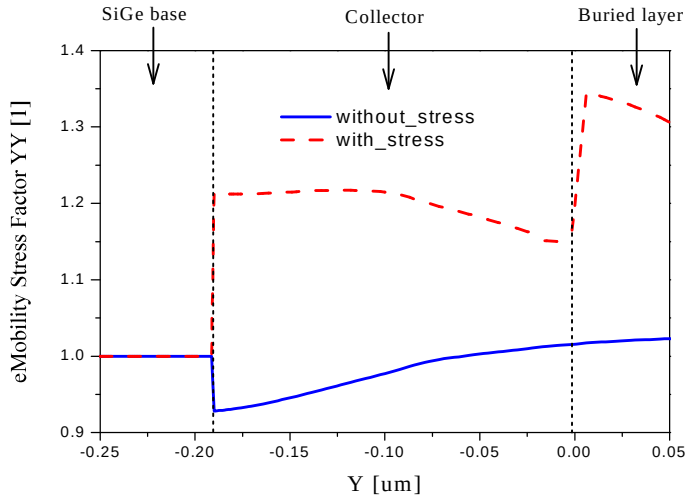

Fig.109: Electron mobility enhancement due to the applied strain.

As shown in the previous figures; introducing Si_3N_4 strain layer at the collector region will create compressive strain along the horizontal direction, and tensile strain along the vertical direction, which causes a reduction in the device's bandgap energy at the collector region. This is related to the reduction of the conduction band energy due to the applied strain. In addition, strain will induce a change in the band structure, and this in turn affects the carrier mobility, resulting in an approximately 20% of mobility improvement in YY direction for electrons in the neutral collector region.

3.3 Electrical Simulation

Sentaurus TCAD software tools have been used to perform the two dimensional device simulations using HD model without taking in consideration self heating effect [10]. The model parameters used in TCAD simulations have been calibrated by BU using Monte Carlo simulation. The stress-induced mobility enhancement has been calculated using the Piezoresistivity model. A detailed description of the models used in the simulation is presented in chapter 3.

A comparison of the forward Gummel plots of IMEC HBT device measurements and the standard conventional HBT device (without strain) simulation results are shown in Fig.110. The cut-off frequency f_T curves as a function of the collector current I_C for the standard NPN-SiGe-

HBT device simulation results, and IMEC NPN-SiGe-HBT device measurement are shown in Fig.111. Simulation results illustrated in Fig.110 and Fig.111 show good agreement between measurement data and simulation results, verifying the validity of the physical models and parameters used in the TCAD simulations.

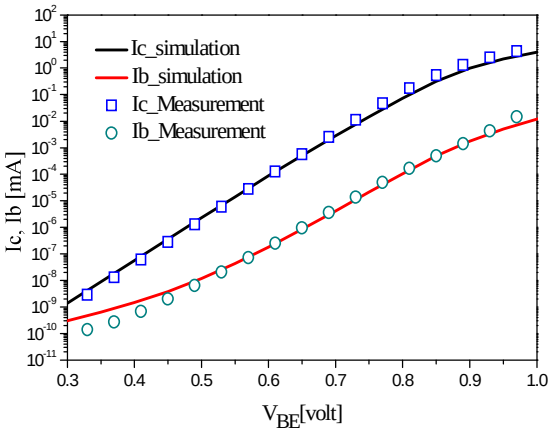

Fig.110: Forward Gummel plots comparison of measurement and simulation results.

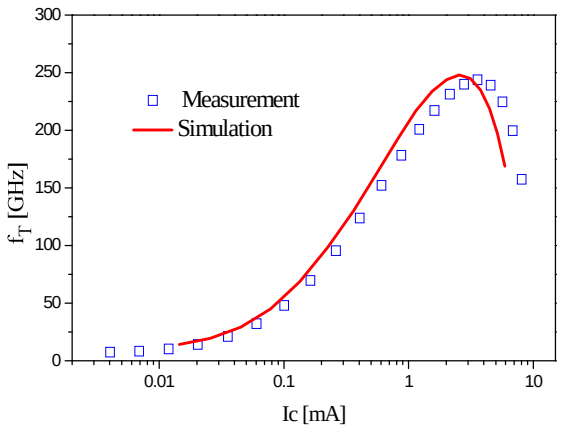

Fig.111: f_T comparison between measurement and simulation results (V_{BC}=0Volt).

The pre- and post-strain f_T and f_{MAX} curves as a function of the collector current are shown in Fig.112. The values of f_T and f_{MAX} are obtained assuming a constant gain-bandwidth product (-20dB/decade slope) with respect to the current gain |h21| and the unilateral gain |U| curves at a spot frequency of 30 GHz. The influence of introducing the nitride strain layer at the collector region is demonstrated. The results show that the post-strain HBT device exhibits better high frequency characteristics in comparison with an equivalent standard conventional device. An approximately 8% of improvement in f_T, and 5% of improvement in f_{MAX} have been achieved for the strained silicon NPN-SiGe-HBT device. Additionally, the transit frequency f_T and the maximum oscillation frequency f_{MAX} have been extracted for different V_{CB} biases for both structures. The obtained results are summarized in Table 12.

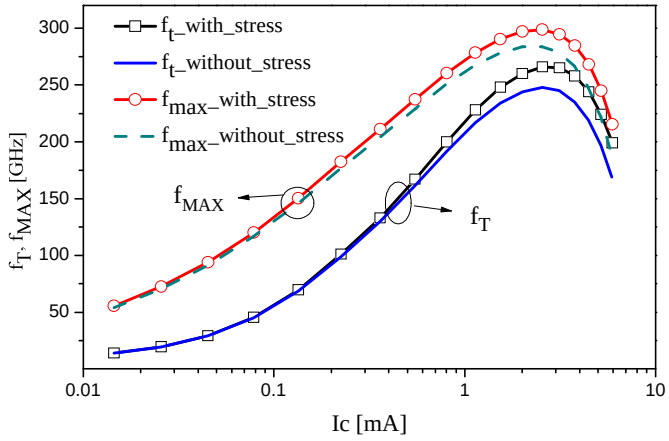

Fig.112: f_T and f_{MAX} vs. I_C for both HBT devices (W_E=130 nm, V_{BC}=0 Volt).

	V_{CB} = -0.2 [Volt]	V_{CB} = 0.0 [Volt]	V_{CB} =0.4 [Volt]	V_{CB} =0.6 [Volt]	V_{CB} =0.8 [Volt]	V_{CB} =1.0 [Volt]
$f_{T\ (without\_stress)}$ [GHz]	314	332	342	342	334.3	345.6
$f_{T\ (with\_stress)}$ [GHz]	338.8	357	370	370	370.4	369.6
$f_{MAX\ (without\_stress)}$ [GHz]	260.92	286	321	321	334.5	346.7
$f_{MAX\ (with\_stress)}$ [GHz]	272.5	299	337	337	349.6	361.8
f_T improvement [%]	7.9	7.8	8.2	8.2	10.8	7
f_{MAX} improvement [%]	4.5	4.5	5	5	4.5	4.4

Table 12: Devices frequency characteristics results extracted for different V_{CB} biases.

These performance improvements are related to the enhanced carrier transport properties. The carrier transport enhancement is mainly due to the enhanced carrier mobility by means of the induced tensile and compressive strains at the collector region. In addition to that, the induced strain at the collector region reduces the intrinsic collector resistance R_C without altering the base-collector junction capacitance C_{BC} as shown in Fig.113, where the base-collector junction capacitance C_{BC} is plotted versus the base-collector bias V_{CB}.

The intrinsic collector resistance R_C has a direct impact on the cut-off frequency f_T value according to the relation:

$$f_T = \frac{1}{2\pi\left(\tau_F + \frac{kT}{qI_C}(C_{BE} + C_{BC}) + (R_E + R_C)C_{CB}\right)} \quad 4.4$$

Where τ_F is the forward transit time, C_{BE} is the base-emitter junction capacitance, and R_E is the emitter resistance. Therefore, any reduction in the intrinsic collector resistance R_C will result in a reduction of the $R_C \times C_{BC}$ product, and consequently enhance the cut-off frequency f_T value.

Furthermore, the breakdown voltage BV_{CEO} has been extracted for both pre- and post-strain HBT devices for different collector doping levels (0.3Nc,0.4Nc, 0.5Nc, 0.65Nc, 0.75Nc, 0.8Nc, 0.9Nc, Nc, 1.5Nc, 2Nc, 3Nc, 4Nc), where Nc is the reference collector doping level taken from IMEC HBT device profile. The BV_{CEO} values have been extracted from the plot of the absolute value of the base current as a function of V_{CE} for both HBT devices for different collector doping level as shown in Fig.114. The obtained results are summarized in Table 13.

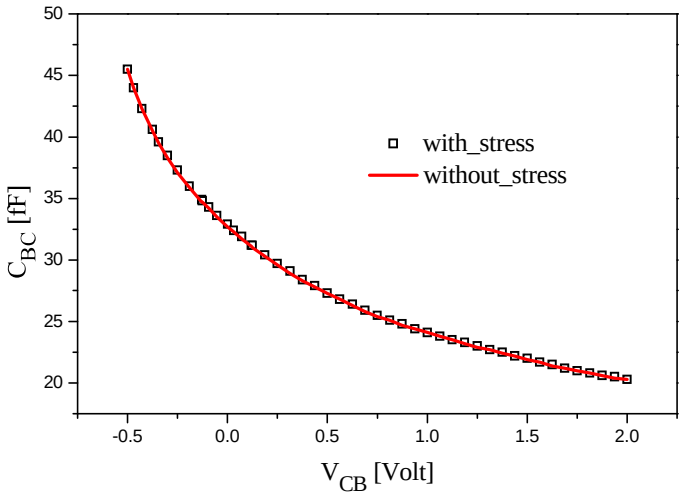

Fig.113: C_{BC} as a function of VCB for both HBT devices.

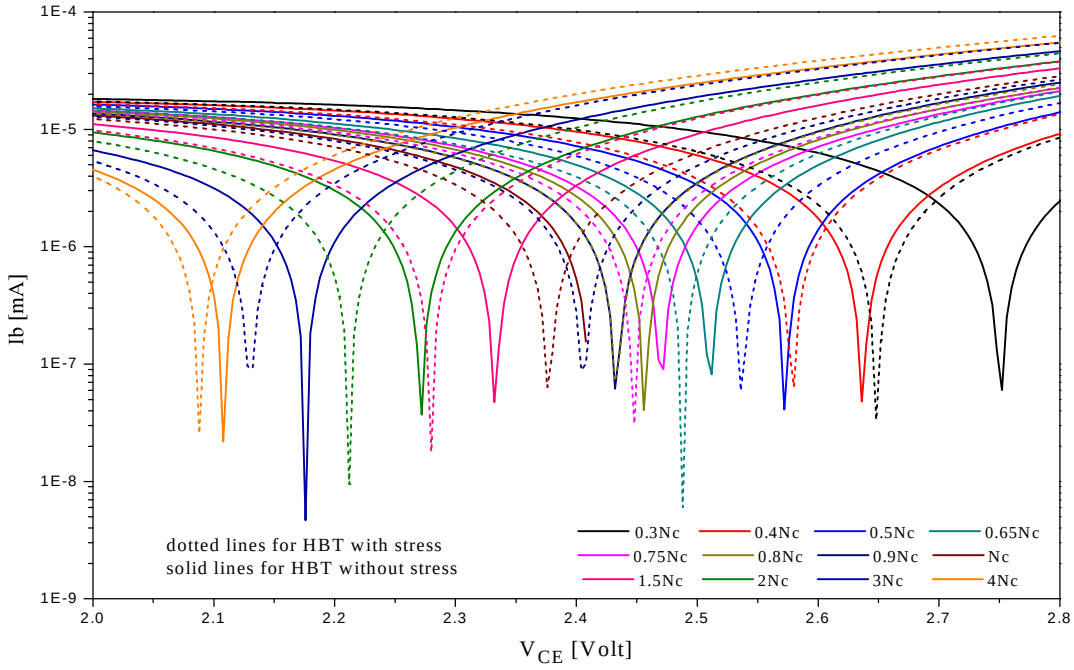

Fig.114: The absolute value of the base (Ib) current as a function of V_{CE} for both HBT devices (W_E=130 nm, V_{BE} = 0.70 V).

N_C Factor	**BV_{CEO} (without_stress) [Volt]**	**BV_{CEO} (with_stress) [Volt]**
0.3	2.75	2.65
0.4	2.63	2.58
0.5	2.57	2.54
0.65	2.51	2.49
0.75	2.47	2.45
0.8	2.45	2.43
0.9	2.43	2.40
1.0	2.41	2.38
1.5	2.33	2.28
2.0	2.27	2.21
3.0	2.18	2.13
4.0	2.11	2.09

Table 13: The extracted BV_{CEO} values for both HBT devices.

Simulation results show a very small reduction in the BV_{CEO} value in the new HBT device's architecture in comparison with an equivalent standard conventional one (1% – 4%). This reduction in the breakdown voltage value is related to the decrease of the collector band gap energy due to the induced strains at the collector region. However, despite of the very small decrease in the BV_{CEO} value, the $f_T \times BV_{CEO}$ product enhancement is about 6% by means of strain engineering ($f_T \times BV_{CEO}$(without-stress) =597.68 GHz.Volt, $f_T \times BV_{CEO}$(with-stress) =633.1 GHz.Volt).

The maximum oscillation frequency f_{MAX}, the relation between f_{MAX} and f_T is given by

$$f_{MAX} = \sqrt{\frac{f_T}{8\pi R_B C_{CB}}} \qquad 4.5$$

Therefore, any enhancement in f_T value will enhance the maximum oscillation frequency f_{MAX} of the device.

Moreover, the f_T, f_{MAX} and BV_{CEO} have been simulated for various collector doping levels. A typical BV_{CEO}–N_C characteristic has been observed for which the BV_{CEO} values decrease with increasing N_C (high f_T value). The peak breakdown voltage value is reached for the lowest collector doping for both devices as shown in Fig.115.

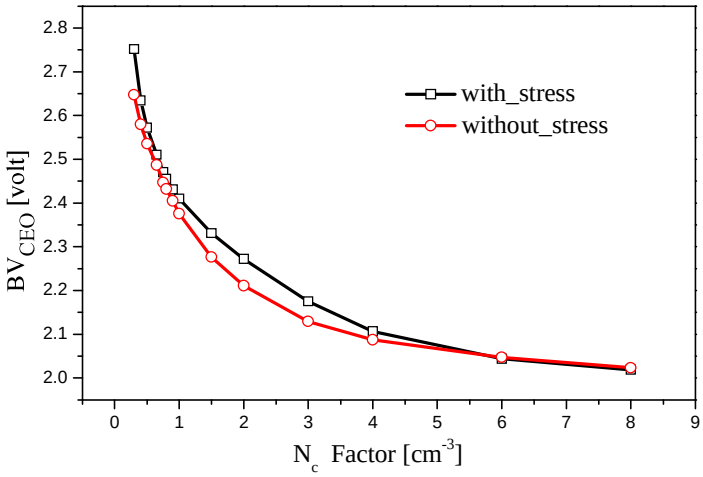

Fig.115: Variation of BV_{CEO} with the collector doping level N_C.

The variations of f_T and f_{MAX} values with the collector doping level Nc are shown in Fig.116 and Fig.117. The variation of f_T and f_{MAX} values with the breakdown voltage BV_{CEO} are

shown in Fig.118 and Fig.119. A trade-off between f_T and f_{MAX} values has been observed. This trade-off is illustrated in Fig.120 where f_T is plotted versus f_{MAX} for different BV_{CEO} values (i.e. different N_C). As shown in Fig.120, an improvement up to 43% in f_T value can be achieved for a given f_{MAX}, and up to 7% of improvement in f_{MAX} value for a given f_T can be achieved by means of strain engineering, as well by choosing the proper collector doping level.

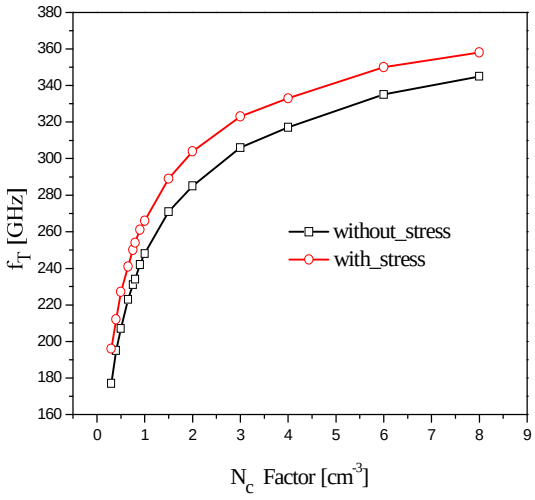

Fig.116: f_T versus collector doping levels characteristics for both devices.

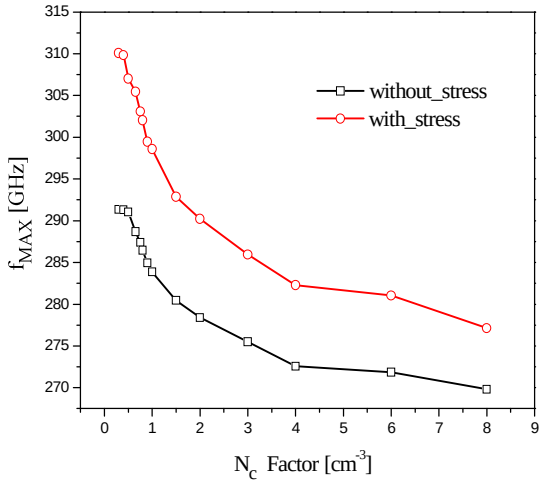

Fig.117: f_{MAX} versus collector doping levels characteristics for both devices.

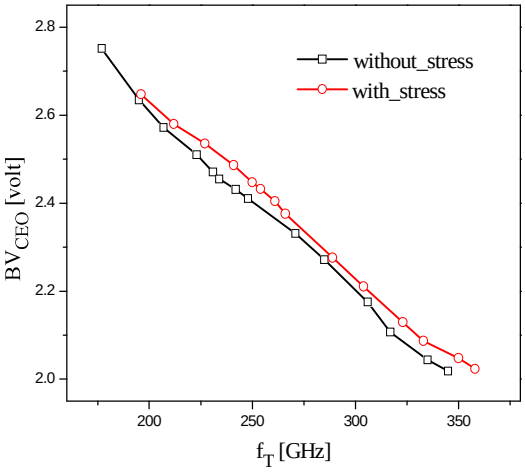

Fig.118: f_T versus BV_{CEO} characteristics for both devices.

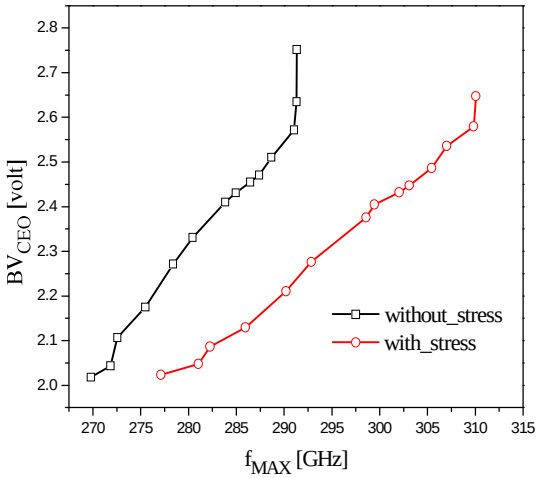

Fig.119: f_{MAX} versus BV_{CEO} characteristics for both devices.

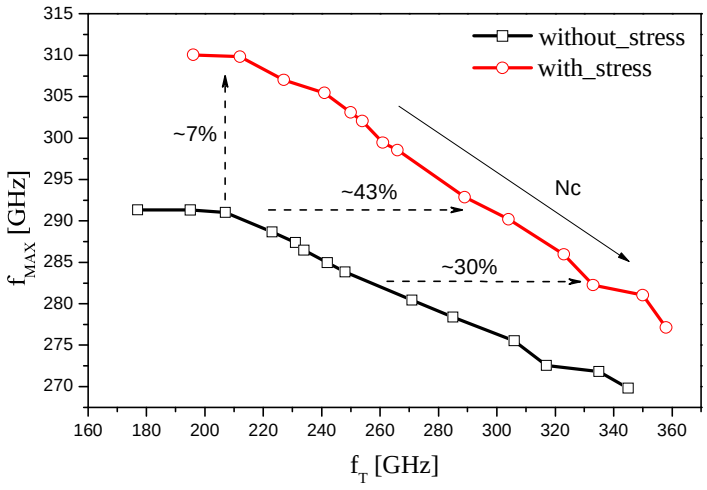

Fig.120: Trade-off between f_T and f_{MAX} for both HBT devices.

The dependence of f_T, f_{MAX} and BV_{CEO} values of the investigated HBT devices on the selected doping distribution at the collector region can be explained by referring to Kirk effect phenomena in HBT devices [20]. Where the Kirk current density J_K is given by

$$J_{Kirk} = qv_{sat}N_c\left[N_c + \frac{2\varepsilon(V_{CB}+V_{BI})}{qW_C^2}\right] \qquad 4.6$$

Where V_{CB} is the applied collector–base voltage, V_{BI} is the base–collector built-in voltage, W_C is the collector layer thickness and v_{sat} is the electron saturation velocity of the collector layer.

The peak f_T is limited by the onset of the Kirk effect or base push-out. Therefore using a high collector doping level suppresses the onset of the Kirk effect. Consequently, higher peak f_T can be obtained, due to the reduced base widening at high current. On the other hand, increasing the collector doping level N_C has two major disadvantages on the device's performance. Firstly, C_{BC} will be increased, leading to a reduced power gain or f_{MAX}. Secondly, the maximum electric field at the base-collector junction will be higher, leading to more impact ionization and thus lower BV_{CEO} value. The dependence of f_{MAX} on f_T and the collector–base capacitance C_{CB} is given by equation (4.5). As the results indicate, HBT devices with high f_T tend to have a high C_{CB} value. The C_{CB} dominates and f_{MAX} decreases despite the rise of f_T value. And that can explain the trade-off between f_T and f_{MAX} illustrated in Fig.120. In the case of strained SiGe-HBT device, the tradeoff is modified through an enhancement of the collector conductivity without changing base-collector capacitance.

Moreover, the transit time has been extracted for both HBT structures as a function of the collector current density Jc as shown in Fig.121. Simulation results show that the new SiGe-HBT device architecture employing nitride strain at the collector region exhibit lower transit time (t_e, t_b, and t_c) in comparison with an equivalent standard conventional HBT device. This enhancement on the device transit time is related to the reduction of the device bandgap energy and mobility enhancement by means of strain at the collector region.

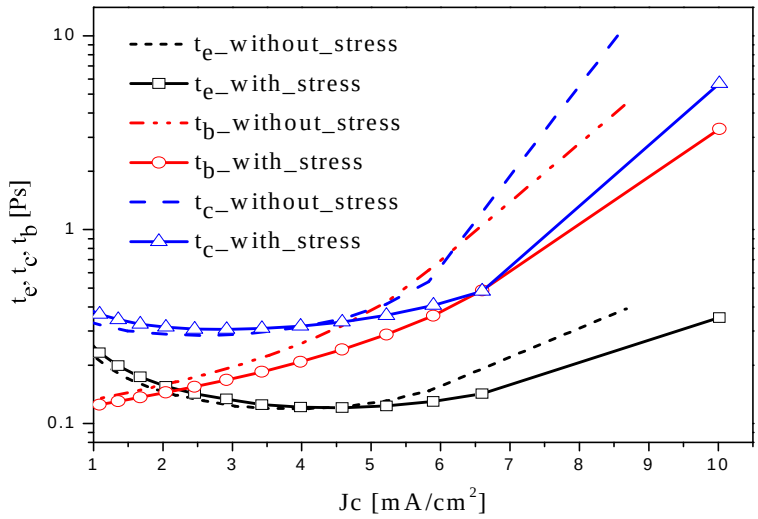

Fig.121: The transit time as a function of the collector current density for both HBT devices.

3.4 Conclusion

A new SiGe-HBT device's architecture employing nitride strain at the collector region has been presented. The device performance parameters have been investigated and compared with an equivalent standard conventional HBT device using TCAD modeling. Simulation results show that the strained silicon HBT device exhibits better high frequency characteristics in comparison with an equivalent standard conventional HBT device. An approximately, 8% of improvement in f_T, and 5% of improvement in f_{MAX} have been achieved for the new HBT architecture. The breakdown voltage BV_{CEO} has been extracted for both devices for different collector doping levels N_C. The obtained results show a very small reduction in the BV_{CEO} values for the strained

HBT device. In addition to that, a trade-off between f_T and f_{MAX} values has been observed from f_{MAX} versus f_T plot for different collector doping level N_C. An improvement up to 43% in f_T value can be achieved for a given f_{MAX}, and up to 7% of improvement in f_{MAX} value for a given f_T can be achieved by means of strain engineering, as well choosing the proper collector doping level

4. NPN-SiGe-HBT Device Employing SiO_2 Strain layer

A second approach to improve the device performance is to introduce the desired strain inside the device through using a silicon oxide (SiO_2) stain layer at the collector region during the device fabrication process. Silicon has a high affinity for oxygen, and an amorphous native oxide film rapidly forms on Si upon exposure to an oxidizing ambient. During thermal processing SiO_2 layer expands and contracts at different rates compared to the silicon substrate according to their thermal expansion coefficients. Because of this thermal expansion coefficients mismatch, as well the growth of oxide on top of silicon substrate, a mechanical strain is induced. The magnitude and the type of the induced mechanical strain are determined by controlling the processing conditions such as temperature, pressure, deposition power, reactant and impurity concentrations [21] [22]. In this work, the oxide strain layer is formed so that the desired strain is induced at the collector region, i.e, a mechanical compressive strain is induced along the horizontal axis, and a mechanical tensile strain is induced along the vertical axis. This approach is different from the normal oxide shallow trenches isolation (STI) which is done with respect to the CMOS device fabrication process.

In what follows, a novel NPN-SiGe-HBT device architecture utilizing SiO_2 stressor at the collector region will be presented. Likewise the NPN-SiGe-HBT device employing nitride strain at the collector region, process simulations are performed using Sentaurus TCAD software tools to build the device structure and to calculate the associated mechanical stress. The major processing steps are described in the following figures.

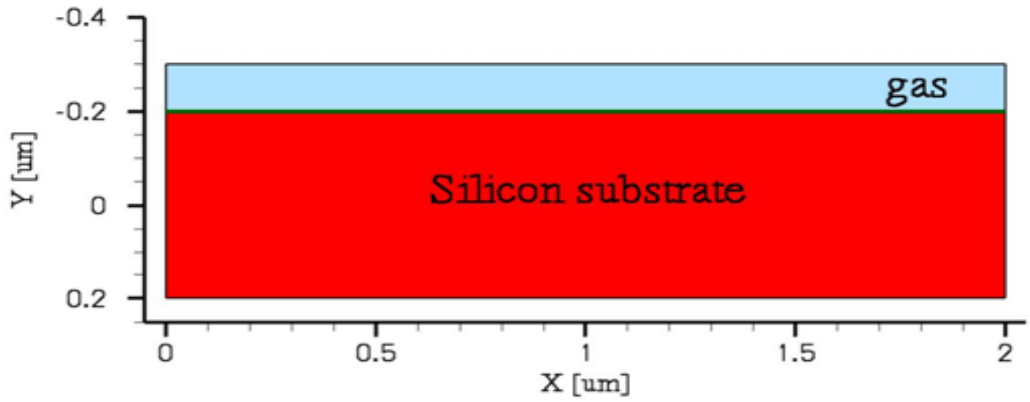

Fig.122: Process simulation : Deposition of silicon substrate.

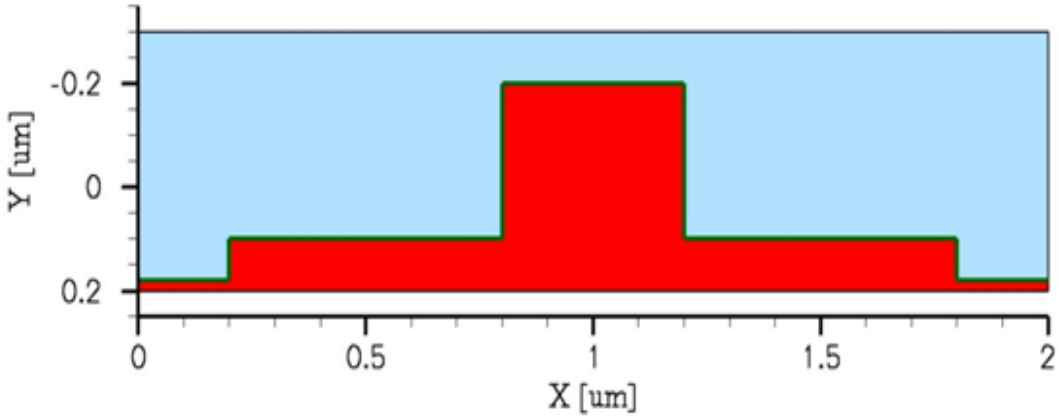

Fig.123: Process simulation: Silicon substrate etching.

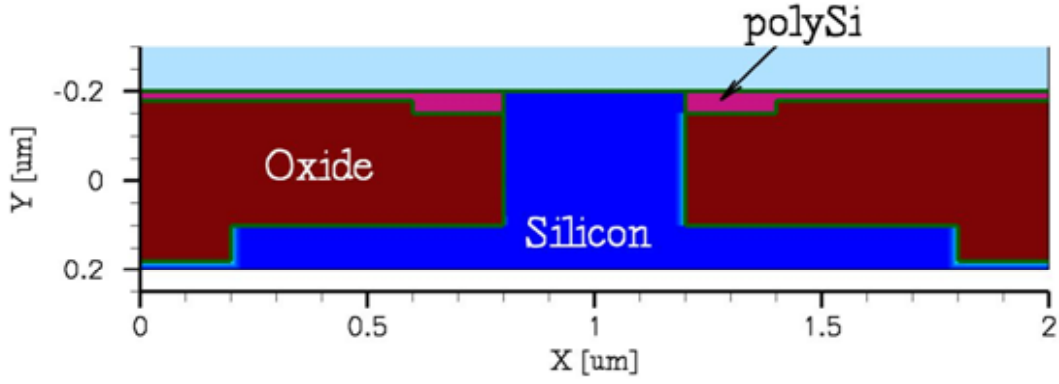

Fig.124: Process simulation: Deposition of oxide, etching of oxide and deposition of polysilicon.

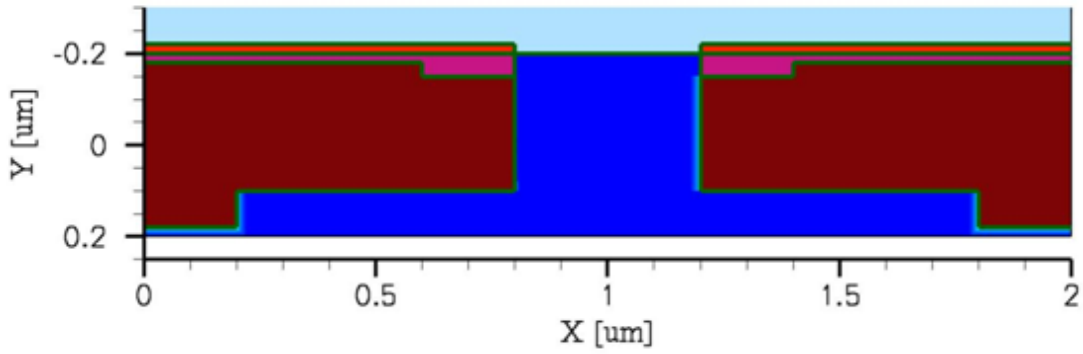

Fig.125: Process simulation: Deposition of silicon followed by etching of oxide and polysilicon layers.

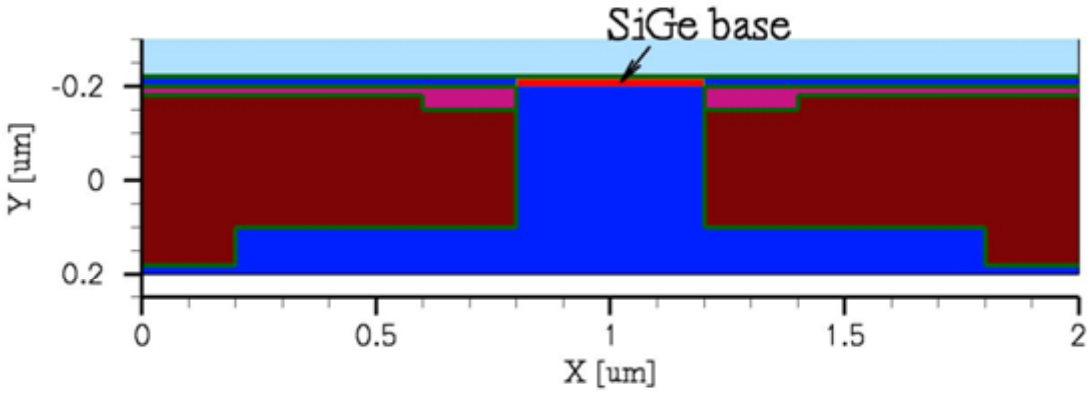

Fig.126: Process simulation: Deposition of SiGe base.

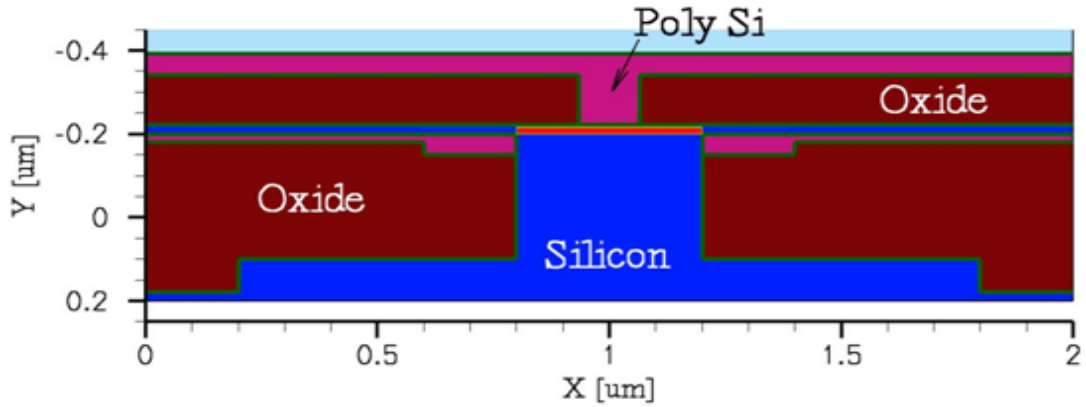

Fig.127: Process simulation: Deposition of oxide layer, formation of emitter opening and deposition of polysilicon layer.

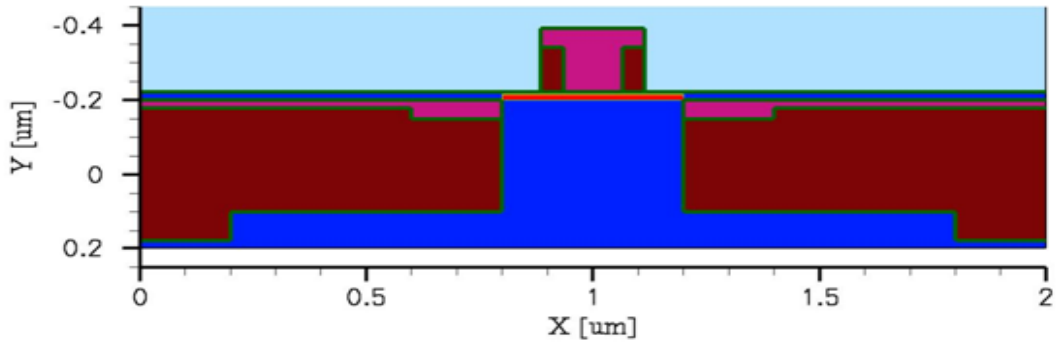

Fig.128: Process simulation: Etching of oxide and polysilicon layers to form the T-shape emitter.

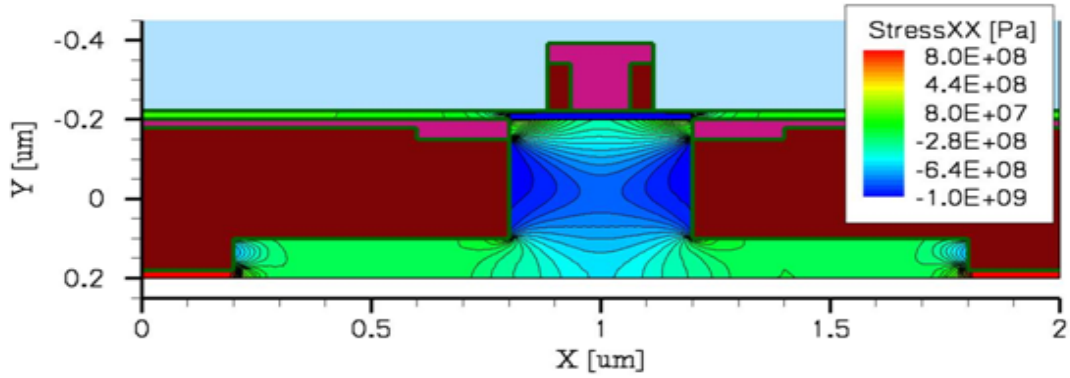

Fig.129: Process simulation: Final device structure showing stress isocontour lines generated inside the device.

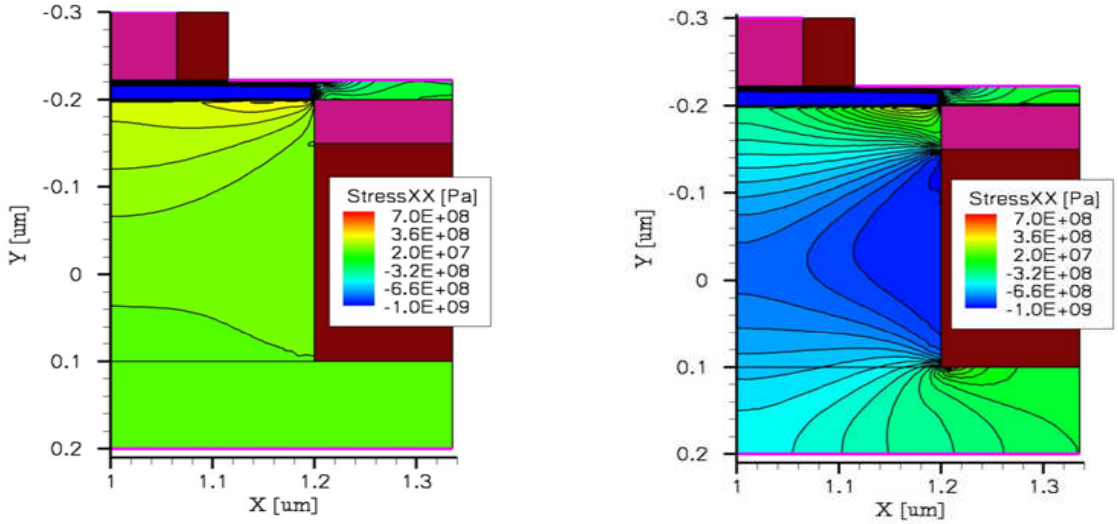

Fig.130: Isocontour lines representing the stress values generated inside the device; standard conventional NPN-SiGe-HBT device (left), and strained silicon NPN-SiGe-HBT device (right).

4.1 Impact of Strain

The impact of introducing a SiO_2 strain layer in the HBT device's architecture on the bandgap energy, and the stress values generated inside the device (Sxx, Syy and Szz) are shown in the following figures (131 to Fig.134). The mobility enhancement due to the applied stress has been calculated using the stress-induced electron mobility model. This model has been calibrated on mobility-strain enhancement results reported by [23] as shown in Fig.135, taking into account the saturation improvement of mobility at high stress values. The impact of SiO_2 strain layer on the device electron mobility is shown in Fig.136.

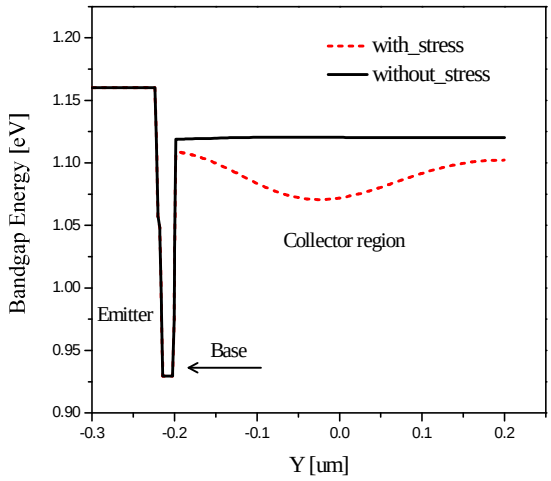

Fig.131: Impact of strain on the bandgap energy.

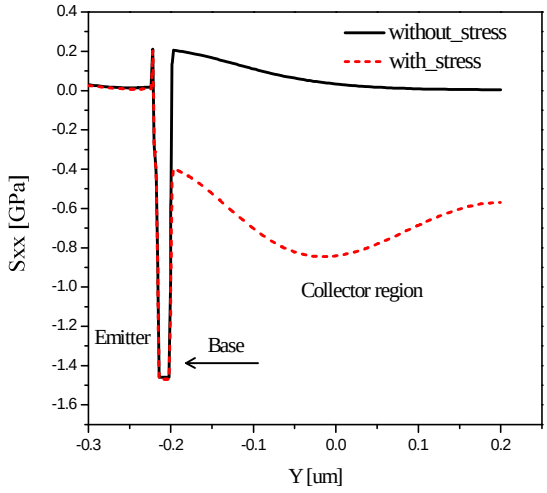

Fig.132: Impact of strain on Sxx.

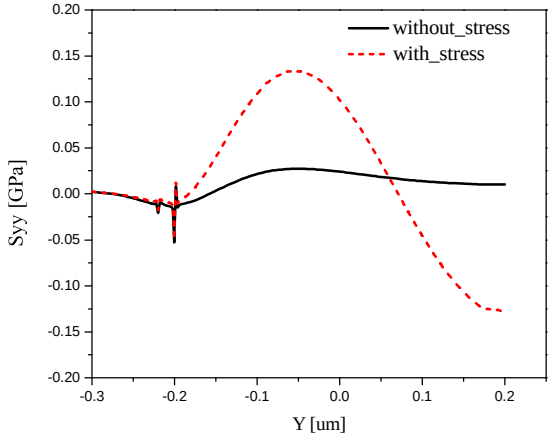

Fig.133: Impact of strain on Syy.

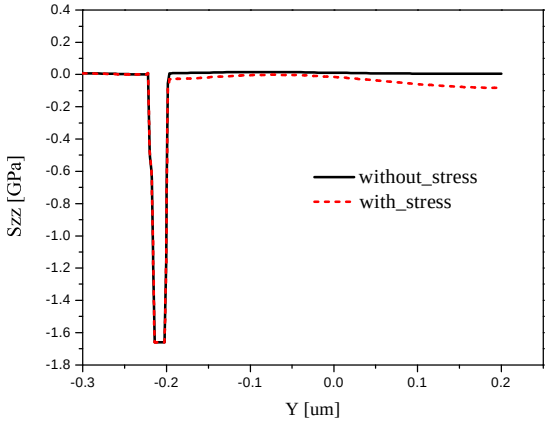

Fig.134: Impact of strain on Szz.

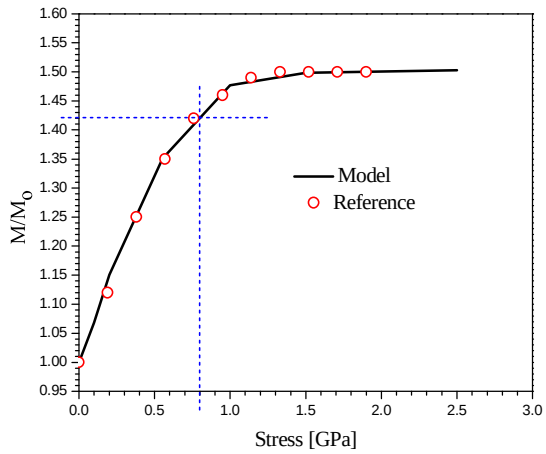

Fig.135: The calibrated mobility model used in TCAD device simulations.

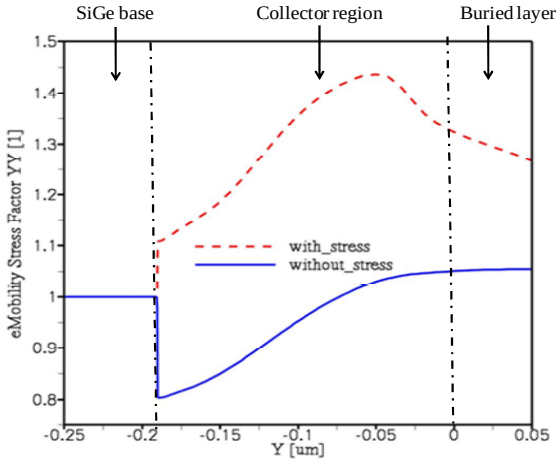

Fig.136: eMobility enhancement due to the applied strain.

As shown in the previous figures; introducing SiO_2 strain layer at the collector region will create compressive strain along the horizontal direction, and tensile strain along the vertical direction which causes a reduction in the device's bandgap energy at the collector region. This is related to the reduction of the conduction band energy due to the applied strain. In addition to that, strain will induce a change in the band structure. This in turn affects the carrier mobility, resulting in an approximately 42% of mobility improvement in Y-direction for electrons in the neutral collector region which is consistent with the obtained results reported by [23].

4.2 Electrical Simulation

TCAD Sentaurus software tools have been used to perform the 2D device simulations using hydrodynamic model (HD) without taking in consideration self heating effect and thermal behavior. However, the device's thermal behavior could be degraded by the formation of the oxide layer through affecting the device's thermal conductivity. The model parameters used in TCAD simulations have been calibrated by BU using Monte Carlo simulations. A detailed description of the models used is presented in chapter 3.

The same procedure carried out for investigating NPN-SiGe-HBT device with nitride strain at the collector region will be performed for NPN-SiGe-HBT device employing oxide

strain at the collector region. Similarly, the first step is to compare the output results of the simulated structure with IMEC HBT device measurement to verify the validity of the models used, and to ensure that the simulated device works properly. A comparison of the forward Gummel plots of IMEC HBT device measurement and the standard conventional HBT device simulation results are shown in Fig.135. The f_T curves as a function of the collector current for the standard NPN-SiGe-HBT device simulation results, and IMEC NPN-SiGe-HBT device measurement results are shown in Fig.138. Simulation results illustrated in Fig.1357 and Fig.138 show good agreement between measurement and simulation results, verifying the validity of the physical models and parameters used in the TCAD device simulations.

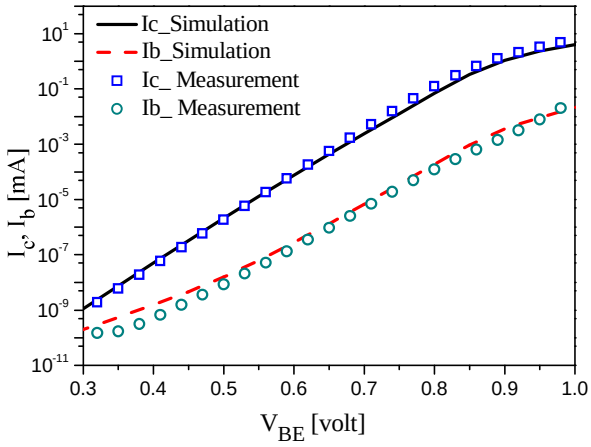

Fig.137: Comparison of Gummel plots of IMEC measurements and the conventional HBT device simulations.

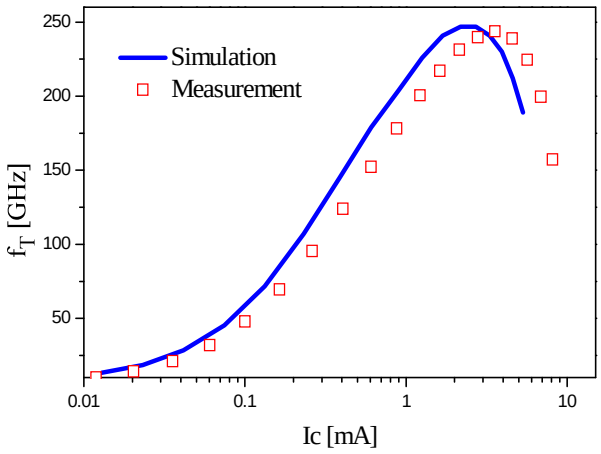

Fig.138: f_T comparison between IMEC HBT device and our conventional HBT device.

The f_T and f_{MAX} curves as a function of the collector current for NPN-SiGe-HBT device utilizing SiO_2 strain layer at the collector region, and a standard conventional NPN-SiGe-HBT device are shown in Fig.139 and Fig.140 respectively. The impact of introducing a SiO_2 strain layer at the collector region on the device's frequency response is demonstrated. An approximately 14% of improvement in f_T, and 9% of improvement in f_{MAX} have been achieved for NPN-SiGe-HBT device architecture employing SiO_2 strain layer at the collector region in comparison with an equivalent standard conventional HBT device.

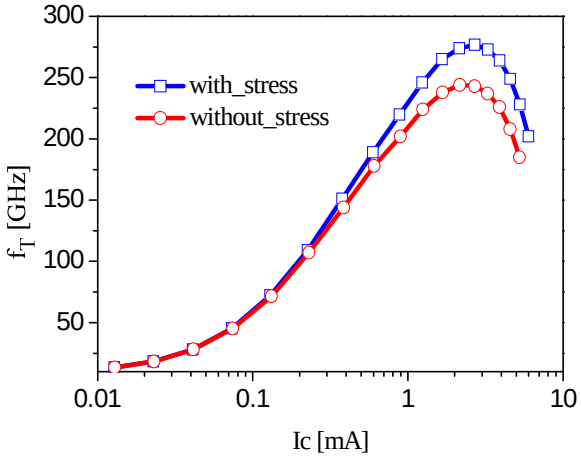

Fig.139: f_T vs. I_C for both HBT devices (V_{CB}=0).

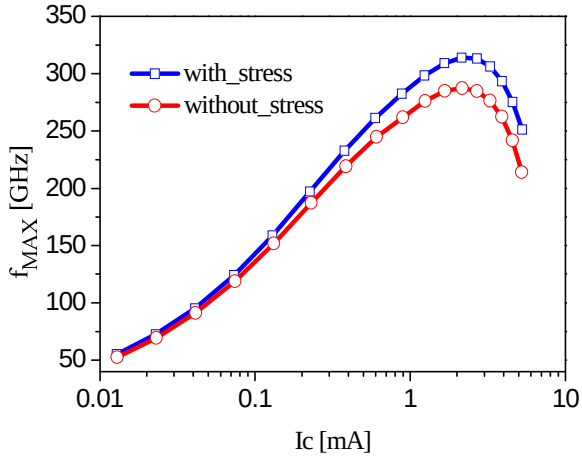

Fig.140: f_{MAX} vs. I_C for both HBT devices (V_{CB}=0).

Similarly, the base collector capacitance has been extracted as a function of collector base bias (V_{CB}) for both HBT devices as shown in Fig.141. The induced tensile and compressive strains due to employing SiO_2 strain layer at the collector region will enhance the electron vertical mobility, which in turn reduces the collector resistance R_C without altering the base collector junction capacitance. Consequently, the $R_C \times C_{BC}$ product will decrease. As a result, f_T will be enhanced due to the decrease of the $R_C \times C_{BC}$ product according to the relation (4.4).

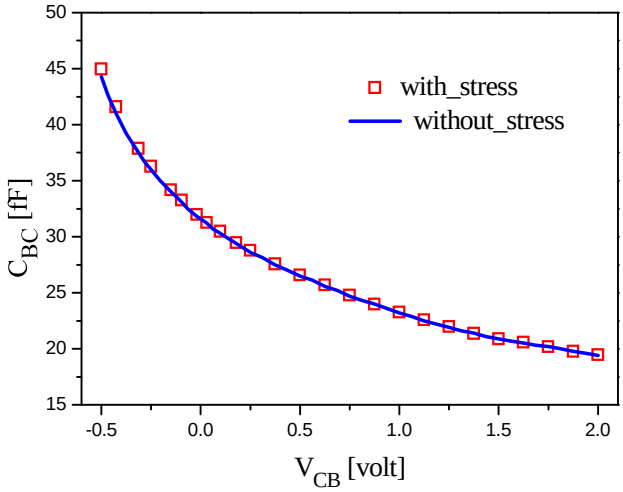

Fig.141: C_{BC} as a function of V_{CB} for both HBT devices.

Moreover, the decrease in the collector bandgap energy is responsible for the very small reduction in the breakdown voltage value shown in Fig.142. Apart from the very small decrease

in the breakdown voltage BV_{CE0} value (~1%), the $f_T \times BV_{CE0}$ product enhancement is approximately 12% by means of strain engineering at the collector region ($f_T \times BV_{CE0\ (without\text{-}stress)}$ =424.6 GHz.Volt, $f_T \times BV_{CE0\ (with\text{-}stress)}$ =476.4 GHz.Volt).

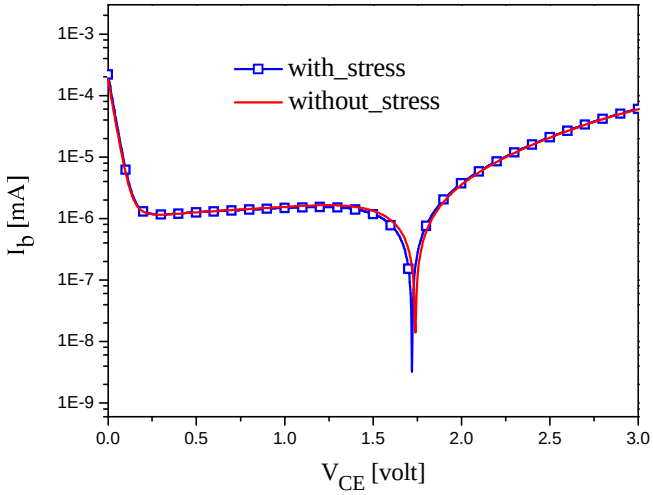

Fig.142: Absolute value of the base current as a function of V_{CE} for both HBT devices.(W_E=130 nm, V_{BE} = 0.70 V).

The f_T, f_{MAX} and the BV_{CEO} have been simulated for different collector doping levels (0.3Nc,0.4Nc, 0.5Nc, 0.65Nc, 0.75Nc, 0.8Nc, 0.9Nc, Nc, 1.5Nc, 2Nc, 3Nc and 4Nc), where Nc is the reference collector doping level taken from IMEC bipolar device profile. A typical BV_{CEO}–N_C characteristic has been observed for which the BV_{CEO} values decrease with increasing N_C (high f_T value). The peak breakdown voltage value is reached for the lowest collector doping for both devices as shown in Fig.143.

The variation of f_T and f_{MAX} values with the collector doping levels is shown in Fig.144 and Fig.145 respectively. The variation of f_T and f_{MAX} values with the breakdown voltage BV_{CEO} is shown in Fig.146 and Fig.147 respectively. A trade-off between f_T and f_{MAX} values has been observed. This trade-off is illustrated in Fig.148 where f_T is plotted versus f_{MAX} for different BV_{CEO} values (i.e. different N_C). As shown in Fig.148, an improvement up to 47% in f_T value can be achieved for a given f_{MAX}, and up to 14% of improvement in f_{MAX} value can be achieved for a given f_T by means of strain engineering technology, as well choosing the proper collector doping level.

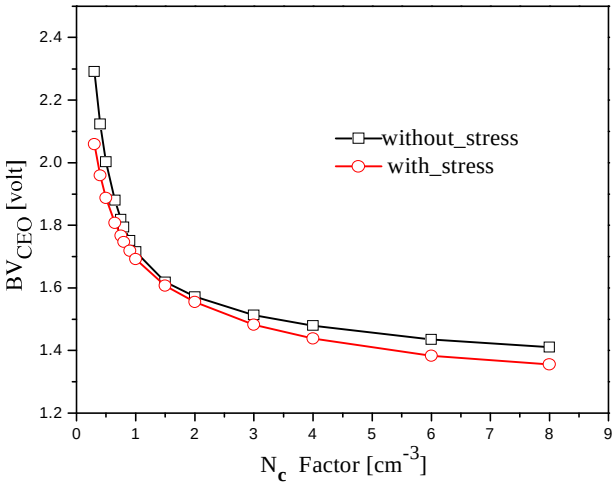

Fig.143: Variation of BV_{CEO} with the collector doping level Nc for both HBT devices.

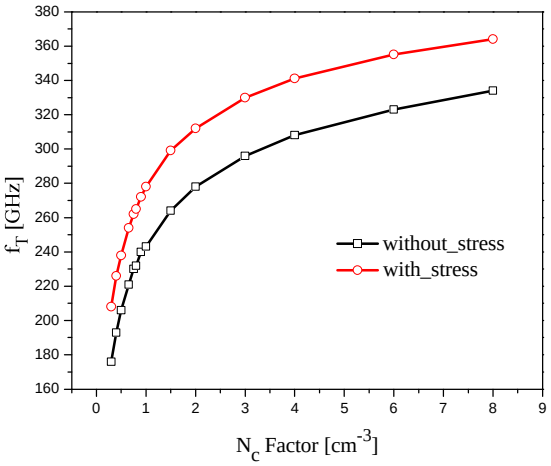

Fig.144: f_T versus collector doping level characteristics for both devices.

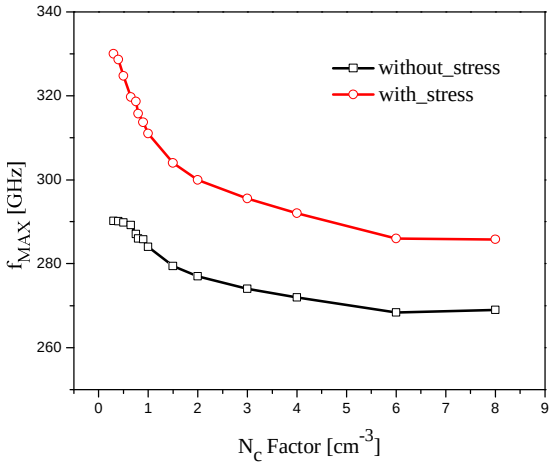

Fig.145: f_{MAX} versus collector doping levels characteristics for both devices.

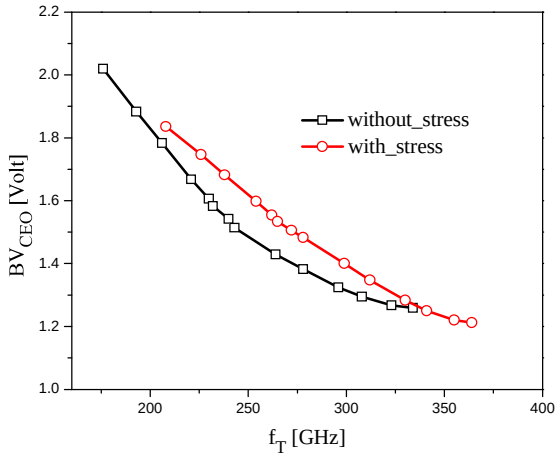

Fig.146: f_T versus BV_{CEO} characteristics for both devices.

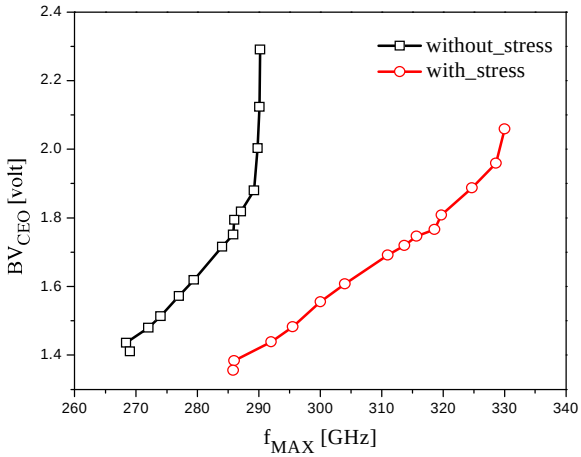

Fig.147: f_{MAX} versus BV_{CEO} characteristics for both devices.

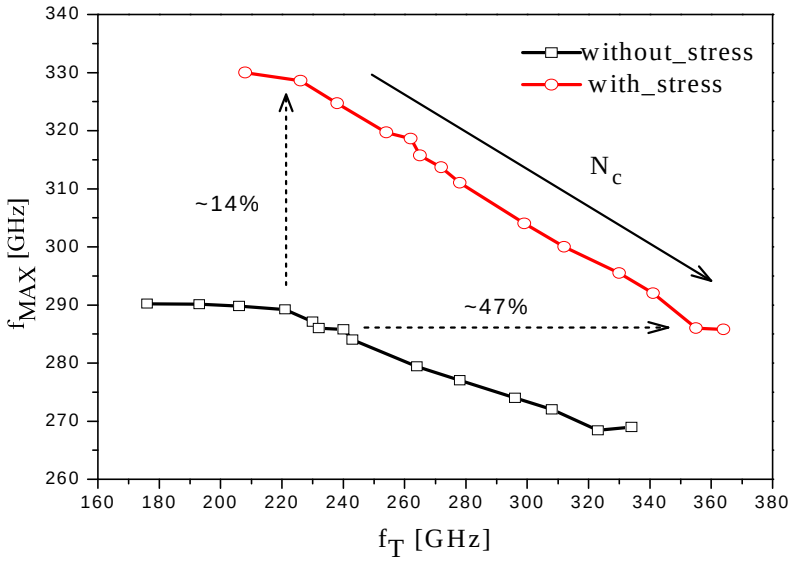

Fig.148: Trade-off between f_T and f_{MAX} for both HBT devices.

The dependence of f_T, f_{MAX} and the BV_{CEO} values of the investigated NPN-SiGe-HBT device on the selected doping distribution at the collector region can be explained by referring to Kirk effect phenomena in HBT devices according to the relation (4.6). The trade-off between f_T

and f_{MAX} in HBT device utilizing SiO_2 strain layer at the collector region can be explained in the same manner as the trade-off between f_T and f_{MAX} in HBT device utilizing Si_3N_4 strain layer at the collector region explained previously.

4.3 Conclusion

A novel NPN-SiGe-HBT device's architecture employing oxide strain layer at the collector region has been presented. The device performance parameters have been investigated and compared with an equivalent standard conventional HBT device using TCAD modeling. Simulation results show that the strained silicon HBT device exhibits better high frequency characteristics in comparison with an equivalent standard conventional HBT device. An approximately 14% of improvement in f_T, and 9% of improvement in f_{MAX} have been achieved for the new device's architecture. The breakdown voltage BV_{CEO} has been extracted for both devices for different collector doping level N_C. The obtained results have shown a very small reduction in the BV_{CEO} values for the strained silicon HBT device (~1%). In addition to that, a trade-off between f_T and f_{MAX} values has been observed from f_{MAX} versus f_T plot for different collector doping level N_C. An improvement up to 47% in f_T value can be achieved for a given f_{MAX}, and up to 14% of improvement in f_{MAX} value can be achieved for a given f_T by means of strain engineering technology, as well choosing the proper collector doping level.

References

[1] D. Chidambarrao, G.G. Freeman, M.H. Khater, "Bipolar Transistor with Extrinsic Stress Layer" U.S. Patent US 7,102,205 B2.Sep-2006.

[2] J.S. Dunn, D.L.Harame, J.B. Johnson, A.B. Joseph, "Structure and Method for Performance Improvement in Vertical Bipolar Transistors" U.S. Patent US 7, 262,484 B2Aug-2007.

[3] Synopsys, "Sentaurus Process User Guide" Dec-2007.

[4] S. Decoutere, A. Sibaja-Hernandez, "IMEC private communication" 2008.

[5] J. Bardeen and W. Shockley, "Deformation Potentials and Mobilities in Non-Polar Crystals" vol. 80, 1950.

[6] G L Bir, G E Pikus, *Symmetry and Strain-induced Effects in Semiconductors*. Wiley, New York, 1974.

[7] J.L. Egley, D. Chidambarrao, "Strain effects on device characteristics: Implementation in drift-diffusion simulators" vol. 36, pp. 1653-1664, 1993.

[8] P.R. Chidambaram, C. Bowen, S. Chakravarthi,C. Machala, R. Wise, "Fundamentals of silicon material properties for successful exploitation of strain engineering in modern CMOS manufacturing" vol. 53, 2006.

[9] M. Schroter, L. Tzung-Yin, "Physics-Based Minority Charge and Transit Time Modeling for Bipolar Transistors" vol. 46, 1999.

[10] Synopsys, "Sentaurus Device User Guide" Dec-2007.

[11] S.L. Salmon, J. D. Cressler, R. C. Jaeger , D. L. Harame, "The Influence of Ge Grading on the Bias and Temperature Characteristics of SiGe HBT's for Precision Analog Circuits" vol. 47, 2000.

[12] B. Tillack, B. Heinemann, D. Knoll, H. Rücker, Y. Yamamoto, "Base doping and dopant profile control of SiGe npn and pnp HBTs" vol. 254, 2008.

[13] F. Ellinger, L. C. Rodoni, G. Sialm, C. Kromer, et al, "Effects of emitter scaling and device biasing on millimeter-wave VCO performance in 200 GHz SiGe HBT technology" vol. 52, 2004.

[14] S. Fregonese, Y. Zhuang, J. N. Burghartz, "Modeling of Strained CMOS on Disposable SiGe Dots: Strain Impacts on Device Electrical Characteristics" vol. 54, 2007.

[15] P. R. Chidambaram, C. Bowen, S. Chakravarthi, C. Machala, R. Wise, "Fundamentals of Silicon Material Properties for Successful Exploitation of Strain Engineering in Modern

CMOS Manufacturing" vol. 53, 2006.

[16] M. Al-Sa'di, S. Fregonese, C. Maneux, T. Zimmer, "Investigation of Electrical BJT Performance Improvement through Extrinsic Stress Layer Using TCAD Modeling" in *Semiconductor conference Dresden proceedings, IEEE SCD 2009*, 2009.

[17] 9. M. Al-Sa'di, S. Fregonese, C. Maneux, T. Zimmer, "TCAD Modeling of NPN-SiGe-HBT Electrical Performance Improvement through Extrinsic Stress Layer" in *27th International Conference on Microelectronics proceedings , IEEE MIEL 2010, 387*, 2010.

[18] C. K. Maiti, L .K. Bera, S. Chattopadhyay, "Strained-Si heterostructure field effect transistors" vol. 13, 1998.

[19] 4. A. A. Bayati, L. Washington, L.-Q. Xia, M. Balseanu, Z. Yuan, M. Kawaguchi, et al, "Production processes for inducing strain in CMOS channels" vol. 26, 2005.

[20] J. D. Cressler, *Silicon heterostructure handbook*. Taylor & Francis, 2006.

[21] 1. J. A. Rodriguez, A. Llobera, C. Dominguez, "Evolution of the mechanical stress on PECVD silicon oxide films under thermal processing" vol. 19, 2000.

[22] N. Verghese, R. Nachman, P. Hurat, "Modeling stress- induced variability optimizes ic timing performance" Cadence Design Systems, Inc., 2010.

[23] K.H. Bach, R. Liebmann, M. Nawaz, C.Jungemann , E. Ungersboeck., "Nonlinear Piezoresistance Effect in Devices with Stressed Etch Stop Liner" vol. 12, 2007.

Summary & Conclusion

SiGe HBTs have proven their capability to support large bandwidth and high data rates for high-speed communication systems. Devices with impressive f_T values have been demonstrated that only a couple of years ago would have been believed to be reserved for III–V technologies. SiGe HBT technologies that exhibit higher operating speed can be leveraged for advanced circuits and systems in different ways; they can open up new applications at very high frequencies (THz). Their speed can be traded for lower power dissipation, or they can be used to mitigate the impact of process, voltage and temperature variations at lower frequencies for higher yield and improved reliability (case of automotive radar application). Due to the continuous demand for devices in areas of technology where high speed and high frequency response are required, it becomes imperative to develop new bipolar device architectures suitable for high frequency and power applications. Among the various techniques and efforts proposed to improve the performance of HBT devices, strain engineering technology provides an additional degree of freedom to enhance the carriers transport properties due to band structure changes and mobility enhancement. The mobility of charge carriers in bipolar devices can be enhanced by creating mechanical tensile strain in the direction of electrons flow to improve electrons mobility, and by creating mechanical compressive strain in the direction of holes flow to improve holes mobility.

This work investigates the effects of introducing strain engineering technology principle on Silicon Bipolar Junction Transistor (BJT) and Silicon-Germanium Heterojunction Bipolar Transistor (HBT) devices as a possible alternative to dimensional scaling. This thesis focuses on how strain affects Si BJT and SiGe HBTs, where compressive and tensile strains are applied during the devices fabrication process. The compressive and tensile strains are created through two approaches. The first approach is through introducing strain engineering technology principle at the device base region using SiGe extrinsic stress layer formed over the extrinsic base layer,

and adjacent to the intrinsic base of NPN-Si-BJT/NPN-SiGe-HBT device. The second approach is through introducing strain engineering technology principle at the device collector region by means of local strain technique using strain layers (Si_3N_4 and SiO_2 strain layers).

The work methodology performed in this study consists of the following steps:

- Virtually fabricate the device using process simulations.
- Study the sensitivity of the device's different zones to strain.
- Propose new methods to incorporate strain in the process and to evaluate the strain level that can be obtained inside the device.
- Define simulation parameters and physical models (the model parameters have been calibrated in collaboration with Universität der Bundeswehr München).
- Perform numerical (device) simulations to analyze the device electrical performance.

Process simulations were performed using Sentaurus TCAD software tools to virtually fabricate the device's structure, and to calculate the associated induced mechanical strain using anisotropic elasticity model. Sentaurus TCAD software tools were used to perform the two-dimensional device's simulations based on Hydrodynamic (HD) and Drift-Diffusion (DD) models. The model parameters used in TCAD simulations were calibrated by Universität der Bundeswehr München using Monte Carlo simulations. The carrier mobilities were calculated using Philips unified mobility model, the high field saturation was calculated through the Canali model by using the carrier temperatures as the driving force. The carrier generation-recombination models used are the Shockley‑Read‑Hall recombination model, and Auger recombination model. As well, the doping induced bandgap narrowing model has been employed. Furthermore, the stress-induced mobility enhancement was calculated using the Piezoresistivity model and the stress-induced electron mobility model. The graded Ge profile and the doping profiles in the emitter, base, and collector regions were taken from IMEC Microelectronics HBT device (f_T/f_{MAX} = 205GHz/275 GHz).

Simulation results show that applying strain engineering concept at the base region of NPN-Si-BJT device using SiGe strain layer can strongly enhance the device's performance due to the sensitivity of silicon material to strain. An approximately 42% of improvement in f_T, and 13% of improvement in f_{MAX} have been achieved. As well, an enhancement of the collector current by

nearly three times in strained silicon NPN-Si-BJT device has been attained. The obtained results for applying the same technique on NPN-SiGe-HBT device have shown that applying strain on the base region of the HBT device is less efficient in comparison with the BJT device, as the SiGe base is already stressed due to the existence of Ge at the base. An approximately, 3% of improvement in f_T, and 5% of improvement in f_{MAX} have been achieved. In addition to that, a decrease in the device's total transit time has been observed. The intensive study of the transit time in the strained NPN-SiGe-HBT device shows that the major modification of the device's total transit time in comparison with the standard conventional NPN-SiGe-HBT device arises from the reduction of the collector transit time. Verifying that silicon material is more sensitive to strain than the SiGe base region, and the device's performance improvements are mainly due to the impact of the induced strain at the collector region.

Consequently, new NPN-SiGe-HBT device architectures utilizing strain layer at the collector region is proposed. Simulation results show that applying strain engineering concept at the collector region of the investigated devices will enhance the device's performance and frequency response characteristic. By using Si_3N_4 as a strain layer, an approximately, 8% of improvement in f_T, and 5% of improvement in f_{MAX} have been achieved in the new NPN-SiGe-HBT device's architecture in comparison with an equivalent standard conventional NPN-SiGe-HBT device. Despite of the very small decrease in the breakdown voltage BV_{CE0} value (1%─4%), the $f_T \times B_{VCE0}$ product enhancement is about 12% by means of strain engineering at the collector region. Moreover, using SiO_2 as a strain layer at the device's collector region will result in 14% of improvement in f_T, and 9% of improvement in f_{MAX} and an enhancement of 12% of $f_T \times BV_{CE0}$ product in comparison with an equivalent standard conventional device.

The performance improvements in the strained BJT/HBT devices are related to the induced tensile and compressive strains inside the device, this in turn will enhance both electron and hole mobility's, and improves the devices electrical performance. The obtained results obviously show that strain engineering technology principle applied to BJT/HBT device can be a promising approach for further devices performance improvements.

However, the work done in this thesis opens new doors for further research and investigations in the field of strained BJT/HBT devices. In addition to that, it provides the suitable background and the calibrated simulator tools for future work on SiGe HBT compact modeling.

Future work

This work explores and investigates the impacts of mechanical strain engineering technology principle on Si bipolar and SiGe heterojunction bipolar devices using TCAD modeling (Process and device simulations). Although TCAD modeling gives a deep insight of the impact of mechanical strain on the devices performance, as well, provides the tool for designing and exploring new device concepts, it would be worth exploring the possibilities of fabricating new strained bipolar devices through a simplified device structure. This can be achieved by considering new device architectures that are based on less complicated fabrication process steps. In addition to that, it would be beneficial to analyze the reliability issues for the proposed devices structures, such as self-heating effect and reliability issues associated with the materials used as stressors (oxide and nitride). Such a study could provide a complete set of information regarding the strained bipolar devices stability and the proper level of the applied strain.

Appendix

1. Mobility Models Calibration

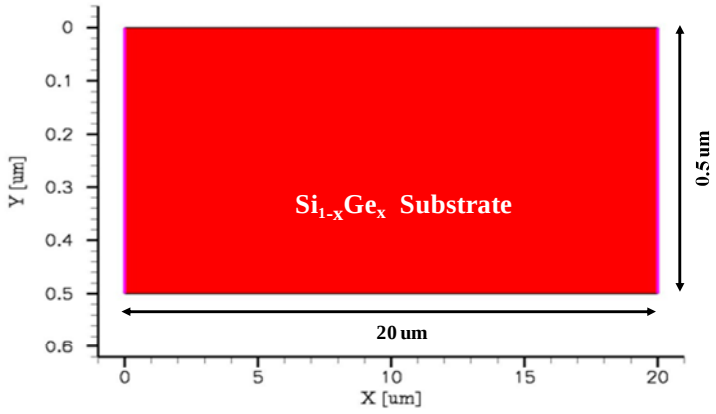

Fig.149 : The $Si_{1-x}Ge_x$ sample used in TCAD simulator calibration process.

Model (I)	Model (II)
Hydrodynamic **EffectiveIntrinsicDensity (Slotboom)** **Mobility** • DopingDep. • eHighFieldSaturation (CarrierTempDrive) • hHighFieldSaturation (CarrierTempDrive)	**Hydrodynamic** **EffectiveIntrinsicDensity (Slotboom)** **Mobility** • PhuMob. • eHighFieldSaturation (CarrierTempDrive) • hHighFieldSaturation (CarrierTempDrive)

Table 14: Physical models used in the calibration models

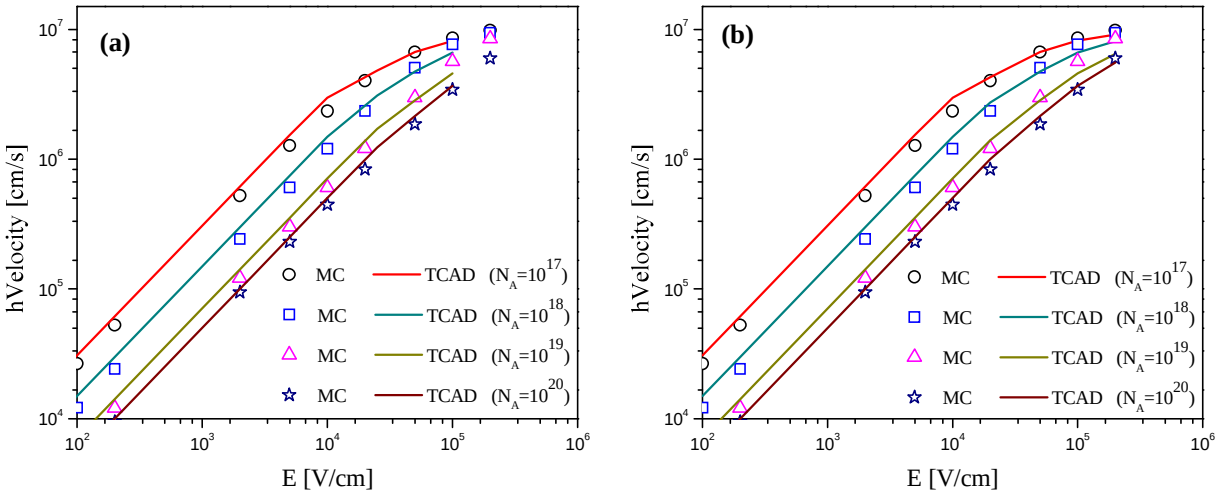

Fig.150: Comparison between MC and TCAD simulation results for hVelocity vs. Electric field using (a) Model (**I**), and (b) Model (**II**), (Ge_content = 4%, T=300K).

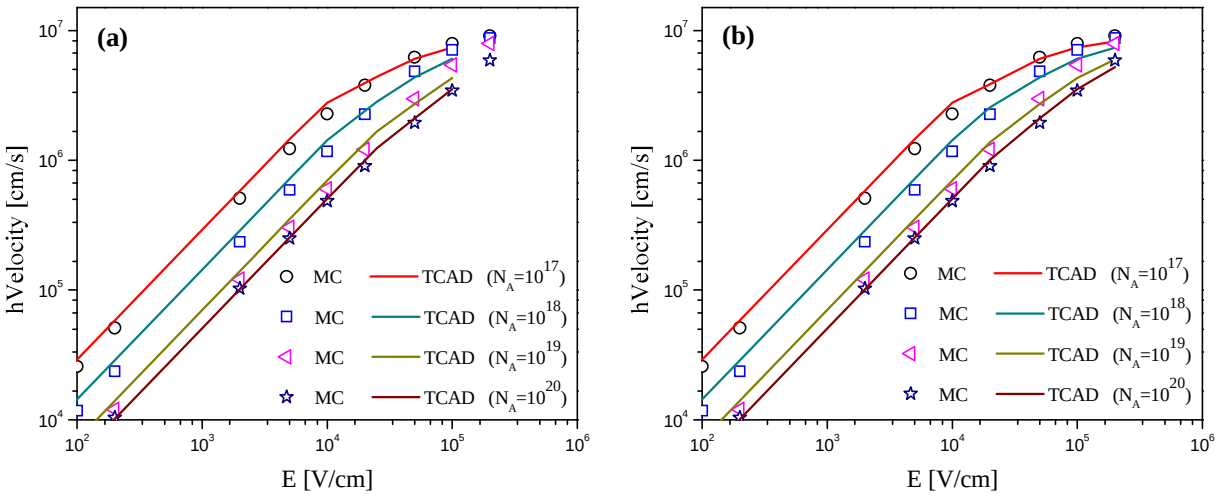

Fig.151: Comparison between MC and TCAD simulation results for hVelocity vs. Electric field using (a) Model (**I**), and (b) Model (**II**), (Ge_content = 12%, T=300K).

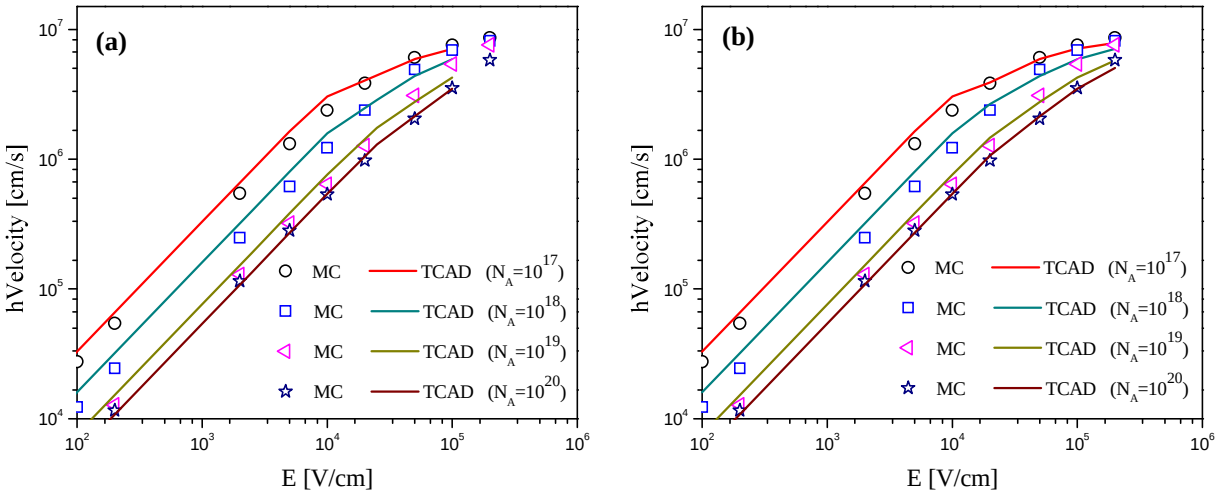

Fig.152: Comparison between MC and TCAD simulation results for hVelocity vs. Electric field using (a) Model (**I**), and (b) Model (**II**), (Ge_content = 18%, T=300K).

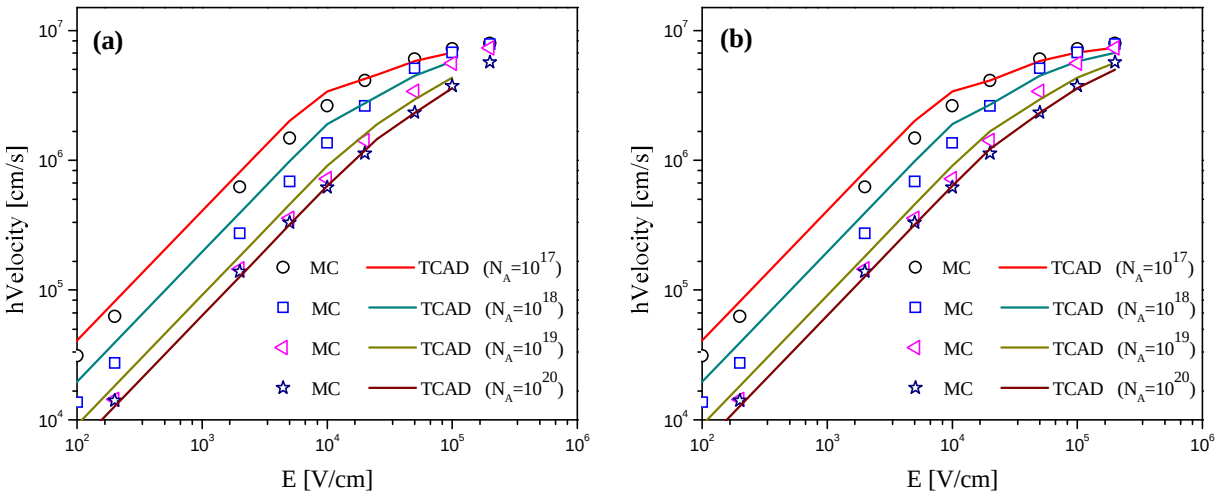

Fig.153: Comparison between MC and TCAD simulation results for hVelocity vs. Electric field using (a) Model (**I**), and (b) Model (**II**), (Ge_content = 28%, T=300K).

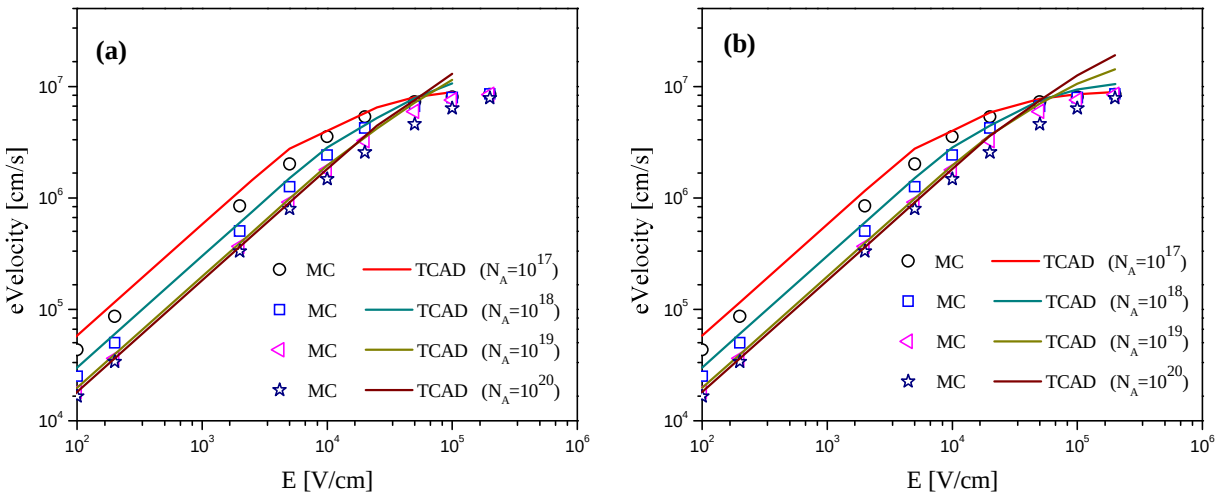

Fig.154: Comparison between MC and TCAD simulation results for eVelocity vs. Electric field using (a) Model (**I**), and (b) Model (**II**), (Ge_content = 4%, T=300K).

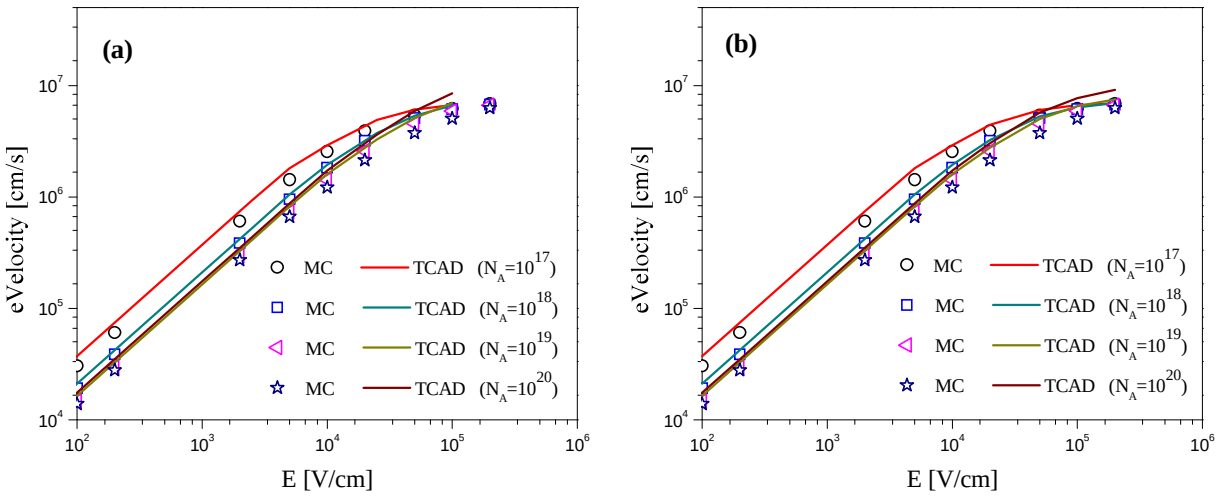

Fig.155: Comparison between MC and TCAD simulation results for eVelocity vs. Electric field using (a) Model (**I**), and (b) Model (**II**), (Ge_content = 12%, T=300K).

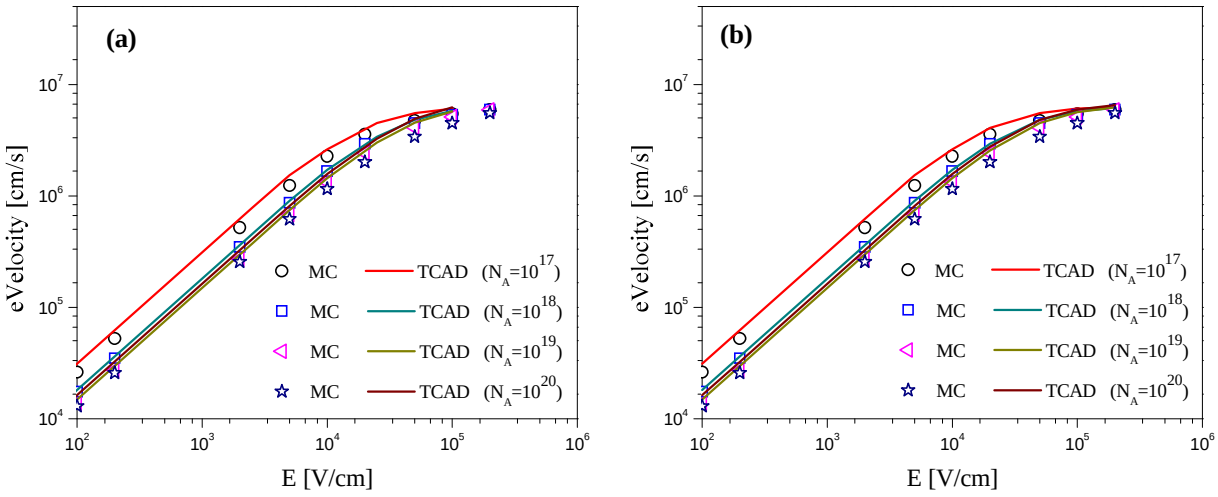

Fig.156: Comparison between MC and TCAD simulation results for eVelocity vs. Electric field using (a) Model (**I**), and (b) Model (**II**), (Ge_content = 18%, T=300K).

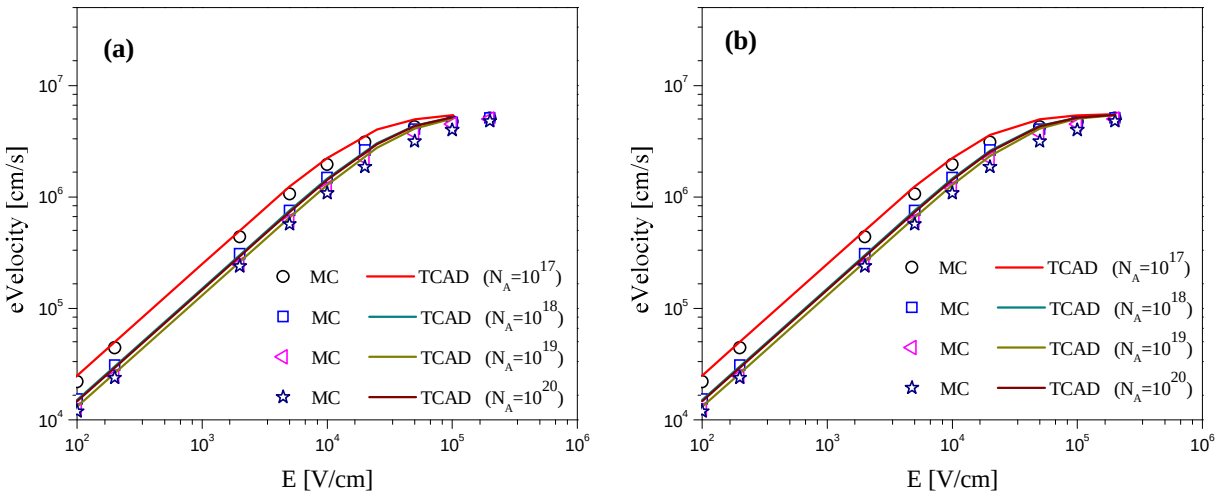

Fig.157: Comparison between MC and TCAD simulation results for eVelocity vs. Electric field using (a) Model (**I**), and (b) Model (**II**), (Ge_content = 28%, T=300K).

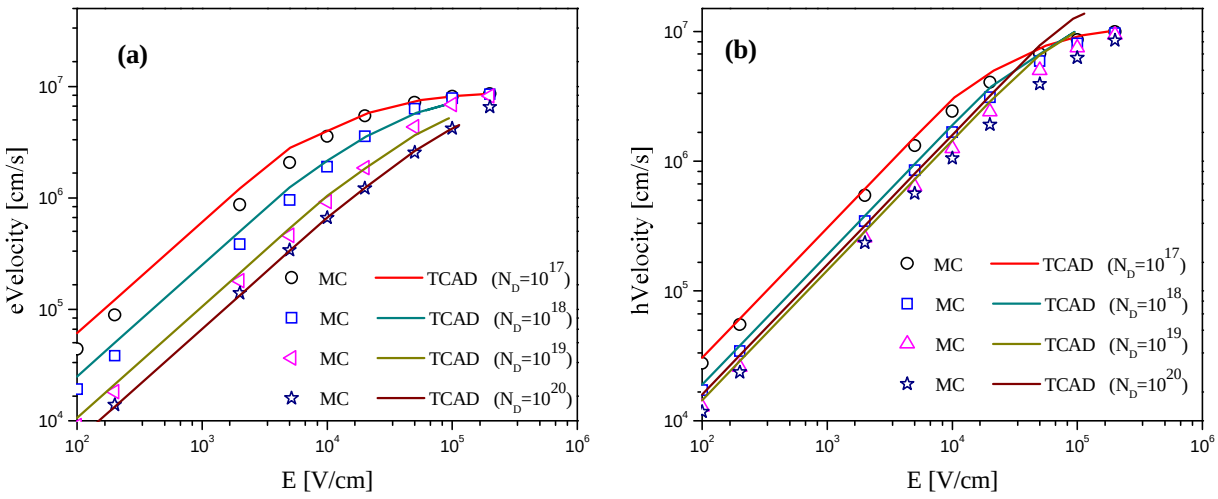

Fig.158: Comparison between MC and TCAD simulation results using Model (**II**) for (a) eVelocity vs. Electric field (b) hVelocity vs. Electric field, (Ge_content = 4%, T=300K).

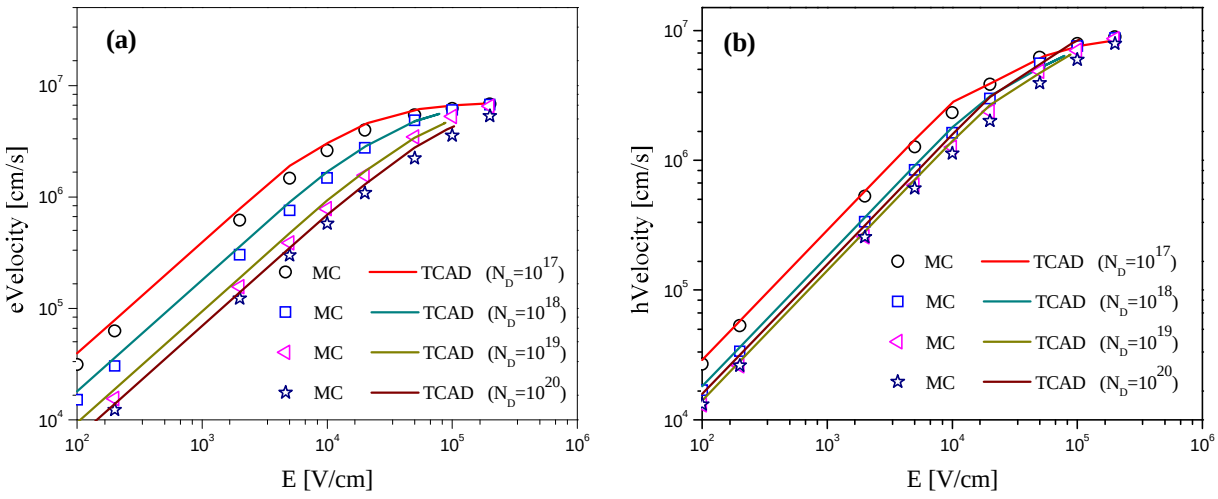

Fig.159: Comparison between MC and TCAD simulation results using Model (**II**) for (a) eVelocity vs. Electric field (b) hVelocity vs. Electric field, (Ge_content = 12%, T=300K).

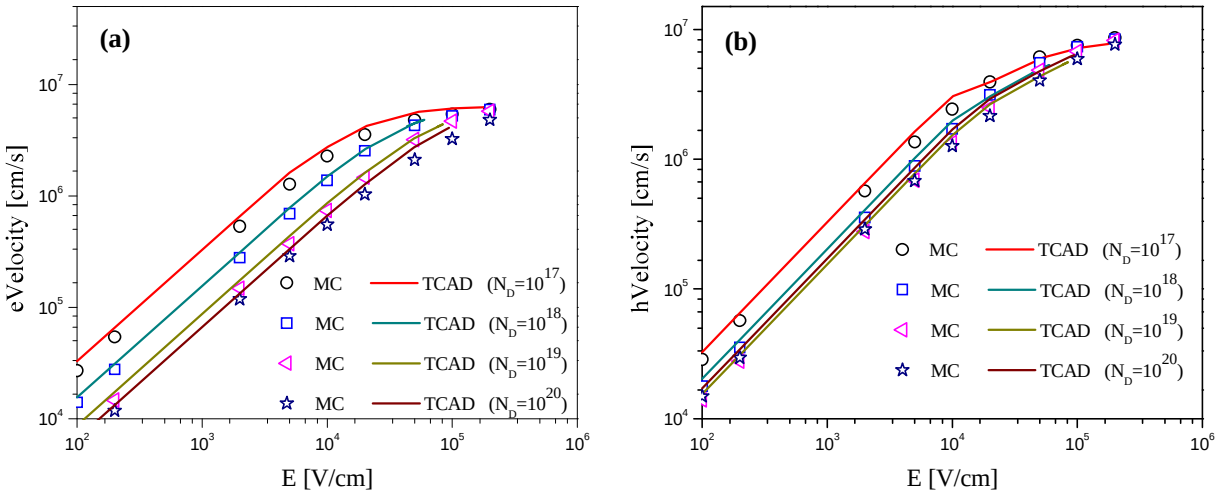

Fig.160: Comparison between MC and TCAD simulation results using Model (**II**) for (a) eVelocity vs. Electric field (b) hVelocity vs. Electric field, (Ge_content = 18%, T=300K).

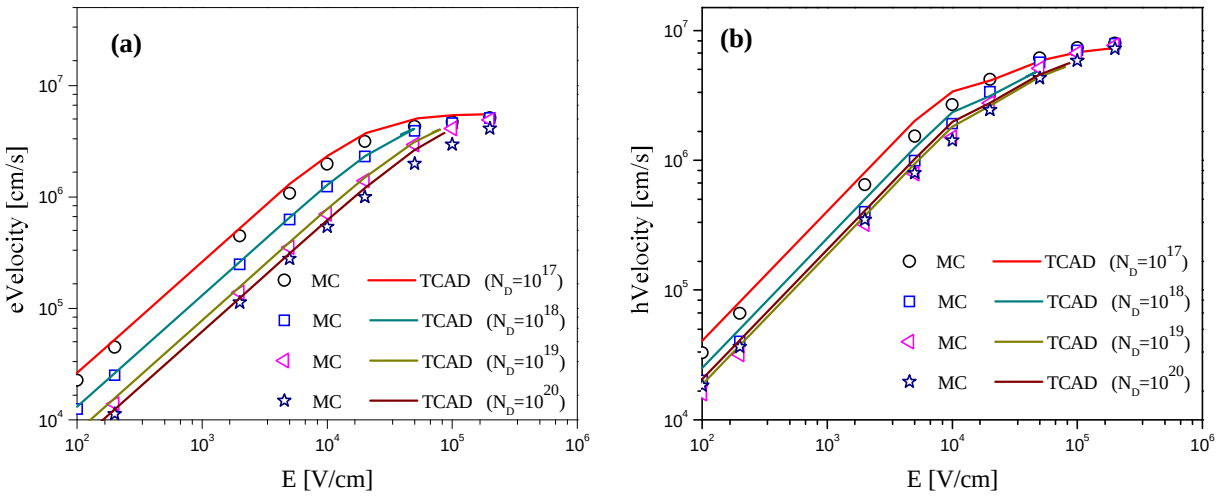

Fig.161: Comparison between MC and TCAD simulation results using Model (**II**) for (a) eVelocity vs. Electric field (b) hVelocity vs. Electric field, (Ge_content = 28%, T=300K).

2. Transport Models Calibration

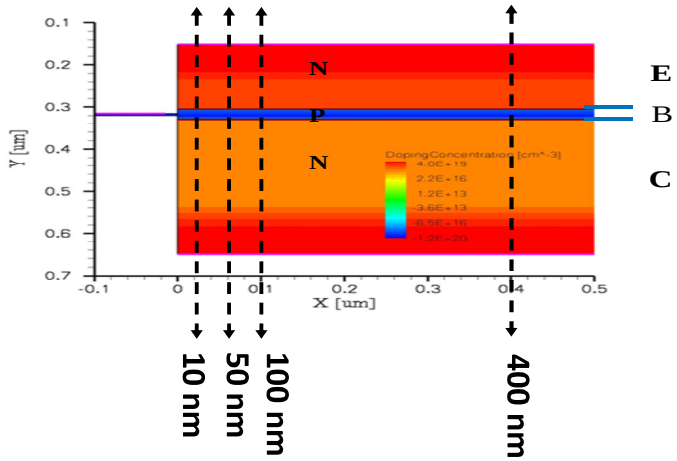

Fig.162 : The reference transistor with different cross sections used in the calibration process.

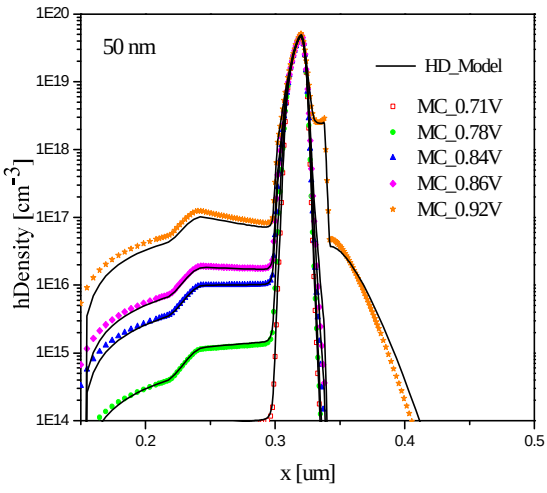

Fig.163: Hole density comparison between MC and HD model simulation results for the cross-section 50 nm

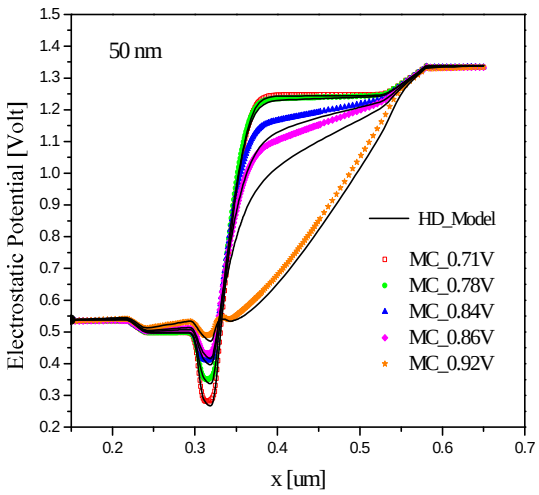

Fig.164 : Electrostatic potential comparison between MC and HD model simulation results for the cross- section 50 nm

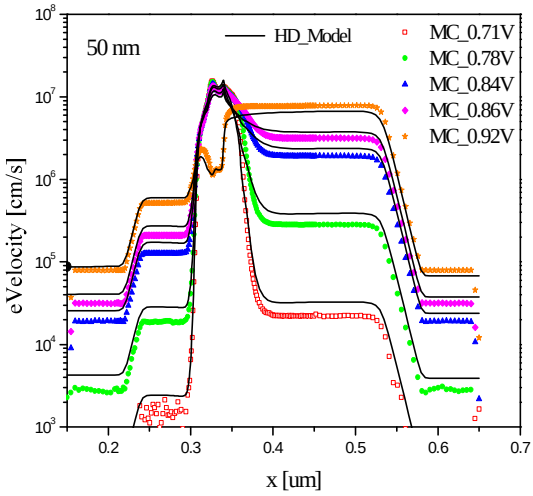

Fig.165: Electron velocity comparison between MC and HD model simulation results for the cross-section 50 nm

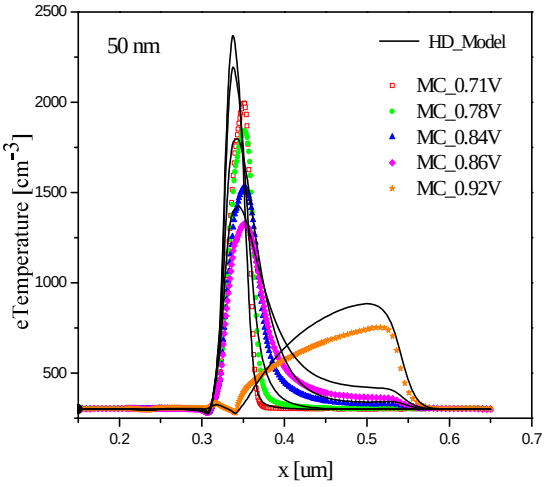

Fig.166: Electron temperature comparison between MC and HD model simulation results for the cross-section 50 nm

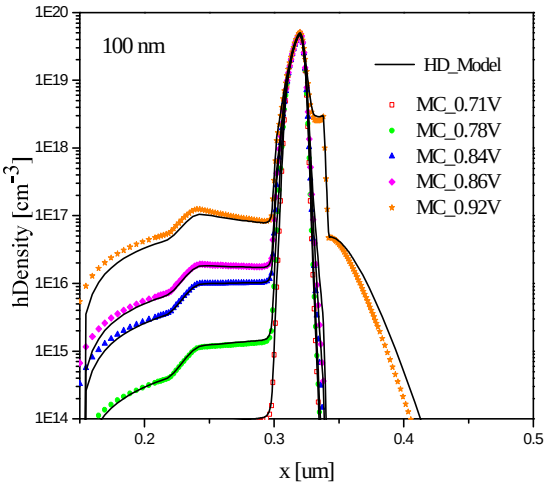

Fig.167: Hole density comparison between MC and HD model simulation results for the cross-section 100 nm

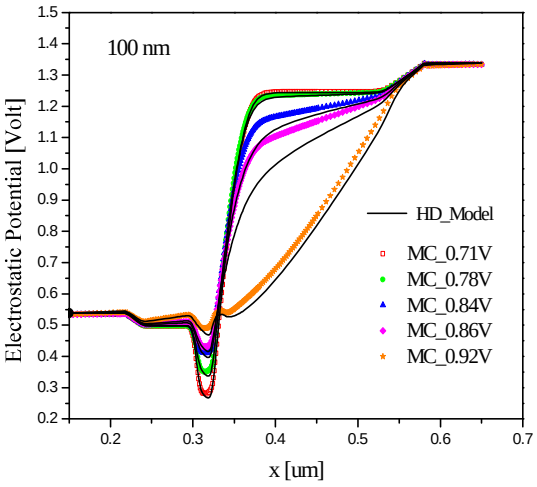

Fig.168 : Electrostatic potential comparison between MC and HD model simulation results for the cross- section 100 nm

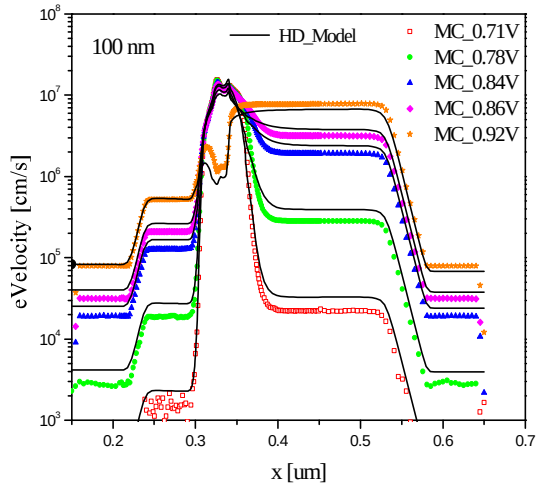

Fig.169: Electron velocity comparison between MC and HD model simulation results for the cross-section 100 nm

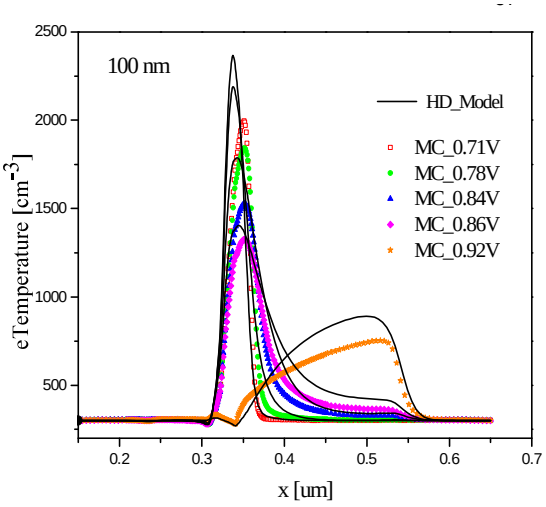

Fig.170: Electron temperature comparison between MC and HD model simulation results for the cross-section 100 nm

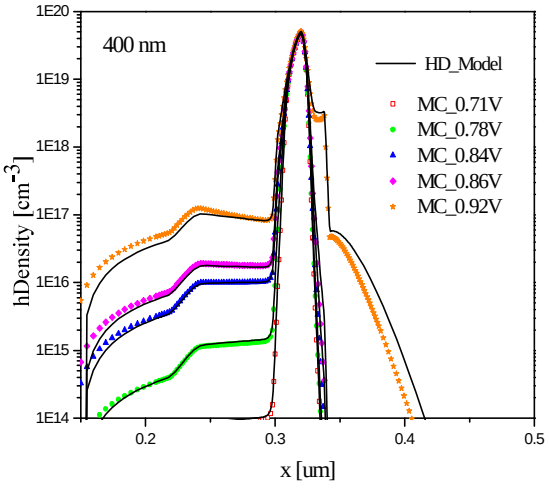

Fig.171: Hole density comparison between MC and HD model simulation results for the cross-section 400 nm

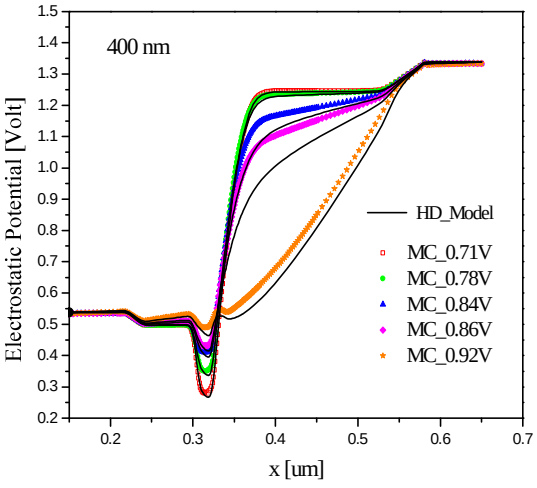

Fig.172 : Electrostatic potential comparison between MC and HD model simulation results for the cross- section 400 nm

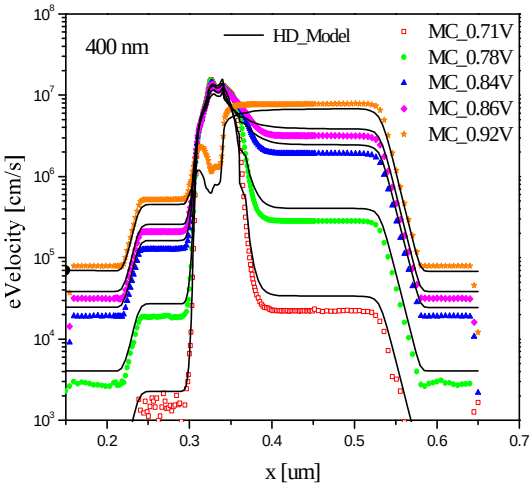

Fig.173: Electron velocity comparison between MC and HD model simulation results for the cross-section 400 nm

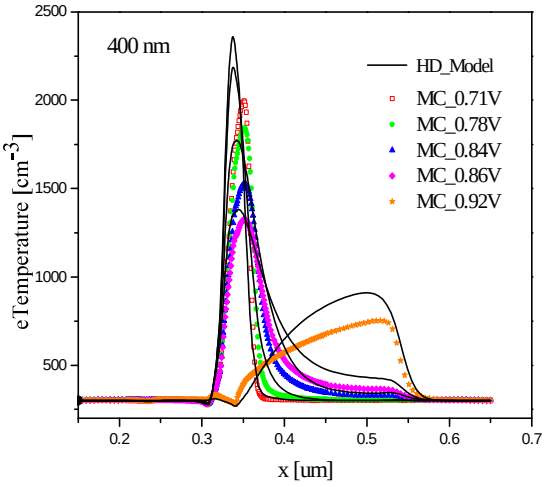

Fig.174: Electron temperature comparison between MC and HD model simulation results for the cross-section 400 nm

List of Publications:

1) **M. Al-Sa'di**, S. Fregonese, C. Maneux, T. Zimmer, "TCAD Modeling of NPN-Si-BJT Electrical Performance Improvement through SiGe Extrinsic Stress Layer", *Mat Sci Semicond Process*, doi:10.1016/j.mssp.2011.03.002, 2011.

2) **M. Al-Sa'di**, S. Fregonese, C. Maneux, T. Zimmer, "Modeling of a Novel NPN-SiGe-HBT Device Structure Using Strain Engineering Technology in the Collector Region for Enhanced Electrical Performance", *Proceedings IEEE BCTM 2010*, p. 216-219, Austin, TX, USA, 2010.

3) **M. Al-Sa'di**, V. d'Alessandro, S. Fregonese, S.-M. Hong, C. Jungemann, C. Maneux, I. Marano, G. Matz, A. Pakfar, N. Rinaldi, G. Sasso, M. Schröter, A. Sibaja-Hernandez, C. Tavernier, and G.Wedel, "TCAD simulation and development within the European DOTFIVE project on 500GHz SiGe:C HBT's", *Proceedings EuMIC 2010*, p. 29-32, CNIT, Paris, France, 2010.

4) **M. Al-Sa'di**, S. Fregonese, C. Maneux, and T. Zimmer, "Modeling of NPN-SiGe-HBT Electrical Performance Improvement through Si3N4 Strain in the Collector Region", *ECS Transactions*, 33 (6) 191-199, 2010.

5) **M. Al-Sa'di**, S. Fregonese, C. Maneux, and T. Zimmer, "Modeling of NPN-SiGe-HBT Electrical Performance Improvement through Si3N4 Strain in the Collector Region", Meet. Abstr. - Electrochem. Soc. 1002, 1873 , *ECS 218th meeting 2010*, Las Vegas, NV, USA, 2010.

6) **M. Al-Sa'di**, S. Fregonese, C. Maneux, and T. Zimmer, "TCAD Modeling of NPN-SiGe-HBT Electrical Performance Improvement through Extrinsic Stress Layer", *Proceedings IEEE MIEL 2010*, p. 387 - 390, Niš, Serbia, 2010.

7) **M. Al-Sa'di**, S. Fregonese, C. Maneux, T. Zimmer, "Investigation of Electrical BJT Performance Improvement through Extrinsic Stress Layer Using TCAD Modeling", *IEEE SCD 2009*. Dresden, Germany, 2009.