Impact of Various Scaling Techniques and Base Region Optimization on the Speed of Si-Ge HBT

ARUN KUMAR
Department of Electrical and Electroncs Engineering
KIET Group of Institution
Ghaziabad Uttar Pradesh
INDIA
aarun115@gmail.com

R.K. CHAUHAN
Department of Electroncs and Communication Engineering
Madan Mohan Malaviya Engineering College
Gorakhpur, Uttar Pradesh – 273010
INDIA
rkchauhan27@gmail.com

,

***Abstract*:-** Nowadays SiGe HBTs are surpassing even the fastest III-V production devices in the GHz speed orbit. Scaling has been the prime source of supremacy behind the successful technology innovations. The impact of different Ge concentration of graded profile as well as The design and optimization of un-scaled and scaled SiGe HBTs on the key performance parameter which affect the speed parameter in GHz frequency range has been explored extensively, which have recently emerged as a tough competitor for RF and mixed-signal applications. Both vertical and lateral scaling affects the device parameters. The impact of lateral scaling on cut off frequency is not as prominent as vertical scaling, but a certain level of lateral scaling needs to go along with vertical scaling Since maximum frequency of oscillation is absolutely associated to cut off frequency, the remuneration of vertical scaling on cut off frequency also apply to maximum frequency of oscillation, although the impact level is comparatively lower. However, maximum frequency of oscillation depends on R_B and C_{CB}, too, which are truly degraded by vertical scaling. The simulated results of un-scaled and scaled SiGe HBT are compared and contrasted.

***Key- words*:-** SiGe HBT, Scaling, Doping Profile (Graded), Speed Parameter Genius Device Simulator.

1 Introduction

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

The principle function of an HBT is alike to that of the BJT, except that the band gap of the emitter region exceeds that of the base region ΔE_g, normally of the order of 0.10–0.2 eV. The resultant $e^{\Delta E_g / kT}$ exponential increase in current gain permits scaling of the base region to slighter thickness and high doping levels.

As the difficulty of the integrated circuits increases, scalable device models are becoming crucial for the optimum circuit design. Scaling of bipolar transistor models is significantly more involving due to current flow in both vertical and lateral guidelines. Moreover, the geometry scaling policy of bipolar devices normally depend on the layout pattern, device cross-section and fabrication progression particulars. As a result, the geometry scalable bipolar transistor models and the analogous parameter drawing out procedures are not usually

available and usually exist as a domestic proprietary. The physical quantities in a bipolar transistor that scales with geometry can be estranged into three categories: (i) current and charge, (ii) ratio of current and charge and (iii) parasitic and thermal resistance.

SiGe HBTs are particularly exciting because of their ability to take immediate advantage of highly developed Silicon processing techniques. Impressive improvements in high-speed SiGe bipolar technology have been made through the growth of device quality strained–Si1-xGex layers. This strain, which occurs because of a ~4 % difference in the lattice constants of Si and Ge, is used to vary the bandgap energy, band discontinuity and other properties of the material.

The SiGe HBT's speed, yield, linear characteristics and high gain make it very appropriate for many applications. The existing applications of SiGe are mostly in the telecommunications industries where high frequency analog and miscellaneous signal circuits are used. As an outcome, contemporary SiGe technology has been fairly optimized for analog applications. By grading Ge concentration in the base fully across the neutral base; HBT transistors have a high early voltage with a comparatively high cut-off frequency (f_T).

In this paper, an analysis has been done to study the importance of scaling as well as doping profile (Graded) on the performance parameter which affects the speed parameter of $Si_{1-x}Ge_x$ HBT. Cut off frequency, maximum oscillation frequency, C_{BC}, C_i and transit time delay of scaled and un-scaled device are calculated for different Ge concentrations with the help of appropriate scaling technique. It is shown that the scaling, both vertical and lateral, has mostly advantageous effects on speed performances of SiGe HBT. Following this motivation, in the next section, need of scaling and various scaling techniques has been discussed. After this section, base region optimization with graded profile and speed parameter has been reviewed. In Section-5 & 6, we discuss the results and concluded with common projections and observations based on Genius Device Simulator.

2 NEED OF SCALING AND SCALING TECHNIQUES

The performance of semiconductor devices tends to perk up as the device scaling is done. This uncomplicated approach of scaling has been the key to the luminous success of semiconductor industry over the last three decades. It has worked for sensibly all types of transistors, including the Si-based bipolar transistor. Scaling has sprint into visible hard limits multiple times in the track of bipolar technology progression, which have been prolifically overcome with help from material and structural innovations, such as the self-aligned base, poly emitter, epitaxial base and most lately the SiGe base [4].

A scaling regulation provides a set of favored device design parameters and expected performance parameters for a specified scaling factor, based on necessities for proper device operation and understood constraints. The scaling factor generally pertains to the lateral dimension which is usually limited by the lithography. For bipolar transistors, the reliance on lateral scaling is less critical and their scaling rule has been of inferior interest than CMOS case. Nonetheless, there have been substantial efforts to establish useful scaling rules for bipolar transistors, too, as is reviewed below [4].

Rosseel took a fairly different approach in developing a scaling rule for bipolar transistors entrenched in BiCMOS logic gates (which consist of both bipolar and CMOS devices as inverter components), in which better performance of BiCMOS gates over CMOS gates was taken as a restraint [5]. Solomon estimated a bipolar scaling rule for transistors in ECL circuits assuming constant voltage (CV) and constant current (CI), and estimated that the gate delay tends to beg off at the same rate as lateral scaling [6].

Bellaouar developed a bipolar scaling rule based on CI format, but with an unmitigated set of self-governing scaling factors, each analogous to lateral dimension, vertical dimension, and voltage crossways base and collector [7]. Raje also recognized a scaling rule for bipolar devices affianced forBiCMOS gates, but took into account the reality that bipolar and CMOS devices in a gate cannot be considered separately and interactions should subsist between these devices in terms of scaling [8]. The bipolar scaling rules from these assorted approaches are summarized in Table 1. Apart from the different constraints and approaches assumed, overall trends suggested by these scaling rules for key bipolar parameters are not significantly far apart. For example, a scaling factor of $1 / \gamma$ - $1/ \gamma^{1.5}$ is expected for the gate delay, while current

density is expected to increase by ~ γ^2 from nearly all scaling rules.

Table 1

	Rosseel [5] $\gamma, \beta, \alpha = 1.15$	Solomon [6] $\gamma = 1.15$	Bellaouar [7] $\gamma_h, \gamma_v = 1.15$	Raje [8] $\gamma = 1.15$
Emitter Width- W_E	$1 / \gamma$	$1 / \gamma$	$1 / \gamma_h$	$1 / \gamma$
Emitter Length- L_E	$1 / \gamma$		$1 / \gamma_h$	$1 / \gamma$
Base Thickness W_B	$\beta^{0.5}/\gamma^{0.5}\ \alpha^{0.5}$	$1 / \gamma^{0.8}$	$1 / \gamma_v$	$1 / \gamma$
Collector Thickness W_C	$1 / \gamma$		$1 / \gamma_v$	
Base Doping N_B	$\gamma\ \beta / \alpha$	$\gamma^{1.6}$		$\gamma^{1.5}$
Collector Doping N_C	$\gamma^2 / \beta k$	γ^2	γ_h^2	γ

3 BASE REGION OPTIMIZATION

The design of SiGe HBT, for a scrupulous technology generation is optimized by suitable scaling of the device. A SiGe HBT offers further design flexibility in that the band gap of the base may be customized by grading the Ge concentration. Reducing the width of the base region reduces the base transit time with related enhancement in cut-off frequency, but inescapably increase overall base resistance. For
effectual design, it is thus vital to use a right Simulation tool. Devices like SiGe HBTs designed with such techniques do have improved performance.

Even though the introduction of Ge in the base increases process integration complication, it offers an extra degree of freedom which relaxes a series of trade-offs affecting device design.

SiGe HBTs are bipolar transistors grown-up on a silicon substrate that reveal speed advantages for various reasons. The transistors are in reality not heterojunction bipolar transistors but rather base

graded transistors. The alloy grading present in the base results in a valence band offset connecting the SiGe and the silicon interface. The valance band offset helps to detain holes in the SiGe layer, thus reducing the reverse injection of holes from the emitter region into the base region during the forward bias of the transistor. The existence of Ge in the base layer also contributes to speed improvements by straining the Si layer; the improved atomic spacing allows for higher electron mobility. The net outcome is that SiGe HBTs offer a lower barrier to electron injection into the base than a similarly configured conformist bipolar junction transistor. Some of this can be traded for lower base resistance. The built in field resulting from Ge grading also enhances speed.

Figure:1 indicates to the important dc effect of adding Ge into the base of SiGe HBT, however, lies with the collector current density. Physically, the barrier to electron injection at decreases the base Gummel number and, therefore, increases collector current density. Imperative in this context is the Ge-induced enhancement in β over a comparably constructed Si BJT.the EB junction is abridged by introducing Ge into the base, yielding more charge transfer from emitter-to collector for a given applied EB bias. Observe from Figure (1)

Figure 1: Graded Ge layer in the base of SiGe HBT

that in this linearly graded base design, the emitter region of the SiGe HBT and Si BJT comparison are primarily identical, implying that the resultant base current density of the two transistors will be approximately the same. The net result is that the introduction of Ge increases the current gain of the transistor (β=collector
current density / base current density).

4 SPEED PARAMETER

Brilliant speed performance is preferential for most practical semiconductor applications of today. The effectiveness of information processing stoutly depends on the speed of devices that compose the system. The speed of a device can be represented by a variety of measures and corresponding speed parameters. Even though some alternatives have been projected [9], the most extensively used speed parameters for transistors are the cutoff frequency and the maximum oscillation frequency, which are defined as the frequency point where the current gain and the power gain become unity, correspondingly.

4.1 Unity gain cut-off frequency, f_T

f_T is defined as the frequency at which the common emitter short circuits ac current gain is unity [10]. It is associated physically to the bipolar device, as the total delay for the minority carrier transversely the device from emitter to collector, τ_{ec}. The whole delay consists of the minority carrier stored charge delay and the junction capacitance charging delay, is often associated to f_T through the equation:

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

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

$$\tau_{ec} = \tau_e + \tau_{eb} + \tau_b + \tau_{bc} + \tau_{je} + \tau_c \qquad (2)$$

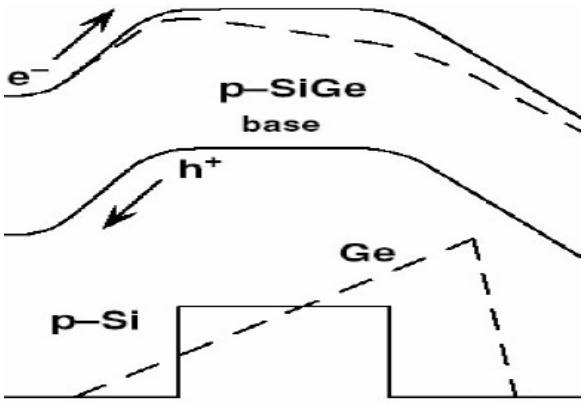

Fig.1

The main components, due to minority carrier stored charge, are τ_e for the neutral emitter and τ_b for the neutral base region. The term τ_{eb} represents minority carrier transit time in the emitter-base depletion region, and is often small adequate to be included in the emitter transit time tenure. The transit time τ_b, the delay due to the surfeit minority carrier storage in the base, is usually the most significant term.

The delay term τ_{je} is the entire charging time allied with emitter base and base collector depletion layers and is given by

$$\tau_{je} = \frac{kT}{qI_c}\left(C_{je} + C_{jc}\right) \qquad (3)$$

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

From equation (3) it is clear that τ_{je} is governing at low collector current, and therefore f_T tends to increase with increase in I_c. However, the influence of $\tau_{je,}$ reduces radically as the collector current continues to increase. Hence the equation of cut-off frequency can be written as

$$f_T = \frac{g_m}{2\pi\left(C_{je} + C_{jc}\right)} \qquad (4)$$

4.2 Maximum oscillation frequency, fmax

The unity gain cut-off frequency provides a good proposal of the intrinsic delay associated with a bipolar transistor. However, it is not a realistic parameter for a circuit surroundings, as it assumes that the output is short circuited. In addition, it is autonomous of base resistance and hence does not seize the base resistance base-collector depletion capacitance time constant into elucidation. These are significant parameters for shaping the transient behavior of bipolar circuits. Therefore, another more realistic and widely predictable figure-of-merit, f_{max}, is usually used, which characterizes the power transfer in and out of the bipolar device. f_{max} is defined as the frequency at which the unilateral current gain becomes unity. Here the output is basically isolated from the input by a suitable exterior matching circuit compromising reactive and

resistive components. The load that it drives is also vague to be conjugate matched to the transistor output impedance. It can be shown that:

$$f_{\max} = \sqrt{\frac{f_T}{8\pi C_{jc} R_b}} \qquad (5)$$

Where, R_b is the base resistance. Equation (5) shows that it is not sufficient to obtain a high value f_T, by declining base width, but that base resistance and base-collector capacitance should be minimized. However, as base width decreases quickly to achieve high f_T, R_b will enlarge except the doping is increased. To contradict that effect, the base desires to be more highly doped, which means that emitter doping has to be lowered to put off emitter-base junction tunneling for very high base doping levels. The amplified current gain ability of a SiGe base enables lowering of emitter doping without exposing sufficient current gain.

5 RESULTS & DISCUSSIONS

To study the impact of dreaded profile doping on cut-off frequency and maximum oscillation frequency, Ge concentration has been taken in the range of 5%-10%, 10%-20%, 15%-30% and 20%-40% respectively. In the range of 5% -10%, Ge concentration varies from 5% to 10% in the base width of .025 µm, in the base width of .075 µm Ge concentration is 10% which is constant and Ge concentration again varies from 10% to 5% in the base width of .030 µm to make a graded profile. In figure:2, cut-off frequency is on left vertical axis and maximum oscillation frequency is on the right vertical axis. Red plot shows cut-off frequency and green plot shows maximum oscillation frequency. At graded Ge concentration range 5%-10%, f_T and f_{max} are 26.33 GHz and 32.68 GHz respectively. At the range of 20%-40% Ge concentration f_T and f_{max} are 87.65 GHz and 96.84 GHz respectively. From figure:2 it is clear that on increasing the Ge concentration of graded profile, cut-off frequency and maximum oscillation frequency increases.

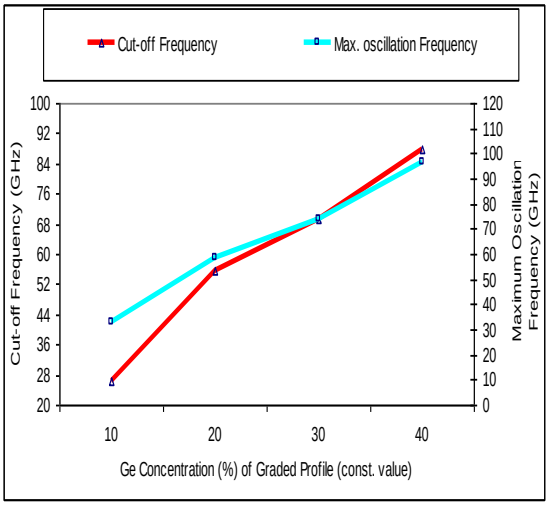

Fig.2

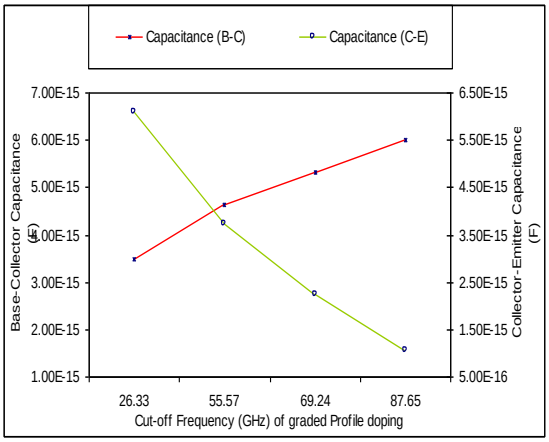

Fig.3

Figure:3 shows the variation of capacitances (Base-Collector Capacitance and Collector-Emitter Capacitance) with respect to cut-off frequency of graded profile doping done in the base region. Red plot shows C_{BC} and green plot shows C_{CE}. At 26.33 GHz cut-off frequency base-collector capacitance is $3.49*10^{-15}$ and collector-emitter capacitance is $6.10*10^{-15}$. After increasing the Ge concentration to the range of 20%-40%, the C_{BC} and C_{CE} are $6.02*10^{-15}$ and $1.06*10^{-15}$ respectively. It is clear from the figure that on increasing the Ge concentration or increasing the cut-off frequency, C_{BC} increases but C_{CE} decreases.

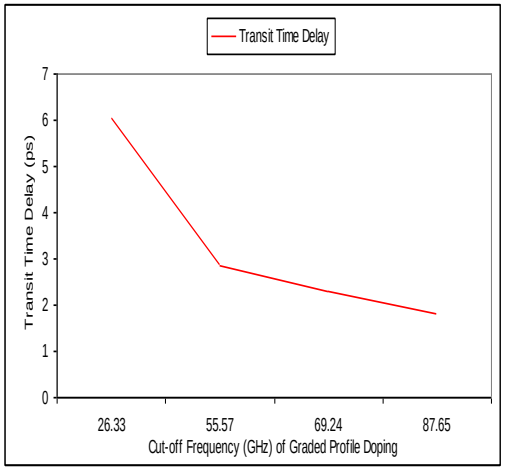

Fig.4

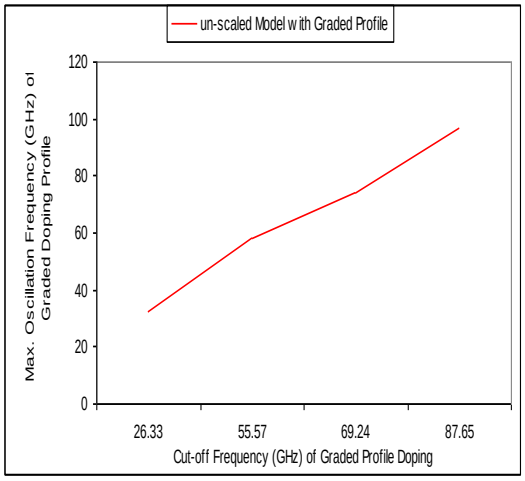

Fig.5

Figure:4 shows the variation of transit time delay with respect to cut-off frequency of graded profile doping. At 26.33 GHz cut-off frequency transit time delay is 6.04 ps and at 87.65 GHz cut-off frequency transit time delay reaches to 1.81 ps. So, on increasing cut-off frequency, transit time delay decreases

Figure:5 shows the plot of un-scaled model with graded profile in which variation of maximum oscillation frequency with respect to cut-off frequency is given. Above figure shows that on increasing cut-off frequency, maximum oscillation frequency increases.

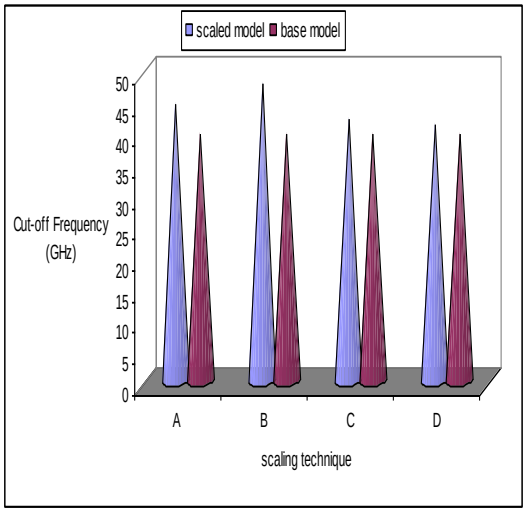

Fig.6

In table:1 various scaling techniques have been given. The scaling of the base model of the device has been done with the help of all the scaling technique given in the table:1. the result of this scaling is given in figure:6 in the term of cut-off frequency. In the figure:6, A stands for Raje's scaling technique, B stands for Solomon's scaling technique, C stands for Bellaouar's Scaling technique and D stands for Rosseel's Scaling technique. Blue cone shows the cut-off frequency of scaled model and red cone shows the cut-off frequency of base model (un-scaled model). Solomon's scaling technique gives highest cut-off frequency (48.09 GHz) of scaled device in comparison to un-scaled device (39.92 GHz). So for the further analysis Solomon's scaling techniques has been considerd.

The analysis of scaled and un-scaled SiGe HBT has been done for constant Ge concentration. Figure:7 shows the variation of cut-off frequency of un-scaled and scaled SiGe HBT with respect to Ge concentration. The range of Ge concentration has been taken in the analysis is 5% to 30% as higher to this are not supported by present epitaxial technologies and beyond it the improvement associated with Ge seizes may be due to lattice constant mismatch. Left vertical axis shows cut-off frequency of un-scaled device and right vertical axis shows cut-off frequency of scaled device. At 10% Ge concentration, cut-off frequency of base model and scaled device is 19.34 GHz and 25.43 GHz respectively. At 30% Ge concentration cut-off frequency of base maodel and scaled model is 43.32

GHz and 47.82 GHz respectively. Hence, it is clear from the figure that on increasing Ge concentration, cut-off frequency increases.

Figure:8 shows the variation of maximum oscillation frequency with respect to Ge concentration. In this figure black dotted line shows plot for scaled model and green dashed line shows plot for scaled model. In the figure scaled model plot is above the base model, so it can be said that on scaling the device with suitable scaling technique, maximum oscillation frequency increases of scaled device. At 5% Ge concentration, max.oscillation frequency of base and scaled model is 28.37 GHz and 37.2 Ghz respectively.

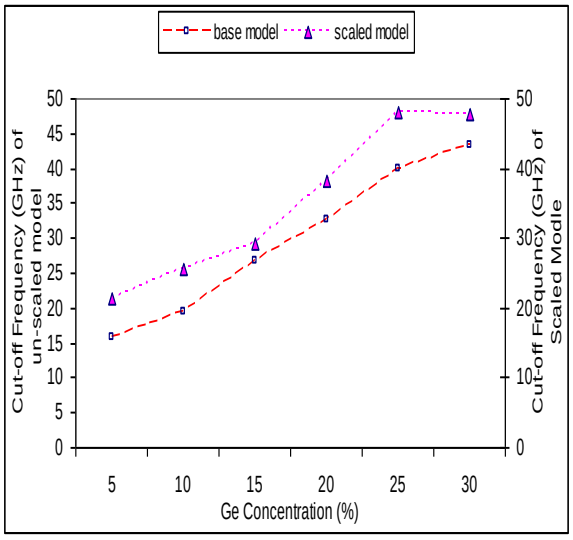

Fig.7

At 25% Ge concentration, max. Oscillation frequency of base and scaled model is 54.99 GHz and 67.8 Ghz respectively. It is clear from above discussion that on increasing Ge concentration, max. oscillation frequency increases.

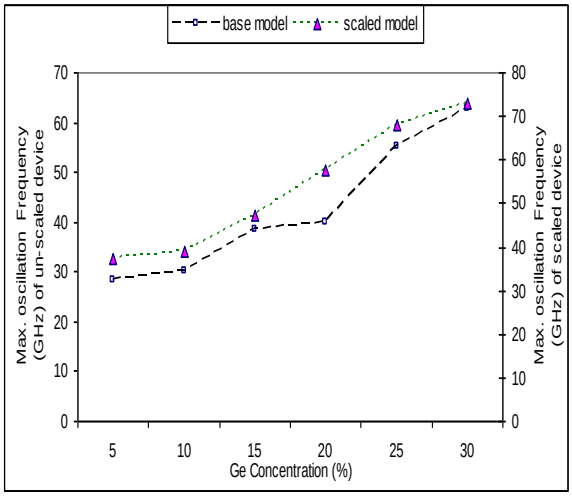

Fig.8

Figure:9 shows the variation of transit time delay with respect to Ge concentration of base model and scaled model. Orange dotted line shows the transit time of base model (left vertical axis) and blue dashed line shows the transit time of scaled model (right vertical axis). At 25% Ge concentration transit time delay of base model and scaled model is 3.92 ps and 3.3 ps respectively. At 30% Ge concentration transit time of base model and scaled model is 3.67 ps and 3.32 ps respectively.

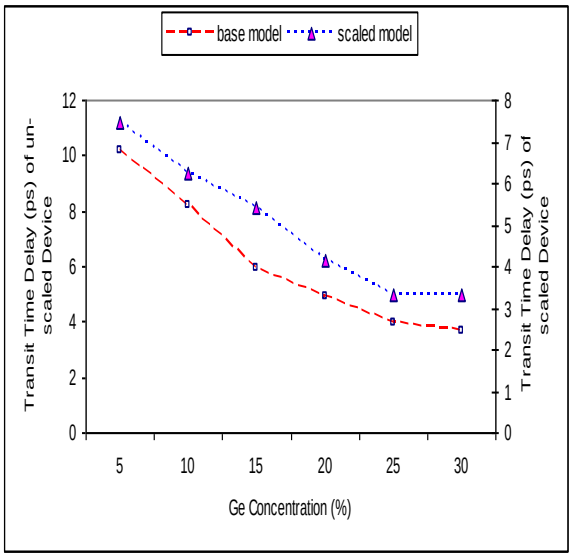

Fig.9

Figure:10 shows the variation of current density of un-scaled and scaled device with respect to Ge concentration. Left vertical axis shows current

density of un-scaled device and right vertical axis shows current density of scaled device. Black dotted line shows the current density of base model which is below the green dashed line (current density of scaled model), hence it can be said that on scaling the device current density of the scaled device increases.

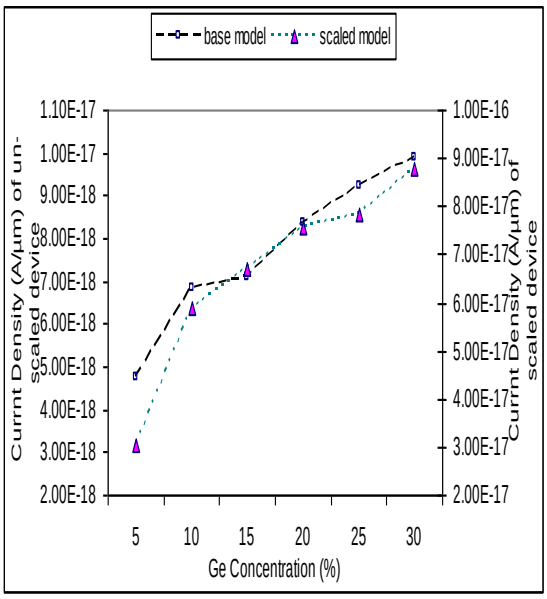

Fig.10

6 CONCLUSION

In this paper, a detailed investigation has been made to study the impact of Ge content of graded profile in the SiGe HBT base and various scaling techniques on the performance parameter of SiGe HBT. An electric field is produced by position in the impartial base region. This effect aids the transportation of minority carriers (electrons) from emitter to collector, which in turn perk up the frequency response. The parameters like C_{BC}, C_{CE}, current density and transit time delay which strongly affect the f_{max}/f_T are calculated of un-scaled and scaled SiGe HBT as well as for graded profile. The scaling of the device and base-region optimizations with graded profile can effectively improve the speed performance of $Si_{1-x}Ge_x$ HBTs in a wide frequency (GHz) range. It can be said that on increasing the Ge content of graded profile in the base of device, transit time delay and C_{CE} decreases and hence cut-off frequency & maximum oscillation frequency increases. On the other hand, after scaling the device with appropriate scaling technique, again transit time delay decreases. From equation (1) & (10) it is clear that on the decrement of transit time delay, cut-off frequency and maximum oscillation frequency increases and the device becomes faster. On comparing different scaling technique in terms of cut-off frequency, Solomon's scaling technique with scaling factor 1.15 has been considered for their better response. After scaling the device, the cut-off frequency increases by 20.64% that of un-scaled device and maximum oscillation frequency increases by 23.29% that of un-scaled device at 25% Ge concentration. The value of transit time delay of scaled model decreases by 18% from mase model. Based on simulation results of un-scaled SiGe HBT and scaled SiGe HBT, it is confirmed that applied scaling technique is effective to obtain better performances of the scaled device.

References:

1. Shockley W, ''US patent, Specification 2569347'', 1951.
2. Kroemer H, ''Theory of a wide-gap emitter for transistor'', Proc. IRE 45, 1535-7, 1957.
3. Larson LE, "Integrated Circuit technology options for RFICs-present status and future directions", IEEE , J.solid-state circuits, **33,** 387-99, 1998.
4. Jae- Sung Rieh, David Greenberg, Andreas Stricker and Greg Freeman " *Scaling of SiGe Heterojunction Bipolar Transistors", Proceedings of the IEEE, VOL. 93, NO. 9, September 2005.*
5. G. P. Rosseel and R.W. Dutton, "Scaling rules for bipolar transistors in BiCMOS circuits," in *Tech. Dig. Int. Electron Devices Meeting*, 1989, pp. 795–798.
6. P. M. Solomon and D. D. Tang, "Bipolar circuit scaling," in *Dig. Int. Solid-State Circuits Conf.*, 1979, pp. 86–87.
7. A. Bellaouar, S. H. K. Embabi, and M. I. Elmasry, "Scaling of digital BiCMOS circuits," *IEEE J. Solid-State Circuits*, vol. 25, no. 4, pp. 932–941, Aug. 1990.
8. P. A. Raje, K. C. Saraswat, and K. M. Cham, "Performance-driven scaling of BiCMOS technology," *IEEE Trans. Electron*

9. *Devices*, vol. 39, no. 3, pp. 685–694, Mar. 1992.
10. R. A. Gosser, O. Foroudi, and S. Flanyak, "New bipolar figure of merit 'fo'," in *Proc. Bipolar/BiCMOS Circuits and Technology Meeting*, 2002, pp. 128–135.
11. Gao G-B, Fan Z-F and Morko H 1991 Analysis of cut-off frequency roll-off at high currents in SiGe double-heterojunction bipolar transistor Appl. Phys. Lett. 58 2951-3.