CHAPTER - 3

Artificial Neural Network based Base Transit Time Prediction for better HBT

3.1 Introduction

3.2 Literature Survey

3.3 Artificial Neural Network Based Prediction Technique

3.3.1 Sequence of Steps in Back-Propagation

3.3.2 Back Propagation Neural Network Training

3.4 Results and Discussions

References
3.1 INTRODUCTION

Heterostructure Devices using strained Si and SiGe layers have shown promise for a better replacement of all Si devices. However, the following practical problems are encountered in Si/SiGe systems. First, in Si/SiGe devices, Boron outdiffusion across the emitter-base and collector-base junctions deteriorates device performance. Second, the lattice mismatch between Si and Ge in the Si base prevents fabrication of a triangular Ge profile, although known to yield the lowest base transit time ($\tau_b$), which leads to poor film stability [3.1]. Si/SiGe systems are not suitable for process integration [3.2]. All the above problems in SiGe base devices are eliminated when a slight amount of C is incorporated in the SiGe alloy. Incorporation of substitutional C increases the base band gap of a strained SiGeC base in Si/SiGe/Si heterojunction bipolar transistors (HBTs) by 26 meV/% of C with the same amount of Ge [3.3]. However, the disadvantage of increase in band gap is more than compensated by the reduction of strain, as 1% of C can reduce the strain produced by 8-10% of Ge [3.4]. In addition to this the measured charge carrier densities at room temperature are not affected substantially by incorporation of small amounts of C (approximately 1%) under identical growth conditions and dopent fluxes [3.5]. In view of the technological interest to use of SiGeC as the base material in Si HBTs, it seems worthwhile to examine how the performance of the devices changes due to incorporation of C. In this work we have considered the model and
the generalized expression for \((\tau_b)\) of a SiGeC-base HBT [3.6] and predict the dependent parameters using artificial neural network (ANN).

An artificial neural network (ANN) is a computational tool having artificial intelligence origins, also known as a neural net. ANN is a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs. Presently they are being used for a wide range of applications covering character recognition, speech recognition, motor control, data framing and many more areas of applications [3.7].

In the present work, an ANN is employed to predict the best amount of carbon to be incorporated in the base of transistors so that the relevant device will exhibit best high frequency response characterized by a desirable cutoff frequency.

### 3.1 Literature survey

Silicon technology has made remarkable progress in the past five decades since Kilby’s invention: \(10^{6-8}\) X in active device count per chip, \(10^{6-8}\) X in switching speed and \(10^{5-7}\) X in cost per function or bit [3.8 – 3.10]. The stated progress has been largely supported by evolution of lithography besides paving the way for the invention of MOS transistors with self aligned gate, followed by CMOS, ICs Better understanding of materials, processing, and device physics that
are leading to continuous improvement in manufacturing technologies. Faster devices are of prime requirements for the present time so that we can get faster processing and lesser access time. In order to improve the speed of device, mobility of the carriers can be enhanced.

These are several ways to improve the carrier mobility. A transistor with enhanced mobility will show a higher on-current than a transistor with lower channel mobility and otherwise identical design. The most common and popular approach to enhance the mobility is to use strained Silicon channels. Unfortunately, replacing Silicon with a new material does not necessarily provide the performances enhancement expected from the higher mobility due to other short comings or performance challenges of the new material. Essentially this is one of the main reasons behind the recent excitement over strained silicon. It is now well established that without changing the channel material, significant enhancements in device performance are possible using strained silicon [3.11 – 3.14]. The mechanism of mobility enhancement in strained Silicon is still evolving. The explanation for such mobility enhancement is that under the biaxial tensile strain, the six-fold degenerate valleys in silicon are split into two groups. In BJT, the base region is strained to make the base transit time less thereby reducing the base current and making the transistor ideal. For bipolar transistor, the base transit time is an important and often used figure of merit. A group of
researchers has taken this area as prime research area and has already produced significant results [3.15 – 3.20].

Comparing the different semiconductor technologies and taking into the account their actual performance/cost ratio, the current understanding is that the silicon technology is and will be the basis of IC fabrication. Due to new applications there is, however, a market’s demand to overcome the cut off frequency limit inherent in silicon technology [3.21 – 3.24]. The existing solutions may be challenged by SiGe or other silicon-based heterostructure devices.

All the problems that are encountered in SiGe base devices like Boron out diffusion, lattice mismatch and complex process integration are eliminated when slight amount of C is incorporated in the base. The value of Avalanche voltage ($V_A$) is found to increase with the incorporation of small amounts of C in the SiGe base due to its ability to prevent boron out-diffusion as reported for a fabricated device. Further improvement in the value of $V_A$ is accounted due to the sharp fall-off of the C profile at the collector –base junction. Also, for the same amount of strain between Si and SiGe, and between Si and SiGeC, the latter yields a larger value of $V_A$. It was demonstrated almost ideal base current characteristics for SiGeC HBTs even with C in the space charges regions. So, non-ideal characteristics are not fundamental property of C containing devices.
Besides comparable 1/f noise, SiGeC devices had shown (I_b driven) early voltages-current gain products greater than 20V, exceeding values known for state-of-the-art SiGe HBTs [3.25 – 3.26]. The high frequency performance of SiGeC also benefits from C incorporation into the SiGe base layer because it allows one to use a high B doping level in a very thin SiGe layer without outdiffusion from SiGe, even when applying post-epitaxial implants. Addition of C can suppress Transient Enhanced Diffusion (TED) caused by a variety of subsequent processing steps in particular, the limitation for lateral emitter scaling in a single poly silicon technology, resulting from TED due to implantation of the external base regions can be eliminated. Recent developments show that a simple, epi-free collector well, single polysilicon Si/SiGeC/Si HBT module offers $f_T$ ($f_{max}$) values of upto 200 (225) GHz range. So whether or not the hetero CMOS will overcome the cutoff frequency limit of silicon or whether they are challenged by other approaches such as gigascale single-electron devices, never the less SiGeC alloys suggest many possibilities for the future.

### 3.3 ARTIFICIAL NEURAL NETWORK BASED PREDICTION TECHNIQUE

ANN generally contains a set of equations associated with a randomly chosen group of coefficients and rules for adjusting those coefficients in the time of training. In order to train the ANN back propagation performs a gradient
descent-learning algorithm for the said purpose. An ANN may be used as an analytical tool on other data if it is trained to a satisfactory level. The user no longer specifies any training runs to perform this and instead allows the network to work in forward propagation model only new inputs are presented to the input pattern where they filter into and are processed by the middle layer as though training were taking place. But, the output of a forward propagation run is the predicted model and can be used for further analysis and interpretation.

There are several soft computing tools. [3.27 – 3.32] Of them ANN is a computational tool with artificial intelligence in origin. ANNs are also referred to as parallel distributed processing systems (PDP) and connectionists systems. The organizational principles of the central nervous system can be modeled by using ANNs, having the exception that the biologically inspired computing capabilities of the ANN will permit the cognitive and sensory tasks to be performed more easily and more satisfactorily than with conventional serial processors. Human-beings and other living creatures tackle practical problems almost effortlessly in comparison with a serial digital computer. The computational and organizational principles exhibited by neural network of the human brain come from the human capabilities like speech understanding, real time visual perception, sensory information processing and adaptivity as well as intelligent decision making. The same computational and organizational principles can be achieved by the neurobiological studies of the brain for
getting faster and better solution deploying [3.33 – 3.36]. An Artificial Neural Network can have a number of learning rules. But in the present work we have considered the delta rule because most common class of ANNs entitled “Back Propagation Neural Networks (BPNNs)” deploy delta rule for their training schemes. Back Propagation Neural Network is also referred as the backward propagation of error. As ANNs are universal approximators, they work best if the systems involved have high tolerance to errors. Within a set of patterns ANNs work very well for capturing associations or discovering regularities when the volume, number of variables or diversity of the data is very great. The relationships between variables are vaguely understood or, the relationships are difficult to describe adequately with conventional approaches. An ANN is an empirical modeling tool and it does operate by “curve-fitting”. However, some notable differences exist between ANN and more typical empirical models. ANN offers distinct advantages in some areas. ANN has a better filtering capacity than empirical models and it has an advantage over empirical models is the ability to adapt. ANN has specified training algorithms by adjusting connection weights. We can generally expect a network to train quite well depending on the nature of the application and the strength of the internal data patterns. Such scheme is suitable for problems where the relationships may be quite dynamic or nonlinear. ANNs provide an analytical alternative to conventional techniques and are often limited by strict
assumptions of linearity, normality, variable independence etc. An ANN can
capture many kinds of relationships and hence it allows the user to quickly and
relatively easily model phenomena which otherwise may have been impossible
or very difficult to explain.

We require an understanding of a number of network design options for
employing a Back Propagation Neural Network. Those options are number of
input nodes, number of output nodes, number of hidden or middle layers, number
of hidden layers, number of nodes per hidden layer, initial connection weights,
initial node biases, learning rate and momentum rate.

3.3.1 Sequence of steps in Back-propagation

1. Measure the actual output after entering a specific input
2. Compare the actual output to the desired output and calculate a quantitative
   error based on the input-output analysis.
3. Iteratively minimize the error(or the squared error) by adjusting the
   connection strength between nodes. Begin at the output nodes and adjust
   their weight.
4. Propagate “backward” to the layer adjacent to the output layer and calculate
   errors in that layer and adjust weights.
5. Continue this backward propagation(from the output side of the network
to the input side) until all errors are calculated and weights are adjusted.
A back propagation neural networks depicting input layers is shown in Fig 3.1. Hidden layer, output layer and their interactions are depicted in Fig 3.1. The analytical equation for back propagation algorithm is expressed as:

\[
F = I \left( a_1 p_1 + a_2 p_2 + \cdots + a_m p_m \right) \quad (1)
\]

\[
F = I \sum_{i=1}^{m} a_ip_i \quad (2)
\]

Where \( p_1, \ldots, p_m \) are the inputs, \( a_1, \ldots, a_m \) are the corresponding weights. The significance of \( F \) is available from the Fig.3.1.

### 3.3.2 Back propagation neural network training:

We can employ several different training techniques in Back propagation neural network scheme and during training for BPNNs several important parameters are set. Training performance of the network is controlled by the complementary learning rate and momentum setting. Hence, the modification of those parameters can create significant impact on the training performance. It is
general practice to set the learning rate within the range of 0.5 to 0.7 and momentum is set at 0.9 during training of Back propagation neural network. However, there are some researchers who have employed different set of those values and achieved success with efficient training performance. The learning rates of 2 or 5 together with the momentum rate of 0.001 have been used in their training.

The learning rate is modified dynamically in the others schemes, particularly adaptive learning, as the RMS error changes and the training is started with a near zero momentum and set it to near one after other number that passes through the data. However, to achieve a good result it is better to do trial and error to predict the best parameters. The training equation that follows in back propagation network is expressed as:

$$W_{predicted} = W_{old} + \beta \frac{E_{p/p_1}}{p_1}$$  \hspace{1cm} (3)

Where $W_{predicted}$ the new value of weight connecting an input component X to the winning neuron. $W_{old}$ the previous value of this weight. $\beta$ = a training rate co-efficient that may vary during the training process. $E$ = the difference between the desired output and the actual computed output.

### 3.4 RESULTS AND DISCUSSIONS

Our requirement here in the present work is to provide the better-predicted system parameters for SiGeC HBT for low base transit time $\tau_b$. System
parameters such as peak doping concentration in the base ($N_{bm}$) located at $x=R_P$, doping concentration in base at the emitter end ($N_{be}$), doping concentration in the base at collector junction ($N_{bc}$), the value of $x$ in the base beyond which the concentration of alloy-composition Ge (C) remains constant ($X_T$) and percentage content of carbon [$Z_C(X_T)$] in the base at $x=X_T$ were considered for optimization.

In the model of SiGeC HBT we have considered saturation drift velocity of electrons ($\tau^v$) along with base transit time due to diffusion ($\tau^i$) [3.37 – 3.41].

In this work an ANN is employed to predict the various system parameters which if used in the device fabrication, the device will give better high frequency response. System parameters such as peak doping concentration in the base ($N_{bm}$) located at $x=R_P$, doping concentration in base at the emitter end ($N_{be}$), doping concentration in the base at collector junction ($N_{bc}$), the value of $x$ in the base beyond which the concentration of alloy-composition Ge (C) remains constant ($X_T$) and percentage content of carbon [$Z_C(X_T)$] in the base at $x=X_T$ were considered for optimization.

For numerical calculations and optimization SiGeC HBT model has been considered and the effect of emitter-base and collector-base depletion widths and the variation of saturation-drift velocity in base-collector depletion region with molar Ge content were incorporated in the calculations [3.42 – 3.44]. Also triangular material (Ge and C) profiles are considered for their low base transit time feature than box type profiles. First, we have done the optimization of SiGeC
HBT with percentage of Germanium \( Y_C(X_T) \) 8.2 and percentage of Carbon \( Z_C(X_T) \) 1.0.

In the plots we have shown each parameter peak doping concentration in the base \( (N_{bm}) \) is shown in fig.3.2, located at \( x=R_P \), doping concentration in base at the emitter end \( (N_{be}) \) is shown in fig.3.3, doping concentration in the base at collector junction \( (N_{bc}) \) is shown in fig.3.4 base transit time \( \tau_b \) variations with respect to the percentage content of carbon \( Z_C(X_T) \) in the base at \( x=X_T \).

Fig. 3.2 - Base Transit Time Vs. % of Carbon for doping Concentration in the Base \( (N_{mb}) \).
Fig. 3.3 - Base Transit Time Vs. % of Carbon for doping concentration in Base at the emitter junction ($N_{be}$)

Fig. 3.4 - Base Transit Time Vs. % of Carbon for doping concentration in the Base at collector junction ($N_{bc}$)
From the graphs we can observe that \( N_{he} \) is having relatively high effect on base transit time \( \tau_b \) compared to \( N_{bm} \) and \( N_{bc} \). This visualizes our theoretical concepts.

Here our work shows one way of getting optimized base transit time in SiGeC HBT by predicting the faster device parameters for efficient switching operations in the devices. New applications in information technology like wireless applications such as wireless local area network (WLAN), hyper LAN and broadband satellite communications, which are of interest, are asking for still faster circuits. Definitely, this work promises in getting faster devices by applying optimized parameter values in the device fabrication of SiGeC HBT [3.45].
References:


[3.5]. Osten H J and Gaworzewski P. 1997 J. Appl. Phys. 82, 4977-81


[3.9]. F. M. Yeh, C. S. Lin, Electron. Lett. 29 (17), 1540, 1993

[3.10]. F. Como, et al., International Test Conference, pp. 39-471996,


[3.16]. L. D. Lanzarotti et al., Techn. Digest IEDM 96, 249, 1996


1993

1994


[3.40] Iyer ss et al., IEEE Electron Devices 36 2043-64, 1989


“Application of Genetic Algorithm to determine the optimized system
parameters of GaAs quantum wells for better high-frequency
performance under hot electron condition” physica B325 189-194, 2003