MODELING OF THE THRESHOLD VOLTAGE IN STRAINED Si/Si$_{1-x}$Ge$_x$/Si$_{1-y}$Ge$_y$ (x ≥ y) CMOS ARCHITECTURES

Y. L. Tsang, Sanatan Chattopadhyay, Suresh Uppal, Enrique Escobedo-Cousin, Hiran K. Ramakrishnan, Sarah H. Olsen, and Anthony G. O’Neill, Senior Member, IEEE

Abstract—In this paper, an analytical model of threshold voltage $V_T$ for globally strained Si/SiGe CMOS devices using a dual-channel architecture is proposed. Since band parameters modify $V_T$, they are calculated and generalized for different Ge contents in a Si$_{1-x}$Ge$_x$ film grown on relaxed Si$_{1-y}$Ge$_y$ virtual substrates ($x, y < 0.7$). A model for predicting $V_T$ is then developed by considering device geometry and material properties, including band parameters, permittivity, and channel and substrate doping concentrations. $V_T$ lowering due to short-channel effects is included by incorporating a voltage-doping transformation. Accuracy of the model is verified by comparing the model with the results of technology computer-aided design simulations and experiments. The model provides a physical insight for the variation of $V_T$ for both n- and p-MOSFETs in a dual-channel architecture, and it can be generalized to be applicable to single-channel devices as well.

Index Terms—CMOS, dual channel, MOSFETs, strained Si/SiGe, threshold voltage.

I. INTRODUCTION

STRAINED Si/Si$_{1-x}$Ge$_x$ on a relaxed Si$_{1-y}$Ge$_y$ (x > y) virtual substrate (VS) dual-channel CMOS is an attractive approach to incorporate SiGe technology into conventional CMOS technology [1], [2]. In a single-channel CMOS device, the tensile strained Si (ε-Si) layer is directly grown on the Si$_{1-y}$Ge$_y$ VS; whereas in a dual-channel architecture, a compressively strained Si$_{1-x}$Ge$_x$ (ε-SiGe) (x > y) and a tensile ε-Si layers are subsequently grown on a relaxed Si$_{1-y}$Ge$_y$ VS. The schematic cross section of the dual-channel CMOS structure is shown in Fig. 1. Such kind of layer architecture leads to a distinctive modification of the energy band structures. The compressively strained Si$_{1-x}$Ge$_x$ and the tensile strained Si exhibit type I and type II band alignments, respectively, with reference to the energy bands of the relaxed Si$_{1-y}$Ge$_y$ VS [3]. It also results in large valence band offsets ($\Delta E_v$) at both the ε-Si/ε-Si$_{1-x}$Ge$_x$ and ε-Si$_{1-x}$Ge$_x$/Si$_{1-y}$Ge$_y$ interfaces and a large conduction band offset ($\Delta E_c$) at the ε-Si/ε-Si$_{1-x}$Ge$_x$ interface, as shown in the schematic band diagrams in Fig. 2(a) and (b). This kind of band structure can be used in creating high-mobility-surface n- and buried p-channel MOSFETs on a single VS [1], [4]–[6] and is helpful in achieving greater confinement of hole in the buried channel as compared to single-channel devices, which is attributed to the large valence band offset at the ε-Si/ε-Si$_{1-x}$Ge$_x$, as shown in the comparison between Fig. 2(b) and (c). Furthermore, the dual-channel architecture can extend the critical thickness of the strained layer, increase thermal budget for process, and, hence, suppress strain relaxation [5], [6]. Although the advantages of dual-channel devices have been experimentally demonstrated [5], [6], one of the upcoming challenges is the precise control of $V_T$.

A reduction in $V_T$ has been reported in ε-Si single-channel CMOS devices when compared with bulk-Si devices and is attributed to the change in electron affinity, presence of band offsets, and narrowed band gap of ε-Si [7], [8]. This $V_T$ reduction is also reported in the experiment for n-MOSFET in a dual-channel architecture [5]. However, there is no analytical model available in the literature for its explanation. Moreover, there is no such model available that can address the variation of $V_T$ for both the single- and dual-channel architectures with a Ge content in various layers in general. Consequently, it has motivated the authors to develop a new analytical generalized $V_T$ model for n- and p-MOSFETs in ε-Si/ε-SiGe on a SiGe VS dual-channel architecture. The model is also applicable to predict $V_T$ for CMOS in a single-channel architecture.

The paper is organized as follows. Section II details the method of calculation of the band parameters. In Section III, an analytical $V_T$ model is developed for both n- and p-MOSFETs using a dual-channel architecture. The short-channel effects are incorporated using voltage-doping transformation (VDT) [9], and a condition has been reached for a specific range of source
and drain voltage difference \( (V_D) \), where the model is valid. In Section IV, the validity of the model is verified by comparing results obtained from the developed model with those from technology computer-aided design (TCAD) simulations and experiments. Section V summarizes the overall findings.

II. CALCULATION OF BAND PARAMETERS

When a thin \( \text{Si}_{1-x}\text{Ge}_x \) film is epitaxially grown on a relaxed \( \text{Si}_{1-y}\text{Ge}_y \) VS \( (x \neq y) \), the grown layer conforms to the lattice spacing of the underlying \( \text{Si}_{1-y}\text{Ge}_y \) and is strained accordingly. The strain modifies the band parameters such as band gap \( (E_g) \) in the \( \varepsilon\text{-Si}_{1-x}\text{Ge}_x \) film and the conduction \( (\Delta E_c) \) and valence \( (\Delta E_v) \) band offsets at the heterointerface. In this section, these band parameters for this SiGe heterojunction are calculated and generalized for a range of \( x \) and \( y (x, y < 0.7) \) by following People’s model [10]–[12]. At first, strain tensors are obtained for the strained layers from the lattice mismatch by taking into account Ge fractions in the strained layer and in the relaxed VS. These strain tensors are then substituted into the equations for calculating the components that lead to band-gap narrowing [10]. The narrowed band gap in the \( \varepsilon\text{-Si}_{1-x}\text{Ge}_x \) layer can therefore be obtained by subtracting these calculated components from the relaxed \( \text{Si}_{1-x}\text{Ge}_x \) band gap for the same Ge fraction. \( \Delta E_v \) are obtained by following the general function in People’s model based on the work of Van de Walle [12]. \( \Delta E_c \) can be obtained after the \( E_g \) of the strained layers is aligned with the \( E_g \) of the relaxed VS in a band diagram. The calculated \( E_g, \Delta E_v, \) and \( \Delta E_c \) for various \( x \) and \( y \) values are shown in Fig. 3(a)–(c), respectively. The extracted band parameters in this paper have been verified with the available results in [11].

III. DERIVATION OF THE MODEL

In the structure shown in Fig. 1, \( x_z, x_x, \) and \( x_{ox} \) represent the thickness (in nanometers) of the \( \varepsilon\text{-Si} \) surface channel, \( \varepsilon\text{-Si}_{1-y}\text{Ge}_y \) buried channel, and oxide layer, respectively. As the strained layers of the surface and buried channels are limited by their critical thickness and are normally thin, carrier depletion is assumed to be deeper than the channel regions and reaches the \( \text{Si}_{1-y}\text{Ge}_y \) VS. The thickness of a buffer layer is \( x_{\text{buff}} \) (in nanometers). This buffer layer is assumed to be undoped, and the intrinsic doping concentrations are calculated to be \( 3.8 \times 10^{10} \text{ cm}^{-3} \) for \( y = 0.1 \) and \( 1.5 \times 10^{11} \text{ cm}^{-3} \) for \( y = 0.3 \) and can be removed from the structure by setting \( x_{\text{buff}} = 0 \) or by assuming its doping concentration to be the same as that of the VS in the model equations. The depletion depth in the substrate after the buffer layer is \( x_D \) (in nanometers). Uniform channel and substrate doping concentrations are assumed to be \( N_{\text{ch}} \) and \( N_b \) (in parts per cubic centimeter), respectively.
respectively. The permittivities $\varepsilon$ in the layers are assumed to linearly vary with Ge fractions.

$V_T$ is defined as the applied gate voltage at which strong inversion occurs in the channel. For conventional bulk-Si MOSFETs, $V_T$ in the absence (zero) of channel (or source/drain) voltage can be expressed as

$$V_T = \phi_s + V_i + V_{FB}$$

(1)

where $\phi_s$ is threshold potential in the semiconductor, $V_i$ is depletion charge voltage, and $V_{FB}$ is the flatband voltage [13].

A. $V_T$ Modeling for n-MOSFETs in a Dual-Channel Architecture

In a dual-channel architecture, the conduction band of the $\varepsilon$-Si layer is always the lowest conduction band, as shown in Fig. 2(a). Therefore, the first strong inversion layer is always created in the surface channel, and hence, $V_T$ for buried channel is irrelevant. $V_T$ for the surface n-channel using a dual-channel architecture ($V_{TN}$) is given by

$$V_{TN} = \phi_{SN} + V_{iN} + V_{FB}$$

(2)

where $\phi_{SN}$ and $V_{iN}$ are $\phi_s$ and $V_i$ for surface n-channel devices in a dual-channel architecture, respectively.

1) Calculation of $\phi_{SN}$: $\phi_{SN}$ is defined as the surface potential required to create strong inversion at the oxide/$\varepsilon$-Si interface. As described in [14], the threshold potential for $\varepsilon$-Si can be approximated by taking the average between the threshold potentials of the respective channels and the relaxed Si$_1$-$y$Ge$_y$ VS. It should be noted that the position of the energy bands of $\varepsilon$-Si is only decided by the amount of strain in the $\varepsilon$-Si channel layer and, hence, by the Ge content ($y$) only. Thus, $\phi_{SN}$ is expressed as

$$\phi_{SN} = \frac{kT}{q} \left[ \ln \frac{N_{ch}}{n_{i,z}} + \ln \frac{N_b}{n_{i,y}} - \frac{1}{2} \left( \frac{\Delta E_{c1} + \Delta E_{o1}}{2} \right) \right]$$

(3)

where $k$ is Boltzmann’s constant, $q$ is the electron charge, $T$ is the temperature in the absolute scale, and $\Delta E_{c1}$ and $\Delta E_{o1}$ denote the conduction and valence band offsets between the $\varepsilon$-Si channel and the relaxed Si$_1$-$y$Ge$_y$ VS, respectively. The notations of $z$ and $y$ next to the intrinsic carrier concentration $n_i$ denote the relevant materials, i.e., $\varepsilon$-Si and Si$_1$-$y$Ge$_y$, respectively. $n_i$ for bulk Si is expressed as

$$n_i = \sqrt{N_c N_v} \exp(-qE_g/2kT)$$

(4)

where $N_c$ and $N_v$ are the density of states for the bulk Si at conduction and valence bands, respectively. In this paper, $N_c$ and $N_v$ are assumed to linearly vary with the Ge fraction for the relaxed SiGe. However, these values for $\varepsilon$-Si are assumed to be one third of the bulk Si, whereas for $\varepsilon$-Si$_1$-$x$Ge$_x$, these values are assumed to be two thirds and one sixth of the bulk Si, respectively [15]. The $E_g$ values for the materials are obtained from Fig. 3(a); hence, the dependence of $n_i$ on the Ge fraction (as well as the strain) is realized.

2) Calculation of $V_{iN}$: In n-MOSFETs, if a positive gate bias is applied, then it induces the negative uncompensated ionized charges in the depletion region. Such charges contribute a depletion charge voltage $V_{iN}$, which can be solved by using a 1-D Poisson’s equation [13].

At the onset of strong inversion, the potential at the SiO$_2$/$\varepsilon$-Si interface $\phi_{sur}$ is equal to $\phi_{SN}$ from (3). Using depletion approximation, the maximum value of $x_D$, i.e., the maximum depletion depth in the substrate $x_{D,max}$, can be derived as follows:

$$x_{D,max} = \sqrt{\frac{2\varepsilon_y}{qN_b} K + (B)^2 - B}$$

(5)

where $K = \phi_{SN} - (qN_{ch}x_D^2/2\varepsilon_y + qN_{ch}x_Dx_Z/\varepsilon_y + qN_{ch}x_D^2/2\varepsilon_x)$, and $B = \varepsilon_y(x_D/x_Z + x_D/x_Z + x_{nuf}/x_y)$.

$V_{iN}$ can be calculated by estimating the total number of ionized charges in the depletion region under a MOS capacitor and is given by

$$V_{iN} = \frac{q}{C_{ox}} (N_{ch}(x_D + x_Z) + 0.5x_{D,max})$$

(6)

where $C_{ox}$ is the capacitance of the oxide layer.

3) Calculation of $V_{FB}$: For bulk-Si devices, $V_{FB} = \phi_{ms} + Q_{ox}/C_{ox}$ [13], where $Q_{ox}$ is a modeling parameter that combines the effects from total charges in the oxide layer (e.g., interface trap density and mobile charge), and $\phi_{ms}$ is the work function difference, which is given by $\phi_{ms} = \phi_m - \phi_{Si}$, where $\phi_m$ and $\phi_{Si}$ are the work functions of the gate material and semiconductor, respectively. In a dual-channel architecture, $\phi_m$ is unchanged, but $\phi_{Si}$ must be modified for a strain-modified band structure, as shown in Fig. 4, where $\chi_{e-Si}$ and $\chi_{Si_1-yGe_y}$ denote the electron affinity of $\varepsilon$-Si and SiGe VS, respectively, and $\phi_{F,e-Si}$ is the potential difference between the Fermi level ($E_F$) and the conduction band in the $\varepsilon$-Si layer ($E_e$). In Fig. 4, it is observed that $\phi_{Si}$ is given by $\phi_{Si} = \chi_{e-Si} + \phi_{F,e-Si}$. However, for modeling $\phi_{Si}$, the average of $\chi_{Si_1-yGe_y}$ and $\chi_{e-Si}$ (instead of $\chi_{e-Si}$) should be assumed since, after Fermi level alignment, the position of $E_e$ of the $\varepsilon$-Si layer depends on $\Delta E_{c1}$ as well as the alignment of the Fermi level. Consequently, $\phi_{Si}$ can be expressed as

$$\phi_{Si} = \left( \chi_{Si_1-yGe_y} - \frac{\Delta E_{c1}}{2} \right) + \left( \frac{E_{g,s} + kT}{2} - \frac{kT}{q} \ln \frac{N_{ch}}{n_{i,z}} \right)$$

(7)
where $\Delta E_{c1}$ should be used with a proper sign, and $E_{g,z}$ is the band gap of the $\varepsilon$-Si layer. The “+” and “−” signs should be chosen for n- and p-MOSFETs, respectively, from the “±” symbol in (7).

The dipole potential ($\phi_{\text{dipole}}$), as described in [14], is also considered for n-MOSFET in a dual-channel architecture. It originates due to the fact that the holes will be accumulated at the $\varepsilon$-Si/Ge interface under flatband conditions due to the presence of the large valence band offset and can be expressed as

$$\phi_{\text{dipole}} = \frac{qN_{\text{ch}}x^2}{2\varepsilon_x} + \frac{qN_{\text{ch}}x_zL_d\sqrt{2}}{2\varepsilon_x},$$

where $L_d$ is the Debye length, and $L_d = \frac{\varepsilon_x}{qkT/\varepsilon_y}$.

A complete expression for $V_{FB}$ can now be written as

$$V_{FB} = \phi_m - \left\{ \frac{E_{g,z}}{2} + \frac{kT}{q} \ln \frac{N_{\text{ch}}}{n_{i,z}} \right\} - \frac{Q_{\text{ox}}}{C_{\text{ox}}} - \phi_{\text{dipole}}.$$

It is noted that $\phi_{\text{dipole}}$ should be excluded from (9) when $V_T$ is calculated for bulk-Si devices. A complete expression for $V_{TN}$ can be obtained by substituting (3), (6), and (9) into (2).

**B. $V_T$ Modeling for p-MOSFETs in a Dual-Channel Architecture**

Due to the large valence band offset at the $\varepsilon$-Si/$\varepsilon$-Si$_{1-x}$Ge$_x$ interface in a dual-channel architecture, transport of holes can take place either along the buried or surface channel or along both channels, as shown in Fig. 2(b). These two channels are switched on at different gate biases, depending on the device design. Therefore, two threshold voltages in the absence of any channel bias are defined as follows:

For the buried channel, $V_{TB} = \phi_{TB} + V_{i,TB} + V_{FB}$

For the surface channel, $V_{TS} = \phi_{TS} + V_{i,TS} + V_{FB}$

where $\phi_{TB}$ and $\phi_{TS}$ are threshold potentials ($\phi_S$), and $V_{i,TB}$ and $V_{i,TS}$ are depletion charge voltages ($V_i$) for buried and surface p-channel devices, respectively.

1) Calculation of $\phi_{TB}$ and $\phi_{TS}$: The definition of $\phi_{TS}$ is similar to $\phi_{TN}$ in Section III-A1. $\phi_{TB}$ is defined as the potential required to create strong inversion in the $\varepsilon$-Si$_{1-x}$Ge$_x$ layer and can be modeled by averaging the threshold potentials of $\varepsilon$-Si$_{1-x}$Ge$_x$ channel and that of the relaxed Si$_{1-y}$Ge$_y$ VS. Thus, $\phi_{TB}$ and $\phi_{TS}$ are expressed as

$$\phi_{TB} = -\frac{kT}{q} \ln \frac{N_{\text{ch}}}{n_{i,x}} + \ln \frac{N_b}{n_{i,y}} + \frac{\Delta E_{c2} + \Delta E_{v2}}{2}$$

$$\phi_{TS} = -\frac{kT}{q} \ln \frac{N_{\text{ch}}}{n_{i,x}} + \ln \frac{N_b}{n_{i,y}} + \frac{\Delta E_{c1} + \Delta E_{v1}}{2}$$

where $\Delta E_{c2}$ and $\Delta E_{v2}$ denote the conduction and valence band offsets between the $\varepsilon$-Si$_{1-x}$Ge$_x$ buried channel and the relaxed Si$_{1-y}$Ge$_y$ VS, respectively.

2) Calculation of $V_{i,TB}$ and $V_{i,TS}$: When the potential at the $\varepsilon$-Si/$\varepsilon$-Si$_{1-x}$Ge$_x$ interface is more negative than $\phi_{TB}$, the strong inversion is created in the buried channel, and holes reside there. These holes, as well as the depletion charges, are both considered in the model.

The potentials at specific interfaces are defined as follows:

$\phi_{\text{sur}}$ at the SiO$_2$/$\varepsilon$-Si interface; $\phi_H$ at the $\varepsilon$-Si/$\varepsilon$-Si$_{1-x}$Ge$_x$ interface; and $\phi(x = x_z + x_x)$ at $\varepsilon$-Si$_{1-x}$Ge$_x$/Si$_{1-y}$Ge$_y$ interface. $\phi(x = x_z + x_x)$ is obtained by using depletion approximation and is expressed as

$$\phi(x = x_z + x_x) = \left(\frac{qN_b x_{\text{bulk}} x_D}{\varepsilon_y} + \frac{qN_b x_D^2}{2\varepsilon_y}\right).$$

The potential difference across the $\varepsilon$-Si channel, i.e., $\phi_{\text{sur}} - \phi_H$, is obtained by solving Poisson’s equation as follows, taking into account the ionized charges and the electric field in the $\varepsilon$-Si layer:

$$\phi_{\text{sur}} - \phi_H = \frac{x_z \varepsilon_E H}{\varepsilon_x} + \frac{qN_{\text{ch}} x^2}{2\varepsilon_x}$$

where $E_H$ is the electric field at the $\varepsilon$-Si/$\varepsilon$-Si$_{1-x}$Ge$_x$ interface. $E_H$ is expressed as

$$E_H = \frac{qN_b x_D}{\varepsilon_x} \sqrt{1 + H(\phi_H) + N_d(\phi_H)}$$

where $H(\phi_H)$ and $N_d(\phi_H)$ are contributions of holes and ionized donors to the electric field, respectively, and are expressed as

$$H(\phi_H) = \frac{2\varepsilon_x N_b kT}{(qN_b x_D)^2} \left[ \exp\left(\frac{\phi_{TB} - \phi_H}{kT/q}\right) - \exp\left(\frac{\phi(x = x_z + x_x) - \phi_H}{kT/q}\right) \right]$$

and

$$N_d(\phi_H) = \frac{2qN_{\text{ch}} x_z}{(qN_b x_D)} \left[ \phi(x = x_z + x_x) - \phi_H \right].$$

Derivation of $E_H$ is detailed in the Appendix.

At the onset of strong inversion at the buried channel, $\phi_H = \phi_{TB}$. By using the depletion approximation, the maximum depletion depth $x_{D,\text{max}}$ in the substrate can be derived as

$$x_{D,\text{max}} = \sqrt{\frac{2\varepsilon_y}{qN_{\text{ch}}} \left(\frac{R_x N_{\text{ch}} x^2}{N_b} + (x_{\text{bulk}} + R_x x)^2\right) - x_{\text{bulk}} - R_x x}$$

where $R_x = \varepsilon_y/\varepsilon_x$. The maximum depletion depth $x_{D,\text{max}}$ is used to replace $x_D$ in all the equations that contain $x_D$ in the $V_{TB}$ and $V_{TS}$ models.

After solving Poisson’s equations, $V_{i,TB}$ is obtained as

$$V_{i,TB} = -\left[ \frac{x_{\text{ox}}}{2\varepsilon_{\text{ox}}} + \frac{x}{\varepsilon_x} \right] qN_{\text{ch}} x_z + \left[ \frac{x_{\text{ox}}}{\varepsilon_{\text{ox}}} + \frac{x}{\varepsilon_x} \right] \varepsilon_x E_H.$$
In (20), the first and second terms represent the voltage contributed by the ionized charges in the \(\varepsilon\)-Si layer and the charges underneath the \(\varepsilon\)-Si/\(\varepsilon\)-Si\(_{1-x}\)Ge\(_{x}\) interface, respectively. To solve \(V_{\text{TS}}\) from (20), \(E_H\) is obtained by setting \(\phi_H = \phi_{\text{TB}}\) in (16). Similarly, a full expression of \(V_{\text{TS}}\) is expressed as

\[
V_{\text{TS}} = -\left(\frac{q\varepsilon_{\text{ox}}}{\varepsilon_{\text{ox}}^*}\right) \left(qN_{\text{ch}}x_z + \varepsilon_x E_H\right).
\]

In (21), \(V_{\text{TS}}\) represents the voltage contributed by the charges underneath the oxide layer. At the onset of strong inversion that occurs at the SiO\(_2\)/\(\varepsilon\)-Si interface, \(\phi_{\text{Sur}} = \phi_{\text{TS}}\) and \(\phi_H \neq \phi_{\text{TB}}\) in (15). \(V_{\text{TS}}\) is solved as follows: 1) the expression for \(E_H\) in (16) is substituted into (15); 2) \(\phi_H\) is now the only unknown and is solved by iterations with a first guess of \(\phi_H = \phi_{\text{TB}}\); and 3) \(E_H\) is obtained after substituting the calculated \(\phi_H\) back into (16). \(V_{\text{TS}}\) in (21) can be solved with this calculated \(E_H\).

3) Calculation of \(V_{\text{FB}}\): Steps for calculation of \(V_{\text{FB}}\) are similar to those in Section III-A3. The differences are that the effect of the dipole potential is calculated to be insignificant in dual-channel p-MOSFETs, and the “+” sign is chosen from the “+” symbol in (7). Thus, \(V_{\text{FB}}\) is expressed as

\[
V_{\text{FB}} = \phi_m - \left(\chi_{\text{Si}_{1-y}\text{Ge}_y} - \frac{\Delta E_{z1}}{2}\right)
+ \left(\frac{E_{g0}}{2} - \frac{kT}{q} \ln \frac{N_{\text{ch}}}{n_{\text{sat}}^*}\right) - \frac{q\varepsilon_{\text{ox}}}{\varepsilon_{\text{ox}}^*}.
\]

The complete expression for \(V_{\text{FB}}\) is obtained by substituting (12), (20), and (22) into (10), and that for \(V_{\text{TS}}\) is obtained by substituting (13), (21), and (22) into (11). It should be pointed out that a measurable \(V_T\) in the experiment is the minimum between \(|V_{\text{FB}}|\) and \(|V_{\text{TS}}|\).

C. Calculation of the Channel-Length-Dependent \(V_T\)

To take into account the short-channel effect in the present model, VDT is incorporated [9] since it is simple and considerably accurate. VDT suggests that the potential barrier height lowering due to the drain–source field can be modeled by assuming a reduction in the net channel doping concentration.

To incorporate VDT in the present model, the first step is to calculate the effective doping concentration \(N_b^*\) in terms of the source–drain voltage \(V_D\), channel and substrate doping concentrations \(N_{\text{ch}}\) and \(N_b\) (considering the effect of advanced doping profile such as retrograde doping), and channel length \(L_{\text{eff}}\). Then, replace \(N_b\) by \(N_b^*\) into the depletion charge voltage equations; i.e., (5), (6), and (14)–(21) for the model of n- and p-MOSFETs, respectively.

Based upon the simplified VDT used in [16], \(N_b^*\) for n-MOSFETs using a dual-channel architecture by taking into account the channel and substrate doping is modified as

\[
N_b^* = (N_b + N_{\text{ch}}) - \frac{2\varepsilon_{\text{ox}}V_D^*}{qL_{\text{eff}}^2},
\]

where \(V_D^*\) is the effective drain voltage, and the “−” sign on the right-hand side represents the product of positive \(V_D^*\) and negative \(q\) for n-MOSFETs. It is worth specifying that the net contribution from \(N_b + N_{\text{ch}}\) in (23) depends on the relative values of the channel and substrate doping concentrations

\[
V_D = V_D^* + 2(V_{\text{bi}} - \phi_{\text{SN}})\frac{(V_{\text{bi}} + V_D - \phi_{\text{SN}})}{2\sqrt{(V_{\text{bi}} - \phi_{\text{SN}})(V_{\text{bi}} + V_D - \phi_{\text{SN}})}}
\]

where \(V_{\text{bi}}\) is the built-in potential of the drain–substrate p–n junction. The second term in (23) indicates that the \(N_b^*\) decreases with decreasing \(L_{\text{eff}}\) and increasing \(V_D^*\). As a result, \(V_T\) decreases.

To express \(N_b^*\) in p-MOSFETs, the negative and positive signs are inverted and \(\phi_{\text{SN}}\) is replaced by \(\phi_{\text{TB}}\) in (24). It is noted that VDT becomes invalid whenever \(N_b^*\) is assumed to have a negative value, indicating that all impurity ions are tied by the drain field and MOSFET reaches the punch-through mode. The term \(V_T\) becomes insignificant under such a condition [16]. It is important to note that VDT depends on \(N_b\) and \(N_{\text{ch}}\) as well as on \(V_D\). Therefore, it can be used in estimating the effect of \(V_D\) on \(V_T\) for a set of given doping values. However, the variation of \(V_T\) with \(V_D\) has been reported in [9], and we have observed a similar variation. In the present model, \(V_D\) is assumed to be 0.1 V, following the values of experimental parameters. Therefore, this model is valid up to a \(V_D\) for which \(N_b^*\) is positive for given values of \(L_{\text{eff}}\), \(N_b\), and \(N_{\text{ch}}\).

IV. RESULTS AND DISCUSSION

In Fig. 5(a), the results extracted from the present analytical model for CMOS devices in a dual-channel architecture are compared with the results obtained from TCAD simulation using Medici [17] for various Ge fractions in the \(\varepsilon\)-Si\(_{1-x}\)Ge\(_x\) relaxed Si\(_{0.85}\)Ge\(_{0.15}\) VS architecture. (b) Schematic energy band moment for increasing Ge fraction \(x\) from 0 to 0.5.
The experimental $V_T$ are defined at the drain current $I_D = 70$ nA/µm in $I_D-V_G$ (gate voltage) characteristics at $V_D = 0.1$ V.

As seen in Fig. 6, the present model shows an excellent agreement to the experiment in the short-channel, as well as the long-channel, regime. Two phases of $V_T$ shifts are observed: 1) from a control-Si to a single-channel n-MOSFET and 2) from a single-channel to a dual-channel n-MOSFET. The present model suggests two different mechanisms for the $V_T$ shifts. The first phase of the $V_T$ shift has been well understood, as reported in the literature [7], [8], [14], and is explained by the presence of the conduction band offset between the $\varepsilon$-Si and Si$_{1-x}$Ge$_y$ VSSs, which brings the conduction band of the surface n-channel closer to the Fermi level and reduces the threshold potential $\phi_{SN}$. Another source that causes this reduction is the change in oxide charges for single-channel devices [18]. To explain the second phase of the $V_T$ shift, initially, it is referred to Fig. 5(a), where $V_{TN}$ is almost independent of the Ge fraction $(x)$. Therefore, theoretically, $V_T$ for dual- and single-channel devices should be the same, which is not consistent with the experiment, as shown in Fig. 6. To explain this apparent inconsistency, two effects are considered: 1) the effect of setting the buried channel thickness ($x_z$) to 12 nm in (6) and 2) the effect of the total oxide charge.

In going from single- to dual-channel structures, a thin layer ($x_z = 12$ nm in this case) with doping lower than the substrate doping is introduced between the surface channel and VS. This additional layer increases $V_{TN}$ by the term $N_{ch}(x_z + x_x)$ in (6). In contrast, this layer provides a higher screening effect due to additional depleted charges in this layer, resulting in the reduction of $x_{D,max}$ in (6). Since $N_b > N_{ch}$, $V_{TN}$ is reduced by this $x_{D, max}$ reduction. Hence, $V_T$ is reduced for dual-channel as compared to single-channel structures.

Additionally, our work in [18] has shown that the fixed oxide charges and interface state density increase with an increasing Ge fraction in the layer underneath the $\varepsilon$-Si layer. Therefore, the variation of $Q_{ox}$ with the Ge fraction $(x)$ is necessarily included when explaining the experimental results. The model uses the measured values $(3.2 \times 10^{11} \text{ cm}^{-2})$ of $Q_{ox}$ for calculating $V_T$ for dual-channel devices, showing a good agreement with the experiment in Fig. 6. However, the major contribution $(\approx 70\%)$ to the reduction of $V_T$ from a single- to a dual-channel architecture comes from the reduction in the depletion charge voltage rather than from the increase in oxide charges. It is worth specifying that the experimental data are not smooth enough compared with the calculated values, which is attributed to the inherent nonuniformity in the device and wafer processing.

Fig. 7(a) compares the $V_T$ values calculated from the present model with experimental data for various Ge fractions in the Si$_{1-y}$Ge$_y$ VS $(0 \leq y \leq 0.3)$ for n-MOSFETs in a single-channel architecture. The model shows an excellent agreement even with the data for a single-channel device. The decreasing trend in Fig. 7(a) can be explained by the fact that increasing the Ge fraction $(y)$ leads to an increase in the conduction band offset between the $\varepsilon$-Si and Si$_{1-y}$Ge$_y$ VSSs, which brings the conduction band of $\varepsilon$-Si closer to the Fermi level and results
tively.
in the reduction of the threshold potential $\phi_{SN}$, as described in (3). It also narrows the band gap in the $\varepsilon$-Si channel, which results in an increase in $n_{i,2}$ and, therefore, decreases the logarithmic term in (3) for $\phi_{SN}$. Hence, the reduced $\phi_{SN}$ allows strong inversion to occur at a lower gate bias and results in a lower $V_T$. Moreover, the model can estimate an increase in $V_T$ by increasing the doping concentration $N_b$, as shown in Fig. 7(b).

In Fig. 8(a), the analytical model is compared with the experiment from [19] for various channel lengths ($L_{eff}$) from 10 to 0.1 $\mu$m, as well as for Si cap thicknesses ($x_z$) from 8 to 1 nm, in p-MOSFETs using Si/$\varepsilon$-Si$_{0.7}$Ge$_{0.3}$ on the bulk-Si architecture. Good agreement between the analytical model and experimental results is obtained. The figure shows the classic $V_T$ roll-off, which is explained by the lowering of the energy barrier with decreasing $L_{eff}$ and is modeled in VDT in (23). Fig. 8(b) shows the variations of $V_{TB}$ and $V_{TS}$ with $x_z$, indicating the increase of $|V_T|$. This increase in $|V_T|$ is due to the fact that the quantum well located in the buried channel [Fig. 2(b)] is shifted away from the oxide layer by increasing $x_z$. For a thicker Si cap, a more negative gate bias is needed to create strong inversion in the buried channel. However, $V_{TS}$ decreases with increasing $x_z$, which results from the significant reduction in $V_{i,TS}$ (21).

The present model is used to predict the variations of $V_T$ for different Ge fractions in the VS ($y$) for the n- and p-MOSFETs with the Si$_{0.5}$Ge$_{0.5}$ buried channel in Fig. 9(a). The same device parameters as for Fig. 5 are used here. From the n-MOSFET, it is observed that $V_{TN}$ decreases with increasing Ge ($y$), and it is consistent with the results shown in Fig. 7(a). In both cases, this $V_{TN}$ lowering is attributed to the increase of the conduction band offset between the $\varepsilon$-Si layer and VS with the Ge fraction ($y$), as shown in Fig. 9(b).

From p-MOSFETs, it is observed that $|V_{TB}|$ is always lower than $|V_{TS}|$, and $|V_{TB}|$ increases with an increase in the Ge fraction ($y$). This increase in $|V_{TB}|$ is attributed to the fact that increasing the Ge fraction ($y$) (from 0 to 0.5) reduces the lattice mismatch between the Si$_{0.5}$Ge$_{0.5}$ buried channel layer and the Si$_1-y$Ge$_y$ VS and, hence, reduces the amount of strain in the buried channel, leading to an increase in the band gap of the buried channel and a decrease in the valence band offset between the buried channel and VS, as shown in Fig. 9(b). Consequently, a more negative gate bias is required to drive the valence band closer to the Fermi level to achieve strong inversion. Another interesting result from the figure is that $|V_{TS}|$ decreases with an increase in the Ge fraction ($y$). It can be explained by the fact that the increase in $y$ reduces the valence band offset between $\varepsilon$-Si and $\varepsilon$-Si$_{1-y}$Ge$_y$, and, hence, reduces hole confinement in the buried channel. The contribution of holes to $V_{i,TS}$ is therefore reduced, as described in (21).

V. SUMMARY

The band parameters have been generalized for different Ge contents in the Si$_{1-y}$Ge$_y$ layer on the Si$_1-y$Ge$_y$ VS ($x, y < 0.7$). An analytical model has been developed and verified with simulations and experimental data for n- and p-MOSFETs in both single- and dual-channel architectures. This model can predict with detailed physical explanations to the variation of $V_T$ with design parameters, in particular, Ge fractions, layer thicknesses, channel lengths, and doping profiles. The present model will allow engineers to predict and optimize $V_T$ for CMOS devices in both single- and dual-channel architectures in its complex design space within a range of source–drain voltage for which the effective doping concentration remains positive for a given set of values for channel length and channel- and substrate-doping concentrations.
APPENDIX
DERIVATION OF THE ELECTRIC FIELD ($E_H$) AT THE $\varepsilon$-Si/$\varepsilon$-Si$_{1-x}$Ge$_x$ INTERFACE

Poisson’s equation is used to solve the electric field across the buried channel. By using depletion approximation, the charges at the onset of strong inversion in the buried p-channel are contributed by holes ($p$) and positive ionized donors ($N_{ch}^+$). Therefore

$$\rho_{\text{charge}}(x) = q \left( p + N_{ch}^+ \right) \quad (A.1)$$

where $\rho_{\text{charge}}$ is the charge density. Since the number of holes $p$ in the buried channel in (A.1) depends on the potential difference $\phi$, the Poisson’s equation to be solved becomes

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{q}{\varepsilon_x} \left( p + N_{ch}^+ \right) = -\frac{q}{\varepsilon_x} \left[ N_h \exp \left( \frac{\phi q}{kT} \right) + N_{ch} \right]. \quad (A.2)$$

Integrating (A.2) with the electric field and potential differences $\phi$ between the $\varepsilon$-Si/$\varepsilon$-Si$_{1-x}$Ge$_x$ and $\varepsilon$-Si$_{1-x}$Ge$_x$/Si$_{1-y}$Ge$_y$ interfaces, we obtain

$$\int_{\phi_1}^{\phi_2} d\phi \left( \frac{\partial \phi}{\partial x} \right) = \frac{q}{\varepsilon_x} \int_{\phi_1}^{\phi_2} \left[ N_h \exp \left( \frac{\phi q}{kT} \right) + N_{ch} \right] d\phi. \quad (A.3)$$

As the electric field $\xi = -\partial \phi / \partial x$, (A.3) is derived to be

$$\xi_1^2 - \xi_2^2 = \frac{2q}{\varepsilon_x} \left\{ \frac{N_h kT}{q} \left\{ \exp \left( \frac{\phi_1 q}{kT} \right) - \exp \left( \frac{\phi_2 q}{kT} \right) \right\} \right\}$$

$$+ \frac{2qN_{ch}}{\varepsilon_x} (\phi_1 - \phi_2). \quad (A.4)$$

The left-hand side in (A.4) is the electric field across the buried channel. Boundary conditions are assumed for the electric fields. These are $\xi_1 = \xi(x = x_1) = E_H$ and $\xi_2 = \xi(x = x_2 + x_x)$, where $\xi_2$ can be solved using depletion approximation as follows:

$$\xi_2(x = x_x + x_x) = \frac{qN_{LD}}{\varepsilon_x}. \quad (A.5)$$

Since holes ($p$) are assumed to reside in the buried channel only when $|\phi_H| > |\phi_{TB}|$, the boundary conditions for the potential $\phi$ for the right-hand side of (A.4) are

$$\phi_1 = \phi_{TB} - \phi_H \quad \phi_2 = \phi_{TB} - \phi(x = dx_x + x_x). \quad (A.6)$$

By substituting (A.5) and (A.6) into (A.4), (16) in Section III-B1 is finally obtained.

REFERENCES


Sanatan Chattopadhyay received the B.Sc. (Hons.) degree in physics and the M.Sc. degree in electronic science from the University of Calcutta, Calcutta, India, in 1992 and 1994, respectively, and the Ph.D. degree from the University of Jadavpur, Calcutta, in 1999. From 1999 to 2001, he was a Post-Doctoral Fellow at the Singapore–Massachusetts Institute of Technology Alliance, Singapore, before joining Newcastle University, Newcastle-upon-Tyne, U.K. In 2007, he joined the Department of Electronic Science, University of Calcutta. His current research interests include the fabrication and characterization of strained Si/SiGe MOSFETs, low-resistance silicides/germanides for contact metallization in very large scale integration, the growth and characterization of ultrathin dielectric gate materials, and the design of low-power high-performance circuits using strained Si CMOS.

Suresh Uppal received the M.Sc.(Hons.) degree from Panjab University, Chandigarh, India in 1997, the M.Tech. degree from the Indian Institute of Technology Delhi (IITD), Delhi, India, in 1999, and the Ph.D. degree from the University of Southampton, Southampton, U.K., in 2003. His Ph.D. dissertation is on the diffusion of B and Si in bulk Ge. During 2003–2005, he was with the University of Southampton, U.K., working on dopant diffusion and self-diffusion in Si and SiGe alloys and modeling of a solid-state infrared modulator using high-purity Ge. He joined Newcastle University, Newcastle-upon-Tyne, U.K., in 2005, where he is currently with the School of Electrical, Electronic and Computer Engineering, involved in the design, fabrication, modeling, and characterization of strained Si/SiGe-based devices. His research interests are dopant diffusion and activation in new channel materials, materials and device engineering for CMOS/BiCMOS and optoelectronics applications, process and device (TCAD) simulation, and reliability of gate dielectrics.

Enrique Escobedo-Cousin received the B.Sc. degree in electrical and electronic engineering from the Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico, in 2001 and the M.Sc. degree in microelectronics from Newcastle University, Newcastle-upon-Tyne, U.K., in 2004, where he is currently working toward the Ph.D. degree in semiconductor technology in the School of Electrical, Electronic and Computer Engineering. In 2001, he joined the Institute of Astronomy, UNAM, where he worked on image processing for infrared photodetectors until 2003. His research involves developing novel characterization methods for strained SiGe devices, including chemical etching, atomic force microscopy, and the development of digital analysis techniques. He is the recipient of a research scholarship from the National Council of Science and Technology (CONACYT), Mexico City.

Hiran K. Ramakrishnan received the B.Tech. degree in electronics from Cochin University of Science and Technology, Kerala, India.

He has been working in the microelectronics design industry since 2000, where he was engaged in the development of cell libraries for EDA tools. His current research interests include low-power high-speed circuits and variability in IC manufacturing.

Sarah H. Olsen received the B.Sc. degree in physics from the University of Bath, Bath, U.K., in 1995 and the Ph.D. degree from Newcastle University, Newcastle-upon-Tyne, U.K., in 2002.

Between 1995 and 1998, she was a Product Engineer with GEC Plessey Semiconductors, Plymouth, U.K., and Siemens Microelectronics, North Tyneside, U.K. In 1998, she joined Newcastle University, where she is currently with the School of Electrical, Electronic and Computer Engineering. Her research interests are in strained Si/Ge materials, devices, and technology.

Anthony G. O’Neill (SM’00) was born in Leicester, U.K., in 1959. He received the B.Sc. degree from the University of Nottingham, Nottingham, U.K., in 1980 and the Ph.D. degree from the University of St. Andrews, St. Andrews, U.K., in 1983.

Between 1983 and 1986, he was with Plessey Research (Caswell) Ltd., Towcester, U.K., before taking up his post at Newcastle University, Newcastle-upon-Tyne, U.K., where he is currently with the School of Electrical, Electronic and Computer Engineering. In 1994, he was a Visiting Scientist at the Massachusetts Institute of Technology. From 2002 to 2005, he was a Royal Society Industry Fellow at Atmel, San Jose, CA. He leads the Si Futures group of leading U.K. universities in Si research. In 2006, he was appointed as the Director of the National Microelectronics Institute, West Lothian, U.K. He is a Siemens Professor of microelectronics. His current interests are strained Si and SiC device technologies and interconnect reliability.