Abstract—A semi-empirical model is proposed to quantify the tunneling currents through ultrathin gate oxides (1–3.6 nm). As a multiplier to a simple analytical model [1], [2], a correction function is introduced to achieve universal applicability to all different combinations of bias polarities (inversion and accumulation), gate materials (N⁺, P⁺, Si, SiGe) and tunneling processes. Each coefficient of the correction function is given a physical meaning and determined by empirical fitting. This new model can accurately predict all the current components that can be observed: electron tunneling from the conduction band (ECB), electron tunneling from the valence band (EVB), and hole tunneling from the valence band (HVB) in dual-gate poly-Si₁₋ₓGeₓ-gated (x = 0 or 0.25) CMOS devices for various gate oxide thicknesses. In addition, this model can also be employed to determine the physical oxide thickness from I–V data with high sensitivity. It is particularly sensitive in the very-thin-oxide regime, where C–V extraction happens to be difficult or impossible (because of the presence of the large tunneling current).

Index Terms—Direct tunneling model, hole tunneling, tunneling current, ultrathin gate oxide.

I. INTRODUCTION

The downsizing of MOSFETs has been accomplished in a large part by decreasing the oxide thickness to obtain high current drive and good short-channel control. MOS transistors featuring sub-1.5-nm gate oxides have been fabricated [3], [4]. With such thin oxides, the impact of gate tunneling current on static power consumption and substrate current can no longer be ignored. Different components of the gate currents have been experimentally recognized [5]–[8] in dual-gate CMOS devices. However, there is not a simple yet accurate analytical model to quantify each component and address the significance of the tunneling currents through ultrathin gate oxide for future MOS technologies.

Assuming independent and elastic electron processes [9], [10], with a one-band parabolic dispersion relation and using the WKB approximation [11], Schuegraf et al. have derived a simple analytical formula to represent direct-tunneling through a trapezoidal barrier [1], [2]. However, the gate current according to this physical model does not approach zero as \( V_g \) goes to zero and does not fit the experimental data very well in the sub-1-V gate bias range. To provide better accuracy, many researchers have turned to quantum-mechanical simulations employing more complicated electron and hole energy band structures to quantify tunneling currents [12]–[14]. This approach requires numerical integration and detail knowledge of the densities of states and effective masses in the oxide and does not yield an analytical model.

The goal of this paper is to develop a simple expression to predict all significant components of the direct-tunneling currents. In this paper, we propose a semi-empirical tunneling current model, which is a function of the oxide thickness \( T_{ox} \), oxide effective mass \( m_{ef} \), and barrier height \( \phi_b \). With the appropriate \( m_{ef} \) and \( \phi_b \), the proposed model can predict all the significant tunneling current components (see Fig. 1): electron tunneling from the conduction band (ECB), electron tunneling from the valence band (EVB), and hole tunneling from the valence band (HVB). The last component can also be described as valence-band electron tunneling into the valence band. EVB generates the substrate current \( I_{sub} \) [7] in both NMOS and PMOS devices. Although \( I_{sub} \) is more than ten times smaller than \( I_g \), it has the profound impacts on the floating-body effects of SOI devices and the performance of analog circuits using cascade devices. Tunneling currents were measured in dual-gate NMOS and PMOS devices with both inversion and accumulation polarities and compared with the model.

Fig. 1. Schematic representation of the different tunneling components in a Si/SiO₂/Si structure considered in this paper: electron tunneling from the conduction band (ECB), electron tunneling from the valence band (EVB), and hole tunneling from the valence band (HVB).
II. BACKGROUND PHYSICS

From [1] and [2], the direct-tunneling current density for an oxide voltage \( V_{ox} \) smaller than the barrier height \( \phi_b \) can be simply expressed as

\[
J_n = A \left( \frac{\phi_b}{V_{ox}} \right) \left( \frac{2}{\phi_b} - 1 \right) \frac{E_{ox}^2}{E_{ox}} \cdot \exp \left( -B \frac{1 - \left( 1 - \frac{V_{ox}}{\phi_b} \right)^{3/2}}{E_{ox}} \right)
\]

where

\[
A = \left( \frac{q^3}{8\pi \hbar \phi_b} \right);
B = \left( \frac{8\pi \sqrt{2m_{ox}} \phi_b^{3/2}}{3\hbar q} \right);
E_{ox} = \left( V_{ox}/T_{ox} \right);
T_{ox} \text{ the oxide thickness;}
\]

\( m_{ox} \) effective mass in the oxide.

The above expression includes a number of approximations that lead to inaccuracies. The use of the WKB approximation can provide a reasonable match to experimental data by fitting the effective mass and barrier height [9], [15], but the adequacy of this approximation for ultrathin oxide is debatable [14]. The finite density of electrons or energy states in the semiconductor needs to be accounted for. The assumption of a constant effective mass for all energies (all locations at any oxide thickness and gate bias) is not accurate either. Furthermore, as oxide becomes thinner, the quantization effects in the semiconductor have to be considered in order to obtain the oxide potentials as an accurate function of the gate voltage. Because of these reasons, Schuegraf’s model falls short of a complete description of the tunneling current and is unable to fit the tunneling current for the entire \( V_g \) range. A correction function is needed in order to cover the second-order effects listed above.

III. THE PROPOSED TUNNELING MODEL

The proposed new model can be formulated as

\[
J_n = \frac{q^3}{8\pi \hbar \phi_b E_{ox}} \cdot C(V_g, V_{ox}, T_{ox}, \phi_b) \cdot \exp \left( -\frac{8\pi \sqrt{2m_{ox}} \phi_b^{3/2}}{3E_{ox}} \left[ 1 - \left( 1 - \frac{V_{ox}}{\phi_b} \right)^{3/2} \right] \right). \tag{2}
\]

Note that we have replaced the \( E_{ox}^2 \) factor from (1) with a correction function \( C(V_g, V_{ox}, T_{ox}, \phi_b) \). Developed by empirical fitting, \( C(V_g, V_{ox}, T_{ox}, \phi_b) \) can be expressed as

\[
C(V_g, V_{ox}, T_{ox}, \phi_b) = \exp \left[ 20 \frac{V_{ox}}{\phi_b} \left( \frac{V_{ox} - \phi_b}{\phi_{bo}} + 1 \right) \right] \cdot \left( 1 - \frac{V_{ox}}{\phi_b} \right) \cdot \left( \frac{V_g}{T_{ox}} \right) \cdot N \tag{3}
\]

where

\( \alpha \) fitting parameter depending on the tunneling process;
\( \phi_{bo} \) Si/SiO\(_2\) barrier height (e.g., 3.1 eV for electron and 4.5 eV for hole);
\( \phi_b \) actual tunneling barrier height (e.g., 3.1 eV for ECB, 4.2 eV for EVB, and 4.5 eV for HVB [7] with Si electrode).

This model can be applied to poly-SiGe gate technology as well. For tunneling from a poly-Si\(_{0.75}\)Ge\(_{0.25}\) gate with an \( E_g \) of 0.9 eV, electron tunneling has a \( \phi_{bo} \) of 3.1 eV and 4.7 eV for hole tunneling while \( \phi_b \) becomes 4.0 eV when electrons are tunneling from the valence band of the poly-Si\(_{0.75}\)Ge\(_{0.25}\) gate. The exponential term in the correction function covers most of the secondary effects which were neglected in Schuegraf’s derivation and affects the curvature of the tunneling characteristics for \( V_{ox} \) between 0 V and \( \phi_b \). The secondary effects come from the unknown densities of states at the electrode interface (Schuegraf et al. assumed a uniform density of states throughout the derivation as most of the other researchers did) and the effective masses in the oxide. The curvature increases as \( \alpha \) decreases and \( \alpha \)'s of 0.6, 1, and 0.4 provide the overall best fit for ECB, EVB, and HVB, respectively.

The use of \( V_g \) term forces the tunneling current to be zero when \( V_g \) is zero; while the auxiliary function \( N \) is used as an indicator of carrier population (for ECB and HVB cases) or transmission probability (for EVB case) and give accurate onset for each tunneling component. This correction function is universal to all combinations of oxide thicknesses and tunneling mechanisms. The following sections describe the detail expressions of \( N \) for the dominant tunneling current components in dual-gate CMOS devices.

A. \( N \) for ECB and HVB

\( N \) represents the density of carriers in the inversion or accumulation layer of the injecting electrode. For ECB and HVB tunneling processes in both the inversion and accumulation regimes \( N \) is

\[
N = \frac{E_{ox}}{T_{ox}} \left( n_{inv} v_t \cdot \ln \left[ 1 + \exp \left( \frac{V_{ge} - V_{th}}{n_{inv} v_t} \right) \right] \right) \tag{4}
\]

where \( n_{inv} \) and \( n_{acc} \) are swing parameters; \( V_{th} \) threshold voltage; \( V_{FB} \) flat-band voltage; \( V_{ge} \) gate voltage minus the gate-depletion voltage \( (V_g - V_{poly}) \) [16].

The default values of \( n_{inv} \) and \( n_{acc} \) are \( S/v_t \) (\( S \) is the sub-threshold swing) and 1, respectively. \( n_{inv} \) and \( n_{acc} \) are positive for NMOS and negative for PMOS. By solving the Poisson equation in the poly gate under depletion approximation [16],
\[ V_{ge} = V_{FB} + \phi_{so} + \frac{q\varepsilon_{Si}N_{poly}T_{ox}^2}{\varepsilon_{ox}} \cdot \left( \sqrt{1 + \frac{2q^2(V_g - V_{FB} - \phi_{so})}{q\varepsilon_{Si}N_{poly}T_{ox}^2} - 1} \right) \] 

where 

- \( V_{FB} \): flat-band voltage; 
- \( \varepsilon_{Si} \) and \( \varepsilon_{ox} \): dielectric constants of Si and SiO\(_2\), respectively; 
- \( N_{poly} \): poly gate doping concentration; 
- \( \phi_{so} \): surface potential equaling twice the Fermi potential \( \phi_f \).

Note that through the \( V_{FB} \) term, (5) is applicable to N\(^+\) and P\(^+\) poly-Si gates as well as other gate materials such as poly-SiGe [8].

Schuegraf’s model tends to overestimate the tunneling current in the low bias regime because it does not consider the lack of free carriers in the tunneling electrode when the channel is in weak inversion or depletion. Since the threshold voltage and the flat-band voltage are generally perceived as the onsets of carrier population in inversion and accumulation operations, respectively, we therefore use \((V_{ge} - V_{th})/n_{inv}\eta_{th}\) and \((V_g - V_{THB})/n_{acc}\eta_{th}\) in the exponential forms (corresponding to the first part and the second part of the \( N \) formula) to describe the trends of inversion and accumulation carrier population. \( n_{inv}\eta_{th} \) and \( n_{acc}\eta_{th} \) represent the increasing rate of carrier population as the gate bias changes. In the subthreshold or subflatband regimes \((V_g < V_{FB}\) or \(V_g < V_{THB}\)), the carrier population depends exponentially on gate bias and the tunneling current is limited by the number of free carriers. Once the channel is in strong inversion or accumulation, the carrier population becomes a linear function of \( (V_{ge} - V_{th}) \) or \( (V_g - V_{THB}) \). The exponential format is consistent with the fact that carrier population is exponentially dependent on the surface potential of the electrode which changes more or less linearly with \( V_g \) or \( V_{ge} \). Also note that the inversion carrier population term dominates when gate bias favors inversion \( (V_g > 0 \text{ for NMOS and } V_g < 0 \text{ for PMOS}) \) while the accumulation term dominates when gate bias favors accumulation \( (V_g < 0 \text{ for NMOS and } V_g > 0 \text{ for PMOS}) \). Fig. 2 shows the plot of \( N \) versus \( V_g \) for a NMOSFET with a 2.5-nm gate oxide.

As a reminder, ECB tunneling dominates in NMOS devices for both bias polarities and in the accumulation regime in PMOS devices [7]. In all these ECB cases, the injecting electrodes are well supplied with conduction band electrons at the surface. These conduction band electrons can tunnel through the relatively small 3.1 eV barrier rather easily. HVB dominates the gate leakage of the inverted PMOS devices in the low bias range [8].

\[ N = \frac{\varepsilon_{ox}}{T_{ox}} \cdot \left( n_{EVB} \eta_{th} \cdot \ln \left[ 1 + \exp \left( \frac{|V_{ox} - \phi_{ek}|}{n_{EVB} \eta_{th}} \right) \right] \right) \] 

where \( \phi_{ek} = E_{ek}/q \) by default and \( n_{EVB} \) is a fitting parameter with a default value of 3. Equation (6) is very similar in format as the aforementioned carrier population terms but with a different physical meaning. Not only the density of tunneling carriers can affect tunneling current, but also the density of the receiving states. In the case of \( |V_{ox}| < \phi_{ek}, \text{ EVB to the energy bandgap of the other side is prohibited since no energy state with the corresponding energy level is available in the band gap to receive the tunneling electrons. Therefore the transmission probability as well as the resulting tunneling current become negligible, as shown in Fig. 3. This fact is not included in (1). The \((|V_{ox}| - \phi_{ek})/n_{EVB} \eta_{th}\) term is therefore used in the exponential form to address this issue. As long as \( |V_{ox}| < \phi_{ek}, \text{ EVB current is negligible} \). \( N \) depends exponentially on \( |V_{ox}| \) and is rather small which corresponds to the negligible transmission probability. As \( |V_{ox}| \) exceeds \( \phi_{ek}, \text{ EVB current becomes a linear function of oxide voltage. The EVB component has a special significance as it appears in inverted NMOS devices as the substrate current. Also EVB produces such a substrate current in an inverted PMOSFET} \) [8] and it dominates the gate leakage in the high bias range.

B. \( N \) for EVB

For EVB tunneling process, \( N \) can be described as

\[ N = \frac{\varepsilon_{ox}}{T_{ox}} \cdot \left( n_{EVB} \eta_{th} \cdot \ln \left[ 1 + \exp \left( \frac{|V_{ox} - \phi_{ek}|}{n_{EVB} \eta_{th}} \right) \right] \right) \] 

IV. \( V_{ge} - V_{ox} \) RELATIONSHIP

According to (2), in order to predict the gate tunneling current, an accurate knowledge of \( V_{ox} \) must be obtained. Using the
effective gate voltage—$V_{ge}$ from (5), $V_{ce}$ can be calculated as

$$V_{ce} = V_{ge} - \phi_s - V_{FB}$$  \hspace{1cm} (7)

$$\phi_s = \frac{\gamma}{2} \left[ -1 + \sqrt{1 + 4\left(\frac{V_{ge} - V_{FB}}{\gamma^2}\right)} \right]$$  \hspace{1cm} (8)

$$\gamma = \sqrt{2\pi/\epsilon N_{sub}}$$  \hspace{1cm} (9)

where

- $\phi_s$: surface band bending of the substrate;
- $N_{sub}$: channel doping concentration;
- $\gamma$: body-effect parameter.

Note that the gate-depletion effect can affect $V_{ce}$ through $V_{ge}$ and therefore impact the tunneling characteristics. Fig. 4 shows the calculated $V_{ce}$-$V_{ge}$ relationship for various gate doping levels and oxide thicknesses. The sensitivity of $V_{ce}$ to $N_{poly}$ increases as $V_{ge}$ increases or oxide thickness decreases due to stronger gate depletion effect.

V. CALIBRATION OF EFFECTIVE MASSES AND OTHER MODEL PARAMETERS

For a given oxide voltage and oxide thickness, the calculation of the tunneling current component depends only on two parameters: the oxide effective mass ($m_{ox}$) and the barrier height ($\phi_B$). We first calibrated these parameters and verified our model by empirical fitting to the measured $I$–$V$ data. For this purpose, we used (10 $\mu$m $\times$ 10 $\mu$m) N$^+$ poly-Si-gated NMOS and P$^+$ poly-Si-gated PMOS transistors with ultrathin gate oxides of ~2.5 nm and ~2.85 nm. Optical measurements, performed with an ellipsometer, indicate oxide thicknesses of 2.45 nm and 2.85 nm with an accuracy of 0.3 nm. From the $C$–$V$ simulation, we estimated the effective channel and gate doping concentrations of $4.7 \times 10^{17}$ cm$^{-3}$ and $9 \times 10^{19}$ cm$^{-3}$ for N$^+$ poly-Si-gated NMOS transistors with a flat-band voltage of $-1.0$ V and they are $5.6 \times 10^{17}$ cm$^{-3}$ and $1 \times 10^{20}$ cm$^{-3}$ for the P$^+$ PMOS transistors with a flat-band voltage of 1.0 V. Oxide thicknesses of 2.5 nm and 2.85 nm were also extracted from the $C$–$V$ fitting which accounts for both quantization and gate depletion effects. The tunneling $I$–$V$ characteristics can then be calculated. Fig. 5 gives an explicit comparison between the proposed model and Schuegraf’s model for various oxide thicknesses. The $I$–$V$ data from N$^+$ NMOSFETs are also included. With $\phi_A = 3.1$ eV assumed for ECB, Schuegraf’s model fails to predict the tunneling characteristics throughout the whole bias range even with a favorably small effective mass, while the proposed model is in excellent agreement with the data using a $m_{ox}$ of 0.4$m_e$. We could obtain the same level of agreement between experiments and the model by fine tuning both $m_{ox}$ and $\phi_A$. However, we adopted the commonly accepted values of $\phi_A = 3.1$ eV for ECB, $\phi_B = 4.2$ eV for EVB and $\phi_B = 4.5$ eV for HVB [7] (considering Si as the tunneling electrode) and consequently effective masses of $m_{ox} = 0.4m_e$ for ECB, $m_{ox} = 0.30m_e$ for EVB and $m_{ox} = 0.32m_e$ for HVB were obtained from the overall best fit. The different masses for ECB, EVB and HVB is not surprising since the exact band structure of the oxide bandgap is not well known as demonstrated by the many relations and mass values proposed in [7], [11]–[14]. The fact that the effective mass is lower for electrons closer to the oxide midgap (EVB tunneling) is consistent with the trend suggested by Mead [19] and the EVB effective mass of 0.30$m_e$ obtained in this work is close to Shanware’s value of 0.33$m_e$ [20]. The hole mass of 0.32$m_e$ also falls between the masses obtained by Schuegraf (0.2$m_e$) [2] and Gritsenko (0.35$m_e$) [21].
VI. VALIDATION OF THE MODEL

Using the same ECB parameters suggested from the previous section \( m_{\text{ECB}} = 0.4 m_0 \) and \( \phi_0 = 3.1 \text{ eV} \), our model can also provide excellent predictions on the tunneling currents in NMOS and PMOS devices with the substrate in accumulation, as shown in Fig. 7. Under accumulation bias, the difference of \( E_g \) between \( V_g \) and \( V_{\text{ox}} \) is assumed in strong accumulation while \( V_{\text{ox}} - V_g \) relationship predicted by (7)–(9) is still good for the depletion region except for the replacement of \( m_{\text{ECB}} \) with \( m_{\text{HV}} \).

We also test this new model on \( P^+ \) poly-Si\(_{0.75}\)Ge\(_{0.25}\)-gated PMOSFETs in the inversion regime. As reported in [8], there are two tunneling components in the gate leakage of \( P^+ \) poly-Si\(_{0.75}\)Ge\(_{0.25}\)-gated PMOSFETs: EVB of the poly-Si\(_{0.75}\)Ge\(_{0.25}\) gate with \( m_{\text{ECB}} = 0.3 m_0 \) and \( \phi_0 = 4.0 \text{ eV} \) (assume the valence band of the poly-Si\(_{0.75}\)Ge\(_{0.25}\) is \( \sim 0.2 \text{ V} \) higher than that of the poly-Si and therefore the valence-band electrons see a reduction of 0.2 V in barrier height), and HVB of the Si substrate with \( m_{\text{ECB}} = 0.3 m_0 \) and \( \phi_0 = 4.5 \text{ eV} \). The same effective masses as those in the poly-Si gate technology are assumed. Plugging these parameters into the model, the total gate currents were calculated as the sum of the two tunneling components. Tunneling currents for both \( P^+ \) poly-Si and \( P^+ \) poly-Si\(_{0.75}\)Ge\(_{0.25}\)-gated PMOSFETs were measured and plotted in Fig. 8 with model predictions superimposed. The HVB currents are separated by \( \sim 0.2 \text{ V} \) due to the difference in the gate work-function. In addition, the EVB current is raised in the poly-Si\(_{0.75}\)Ge\(_{0.25}\)-gated device because the barrier against EVB is 0.2 V lower.

In order to validate this model against the data and simulations in the literature, we reviewed the quantum-mechanical modeling of the electron tunneling currents reported by Lo et al. for ECB [12] and Shanware et al. for EVB [20]. In Lo’s work, taking into account the details of the NMOS structures, the simulation results showed good agreement with the experimental results especially for \( V_g \geq 1.5 \text{ V} \). However, the misfit in the sub-1-V regime becomes quite large as oxide gets thinner. With Tunnel-PISCES simulations using the modified Gundlach method, Shanware predicted the EVB tunneling currents with only fair fitting. Using the channel and gate doping concentrations \( (4.3 \times 10^{17} \text{ cm}^{-3} \) and \( 9 \times 10^{19} \text{ cm}^{-3} \), respectively, for Lo’s experiment while they are \( 4.7 \times 10^{17} \text{ cm}^{-3} \) and \( 1.5 \times 10^{20} \text{ cm}^{-3} \) for Shanware’s work) estimated by the \( C-V \) simulation, the oxide potential profiles were calculated and the corresponding tunneling current for each oxide thickness was calculated using the appropriate \( m_{\text{ECB}} \) and \( \phi_0 \) combination. Their simulation and experimental results are plotted in Figs. 9 and 10 superimposed with our model predictions. All the extracted oxide thicknesses are in good agreement with those from the simulations (in parenthesis). Not only our model gives consistent current levels as compared to the simulation results, but also our calculations fit the experimental data with great accuracy over the entire \( V_g \) range. These comparisons support the universality as well as the robustness of the proposed model.

Finally, we address the case of \( P^+ \) poly-Si-gated PMOS transistors under negative gate bias, whose calculated and measured
Fig. 9. Simulation (dashed lines) and experimental (dots) results adapted from Lo et al. [12] for gate currents in inverted NMOSFETs are compared with our calculations (solid lines) for various oxide thicknesses. In most case, our analytical model agrees with Lo’s data better than the quantum simulation.

Fig. 10. Simulation (dashed lines) and experimental (dots) results adapted from et al. [20] for substrate currents in inverted NMOSFETs are compared with our calculations (solid lines) for various oxide thicknesses.

Fig. 11. Calculation (solid lines) and experimental (hollow dots) results for gate currents in inverted PMOSFETs for various oxide thicknesses. The dashed line marks the HVB-to-EVB transition predicted from our model. For gate bias values of 1.0 V, 1.5 V, 2.0 V, 2.5 V, and 3.0 V, (A, B)’s are (1.35, 0.195), (1.67, 0.2), (1.832, 0.207), (1.965, 0.215), and (2.096, 0.224), respectively, for accumulated NMOSFETs.

Fig. 12. Tunneling currents predicted by the new model with various oxide thicknesses. Due to the high sensitivity, error as small as of ±0.01 nm in thickness extraction can be easily attained.

Fig. 13. Gate oxide thickness obtained from fitting the tunneling currents using our model, designated by $T_{ox}$, is compared to $T_{ox}$ extracted by C–V fitting.

Fig. 14. $T_{ox}$ can be extracted from the gate currents using a simple linear equation: $T_{ox} = A - B \cdot \log\left(\frac{I_g}{(\text{A/cm})^2}\right)$. For gate bias values of 1.0 V, 1.5 V, 2.0 V, 2.5 V, and 3.0 V, (A, B)’s are (1.35, 0.195), (1.67, 0.2), (1.832, 0.207), (1.965, 0.215), and (2.096, 0.224), respectively, for accumulated NMOSFETs.

$V_I$ characteristics are shown in Fig. 11. For the measured data, the oxide thicknesses were determined by best fitting the $I-V$ curves. It can be seen that a good match between the measurements and calculations is reached again. The gate tunneling currents for oxide thicknesses down to 1 nm were also predicted. In Fig. 11, we mark the HVB-to-EVB transition with a dashed line and the transition seems to occur at a higher gate bias value as oxide becomes thinner. It is ~1.5 V for oxide of 3.5 nm and exceeds 3 V when oxide approaches 1 nm. Below 2.5 nm, the PMOSFET on-state $I_g$ will be dominated by hole tunneling under normal device operation. Compared with the $I-V$ tunneling characteristics in Fig. 9, PMOSFET gate current is roughly ten times smaller than NMOS gate leakage. Although the oxide scaling limit will still be first reached by NMOS devices in terms of leakage level, we should start paying more attention to the conceptually more damaging hole tunneling current in PMOSFETs in consideration of device reliability.

VII. EXTRACTION OF $T_{ox}$ USING $I_g-V_g$ CHARACTERISTICS

The direct-tunneling current has been known to have the exponential dependence on oxide thickness as also indicated in our model. Fig. 12 demonstrates its sensitivity. $I_g-V_g$ data is...
plotted with several $I-V$ curves calculated by the model superimposed over a wide bias range of 1.5 V. The model predictions span over a narrow range of $T_{\text{ox}}$ (from 2.46 nm to 2.5 nm) and all the curves are basically parallel to each other. As can be seen, not only excellent agreement between the data and the prediction can be achieved with an oxide thickness of 2.48 nm, but also all the data points easily fall between the 2.47-nm and 2.49-nm curves which implies a high accuracy of ±0.01 nm in the $T_{\text{ox}}$ extraction. The gate oxide thickness obtained from fitting the tunneling currents using our model, designated by $T_{\text{ox},\text{DT}}$, is compared to $T_{\text{ox},\text{CV}}$ extracted by $C-V$ fitting in Fig. 13. As can be seen, good agreement between the two methods is achieved. $C-V$ measurement becomes increasingly difficult at smaller $T_{\text{ox}}$, while the $T_{\text{ox},\text{DT}}$ extraction is not limited by the scaling of gate oxide. $T_{\text{ox}}$ extraction from $I_g$ measurement in the accumulation region is insensitive to $V_{th}$ and gate depletion and therefore convenient and recommended [4].

Fig. 14 shows $I_g$ in accumulated NMOS at different gate biases using the proposed model. A simple linear equation in the form of $T_{\text{ox}}(\text{nm}) = A - B \log(J_g (A/cm^2))$ can be used to extract $T_{\text{ox}}$ from the gate currents as shown in Fig. 14. For gate bias values of 1.0 V, 1.5 V, 2.0 V, 2.5 V, and 3.0 V ($A, B$)'s of (1.35, 0.195), (1.67, 0.2), (1.832, 0.207), (1.965, 0.215), and (2.096, 0.224) are recommended, respectively, for best $T_{\text{ox}}$ predictions in accumulated NMOSFETs.

VIII. CONCLUSIONS

With the increasing importance of the gate tunneling current, a good quantitative understanding of direct-tunneling currents is essential. In this work, we have presented a physically-based model and semi-empirical equations for a large variety of experimental conditions including injection of electrons and holes from accumulation and inversion layers, poly-Si$_{1-x}$Ge$_x$ gates, and different bias polarities. Although limited when comes to the detail physics of tunneling, this proposed model provides excellent predictions of the electron tunneling from conduction band, electron tunneling from valence band, and hole tunneling from valence band taking into account both quantization and gate-depletion effects. We have adopted the well-known barrier heights and determined the less known effective mass as relevant for each tunneling process.

For devices with ultrathin gate oxides, this semi-empirical model can also help to address two very important issues: 1) the relative significance of the different tunneling components and 2) the sensitivity of the tunneling currents on oxide thickness. The second point, in particular, might have greater impact on the use of the $I-V$ tunneling currents instead of conventional $C-V$ characteristics to assess the physical oxide thickness for future sub-0.1-$\mu$m CMOS technologies.

REFERENCES


Wen-Chin Lee (S’97–M’99) received the B.S. degree in electrical engineering from National Tsing-Hua University, Hsinchu, Taiwan, R.O.C., in 1993, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1997 and 1999, respectively. His research involved poly-SiGe gate for dual-gate CMOS application, modeling of direct-tunneling current through ultrathin gate oxide, development of sub-50-nm CMOS FinFET, and other deep submicrometer technologies. He joined Intel Corporation, Hillsboro, OR, in 2000 as a Senior Process Engineer and is currently involved with the development of 0.1-$\mu$m CMOS technology node and novel process modules.
Chenming Hu (S’71–M’76–SM’83–F’90) received the B.S. degree from Taiwan University, Taiwan, R.O.C., in 1968, and the M.S. and Ph.D. degrees from the University of California, Berkeley. Currently, he is the TSMC Distinguished Professor of Electrical Engineering and Computer Sciences at the University of California, Berkeley. His research areas include microelectronic devices and technology and device modeling for circuit simulation. He has authored or coauthored five books and over 600 research papers. He is a Member of the editorial boards of the Journal of Semiconductor Science and Technology and of the Journal of Microelectronics Reliability.

Dr. Hu is a Member of the National Academy of Engineering, a Fellow of the Institute of Electrical and Electronics Engineers, a Fellow of the Institute of Physics, an Honorary Professor of the Chinese Academy of Science, and an Adjunct Professor of Peking University. He leads the development of the MOSFET model BSIM3v3, the industry standard model of IC simulation and the recipient of an R&D 100 Award. He received the 1997 Jack A. Morton Award for contributions to the physics of MOSFET reliability. He has received UC Berkeley’s highest honor for teaching—the Distinguished Teaching Award, the Monie A. Ferst Award of Sigma Xi, the W.Y. Pan Foundation Award, the National Technology University Outstanding Teaching Award, and DARPA Most Significant Technological Accomplishment Award for codeveloping the FinFET transistor structure.