Analytical Modeling of Tunneling Current through SiO2-HfO2 Stacks in MOS Structures

J. Coignus, R. Clerc, Cédric Leroux, G. Reimbold, G. Ghibaudo, F. Boulanger

Analytical modeling of tunneling current through SiO_2-HfO_2 stacks in metal oxide semiconductor structures

J. Coignus[a)]
IMEP-LAHC, MINATEC-INPG, 3 Parvis Louis Néel, 3806 Grenoble Cedex 1, France and CEA-LETI MINATEC, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France

R. Clerc
IMEP-LAHC, MINATEC-INPG, 3 Parvis Louis Néel, 3806 Grenoble Cedex 1, France

C. Leroux and G. Reimbold
CEA-LETI MINATEC, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France

G. Ghibaudo
IMEP-LAHC, MINATEC-INPG, 3 Parvis Louis Néel, 3806 Grenoble Cedex 1, France

F. Boulanger
CEA-LETI MINATEC, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France

(Received 16 July 2008; accepted 27 October 2008; published 9 February 2009)

This work presents an original approach to model direct tunneling current through high-κ dielectrics including SiO_2 interfacial oxide from electron inversion layers. Quantum confinement is taken into account by means of an improved triangular well approximation including physically-based analytical corrections of subband energy levels. An efficient way to compute tunnel transmission probability is also proposed, taking into account the reflections on discontinuous dielectrics interfaces. Finally, this model has been successfully validated by comparison to both numerical simulations and experimental results. © *2009 American Vacuum Society.*
[DOI: 10.1116/1.3043539]

I. INTRODUCTION

Ultrascaled complementary metal oxide semiconductor (CMOS) technology, for 45 nm node and beyond, requires the replacement of conventional SiO_2 gate dielectrics by high-κ materials, to achieve the target in term of leakage current through the gate.[1] Recently, HfO_2-based dielectrics have been introduced in industry for a 45 nm product by Intel Corp.[2]

In this context, simulation of direct tunneling current through the gate insulator is needed for both process technology optimization and parameter extraction from I_g-V_g measurements. If the issue of tunneling current modeling in high-κ stacks (Fig. 1) has been extensively studied by means of numerical solution of Schrödinger equation,[3,4] the development of physically based and accurate analytical models still require improvements.

Accurate analytical model needs to account for quantum confinement at the interface of carriers responsible for the transport in dielectric (i.e., tunneling). Such quantum calculation usually requires numerical solution of the coupled Schrödinger and Poisson equations.[5] Various analytical alternatives have been proposed in the literature.[5] If these approaches correctly reproduced the total charge versus voltage characteristics, they fail, however, to model accurately the position of the first subband energy levels, leading in the case of tunneling in high-κ dielectrics, to erroneous current prediction.

Moreover, tunneling current modeling usually also requires the calculation of tunneling lifetime of confined carriers. Several approaches have been proposed to model efficiently these tunneling lifetimes in pure SiO_2 dielectrics,[6] beyond the conventional Wentzel-Kramers-Brillouin (WKB) approximation, known to be quite inaccurate.[7] The situation is more complicated in the case of high-κ stacks, as they usually also includes a SiO_2 interfacial layer. To our knowledge, no satisfactory solution has been found so far to achieve an analytical solution in multilayer dielectrics beyond the WKB approximation.

In this work, an original approach to compute electron tunneling current from inversion layer through high-κ stack is proposed, improving the state of the art of analytical modeling of both quantum confinement and tunneling through multilayer dielectrics. Results have successfully been compared to experimental data in HfO_2/SiO_2 stacks.

II. MODEL DESCRIPTION

The current flowing from a confined gas of carriers in close-to-equilibrium condition through a tunneling dielectric is usually expressed as[8]

$$J = \sum_n \frac{Q_n}{\tau_n}, \tag{1}$$

where J is the current density (A m^{-2}), Q_n the charge density contained in nth subband, and τ_n the tunneling lifetime. While Q_n is usually computed by solving Poisson and Schrödinger (PS) equations, τ_n can be modeled using differ-

a)Electronic mail: jean.coignus@minatec.inpg.fr

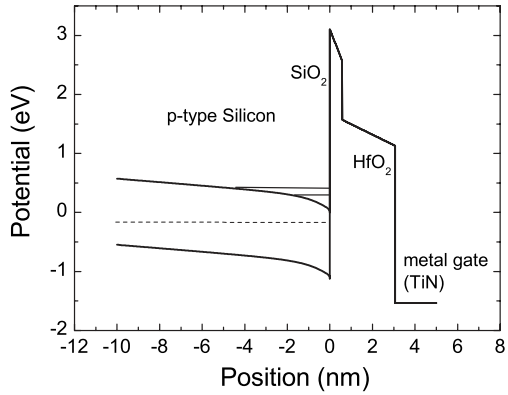

FIG. 1. Illustration of the double-layer stack considered in this work: *p*-type silicon substrate, SiO_2–HfO_2 gate oxides, and TiN metal gate.

ent methodologies such as Bardeen's approach,[9,10] resonant transfer matrix,[11] or transparency-based approach.[12–18] In the latter, the tunneling lifetime is computed as $1/\tau_n = f_n T_n$, where f_n is the impact frequency and T_n the tunnel barrier transparency. This technique has been chosen in this work, as, contrary to the other ones, it offers the possibility of analytical solution in the multilayer dielectrics case.

A. Modeling of quantum confinement

The modeling of quantum confinement at silicon-oxide interface typically requires numerical self-consistent Poisson–Schrödinger simulations.[3,19] Analytical alternatives[5] (such as the variational approach or triangular well approximation) are accurate enough to model only the charge located in the ground state (the first unprimed subband), which is usually satisfactory for charge and *C-V* modeling. However, these approximations can lead to significant errors when applied to tunneling current computation, where the first primed subband, as shown later, also plays a significant role. For this reason, the conventional triangular well approximation (TWA) has been improved to account for several subbands, following the approach proposed by Ferrier *et al.*[20]

In the TWA model, the potential energy profile is supposed to be given by $V(x)=qFx$, where F denotes a constant field. In such potential, the energy levels E (referenced to the silicon conduction band edge) and wave function Ψ solution of Schrödinger's equation are given by

$$E_{0,i}^{L,T} = \left(\frac{\hbar^2}{2m_{L,T}}\right)^{1/3} \cdot \left[\frac{3\pi qF}{2} \cdot \left(i + \tfrac{3}{4}\right)\right]^{2/3}, \tag{2}$$

$$\Psi_i^{L,T}(x) = \mathrm{Ai}\left[\left(\frac{2m_{L,T}qF}{\hbar^2}\right)^{1/3} \cdot \left(x - \frac{E_i^{L,T}}{qF}\right)\right], \tag{3}$$

where Ai is the Airy function of the first kind.[21] To achieve a better comparison with PS results, and to compensate for the nonlinearity of the real potential energy profile, the field F used in Eqs. (2) and (3) is not taken equal to the surface field F_s, but rather to

$$F = \frac{Q_{\mathrm{dep}} + fQ_{\mathrm{inv}}}{\varepsilon_s}. \tag{4}$$

Q_{dep} and Q_{inv} stands, respectively, for depletion and inversion charge expressed as

$$Q_{\mathrm{dep}} = \sqrt{2qN_a\varepsilon_s V_s},$$

$$Q_{\mathrm{inv}} = q\sum_i n_i^{L,T}, \tag{5}$$

where V_s is the band bending in the silicon substrate, N_a the *p*-silicon doping level, and $n_i^{L,T}$ the electron density on the ith L,T-subband.

The f parameter in Eq. (4) has been found to be equal to 0.5 for electron, F being in this case the effective field, i.e., the average transverse field seen by electrons in the inversion layer. Using $f=1$ in Eq. (4) would have given the surface effective field F_s. Note that F and F_s only coincide in weak inversion regime, where $Q_{\mathrm{dep}} \gg fQ_{\mathrm{inv}}$.

In strong inversion regime, Eq. (2) is inaccurate to model properly the energy levels. To improve the TWA, different fields $F_i^{L,T}$ have been introduced for each subband, using a set of parameters $f_i^{L,T}$, instead of a single f.[12,20] This modification accounts for the fact that each different subbands see a different average transverse field, as the centroid of charge changes from a subband to another.

The set of parameter $f_i^{L,T}$ have been extracted from self-consistent solution of Poisson–Schrödinger equations, and are given by $f_{0L}=0.58$, $f_{0T}=0.47$, $f_{1L}=0.39$, $f_{1T}=0.24$, $f_{2L}=0.28$, and $f_{2T}=0.18$. Detailed comparisons with Poisson–Schrödinger equations have shown that these parameters do not significantly depend on substrate doping (in a tested range of 10^{23}–5×10^{24} m^{-3}) and oxide thickness.

In weak inversion regime, the nonlinearity of the potential well has been taken into account using the perturbation theory, as explained by Ferrier *et al.*[20] Energy levels are corrected by a ΔE_{NL} shift,

$$\Delta E_{\mathrm{NL},i}^{L,T} = \frac{qN_a}{\varepsilon_s}\frac{4}{15}\left(\frac{\hbar^2}{2m_{L,T}qF_i^{L,T}}\right)^{2/3} \cdot \left[\frac{3\pi}{2} \cdot \left(i + \tfrac{3}{4}\right)\right]^{4/3}. \tag{6}$$

Finally, the wave function penetration in the tunneling dielectric has been also considered, leading to an additional ΔE_{WFP} correction factor, expressed by

$$\Delta E_{\mathrm{WFP},i}^{L,T} = \frac{\hbar m_{\mathrm{ox}} F_i^{L,T}}{m_i^{L,T} \cdot \sqrt{2 \cdot m_{\mathrm{ox}} \cdot (\Phi_{\mathrm{ox}} - E_{0,i}^{L,T})}}. \tag{7}$$

Using these two corrections, the energy levels are given by

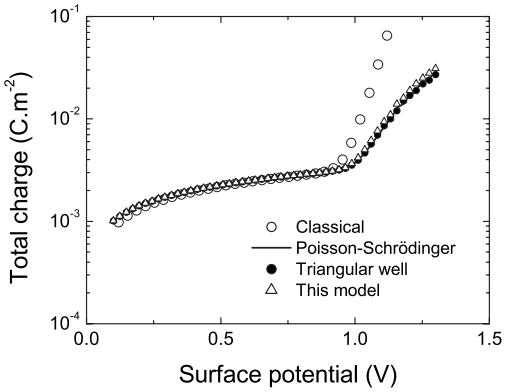

FIG. 2. Total charge versus surface potential computed with classical and quantum approaches: Poisson–Schrödinger resolution, triangular well approximation, and this model. ($N_a = 3 \times 10^{23}$ m^{-3})

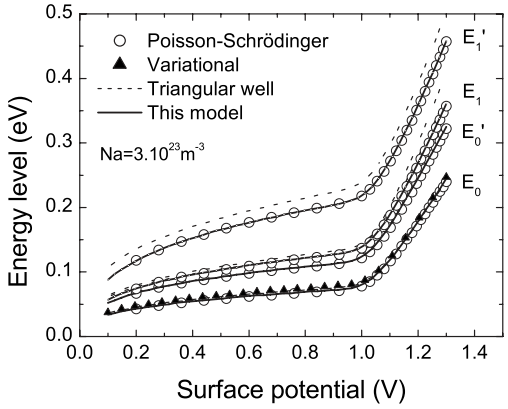

FIG. 3. First energy levels E_0, E_0', E_1, and E_1' in quantized inversion layer vs surface potential computed with Poisson–Schrödinger solution, variational method, triangular well approximation, and improved triangular well approximation (this model) ($m = 0.92m_0$ and $m' = 0.19m_0$).

$$E_i^{L,T} = E_{0,i}^{L,T} - \Delta E_{\mathrm{NL},i}^{L,T} - \Delta E_{\mathrm{WFP},i}^{L,T}. \tag{8}$$

Finally, the electron density n_i for each applied gate voltage is given by[18,20]

$$n_i^{L,T} = \frac{m_d^{L,T}}{\pi\hbar^2} k \cdot T \ln\left[1 + \exp\left(\frac{E_F - E_i^{L,T}(n)}{kT}\right)\right]. \tag{9}$$

E_F is the Fermi level position with respect to the silicon conduction band,

$$E_F = V_s - kT \ln\left(\frac{N_a}{n_i}\right) - \frac{E_g}{2}, \tag{10}$$

where E_g denotes the silicon band gap and n_i the intrinsic silicon doping level.

For each applied voltage, self-consistency is achieved by numerically finding the root solution of Eq. (9), less time consuming than standard Poisson–Schrödinger resolution.

Figures 2 and 3 show the very good agreement with self-consistent resolution of both Poisson and Schrödinger equations: this approach allows us to model accurately not only the total charge (Fig. 2), but also the energy levels (Fig. 3) and charge occupancy for all relevant subbands.

B. Modeling of tunneling current

As previously mentioned, the computation of tunneling current typically requires the calculation of the transmission probability through the gate insulator. In the case of high-κ stacks (Fig. 1), as a SiO_2 interfacial layer is usually present, the transmission probability has to be computed through a stack of two different dielectric layers. In absence of charges trapped in the dielectrics, as all barriers have a triangular shape, the exact transmission probability can be computed using the Airy matrix formalism.[17] Another numerical approach, called the transfer matrix approach,[15] consists in approximating the real barrier by a succession of N constant potential energy barriers. Compared to the previous method, no special function (such as Airy function, requiring special library code not available in standard compact model) is required to apply this technique. Nevertheless, a large number of mesh points (typically larger than 20) is needed to ensure a good accuracy.

The only analytical alternative to model transparency through multitunnel barrier is to use the Wentzel–Kramers–Brillouin approximation, leading, however, to significant errors. To improve this method, it has been proposed by Register and co-workers to introduce a pre-exponential corrective term. This technique has been first discussed in the single barrier case[7] and then extended to the double barrier case using an *ad hoc* correction and neglecting reflections on SiO_2/HfO_2 and HfO_2/metal interfaces.[13] Although empirical, this procedure has been shown to significantly improve WKB formula accuracy.

In this work, an effective-barrier-height method is proposed, based on the approximation carried out by Simmons when studying tunneling in metal-insulator-metal structures.[22] It consists in approximating the double triangular barrier by double squared barrier, with an effective barrier height ϕ_{eff}, given by $\phi_{\mathrm{eff}} = \phi_0 - V/2$, where ϕ_0 is the real barrier height, and V the considering oxide voltage drop. This approximation leads to an analytical expression of the tunnel transparency, without requiring the use of any special function.

The different methods mentioned above (Airy, transfer matrix, WKB with and without Register's corrective term, and this method) are compared first in the case of a single tunnel barrier in Sec. II B 1 and then in the case of a double layer barrier in Sec. II B 2.

1. Single-barrier transmission probability

Considering a squared barrier with an effective barrier height ϕ_{eff}, the calculation of the transmission coefficient T can be achieved by applying the standard procedure of matching wave functions and probability currents at the different substrate-dielectric-gate interfaces, leading to the following formula:

$$T = \frac{16 v_{in} (v_{ox}^*)^2 v_{out}}{|(v_{ox}^* - i v_{in}) \cdot (i v_{out} - v_{ox}^*) \exp(K_{ox} t_{ox}) + (v_{ox}^* + i v_{in}) \cdot (i v_{out} + v_{ox}^*) \exp(-K_{ox} t_{ox})|^2}, \quad (11)$$

where t_{ox} is the barrier thickness, v_{in}, v_{ox}^*, and v_{out} the carrier velocities in silicon substrate, squared barrier, and metal gate, respectively, and K_{ox} the wave vector into the dielectric. Energy reference has been taken equal to the silicon conduction band. Assuming a parabolic band structure, carrier velocities are given by

$$v_{in}(E) = \sqrt{\frac{2E}{m_{L,T}}},$$

$$v_{ox}^*(E) = \frac{\hbar K_{ox}}{m_{ox}} = \sqrt{\frac{2}{m_{ox}}(\Phi_{eff} - E)},$$

$$v_{out}(E) = \sqrt{\frac{2}{m_0}(E + V_{ox})}. \quad (12)$$

As mentioned in Sec. II A, tunneling current is mainly composed of electrons emitted from the first lowest subbands. As a consequence, in this low energy range, the product $K_{ox}t_{ox}$ is usually greater than 1, and the factor $\exp(-K_{ox}t_{ox})$ in Eq. (11) can be neglected compared to $\exp(K_{ox}t_{ox})$. Thus, in this case, Eq. (11) simply reduces to

$$T = \frac{16 v_{in} (v_{ox}^*)^2 v_{out}}{[v_{in}^2 + (v_{ox}^*)^2] \cdot [(v_{ox}^*)^2 + v_{out}^2]} \exp(-2K_{ox}t_{ox}). \quad (13)$$

This formula is in good accordance with the result obtained empirically by Register *et al.* in Ref. 7, except that the $\exp(-K_{ox}t_{ox})$ term has not been replaced by its WKB equivalent. The pre-exponential term in Eq. (11) stands for reflections at both silicon-dielectric and dielectric-gate interfaces and has been rigorously derived without any other assumptions. As Eq. (13) as well as the Register corrective term are no longer valid when K_{ox} tends to 0, it may be convenient to use Eq. (11) instead, especially when tunneling through relatively thick dielectric by thermionic or Fowler–Nordheim mechanisms is concerned.

Figure 4 shows the transmission probability through a thin SiO_2 layer versus incoming carriers energy level. SiO_2 layer parameters (effective tunneling mass, barrier height) have been taken equal to their standard values.

In the low energy range, where direct tunneling usually occurs, all methods show a good agreement with the exact Airy formalism results, except the uncorrected WKB approximation, which exhibits an error of more than 100% in the worst case (Fig. 5). Correction of reflections at dielectric interfaces allows decreasing the WKB error. As far as single barriers are concerned, the effective potential approximation proposed in this work and Register's corrected WKB method appear to be the best approaches to compute analytically the transmission probability.

2. Double-barrier transmission probability

The previous techniques have been extended to the case of tunneling through a double dielectric barrier. Using the concept of effective barrier height, the transmission coefficient T of the double tunnel barrier can be analytically derived, leading to

$$T = \frac{v_{in}}{v_{out}} \frac{4}{|\text{ch } 2(A \text{ ch } 1 - B \text{ sh } 1) - \text{sh } 2(C \text{ ch } 1 - D \text{ sh } 1)|^2}, \quad (14)$$

where

$$A = \frac{v_{in}}{v_{out}} + 1,$$

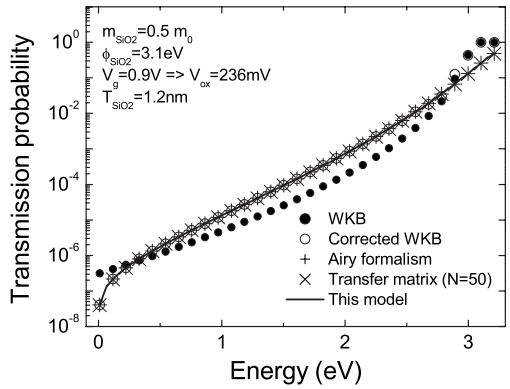

FIG. 4. Transmission coefficient of a single SiO_2 layer versus incoming carriers energy. Five methods are compared (uncorrected and corrected WKB, Airy formalism, transfer matrix method, and this model).

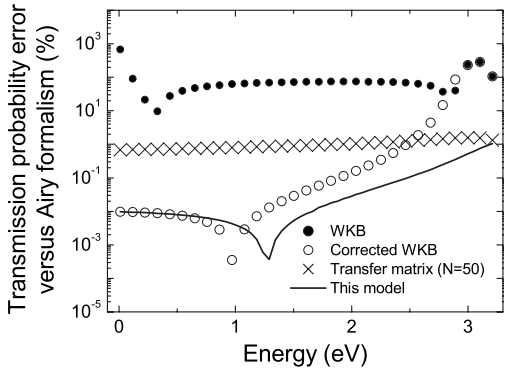

FIG. 5. Relative error of previously plotted methods versus Airy formalism (exact). Parameters are the same than Fig. 4.

$$B = \frac{i \cdot v_{\text{in}}}{V_1^*} + \frac{V_1^*}{i \cdot v_{\text{out}}},$$

$$C = \frac{i \cdot v_{\text{in}}}{V_2^*} + \frac{V_2^*}{i \cdot v_{\text{out}}},$$

$$D = \frac{V_1^*}{V_2^*} + \frac{v_{\text{in}} \cdot V_2^*}{v_{\text{out}} \cdot V_1^*},$$

$$\text{ch } 1 = \cosh(K_1 t_1),$$

$$\text{ch } 2 = \cosh(K_2 t_2),$$

$$\text{sh } 1 = \sinh(K_1 t_1),$$

$$\text{sh } 2 = \sinh(K_2 t_2). \tag{15}$$

V^*, K, and t stand, respectively, for carriers velocities, wave vectors, and thicknesses of dielectrics 1 (SiO_2 interfacial oxide) and 2 (high-κ).

As in Sec. II B 1, Eq. (14) can be further simplified in the case of low energy tunneling, as long as $K_{1,2}t_{1,2} \gg 1$, i.e., $E \ll \Phi_{\text{eff}1,2}$, leading to

$$T = \frac{64 v_{\text{in}} (V_1^*)^2 (V_2^*)^2 v_{\text{out}}}{[v_{\text{in}}^2 + (V_1^*)^2][(V_1^*)^2 + (V_2^*)^2][(V_2^*)^2 + v_{\text{out}}^2]} \times \exp[-2 \cdot (K_1 t_1 + K_2 t_2)], \tag{16}$$

with

$$v_{\text{in}}(E) = \frac{\hbar k_{\text{in}}(E)}{m_{L,T}} = \frac{1}{m_{L,T}} \cdot \sqrt{2 m_{L,T} E},$$

$$V_1^*(E) = \frac{\hbar K_1(E)}{m_{\text{SiO}_2}} = \frac{1}{m_{\text{SiO}_2}} \cdot \sqrt{2 m_{\text{SiO}_2} \left(\phi_{\text{SiO}_2} - \frac{V_{\text{SiO}_2}}{2} - E \right)},$$

$$V_2^*(E) = \frac{\hbar K_2(E)}{m_{\text{High-}K}} = \frac{1}{m_{\text{High-}K}} \cdot \sqrt{2 m_{\text{High-}K} \left(\phi_{\text{High-}K} - V_{\text{SiO}_2} - \frac{V_{\text{High-}K}}{2} - E \right)},$$

$$v_{\text{out}}(E) = \frac{\hbar k_{\text{out}}(E)}{m_0} = \frac{1}{m_0} \cdot \sqrt{2 m_0 (E + V_{\text{SiO}_2} + V_{\text{High-}K})}. \tag{17}$$

Equation (16) can be interpreted in the same way as Eq. (13) in the single barrier case. The pre-exponential factor accounts for reflections at the Si/SiO_2, SiO_2/high-κ and high-κ/gate potential discontinuities and is only valid when the condition $K_{1,2} \cdot t_{1,2} \gg 1$ is satisfied. This expression can be considered as an extension of Register's single-barrier correction factor applied to double square barriers.

Figures 6 and 7 show the transmission probability and respective error versus Airy formalism for the whole

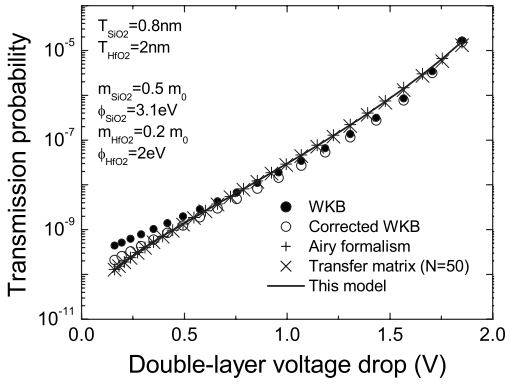

FIG. 6. Transmission probability vs SiO_2–HfO_2 double-layer stack as a function of voltage drop in double-layer system. Five models have been considered: transfer matrices, Airy formalism, square barrier approximation, and WKB method (corrected and uncorrected).

SiO_2–HfO_2 gate stack. SiO_2 and HfO_2 layers parameters have been taken to their standard values, in accordance with data extracted from the literature.[3,23,24]

The application of the standard WKB approximation leads to strong discrepancy with the exact solution (more than 100% in the worst case), even in the low energy range, particularly important to model direct tunneling. In the low energy range, the application of Register's corrective term taking into account the discontinuity at the three interfaces, and this method lead both to very good results compared to the exact solution. At high energy, however, as previously explained, Register's corrective term diverges, while these techniques [Eq. (14)] still lead to acceptable results.

Moreover, it has been suggested by Li *et al.*[13] to apply Register's corrective term to the first discontinuity Si/SiO_2 only, introducing artificial corrections to compensate the previously mentioned divergence at high energy. We found, however, that it is necessary to consider the three discontinuities to achieve a good accuracy at low energy range, and that at high energy, Li's model is less accurate that the results obtained by Eq. (14).

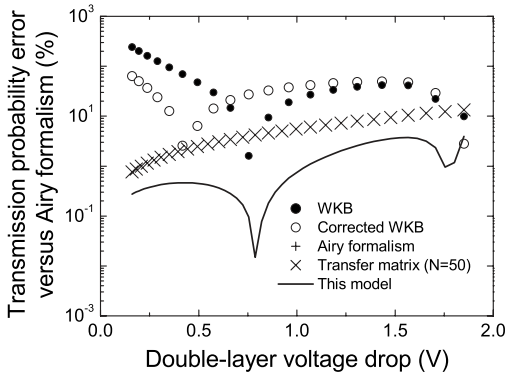

FIG. 7. Relative error of WKB (corrected and uncorrected), square barrier approximation and transfer matrices vs Airy matrices (exact) method. Parameters are the same as in Fig. 6.

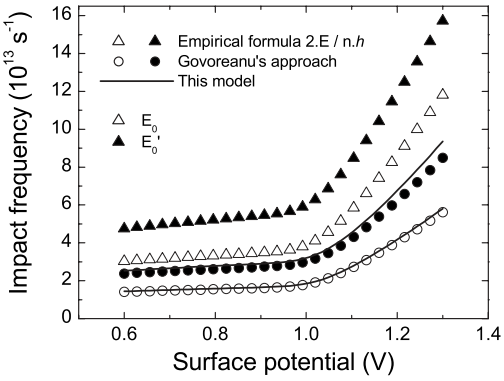

FIG. 8. Incoming carriers impact frequency versus surface potential for the first unprimed and primed energy levels. Three methods are compared: the empirical $2E_n/nh$ formula, and one derived from the uncorrected/corrected triangular well approximation.

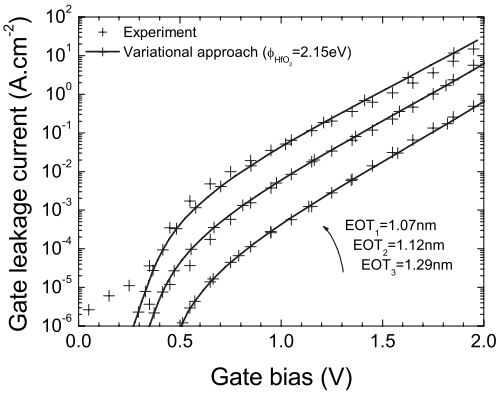

FIG. 9. Experimental and simulated gate current densities vs gate bias for different SiO_2–HfO_2 gate stacks. (See the text for details). Data can be fitted using the variational approach with $\phi_{HfO_2}=2.15$ eV barrier.

3. Impact frequency

Last element to model tunneling current in the transparency formalism, the impact frequency (standing for incoming flux of carrier) is usually taken equal to the phenomenological equation[7,18]

$$f_n^{L,T}=\frac{2E_n^{L,T}}{nh}. \tag{18}$$

This formula is, in fact, correct when carriers are emitted by a square well, which is not the case considered here, as carriers are emitted by an inversion layer. Using the approximated variational approach to model quantum confinement in the inversion layer, a similar formula can be derived, with an additional 2/3 factor.[25]

The impact frequency of a particle confined in a potential well extending from 0 to x_{well} can be computed, using the WKB approximation, by[12,13]

$$f_n^{L,T}=\frac{1}{\int_0^{x_{\text{well}}}\frac{2m_{L,T}}{\hbar k_n^{L,T}(x)}\mathrm{d}x}. \tag{19}$$

This formula has been found in excellent agreement with more exact quantum results in Ref. 26.

In the framework of the triangular well approximation [using Eq. (2)], and as demonstrated by Govoreanu *et al.*,[12] Eq. (19) leads to

$$f_n^{L,T}=\frac{(qF_n^{L,T})^{2/3}}{2\left(2m_{L,T}\cdot\hbar\frac{3\pi}{2}\left(n+\frac{3}{4}\right)\right)^{1/3}}. \tag{20}$$

However, as demonstrated in Sec. II A, the triangular well approximation leads to an erroneous estimation of the different energy levels. To account for the energy levels corrections used in Sec. II A, Eq. (20) has been rewritten as a function of the discrete energies levels, leading to

$$f_n^{L,T}=\frac{qF_n^{L,T}}{2\sqrt{2m_{L,T}E_n^{L,T}}}, \tag{21}$$

where F_n and E_n includes the various improvements introduced in Sec. II A.

The impact frequency of carriers located on first unprimed and primed energy levels versus surface potential is plotted on Fig. 8. While Eq. (18) leads to at least 10% errors, this model and Govoreanu's approach give similar results for the first unprimed subband, where the impact of energy corrective terms to the standard triangular well approximation is negligible. For higher energies levels, however, the impact frequency given by Eq. (21) appears to be more accurate.

III. RESULTS AND DISCUSSION

Model predictions of gate current I_g versus applied gate voltage V_g have been compared with measurements in polysilicon/TiN/HfO_2/SiO_2/p-type silicon MOS transistors in inversion regime. Silicon substrate is ⟨100⟩-oriented and doping level is roughly equal to $N_a=3\times10^{23}$ m^{-3}. A 8 Å SiO_2 interfacial layer has been thermally grown using rapid thermal oxidation (RTO). Different HfO_2 layers, from

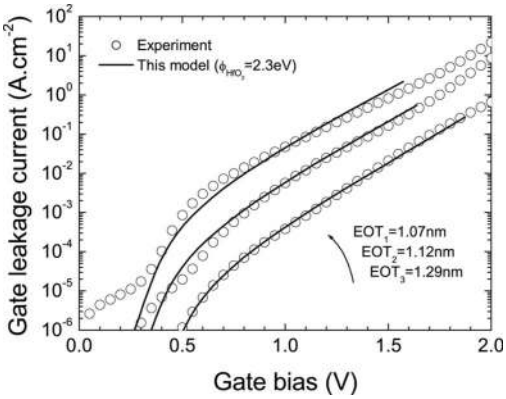

FIG. 10. Experimental and simulated gate current densities vs gate bias for different SiO_2–HfO_2 gate stacks. Data can be fitted using this model, with $\phi_{HfO_2}=2.3$ eV barrier.

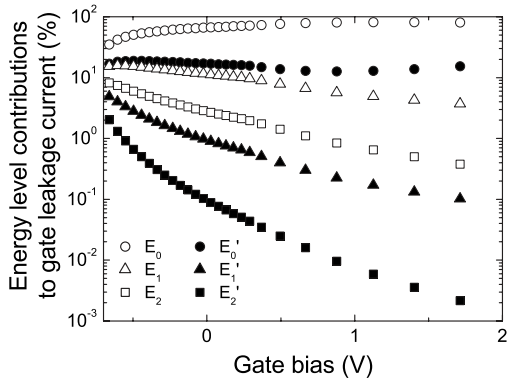

FIG. 11. Contribution of each subband to tunneling current versus gate bias in a pure SiO_2 dielectric. In this case, tunnel current is mainly composed of electrons from the first *unprimed* subband ($m_{SiO_2}=0.5m_0$, $\phi_{SiO_2}=3.1$ eV, $V_{FB}=-0.6$ V, EOT=1.3 nm).

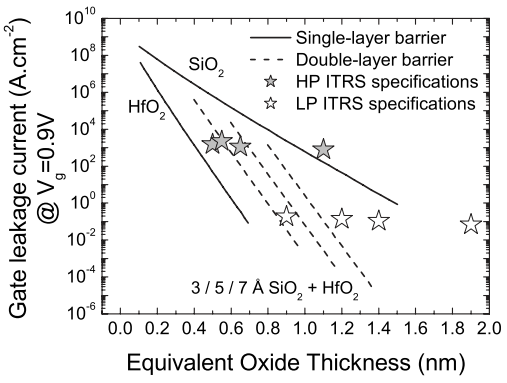

FIG. 13. Theoretical gate leakage current vs EOT at $V_g=0.9$ V [$V_g=0.9$ V being the ITRS specification for 32 nm high performance (HP) node]. Single-layer stacks are compared (SiO_2 or HfO_2) to double-layer stacks with a varying interfacial SiO_2 oxide. ITRS specifications for HP and LP devices are shown (65, 45, 32, and 22 nm nodes). Material parameters used in this plot are the same as in Fig. 10.

2 to 4.5 nm, have been deposited using atomic layer deposition (ALD), with a final 600 °C anneal. Finally, a 10 nm chemical vapor deposition (CVD) TiN metal gate has been processed.

This model has been compared to experiments using the following procedure: equivalent oxide thickness (EOT), *p*-substrate doping level (N_a), and flatband voltage (V_{FB}) have been extracted from *C-V* measurements and the following parameters: $m_{SiO_2}=0.5m_0$, $m_{HfO_2}=0.18m_0$, $\varepsilon_{HfO_2}=20\varepsilon_0$, and $\phi_{SiO_2}=3.1$ eV, have been assumed equal to their standard values.[3] Finally, the high-κ thickness t_{HfO2} and barrier height ϕ_{HfO2} have been extracted by fitting the *I-V* curves. In the low-voltage range (typically 0–0.4 V), direct tunneling model fails to reproduce experiments. Accounting for trapping mechanisms in Hf-based gate stacks (out of the scope of this work) would allow to reliably fit the whole *I-V* curve.

The extracted value of ϕ_{HfO_2} has been found to depend significantly on the model. As an example, experimental data have been fitted using the variational approach (which only account for one quantized level) and this model (which account for all the relevant subbands). In consequence, even if both model are able to nicely reproduce the experimental *I-V* curve, a lower ϕ_{HfO_2} barrier has been found in the variational approach case ($\phi_{HfO_2}=2.15$ eV) (Fig. 9),, than in this approach ($\phi_{HfO_2}=2.3$ eV) (Fig. 10). This result can be explained by the particular role of tunneling of carriers emitted by the first primed subband. Indeed, contrary to the case of tunneling in pure SiO_2 (Fig. 11) where current is mainly composed of carriers from the first unprimed subband,[27] in the case of SiO_2/HfO_2 gate stacks, however, primed subbands also contributes significantly to the total current (Fig. 12). This fact justifies the use of a multisubband model and explains the different extracted values of barrier height ϕ_{HfO_2} using this model or a less accurate single subband model.

Finally, prediction of gate leakage currents in HfO_2/SiO_2 stacks using this model have been compared with ITRS requirements[1] (Fig. 13). Using the parameters extracted on experiments, 3 Å of SiO_2 interfacial layer seems to be the minimum thickness acceptable to fulfill HP ITRS requirements for 22 nm node.

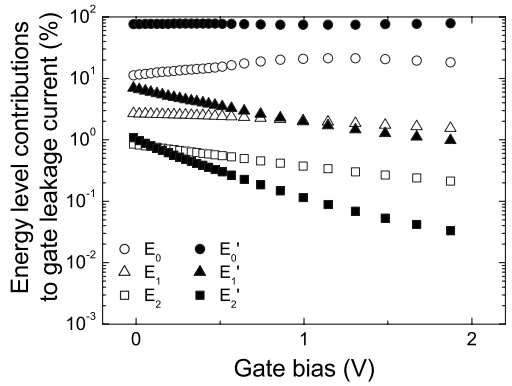

FIG. 12. Contribution of each subband to tunneling current vs gate bias in a SiO_2/HfO_2 stack. Tunnel current is mainly composed of electrons from the first *primed* subband. ($m_{SiO_2}=0.5m_0$, $m_{HfO_2}=0.18m_0$, $\phi_{SiO_2}=3.1$ eV, $\phi_{HfO_2}=2.3$ eV, $V_{FB}=-0.6$ V, EOT=1.3 nm).

IV. CONCLUSIONS

An original and efficient analytical model of gate leakage current through HfO_2/SiO_2 gate stacks has been presented, as well as example of application, to model experimental data, or estimate the evolution of gate leakage current along scaling.

An accurate way to compute quantum confinement at silicon interface has been used, allowing a much faster resolution than the exact self-consistent Poisson–Schrödinger resolution. This method, physically based, allows to consider multiple subbands and valleys. It has been shown that this last point is of particular importance in the case of SiO_2/HfO_2 gate stacks, as, contrary to conventional SiO_2 tunneling, not only the lowest subband contributes to the total tunnel current.

Moreover, an original approach has been proposed to calculate transmission probability across a double-layer system, accounting for reflections at dielectric interfaces. This method has been successfully validated by comparison with

more complex transfer matrix and Airy function formalisms. Finally, this techniques has be found to be an attractive alternative to Register's corrective term method, leading to similar accurate results in the low energy range, without diverging at high energy.

[1]ITRS Roadmap, 2007 Edition, www.itrs.net.
[2]K. Mistry *et al.*, Tech. Dig. - Int. Electron Devices Meet. 247 (2007).
[3]P. Palestri *et al.*, IEEE Trans. Electron Devices **54**, 106 (2007).
[4]V. Nam Do and P. Dollfus, J. Appl. Phys. **101**, 073709 (2007).
[5]F. Stern, Phys. Rev. B **5**, 4891 (1972).
[6]R. Clerc, A. Spinelli, G. Ghibaudo, and G. Pananakakis, J. Appl. Phys. **91**, 1400 (2002).
[7]L. F. Register, E. Rosenbaum, and K. Yang, Appl. Phys. Lett. **74**, 457 (1999).
[8]S. H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, IEEE Electron Device Lett. **18**, 209 (1997).
[9]R. Clerc, G. Ghibaudo, and G. Pananakakis, Solid-State Electron. **46**, 407 (2002).
[10]J. Cai and C. T. Sah, J. Appl. Phys. **89**, 2272 (2001).
[11]J. Wang, Y. Ma, L. Tian, and Z. Li, Appl. Phys. Lett. **79**, 1831 (2001).
[12]B. Govoreanu, P. Blomme, K. Henson, J. Van Houdt, and K. De Meyer, Solid-State Electron. **48**, 617 (2004).
[13]F. Li, S. P. Mudanai, Y. Y. Fan, L. F. Register, and S. K. Banerjee, IEEE Trans. Electron Devices **53**, 1096 (2006).
[14]K. F. Brennan and C. J. Summers, J. Appl. Phys. **61**, 614 (1986).
[15]Y. Ando and T. Itoh, J. Appl. Phys. **61**, 1497 (1987).
[16]M. O. Vassell, J. Lee, and H. F. Lockwood, J. Appl. Phys. **54**, 5206 (1983).
[17]S. S. Allen and S. L. Richardson, J. Appl. Phys. **79**, 886 (1996).
[18]N. Yang, W. K. Henson, J. R. Hauser, and J. J. Wortman, IEEE Trans. Electron Devices **46**, 1464 (1999).
[19]T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[20]M. Ferrier, R. Clerc, G. Ghibaudo, F. Boeuf, and T. Skotnicki, Solid-State Electron. **50**, 69 (2006).
[21]*Handbook of Mathematical Functions*, edited by M. Abramowitz and A. Stegun (NBS, Washington DC, 1964).
[22]J. G. Simmons, J. Appl. Phys. **34**, 2581 (1963).
[23]J. Buckley, B. De Salvo, G. Molas, M. Gely, and S. Deleonibus, ESSDERC Proceedings, 2005.
[24]Y. T. Hou, M. F. Li, H. Y. Yu, and D. L. Kwong, IEEE Electron Device Lett. **24**, 96 (2003).
[25]R. Clerc, B. De Salvo, G. Ghibaudo, G. Reimbold, and G. Pananakakis, Solid-State Electron. **46**, 407 (2002).
[26]A. Dalla Serra, A. Abramo, P. Palestri, L. Selmi, and F. Widdershoven, IEEE Trans. Electron Devices **48**, 1811 (2001).
[27]Z. A. Weinberg, Solid-State Electron. **20**, 11 (1976).