A Physically Based Compact Gate C–V Model for Ultrathin (EOT $\sim$ 1 nm and Below) Gate Dielectric MOS Devices

Fei Li, *Member, IEEE*, Sivakumar Mudanai, *Member, IEEE*, Leonard Franklin Register, *Senior Member, IEEE*, and Sanjay K. Banerjee, *Fellow, IEEE*

***Abstract*—A computationally efficient and accurate physically based gate capacitance model of MOS devices with advanced ultrathin equivalent oxide thickness (EOT) oxides (down to 0.5 nm explicitly considered here) is introduced for the current and near future integrated circuit technology nodes. In such a thin gate dielectric regime, the modeling of quantum–mechanical (QM) effects simply with the assumption of an infinite triangular quantum well at the Si-dielectric interface can result in unacceptable underestimates of calculated gate capacitance. With the aid of self-consistent numerical Schrödinger–Poisson calculations, the QM effects have been reconsidered in this model. The 2/3 power law for the lowest quantized energy level versus field relations $\left(E_1 \propto F_{\mathrm{ox}}^{2/3}\right)$, often used in compact models, was refined to 0.61 for electrons and 0.64 for holes, respectively, in the substrate in the regimes of moderate to strong inversion and accumulation to address primarily barrier penetration. The filling of excited states consistent with Fermi statistics has been addressed. The quantum-corrected gate capacitance–voltage $(C$–$V)$ calculations have then been tied directly to the Fermi level shift as per the definition of voltage (rather than, for example, obtained indirectly through calculation of quantum corrections to the charge centroids offset from the interface). The model was implemented and tested by comparisons to both numerical calculations down to 0.5 nm, and to experimental data from n-MOS or p-MOS metal-gate devices with SiO_2, Si_3N_4 and high-κ (e.g., HfO_2) gate dielectrics on (100) Si with EOTs down to ~ 1.3 nm. The compact model has also been adapted to address interface states, and poly depletion and poly accumulation effects on gate capacitance.**

***Index Terms*—Compact gate capacitance–voltage $(C$–$V)$ modeling, Fermi–Dirac statistics, interface states, quantum–mechanical (QM) effects, wave function penetration.**

I. Introduction

As THE aggressive scaling of CMOS technology continues, the equivalent physical oxide thickness (EOT) of gate dielectrics is projected by the International Technology Roadmap for Semiconductors (ITRS 2003 Edition) to become as thin as 0.5 nm in some cases in the next 15 years [1]. To overcome the well-known increasingly serious technology difficulties of large gate leakage current and polysilicon related shortcomings (polydepletion, B penetration, etc.), alternative high dielectric constant (high-κ) materials and metal-gate technology are being considered as replacements for SiO_2 and polysilicon gates, respectively [1]. However, conventional electrical analysis techniques for oxide and MOS devices such as gate capacitance–voltage $(C$–$V)$ and gate current–voltage $(I$–$V)$ measurements remain important in characterization of gate dielectrics and MOS devices to provide critical information such as dielectric EOT, oxide charge, metal-gate work function, surface doping density, and interface states.

Interpretation of C–V data, the subject of this paper, becomes increasingly challenging with scaling, however, as the nonzero thicknesses of charge accumulation and strong inversion layers becomes more significant, reducing the capacitance relative to that otherwise expected. This problem is exacerbated by quantum confinement and Pauli exclusion effects that tend to further push carriers away from the dielectric interface and raise the carrier energies relative to what would be expected classically [2], and barrier penetration effects that can mitigate both of the above consequences [3]–[7]. Although such behavior can be somewhat readily addressed in numerical studies on gate stacks e.g., [3], [6], the increasing physical complexity of the problem has made it difficult to create compact models applicable to $\sim$ 1 nm and below EOTs. And while such numerical C–V simulators can provide a physically accurate and comprehensive understanding of these effects, efficient analytic C–V models with similar accuracy are required for practical everyday device and circuit studies.

In this paper, guided by basic physics, analytic theory [7] and self-consistent numerical Schrödinger–Poisson calculations [4]–[6], [8], [9], a new compact model of C–V has been developed for $\sim$ 1 nm and below EOTs. Comparisons to numerical simulations for gate dielectrics down to 0.5-nm EOTs and to experimental data for n-MOS and p-MOS devices with gate dielectrics down to 1.29-nm EOTs, including devices with SiO_2, Si_3N_4 and high-κ (e.g., HfO_2) gate dielectrics, exhibit excellent accuracy and efficiency for this compact model for metal-gate devices on (100) Si. Computation times for C–V curves are around a few 10 s of milliseconds on personal computers. Although not the emphasis of this paper, for completeness, polydepletion and polyaccumulation effects on gate capacitance were also added to this new compact model and tested by comparison to numerical simulations and experimental data. Surface defect states were also added to the model and tested by comparison to numerical simulation. Finally, while this paper on compact

Manuscript received October 7, 2004; revised February 25, 2005. This work was supported in part by the Semiconductor Research Corporation and in part by the Advanced Materials Research Center under Sematech. The review of this paper was arranged by Editor C. McAndrew.

F. Li is with the Nassda Corporation, Santa Clara, CA 95035 USA.

S. Mudanai is with the Intel Corporation, Hillsboro, OR 97124 USA.

L. F. Register and S. K. Banerjee are with the Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas, Austin, TX 78758 USA.

Digital Object Identifier 10.1109/TED.2005.848079

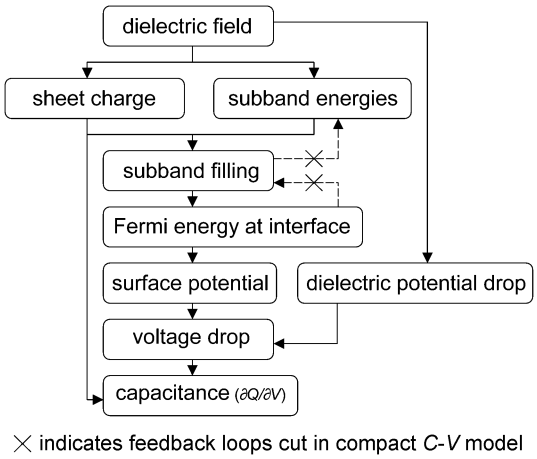


Fig. 1. Flowchart for the compact model (and numerical calculations).

C–V modeling is complete in itself, we point out that this approach provides a foundation for self-consistent compact calculation of gate tunneling currents $(I$–$V)$ using the same subband filling (tunneling charge source) subband energies and overall barrier shape (surface impact frequencies and tunneling probabilities), as to be addressed in planned work to follow.

II. Description of the Model for Metal-Gate Devices

A. Algorithm for Gate Capacitance Calculation in Accumulation and Strong Inversion

A flowchart that summarizes the algorithm used in this compact model in strong inversion or accumulation in the substrate is shown in Fig. 1. First, the subband energies and the total sheet charge on the capacitor in the substrate are determined as function of the dielectric field. Subband filling within the energy bands subject to Fermi–Dirac statistics then determines the Fermi energy and, thus, by definition, the voltage $(V = -E_F/q)$. Finally, the quasi-static capacitance is obtained, also as per definition, by comparing changes in the sheet charge to changes in the voltage $(C = dQ/dV)$. This algorithm, itself, is thus firmly grounded in first-principles, and no explicit or implicit approximations have been made. Indeed, not coincidentally, this flow chart is essentially the same as that for the numerical simulator used in the development of this model. However, to apply this algorithm within the framework of a compact model requires that the subband energies and relative distributions of carriers among the subbands be quickly obtainable with little or none of the feedback illustrated in Fig. 1 that is used to obtain self-consistency in the computationally intensive numerical simulations.

B. Subband Energies in Finite Quasi-Triangular Wells

Stern [7] showed that if the channel potential could be approximated as an infinite triangular well—the substrate dielectric approximated as an infinite potential barrier and the slope of the potential in the substrate defined by the electric field at the interface, F_s as in Fig. 2(a)—then the magnitudes of the bound-state eigenenergies relative to the band edge could be obtained as

$$E_i \approx \left(\frac{\hbar^2}{2m^*}\right)^{1/3}\left[\frac{3}{2}\pi e F_s\left(i+\frac{3}{4}\right)\right]^{2/3} \quad (1)$$

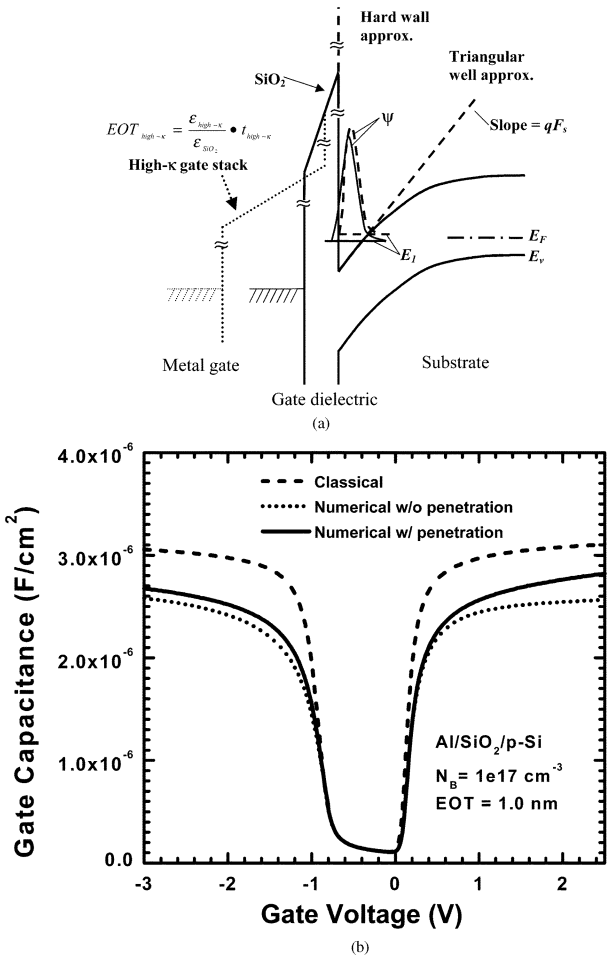


Fig. 2. (a) Schematic diagram for an MOS structure showing the potential profile used in self-consistent numerical calculations (solid lines) and the hard-wall triangular well approximation (dashed lines) often employed for analytic calculations of subband energies, along with the differences in energies and probability distributions for the groundstates. High-κ dielectrics (dotted line to the left) have larger physical thickness for the same equivalent oxide thickness (EOT). Here $t_{high-\kappa}$ is the physical thickness of gate dielectrics. The carrier centroid is shifted closer to interface and the quantization energy is lower due to wavefunction penetration. (b) The numerically calculated MOS C–Vs for a device with 1 nm EOT within a classical approximation (particles and Boltzmann statistics), calculated quantum mechanically within a hard wall approximation to the dielectric-Si interface, and calculated quantum mechanically with wave function penetration into the dielectric.

using Airy functions with negligible difference as compared to numerical calculations. Here, m^* is the quantization effective mass and i labels the eigenstate with $i = 0$ corresponding to the groundstate. With this 2/3 power law van Dort *et al.* [10], [11] developed the first compact model with QM effects for MOS devices and estimated the threshold voltage shift due to QM effects efficiently. CVC, a popular compact C–V simulator developed later by Hauser *et al.* [12], [13] that addresses QM effects and has been verified for polysilicon MOS capacitors with oxides about 2 nm or thicker, also makes use of this 2/3 power law relation to predict the lowest quantized energy level, although the quantization effective mass in the lead coefficient was used as a fitting parameter to deal with nonidealities in the true system. (Notably, this information was only used to effectively shift the band-edge in CVC; the centroid shift of quantum-confined carriers is fixed to be 1.2 nm from [14]. In contrast, the centroid

shift is not explicitly required here, but its nonconstant nature is apparent in the results to follow.)

However, as revealed by numerical calculations such as those of Fig. 2(b) and as noted in the introduction, barrier penetration effects in real devices can partially mitigate the effects of quantum confinement eventually leading to unacceptable underestimates of gate capacitance [3], [6], [15]–[17], as well as gate current [18], in ultrathin EOT devices. Unfortunately for the system of Fig. 2(a) with barrier penetration and variations in the potential function from triangular as the field falls off through the charge layer itself, no derived explicit analytic expression such as in (1) is available. However, with the aid of self-consistent Schrödinger-Poisson numerical simulation we have found that the power law *form* of (1) nevertheless remains reliable over the regions of interest; only the parameters require refinement. That is, to an excellent approximation even after allowing for barrier penetration and any deviations from triangular wells, the subband energies relative to the band edges at the Si-dielectric interface can be written in the form,

$$E_i - E_{c,v} \cong \pm\gamma_i \left(\frac{|F_{\rm ox}|\,{\rm cm}}{\rm MV} \right)^{\lambda_i} \tag{2}$$

for electrons and holes, respectively, where E_c and E_v are the (nominal) conduction and valence band edges at the Si-dielectric interface, respectively, i again labels the subband, and $F_{\rm ox}$ is the effective oxide field. Note that energy signs used here correspond to the electron energies even for "hole" subbands. Fig. 3(a) and (b) show the groundstate subband energies versus the oxide field $F_{\rm ox}$ for electrons in strong inversion and holes in accumulation for n-channel MOS capacitors with a SiO_2 dielectric and substrate doping levels varying from 10^{16} to 10^{18} cm^{-3}, along with curve fits of the form of (2) which show up as straight lines on these log-log scales. Via these simulations and similar calculations for p-MOS devices we obtained $\gamma_1 = 77$ meV and $\lambda_1 = 0.61$ for electrons in both inversion (n-MOS) and accumulation (p-MOS), and $\gamma_1 = 88$ meV and $\lambda_1 = 0.64$ for holes in both inversion (p-MOS) and accumulation (n-MOS) essentially independent of doping in all cases. These seemingly small deviations of the exponents from 2/3 due to the nonidealities are, nevertheless, quite significant and have been noted for n-MOS devices previously [16]. Furthermore, unlike for calculation of tunneling currents, only the near-surface dielectric properties are of importance for determining quantum confinement and barrier penetration effects. Therefore, γ and λ parameter values obtained for SiO_2 should remain reliable when high-κ gate stacks are employed with, as typical, a thin SiO_2 layer incorporated adjacent to the interface [19].

Of course when the interfacial layer properties vary or for high-κ directly on substrate such as Ge channel devices [20], the degree of barrier penetration might change. Indeed different models of the oxide and, perhaps, the hole bandstructure in Si could lead to slightly altered expectations of penetration for SiO_2. However, as the power law form remains reliable here for both electrons and holes with different degrees of barrier penetration (and, of course, in the limit of no barrier penetration where the exponent is 2/3) it therefore seems likely that the power-law form would remain reliable for different surface dielectric materials; only the parameters γ and λ would be expected to vary. In short, although first-principles-based analytically derived explicit mathematical expression are not yet available to address the nonidealities of finite quasi-triangular potentials, it appears that the nonidealities of real devices can be reliably addressed via only small modifications to the first-principles-based analytically derived explicit mathematical expression for the idealized limiting case of infinite triangular well, [7].

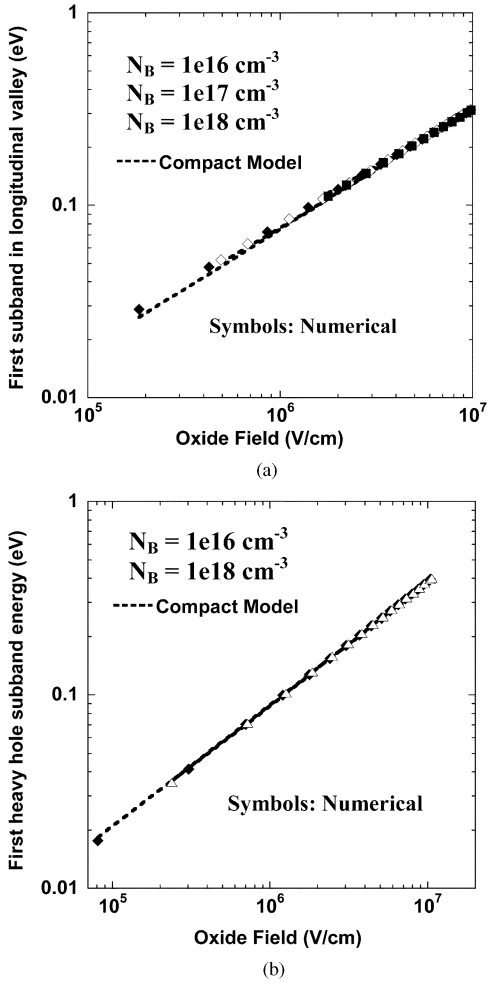


Fig. 3. Lowest subband energy for (a) electrons in inversion and (b) holes in accumulation, showing dependencies of $F_{\rm ox}^{0.61}$ and $F_{\rm ox}^{0.64}$, respectively, and negligible dependence on doping. In this paper, the barrier heights in the numerical simulations are 3.15 eV for electrons and 4.8 eV for SiO_2, respectively, and the oxide effective masses were taken to be 0.53 times the electron rest mass.

Finally, note that the higher electron and hole subband minimum energies have close to the same power law dependencies with the oxide field as the groundstate subbands although with more noticeable doping dependencies and of course higher lead coefficients, as shown in Fig. 4. Although this information is not explicitly required in this paper, it would be pertinent for, e.g., tunneling leakage current calculations.

C. Subband Filling in Accumulation and Strong Inversion

In strong inversion or accumulation of the channel, once the eigenenergies have been calculated, the Fermi level E_F is established by the charge occupation of these levels. The total field-induced mobile surface charge density is approximated by

$$Q_s(F_{\rm ox}) = -\varepsilon_{\rm ox} F_{\rm ox} \tag{3}$$

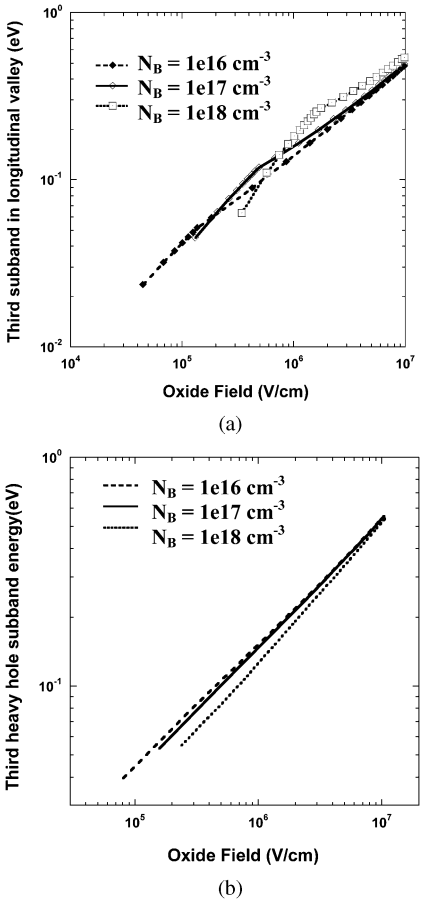


Fig. 4. Third subband energy of (a) the longitudinal conduction band valley and (b) the heavy-hole band showing simple power-law field dependencies and moderate doping dependencies.

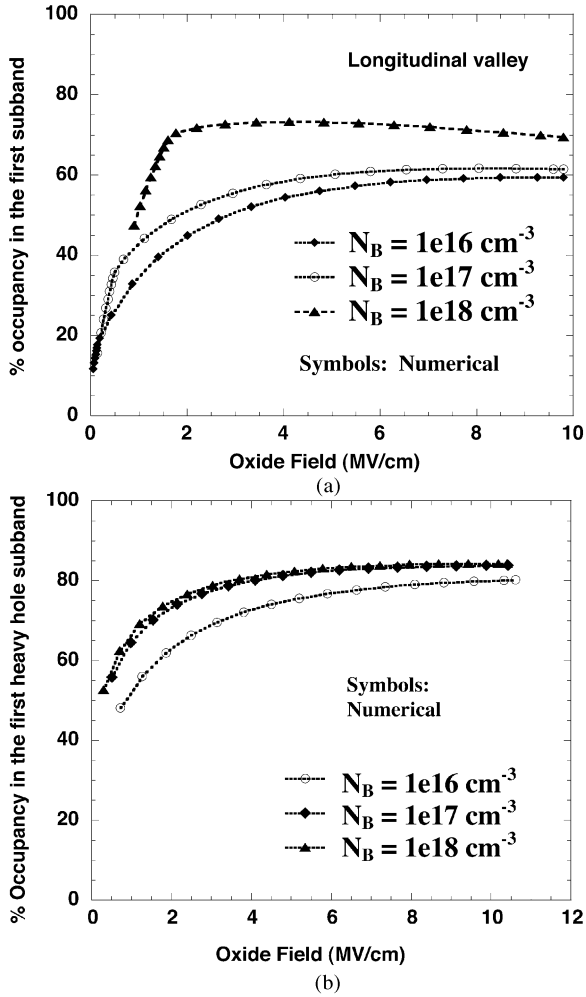


Fig. 5. Percentage occupancy of (a) electrons in the lowest subband (in the valleys with longitudinal mass normal to the interface) and (b) holes in the lowest (heavy-hole) subband. Occupancies are roughly constant at moderate to high fields and show weak doping dependence.

in accumulation, and

$$Q_s(F_{\rm ox}) = -\varepsilon_{\rm ox}F_{\rm ox} - Q_{\rm depletion}(F_{\rm ox}) \tag{4}$$

in strong inversion, where $\epsilon_{\rm ox}$ is the dielectric constant of the oxide and $Q_{depletion}$ is the depletion region charge as a function of the oxide field as discussed below. Applying Fermi–Dirac statistics within an effective mass approximation such that there is a constant density of states in each subband, we get [21]

$$\mp Q_s(F_{\rm ox}) = \sum_i \frac{\eta_i m_i k_B T}{\pi\hbar^2} \ln\left(1 + \exp\frac{\pm[E_F - E_i(F_{\rm ox})]}{k_B T}\right) \tag{5}$$

for electrons and holes, respectively, (where all energies are given for electrons even for valence band states), η_i is the degeneracy of the ith eigenenergy level, m_i is the two-dimensional interface-parallel density of states effective mass which depends on which valley the ith eigenenergy level is in, and E_F is the Fermi level. For temperature $T > 0$, (5) cannot be solved explicitly for E_F if more than one valley is considered.

To explore the possibility of using just a single eigenenergy level to calculate the Fermi-level without neglecting the charges in the excited subbands, the occupancy of the groundstate subband was studied as a function of the oxide field via numerical simulations. As shown in Fig. 5(a) electrons in inversion and Fig. 5(b) for holes in accumulation, the percentage occupancy was a roughly constant fraction of the total induced charge for moderate to high fields, and was only weakly dependent on doping. This rough saturation of the ground-state occupation probability is the consequence of two opposing effects: the increase in the subband splitting and the increasing degenerate nature of the carrier distribution. The increasing separation between the ground and excited state subbands with increasing field would result in a continuing increase in not only the absolute but also the *relative* occupation of the ground state subband if Boltzmann statistics were applicable. However, the carrier distribution quickly becomes degenerate in strong inversion and Fermi statistics must be used. As a result, the lower the subband energy is, the less *relative* change in the subband charge there is with changes in $E_F - E_i$, particularly with the Fermi-level E_F above the bottom of the subband E_i. At low fields in Fig. 5(a) where the occupation percentage varies abruptly the device is in depletion/weak inversion, the subband occupancies are negligible, and the Fermi level shift/surface potential is handled in a different manner to be described in Section II-E. At low fields in Fig. 5(b), the device is in accumulation but near flatband and nonfield-induced charge, charge that is already present at flatband raises the Fermi level increasing the occupancy of higher eigenenergy levels, and again a different approach is used as described in Section II-G.

Assuming a known occupancy factor for the groundstate subband κ_1 (5) reduces to

$$\mp\kappa_1 Q_s(F_{\rm ox}) = \frac{\eta_1 m_1 k_B T}{\pi\hbar^2}\ln\left(1+\exp\frac{\pm[E_F - E_1(F_{\rm ox})]}{k_B T}\right). \quad (6)$$

This latter equation can be solved explicitly for $E_F - E_{c,v}(F_{\rm ox})$ in terms of $E_1(F_{\rm ox}) - E_{c,v}(F_{\rm ox})$ [from (1)] and $Q(F_{\rm ox})$ [from (3) or (4) and (8) to follow]

$$E_F - E_{c,v}(F_{\rm ox}) = E_1(F_{\rm ox}) - E_{c,v}(F_{\rm ox}) \pm k_B T\ln\left[\exp\left(\frac{\kappa_1|Q_s(F_{\rm ox})|\pi\hbar^2}{\eta_1 m_1 k_B T}\right)-1\right] \quad (7)$$

for electrons ($\eta_1 = 4$ including spin degeneracy) and holes ($\eta_1 = 2$), respectively, for the compact model (and where the reader is reminded that it is not the Fermi level that is a function of field in the assumed grounded substrate but the band edges). The constant values of κ_1 used in this paper are 0.65 for electrons in inversion (n-MOS) and 0.57 in accumulation (p-MOS), and 0.86 for holes in inversion (p-MOS) and 0.83 in accumulation (n-MOS). Note that these occupation probabilities are close to those seen long ago by Stern [22]. Note also that the occupancy factor κ_1 actually could be modeled as a function (if empirical) of the oxide field and doping for better accuracy while still allowing for explicit calculation of E_F, but to date there has been no apparent need. A change of about $\pm 10\%$ in κ_1 was found to result in a negligible error in calculated gate capacitances even for devices with dielectric thickness of 0.5 nm EOT, as shown C–V simulation to follow. The values of m_1 used in this paper are $0.19m_e$ for electrons (transverse mass) and $0.49m_e$ for the heavy holes, as in [23].

The total band-bending in the substrate, the surface potential, is then given by

$$q\varphi_{\rm sub}(F_{\rm ox}) = [E_F - E_{c,v}(F_{\rm ox})] - [E_F - E_{c,v}(0)] \quad (8)$$

where $E_F - E_{c,v}(0)$ is the relative position of the Fermi-level at the interface at flatband.

D. Depletion Layer Charge in Strong Inversion

Application of (4) above requires knowledge of the depletion layer charge. For this purpose and this purpose only, in strong inversion, the potential energy drop across the depletion region (but not including the strong inversion layer) is approximated as

$$q\psi_{d,si}(F_{\rm ox}) = \pm 2q\phi_F + [E_1(F_{\rm ox}) - E_{c,v}(F_{\rm ox})] \quad (9)$$

where $q\phi_F$ is, as usual, the magnitude of the separation between the Fermi level and the intrinsic Fermi level at flatband (and the expression in "[]"s is negative for holes). The addition of E_1 approximates the effective increase in magnitude of the band edge due to quantum confinement. The required depletion layer charge in strong inversion required for (4) is then simply calculated as [23]

$$\mp Q_{\rm depletion}(F_{\rm ox}) \cong \sqrt{2q|\psi_{d,si}(F_{\rm ox})|\varepsilon_s N_{\rm dop}} \quad (10)$$

for electrons and holes, respectively, where ε_s is the dielectric constant in Si. Note that, through E_1, the depletion layer potential drop and associated depletion layer charge are not modeled as a constant under strong inversion conditions, in contrast to other works e.g., [25].

Although not modeled here, the insensitivity of the Fermi level of (6) to the doping concentration suggests that more complicated doping profiles (e.g., retrograde) can be considered without modification to the accumulation and strong inversion layer models. Only the model for the depletion region would have to be refined.

E. Band Bending in Depletion and Weak Inversion

In weak inversion, in the simplest approach, the surface charge layer is neglected along with the associated challenges considered above. In this regime, our model becomes quite conventional; the total band bending in the substrate is calculated via [23]

$$q\varphi_{\rm sub}(F_{\rm ox}) = \pm\frac{\varepsilon_{\rm ox}^2 F_{\rm ox}^2}{2q\varepsilon_s N_{\rm dop}} \quad (11)$$

for n-MOS and p-MOS, respectively. Alternatively, we can take account of weak inversion layer charge iteratively but the effect is only significant *immediately* before transition to strong inversion, as described in Section II-G.

F. Gate Voltage, Charge, and Capacitance

For metal gates, once the $\varphi_{\rm sub}(F_{\rm ox})$ is known in accumulation, depletion and weak inversion, or strong inversion, the gate voltage, $V_G = -[E_{F,\rm Gate} - E_F]/q$, is given by

$$V_G = V_{\rm FB} + F_{\rm ox}t_{\rm ox} + \varphi_{\rm sub}(F_{\rm ox}) \quad (12)$$

where $t_{\rm ox}$ is the equivalent oxide thickness, $F_{\rm ox}$ is the effective oxide field, and $\varphi_{\rm sub}(F_{\rm ox})$ is the total band bending/surface potential in the substrate. As per the algorithm of Fig. 1, in accumulation or strong inversion $\varphi_{\rm sub}(F_{\rm ox})$ is obtained from, sequentially: (2), (9), and (10) for strong inversion only; (3) or (4) for accumulation or strong inversion, respectively; (7); and, finally, (8). In depletion and weak inversion $\varphi_{\rm sub}(F_{\rm ox})$ is obtained simply from (11). Again as per Fig. 1, the total induced charge on the gate (equals negative of total induced charge in the substrate) is given simply by

$$Q_{\rm tot} = \varepsilon_{\rm ox}F_{\rm ox}. \quad (13)$$

The quasi-static capacitance is then obtained using (12) and (13) from the definition of capacitance.

$$C = \frac{\partial Q_{tot}}{\partial V_G}. \quad (14)$$

G. Transitions Between Weak and Strong Inversion, and Accumulation and Depletion

To determine whether the device is in strong or weak inversion, the band-bending is simply calculated under both assumptions and the smallest result is kept. That is, if the magnitude of the result of (11) is smaller than that of (9) the device is deemed to be in depletion or weak inversion and the result of (11) for the band-bending is used; if the magnitude of (9) is smaller, the device is deemed to be in strong inversion and the result of (9) is used. In this way the band bending across the transition between the two regimes is at least continuous if not entirely smooth. The transition can, however, be smoothed if desired using a simple iterative method just around the transition: The Fermi level is calculated assuming no inversion layer

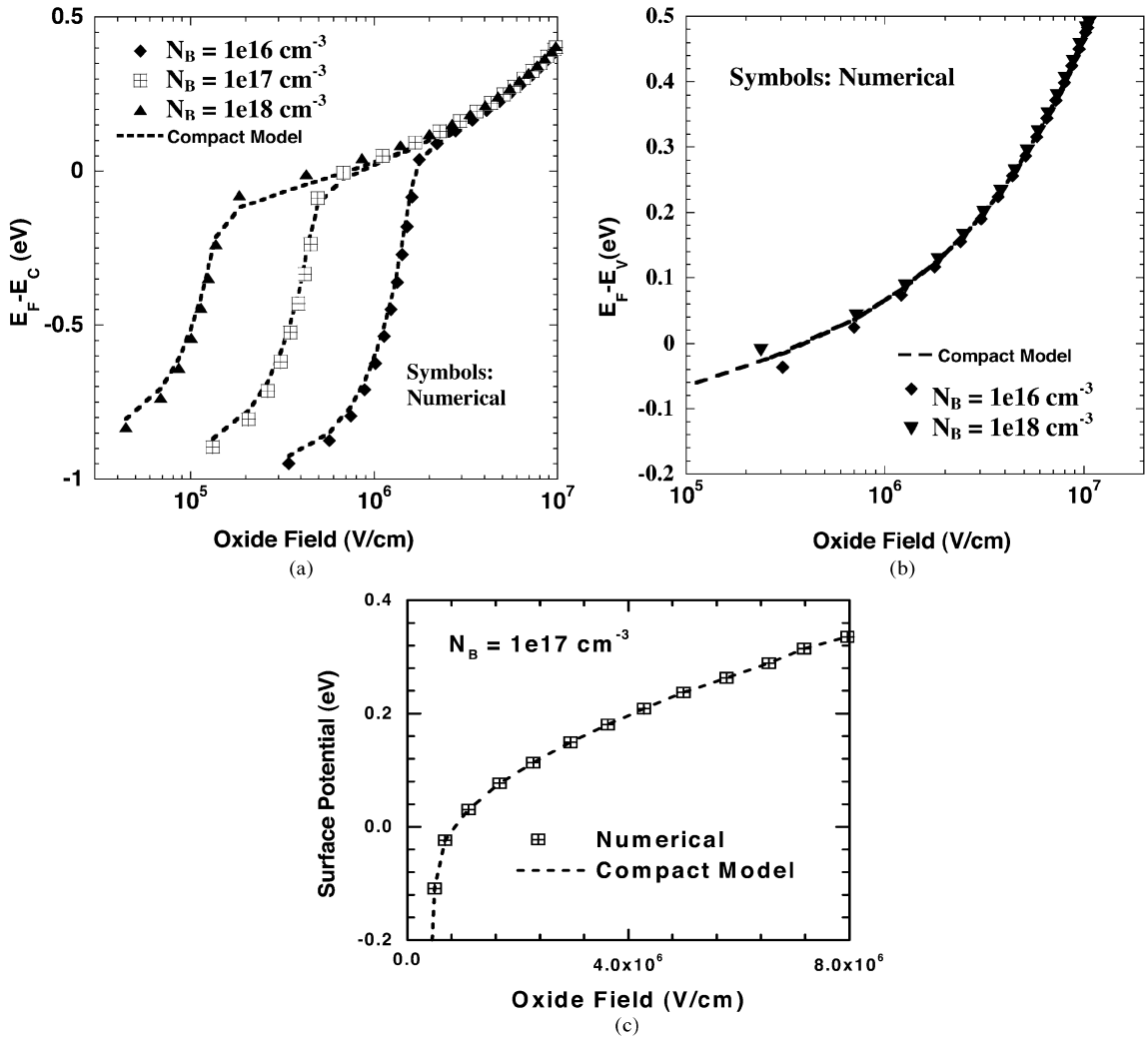


Fig. 6. Comparison of the Fermi levels at the interface of dielectric-Si in an n-MOS device obtained from the compact model and obtained from self-consistent numerical calculations with wave function penetration into the dielectric (a) relative to the conduction band edge at the interface under depletion or weak inversion and strong inversion conditions, and (b) relative to the valence band edge at the interface under accumulation conditions. (c) is a plot on a linear-linear scale of the data of (a) for the device with doping of 10^{17} cm^{-3} in strong inversion.

charge as before. The weak inversion layer charge Q_s is then calculated from (6) based on the estimated Fermi level in weak inversion. The depletion layer charge in then calculated from (11) with the "oxide" field reduced by the weak inversion layer charge according to $F_{\rm ox} \rightarrow F_{\rm ox}+Q_s/\varepsilon_{\rm ox}$ (remembering that for electrons Q_s is negative). The Fermi level is then recalculated and the process repeated to convergence. Because only a few iterations (less than five typically) of very simple equations are required and only very near the transition from strong to weak inversion, the additional computational cost is negligible. However, the effects of this iterative smoothing also are only visually apparent as the "corner" is rounded from weak inversion to strong inversion about the capacitance minimum and then only when the field is highly resolved.

The transition from accumulation to weak inversion is handled even more simply. The approaches of Sections II-C and C above are used only outside of typically ± 0.1 V about flatband. Then the C–V curves are simply connected using a Hermite cubic interpolation method [24]. This is a practical method that both seems to work and to avoid detailed treatment of physics that is still more complicated than that considered above: Below this field magnitude in accumulation, the approximations of Section II-C become questionable as already noted therein, and below this field magnitude in depletion the model of Section II-E becomes unreliable as it ignores the classical Debye length and additional contributions from quantum mechanical repulsion of unbound carriers from the dielectric.

III. Compact Gate C–V Modeling Results for Metal-Gate Devices

Fig. 6(a) and (b) provides comparisons between the compact model and numerical simulations for the Fermi level relative to the band edge at the Si–SiO_2 interface as a function of field in (a) weak and strong inversion and (b) accumulation for an n-MOS capacitor. The accuracy of the results is evident, along with the insensitivity of the results to doping in strong inversion and, except near flatband, in accumulation. The continued increase in the relative Fermi level position in strong inversion and in accumulation leads to the decrease in capacitance $C = \partial Q_{tot}/\partial V_G$ relative that for a true surface charge. This increase is *sub*linear in the substrate in accumulation and strong inversion [Fig. 6(c)],

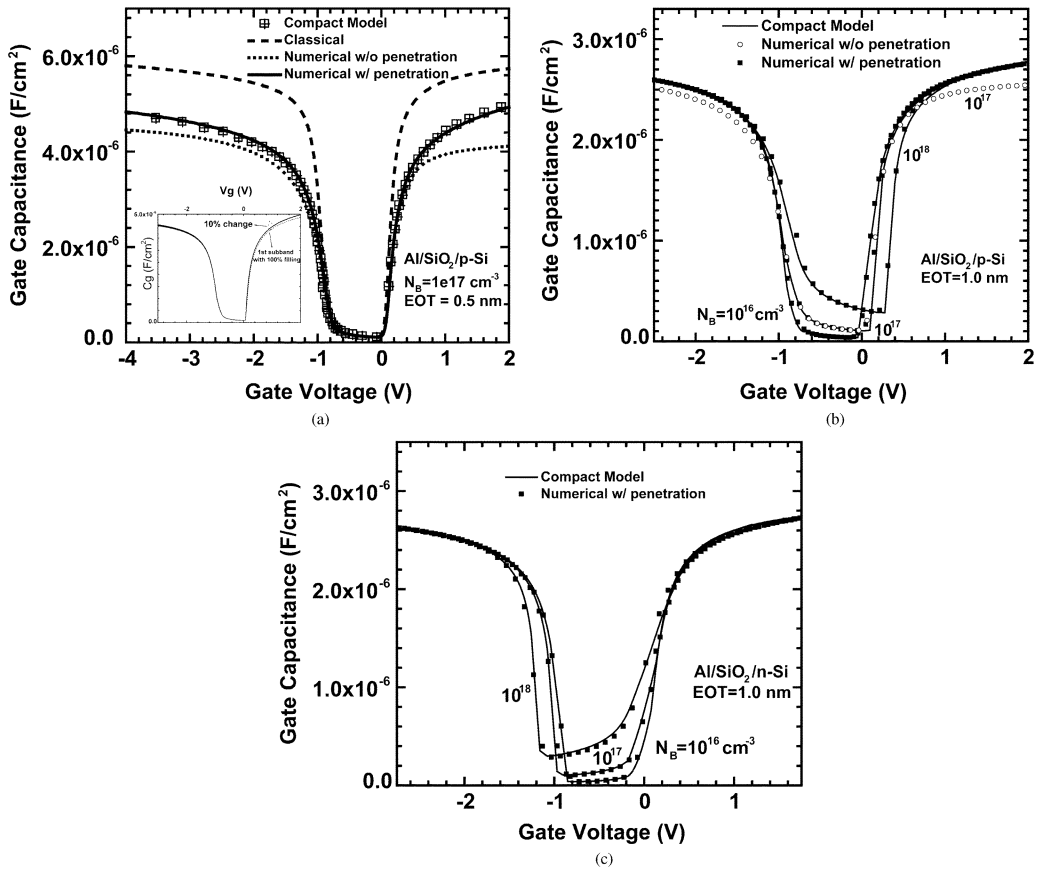


Fig. 7. (a) Comparison of C–V for 0.5-nm EOT device obtained from the compact model and obtained from numerical simulations within a classical model, and within quantum-mechanical treatments with and without wave function penetration into the dielectric. The inset shows that the degree of subband filling is less critical to the results than barrier penetration and that small changes in subband occupancy produce almost indiscernible changes. However, an assumption of 100% groundstate subband occupancy still produces an at least noticeable error. Similar behavior was observed for p-MOS devices. (b), (c) Comparisons of compact and numerical calculations for n-MOS and p-MOS metal-gate capacitors, respectively, with 1-nm EOT dielectrics and doping concentrations of 10^{16}, 10^{17} and 10^{18} cm^{-3}, respectively.

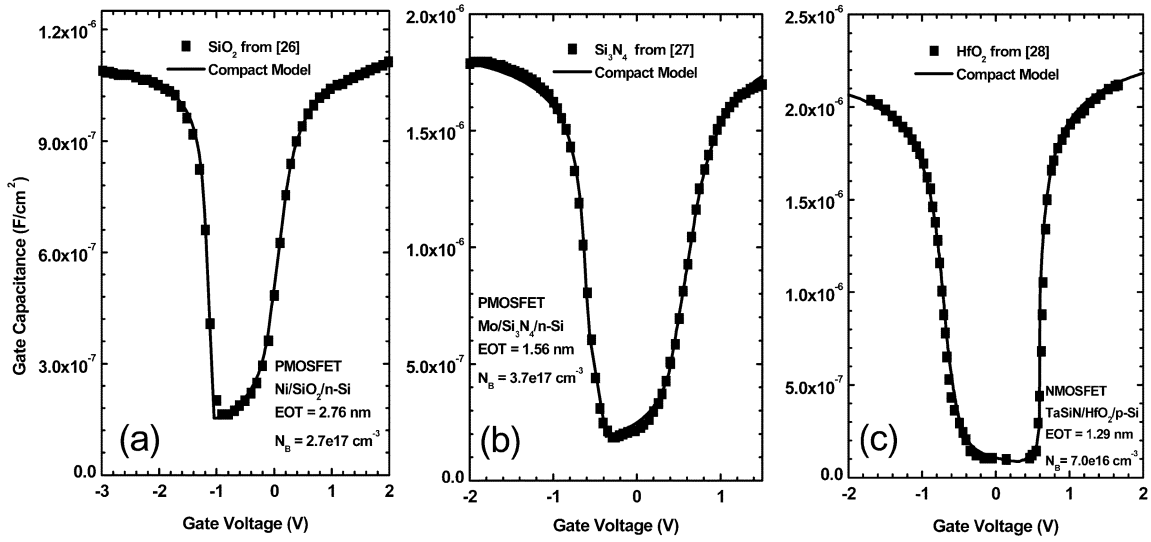


Fig. 8. Comparing experimental data of metal-gate MOS devices: (a) p-MOSFET with SiO_2 [26], (b) p-MOSFET with Si_3N_4 [27], and (c) n-MOSFET with HfO_2 [28] as gate dielectrics with EOTs of 2.76, 1.53, and 1.29 nm, respectively.

which means that the strong inversion/accumulation layer becomes effectively thinner with increasing field, largely due to barrier penetration in this case. As a result, the capacitance will continue to increase somewhat even well into strong inversion. Similar behavior including the insensitivity of Fermi level to doping concentration is also observed in p-MOS capacitors.

In Fig. 7(a), the compact model results of C–V for a 0.5-nm EOT n-MOS device are compared to numerical results for a

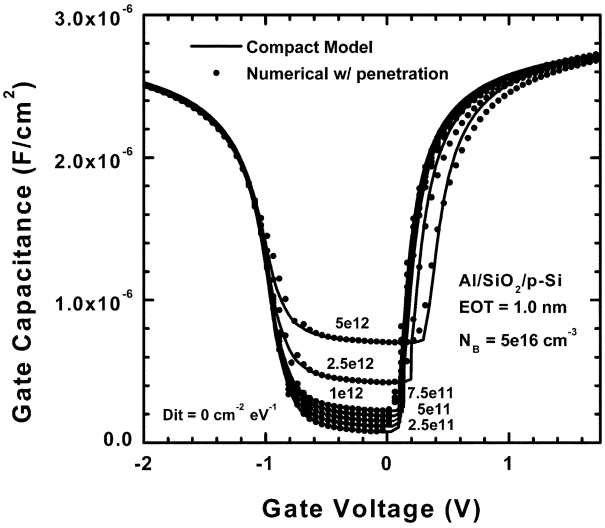


Fig. 9. Comparison of compact model predictions with self-consistent solutions with wave function penetration effects for n-MOS metal-gate capacitors with dielectrics of 1-nm EOT and uniform $D_{\rm it}$ profiles.

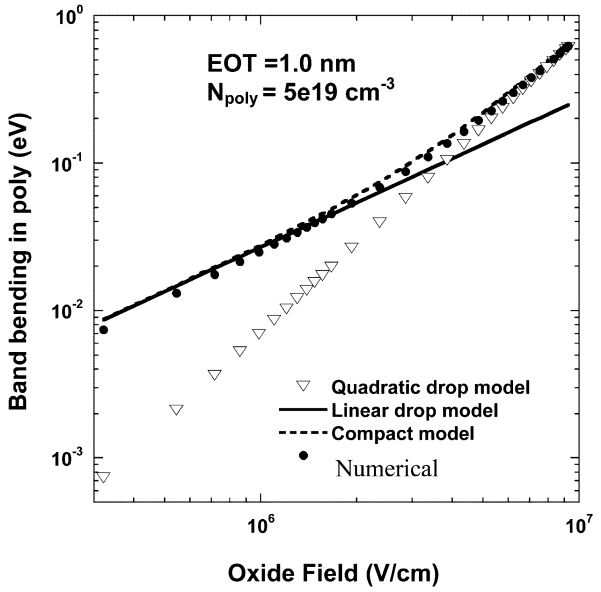


Fig. 10. Numercial calculations of polydepletion-induced band bending in the gate compared with the linear, quadratic and composite compact models.

classical treatment of the carriers—particles and Boltzmann statistics—in the inversion and accumulation layers, for a quantum mechanical treatment neglecting barrier penetration, and for a quantum mechanical treatment including barrier penetration. Here, the accuracy of the compact model and the importance of barrier penetration are both evident. Also evident via the continued approach of the quantum results with barrier penetration to the classical results are the limits of accuracy of modeling quantum mechanical effects via a fixed charge centroid shift or "dark space" for such small devices. In Fig. 7(b) and (c) excellent agreement is demonstrated between numerical and compact modeling of both 1.0-nm EOT n-MOS and p-MOS devices, respectively, and with varying doping concentrations. Furthermore, the compact model requires only 50 ms on a 1.1 GHz clock speed computer for a C–V curve consisting of 200 individual data points. By contrast the numerical simulations (with obviously fewer data points) required ~ 30 minutes to 1.5 h depending on speed of convergence for a relative speed difference of over four orders of magnitude.

Fig. 8 shows the results of fitting the compact model to experimental data: (a), a p-MOS device with 2.76 nm of SiO_2 [26], (b) a p-MOS device with a 1.53-nm EOT Si_3N_4 dielectric [27], and (c) an n-MOS device with a 1.29 EOT HfO_2 gate stack [28]. Note that while true EOTs can always be argued and particularly so for high-κ gate stacks (which is one of the reasons comparison to numerical results are also important), we could not fit the shape of the data curves as well without the consideration of barrier penetration that is built into this compact model.

Note that the effects of barrier penetration, obvious here for metal-gate devices, could be obscured by underestimation of polydepletion effects and associated overestimation of poly doping in poly gate devices to be considered.

Large densities of interface states $(D_{\rm it})$ degrade the device performance and the behavior of gate C–V [29]. Hence, it is useful to characterize the $D_{\rm it}$ values and suppress them in device fabrication, because the quality of dielectric-substrate interfaces for MOS devices with high-κ materials is inferior to that

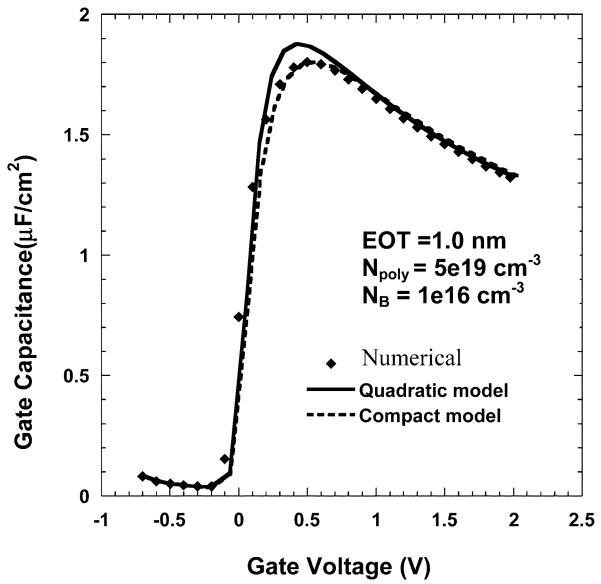


Fig. 11. Effects of polydepletion on gate capacitance in weak and strong inversion as calculated numerically, and with the quadratic and composite models of band-bending in the polysilicon.

with SiO_2. The effects of a fixed uniform $D_{\rm it}$ distribution within the bandgap of the substrate Si were also introduced in this compact model producing results consistent with our numerical simulations, as shown in Fig. 9. It was simply assumed that at the interface states with energies below the gate voltage-dependent Fermi level are filled with electrons, while those above the Fermi level are empty. The discontinuity between the dielectric field and the substrate field—where it is the latter that determines subband energies, band filling and the Fermi level at the surface, of course—was then adjusted accordingly. For specificity only, flatband was chosen as the charge-neutral condition for this calculation.

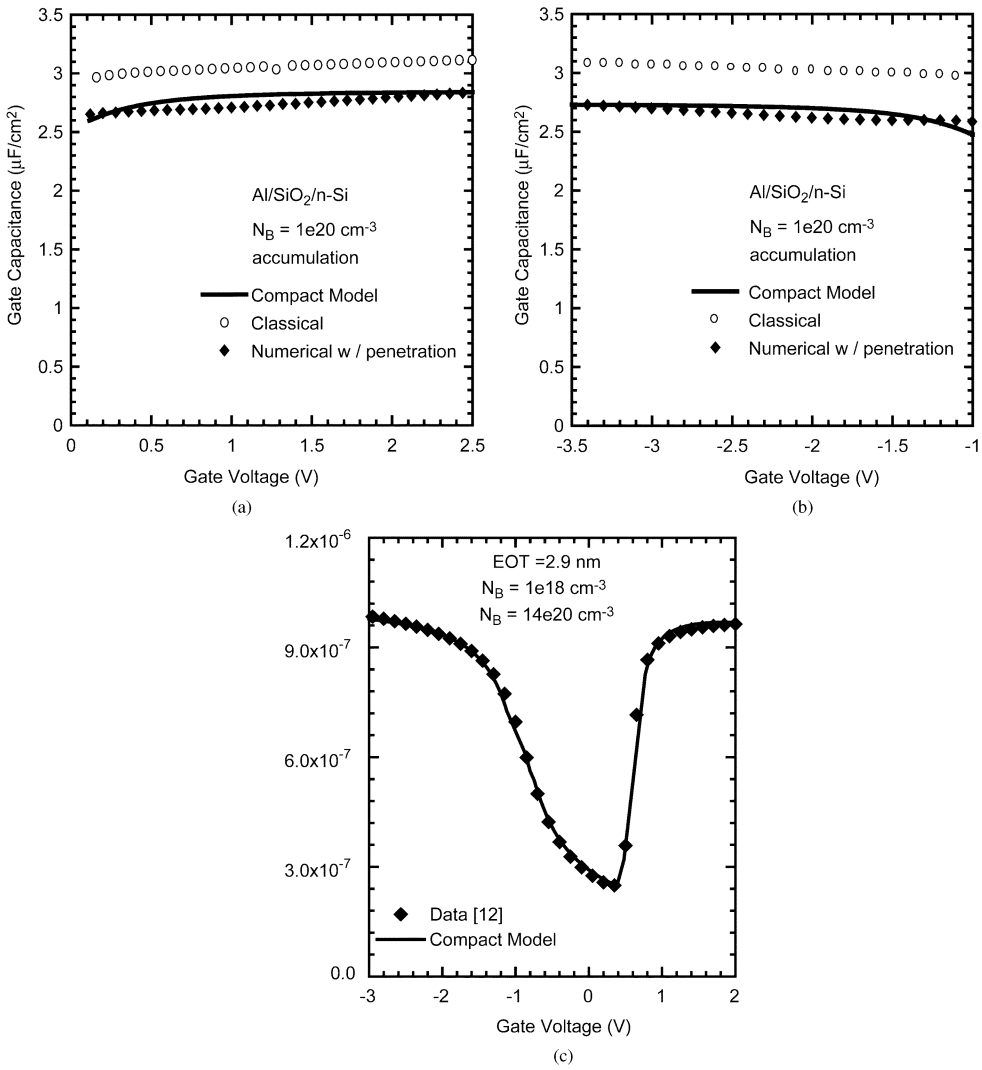


Fig. 12. Capacitance for a metal/SiO_2/highly doped-Si (polysilicon-like) n-substrate structure (a) [a metal/SiO_2/highly doped-Si (polysilicon-like) p-substrate structure (b)] used to isolate the effect of band-bending during accumulation in the polysilicon gate of a n-MOS (p-MOS) device, as calculated with the compact model and numerically within classical and quantum mechanical models. (c) Fit of C–V from the compact model to experimental data of a poly gate n-MOS device with, according to our calculations, a 2.9-nm SiO_2.

IV. Gate Capacitance and Polysilicon Gate Devices

When polysilicon gates are used, polydepletion effects result in reduced gate capacitance when the substrate is in strong inversion. This reduction becomes relatively more important as the oxide EOT decreases, which is at least one of the motivating factors for a return to true metal-gates. Even in accumulation, the thickness of the surface charge layer in the polysilicon must be considered. Therefore, for completeness in this paper we have also provided a simple yet still accurate compact model for these effects as well. However, polysilicon gate modeling is not a major focus of this paper and the approach here is clearly more empirical as we have little expectation that sub-nanometer EOT devices will widely employ polysilicon gates. Also, one could easily substitute other models, such as that of [32] for polyaccumulation affects, without affecting the remainder of this model.

In polydepletion, for very high doping concentrations we have found the electrostatic potential drop predicted by the simple quadratic model characteristic of an abrupt transition approximation between the charge neutral region and the depletion region, (11) above, to be insufficient even as the transition to strong inversion in the substrate is occurring. At low fields the relationship becomes linear and it can be better approximated as [23], [30]

$$q\,|\varphi_{\mathrm{poly,lin}}| = q\frac{\varepsilon_{\mathrm{ox}}}{\varepsilon_s}\,|F_{\mathrm{ox}}|\,\sqrt{2}\,L_{\mathrm{Debye}} \tag{15}$$

where $L_{\mathrm{Debye}} = \left(\varepsilon_s K_B T/q^2 N_{\mathrm{poly}}\right)$ is the doping dependent classical Debye length and $F_{\mathrm{ox}}\varepsilon_{\mathrm{ox}}/\varepsilon_s$ is field on the polysilicon side of the interface with the gate dielectric and the field. For the compact model, therefore, a simple composite model was created

$$q|\varphi_{\mathrm{poly}}| = q(|\varphi_{\mathrm{poly,lin}}|^2 + |\varphi_{\mathrm{poly,quad}}|^2)^{1/2} \tag{16}$$

that captures the asymptotic behavior of the voltage drop in the polysilicon as a function of the oxide field at low and high fields, as illustrated in Fig. 10. The total voltage drop then becomes

$$V_G = V_{\mathrm{FB}} + F_{\mathrm{ox}}t_{\mathrm{ox}} + \varphi_{\mathrm{poly}}(F_{\mathrm{ox}}) + \varphi_{\mathrm{sub}}(F_{\mathrm{ox}}) \tag{17}$$

in depletion, weak inversion, and strong inversion. The improvement in C–V modeling resulting from this more careful treatment of the polydepletion is illustrated in Fig. 11.

Traditionally, if surface potential in the polysilicon in accumulation is considered at all, carriers in polysilicon have been modeled classically [25]. However, the carrier wavelengths remain those of a semiconductor so that the charge centroid should be shifted back from the interface by quantum effects even at flatband [31]. Neither, however, can the band bending in the now heavily doped accumulation layer be accurately estimated assuming the single eigenenergy level approach of (6) and (7) as the groundstate occupancy becomes significantly and nontrivially field-dependent. The groundstate occupancy, however, also becomes relatively small, which is to say that many eigenenergy levels are significantly occupied. However, in the end, and in stark contrast to the case for substrate, a fixed charge centroid shift/dark-space such as in [31], in series with the classical Debye length at low fields, seems to work well, as illustrated in Fig. 12 where a metal/SiO_2/high-doping-Si polysilicon-like structure was used to isolate this behavior. By comparison to such numerical simulations, the necessary dark space in the polysilicon was found to be 0.65 and 0.80 nm for n-type and p-type polysilicon, respectively, essentially independent of voltage, polysilicon doping concentration $N_{\rm poly}$ (over a range of 10^{19} cm^{-3} to 5×10^{20} cm^{-3}) and even temperature. After correcting for the differences in dielectric constants, the dark-space and the classical screening length are simply added to the EOT of (11), so that the gate voltage becomes

$$V_G = V_{\rm FB} + F_{\rm ox}\left[t_{\rm ox} + \frac{\varepsilon_{\rm ox}}{\varepsilon_{\rm Si}}(t_{\rm cl} + t_{\rm dark})\right] + \varphi_{\rm sub}(F_{\rm ox}) \quad (18)$$

in accumulation. The classical contribution [23] can be accurately fitted as a function of oxide field by

$$t_{cl} = \frac{\sqrt{2}\,L_D}{1 + 5.83\times10^8\left(\dfrac{|F_{\rm ox}|\frac{\varepsilon_{\rm ox}}{\varepsilon_{Si}}\left(\frac{\rm MV}{\rm cm}\right)}{\sqrt{\frac{T\cdot N_{\rm poly}}{({\rm K}\cdot{\rm cm}^3)}}}\right)^{1.1}} \quad (19)$$

(which is of the form of a simple Fermi function in the log of the field in the polysilicon gate) which again reduces to $2^{1/2}L_D$ at flatband. Note that the 0.65-nm dark-space in the Si for electrons produces about a 0.2 nm increase of apparent oxide thickness of the classical result. Fig. 12(c) shows excellent agreement between the compact model and measured data [12] during both accumulation and depletion in the polysilicon for a MOSFET with EOT of 2.9 nm. However, the models are still discontinuous across flatband and interpolation is again used as described in Section II-G.

V. Conclusion

Guided by basic physics, analytic theory [7], and self-consistent numerical Schrödinger–Poisson calculations [4]–[6], [8], [9], a new compact model of C–V for gate dielectrics with EOTs of ~ 1 nm and below and metal gates has been developed that addresses quantum mechanical effects including subband formation and barrier penetration. This basic algorithm used in the compact C–V simulations is firmly grounded in first-principles. Comparisons to numerical simulations have been used to refine the parameters of otherwise analytically derived expressions for subband energies and subband occupation probabilities within the substrate in strong inversion and accumulation. Comparison to both numerical simulation and experimental data show excellent agreement. Also, interface states have been considered within the model. For nonuniform substrate doping profiles, although not explicitly considered here, the insensitivity of the remainder of the compact C–V calculations to doping suggest that only the model for band-bending in the depletion region would have to be refined. Although the trend is back to metal-gates for the thinnest of gate dielectrics, for completeness models of polydepletion and polyaccumulation effects have been provided that also compare well to numerical simulation. All equations are simple explicit functions of the oxide field allowing entire C–V curves to be produced in a few tens of milliseconds on personal computers with accuracy comparable to fully numerical calculations.

References

[1] (2003) The International Technology Roadmap for Semiconductors (2003 Edition) Process Integration, Devices and Structures, (Table 47 for Future EOTs). [Online]. Available: http://public.itrs.net/Files/2003ITRS/Home2003.htm

[2] F. Stern and W. E. Howard, "Properties of semiconductor surface inversion layers in the electric quantum limit," *Phys. Rev.*, vol. 163, pp. 816–835, 1967.

[3] A. Haque and M. Z. Kauser, "A comparison of wave-function penetration effects on gate capacitance in deep submicron n- and p-MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 9, pp. 1580–1587, Sep. 2000.

[4] W.-K. Shih, E. X. Wang, S. Jallepalli, F. Leon, C. M. Maziar, and A. F. Tasch, "Modeling gate leakage current in nMOS structures due to tunneling through an ultrathin oxide," *Solid State Electron.*, vol. 42, pp. 997–1006, 1998.

[5] S. Mudanai, Y. Y. Fan, Q. Ouyang, A. F. Tasch, L. F. Register, and S. K. Banerjee, "Modeling of direct tunneling current through gate dielectric stacks," *IEEE Trans. Electron Devices*, vol. 47, no. 10, pp. 1851–1857, Oct. 2000.

[6] S. Mudanai, L. F. Register, A. F. Tasch, and S. K. Banerjee, "Understanding the effects of wave penetration on the inversion layer capacitance of NMOSFETs," *IEEE Electron Device Letters*, vol. 22, pp. 145–147, Mar. 2001.

[7] F. Stern, "Self-consistent results for n-type Si inversion layers," *Phys. Rev. B, Condens. Matter*, vol. 5, pp. 4891–4899, 1972.

[8] "UTQUANT.3.0," Univ. Texas, Austin, TX, 2004.

[9] S. A. Hareland, S. Krishnamurthy, S. Jallepalli, C.-F. Yeap, A. F. Tasch, and C. M. Maziar, "A computationally efficient model for inversion layer quantization effects in deep submicron n-channel MOSFETs," *IEEE Trans. Electron Devices*, vol. 43, no. 1, pp. 90–96, Jan. 1996.

[10] M. J. Van Dort, P. H. Woerlee, A. J. Walker, C. A. H. Juffermans, and H. Lifka, "Quantum mechanical threshold voltage shifts of MOSFETs caused by high levels of channel doping," in *IEDM Tech. Dig.*, 1991, session 18.

[11] M. J. Van Dort, P. H. Woerlee, and A. J. Walker, "A simple model for quantization effects in heavily doped silicon MOSFETs at inversion conditions," *Solid State Electron.*, vol. 37, no. 3, pp. 411–414, 1994.

[12] J. R. Hauser and K. Ahmed, "Characterization of ultrathin oxides using electrical C–V and I–V measurements," in *Proc. Characterization Methology ULSI Technology Conf.*, 1998, pp. 235–239.

[13] K. Ahmed, E. Ibok, G. Bains, D. Chi, B. Ogle, J. Wortman, and J. R. Hauser, "Comparative physical and electrical metrology of ultrathin oxides in the 6 to 1.5 nm regime," *IEEE Trans. Electron Devices*, vol. 43, no. 1, pp. 1349–1354, Jan. 2000.

[14] Y. Ohkura, "Quantum effects in Si n-MOS inversion layer at high substrate concentration," *Solid State Electron.*, vol. 33, pp. 1581–1585, 1990.

[15] F. Stern, "Effect of a thin transition layer at a Si-SiO_2 interface on electron mobility and energy levels," *Solid State Comm.*, vol. 21, pp. 163–166, 1977.

[16] S. Mudanai, L. F. Register, A. F. Tasch, and S. K. Banerjee, "A new and accurate quantum mechanical compact model for nMOS gate capacitance," in *Proc. Device Research Conf.*, Jun. 2001, pp. 87–88.

[17] W. K. Chim, J. X. Zheng, and B. H. Koh, "Modeling of charge quantization and wave penetration effects in a metal-oxide-semiconductor system with ultrathin gate oxide," *J. Appl. Phys.*, vol. 94, no. 8, pp. 5273–5277, 2003.

[18] M. M. A. Hakim and A. Haque, "Effects of neglecting carrier tunneling on electrostatic potential in calculating direct tunneling gate current in deep submicron MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 9, pp. 1669–1671, Sep. 2000.
[19] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High-κ gate dielectrics: Current status and materials properties considerations," *J. Appl. Phys.*, vol. 89, no. 10, pp. 5243–5275, Oct. 2001.
[20] H. L. Shang, K.-L. Lee, P. Kozlowski, C. D'Emic, I. Babich, E. Sikorski, M. Ieong, H. S. P. Wong, K. Guarini, and W. Haensch, "Self-aligned n-channel germanium MOSFETs with a thin Ge oxynitride gate dielectric and tungsten gate," *IEEE Electron Device Lett.*, vol. 25, pp. 135–137, Dec. 2004.
[21] T. Ando, B. Fowler, and F. Stern, "Electronic properties of two-dimensional systems," *Rev. Modern Phys.*, vol. 54, no. 2, pp. 437–672, Apr. 1982.
[22] F. Stern, "Quantum properties of surface-charge layer," *CRC Critical Rev. Solid State Sci.*, pp. 499–514, Mar. 1974.
[23] S. M. Sze, *Physics of Semiconductor Devices*. New York: Wiley, 1981, ch. 7.
[24] M. T. Heath, *Scientific Computing an Introductory Survey*, 2nd ed. New York: McGraw-Hill, ch. 7.
[25] L. Larcher, P. Pavan, F. Pellizzer, and G. Ghidini, "A new model of gate capacitance as a simple tool to extract MOS parameters," *IEEE Trans. Electron Devices*, vol. 48, no. 5, pp. 935–945, May 2001.
[26] P. Q. Xuan and J. Bokor, "Investigation of NiSi and TiSi as CMOS gate materials," *IEEE Electron Device Lett.*, vol. 24, no. 3, pp. 634–638, Mar. 2003.
[27] Q. Lu, R. Lin, P. Ranade, Y. C. Yeo, X. F. Meng, H. Takeuchi, T.-J. King, C. M. Hu, H. F. Luang, S. J. Lee, W. P. Bai, C.-H. Lee, D.-L. Kwong, X. Guo, X. W. Wang, and T.-P. Ma, "Molybdenum metal-gate MOS technology for post-SiO_2 gate dielectrics," in *IEDM Tech. Dig.*, Dec. 2000, pp. 641–644.
[28] S. B. Samavedam *et al.*, "Dual-metal gate CMOS with HfO_2 gate dielectric," in *IEDM Tech. Dig.*, 2002, pp. 433–436.
[29] S. Mudanai, F. Li, S. B. Samavedam, P. J. Tobin, C. S. Kang, R. Nieh, J. C. Lee, L. F. Register, and S. K. Banerjee, "Interfacial defect states in HfO_2 and ZrO_2 nMOS capacitors," *IEEE Electron Device Lett.*, vol. 23, no. 12, pp. 728–730, Dec. 2002.
[30] S. Mudanai, "Gate Current Modeling Through High-κ Materials and Compact Modeling of Gate Capacitance," Ph.D. dissertation, Univ. Texas, Austin, TX, 2001.
[31] A. Pacelli, A. Spinelli, and L. Perron, "Carrier quantization at flat bands in MOS devices," *IEEE Trans. Electron Devices*, vol. 46, no. 2, pp. 383–387, Feb. 1999.
[32] A. Spinelli, R. Clerc, and Ghibaudo, "An analytic model for flat-band polysilicon quantization in MOS devices," *IEEE Trans. Electron Devices*, vol. 49, no. 7, pp. 1314–1316, Jul. 2002.

Fei Li (M'04) received the B.S. and M.S. degrees in physics from Peking University, Beijing, China, in 1996 and 1999, respectively, the M.S.E.E. degree from University of Notre Dame, Notre Dame, IN, in 2001, and the Ph.D. degree in electrical engineering based on a study of compact modeling and Flash memories at University of Texas at Austin, Austin, in 2004.

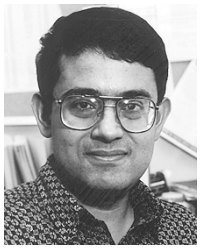
He joined Nassda Corporation, Santa Clara, CA, in January, 2005.

Sivakumar Mudanai (M'03) was born in Chennai (formerly Madras), India, in 1975. He received the B.Tech degree from the Indian Institute of Technology, Chennai, India, in 1996, and the M.S.E. and Ph.D. degrees in electrical engineering in 1998 and 2001, respectively, from the University of Texas, Austin.

His research interests are in the area of deep submicrometer MOS device modeling and analysis, and process integration. He joined Intel Corporation, Portland, OR, in 2001.

Leonard Franklin Register (SM'95) received the B.S. degrees in electrical and computer engineering and in physics, and the Ph.D. degree in electrical and computer engineering, all from North Carolina State University, Raleigh.

He then joined the Computational Electronics Group at the University of Illinois' Beckman Institute where he became a research scientist. There he reformed theoretical research and published in diverse areas including quantum transport, high-energy transport, single electronics, laser theory and device reliability. He joined the faculty of The University of Texas at Austin, Austin, within the Department of Electrical and Computer Engineering and the Microelectronics Research Center in the spring of 2000. His current research has an added emphasis on quantum and high-energy transport effects within sub-tenth-micrometer and non-classical CMOS.



Sanjay K. Banerjee (F'95) received the B.Tech. degree from the Indian Institute of Technology, Kharagpur, India, in 1979 and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign in 1979, 1981, and 1983, respectively, all in electrical engineering.

He is the Cockrell Family Regents Chair Professor of Electrical and Computer Engineering and Director, Microelectronics Research Center, University of Texas (UT) at Austin, Austin. As Member of the Technical Staff, Corporate Research, Development, and Engineering, Texas Instruments, Inc., Dallas, TX, from 1983 to 1987, he worked on polysilicon transistors and dynamic random access trench memory cells used by Texas Instruments in the worlds first 4-Mb DRAM. He has been Assistant Professor (1987–1990), Associate Professor (1990– 1993), and Professor (1993 to the present) at UT. He has more than 400 archival referee d publications/talks, five books/chapters, and over 20 U.S. patents. He is currently active in the areas of ultrahigh vacuum and remote plasma-enhanced chemical vapor deposition for silicon-Ge-carbon heterostructure MOSFETs and nanostructures. He is also interested in the areas of ultrashallow junction technology and semiconductor device modeling.

Dr. Banerjee received the Best Paper Award from the IEEE International Solid State Circuits Conference in 1986, the Engineering Foundation Advisory Council Halliburton Award in 1991, the Texas Atomic Energy Fellowship (1990 to 1997), the Cullen Professorship (1997 to 2001), the NSF Presidential Young Investigator Award in 1988, the ECS Callinan Award, in 2003 the IEEE Millennium Medal, in 2000, and the SRC Inventor Recognition Award in 2000. He is a Distinguished Lecturer for the IEEE Electron Devices Society and was the General Chair of the IEEE Device Research Conference in 2002.