The Influence of Space Quantization Effect on the Threshold Voltage, Inversion Layer and Total Gate Capacitance in Scaled Si-MOSFETs

Dragica Vasileska and David K. Ferry

Center for Solid State Electronics Research
Arizona State University, Tempe AZ 85287-5706, USA
vasilesk@imap2.asu.edu

ABSTRACT

We investigate the influence of poly-gate depletion on the inversion layer capacitance C_{inv}, total gate capacitance C_{tot} and threshold voltage V_T in scaled Si MOSFETs using both semiclassical (SC) and quantum-mechanical (QM) description of the charge density in the channel. We also present an analytical expression for the total gate capacitance C_{tot} that uses SC charge description and takes into account the depletion of the poly-silicon gates. Our simulations show that poly-gate depletion has larger influence on C_{tot} than the QM charge description. On the contrary, V_T is significantly affected by both the poly-gate depletion and the QM effects in the channel.

Keywords: scaled Si-MOSFETs, threshold voltage, inversion layer capacitance, total gate capacitance, poly-gate depletion.

INTRODUCTION

Successful scaling of MOSFETs toward shorter channel lengths requires thinner gate oxides and higher doping levels in order to have high drive currents and minimized short-channel effects [1,2]. For state-of-the art devices, it was demonstrated a long time ago that, as the gate oxide thickness is scaled down to 10 nm and below, the total gate capacitance is smaller than the oxide capacitance due to the comparable values of the oxide and the inversion layer capacitances. As a consequence, the device transconductance is degraded relative to the expectations of scaling theory [3]. The inversion layer capacitance was also identified as being the main cause of the second-order thickness-dependence of MOSFET *I-V* characteristics [4].

The finite inversion layer thickness was estimated experimentally by Hartstein and Albert [5]. The high levels of substrate doping needed in deep-submicrometer devices leads to a pronounced quasi-two-dimensional (Q2D) nature of the carrier transport and this was found responsible for the increased threshold voltage and decreased channel mobility. A simple analytical model that accounts for these effects was proposed [6,7]. Two physical origins of inversion layer capacitance (due to finite density of states and due to finite inversion layer thickness) were demonstrated experimentally by Takagi and Toriumi [8].

A computationally efficient three-subband model, that predicts both the quantum-mechanical effects in the inversion layer and the electron distribution within the inversion layer, was proposed and implemented into the PISCES simulator [9]. The influence of the image and exchange-correlation effects on the inversion layer and total gate capacitance was studied by Vasileska *et al.* [10]. Very recently, it was also pointed out [11] that the depletion of the poly-silicon gates considerably influences the magnitude of C_{tot}. However, nobody has yet examined the influence of the poly-gate depletion on the threshold voltage or on the inversion layer capacitance, two very important issues that we focus on in this work.

NUMERICAL CALCULATION OF C_{tot}

Detailed investigation of the total gate capacitance C_{tot} requires a choice for an accurate model for the mobile charge distribution in the surface inversion region of the device. There are two possibilities: One can use either a quantum-mechanical or a classical description of the electron density in the inversion layer. Using a QM description of the inversion layer electron density, the procedure that we follow when calculating the total gate capacitance is the following: We first solve the non-linear 1D Poisson equation,

$$\frac{\partial}{\partial z}\left[\frac{1}{\varepsilon(z)}\frac{\partial \varphi}{\partial z}\right] = -e\left[N_D^+(z) - N_A^-(z) + p(z) - n(z)\right], \qquad (1)$$

self-consistently with the 1D Schrödinger equation

$$\left[-\frac{\hbar^2}{2m_\perp^i}\frac{\partial^2}{\partial z^2}+V(z)\right]\psi_{ij}(z)=E_{ij}\psi_{ij}(z)\,. \tag{2}$$

In (1) and (2), $\varphi(z)$ is the electrostatic potential, $\varepsilon(z)$ is the dielectric constant, $N_D^+(z)$ and $N_A^-(z)$ are the ionized donor and acceptor concentrations, $n(z)$ and $p(z)$ are the electron and hole densities, $V(z)$ is the potential energy, $m_\perp^i$ is the effective mass normal to the semiconductor-oxide interface of the i-th valley, and E_{ij} and $\psi_{ij}(z)$ are the energy level and the corresponding wavefunction of the electrons residing in the j-th subband from the i-th valley. We want to point out that, in the calculations presented here, we assume that the SiO_2/Si interface is parallel to the [100] plane. For this particular case, the six equivalent minima of the bulk silicon conduction band split into two sets of subbands. The first set (Δ_2-band) consists of the two equivalent valleys with in-plane effective mass $m_\parallel$=0.19m_o and perpendicular effective mass $m_\perp$=0.91m_o. The second set (Δ_4-band) consists of the four equivalent valleys with $m_\parallel$=0.42m_o and $m_\perp$=0.19m_o. The energy levels associated with the Δ_2-band comprise the so-called *unprimed* ladder of subbands, whereas those associated with the Δ_4-band comprise the *primed* ladder of subbands.

Once the eigenfunctions and the eigenvalues that characterize the electrons in the inversion layer are determined, the inversion layer electron density is obtained by summing over all subbands to get

$$n(z)=\sum_{i,j}\frac{m_\parallel^i k_B T}{\pi\hbar^2}\ln\left[1+\exp\left(\frac{E_F-E_{ij}}{k_B T}\right)\right]\psi_{ij}^2(z)\,, \tag{3a}$$

where E_F is the Fermi level, k_B is the Boltzmann constant, T is the temperature and $m_\parallel^i$ is the in-plane effective mass of the i-th valley. We want to point out that the inversion layer electrons are treated quantum-mechanically only when confined by the surface field. Otherwise, we skip the solution of the 1D Schrödinger equation and use the SC description

$$n(z)=N_C F_{1/2}\left(\frac{E_F-E_C(z)}{k_B T}\right), \tag{3b}$$

where N_C is the effective density of states of the conduction band. For holes, which are always treated SC, we use

$$p(z)=N_V F_{1/2}\left(\frac{E_C(z)-E_G-E_F}{k_B T}\right), \tag{4}$$

where N_V is the effective densities of states of the valence band and E_G is the semiconductor bandgap. Analytical approximations for the Fermi-Dirac integrals, which appear in expression (3b) and (4), are given in [12].

Once we have determined the charge distribution on the semiconductor side of the MOS capacitor, we can evaluate the magnitude of C_{tot} by differentiating the total induced charge density in the channel with respect to the gate voltage V_G. The inversion layer and the poly-gate capacitances are evaluated analogously.

We point out that, in both our numerical and our analytical model (described in the next section), the poly-silicon gates are modeled as heavily-doped single-crystal silicon. Both the electrons and holes are treated classically using equations (3b) and (4), assuming general Fermi-Dirac statistics valid for degenerate semiconductors.

ANALYTICAL MODEL

In our extended analytical model, the poly-gate capacitance C_{poly} is found by the approach in [13]. Briefly, we first solve for the surface potential φ_p from

$$E_{ox}=\frac{\varepsilon_{sc}}{\varepsilon_{ox}}\sqrt{\frac{2eN_D}{\varepsilon_{sc}}}G(\varphi_p)\,, \tag{5}$$

where

$$G(\varphi_p)=\left[\varphi_p+\frac{N_C}{N_D}\int_0^{-\varphi_p}d\varphi F_{1/2}\left(\frac{e\varphi-\Delta E_{CP}}{k_B T}\right)-\frac{N_V}{N_D}\int_0^{-\varphi_p}d\varphi F_{1/2}\left(\frac{\Delta E_{CP}-E_G-e\varphi}{k_B T}\right)\right]^{1/2}, \tag{6}$$

and then calculate the poly-gate capacitance C_{poly} from

$$C_{poly}=\frac{\sqrt{eN_D\varepsilon_{sc}}}{\sqrt{2}G(\varphi_p)}\times\left[1-\frac{N_C}{N_D}F_{1/2}\left(-\frac{e\varphi_p+\Delta E_{CP}}{k_B T}\right)+\frac{N_V}{N_D}F_{1/2}\left(\frac{\Delta E_{CP}-E_G+e\varphi_p}{k_B T}\right)\right]. \tag{7}$$

In (5-7), ε_{sc} and ε_{ox} are the semiconductor and oxide dielectric functions, N_D is the doping of the poly-silicon

gates, and $\Delta E_{CP}=(E_C\text{-}E_F)_{poly}$ is found from the charge neutrality condition in the poly-silicon away from the interface. The corresponding gate voltage is found from

$$V_G = \varphi_{sc} + \varphi_p + (\Delta E_{CP} - \Delta E_{CB})/e + V_{ox}, \qquad (8)$$

where φ_{sc} is the surface potential in the semiconductor side of the semiconductor/oxide interface, $\Delta E_{CB}=(E_C\text{-}E_F)_{sc}$ is found from the charge neutrality condition in the semiconductor bulk region, and $V_{ox}=E_{ox}t_{ox}$, where t_{ox} is the oxide thickness. The total gate capacitance is then evaluated as a series combination of C_{poly}, $C_{ox}=\varepsilon_{ox}/t_{ox}$ and C_{inv}.

RESULTS AND DISCUSSION

To demonstrate the existence of the two physical origins of the inversion layer capacitance C_{inv}, discussed in Refs. [8] and [10], we show in Fig. 1 the variation of C_{inv} with inversion charge density N_s in the channel.

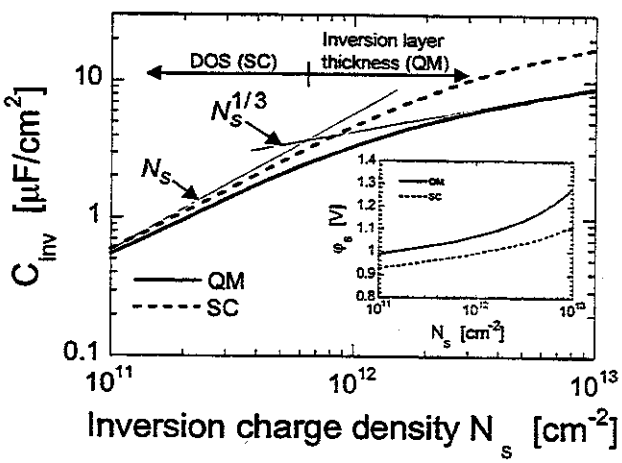

Figure 1. Variation of the inversion layer capacitance with inversion charge density at T=300 K for a MOS capacitor with substrate doping $N_A=5\times10^{17}$ cm^{-3}, oxide thickness t_{ox}=4 nm and metal gates. In the inset we show the self-consistent results for the variation of the surface potential with N_s when using both SC and QM descriptions of the electron density in the inversion layer.

A pronounced double-slope behavior of the quantum-mechanically calculated C_{inv} comes from the fact that the total inversion layer capacitance can be represented as a series capacitance of two contributions. The first contribution is the classical one and comes from the finite density of states, i.e. due to the fact that a finite change in the surface potential is always necessary to increase N_s (inset of Fig. 1), which leads to finite value of C_{inv}. This term dominates at low values of N_s (low gate voltages). The second contribution to C_{inv} is due to the finite inversion layer thickness, which effectively increases the oxide thickness in terms of the total gate capacitance, thus providing an additional capacitance component. As shown in Fig. 2, the origin of this contribution is the quantum-mechanical space quantization effect in the inversion layer that leads to larger inversion layer thickness and larger displacement of the charge from the interface when compared to what the SC calculations give (inset of Fig. 2). This term dominates at large gate voltages, where the inversion charge density N_s significantly influences the band-bending and leads to steeper rise of the conduction band near the SiO_2/Si interface.

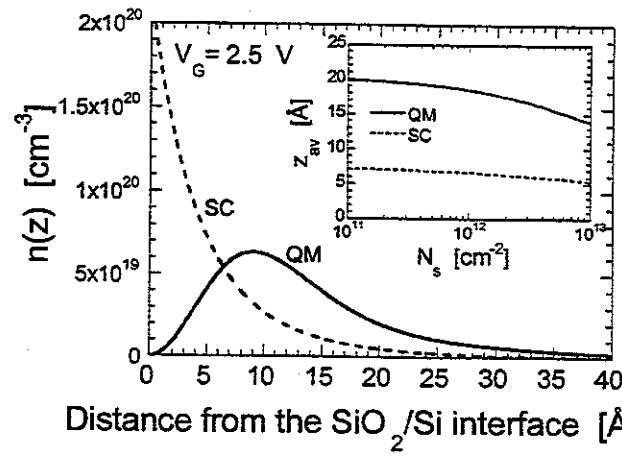

Figure 2. Classical and quantum-mechanically calculated charge distribution in the inversion layer for the device from Fig. 1, for V_G=2.5 V and T=300K. In the inset, we show the variation of the centroid of the electron density distribution in the inversion layer when using SC and QM descriptions of the inversion layer electrons. The quantum mechanical space-quantization effect leads to approximately three times larger average displacement of the inversion layer electrons from the SiO_2/Si interface.

As shown in Fig. 3, the inclusion of the poly-gate depletion only slightly modifies the magnitude of C_{inv} but leads to significant reduction of the inversion charge density in the channel at large gate biases (inset of Fig. 3), primarily due to the smaller density of states function for the Q2D case and significant degradation of the total gate capacitance due to the finite poly-gate capacitance shown in the inset of Fig. 4. For instance, for capacitors with metal gates and applied gate bias V_G=2.5 V, the SC (QM) models predict that C_{tot} is 95% (91%) of C_{ox}. For capacitors with poly-gates, the magnitude of C_{tot} is dramatically reduced

down to 80% (78%) of C_{ox} when using SC (QM) description of the charge in the channel. The above observations suggest that, even though the quantum-mechanical space-quantization effect considerably influences the magnitude of C_{inv}, it leads to no more than 10% degradation of the total gate capacitance C_{tot} in strong inversion. In this region, the dominant degradation mechanism for the total gate capacitance is depletion of the poly-silicon gates.

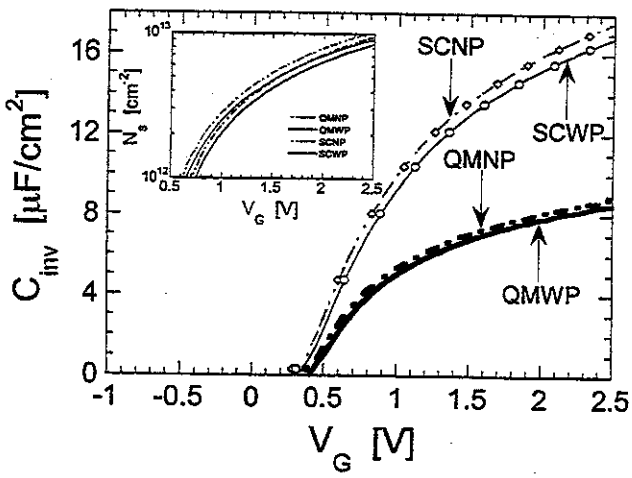

Figure 3. Variation of C_{inv} with V_G when using SC and QM description of the charge in the channel, with (WP) and without (NP) the inclusion of the poly-gate depletion. We use N_A=5×10^{17} cm^{-3}, t_{ox}=4 nm and N_D=5×10^{19} cm^{-3}. Also shown here are the semiclassical results obtained with our analytical model (symbols). In the inset, we show the variation of inversion layer electron density N_s with V_G.

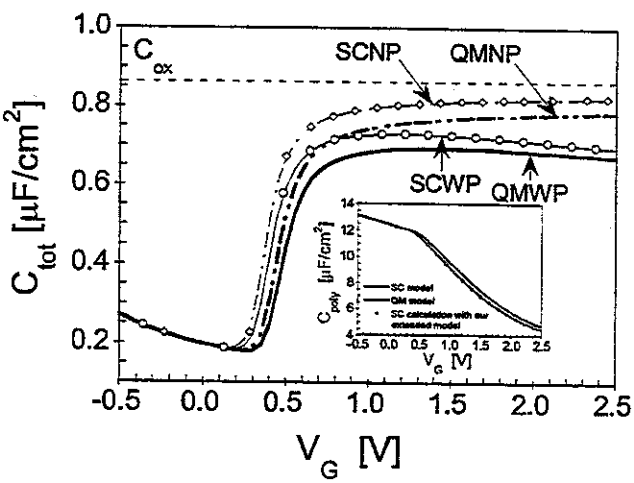

Figure 4. Simulated low-frequency CV-curves for the device from Fig. 3, which illustrate the change in the total gate capacitance C_{tot} due to finite C_{inv} and finite poly-gate capacitance C_{poly}. In the inset we show the variation of C_{poly} with V_G. Symbols have the same meaning as in Fig. 3.

The excellent agreement between our self-consistent (numerical) results for C_{inv}, C_{poly} and C_{tot} with those obtained with our extended analytical model suggest that our analytical model can be successfully used in predicting the errors which will result when neglecting the effect of C_{poly} on the terminal device characteristics. Another important observation that follows from the results shown in Fig. 4 is that, in contrast to the QM space-quantization effect, which in strong inversion leads to almost constant deviation of C_{tot} from C_{ox}, the depletion of the poly-silicon gates is a bias-dependent phenomenon, and one cannot use a single correction and accurately predict C_{tot} over the entire voltage range.

Until now, very little attention has been paid to careful modeling of the threshold voltage. Yet, the lowering of the supply voltage in deep-submicron devices must be accompanied by lowering of the threshold voltage to maintain optimum circuit performance. A careful investigation of the influence of the depletion of the poly-silicon gates and space-quantization effects on the magnitude of the threshold voltage V_{th} is thus mandatory.

To accomplish this task, we do a systematic study of the shift in the threshold voltage due to space quantization effects in the channel region of the device and the depletion of the polysilicon gates. A range of substrate doping densities and doping of the polysilicon gates that are representative of the deep-submicron technology is considered.

In Fig. 5 we show the linear region threshold voltage shift between the QM and SC predictions for a device with N_A=5×10^{17} cm^{-3} and t_{ox}=4 nm as a function of the doping of the poly-silicon gate. The threshold voltage V_{th} equals the gate voltage for which Q_{inv}=10$^{-3}Q_{depl}$. As expected, the QM description of the charge in the channel increases V_{th}. This is due to the fact that the QM picture differs from the SC one in two ways: First, the energy spectrum is not continuous, but consists of discrete energy levels which, in turn, reduces the DOS function. Second, the energy of the ground subband from the unprimed ladder of subbands does not coincide with the bottom of the conduction band (Fig. 6), and the energy difference $\Delta E = E_{11} - E_C$ increases with increasing substrate doping (not shown on this figure). The depletion of the poly-silicon gate (due to the insufficient doping) also increases the threshold voltage due to the significant degradation of the total gate capacitance.

The linear region threshold voltage shift for the device with t_{ox}=4 nm, N_D=10^{20} cm^{-3}, and different substrate doping is shown in Fig. 7. Also shown in this figure are

van Dort *et al.* [6] experimental data for a device with metal gates and oxide thickness t_{ox}=14 nm. Very close agreement between the experimentally derived threshold voltage shifts and our simulation results for the device with 14 nm thick oxide can be observed. A major difference from the results shown in Fig. 5 is that the inclusion of both the QM effects in the channel and poly-gate depletion leads to a strong dependence of the threshold voltage shift upon the substrate doping N_A. This suggests that both a QM description of the charge density distribution in the channel and the poly-gate depletion must be accounted for if accurate results for the threshold voltage are desired.

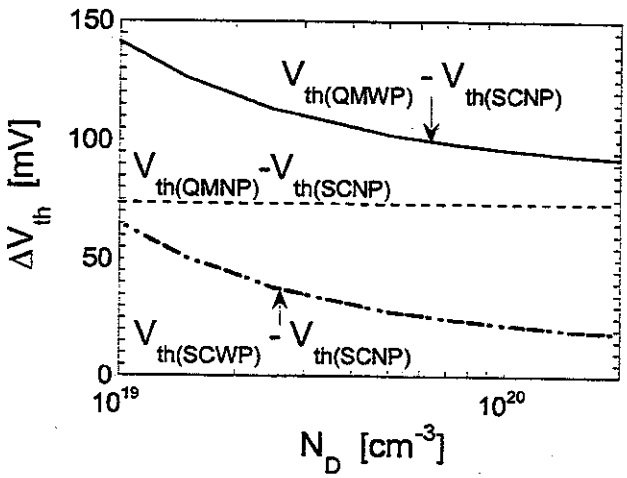

Figure 5. Linear region threshold voltage shift between the QM and the SC predictions versus N_D. We use N_A=5×10^{17} cm^{-3} and t_{ox}=4 nm.

CONCLUSIONS

In conclusion, we have investigated the influence of space quantization effects and poly-gate capacitance on the inversion layer and total gate capacitance, as well as their influence on the threshold voltage. We have demonstrated that for devices with thin gate oxides, poly-gate depletion is dominant degradation mechanism of the total gate capacitance. An analytical model for the calculation of the poly-gate capacitance, based on SC charge description in the channel, was also presented. The results obtained by using this model are in excellent agreement with our numerical self-consistent results. Our simulation results for the threshold voltage imply that the omission of the quantum-mechanical space-quantization effect can lead to erroneous predictions for the threshold voltage for deep submicrometer devices.

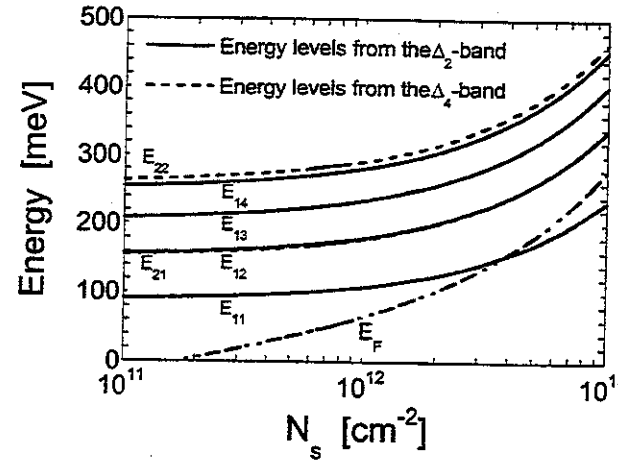

Figure 6. Variation of the energies of the lowest six subbands (four from the Δ_2-band and two from the Δ_4-band) and the position of the Fermi level with inversion charge density in the channel for the device from Fig. 1. All energies are measured from the bottom of the conduction band at the SiO_2/Si interface.

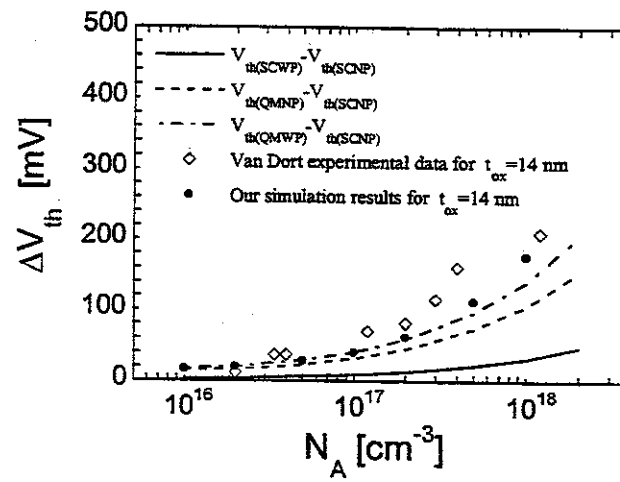

Figure 7. Linear region threshold voltage shift between the QM and the SC predictions versus N_A.

ACKNOWLEDGEMENTS

The authors would like to thank Dieter K. Schroder (Arizona State University) for pointing out the importance of this issue. The financial support by the Office of Naval Research is also acknowledged.

REFERENCES

1. R.H. Dennard, F.H. Gaensslen, H.-N. Yu, V.L. Rideout, E. Bassous, and A.R. leBlanc, "Design of

ion-implanted MOSFET's with very small physical dimensions," *IEEE J. Solid-State Circuits*, Vol. 9, pp. 256-268, 1974.

2. J.R. Brews, W. Fichtner, E.H. Nicollian, and S.M. Sze, "Generalized guide for MOSFET miniaturization," *IEEE Electron Dev. Lett.*, Vol. 1, pp. 2-4, 1980.

3. G. Bacarani, and M.R. Worderman, "Transconductance degradation in thin-oxide MOSFET's," in *Proceedings of the IEDM*, pp. 278-281, 1982.

4. M.-S. Liang, J.Y. Choi, P.-K. Ko, and C. Hu, "Inversion-layer capacitance and mobility of very thin gate-oxide MOSFET's," *IEEE Trans. Electron Devices*, Vol. 33, pp. 409-413, 1986.

5. A. Hartstein and N.F. Albert, "Determination of the inversion-layer thickness from capacitance measurements of metal-oxide-semiconductor field-effect transistors with ultrathin oxide layers," *Phys. Rev. B*, Vol. 38, pp. 1235-1240, 1988.

6. M.J. van Dort, P. H. Woerlee , A.J. Walker, C.A.H. Juffermans, and H. Lifka, "Influence of high substrate doping levels on the threshold voltage and the mobility of deep-submicrometer MOSFET's," *IEEE Trans. Electron Devices*, Vol. 39, pp. 932-938, 1992.

7. M. J. van Dort, P. H. Woerlee and A. J. Walker, "A simple model for quantization effects in heavily-doped silicon MOSFET's at inversion conditions," *Solid-State Electronics*, Vol. 37, pp. 411-414, 1994.

8. S. Takagi and A. Toriumi, "Quantitative understanding of inversion-layer capacitance in Si MOSFET's," *IEEE Trans. Electron Devices*, Vol. 42, pp. 2125-2130, 1995.

9. S. A. Hareland, S. Krishnamurthy, S. Jallepalli, C.-F. Yeap, K. Hasnat, A. F. Tasch, Jr. and C. M. Maziar, "A computationally efficient model for inversion layer quantization effects in deep submicron n-channel MOSFET's," *IEEE Trans. Electron Devices*, Vol. 43, pp. 90-96, 1996.

10. D. Vasileska, D. K. Schroder and D. K. Ferry, "Scaled silicon MOSFET's: Degradation of the total gate capacitance," *IEEE Trans. Electron Devices*, Vol. 44, pp. 584-587, 1997.

11. K. S. Krisch, J. D. Bude and L. Manchanda, "Gate capacitance attenuation in MOS devices with thin gate dielectrics," *IEEE Electron Device Letters*, Vol. 17, pp. 521-524, 1996.

12. J.S. Blakemore, "Approximations for Fermi-Dirac integrals, especially the function $F_{1/2}(\eta)$ used to describe electron density in a semiconductor," *Solid-State Electronics*, Vol. 25, pp. 1067-1076, 1982.

13. K. Chen, G. Zhang, J. Duster, C. Hu, J. Huang, Z. Liu and P. K. Ko, "MOSFET inversion layer capacitance model based on Fermi-Dirac statistics for wide temperature range," *Solid-State Electronics*, Vol. 41, pp. 507-509, 1997.