Modeling and characterization of quantization, polysilicon depletion, and direct tunneling effects in MOSFETs with ultrathin oxides

by S.-H. Lo
D. A. Buchanan
Y. Taur

The electrical characteristics (*C–V* and *I–V*) of n^+- and p^+-polysilicon-gated ultrathin-oxide capacitors and FETs were studied extensively to determine oxide thickness and to evaluate tunneling current. A quantum-mechanical model was developed to help understand finite inversion layer width, threshold voltage shift, and polysilicon gate depletion effects. It allows a consistent determination of the physical oxide thickness based on an excellent agreement between the measured and modeled *C–V* curves. With a chip standby power of ≤0.1 W per chip, direct tunneling current can be tolerated down to an oxide thickness of 15–20 Å. However, transconductance reduction due to polysilicon depletion and finite inversion layer width effects becomes more severe for thinner oxides. The quantum-mechanical model predicts higher threshold voltage than the classical model, and the difference increases with the electric field strength at the silicon/oxide interface.

Introduction

With the continuing application of device scaling, gate oxides below 30 Å will be needed for CMOS devices beyond 0.1 μm. In such small devices, the oxide field reaches a maximum of 5 MV/cm, while the field in silicon exceeds 1 MV/cm [1]. The operation of deep-submicron MOSFETs is now entering a regime in which quantum-mechanical effects become noticeable and classical physics

0018-8646/99/$5.00 © 1999 IBM

is no longer sufficient for accurate modeling of operating characteristics. The finite thickness of the inversion/accumulation layer (mostly due to quantum-mechanical effects) not only causes a discrepancy between the oxide capacitance and the measured capacitance but also degrades the transconductance [2, 3]. It has been shown that the inversion charge density calculated quantum-mechanically is smaller than that calculated classically for a given gate voltage, thus affecting the shift of the subthreshold curves [4]. As the surface electric field has increased with increased scaling, this has led to increased polysilicon depletion effects [5]. This depletion further reduces the gate capacitance and inversion charge density for a given gate bias. Furthermore, substantial electron tunneling through the gate insulator takes place even at operating bias conditions as low as 1–1.5 V. The gate leakage current increases exponentially as the oxide thickness is decreased. Consequently, future low-voltage circuits may operate with considerable gate-oxide tunneling [6]. With respect to the oxide degradation, hot electrons injected into the gate insulator result in a reduction of device reliability. Therefore, the tunneling characteristics of electrons through ultrathin oxides must be well understood if the technology is to be better understood and controlled.

In this paper, a full quantum-mechanical scheme is developed in order to study the effects described above for ultrathin-oxide MOS structures. The Poisson and the effective-mass Schrödinger equations are solved self-consistently. A transverse-resonant method [7] is utilized to calculate the lifetime associated with the quasi-bound states in the electron accumulation and inversion layers. The tunneling probability associated with the continuum states is also taken into account. Simulated results are compared with experimental results for verification.

Physical models and numerical techniques

- *Schrödinger's and Poisson's equations*

When energy bands are bent strongly near a semiconductor–insulator interface, the potential well formed by the interface barrier and the electrostatic potential in the semiconductor can be narrow enough that quantum-mechanical effects become important. Only a given carrier type is treated quantum-mechanically when confined by the surface potential. When electrons are confined, the electrical characteristics of an MOS structure are modeled by solving the coupled effective-mass Schrödinger (in the oxide and silicon regions) and Poisson equations (in the polysilicon, oxide, and silicon regions) self-consistently without taking tunneling current into account [8, 9]:

$$\left[-\frac{\hbar^2}{2}\frac{d}{dz}\frac{1}{m_i^*(z)}\frac{d}{dz}+V(z)-E_{i,j}\right]\psi_{i,j}(z)=0 \tag{1}$$

and

$$\frac{d}{dz}\left[\varepsilon(z)\frac{d}{dz}\right]\phi(z)=\frac{q}{\varepsilon_0}[n(z)-p(z)+N_A^- - N_D^+], \tag{2}$$

where z is the coordinate along the perpendicular direction, m_i^* is the electron effective mass in the ith valley, $\psi_{i,j}$ is the envelope wave function for the jth subband in the ith valley, V is the potential energy, and N_A^- and N_D^+ are respectively the ionized acceptor concentration and donor concentration. The potential energy $V(z)$ in Equation (1) is related to the electrostatic potential $\phi(z)$ in Equation (2) as follows:

$$V(z)=-q\phi(z)+\Delta E_C(z), \tag{3}$$

where $\Delta E_C(z)$ is the pseudopotential energy due to the band offset at the Si/SiO_2 interface. The wave function $\psi(z)$ in Equation (1) and the electron density $n(z)$ (Si and SiO_2 regions) in Equation (2) are related by

$$n(z)=\frac{k_BT}{\pi\hbar^2}\sum_i g_i m_{di}^*\sum_j \ln[1+e^{(E_F-E_{i,j})/k_BT}]\psi_{i,j}^2(z), \tag{4}$$

where g_i and m_{di}^* are the ith valley degeneracy and the ith density-of-states effective mass. The basic equations describing hole quantization are similar to Equations (1)–(4). For simplicity, the light and heavy hole bands are also modeled as parabolic, with $m_{h,l}^* = 0.20m_0$ and $m_{h,h}^* = 0.29m_0$ [10].

The dielectric constant $\varepsilon(z)$ is 11.7 for Si and polysilicon, and 3.9 for SiO_2. The boundary condition at the polysilicon/SiO_2 interface sets electron or hole wave functions to zero; i.e., there is no tunneling current. The electron wave-function solution obtained here is a starting approximation for the next-step electron tunneling current calculation, and the zero-wave-function assumption at the polysilicon/SiO_2 interface will be removed. To accurately model the polysilicon depletion effect, the free-carrier density in the polysilicon region is described by Fermi–Dirac statistics and an incomplete ionization model [11]. The charge-balance equation for heavily doped n-type polysilicon is

$$\frac{2}{\sqrt{\pi}}N_C F_{1/2}(\eta_f)=\frac{N_D}{1+2e^{\eta_f}}, \tag{5}$$

where N_C is the effective density of states at the conduction-band edge, $F_{1/2}(\eta_f)$ is the Fermi–Dirac integral, and $\eta_f = (E_F - E_C)/k_BT$. Analogous reasoning applies to heavily doped p-type polysilicon.

• *Electron tunneling from quasi-bound states*

Many models for the current–voltage characteristics of MOS devices in the direct and the F–N (Fowler–Nordheim) tunneling regimes are available [12–14]. For a two-dimensional gas system of electrons, the tunneling probability, which is applicable only for an incident Fermi gas system of electrons, may no longer be a valid concept. Weinberg [12] proposed a model based on the triangular well approximation and electrons being confined within the lowest subband. However, the contribution from higher subbands was neglected, which might not be entirely correct for the higher temperature. Rana et al. [15] calculated the lifetimes associated with the quasi-bound bands using a path integral expansion of the resolvent operator, which gave quite good results for electron tunneling from an accumulation layer.

For the calculation of the tunneling current from the quasi-bound states at a given bias, the conduction-band profile and the discrete sheet charge densities, previously obtained by assuming $\psi_{i,j} = 0$ at the polysilicon/SiO_2 interface, are assumed to be unchanged; i.e., the inversion and accumulation layers act as a reservoir within which the electrons are all in equilibrium with a constant temperature and Fermi level. The closed boundary condition at polysilicon/SiO_2, i.e., $\psi_{i,j} = 0$ (assumed in the previous calculation), is relaxed such that the electron wave function is no longer nonzero in the polysilicon gate region. Only the Schrödinger equation throughout the MOS structure (including the polysilicon region) is solved again.

The close analogy between these confined electrons in a varying potential and electromagnetic waves in a wave guide with a varying refractive index provides for the utilization of the transverse-resonant method [7], which is often used for finding the eigenvalue equation for inhomogeneously filled wave guides and dielectric resonators. Consider the conduction-band edge profile shown in **Figure 1**. The transverse-resonant method defines the intrinsic impedance $\eta_l = m_l^*/k_l$ and $\tilde{Z}_l$ and $\tilde{\tilde{Z}}_l$, the terminal impedances for any interface l in which m^* is the electron effective mass, k is the wave number, and the arrow superscripts denote looking to the left or looking to the right for any given position. The boundary conditions at interface 1 and interface $N - 1$ imply that $\tilde{Z}_1 = \eta_1$ and $\tilde{Z}_{N-1} = \eta_N$. Let region i be approximately the region with the highest electron charge density. We repeatedly apply the transmission-line transformation to express $\tilde{Z}_i$ in terms of $\tilde{Z}_1, \tilde{Z}_2, \cdots$, and $\tilde{Z}_{i-1}$, and similarly for $\tilde{\tilde{Z}}_i$ in terms of $\tilde{\tilde{Z}}_{N-1}, \tilde{\tilde{Z}}_{N-2}, \cdots, \tilde{\tilde{Z}}_{i+1}$; i.e.,

$$\tilde{Z}_m = \eta_m \frac{\tilde{Z}_{m-1} - j\eta_m \tan(k_m d_m)}{\eta_m - j\tilde{Z}_{m-1} \tan(k_m d_m)} \qquad \text{for } m = 2, 3, \cdots, i. \tag{6}$$

and

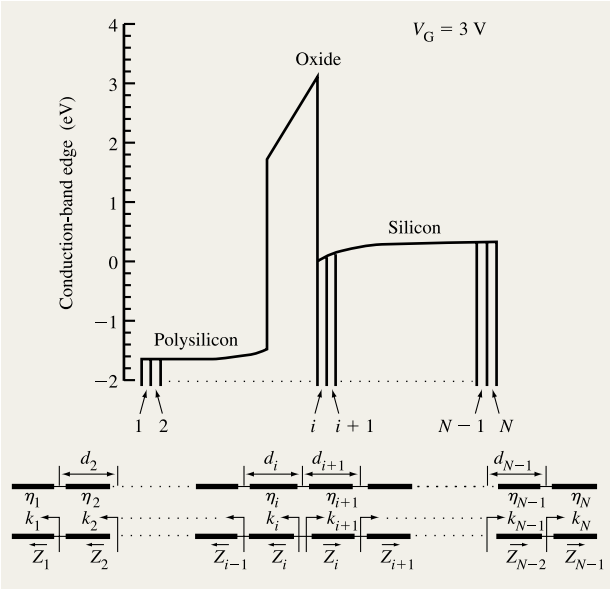

Figure 1

Illustrations for the electron tunneling current analysis using the transverse-resonant method. The conduction-band edge profile $E_C(z)$ and its transmission-line network representation are shown.

$$\tilde{\tilde{Z}}_m = \eta_{m+1} \frac{\tilde{\tilde{Z}}_{m+1} - j\eta_{m+1} \tan(k_{m+1} d_{m+1})}{\eta_{m+1} - j\tilde{\tilde{Z}}_{m+1} \tan(k_{m+1} d_{m+1})}$$

$$\text{for } m = N - 2, N - 3, \cdots, i. \tag{7}$$

The final eigenvalue equation for the leaky quasi-bound state follows by imposing the resonance condition

$$\tilde{Z}_i + \tilde{\tilde{Z}}_i = 0; \tag{8}$$

i.e., the input impedance looking to the right must be equal to the negative value of the input impedance seen at this point looking to the left. As a result, we obtain the complex eigenenergy $E_i = E_{iR} - j\Gamma_i$, where the resonance width Γ_i is related to the lifetime of the ith quasi-bound state by

$$\tau_i = \frac{\hbar}{2\Gamma_i}. \tag{9}$$

The tunneling current density from the ith quasi-bound state is given by

$$J_i = \frac{Q_i}{\tau_i}, \tag{10}$$

where Q_i is the sheet charge density associated with the ith quasi-bound state.

In the case of three-dimensional (3D) states, the transmission probability through the SiO_2 barrier is a well-defined concept and has a value equal to the ratio of

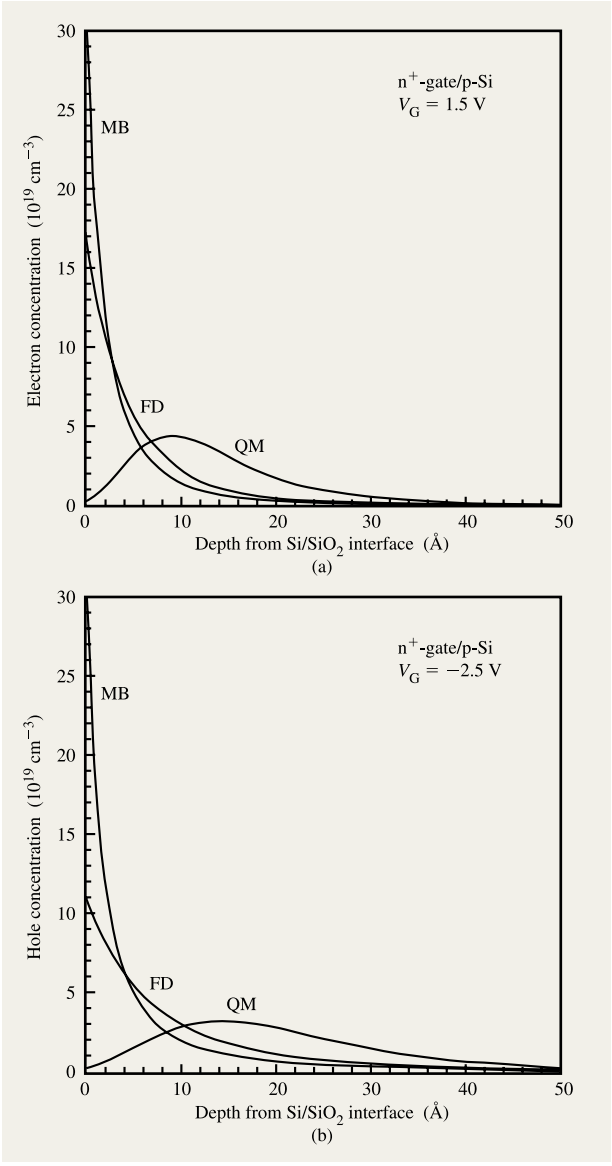

Figure 2

(a) Electron and (b) hole distributions calculated on the basis of classical [Maxwell–Boltzmann (MB) and Fermi–Dirac (FD) statistics] and quantum-mechanical (QM) models. The temperature is 25°C, the substrate concentration is 5×10^{17} cm^{-3}, and the oxide thickness is 25 Å.

transmitted and incident fluxes. The current density from the 3D states is given [15] by

$$J_{3D}(E_z)dE_z = \sum_i \frac{qD_i m_i^* k_B T}{2\pi^2 \hbar^3} \times T_i(E_z) \times \ln\left[1 + \exp\left(\frac{E_F - E_z}{k_B T}\right)\right] dE_z, \quad (11)$$

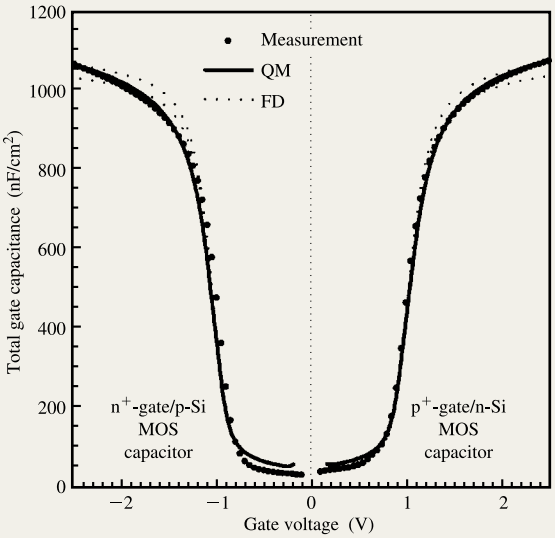

Figure 3

Measured and simulated C–V curves of an n$^+$-gate/p-Si and a p$^+$-gate/n-Si device ($t_{ox} \approx 25$ Å). The quantum-mechanical model (QM) and Fermi–Dirac statistics (FD) are compared. For the FD model, 27.7-Å and 28.7-Å oxides are assumed for the n$^+$-gate/p-Si device, and 27.1-Å and 28.1-Å oxides for the p$^+$-gate/n-Si device. The two classical curves fail to match the entire range of C–V characteristics regardless of the oxide thickness assumed.

where m_i^* is the effective mass, D_i the valley degeneracy associated with the direction i along or across the symmetry axes, k_B the Boltzmann constant, and E_F the Fermi level. The integral has been taken over all states with the same energy E_z in the direction normal to the Si/SiO$_2$ interface. The transmission probability $T_i(E_z)$ is calculated using the transfer matrix approach [16].

The energy-band structure of SiO$_2$ is modeled using a Franz-type $E(k)$ dispersion [13] with the effective mass at the conduction-band edge, $m^* = 0.58m_0$ and $0.55m_0$, in the accumulation and the inversion regions, respectively. The barrier height due to conduction-band discontinuity at the Si/SiO$_2$ interface is taken to be 3.15 eV. The barrier lowering due to the image force and the conservation of transverse crystal momentum is neglected because it is questionable whether they have any significant effect in the direct tunneling regime of interest here and because none of the existing theories can be readily applied [12].

SiO$_2$/Si interface quantization effects

Figures 2(a) and **2(b)** respectively show the electron and hole distributions in an n$^+$-gate/p-Si MOS device. They are calculated using classical [Maxwell–Boltzmann (MB), Fermi–Dirac (FD) statistics] and quantum-mechanical

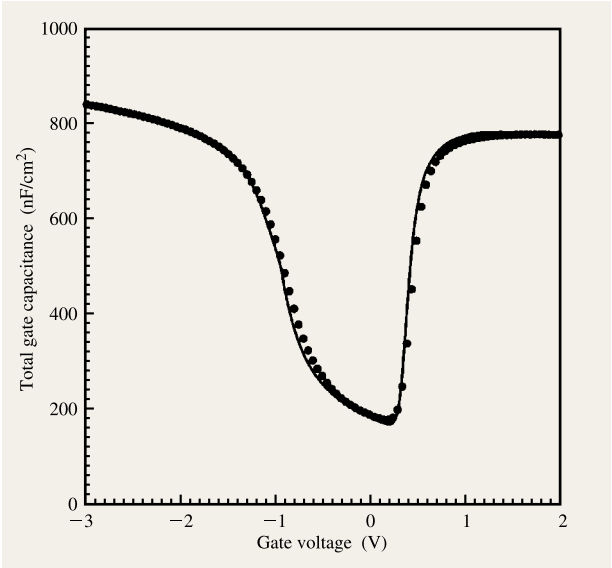

Figure 4

Quantum-mechanically simulated (line) and measured (dots) full *C–V* characteristics of an n$^+$-gate n-FET device on IBM 1.8-V and 0.20-μm technology. The oxide thickness is 35.1 Å.

(QM) models. Both the effective inversion and accumulation layer thicknesses in Si are increased by several angstroms owing to quantization effects [8, 9].

• *Oxide thickness determination from* C–V

Capacitance–voltage (*C–V*) characteristics have been used extensively to extrapolate the oxide thickness of MOS devices [17–20]. Thin-oxide MOS capacitance in the strong accumulation region does not saturate but rather increases with gate voltage, as it would for thicker dielectric film. This is partly due to the quantization effects causing the change of the effective thickness of the silicon surface accumulation layer with the gate voltage [see Figure 2(b)] for n$^+$-gate/p-Si MOS devices. **Figure 3** shows the agreement between the quantum-mechanically calculated and the measured *C–V* curves of an n$^+$-gate/p-Si and a p$^+$-gate/n-Si MOS capacitor ($t_{ox} \approx 25$ Å). The classical model fails to match the entire range of *C–V* characteristics regardless of the oxide thickness used. The quantum-mechanical model has proved itself able to fit a full *C–V* curve of an n-type FET on IBM CMOS 1.8-V, 0.2-μm technology very well, as shown in **Figure 4**.

On the basis of the quantum-mechanical simulation and experimentally, it was found that the gate capacitance of an n$^+$ (p$^+$)-gate/p (n)-Si capacitor in the strong accumulation region (typically for an oxide field $\geq$3 MV/cm) is insensitive to both the polysilicon and the substrate doping concentrations (**Figures 5** and **6**),

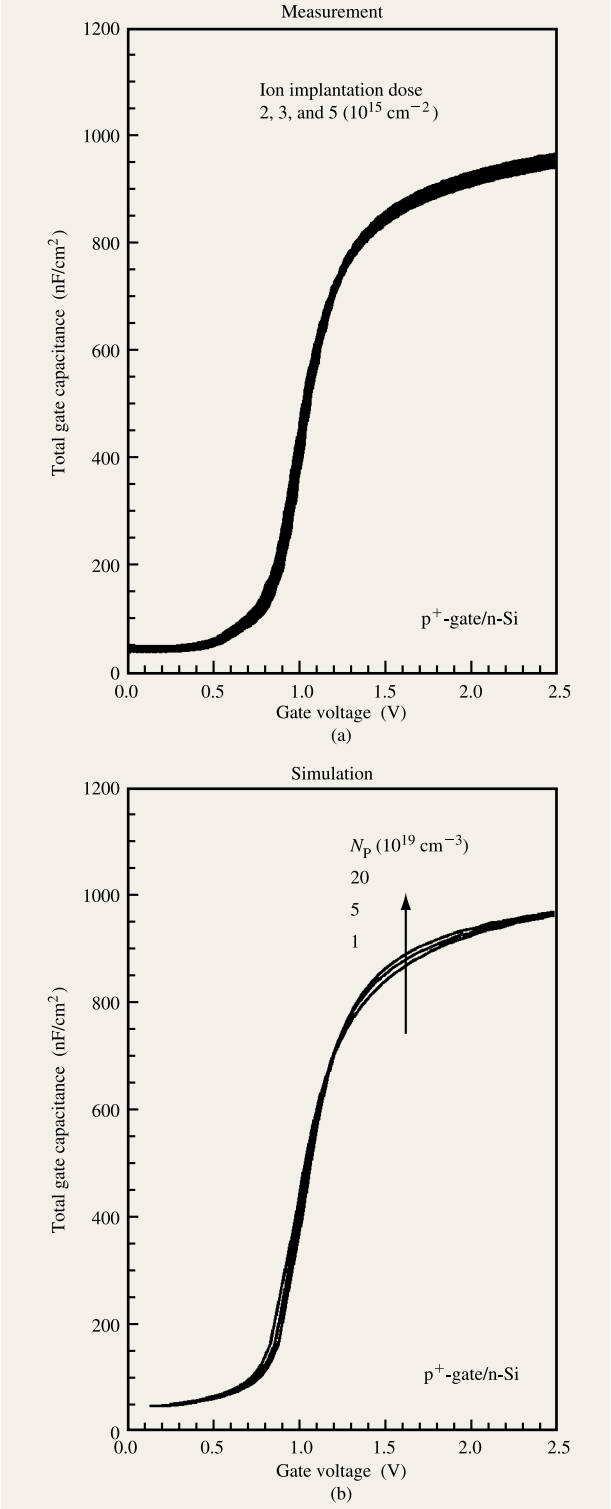

Figure 5

Sensitivity of *C–V* to the polysilicon doping concentration (N_P): (a) experimentally and (b) theoretically. The oxide thickness is 28.7 Å.

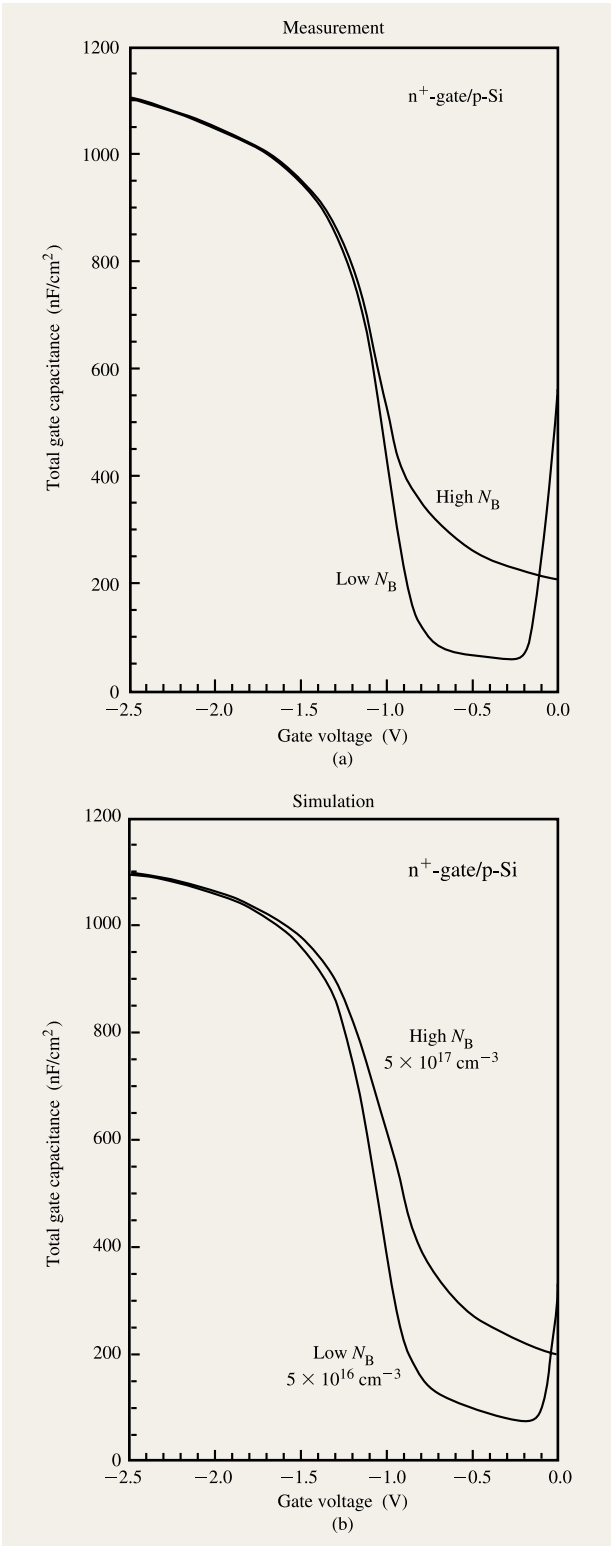

Figure 6

Sensitivity of *C*–*V* to the substrate doping concentration (N_B): (a) experimentally and (b) theoretically. The oxide thickness is 24.9 Å.

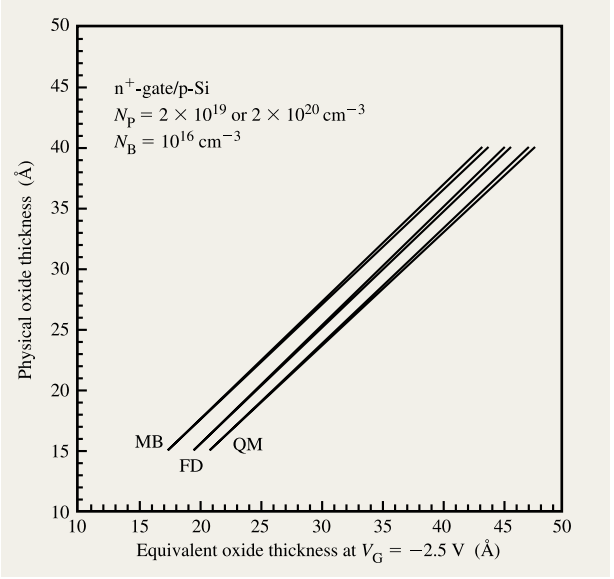

Figure 7

Quantum-mechanically (QM) calculated t_{ox}–t_{eq} curves for n^+-gate/p-Si MOS devices. Two other groups are based on the classical model with Fermi–Dirac (FD) and Maxwell–Boltzmann (MB) statistics, respectively, for comparison. No substrate doping dependence was observed for concentrations between 10^{16} and 10^{18} cm^{-3}.

provided that both the substrate and polysilicon are accumulated under the same bias polarity. Therefore, a well-defined QM calculated oxide thickness can be extracted from one measured point in a strong accumulation region, even when the other device parameters are unknown. The QM calculated oxide thickness can be determined from the precalculated oxide thickness versus the measured equivalent oxide thickness curve, as shown in **Figure 7**. The figure clearly shows that a well-defined oxide thickness can be accurately extracted from the curve independent of polysilicon and substrate doping levels. If the classical models, Maxwell–Boltzmann and Fermi–Dirac statistics, are used instead, the oxide thickness will be overestimated by 2–4 Å as compared with the quantum-mechanical model. Optical measurements have been conducted to verify this extraction scheme. **Figure 8** shows that the oxide thicknesses as determined using the above method are within 1 Å of those measured by ellipsometry with multiple measurement angles (wavelength = 6328 Å and index of refraction = 1.458).

- *Threshold voltage shift with high surface field*

In the quantum-mechanical treatment, carriers in the inversion layer not only are distributed away from the

surface, but also occupy discrete subband energy levels. The inversion carrier sheet densities calculated both quantum-mechanically and classically (FD) for three impurity concentrations (10^{17}, 5×10^{17}, and 10^{18} cm^{-3}) as a function of gate voltage at room temperature are shown in **Figure 9**. Since the two lowest subbands (longitudinal and transverse modes) are at finite energies above the bottom of the conduction band, more band bending or a larger surface potential is required to populate the inversion layer. This has the effect of shifting the threshold voltage to a higher value, particularly with high-doping substrate. As the substrate doping increases, the two lowest subband energies increase because of the stronger surface field, and this results in an increasing threshold voltage shift. This shift can be as large as 0.1 V when the substrate impurity concentration reaches 10^{18} cm^{-3}. The threshold voltage shift also increases with the oxide thickness [1, 21].

High substrate doping is needed to suppress the short-channel effects for devices beyond 0.1 μm. It becomes difficult to design a channel profile that controls the short channel while still having a V_T low enough to be compatible with an ~1-V power supply. Therefore, the quantum-effect-induced threshold voltage shift becomes critical in designing deep-submicron devices in which substrate doping $>10^{17}$ cm^{-3} and surface field $>10^5$ V/cm at threshold voltage are required [1].

Finite inversion layer width and polysilicon depletion

Figures 10(a) and **10(b)** respectively show the simulated conduction-band edge and the free-electron concentration profiles and *C–V* characteristics in the inversion region for an n-type MOSFET. The oxide thickness is 25 Å, and the active polysilicon doping concentration is assumed to be 5×10^{19} cm^{-3}. The peak electron concentration in the silicon is about 10 Å away from the silicon/oxide interface. Besides, the surface field is very strong, so that there is a 30-Å-wide depletion region around the polysilicon/oxide interface. The two finite capacitances are in series with the oxide capacitance. The difference between the total capacitance and the oxide capacitance due to the two effects becomes more significant with decreasing oxide thickness. Polysilicon depletion effects are directly related to the high fields as well as insufficient activation of dopants at the polysilicon/oxide interface. Even if the polysilicon gate is assumed to behave like a metal gate, i.e., with no polysilicon depletion effect, the difference due to the finite width of the inversion layer only is still 15%. **Figure 11** shows the quantum-mechanically simulated C_G/C_{ox} ratio at $V_G = 1.5$ V versus polysilicon doping concentration for oxide thicknesses ranging from 15 to 35 Å. Note that as the gate oxide is made thinner, the

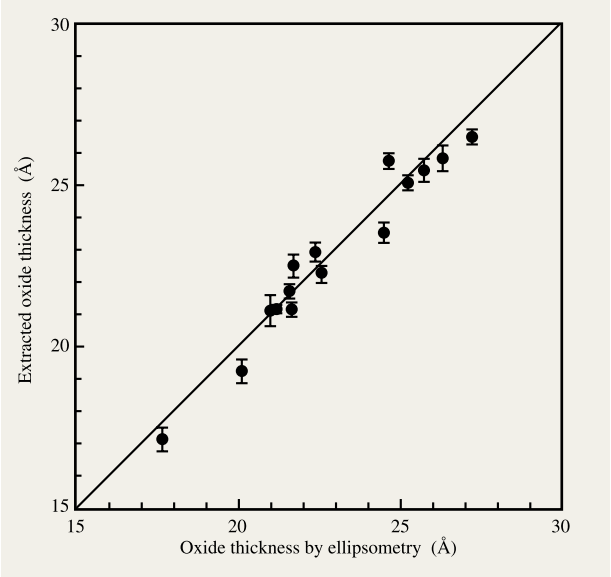

Figure 8

Comparison between *C–V*-extracted oxide thickness and thickness measured by ellipsometry. Every ellipsometry thickness represents the average of nine points and every *C–V*-extracted thickness represents the average of 64 capacitors (area = 10^{-4} cm^2) over an 8-in. wafer.

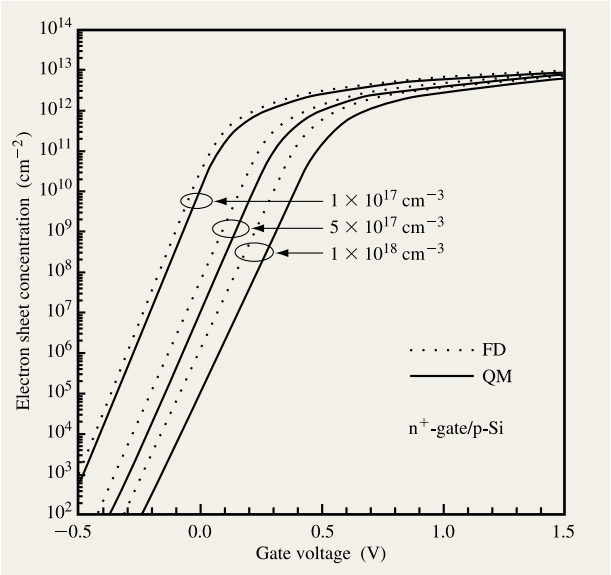

Figure 9

Inversion carrier density calculated both quantum-mechanically (QM) and classically (FD) for three impurity concentrations (10^{17}, 5×10^{17}, and 10^{18} cm^{-3}) as a function of gate voltage at room temperature. The oxide thickness is 25 Å.

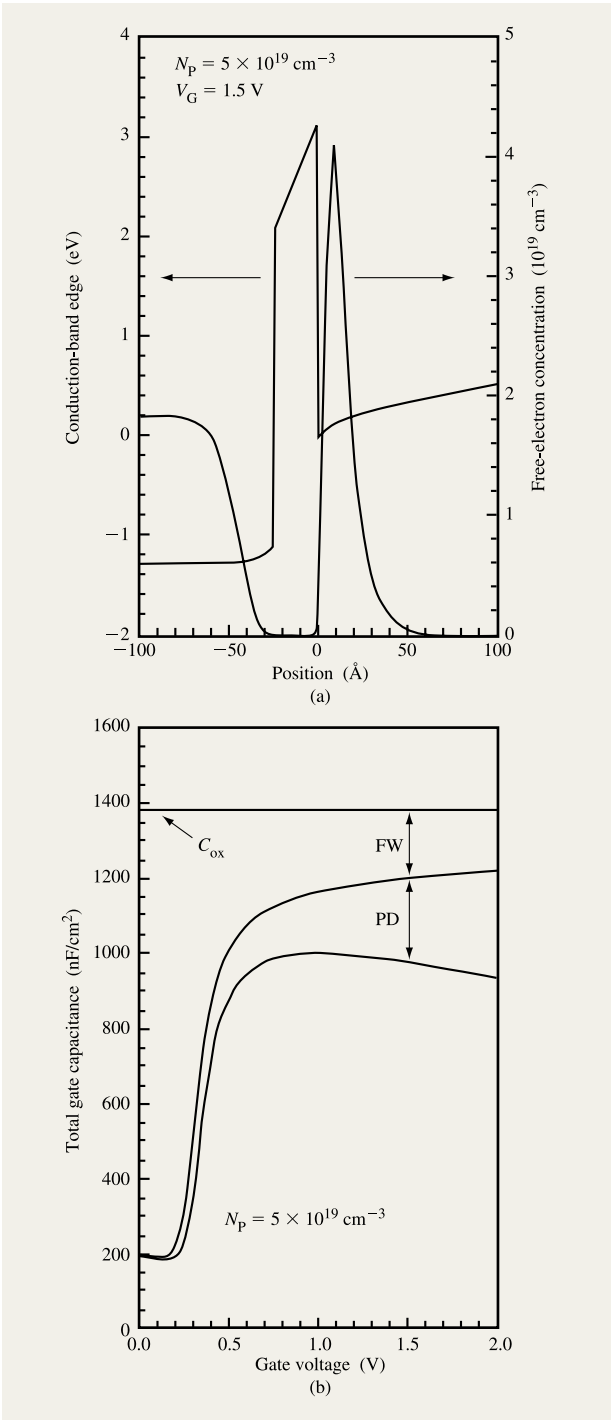

Figure 10

(a) Simulated conduction-band edge and free-electron concentration profiles and (b) finite-width (FW) and polysilicon depletion (PD) effects on *C–V* characteristics of an n-FET in the inversion region. The free-electron concentration in the polysilicon far from the interface is lower than the impurity doping concentration because of the incomplete ionization effect assumed in the simulation. The oxide thickness is 25 Å.

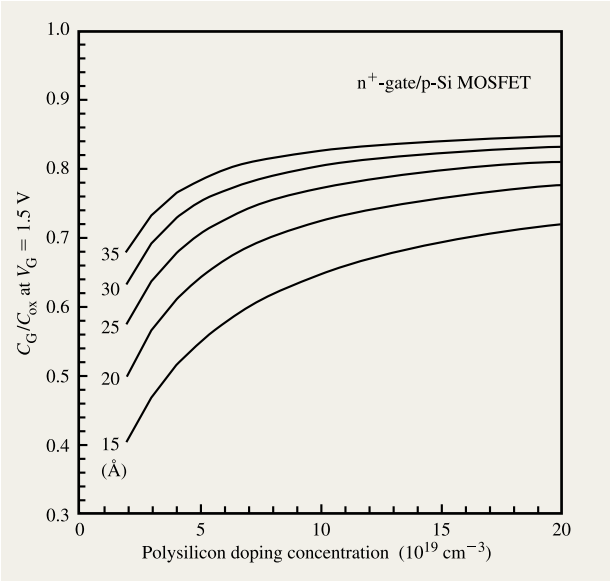

Figure 11

Calculated ratio of total inversion capacitance to oxide capacitance at 1.5 V vs. polysilicon doping concentration with oxide thickness as a parameter (15–35 Å). Note that the capacitance is calculated in the inversion region, not the accumulation region.

C_G/C_{ox} calculated in the inversion region becomes more sensitive to the polysilicon doping concentration. As shown in Figure 5, the results for a given oxide thickness are less sensitive to the polysilicon doping concentration in the accumulation region. This effect has also been studied experimentally using high-frequency *C–V* curves of both n$^+$-polysilicon/n-Si and p$^+$-polysilicon/p-Si MOS capacitors for different rapid thermal anneal (RTA) times at 1000°C in Ar, as shown in **Figures 12** and **13**. The 150-nm polysilicon was implanted to a dose of 3×10^{15} cm^{-2} with P^- at 15 keV for n$^+$-polysilicon MOS devices, and B^+ at 5 keV for p$^+$-polysilicon MOS devices. A significant degradation due to polysilicon depletion is found with even ten seconds annealing time. For a 26-Å oxide, the normalized gate capacitance C_G/C_{ox} of the n$^+$-polysilicon MOS device is only 0.7, as opposed to 0.87 expected with no polysilicon depletion. This corresponds to an active phosphorus concentration of only 5×10^{19} cm^{-3} at the polysilicon/SiO_2 interface. The p$^+$-polysilicon MOS has about the same active doping concentration at the polysilicon/SiO_2 interface.

Gate direct tunneling current from quantization states

The gate tunneling current, when electrons are confined in the inversion or accumulation layers, is calculated using a

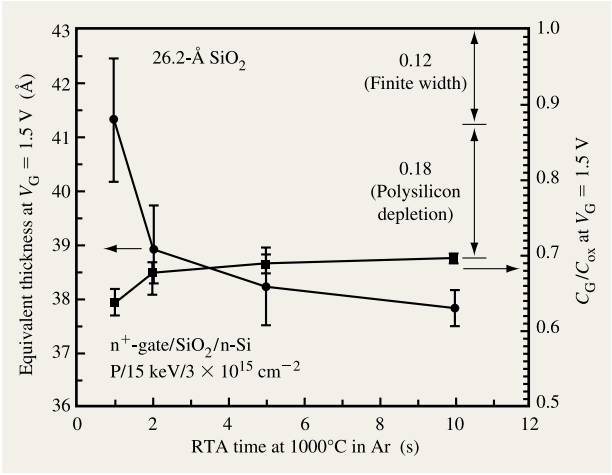

Figure 12

Equivalent thickness and C_G/C_{ox} of n^+-gate/n-Si capacitors for different post-implant RTA times at 1000°C in Ar. The O_2-grown oxides are ~26 Å thick. An 18% reduction in C_G/C_{ox} is due to polysilicon depletion even with ten seconds annealing time.

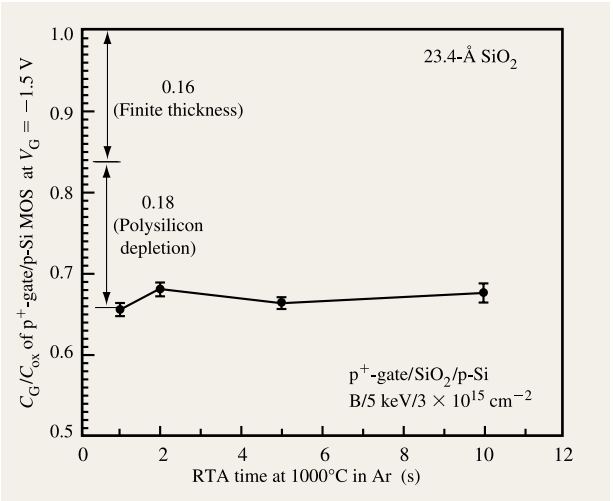

Figure 13

Measured C_G/C_{ox} for p^+-gate/p-Si capacitors. The O_2-grown oxides are ~23 Å. An 18.3% reduction in C_G/C_{ox} is due to the polysilicon depletion effect.

transverse-resonant method described in Section 2. The physical oxide thickness is determined using the QM scheme mentioned previously. After the oxide thickness is determined, the polysilicon and the substrate doping concentrations, respectively, can be extracted by fitting the measured C–V curve in the depletion and the inversion regions.

• *Electron tunneling in accumulation and inversion regimes*

Figure 14 shows a comparison between the calculated and measured gate tunneling current characteristics of seven p^+-gate/n-Si p-MOS capacitors (t_{ox} = 22.9–41.8 Å) in the electron-accumulation region. As the electron accumulation is strong enough, the tunneling current from the free-electron states (3D) is negligible compared with that from quasi-bound states (2D). **Figure 15** also shows the excellent agreement between the calculated and the measured I_G–V_G characteristics of six n^+-gate/p-Si n-FETs (t_{ox} = 21.9–36.1 Å). More than 90% of the current density comes from the lowest two subbands associated with the twofold-degenerate and the fourfold-degenerate valleys in the conduction band [22]. Simulated I_G–V_G characteristics for oxides down to 15 Å are also shown in **Figure 15**. In the inversion region, the current density at a gate bias of 1.5 V increases by ten orders of magnitude as the oxide thickness decreases from 36 Å to 15 Å. The tunneling current density calculated for a 15-Å oxide agrees reasonably with the reported experimental data [6].

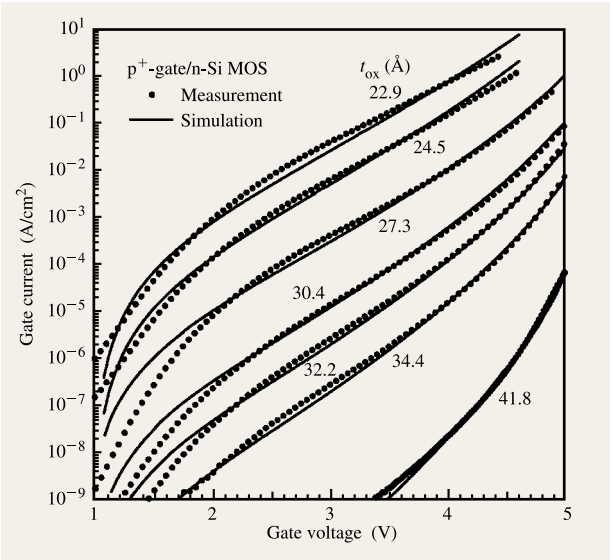

Figure 14

Measured and simulated I_G–V_G characteristics under accumulation conditions of p^+-gate/n-Si MOS devices with oxides ranging from 22.9 to 41.8 Å. The thickness is determined using the QM scheme.

• *Standby power consumption*

While the gate leakage current may be at a negligible level compared with the on-state current of a Si MOSFET, it

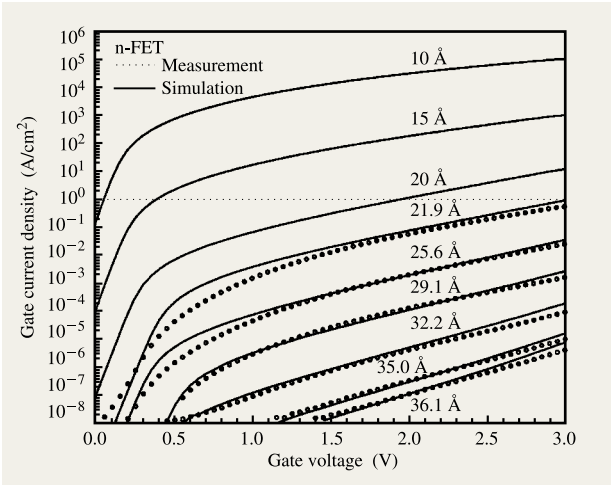

Figure 15

Measured and simulated I_G–V_G characteristics under inversion conditions of n-FETs with oxides ranging from 21.9 to 36.1 Å. The thickness is determined using the QM scheme. The dotted line indicates the 1-A/cm^2 limit for leakage current, as discussed in the text.

will first have an effect on the chip standby power. Note that the tunneling leakage will be dominated by n-MOS, since p-MOS has a higher barrier for hole tunneling and therefore a lower leakage current. High-performance CMOS logic chips can tolerate a standby power in the 100-mW range. If one assumes that the total active gate area per chip is of the order of 0.1 cm^2 for future-generation technologies, the maximum tolerable gate current would be of the order of 1 A/cm^2 for a power supply of 1 V. At higher temperatures (105°C), the tunneling current should be at least ten times lower than the short-channel leakage, since the latter increases rapidly with temperature while the former does not. From Figure 15, it is projected that gate oxide can be scaled to 15–20 Å before running into such a limit. Below 15 Å, however, the oxide tunneling current quickly becomes problematic. Unless a new gate dielectric material is discovered, this sets a limit for bulk CMOS scaling. Dynamic memory devices have a more stringent leakage requirement and therefore a thicker oxide limit.

Conclusions

In this paper, a total quantum-mechanical treatment of accumulated and inverted silicon layers and the electron tunneling current of ultrathin-oxide MOS structures has been presented. A quantum-mechanical scheme is proposed to accurately determine the physical thickness of ultrathin oxides. As the gate length is scaled ≤ 0.1 μm, where substrate doping concentration $>10^{17}$ cm^{-3} or surface field $>10^5$ V/cm are required, the shift of the threshold voltage due to the surface quantization becomes substantial. The finite thickness effects degrade the gate capacitance by 13% or more for an oxide thickness of 25 Å or less. Polysilicon depletion effects significantly degrade the capacitance and therefore the transconductance of CMOS devices. It is projected that gate oxides can be scaled down to 15–20 Å in terms of the chip standby power consumption due to direct tunneling currents.

Acknowledgments

The authors wish to thank F. Stern, S. Tiwari, M. Fischetti, K. Han, and E. Wu for many stimulating discussions. They also wish to thank the Yorktown Heights, New York, Silicon Facility and the Advanced Silicon Technology Center (ASTC) in East Fishkill, New York, for device fabrication.

References

1. Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S.-H. Lo, G. A. Sai-Halasz, R. G. Viswanathan, H.-J. C. Wann, S. J. Wind, and H.-S. Wong, "CMOS Scaling into the Nanometer Regime," *Proc. IEEE* **85,** No. 4, 486–504 (1997).
2. G. Baccarani and M. R. Wordeman, "Transconductance Degradation in Thin-Oxide MOSFET's," *IEEE Trans. Electron Devices* **30,** No. 10, 1295–1304 (1983).
3. K. S. Krisch, J. D. Bude, and L. Manchanda, "Gate Capacitance Attenuation in MOS Devices with Thin Gate Dielectrics," *IEEE Electron Device Lett.* **17,** 521–524 (1996).
4. Y. Ohkura, "Quantum Effects in Si n-MOS Inversion Layer at High Substrate Concentration," *Solid-State Electron.* **33,** No. 12, 1581–1585 (1990).
5. C. Y. Wong, J. Y.-C. Sun, Y. Taur, C. S. Oh, R. Angelucci, and B. Bavari, "Doping of n^+ and p^+ Polysilicon in a Dual-Gate Process," *IEDM Tech. Digest*, pp. 238–241 (1988).
6. H. S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Saito, and Hiroshi Iwai, "Tunneling Gate Oxide Approach to Ultra-High Current Drive in Small-Geometry MOSFETs," *IEDM Tech. Digest*, pp. 593–596 (1994).
7. R. E. Collin, *Field Theory of Guided Waves*, 2nd Ed., IEEE Press, New York, 1991.
8. F. Stern, "Self-Consistent Results for n-type Si Inversion Layers," *Phys. Rev. B* **5,** No. 12, 4891–4899 (1972).
9. J. Suñé, P. Olivo, and B. Riccò, "Quantum-Mechanical Modeling of Accumulation Layers in MOS Structure," *IEEE Trans. Electron Devices* **39,** 1732–1739 (1992).
10. C. Moglestue, "Self-Consistent Calculation of Electron and Hole Inversion Charges at Silicon–Silicon Dioxide Interfaces," *J. Appl. Phys.* **59,** No. 9, 3175–3183 (1986).
11. Sorab K. Ghandhi, *The Theory and Practice of Microelectronics*, John Wiley and Sons, Inc., New York, 1968.
12. Z. A. Weinberg, "On Tunneling in Metal–Oxide Silicon Structures," *J. Appl. Phys.* **53,** No. 7, 5052–5056 (1982).
13. J. Maserjian, "Tunneling in Thin MOS Structures," *J. Vac. Sci. Technol.* **11,** No. 6, 996–1003 (1974).
14. J. G. Simmons, "Generalized Formula for the Electric Tunneling Effect Between Similar Electrodes Separated by a Thin Insulating Film," *J. Appl. Phys.* **34,** No. 6, 1793–1803 (1963).

15. F. Rana, S. Tiwari, and D. A. Buchanan, "Self-Consistent Modeling of Accumulation Layers and Tunneling Currents Through Very Thin Oxides," *Appl. Phys. Lett.* **69,** No. 8, 1104–1106 (1996).
16. S. Datta, *Quantum Phenomena*, Addison-Wesley Publishing Co., Reading, MA, 1989.
17. M. J. McNutt and C. T. Sah, "Determination of the MOS Oxide Capacitance," *J. Appl. Phys.* **46,** No. 9, 3909–3913 (1975).
18. K. Lehovec and S.-T. Lin, "Analysis of C–V Data in the Accumulation Regime of MIS Structures," *Solid-State Electron.* **19,** 993–996 (1976).
19. B. Ricco, P. Olivo, T. N. Nguyen, T.-S. Kuan, and G. Ferriani, "Oxide-Thickness Determination in Thin-Insulator MOS Structures," *IEEE Trans. Electron Devices* **35,** No. 4, 432–438 (1988).
20. R. Rios and N. D. Arora, "Determination of Ultra-Thin Gate Oxide Thicknesses for CMOS Structures Using Quantum Effects," *IEDM Tech. Digest*, pp. 613–616 (1994).
21. M. J. van Dort, Pierre H. Woerlee, A. J. Walker, 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* **39,** No. 4, 932–938 (1992).
22. S.-H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, "Quantum-Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFET's," *IEEE Electron Device Lett.* **18,** No. 5, 209–211 (1997).

Received September 2, 1998; accepted for publication January 29, 1999

Shih-Hsien Lo *IBM Microelectronics Division, East Fishkill facility, Hopewell Junction, New York 12533 (shlo@us.ibm.com).* Dr. Lo received his B.S. degree in electrical engineering from the National Cheng-Kung University in 1986 and his M.S. and Ph.D. degrees in electronics from the National Chiao-Tung University in 1988 and 1991, respectively. From July 1991 to May 1993, he served in the military as a Second Lieutenant. From 1993 to 1995, he was with the National Nano Device Laboratories, Taiwan, where he was a Research Scientist working on 0.25-μm CMOS device design and process integration. In 1995 Dr. Lo was recruited by the Exploratory Devices and Circuits group of the Silicon Technology Department at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York. His researches focused on the areas of ultrathin oxide reliability characterization and modeling, and electrical characterization of Si-based small geometry devices. Since 1997, Dr. Lo has been with the IBM Semiconductor Research and Development Center, East Fishkill, New York. He first joined the CMOS 7S team working on sub-0.25-μm, 1.8-V CMOS technology development, which first offered the Cu BEOL in the industry. He is currently working on 0.18-μm, 1.5-V CMOS 8S2 device design and characterization. Dr. Lo is a member of the IEEE. He has received the IBM Microelectronics General Manager's Excellence Award three times, in 1997, 1998, and 1999.

Douglas A. Buchanan *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (dabuchan@us.ibm.com).* Dr. Buchanan received his B.Sc. and M.Sc. degrees in electrical engineering from the University of Manitoba, Winnipeg, Canada, in 1981 and 1982, respectively. In 1986 he received his Ph.D. from Durham University, Durham, England. Following a two-year postdoctoral fellowship at the IBM Thomas J. Watson Research Center, he spent three years in the CVD thin-film technology group in the IBM Microelectronics Division. Currently Dr. Buchanan works at the Thomas J. Watson Research Center on issues that relate to the growth, characterization, and integration of ultrathin and high dielectrics.

Yuan Taur *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (taur@us.ibm.com).* Dr. Taur joined IBM Research at Yorktown Heights as a Research Staff Member in 1981. He received the B.S. degree in physics from National Taiwan University, Taipei, Taiwan, in 1967 and the Ph.D. degree in physics from the University of California, Berkeley, in 1974. Since 1981 Dr. Taur has been with the Silicon Technology Department of the IBM Thomas J. Watson Research Center, Yorktown Heights, New York. His research activities include latchup-free 1-μm CMOS, self-aligned $TiSi_2$, 0.5-μm CMOS and BiCMOS, shallow-trench isolation, 0.25-μm CMOS with n^+/p^+ poly gates, SOI, low-temperature CMOS, and 0.1-μm CMOS. Dr. Taur has authored or coauthored more than 100 technical papers; he holds nine U.S. patents. Over his IBM career, he has received four Outstanding Technical Achievement Awards and Invention Achievement Awards. Dr. Taur was an editor of the *IEEE Electron Device Letters* from 1996 to 1998; he was elected a Fellow of the IEEE in 1998.