MOS *C–V* Characterization of Ultrathin Gate Oxide Thickness (1.3–1.8 nm)

Chang-Hoon Choi, Jung-Suk Goo, Tae-Young Oh, Zhiping Yu, Robert W. Dutton, Amr Bayoumi, Min Cao, Paul Vande Voorde, Dieter Vook, and C. H. Diaz

***Abstract*— An equivalent circuit approach to MOS capacitance–voltage (C–V) modeling of ultrathin gate oxides (1.3–1.8 nm) is proposed. Capacitance simulation including polysilicon depletion is based on quantum mechanical (QM) corrections implemented in a two-dimensional (2-D) device simulator; tunneling current is calculated using a one-dimensional (1-D) Green's function solver. The sharp decrease in capacitance observed for gate oxides below 2.0 nm in both accumulation and inversion is modeled using distributed voltage-controlled RC networks. The imaginary components of small-signal input admittance obtained from AC network analysis agree well with measured capacitance.**

***Index Terms*—Device simulation, MOS C–V modeling, quantum mechanical corrections, SPICE, ultrathin gate oxide,**

I. Introduction

FOR MODERN CMOS devices with gate length of 0.1 μm and operating at 1 V, an oxide thickness of the order 3.0 nm is needed, which corresponds to five to six layers of silicon atoms. With such thin oxides, direct tunneling results in exponential increases in gate leakage current, increased power consumption and deteriorated device performance [1]. Another issue is the loss of inversion charge and transconductance due to the quantization of carriers in the channel as well as polysilicon gate depletion effects. In the QM model, the density of inversion charge peaks at around 10 Å below the silicon surface such that gate capacitance and inversion charge are both effectively reduced to those of an equivalent oxide a few angstroms thicker than the physical oxide. Thus, characterization is needed to account for these effects in the choice of design parameters, especially for future sub-0.1-μm MOS technology.

The goal of this work is to characterize MOS capacitance for gate oxides in the thickness range of 1.3–1.8 nm. In particular, modeling of the sharp decrease both in capacitance in the inversion and in the accumulation is emphasized; an equivalent circuit approach considers gate tunneling current as well as other QM effects.

Manuscript received November 25, 1998; revised February 22, 1999. This work was supported by DARPA (Contract DABT 63-94-C-0055) and NSF (Contract ECS 9526378) under the National Center for Computational Electronics (NCCE).

C.-H. Choi, J.-S. Goo, T.-Y. Oh, Z. Yu, and R. W. Dutton are with the Center for Integrated Systems, Stanford University, Stanford, CA 94305-4075 USA (e-mail: chchoi@leland.stanford.edu).

A. Bayoumi, M. Cao, P. Vande Voorde, and D. Vook are with ULSI Laboratory, HP Laboratory, HP Company, Palo Alto, CA 94303 USA.

C. H. Diaz is with Taiwan Semiconductor Manufacturing Company, Taiwan R.O.C.

Publisher Item Identifier S 0741-3106(99)05057-0.

II. Capacitance Simulation

The physical phenomena for QM effects involve a shift of the carrier distribution in the channel region away from the Si–SiO_2 interface as the carrier distribution becomes quantized into discrete levels rather than having a distribution based on a single band approximation. An empirical and hybrid model for the QM correction in MOS structures is proposed and implemented in a two-dimensional (2-D) device simulator, which combines both the van Dort and Hansch models [2].

In the vicinity of the Si/SiO_2 interface, the density of conduction band states can be expressed as

$$N_C(z) = N_C[1 - e^{-(z+z_0)^2/\lambda_{th}^2}] \tag{1}$$

where the z-axis is perpendicular to the Si–SiO_2 interface and z_0 is introduced for the finite carrier concentration at the surface ($z = 0$). N_C is the normal density of conduction band states, which is position-independent, and λ_{th} is a thermal wavelength.

To eliminate the flatband singularity that occurs in the van Dort model [3], a novel bandgap broadening expression is used. The bandgap correction term, depending on the 2/3 power of the transverse surface electric field, F_S, in the van Dort model is given as

$$\Delta E_g(z) = \frac{\kappa\beta}{2k_BT} F_S^{2/3} \left[\frac{2e^{(z/L)^2}}{1 + e^{-2(z/L)^2}}\right] \tag{2}$$

where β is a physical constant, κ is the fitting parameter with the theoretical value of unity and L is a characteristic decay length. The singularity originates from the derivative of this quantity with respect to $F_S = 0$ in the gate capacitance computation. To eliminate this flatband singularity, the $F_S^{2/3}$ term in (2) is now replaced with

$$\frac{F_S^2}{ae^{-(F_S/\sigma)^2} + F_S^{4/3}} \tag{3}$$

where both a and σ are adjustable parameters. It preserves the asymptotic dependence of the bandgap correction on $F_S^{2/3}$ when $F_S \to \pm\infty$ while eliminating the flatband singularity.

In addition, the intrinsic carrier concentration, n_i, is evaluated by introducing a bandgap broadening mechanism into the Hansch formulation. Polysilicon depletion effects are considered by specifying the polysilicon gate as a heavily doped n-type silicon region.

Fig. 1 shows the comparison between experimentally measured and simulated nMOS C–V data for an oxide thickness

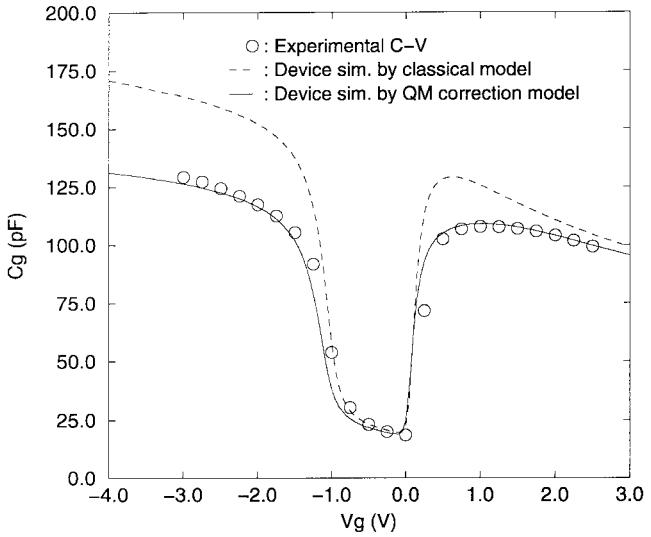


Fig. 1. Simulated and experimental C–V for the n-type MOS capacitor with $t_{ox} = 2.0$ nm. Circles denote measurements and solid and dashed lines indicate simulated data by the QM correction and the classical model, respectively.

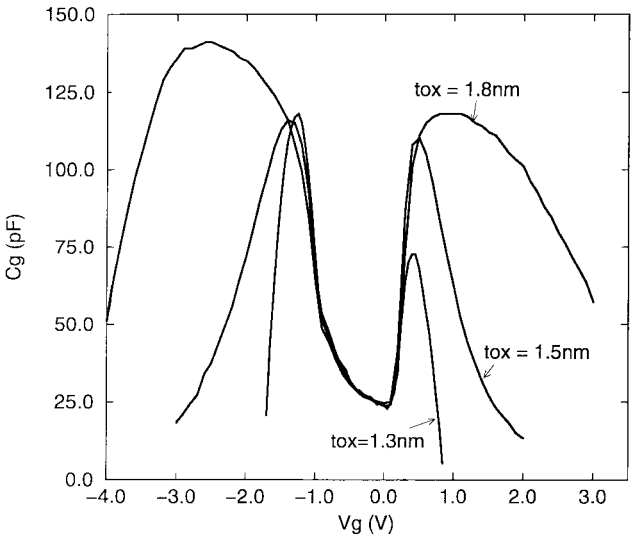


Fig. 2. Experimental MOS C–V for $t_{ox} = 1.3$, 1.5, and 1.8 nm, respectively. Sharp decrease of gate capacitance occurs by the gate tunneling current.

of 2.0 nm and area of 100 μm $\times$ 100 μm. For this analysis, the substrate doping profile is obtained using a process simulation that includes TED (transient enhanced diffusion) effects [4]. The reduction of gate capacitance in the accumulation region is related to the QM effect; the reduction in the inversion region is related to both the QM and poly depletion effects.

III. Gate Tunneling Simulation

Thin gate oxides were grown using rapid thermal oxidation (RTO) in the temperature range of 900–950 °C. Fig. 2 shows measured C–V curves of the n^+-poly/SiO_2/p-Si MOS capacitors for oxide thicknesses of 1.3, 1.5, and 1.8 nm, respectively. The capacitor area is 100 μm $\times$ 100 μm, and the small-signal frequency is 100 kHz. It should be noted that there are sharp transitions in the capacitance both in the inversion and accumulation regions. Also, the peak capacitance for a gate oxide thickness (t_{ox}) of 1.3 nm is less than that of $t_{ox} = 1.8$ nm. This is caused by the gate tunneling current across the oxide. As a result, measured capacitance for oxides thinner than 2.0 nm shows qualitatively different behavior compared to thicker oxides. In order to model this behavior quantitatively, an accurate gate current calculation needs to be coupled with the QM effects.

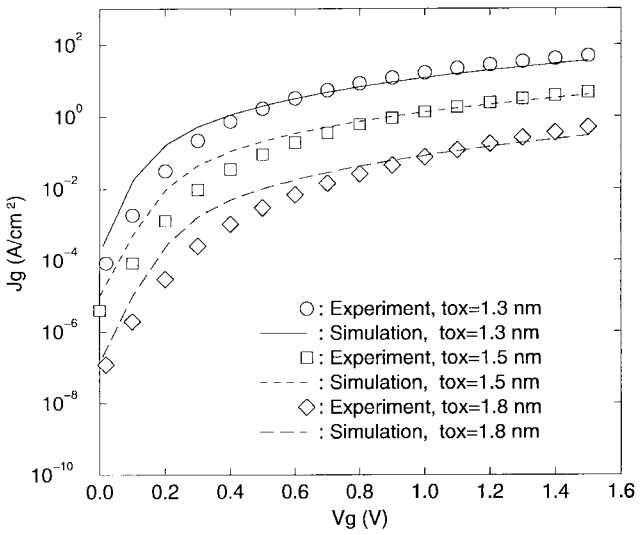


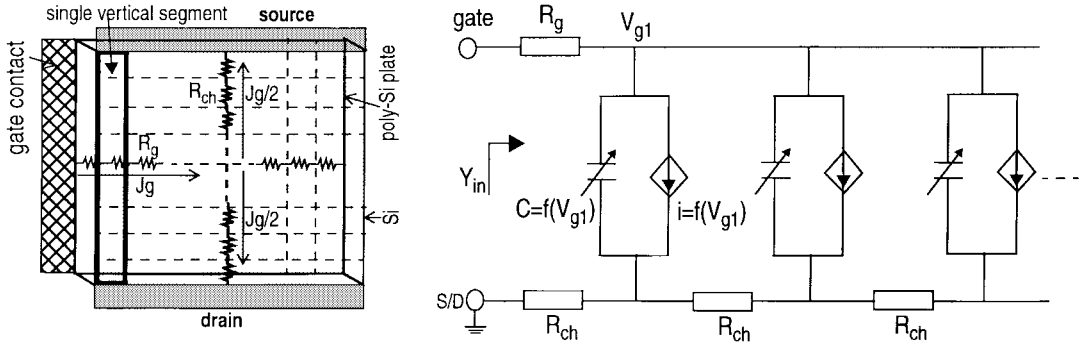
Fig. 3. Comparisons of gate tunneling current between measured and simulated using 1-D Green's function solver for $t_{ox} = 1.3$, 1.5, and 1.8 nm, respectively.

In this work the tunneling current is calculated using a one-dimensional (1-D) Green's function simulator, NEMO [5]. Because the multiple-scattering effects become extremely important, the WKB approximation is no longer satisfied for the thinner SiO_2 layers [6]. NEMO considers the injection from quasibound states and continuum states; hence an accurate calculation can be achievable for the density of states. Carrier density is calculated quantum mechanically in the device; scattering effects at the contacts can be approximated by the potential used in calculation of the surface Green's function.

Fig. 3 shows the measured and simulated gate current for three gate oxide thicknesses. The discrepancies between the measured and simulated data are probably caused by either surface roughness or uncertainty in determining the effective oxide thickness.

IV. Equivalent Circuit Modeling

To characterize the MOS C–V for gate oxide thicknesses ranging from 1.3 to 1.8 nm, an equivalent circuit is introduced. First, the silicon surface region is divided into small rectangular segments perpendicularly to the channel-current flow in order to consider the channel resistance, as in Fig. 4(a). In addition, the polysilicon area is vertically divided into small pieces to consider the polysilicon resistance effect. R_{ch} represents the channel resistance for each segment; this channel resistance also varies as a function of gate bias. R_g



(a) (b)

Fig. 4. An equivalent circuit approach of the MOS capacitor for gate oxide thickness thinner than 2.0 nm: (a) top view of the n-channel MOS capacitor with source/drain grounded (area = 100 μm × 100 μm) and (b) distributed RC network for the single vertical segment in Fig. 4(a).

represents a distributed resistance in the polysilicon gate, and the measured sheet resistance of the polysilicon is 5 $\Omega/\square$, which is much smaller than the channel resistance ($R_{ch} \simeq$ 1.5 K$\Omega/\square$). Fig. 4(b) is the equivalent circuit for the single vertical segment represented in Fig. 4(a). Then the equivalent circuit for the whole device structure is constructed by incorporating all the vertical segments. The tunneling current is modeled using a nonlinear, voltage-controlled current source ($i = f(V_g)$) derived from the gate current simulation discussed previously. The current values of each segment are then scaled according to the number of RC stages used. Also, the gate capacitance without the gate tunneling effect is modeled using a nonlinear, voltage-controlled capacitance ($C = f(V_g)$). These voltage-controlled elements contain 1-D and piecewise linear voltage-dependent capacitance and voltage-dependent current tables which are built from C–V and gate tunneling simulations, respectively. SPICE AC small-signal analysis is then performed to find the input admittance ($Y_{\rm in} = G + j\omega C$) of the circuit at a given frequency of 100 kHz. The C–V data including gate tunneling effects is finally obtained by taking the imaginary component of $Y_{\rm in}$. Fig. 5 compares the modeling with the measurements for oxide thicknesses of 1.3, 1.5 and 1.8 nm, respectively, when the number of RC segments is 20. When it exceeds five, the number of RC segments makes no difference in terms of the final capacitance values. The extracted channel resistance values in the inversion and accumulation regions are 2.5 K and 1.1 K $\Omega/\square$, respectively. The sharp transition point in capacitance becomes lower as gate oxide thickness is reduced, which results from the increase in gate tunneling current with an exponential-like dependence between tunneling current and oxide thickness.

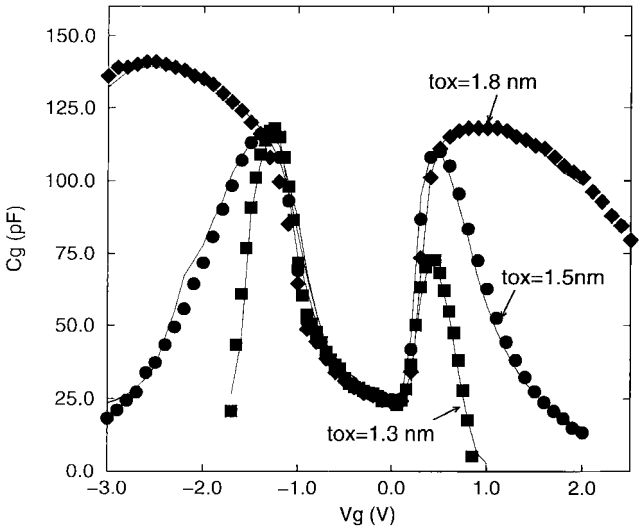


Fig. 5. Results of imaginary components of input admittance by the AC network analysis for the equivalent circuit in Fig. 4. Symbols represent measured C–V for t_{ox} = 1.3, 1.5, and 1.8 nm.

V. Summary

MOS C–V characteristics of gate oxide in the sub-2.0-nm region are modeled with an empirical, hybrid formulation for QM corrections implemented in a 2-D device simulator. Discrepancies in C–V modeling exist as the gate oxide becomes thinner than 2.0 nm due to the effects of the gate tunneling current. The sharp decrease in capacitance for gate oxides below 2.0 nm in both the accumulation and inversion regions is modeled by using a distributed RC network that includes the gate tunneling current which is calculated using a Green's function solver.

References

[1] Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S. H. Lo, *et al.*, "CMOS scaling into nanometer regime," *Proc. IEEE*, vol. 85, no. 4, p. 486, 1997.
[2] Z. Yu, Stanford Univ., Stanford, CA, private communication.
[3] 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 Electron*, vol. 37, no. 3, p. 411, 1994.
[4] P. Vande Voorde, P. B. Griffin, Z. Yu, S-Y Oh, and R. W. Dutton, "Accurate doping profile determination using TED/QM models extensible to sub-quarter-micron nMOSFET's," in *IEDM Tech. Dig.*, 1996, p. 811.
[5] C. Bowen, *et al.*, "Physical oxide thickness extraction and verification using quantum mechanical simulation," in *IEDM Tech. Dig.*, 1997, p. 869.
[6] S. Nagano, M. Tsukiji, E. Hasegawa, and A. Ishitani, "Mechanism of leakage current through the nano-scale SiO_2 layer," *J. Appl. Phys.*, vol. 75, no. 7, p. 3530, 1994.