Implications of gate tunneling and quantum effects on compact modeling in the gate-channel stack

Robert W. Dutton and Chang-Hoon Choi
Center for Integrated Systems
Stanford University, Stanford, CA, USA, rdutton@stanford.edu

ABSTRACT

Simulation and modeling of gate tunneling current for thin-oxide MOSFETs and Double-Gate SOIs are discussed. Guidelines for design of leaky MOS capacitors are proposed. Resonant gate tunneling current in DG SOI is simulated, based on quantum-mechanical models, and shown to be an issue of growing concern.

Keywords: gate tunneling, DG SOI, quantum effect, CV.

1 Introduction

As MOSFET dimensions are scaled below the 100 nm regime with gate oxides thinner than 2.0 nm, exponentially higher gate tunneling current occurs, leading to loss of inversion charge, gate capacitance and transconductance. Moreover, this gate current causes significant off-state leakage current, which may result in faulty circuit operation, since designers may assume that there is no appreciable gate current. Circuit models are needed in design for scaled CMOS, since designers must now consider the impact of gate tunneling current on devices and circuits.

2 Design for Leaky MOS Capacitor

Measured $C - V$ for oxides thinner than 2.0 nm have shown a sharp decrease of capacitance both in the inversion and accumulation regions [1], as reflected in Figure 1. This capacitance attenuation was found to increase with increasing gate area and thinning gate oxide thickness. This becomes problematic in the design of large-area MOS capacitors, as gate capacitance is less than the expected value due to the gate tunneling current associated with the channel resistance.

Consider a long channel MOS capacitor with highly-leaky gate dielectrics when the bias is in the inversion condition. As shown in Figure 2, the current ($I(x)$) along the channel direction is not uniform in the presence of the gate tunneling current ($I_g(x)$); it has a peak near the source ($x = 0$) and decreases towards the center of the channel ($x = L/2$). This results from the channel resistance across the long channel, which in turn raises the channel potential ($V_C(x)$) at the position near $x = L/2$. Since gate tunneling current decreases exponentially with decreasing gate bias, we can assume that $I(x)$ (A/μm^2) is exponentially reduced from the source along the channel direction given by:

$$I(x) = I(x=0)e^{-ax} \equiv I_0 e^{-ax} (\simeq 0.5 I_G e^{-ax}) \quad (1)$$

where 'a' is a decay length, which depends on the dopant concentration. Now $V_C(x)$ can be represented as:

$$V_C(x) = \int_0^x \frac{\rho_s}{W} I_0 e^{-ax} dx = \frac{\rho_s}{aW} I_0 [1 - e^{-ax}] \quad (2)$$

where ρ_s is the channel sheet resistance. The effective gate-to-channel bias (V_{GC}) is minimized at the center of the channel due to a peak $V_C(x)$, leading to a reduced MOS capacitor. The average voltage drop along the channel for the the gate length (L) is expressed as,

$$\overline{V_C}(L) = \frac{\rho_s}{aW} I_0 [1 - \frac{1}{aL}(1 - e^{-aL})] \quad (3)$$

The modeled $I(x)$, $V_C(x)$ and $V_C(L)$ are shown in Figure 3, compared with device simulation results using MEDICI, when tox=1 nm, ρ_s = 1.15K Ω/□, $I_0 = 1.33 \times 10^{-4}$ A/μm^2 and $a = 0.16$. The capacitance roll-off can be reduced by using shorter gate lengths, owing to the reduced channel resistance, as shown in Figure 4. Thus, selective use of the thicker oxide in the dual-gate oxide process or use of a multi-fingered gate structure (instead of a long channel gate structure) is more desirable in the design of thin-oxide, large MOS capacitors, as shown in Figure 5. However, parasitic gate overlap capacitance becomes dominant when the length is too short. Therefore, selection of capacitor geometry requires consideration of gate tunneling current, channel resistance and position of well contacts for proper design of thin-oxide MOS capacitors.

3 Gate Current for Circuit Simulation

This direct tunneling current model is expressed as [2],

$$J_{DT} = C(V_{ox}/t_{ox})^2 e^{\frac{-B(1-(1-V_{ox}/\Phi_B)^{3/2})}{V_{ox}/t_{ox}}} \quad (4)$$

where the pre-exponent C and slope B are given in [2]: In order to apply this equation to the calculation of gate tunneling current, it is necessary to relate the oxide voltage (V_{ox}) to the applied voltage (V_{GS}), since V_{ox} depends on V_{GS} as well as the surface potential (ψ_s) as follows:

$$V_{ox} = V_{GS} - V_{FB} - \psi_s \quad (5)$$

The surface potential in the weak inversion region ($0 < \psi_s < 2\phi_f$) can be approximated as:

$$\psi_{s,weak} = V_{GS} - V_{FB} + \frac{\gamma^2}{2} - \gamma\sqrt{V_{GS} - V_{FB} + \frac{\gamma^2}{4}} \quad (6)$$

where γ is the body factor. In the strong inversion region ($\psi_s > 2\phi_f$) ψ_s becomes

$$\psi_{s,strong} = 2\phi_f + V \quad (7)$$

where V denotes the electron quasi-Fermi potential, which ranges from V_{SB} at the source to V_{DB} at the drain side. For calculation of the surface potential, considering the QM effects, the intrinsic carrier concentration (n_i), is evaluated by introducing a bandgap broadening mechanism which is approximated using van Dort's approach [3]. Figure 6 shows calculated gate currents from the compact model, compared with those from a Schrödinger equation solver using Green's function method, NEMO [4].

4 Impact of Gate Current On Circuits

A recent study has shown that direct tunneling current appearing between the Source-Drain Extension (SDE) and the gate overlap, the so-called Edge Direct Tunneling (EDT) effect, dominates off-state drive current, especially in very short channel devices. This results from the fact that the ratio of the gate overlap to the total channel length becomes large in short channel devices compared to that of the long channel device. Thus, the gate current effect is expected to become appreciable in ultra-thin oxide, sub-100 nm channel length MOS circuits. In order to evaluate circuit performance by considering gate direct tunneling effects, a two-lump (macro) circuit model has been constructed in the circuit simulator, HSPICE. The macro-circuit model has been applied to the sample and hold (S/H) circuit, which is formed by a sampling CMOS switch followed by a hold capacitor, as shown in Figure 7(a). When the clock (*Phi*) is high, V_{out} follows V_{in}; when *Phi* goes low, V_{out} ideally remains at a constant level. However, V_{out} will not hold this sampled value if leakage paths exist. This tunneling current-induced decay in V_{out} during the hold period can be modeled using the RC circuit shown in Figure 7(a). V_{out} decays as a function of time as,

$$V_{out}(t) = V_{out}(t_0)\, e^{-\frac{t-t_0}{R_{gd,n}C_{out}}} \approx V_{dd}e^{-\frac{I_{gd,n}\ (t-t_0)}{V_{dd}C_{out}}} \quad (8)$$

Figure 7(b) shows simulation results of a S/H switch for three gate oxide thicknesses. During the holding period, the output node does not maintain the sampled value due to gate leakage current, and degradation becomes increasingly severe as the oxide thickness is scaled down.

5 Gate Tunneling in Double-Gate SOI

To overcome an undesirable, negative threshold voltage in Symmetrical DG NMOS with two n^+ poly-gates, Symmetrical DG with Ground-Plane (SGP) has been proposed as means to adjust threshold voltage by using different back-gate bias. Schematic and band diagram representations for a SGP NMOS device are shown in Figure 8. Consider a one-dimensional device geometry, ignoring source and drain effects. As shown from the band diagram, a potential difference exists between the two gates for $V_{FG} > V_T$. Thus, electrons can tunnel through the gate structures, resulting in substantial gate currents.

Figure 9 shows single- and double-barrier structures with incident electrons, in which the oxide layers act as potential barriers. Current transmission coefficients are denoted as T_B and T_{2B} for the single- and double-barrier, respectively. T_B is increases exponentially as the energy increases, while T_{2B} is written as [5],

$$T_{2B} \simeq \left[1 + \frac{4R_B}{{T_B}^2} sin^2(k_1 L - \theta)\right]^{-1} \quad (9)$$

where $k_1 \simeq \sqrt{\frac{2m^*(E-E_{c1})}{\hbar^2}}$ (m^* is the effective mass), R_B is the current reflection coefficient ($T_B + R_B = 1$) and θ is the phase. Eq. (9) implies that T_{2B} can reach unity when the second term is zero; resonance can occur based on the relationship between the electron wavelength and the quantum-well width (L) in the double-barrier.

In the simulation model it is assumed that T_{si} = 5 – 40 nm, t_{ox} = 1.5 nm, N_{p-Si} = 1.0×10^{18} cm^{-3} and N_{n-poly} = 2.0×10^{20} cm^{-3}. Figure 10 compares simulated gate tunneling currents between double-gate SOI (T_{si} = 5 nm) and single-gate (bulk) MOS. The gate current for the double-gate structure is higher than that for the single-gate MOS due to resonant tunneling in the double-barrier. As a result, this gate tunneling for the DG has a negative-differential resistance like that of resonant-tunneling diodes (RTD) at the gate bias $\sim$1.4 V, which has recently received considerable attention for possible circuit applications.

6 Summary

Gate tunneling effects including Edge-Direct Tunneling should be considered in scaled MOS circuits. Compact models are proposed that provide guidelines in the design of large-area MOS capacitors to avoid the capacitance roll-off. A gate current model for circuit simulation is also proposed. Gate tunneling current for DG SOI can be significant compared to a bulk MOS, owing to resonant tunneling effects.

REFERENCES

[1] C.-H. Choi, and et al, *IEEE Electron Device Lett.*, p.292, June 1999.

[2] C.-H. Choi, and et al., in *IEDM Tech. Dig.*, p.735, 1999.

[3] M.J. van Dort, and et al, *Solid-State Elec.*, no.3, p.411, 1994.

[4] C. Bowen, and et al., in *IEDM Tech. Dig.*, p.869, 1997.

[5] S. Datta, "Quantum Phenomena," Addison-Wesley, Reading, MA, 1989.

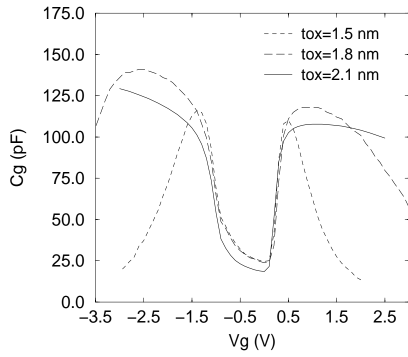


Figure 1: Measured gate capacitance curves for thin gate oxide (tox = 1.5, 1.8, and 2.1 nm) N-MOSFETs.

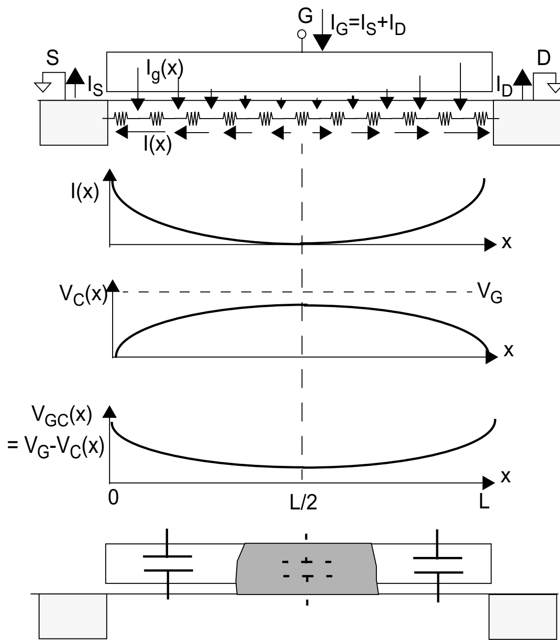
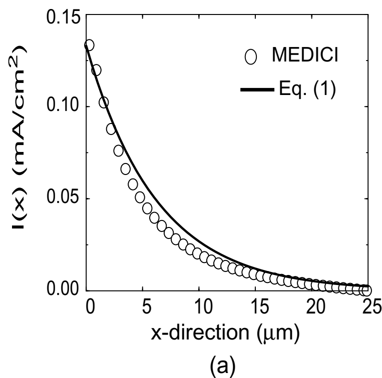
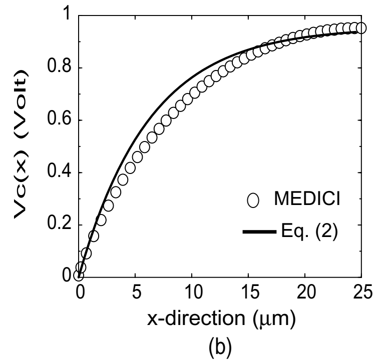


Figure 2: Current and potential distributions for a long channel MOS capacitor in the presence of gate tunneling current. Note the reduced MOS capacitor for highly leaky gate dielectrics, because of the decreased gate-to-channel bias (V_{GC}) near the center of the channel.





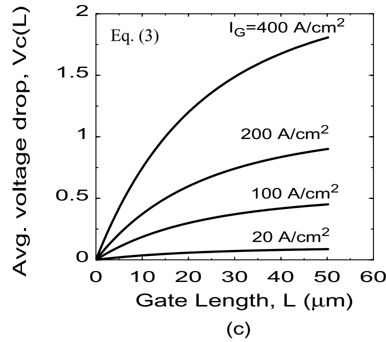


Figure 3: Current and Potential distributions in a large, leaky MOS capacitor, obtained from Eqs.(1)-(3), compared with numerical simulation using MEDICI. tox=1 nm, ρ_s=1.15K Ω/sq., I_0=1.33x10^{-4}A/um^2 and a=0.16, (a) current I(x) for L=50um, (b) potential Vc(x) for L=50um, (c) average voltage drop, Vc(L) with respect to gate lengths (L).

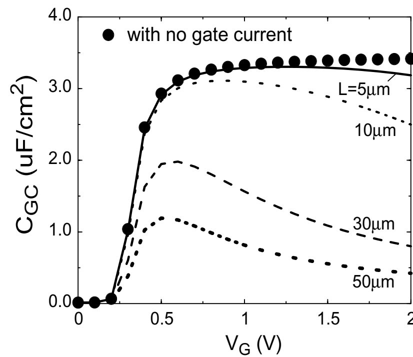


Figure 4: Simulated gate-to-channel capacitance (C_{GC}) curves by using MEDICI for different gate lengths. Note the less capacitance roll-off for the shorter gate lengths. tox=1 nm and I_G=177 A/cm^2.

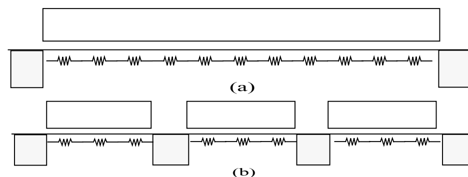


Figure 5: Design of highly-leaky, thin-oxide MOS capacitor. (a) long-channel structure is not desirable for capacitor, (b) multi-fingered gate structure to reduce capacitance roll-off.

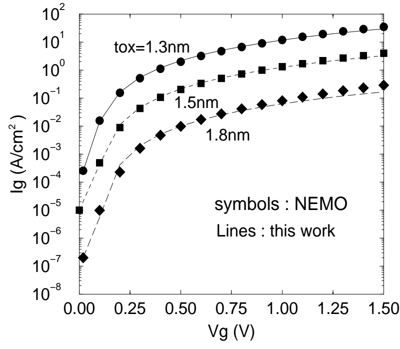
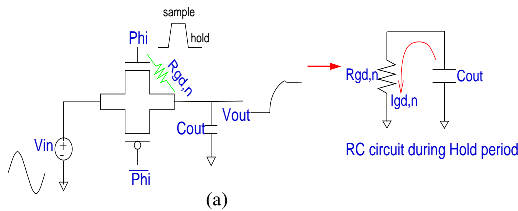
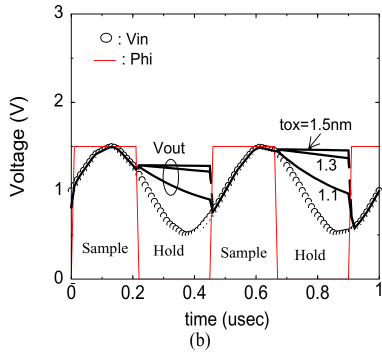


Figure 6: Gate currents calulated from the compact model, compared with those from a Green's function solver, NEMO [4].



(a)



(b)

Figure 7: CMOS sample and hold (S/H) circuit and simulated waveforms. (a) S/H circuit, (b) circuit simulation results with gate tunneling effects. Note the reduced Vout during the Hold Period for the thinner tox due to higher gate leakage current.

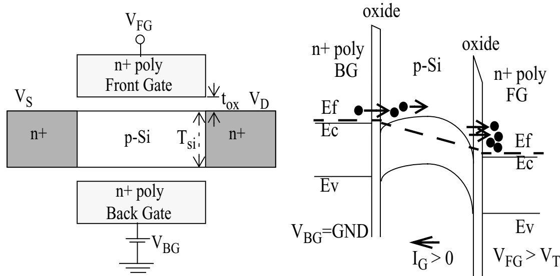
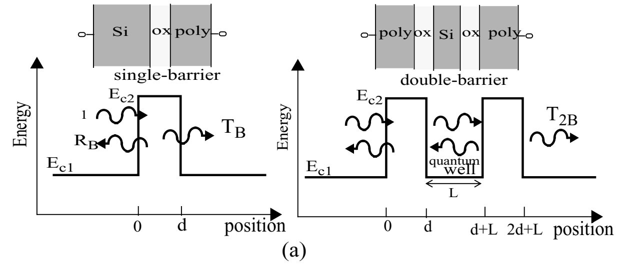
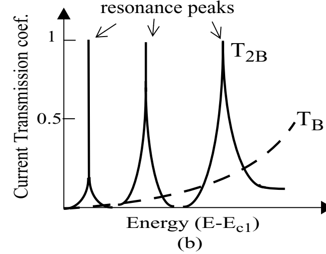


Figure 8: Cross-section of a Symmetrical DG with Ground-Plane (SGP) and corresponding band diagram for $V_{FG} > V_T$. tox = 1.5 nm, Tsi = 5-40 nm, $N_{p\text{-}Si} = 1.0\text{x}10^{18}\text{cm}^{-3}$ and $N_{n\text{-}poly} = 2.0\text{x}10^{20}\ \text{cm}^{-3}$ are assumed for simulation.



(a)



(b)

Figure 9: Current transmission coefficients for single- and double-barrier structures [5]. (a) single- and double-barrier with incident electrons (b) current transmission coefficients (T_B and T_{2B}) versus the electron energy, note the higher T_{2B} than T_B.

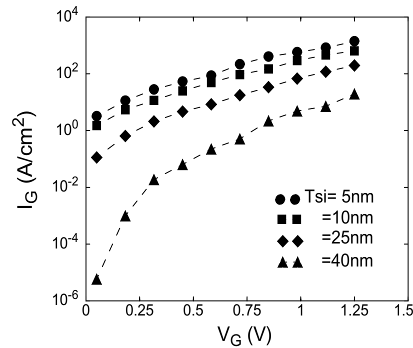


Figure 10: Simulated gate tunneling current using NEMO for a SGP DG structure with different silicon layer thicknesses (Tsi).

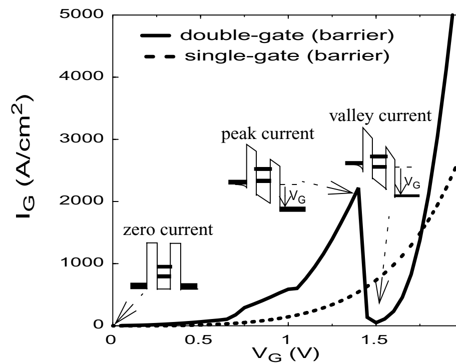


Figure 11: Simulated gate tunneling currents for double-gate (barrier) (tox = 1.5 nm and Tsi= 5 nm) and single-gate MOS (tox = 1.5 nm). Note the resonant tunneling current for double-gate structure due to the quantized energy states in the Si layer (quantum well).