

# A Compact Physical Model for Critical Quantum Mechanical Effects in Bulk MOSFETs

Lihui Wang and James D. Meindl

Microelectronic Research Center of  
Georgia Institute of Technology 791 Atlantic Drive NW, Atlanta GA 30332 USA  
Email: gt0463c@mail.gatech.edu

## ABSTRACT

In this paper, a compact physical model is developed to incorporate the critical quantum mechanical effects (QME) in bulk MOSFETs. This new model gives close agreement with the numerical simulation down to 20 nm effective channel length. For the 20 nm channel length MOSFET with the gate dielectric equivalent oxide thickness (EOT) of 0.5 nm, QME lead to the increase of subthreshold swing ( $S$ ) from 74 mV/decade to 105 mV/decade. The scaling limit of bulk MOSFETs is then discussed based on two considerations: short-channel effects (SCE) controllability and power consumption constraint for tunneling current. For 30 nm channel length MOSFETs with tolerable SCE and tunneling current density, the appropriate range of EOT for SiO<sub>2</sub> is only 0.1 nm and less than the thickness of one atom layer of SiO<sub>2</sub>. This indicates high dielectric constant ( $\kappa$ ) gate material should be applied to extend the scaling of bulk MOSFETs for the 30 nm technology generation and beyond.

**Keywords:** MOSFET, scaling limit, quantum mechanical effects

## 1 INTRODUCTION

For years, CMOS technology advances by continuous size reduction of MOSFET. According to the scaling theory of MOSFETs, a reduction of the channel length ( $L$ ) requires a correspondent decrease in the gate dielectric thickness. In addition, further reduction in gate dielectric thickness is necessary due to SCE suppression and threshold voltage scaling. As MOSFETs are scaled down to the sub-50 nm range, the suitable EOT is only about 1 nm [1]. At such small dimensions, QME including energy quantization and electron tunneling through the gate oxide play significant roles in determining device characteristics.

One consequence of the ultra-thin oxide layer is the width reduction of the potential barrier separating the gate from the channel, thus making it easier for electrons to tunnel through the insulator layer. Gate tunneling gives rise to MOSFET leakage current, which causes an increase of power consumption and a decrease of logic operating margins.

Besides the tunneling current, the ultra-thin oxide layer exaggerates the influence of energy quantization which

leads to effective oxide thickness increase and total gate capacitance degradation [2]. As the result, transconductance and driving ability of MOSFETs in the super-threshold region is degraded [3,4]. Also, the increased EOT exacerbate SCE which lead to the degraded  $S$  [2]. Since increased EOT is non-scalable with the reduction of physical oxide thickness, it is particularly important to include the energy quantization effect in short-channel devices with ultra-thin gate dielectrics. A new compact physical model for energy quantization is developed in this paper, and incorporated into the short-channel  $S$  model [5].

## 2 QME MODEL

To investigate the QME in MOSFETs, it is necessary to introduce the wavefunction to characterize inversion charges based on the Schrödinger Equation

$$-\frac{\hbar^2}{2m^*}\nabla^2\psi(x,y)-q\phi(x,y)\psi(x,y)=E\psi(x,y) \quad (1)$$

where  $\hbar$  is the reduced Planck constant,  $m^*$  is the effective mass of the electrons in silicon,  $\psi(x,y)$  is the wavefunction of the electrons,  $q$  is the electron charge,  $\phi(x,y)$  is the electric potential and  $E$  is the energy of the electrons. The electric potential  $\phi(x,y)$  is given by the Poisson Equation:

$$\frac{\partial^2\phi(x,y)}{\partial x^2}+\frac{\partial^2\phi(x,y)}{\partial y^2}=\frac{q}{\varepsilon_{Si}}(N_A+n(x,y)) \quad (2)$$

where  $\varepsilon_{Si}$  is the permittivity of silicon,  $N_A$  is the doping concentration of the channel and  $n(x,y)$  is the electron density.

While many research efforts are focused on the numerical simulations, compact physical model is highly desired to obtain physical insight into device operating principles and project QME on scaling capacity.

### 2.1 Gate tunneling Model

A physical model for gate direct tunneling current density is derived from the simplified solution of the Schrödinger equation [6],

$$J_T = \frac{4\pi m^* q}{h^3} (kT)^2 \left(1 - \frac{\gamma kT}{2\sqrt{q(\chi - V_{ox}/2)}}\right)^{-1} \times \exp\left(\frac{q\phi_s - q\phi_b - E_g/2}{kT}\right) \exp\left(-\gamma\sqrt{q(\chi - V_{ox}/2)}\right) \quad (3)$$

where  $\phi_b = \frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right)$ ,  $\gamma = \frac{4\pi t_i \sqrt{2m_i}}{h}$ ,  $k$  is the Boltzmann constant,  $T$  is the absolute temperature,  $h$  is the Planck constant,  $t_i$  is the insulator layer thickness in the MOSFET,  $m_i$  is the effective electron mass in the insulator,  $E_g$  is the energy band gap,  $\chi$  is the modified electron affinity in silicon,  $V_{ox}$  is the voltage drop on the oxide layer and  $n_i$  is the intrinsic electron density.

Replacing gate material of SiO<sub>2</sub> with high- $\kappa$  material is a potential solution to reduce the tunneling current. Within the high- $\kappa$  material family, HfO<sub>2</sub> and HfSiO<sub>4</sub> are the two most promising candidates. A predicted 2~3-order of magnitude reduction of the tunneling current density for an EOT of 0.5 nm is obtained according to the tunneling current model as shown in Figure 1.

## 2.2 S Model

The subthreshold swing can be used as a measure of SCE which is defined by

$$S = 2.3 \left( \frac{\partial \ln I_{sub}}{\partial V_G} \right)^{-1} \quad (4)$$

where  $I_{sub}$  is the subthreshold current in the channel and  $V_G$  is the voltage applied on the gate. From the classical viewpoint, distribution of inversion charges is characterized by the “charge sheet” at the channel surface and its density varies exponentially with the surface potential. Based on this assumption, the subthreshold swing can be written as

$$S = 2.3 \frac{kT}{q} \left( 1 + \frac{C_d}{C_{ox}} \right) \quad (5)$$

where  $C_d$  is the capacitance of the depletion layer and  $C_{ox}$  is the oxide layer capacitance. However, as quantum mechanics considers inversion charges as a wave and the oxide layer as an abrupt potential barrier keeping the charges away, it is unlikely for inversion charges to concentrate on the substrate surface. Applying quantum

mechanical description of the inversion charges leads to the new expression of  $S$  [4] as

$$S = 2.3 \frac{kT}{q} \left( 1 + \frac{C_d'}{C_{ox}'} \right) \quad (6)$$

where

$$C_d' = \varepsilon_{si} / (d - \Delta) \quad (7)$$

$$C_{ox}' = \left( \frac{1}{C_{ox}} + \frac{\Delta}{\varepsilon_{si}} \right)^{-1} \quad (8)$$

$d$  is the depletion depth and  $\Delta$  is the average depth of the inversion charges. If comparing with the classical model, the only significant difference is the position of the inversion charge sheet which is modified from the channel surface to its centroid below the surface. The shift of the charge sheet position leads to the increase of the EOT by  $\Delta \cdot \varepsilon_{ox} / \varepsilon_{si}$  and the corresponding increase in  $S$  as well. As shown in Figure 2, EOT is effectively increased by 0.2~1.1 nm for various channel doping.

A more comprehensive  $S$  model is developed by incorporating QME into the existing short-channel  $S$  model based on the classical physics[5]. Figure 3a&b show that the new model gives close agreement with the numerical simulation results obtained from Dessis® of ISE TCAD [7] for short-channel MOSFETs with effective channel length down to 20 nm. For the 20 nm MOSFET with EOT=0.5 nm, the QME cause  $S$  increased from 74 mV/decade to 105 mV/decade, as shown in Figure 3b. Meanwhile, with the decreasing of  $L$ , the value of subthreshold swing increases more rapidly when the QME is taken into account. Classical models underestimate MOSFETs' susceptibility to SCE according to the effectively increased EOT from QME.

## 3 SCALING LIMIT

The scaling limit of bulk MOSFET is discussed based on two considerations: the SCE controllability and the power constraint. The SCE control requires keeping  $S < 100$  mV/decade which leads to the reduction of EOT along with the evolution of technology towards the shorter channel length. From the power constraint, the maximum allowed tunneling current density is specified by ITRS [1]. Because SiO<sub>2</sub> is preferred to be kept as long as possible to delay the major technology transition to high- $\kappa$  material, an accurate EOT scaling projection incorporating QME is expedient for exploring the ultimate scaling limit of SiO<sub>2</sub>. Illustrated by Figure 4, for 30 nm channel length MOSFET with the tolerable SCE and tunneling current density, the appropriate range for SiO<sub>2</sub> EOT is only 0.1 nm which is less than the thickness of one atom layer of SiO<sub>2</sub>. This indicates that the fundamental “atom layer exhaustion limit” prevents the

continuing scale-down from using SiO<sub>2</sub> for the gate insulation in the bulk MOSFET. The high- $\kappa$  material should be applied to extend the scaling of the MOSFET for the 30 nm technology generation and beyond.

#### 4 CONCLUSION

In conclusion, we have investigated the impact of key QME including gate tunneling and energy quantization on bulk MOSFETs through compact physical models. We find that the increase EOT due to the energy quantization deteriorate SCE which causes the degraded  $S$  in the short-channel MOSFET. For the 20 nm channel length MOSFET with 0.5 nm EOT, QME lead to degradation of  $S$  from 74 mV/decade to 105 mV/decade. Considering the QME and SCE comprehensively, minimum channel length for bulk MOSFETs of SiO<sub>2</sub> gate insulation is 30 nm. To extend the scaling of bulk MOSFETs, high- $\kappa$  material should be applied to replace SiO<sub>2</sub> as the gate dielectric.

#### REFERENCES

- [1] "International Technology Roadmap for Semiconductors (ITRS)," SIA, 2003.
- [2] S. Jin, Y. J. Park, and H. S. Min, "A numerically efficient method for the hydrodynamic density-gradient model," 2003, pp. 263-266.
- [3] S. A. Harelend, S. Krishnamurthy, S. Jallepalli, Y. Choh-Fei, K. Hasnat, A. F. Tasch, Jr., and C. M. Maziar, "A computationally efficient model for inversion layer quantization effects in deep submicron N-channel MOSFETs," *Electron Devices, IEEE Transactions on*, vol. 43, no. 1, pp. 90-96, 1996.
- [4] L. Wang, Q. Chen, R. Murali, and J. D. Meindl, "Quantum mechanical effects on CMOS SOC performance," 2003, pp. 109-112.
- [5] B. Agrawal, V. K. De, and J. D. Meindl, "Three-dimensional analytical subthreshold models for bulk MOSFETs," *Electron Devices, IEEE Transactions on*, vol. 42, no. 12, pp. 2170-2180, 1995.
- [6] K. A. Bowman, W. Lihui, T. Xinghai, and J. D. Meindl, "A circuit-level perspective of the optimum gate oxide thickness," *Electron Devices, IEEE Transactions on*, vol. 48, no. 8, pp. 1800-1810, 2001.
- [7] Integrated Systems Engineering Inc., "ISE TCAD Manuals," San Jose, CA 95113: 2004.

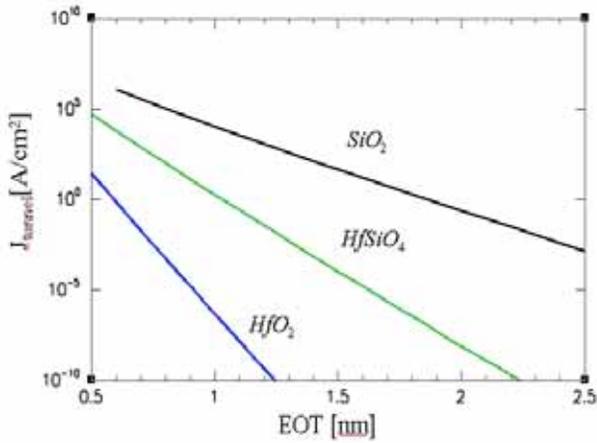


Figure 1: Gate tunneling current density as a function of the EOT for SiO<sub>2</sub>, HfSiO<sub>4</sub> and HfO<sub>2</sub>

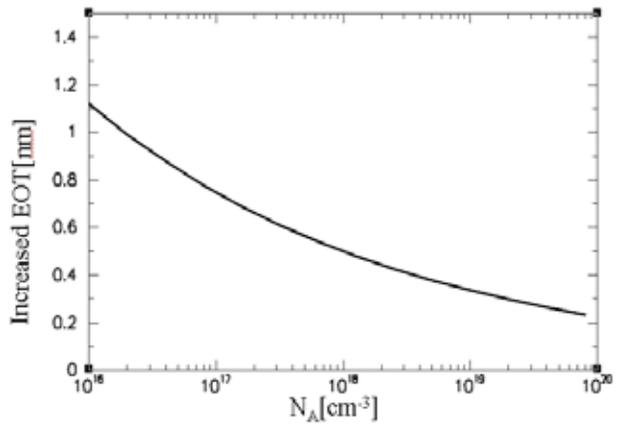


Figure 2: Increased EOT from the inversion charge energy quantization

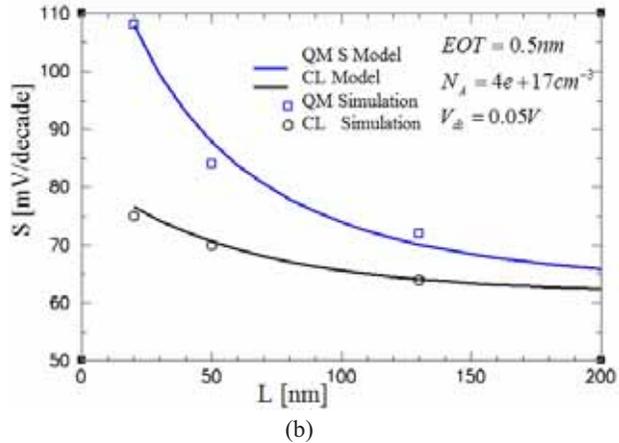
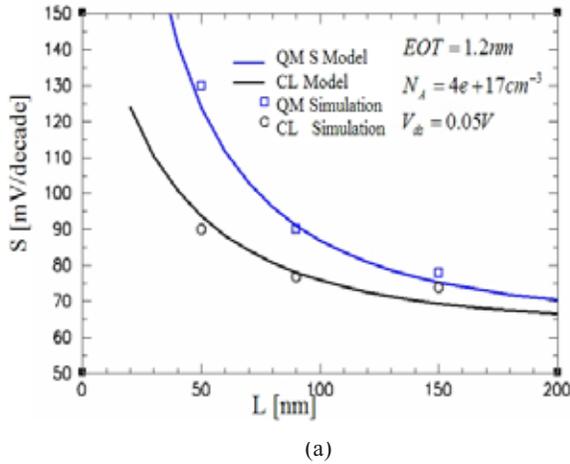


Figure 3: *S* model incorporated with QME for short-channel devices. (a) EOT=1.2nm as 2004 technology (b) EOT=0.5nm as technology for 2018. This model is compared with DESSIS simulation results.

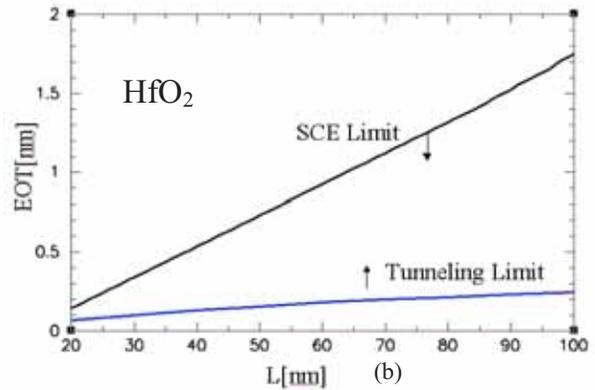
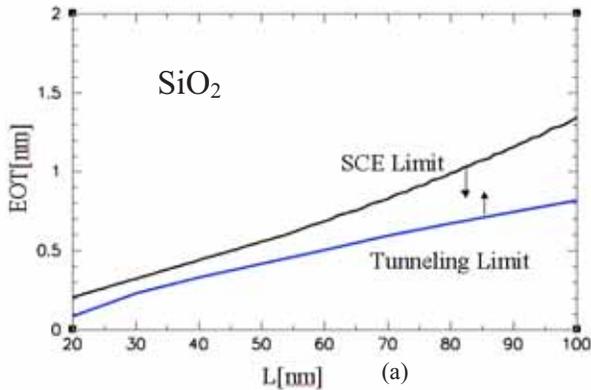


Figure 4: Design spaces for future MOSFETs with two constraints: 1) the SCE limit as  $S < 100$  mV/decade 2) the gate tunneling density limit as  $J_{\text{tunnel}} < I_{\text{sub}}/L * 10$  from ITRS for high performance devices. Figure (a) and (b) are projections using SiO<sub>2</sub> and HfO<sub>2</sub> as the gate insulator respectively.