Mobility Extraction and Compact Modeling for FETs Using High-k Gate Materials

R. Dutton, Y. Liu, C.-H. Choi and T. W. Chen

Center for Integrated Systems
Stanford University, Stanford, CA, USA, rdutton@stanford.edu

ABSTRACT

This simulation work discusses the impact of direct gate tunneling on effective mobility extraction methods, particularly the Inversion Charge Pumping (ICP) method proposed recently for transistors with high-k gate dielectrics. The valence-band electron gate tunneling (VBET)-induced substrate current is found to be critical to correctly reconstructing the drain current at high gate biases. The ICP technique is shown to be error-prone in the presence of non-negligible gate tunneling currents. The importance of the compact modeling of the charge pumping effect is also discussed for transistors operating in the large signal, non-quasi-static (NQS) regime. The channel segmentation approach is demonstrated to be suitable for such a purpose.

Keywords: effective mobility, gate tunneling, charge pumping, non-quasi-static, compact model

1 INTRODUCTION

The determination of effective channel mobility is of central importance to the characterization and modeling of field effect transistors. As gate oxide thickness scales down drastically, significant gate current is induced due to direct tunneling, which poses a challenge to correct channel mobility extraction in such devices. Furthermore, in the quest to understand and solve the mobility degradation problem associated with alternative high-k dielectric transistors, the existence of high density, fixed interface charge introduces additional complexity in mobility measurement and modeling. The effective channel mobility is obtained by accurate reconstruction of drain current I_d and inversion carrier density N_{inv} for long channel devices. Zeitzoff et al developed a physically based technique to reconstruct I_d in the presence of gate leakage for both ultra-thin gate oxide and high-k gate dielectric devices [1]. However, their correction technique breaks down at high gate bias. We demonstrate that this failure is due to the omission of non-negligible substrate current induced by valence band electron tunneling (VBET). By taking the substrate current into account, the validity range of the mentioned technique is shown to be greatly extended. In the work of Kerber et. al [2], the inversion charge-pumping (ICP) method is introduced to measure N_{inv} in high-k dielectric transistors. Although the errors due to interface traps are eliminated in their pulsed method, numerical simulations now presented show that its accuracy is greatly affected when the gate tunneling current is non-negligible.

Due to the increasing importance of RF CMOS technologies in wireless applications, it is desirable to develop accurate compact models for MOS transistors in the large signal, non-quasi-static (NQS) operation regime. Most available models are limited by the assumption of the inversion charge partition between the source/drain nodes. At the large signal, NQS condition, however, the substrate acts as a competing "channel" to remove the inversion charge due to the charge pumping effect. In this work, we show that a charge-pumping compact model can be achieved by using the channel segmentation approach developed in the work of [3].

2 EFFECTIVE MOBILITY EXTRACTION

2.1 Drain Current Reconstruction

For an inverted NMOS with ultra-thin gate dielectrics, there are various components contributing to gate tunneling current I_g, as illustrated in Fig. 1. In particular, the VBET contribution from the substrate into the conduction band leads to hole generation in the substrate, resulting in the substrate current (I_{sub}). Figure 2 shows the correlation between measured substrate current density as a function of the gate bias for different oxide thicknesses [4] and numerical results using a device simulator, MEDICI. It is observed that the VBET-induced I_{sub} increases abruptly at high gate voltage. For leaky gate dielectrics, the presence of significant gate current causes a droop in the drain current at high gate biases. This is evident in the simulated Id-Vg curve as shown in Fig. 3. In order to obtain the corrected dain current, a correction expression is suggested in [1] to eliminate the gate current as $I_{d0}=I_d+0.5I_g$. We plot the corrected I_{d0} in Fig. 3. As a reference of the "true" drain current, a simulated Id-Vg curve with the gate tunneling model turned off is also shown. An unrealistic rise of the corrected drain current is evident at high gate biases, as already noted in [1]. Such a serious limitation on the validity of the I_d correction technique is due to the failure to account for the VBET-induced I_{sub}. That part of the gate current is removed by the substrate and should be excluded

from the gate current. Hence, the following improved correction should be used instead:

$$I_{d0}^{*} = I_d + 0.5(I_g - I_{sub}).$$

It is clearly observed in Fig. 3 that this method correctly reproduces the "true" drain current up to high gate biases.

2.2 Inversion Charge Extraction

A schematic plot of the ICP technique developed in [2] is shown in Fig. 4, with the addition of a tunneling current component. The device symmetry has been employed and results of the two setups have been used to measure charge pumping component and to suppress the charge-trapping effect as well. The charge pumping carriers, $N_{ICP}^{(a)}$ and $N_{ICP}^{(b)}$, are obtained through integration of the charge pumping current for the two setups, respectively. Therefore, the full inversion charge has been obtained as follows [2]: $N_{inv}=2N_{ICP}^{(a)}-N_{ICP}^{(b)}$. However, in order to account for the gate tunneling current, the expression for the "true" inversion charge should be corrected as $N_{inv0}=2N_{ICP}^{(a)}-N_{ICP}^{(b)}+N_{tun}$, where N_{tun} is the portion of inversion charge that tunnels through the gate.

As a demonstration of the gate tunneling effect on ICP, we have conducted turn-off transient simulations using MEDICI for a long channel NMOS with a leaky gate stack. For comparison purposes, two cases are simulated with the gate tunneling model turned on and off, respectively; the simulated substrate current are plotted in Fig. 5. The difference between the two simulated curves is evident, indicating a strong influence of the gate tunneling current component. Figure 6 compares N_{inv} obtained from simulations of the ICP technique with that directly extracted from device simulation outputs. We firstly conduct ICP simulations with the gate tunneling model turned off; the successful reproduction of the directly extracted N_{inv} validates ICP in the absence of gate tunneling. After turning on the gate tunneling model, we again simulate the ICP for different fall times of the gate signal. As shown in Fig. 6, considerable errors in extracted N_{inv} are observed for long fall time. Such errors are suppressed by reducing the fall time. This is because the gate tunneling term is expressed as

$$N_{inv} = \int I_{tun}(V_g(t), N_{inv}(t))dt,$$

where the integration is over the rising and falling edges only. The transient tunneling current in the integrand has an upper bound equal to its steady-state value $I_g(V_g)$. Therefore, N_{tun} decreases with shorter fall times. We note that N_{tun} cannot be obtained by replacing the transient tunneling current with its corresponding quasi-static values because the ICP operation is in the NQS regime.

3 COMPACT MODELING OF CHARGE PUMPING CURRENT

The charge pumping effect occurs under large signal, NQS operations. When a transistor is switched off at a time scale shorter than the channel transit time, the inversion charge cannot be efficiently removed from the source/drain due to the finite channel resistance. A portion of it flows to the substrate and gets recombined with the majority carriers. In the work of [5], this effect and its dependence on the channel length and the fall time of the gate signal have been investigated. The charge-pumping effect is more significant for lower substrate resistance, particularly in PMOS with high substrate doping densities. We simulate the turn-off transients for two 1.5µm PMOS transistors with different substrate doping concentrations and plot the drain and substrate current in Fig. 7. The fall time of the gate signal is 0.5ns and both the source and drain are grounded. The substrate current becomes significant compared to the drain current for high substrate doping concentrations.

The non-negligible charge-pumping current complicates the compact modeling of the transistor's large signal, NQS operation. As illustrated by the equivalent circuit in Fig. 8(a), the NQS effect is modeled by a distributed channel resistance. In addition, the substrate resistance and the forward-biased PN diodes formed between the channel and the substrate during charge-pumping need to be accounted for. This invalidates one-dimensional treatments of the channel as well as the assumption of inversion charge partition solely between the source and the drain. Based on such considerations, we find that the channel segmentation approach proposed in the work of [3] particularly suitable to address the problem. As shown in Fig. 8(b), the channel is partitioned into several small segments; each of them is modeled by the conventional BSIM3 or MOS9 QS model.

Based on the channel segmentation model, we have conducted Hspice simulations of the turn-off transients for an NMOS with t_{ox}=6.5nm, L_g=5µm, W=10µm and a substrate depth of 2µm. The channel is partitioned into 10 segments. Each segment is modeled by the BSIM3v3 model with W/L=10/0.5µm. The parasitic source/drain resistance of the small transistors is set to zero except at the two ends. The substrate doping concentration is 6E16 cm^{-3}, which gives a total substrate resistance R_{sub}=120Ω. In the transient simulation, the gate bias is switched from 4V to 0V and the fall time of the signal is 0.2ns. The source and drain are both grounded. The simulated substrate and drain currents are plotted in Fig. 9, respectively. For comparison purpose, simulation results using MEDICI are also presented. Both the substrate and drain transients from the Hspice and MEDICI simulations show a generally good agreement.

4 SUMMARY

We have conducted device-level simulations to investigate the impact of gate tunneling current on two recently proposed mobility extraction methods. By taking

into account the VBET-induced substrate current, the drain current reconstruction method has been improved. It is also demonstrated that the gate tunneling current causes serious errors in the ICP technique for the extraction of inversion charge. Furthermore, we have discussed the compact modeling of the charge-pumping effect for transistors under large signal, NQS operations, particularly using the channel segmentation approach.

REFERENCES

[1] P. M. Zeitzoff et al, *IEEE Electron Device Lett.*, vol. 24, p.275, 2003.

[2] A. Kerber et al, in *Proc. VLSI Symp.*, p.159, 2003

[3] A. J. Scholten et al, in *IEDM Tech. Dig.* p.163, 1999.

[4] A. Shanware et al, in *IEDM Tech. Dig.,* p.815, 1999

[5] S.-Y. Oh et al, *IEEE Electron Device Lett.,* vol. ED-27, p. 1571, 1980.

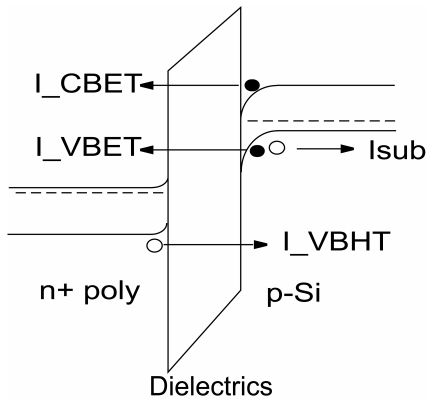


Figure 1: Gate tunneling component of an inverted NMOS transistor with leaky gate dielectrics. The VBET component results in a substrate current.

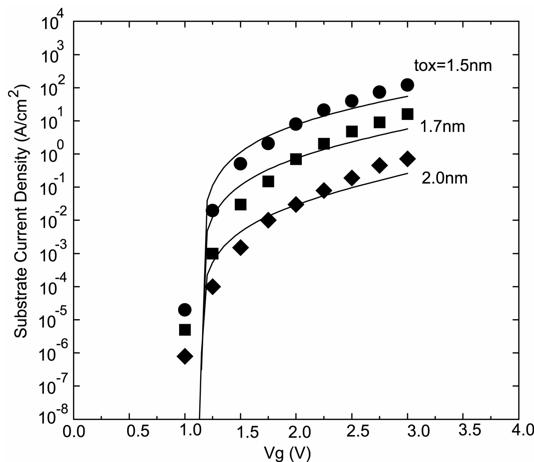


Figure 2: Measured [4] and simulated substrate current density as function of the gate bias for NMOS transistors with various gate oxide thicknesses.

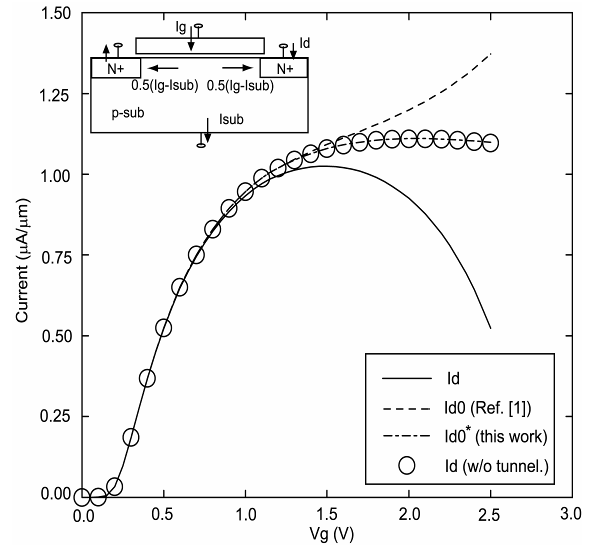


Figure 3: Simulated and reconstructed Id-Vg for an NMOS of 2µm channel length and 1.5nm gate oxide thickness. The drain bias is 10mV. The inset illustrates various current components in such a device.

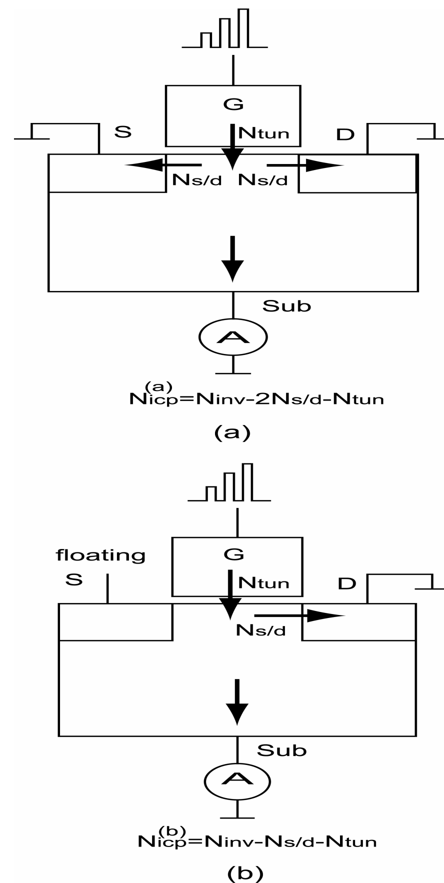


Figure 4: Schematic plots of the ICP setups: (a) both the source and the drain are grounded; (b) only the drain is grounded and the source is floating.

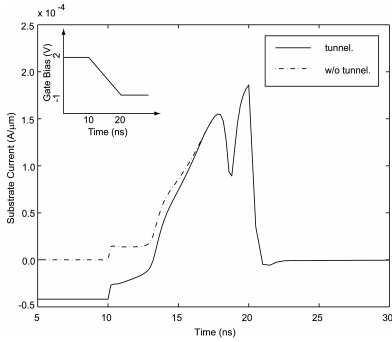


Figure 5: Simulated substrate currents at the falling edge of an ICP pulse with the gate tunneling model turned on and off, respectively. The device is an NMOS of 30μm channel length and 1.5nm gate oxide thickness.

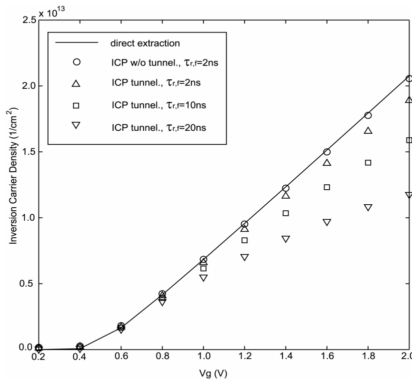


Figure 6: Extracted Ninv as a function of Vg from the following methods: direct extraction from device simulation outputs; ICP simulation with the gate tunneling turned off and on for various rise/fall times.

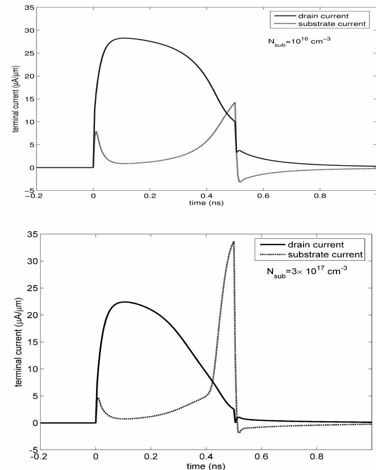


Figure 7: Simulated substrate and drain transient currents for PMOS of 1.5μm channel length and two substrate concentrations. The fall time of gate signal is 0.5ns. The charge-pumping is significant at large N_{sub}.

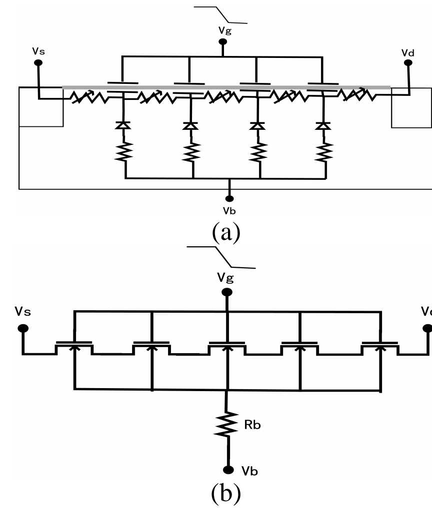
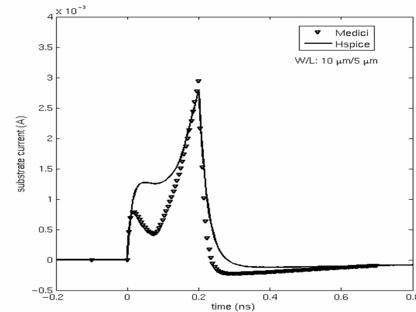
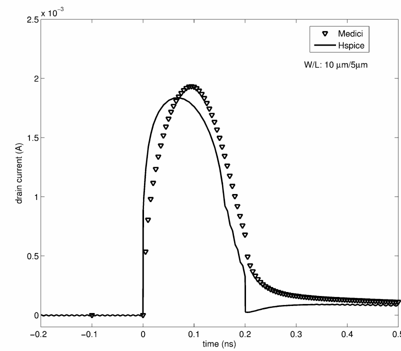


Figure 8: (a) Distributed equivalent circuit model of a long channel NMOS under large signal, NQS operations. (b) The equivalent channel segmentation model. The small component transistors are modeled by the BSIM3v3 QS model.



(a)



(b)

Figure 9: Turn-off transient simulations using MEDICI and Hspice for an NMOS of 5μm channel length: (a) substrate current; (b) drain current. The fall time of the gate signal is 0.2ns. In the Hspice simulation, the channel is divided into 10 segments. A total substrate resistance of 120Ω is used corresponding to a substrate doping concentration of $6E16cm^{-3}$.