Impact of Gate Direct Tunneling Current on Circuit Performance: A Simulation Study

Chang-Hoon Choi, *Student Member, IEEE*, Ki-Young Nam, *Student Member, IEEE*, Zhiping Yu, *Senior Member, IEEE*, and Robert W. Dutton, *Fellow, IEEE*

***Abstract*—The influence of gate direct tunneling current on ultrathin gate oxide MOS (1.1 nm $\leq t_{ox} \leq$ 1.5 nm, L_g = 50–70 nm) circuits has been studied based on detailed simulations. For the gate oxide thickness down to 1.1 nm, gate direct tunneling currents, including the edge direct tunneling (EDT), show only a minor impact on low V_{dd} static-logic circuits. However, dynamic logic and analog circuits are more significantly influenced by the off-state leakage current for oxide thickness below 1.5 nm, under low-voltage operation. Based on the study, the oxide thicknesses which ensure the International Technological Roadmap for Semiconductors (ITRS) gate leakage limit are outlined both for high-performance and low-power devices.**

***Index Terms*—Circuit simulation, CMOS inverters, device simulation, gate direct tunneling, static and dynamic circuits, thin gate oxide.**

I. Introduction

ACCORDING to the International Technological Roadmap for Semiconductors (ITRS), gate oxide thicknesses of 1.2–1.5 nm will be required by 2004 for sub-100-nm CMOS [1]. In this thin gate oxide regime, direct tunneling current increases exponentially with decreasing oxide thickness [2], which is of primary concern for CMOS scaling. For conventional CMOS devices, the dominant leakage mechanism is mainly due to short-channel effects owing to drain-induced barrier lowering (DIBL). In the ultrathin gate oxide regime, however, the gate leakage current can contribute significantly to off-state leakage, which may result in faulty circuit operation since designers may assume that there is no appreciable gate current. A recent study has shown that direct tunneling current appearing between the source–drain extension (SDE) and the gate overlap, so-called the edge direct tunneling (EDT), dominates off-state drive current, especially in very short channel devices [3], [4]. This results from the fact that the ratio of the gate overlap to the total channel length becomes large in the short-channel device compared to that of the long-channel device. Thus, the gate current effect is expected to become appreciable in ultrathin oxide, sub-100-nm MOS circuits. Even though many researchers have discussed the effects of gate leakage current, scaling limitations due to gate tunneling current from the viewpoint of circuit operation have not been critically addressed. Assessment of circuit immunity against the gate tunneling current, depending on various device structures and bias conditions, is of great importance in determining directions for future gate oxide scaling. This article considers circuit operation stability and oxide scaling limitations for several typical logic and nonlogic CMOS circuits using both device- and circuit-level simulation models.

Manuscript received February 27, 2001; revised July 9, 2001. This work was supported by the National Science Foundation (NSF) under the Distributed Center for Advanced Electronics Simulations (DesCArtES). The review of this paper was arranged by Editor G. Baccarani.

The authors are with the Center for Integrated Systems, Stanford University, Stanford, CA 94305 USA (e-mail: chchoi@stanford.edu).

Publisher Item Identifier S 0018-9383(01)10129-2.

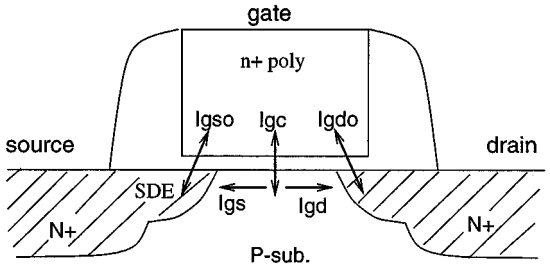


Fig. 1. Illustration of gate direct tunneling components of a very short-channel NMOSFET (I_{gso} and I_{gdo}) are EDT currents.

II. Gate Current Modeling

A. Edge Direct Tunneling (EDT)

Gate direct tunneling current is produced by the quantum-mechanical wavefunction of a charged carrier through the gate oxide potential barrier into the gate, which depends not only on the device structure but also bias conditions. Fig. 1 illustrates various gate tunneling components in a scaled NMOSFET; the gate-to-channel current (I_{gc}), and the EDT currents (I_{gso} and I_{gdo}) are shown. In long-channel devices, I_{gso} and I_{gdo} are less important than I_{gc} because the gate overlap length is small compared to the channel length. In very short channel devices, the portion of the gate overlap compared to the total gate length becomes larger.

Fig. 2 illustrates the band diagrams and electron tunneling directions along the gate-to-channel and gate-to-SDE directions for a highly doped drain (HDD) NMOSFET. For $V_g > 0$ V, the gate-to-channel tunneling current (I_{gc}) is the dominant current component, since a higher gate oxide voltage (V_{ox}) appears between the gate and the channel, as shown in Fig. 2(a). Here, the V_{fb} of an NMOSFET with an n-type polysilicon gate (i.e., n^+-poly/SiO_2/p-substrate) is approximately -1 V, while the V_{fb} along the gate-to-SDE (i.e., n^+-poly/ SiO_2/n^+ SDE) is approximately 0 V. On the contrary, the EDT currents (I_{gso} and I_{gdo}) can become dominant for bias conditions of $V_{fb} < V_g < 0$ V. For the gate-to-SDE case, electrons accumulated in the n^+-poly gate tunnel to the SDE region can lead to an appreciable off-state current. Meanwhile, operating in the depletion

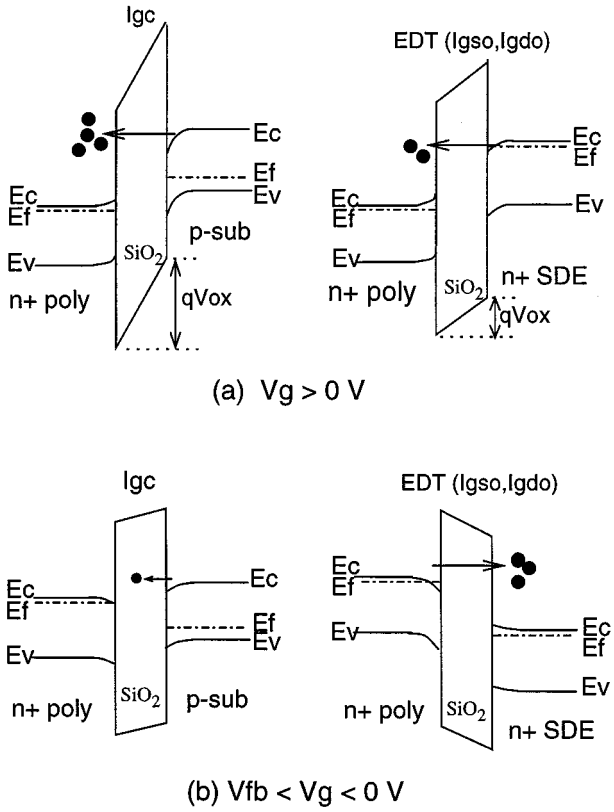


Fig. 2. Gate bias dependent band diagrams and electron tunneling in the channel ($I_{\rm gc}$) and the gate edge ($I_{\rm gso}$ and $I_{\rm gdo}$). (a) $V_{\rm g} > 0$ V (inversion mode). (b) $V_{\rm fb} < V_{\rm g} < 0$ V (depletion mode).

mode along the n^+-poly/SiO_2/p-substrate surface, few electrons are present in the channel that can in turn tunnel into the gate, as shown in Fig. 2(b).

B. Direct Tunneling Device Simulation

The EDT in the gate-to-SDE region must be treated as a two-dimensional problem in very short channel devices [4], owing to the laterally finite doping gradient in the SDE region and the drain electric field effects.

In order to model the EDT behavior, MEDICI [5] was used. For conduction band electron tunneling, the net direct tunneling current is calculated for the conduction band electrons, using the independent electron approximation given by [6]

$$J_{\rm DT} = \frac{4\pi q m_1 k_B T}{h^3} \times \int_0^{E_{\rm b}} TC(E) \ln\left[\frac{e^{(E_{\rm Fn1}-E_{\rm c1}-E)/k_BT}+1}{e^{(E_{\rm Fn3}-E_{\rm c3}-E)/k_BT}+1}\right] dE \quad (1)$$

where $E_{\rm Fn1}$, $E_{\rm c1}$, and m_1 are the electron quasi-fermi level, the conduction bandedge, and the electron effective tunneling mass in the silicon region, respectively. $E_{\rm Fn3}$ and $E_{\rm c3}$ are the electron quasi-fermi level and the conduction band in the polysilicon region, respectively. The endpoint for integration is determined by the barrier height in the silicon region $E_{\rm b}$. $TC(E)$ is the tunneling coefficient of an electron with kinetic energy of the incident electron (E) that describes the probability that an electron with a certain energy can tunnel through the oxide.

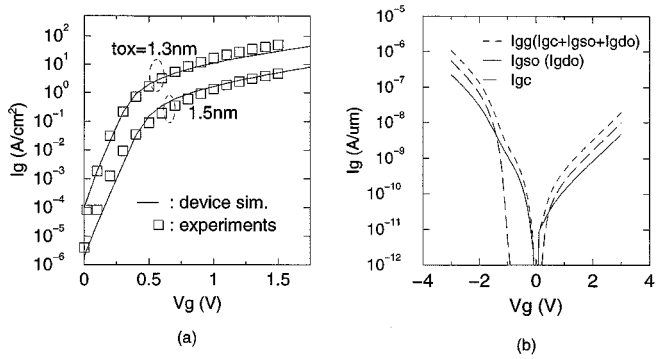


Fig. 3. Simulated gate currents by using MEDICI. (a) Simulated gate currents for a long-channel ($L_{\rm g} = 100\ \mu$m) NMOSFET, compared with measured data ($V_{\rm ds} = 0$ V). (b) Simulated $I_{\rm gc}$, $I_{\rm gso}$ and $I_{\rm gg}$ ($= I_{\rm gc} + I_{\rm gso} + I_{\rm gdo}$) for an NMOSFET with $t_{\rm ox} = 1.5$ nm and $L_{\rm g} = 50$ nm ($V_{\rm ds} = 0$ V).

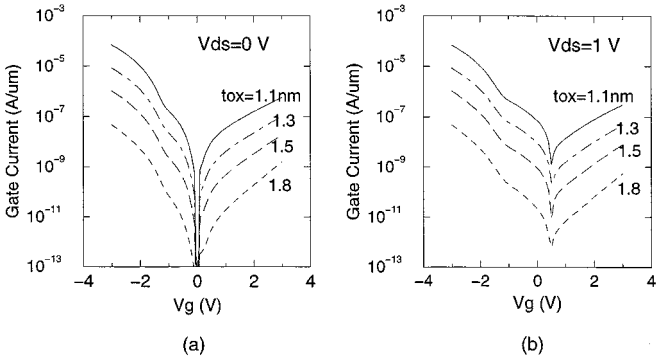


Fig. 4. Simulated $I_{\rm gg}$ ($= I_{\rm gc} + I_{\rm gso} + I_{\rm gdo}$) for different $t_{\rm ox}$s ranging from 1.1 to 1.8 nm and $L_{\rm g} = 50$ nm. (a) $V_{\rm ds} = 0$ V. (b) $V_{\rm ds} = 1$ V.

To validate the direct tunneling model, gate currents were simulated and compared to the experimental data of long-channel NMOSFETs (i.e., $W/L = 100\ \mu$m/$100\ \mu$m and $t_{\rm ox} = 1.3$ and 1.5 nm). Though agreement is not perfect, simulated gate currents from MEDICI show reasonable correspondence to the measurements, as reflected in Fig. 3(a). Device simulations were also performed for a very short channel NMOSFET with 50 nm gate length; Fig. 3(b) illustrates resulting gate currents for an NMOSFET with $t_{\rm ox} = 1.5$ nm. The source and drain are tied to ground and the gate bias is forced from negative to positive values. Note that the EDT current ($I_{\rm gso}$, $I_{\rm gdo}$) is higher than the gate-to-channel current ($I_{\rm gc}$) for gate biases of -1.5 V $< V_{\rm g} < 0$ V, implying that the EDT is the dominant leakage source for the off-state current in the low-voltage range of operation for MOS circuits. Fig. 4 shows the total simulated gate current ($I_{\rm gg} = I_{\rm gc} + I_{\rm gso} + I_{\rm gdo}$) for different gate oxide thicknesses, ranging from 1.1 to 1.8 nm ($V_{\rm ds} = 0$ and 1 V); gate current increases exponentially as the gate oxide thicknesses are scaled down.

III. Circuit Application

In order to evaluate circuit performance by considering gate direct tunneling effects, a macro-circuit model has been constructed in the circuit simulator HSPICE [7]. Gate direct tunneling currents, obtained from the device simulation for the gate oxide thicknesses of 1.1, 1.3, and 1.5 nm, are described using voltage-dependent current sources as a function of the

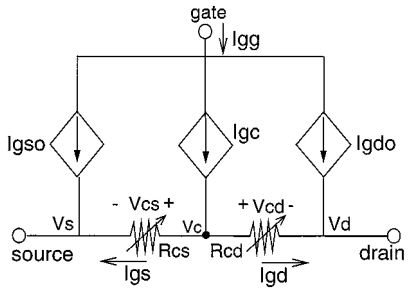


Fig. 5. Macro-circuit model for direct tunneling current combined circuit simulation.

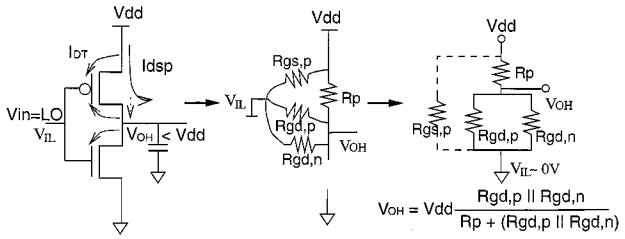


Fig. 6. $V_{\rm OH}$ modeling considering gate tunneling currents of a CMOS inverter.

terminal voltage, as shown in Fig. 5. The partitioning of $I_{\rm gc}$ into $I_{\rm gs}$ and $I_{\rm gd}$ is modeled by using variable resistances $R_{\rm cs}$ and $R_{\rm cd}$, respectively, in each part of the channel. $R_{\rm cs}$ and $R_{\rm cd}$ are the channel resistance corresponding to $0.5L_{\rm ch}$, where $R_{\rm cs} = V_{\rm cs}/I_{\rm cs}$ and $R_{\rm cd} = V_{\rm cd}/I_{\rm cd}$. $I_{\rm cs}$ and $I_{\rm cd}$ are channel currents of each region; they have been obtained by adjusting the BSIM3-model parameters to fit the current-voltage (I–V) curves generated from device simulation. The macro-circuit model has been applied to several MOS circuits—CMOS inverter, dynamic AND gate, and sample and hold (S/H).

A. Static CMOS Inverter

For the CMOS inverter application we assumed the amount of hole direct tunneling of the PMOSFET is the same as that of the NMOSFET. The magnitude of the channel current is assumed to be identical, regardless of the gate oxide thickness, in order to focus on the circuit performance difference based on the oxide thickness dependent gate tunneling current contributions. Estimated gate current paths during the operation are shown schematically in Fig. 6.

When the input is *low* and the gate tunneling current is significant, $V_{\rm out}$ (i.e., $V_{\rm OH}$) will not reach $V_{\rm dd}$ due to the leakage current that flows from the output node. Here, direct tunneling current components can be modeled as resistors and $V_{\rm OH}$ is approximated by the voltage divider

$$V_{\rm OH} \approx V_{\rm dd} \times \frac{R_{{\rm gd},p} \| R_{{\rm gd},n}}{R_p + (R_{{\rm gd},p} \| R_{{\rm gd},n})} \tag{2}$$

where R_p is the on-state channel resistance of the PMOS, and $R_{{\rm gd},n}$ and $R_{{\rm gd},p}$ are gate-to-drain resistances of the N- and P-MOSFET, respectively, modeling gate direct tunneling effects. The resistor values are approximately $R_{{\rm gd},p} \approx V_{\rm dd}/I_{{\rm gd},p}$ and $R_{{\rm gd},n} \approx V_{\rm dd}/I_{{\rm gd},n}$. As an example, if the ratio of $R_{{\rm gd},n}$ (or $R_{{\rm gd},p}$) to R_p is 100, then $V_{\rm OH}$ will drop by 2% from the $V_{\rm dd}$ level.

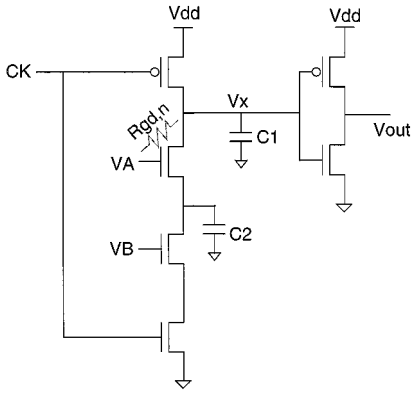


Fig. 7. Domino CMOS AND-2 gate.

$V_{\rm OL}$ is again approximated using a voltage divider

$$V_{\rm OL} \approx V_{\rm dd} \times \frac{R_n}{R_n + (R_{{\rm gd},n} \| R_{{\rm gd},p})} \tag{3}$$

where R_n is the on-state channel resistance of the NMOS. As a result, when tunneling current is significant $V_{\rm OL}$ will not fall to the GND level and the $V_{\rm out}$ swing $({\rm GND} < V_{\rm out} < V_{\rm dd})$ is reduced for very leaky, thin gate oxide CMOS inverters.

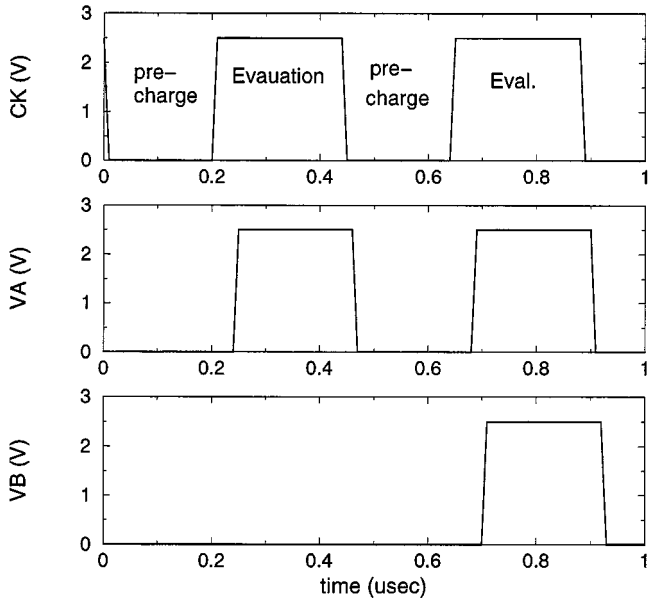
Again considering an example, assuming that $V_{\rm dd} = 2.5$ V, $I_{{\rm ds},n} = 0.50$ mA/μm, $I_{{\rm ds},p} = 0.25$ mA/μm, and $I_{{\rm gd},p} = I_{{\rm gd},n} = 5.0 \times 10^{-6}$ A/μm (i.e., $W_p = 2W_n = 20\ \mu$m), based on the simulations for $t_{\rm ox} = 1.1$ nm and $L_{\rm g} = 50$ nm, then $R_{{\rm gd},n}/R_n$ or $R_{{\rm gd},p}/R_p$ is about 100. In this case, the estimated $V_{\rm OH}$ and $V_{\rm OL}$ values are $V_{\rm OH} = 2.42$ V and $V_{\rm OL} = 0.08$ V, respectively, from (2) and (3), or a total reduced logic swing of 160 mV.

When $V_{\rm dd} = 2.5$ V, simulated $V_{\rm OH}$ and $V_{\rm OL}$ values using the macro-circuit model are 2.45 V and 0.04 V for $t_{\rm ox} = 1.1$ nm, respectively. The full logic-high $(V_{\rm dd})$ and logic-low (GND) levels are achieved for $t_{\rm ox} = 1.3$ and 1.5 nm. The average power consumption is 0.75, 0.16, and 0.09 mW for $t_{\rm ox} = 1.1$, 1.3, and 1.5 nm, respectively.

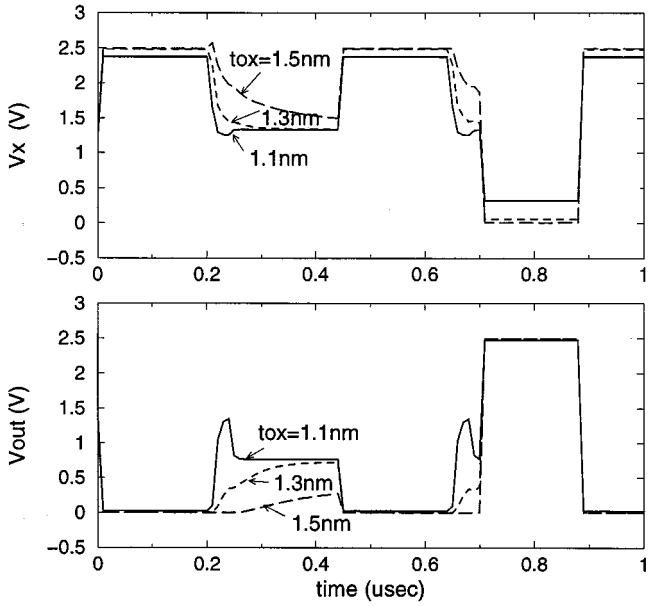
When $V_{\rm dd} = 1.5$ V, the output node of the inverter swings between full logic-high and logic-low (i.e., only 0.1% of $V_{\rm OH}$ reduction even for $t_{\rm ox} = 1.1$ nm). The power consumption for $V_{\rm dd} = 1.5$ V is exponentially reduced compared to the case for $V_{\rm dd} = 2.5$ V, due to the exponential decrease in gate current; a reduction of 20–100 times is realized, compared to the $V_{\rm dd} = 2.5$ V case. According to the ITRS, $V_{\rm dd}$ of 1.0–1.5 V is required for 70-nm CMOS technology. For the low $V_{\rm dd}$, gate direct tunneling current effects on static-logic circuits will be less serious for oxide thicknesses down to 1.1 nm.

B. Dynamic AND Gate

Consider the domino CMOS AND-2 gate shown in Fig. 7. Assume all inputs are low initially and the intermediate node voltage across C_2 has an initial value of 0 V. During the precharge phase, the output node capacitance (C_1) is charged up to its logic-high level of $V_{\rm dd}$ through the PMOS transistor. In the next phase, CK switches logic-high and the evaluation begins. If the input V_A switches from low to high during the evaluation phase, charge stored on C_1 will be shared with C_2, and the node voltage V_X drops after the charge-sharing.



(a)



(b)

Fig. 8. Simulated waveforms of the domino AND-2 gate. (a) Clock and input signals. (b) Output waveforms for $V_{\rm dd} = 2.5$ V.

When gate tunneling current becomes significant, V_X may be less than the logic-high level during the precharge phase, so that V_X can even drop to less than $V_{\rm dd}/2$ during the first evaluation period. As a result, $V_{\rm out}$ will inadvertently switch to a logic-high, resulting in a logic error.

Fig. 8 shows the simulated input and output waveforms of a domino AND-2 gate for $V_{\rm dd} = 2.5$. When $t_{\rm ox} = 1.1$ nm and $V_{\rm dd} = 2.5$ V, V_X drops to about 1.2 V during the evaluation phase due to gate tunneling current effects. As a result, $V_{\rm out}$

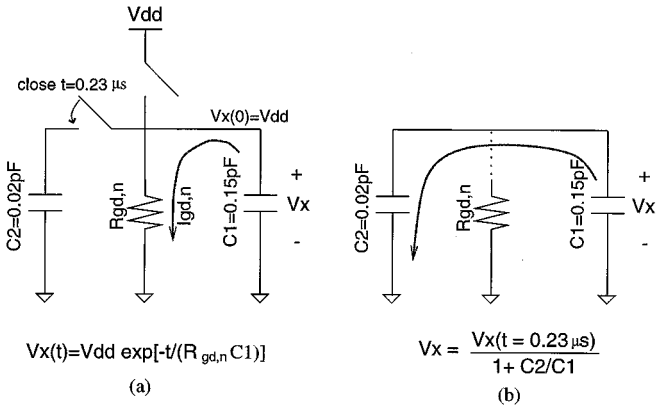


Fig. 9. Modeling of discharge and charge sharing behaviors during first evaluation period of domino AND-2 gate. (a) Discharge through tunneling resistance before V_A switches to logic-high ($t = 0.20$–0.23 μs). (b) Charge sharing of C_1 with C_3 after the V_A switches to logic-high ($t > 0.23$ μs).

erroneously switches during the first evaluation period and a glitch appears prior to the second evaluation phase.

These phenomena can be simply modeled as RC circuits, as shown in Fig. 9. Initially, V_X has a value of $V_{\rm dd}$ during the precharge phase. When the evaluation begins at $t = 0.2$ μs by switching of the CK signal to logic-high (i.e., V_A remains logic-low until $t = 0.23$ μs), charge stored in C_1 flows to ground through the tunneling resistance $(R_{{\rm gd},n})$, as illustrated in Fig. 9(a). During the discharge process the V_X level drops as a function of time according to the following:

$$V_X(t) = V_X(t_0)e^{-(t-t_0)/R_{{\rm gd},n}C_1} \approx V_{\rm dd}e^{-(I_{{\rm gd},n}(t-t_0))/V_{\rm dd}C_1} \tag{4}$$

where $t_0 = 0.20$ μs in the case. With a tunneling current of $I_{{\rm gd},n} = 4 \times 10^{-6}$ A, V_X drops to 1.60 V from its initial voltage of $V_{\rm dd} = 2.5$ V during 0.03 μs ($t = 0.20$–0.23 μs), according to (4). When V_A switches to logic-high at $t = 0.23$ μs, direct tunneling current is reduced due to a smaller voltage difference between V_X and V_A. Thus, instead of the discharge process, charge sharing begins; charge in C_1 is shared with C_2, as shown in Fig. 9(b). The final V_X after the charge sharing is approximated as follows:

$$V_X \approx \frac{V_X\ (t = 0.23\ \mu{\rm s})}{1 + \frac{C_2}{C_1}}. \tag{5}$$

When $V_X\ (t = 0.23\ \mu{\rm s}) = 1.60$ V, $C_1 = 0.15$ pF and $C_2 = 0.04$ pF, V_X after charge sharing estimated by (4) is about 1.26 V, which corresponds to $\sim V_{\rm dd}/2$ level. Hence, $V_{\rm out}$ may erroneously switch during the evaluation phases, as shown in the simulation results of Fig. 8(b). Even though these spurious results can be reduced by lowering $V_{\rm dd}$, the dynamic logic circuit may have potential problems due to gate tunneling-induced off-state current during the precharge-and-evaluation phases of operation.

C. Sample and Hold (S/H) Circuit

The S/H circuit is an important analog building block in data-converter systems used to acquire analog signals and to store the

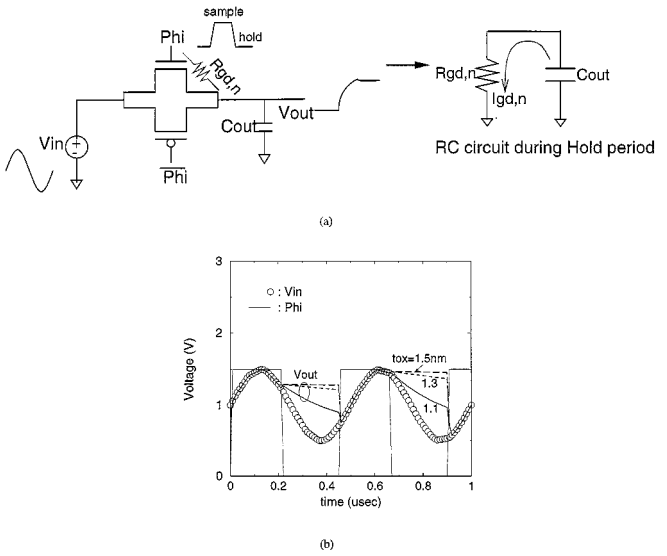


Fig. 10. CMOS S/H circuit and simulated waveforms for different $t_{\rm ox}$s. (a) CMOS S/H circuit schematic and RC circuit model during hold period. (b) Waveforms for $V_{\rm dd} = 1.5$ V.

value for some length of time. A simple S/H circuit is formed by a sampling CMOS switch followed by a hold capacitor, as shown in Fig. 10(a). When the clock (*Phi*) is high, $V_{\rm out}$ follows $V_{\rm in}$; when *Phi* goes low, $V_{\rm out}$ will ideally remain at a constant level. However, $V_{\rm out}$ will not hold this sampled value if leakage paths exist. This tunneling current-induced decay in $V_{\rm out}$ during the hold period can be modeled using the RC circuit shown in Fig. 10(a). As for the previous dynamic AND gate, $V_{\rm out}$ decays as a function of time, again using the expression given in (4).

Fig. 10(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. This implies that the S/H circuit has poor robustness in the face of gate leakage current, limiting its operation to oxide scaling only to the 1.5-nm regime.

D. Impact of Alternative Gate Dielectrics

Alternate insulating materials with a dielectric constant larger than that of SiO_2 are under evaluation to replace the conventional SiO_2 gate stack. Using such gate dielectrics, devices with lower gate-leakage current can be achieved as a result of the increased film thickness resulting from the increased dielectric constant of nitride-based layers ($\epsilon_{\rm nitride} \sim 7.8$). Device simulation for an alternative gate dielectric of Si_3N_4, assuming a dielectric constant of $\kappa = 2\epsilon_{\rm ox}$ and thickness of 2.6 nm ($t_{\rm ox,eq} = 1.3$ nm), shows about five orders of magnitude lower gate current compared to the pure oxide device with the same equivalent oxide thickness of $t_{\rm ox} = 1.3$ nm. Secondary effects such as surface roughness and interface traps are not considered.

Voltage bootstrapping is used to overcome threshold voltage drops in digital circuits. Fig. 11(a) shows a schematic of the bootstrapping circuit, including the bootstrap MOS capacitor; the voltage V_X is increased during the $V_{\rm in}$ switching event. As a result, the threshold voltage drop can be compensated for at the output node $V_{\rm out}$.

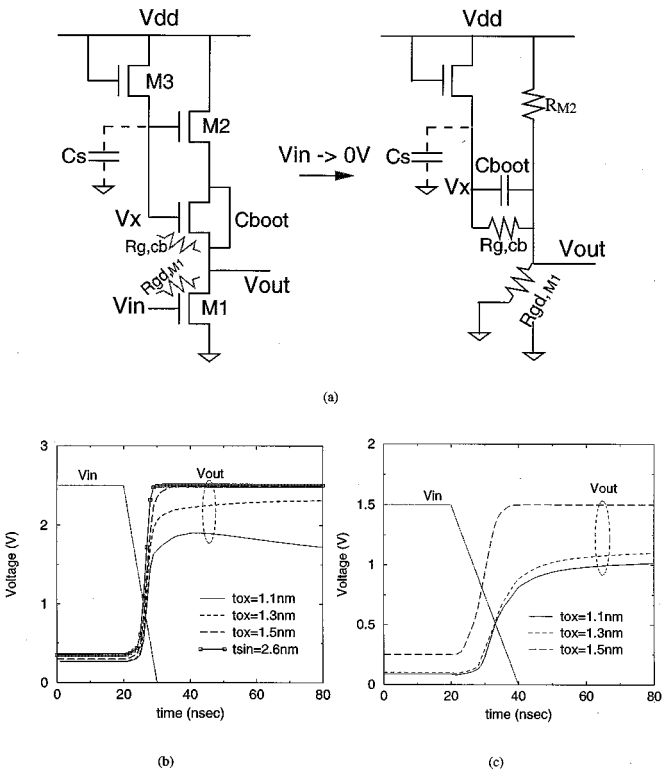


Fig. 11. Voltage bootstrapping circuit and simulation results. (a) Circuit schematic and its equivalent circuit when $V_{\rm in}$ switches to 0 V. (b) Simulated waveforms with an alternative gate dielectric (Si_3N_4) and pure oxides ($V_{\rm dd} = 2.5$ V). (c) Simulated waveforms for $V_{\rm dd} = 1.5$ V.

When $V_{\rm in}$ switches to logic-low, $V_{\rm out}$ and V_X are approximated as follows [8]:

$$V_{\rm out} \approx V_X - V_{{\rm th},M2} \tag{6}$$

where

$$V_X \approx (V_{\rm dd} - V_{{\rm th},M3}) + V_{\rm dd}\frac{C_{\rm boot}}{C_{\rm s} + C_{\rm boot}}. \tag{7}$$

$(V_{\rm dd} - V_{{\rm th},M3})$ is the initial condition of V_X and the second term of (7) represents the increase in V_X after the $V_{\rm in}$ switches to 0 V. However, this V_X expression should be modified to account for the gate tunneling current. First, the initial V_X is reduced due to the discharge via the gate leakage resistor (i.e., $R_{\rm g,cb}$)

$$(V_{\rm dd} - V_{{\rm th},M3}) \longrightarrow (V_{\rm dd} - V_{{\rm th},M3})e^{-(t/R_{\rm g,cb}C_{\rm boot})}. \tag{8}$$

In addition, the second term of (7) is modified due to the gate-to-drain resistance effects of $M1$

$$V_{\rm dd}\frac{C_{\rm boot}}{C_{\rm s} + C_{\rm boot}} \longrightarrow \left(V_{\rm dd}\frac{R_{{\rm gd},M1}}{R_{M2} + R_{{\rm gd},M1}}\right) \cdot \frac{C_{\rm boot}}{C_{\rm s} + C_{\rm boot}}. \tag{9}$$

Thus, both V_X and $V_{\rm out}$ will be reduced in the presence of substantial gate tunneling current.

Fig. 11(b) illustrates the simulated input and output waveforms of the bootstrapping circuit. For $t_{\rm ox} = 1.1$ and 1.3 nm, the final $V_{\rm out}$ does not reach $V_{\rm dd}$ because of the gate leakage resistors in both the MOS capacitor ($C_{\rm boot}$) and the driving transistor ($M1$) (i.e., $R_{\rm g,cb}$ and $R_{{\rm gd},M1}$). In contrast, a full $V_{\rm dd}$ output

TABLE I
ITRS LOGIC TECHNOLOGY ROADMAP (2000 EDITION) [1] AND THICKNESSES TO ENSURE THE ROADMAP REQUIREMENT CONSIDERING THE EDGE DIRECT TUNNELING LEAKAGE (DIELECTRICS WITH $\kappa = 2\epsilon_{ox}$ ARE ASSUMED FOR 2005–2011 NODE)

	Year	1999	2001	2004	2005	2008	2011
1	MPU gate length(nm)	120	100	70	65	45	30
2	Logic V_{dd} (V)	1.5–1.8	1.2–1.5	0.9–1.2	0.9–1.2	0.6–0.9	0.5–0.6
3	t_{ox} equivalent (nm)	1.9–2.5	1.5–1.9	1.2–1.5	1.0–1.5	0.8–1.2	0.6–0.8
4	I_{gate} limit (A/μm) for low-power device	0.07p	0.10p	0.16p	0.2p	0.4p	0.8p
5	I_{gate} limit (A/μm) for high-performance	70p	100p	160p	200p	400p	800p
6	Simulated I_{gate} (A/μm) for low-power device	150p	2.5n	19n	2p	65p	4.1n
7	Simulated I_{gate} (A/μm) for high-performance	450p	9.5n	50n	6p	240p	6n
8	t_{ox} (nm) to ensure low-power I_{gate} limit	~2.6	~2.4	~2.0	~1.1	~1.0	~0.9
9	t_{ox} (nm) to ensure hi-perform I_{gate} limit	~2.2	~1.9	~1.7	~1.0	~0.8	~0.7

voltage can be achieved by adopting a Si_3N_4 gate dielectric due to the substantial reduction in leakage current. Note that the equivalent oxide thickness of $t_{Si_3N_4} = 2.6$ nm is $t_{ox} = 1.3$ nm. Alternative gate dielectrics will be necessary to replace leaky gate oxides in MOS circuits, especially where charge conservation or charge bootstrapping techniques are required.

IV. DISCUSSION

It is instructive to estimate how the gate direct tunneling current (i.e., EDT current) will affect oxide scaling in the technology roadmap. A comparison has been made between the I_{gate} limit of the ITRS and the device simulation results. I_{gate} simulation for the low-power device was performed with a combination of the lower value of the V_{dd} range and the thinner oxide of the equivalent t_{ox} range to produce a maximum gate current value; the higher value of the V_{dd} range and the thinner oxide of the equivalent t_{ox} range were used for the high-performance case. As a result, the maximum I_{gate} limit of the technology roadmap across the 1999–2004 nodes is far below the simulated I_{gate}, as shown in Table I. In particular, the simulated I_{gate} for the low-power device is three to five orders of magnitude higher than the required limit. Namely, the maximum I_{gate} limits in Table I—Rows 4 and 5 are too strict for the oxide thickness range in the technology roadmap. The simulated I_{gate} beyond the 65-nm node (2005) is rather close to the roadmap requirement owing to the use of a nitride-based dielectric with $\kappa = 2\epsilon_{ox}$, but dielectrics with $\kappa > 2\epsilon_{ox}$ will be necessary beyond the 45-nm node technology, especially for the low-power device. In order to satisfy the roadmap requirements for the low-power device, the use of oxides thicker than the equivalent t_{ox} range of the roadmap is desirable. For example, an oxide thickness about 2.0 nm is required to ensure the I_{gate} limit for the 70-nm technology node (2004), as shown in Table I—Row 8. In conclusion, the oxide scaling suggested by the roadmap may be little aggressive, especially for the low-power devices. In order to ensure the I_{gate} limit of the technology roadmap, an early use of high κ dielectric materials or a more conservative oxide scaling (Table I—Rows 8 and 9) is necessary.

V. CONCLUSIONS

CMOS circuit robustness in the presence of gate tunneling currents has been studied using circuit simulation, combined with a macro-circuit model of gate tunneling current and analytic estimation of the effects. CMOS static inverters at $V_{dd} = 1.5$ V show acceptable noise margins with low-power consumption for the oxide thicknesses down to 1.1 nm, while dynamic AND gates have a potential weakness in the presence of gate current during the precharge and evaluation phases. For circuits that require charge-conservation or charge-bootstrapping, including the S/H circuit, significant performance degradation can be expected for $t_{ox} < 1.5$ nm, even considering low-voltage operation. A dual-gate oxide process or use of high-κ dielectric will be necessary on these circuits to continue device scaling. Based on the simulation studies, the oxide thicknesses to ensure the off-state gate leakage requirement of the ITRS roadmap are outlined.

REFERENCES

[1] *International Technology Roadmap for Semiconductors 2000 Edition*, Semiconductor Industry Assoc., Austin, TX, 2000.

[2] H. S. Momose, M. Ono, T. Yoshitomo, T. Ohguro, S. Nakamura, M. Saito, and H. Iwai, "1.5-nm direct-tunneling gate oxide Si MOSFETs," *IEEE Trans. Electron Devices*, vol. 43, pp. 1233–1242, Aug. 1996.

[3] K. N. Yang, H. T. Huang, M. J. Chen, Y. M. Lin, M. C. Yu, S. M. Jang, C. H. Yu, and M. S. Liang, "Edge hole direct tunneling in off-state ultrathin gate oxide p-channel MOSFETs," in IEDM Tech. Dig., 2000.

[4] N. Yang, W. K. Henson, and J. Wortman, "A comparative study of gate direct tunneling and drain leakage currents in N-MOSFETs with sub-2100-nm gate oxides," *IEEE Trans. Electron Devices*, vol. 47, pp. 1636–1644, Aug. 2000.

[5] *MEDICI: Two-Dimensional Semiconductor Device Simulation*, AVANT! Corp., Fremont, CA, 2000.

[6] A. Shanware, J. P. Shiely, and H. Z. Massoud, "Extraction of the gate oxide thickness of N- and P-channel MOSFETs below 20 Å from substrate current resulting from valence-band electron tunneling," in *IEDM Tech. Dig.*, 1999, pp. 815–818.

[7] *HSPICE User's Manual*, AVANT! Corp., Fremont, CA, 1999.

[8] S. M. Kang and Y. Leblebici, *CMOS Digital Integrated Circuits*. New York: McGraw-Hill, 1996.

Chang-Hoon Choi (S'97) received the B.S. and M.S. degrees in electronic engineering from Sogang University, Seoul, Korea, in 1988 and 1990, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at the Center for Integrated Systems, Stanford University, Stanford, CA.

From January 1990 through May 1997, he was with Samsung Electronics Co., Ltd., Kyungki-Do, Korea, where he was engaged in modeling and simulation on the IC circuits, devices, and TCAD. His research interests include characterization, modeling, and simulation of processes/devices for future CMOS generations.

Mr. Choi is a reviewer for IEEE TRANSACTIONS ON ELECTRON DEVICES. He is listed in *Who's Who in the World.*

Ki-Young Nam (S'97) was born in Seoul, Korea, in 1970. He received the B.S. degrees in electronic engineering and metallurgical engineering from Yonsei University, Seoul, Korea, in 1997, and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1999, where he is currently pursuing the Ph.D. degree.

During the summer of 1999, he worked on the modeling of the sigma–delta modulated fractional-N frequency synthesizer at Texas Instruments, Dallas, TX. His current research interests include the design of CMOS low-power analog-to-digital converters for wideband applications.

Zhiping Yu (M'90–SM'94) received the B.S. degree from Tsinghua University, Beijing, China, in 1967, and the M.S. and Ph.D degrees from Stanford University, Stanford, CA, in 1980 and 1985, respectively.

He is presently a Senior Research Scientist in the Department of Electrical Engineering, Stanford University, and also holds a full professorship at Tsinghua University. His research interests focus on IC process, device, and circuit simulation, and in particular, the numerical techniques and modeling of RF and heterostructure devices. He has been involved in efforts to develop a simulation package for optoelectronic devices and 3-D solid modeling for ICs. Besides the full-time university research, he is a Consultant to HP Computer System and Technology Lab, HP, developing advanced transport models for sub-quarter-micrometer CMOS technology, including quantum mechanical effects.

Dr. Yu is currently serving as the Associate Editor of IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS responsible for TCAD related field.

Robert W. Dutton (S'67–M'70–SM'80–F'84) received the B.S., M.S., and Ph.D. degrees from the University of California, Berkeley, in 1966, 1967, and 1970, respectively.

He is currently Professor of Electrical Engineering at Stanford University, Stanford, CA, and Director of Research in the Center for Integrated Systems. He has held summer staff positions at Fairchild, Bell Telephone Laboratories, Hewlett-Packard, IBM Research, and Matsushita during 1967, 1973, 1975, 1977, and 1988, respectively. His research interests focus on integrated circuit process, device, and circuit technologies, especially the use of computer-aided design and parallel computational methods. He has published more than 200 journal articles and graduated more than four dozen doctorate students.

Dr. Dutton was Editor of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS from 1984 to 1986. He was the recipient of the 1987 IEEE J. J. Ebers Award and the 1988 Guggenheim Fellowship to study in Japan, was elected to the National Academy of Engineering in 1991, and the winner of the Jack A Morton Award for 1996.