Predicting CMOS speed with gate oxide and voltage scaling and interconnect loading effects

Article *in* IEEE Transactions on Electron Devices · December 1997

DOI: 10.1109/16.641365 · Source: IEEE Xplore

5 authors, including:

Chenming hu
University of California, Berkeley
1,148 PUBLICATIONS **52,867** CITATIONS

SEE PROFILE

Poya Fang
Chia Nan University of Pharmacy and Science
18 PUBLICATIONS **484** CITATIONS

SEE PROFILE

Predicting CMOS Speed with Gate Oxide and Voltage Scaling and Interconnect Loading Effects

Kai Chen, *Member, IEEE,* Chenming Hu, *Fellow, IEEE,* Peng Fang, *Member, IEEE,* Min Ren Lin, and Donald L. Wollesen, *Associate Member, IEEE*

***Abstract*—Sub-quarter micron MOSFET's and ring oscillators with 2.5–6 nm physical gate oxide thicknesses have been studied at supply voltages of 1.5–3.3 V. I_{dsat} can be accurately predicted from a universal mobility model and a current model considering velocity saturation and parasitic series resistance. Gate delay and the optimal gate oxide thickness were modeled and predicted. Optimal gate oxide thicknesses for different interconnect loading are highlighted.**

I. Introduction

TECHNOLOGY road map and strategic planning of future MOS device and IC performance has been in high demand. So far there is no widely accepted tools or models that can deliver such capability. Empirical SPICE models can only do well in curve fitting the devices already fabricated and characterized. They are unreliable to predict future IC performance with reasonable accuracy. Some physical compact models such as BSIM3v3 *Berkeley Short-channel IGFET Model* 3 version 3.0) [1] may be able to do so, but this is yet to be confirmed. Even some two-dimensional (2-D) device simulators have had difficulty in making predictions. They are even less trustworthy when evaluating IC speed.

This paper attempts to address this important issue. Predicting CMOS ring oscillator (RO) propagation delay, t_{pd}, with gate oxide and voltage scaling is the goal. To achieve this goal, analytical equations have been developed for 1) universal MOSFET inversion layer carrier mobility model solely expressed in terms of V_{gs}, V_{th}, and T_{ox}, 2) the resulting accurate drain saturation current, I_{dsat}, including effects of velocity saturation, mobility degradation and LDD parasitic resistance, 3) load capacitance, C_L, and 4) propagation delay of CMOS RO, t_{pd}. To confirm the new I_{dsat} and t_{pd} models, sub-quarter micron CMOSFET's and ring oscillators were fabricated with gate oxide thicknesses of 2.5–5.9 nm and effective channel length down to 0.22 μm and were characterized at supply voltages from 1.5–3.3 V. It was shown that gate oxide and voltage scaling as well as interconnect loading effects on CMOS speed can be modeled and predicted.

Manuscript received December 13, 1997; revised May 2, 1997. This work was supported by the Joint Services Electronics Program, F49620-94-C-0038, and AFOSR F49620-94-1-0464, SRC Contract 96-SJ-417.

K. Chen was with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720-1772 USA. He is now with IBM Corporation, Hopewell Junction, NY 12533 USA.

C. Hu is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720-1772 USA.

P. Fang, M. R. Lin, and D. L. Wollesen are with Advanced Micro Devices, Sunnyvale, CA 94088 USA.

Publisher Item Identifier S 0018-9383(97)07733-2.

The paper is organized as follows. Section II, Modeling and Characterization, is divided into four sub-sections to discuss the development of analytical equations of electron and hole mobility, the drain saturation current, load capacitance, and propagation delay, respectively, with measurement data. Section III illustrates the predictive capability of the new gate delay model and its application in studying the impact of interconnect loading effect on IC speed. Section IV summarizes the results.

II. Modeling and Characterization

A. Universal Mobility Model Solely Dependent on V_{gs}, V_{th}, and T_{ox}

It has been known that MOSFET's carrier mobility depends on gate voltage, V_{gs}, body bias, V_{bs}, gate oxide thickness, T_{ox}, and channel doping concentration, N_{sub}. The universal dependence can be represented by a single parameter, effective vertical electric field in the MOSFET inversion layer [3]

$$E_{eff} = \frac{\eta Q_{inv} + Q_b}{\varepsilon_s} \tag{1}$$

where Q_{inv} and Q_b are the inversion and body charge densities, respectively, ε_s is the relative permitivity of silicon material, and η is a dimensionless constant. Unfortunately, this E_{eff} is not a convenient quantity to evaluate.

Based on this concept, a new universal mobility model, solely based on commonplace parameters such as T_{ox}, V_{th}, and V_{gs} has been developed recently [2]. The E_{eff} for NMOS electrons and PMOS holes can be expressed as

$$E_{eff} = C_{ox}\frac{\frac{V_{gs}-V_{th}}{2}+V_{th}}{\varepsilon_s} = \frac{\varepsilon_{ox}}{T_{ox}}\frac{V_{gs}+V_{th}}{2\varepsilon_s} = \frac{V_{gs}+V_{th}}{6T_{ox}} \tag{2}$$

and

$$E_{eff} \approx \frac{\frac{Q_p}{2.5}+Q_b}{\varepsilon_{\mathrm{Si}}} \approx \frac{V_{gs}+1.5V_{th}-\alpha}{7.5T_{ox}} \tag{3}$$

respectively, where C_{ox} is the gate oxide capacitance per unit area, V_{gs} and V_{th} are the gate bias and threshold voltage of MOSFET, respectively, V_{fb} is the flat-band voltage, Ψ_s is the surface potential of MOSFET substrate, ε_{ox} and $\varepsilon_{\mathrm{Si}}$ are the relative permitivity of silicon dioxide and Si, respectively. For PMOS holes, V_{gs} and V_{th} are taken as positive numbers and $\alpha = 0$ and 2.3 for surface channel PMOSFET with p^+-poly

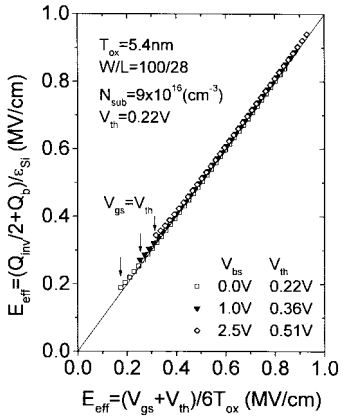

Fig. 1. $E_{eff} = (V_{gs}+V_{th})/6T_{ox}$ is a very good substitute for the original $E_{eff} = (Q_{inv}/2 + Q_b)/\varepsilon_s$ for NMOSFET.

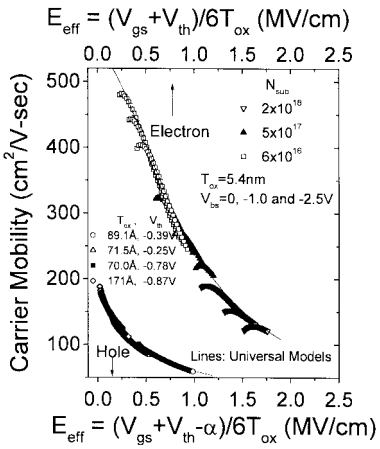

Fig. 2. New carrier universal mobility model for both NMOS electrons and PMOS holes are shown to fit experimental data of different technologies. Wafers were fabricated in six different laboratories.

and buried channel PMOSFET with n$^+$-poly, respectively (and $\alpha = 2.7$ for n$^+$-poly PMOSFET without boron implantation). Fig. 1 shows the comparison between the two expressions of E_{eff} in (1) and (2). Q_{inv} and Q_b for (1) was obtained from C–V measurement.

Using the new expressions for E_{eff}, new empirical MOSFET carrier mobility equations similar to that of [4] is found as follows:

$$\mu_n(V_{gs}, V_{th}, T_{ox}) = \frac{540}{1+\left(\frac{E_{eff}}{0.9}\right)^{1.85}} = \frac{540}{1+\left(\frac{V_{gs}+V_{th}}{5.4T_{ox}}\right)^{1.85}} \quad (4)$$

for NMOS electrons and

$$\mu_p(V_{gs}, V_{th}, T_{ox}) = \frac{185}{1+\left(\frac{E_{eff}}{0.45}\right)} = \frac{185}{1+\frac{\left(\frac{V_{gs}+1.5V_{th}-a}{7.5T_{ox}}\right)}{0.45}} \quad (5)$$

for holes, where the units for μ_n, E_{eff}, V_{gs} and V_{th}, and T_{ox} are cm^2/(V $\cdot$ s), MV/cm, MV, and cm, respectively. For the first time, hole mobility for all types of PMOSFET's was found to observe the same universal mobility equation (5), with only a switch of "α."

These expressions have been verified with measurement data taken from numerous MOSFET's fabricated in different laboratories worldwide [2], as shown in Fig. 2.

The mobility being modeled here in (4) and (5) is low lateral field mobility. Effect of drain bias is considered through current modeling to be presented later. Measured mobility is obtained from measured inversion channel charge (integration of split C_{gc}–V_{gs} curve) and the drain current in linear region.

It has been shown that carrier mobility of both NMOSFET electrons and all types of PMOSFET holes can be predicted if physical and process parameters such as T_{ox}, V_{th}, and V_{gs} are given. For the first time, it was found that holes of both surface and buried channel PMOSFET's observe the same universal mobility model except that a buried channel PMOSFET has the same mobility as a surface PMOSFET at $\sim$2 V lower V_{gs} assuming the same V_{th} and T_{ox} [2]. The closed form expressions of carrier mobility degradation as a function of V_{gs}, V_{th}, and T_{ox} are the corner stones of analytical I_{dsat} model to be presented next.

B. Drain Saturation Current, I_{dsat}

Among all MOSFET parameters, saturation drain current I_{dsat} has the strongest impact on circuit speed, therefore, it is one of the most important device parameters. Yet, the following poor approximation for I_{dsat} has been used by people in many circumstances to analyze or even to predict the effect of T_{ox}, L_{eff}, and V_{dd} because no analytical I_{dsat} model for deep sub-micron MOSFET's was available:

$$I_{dsat} = \frac{1}{2}\left(\frac{W_{eff}}{L_{eff}}\right)\mu_{eff}\left(\frac{\varepsilon_{ox}}{T_{ox}}\right)(V_{gs}-V_{th})^2 \quad (6)$$

where W_{eff}, L_{eff}, T_{ox}, ε_{ox}, V_{th}, V_{gs}, and μ_{eff} are effective channel width, length, gate oxide thickness, dielectric constant of silicon dioxide, threshold voltage, gate bias, and a constant mobility, respectively. This basic textbook model is very inadequate for today's MOSFET's because the effects of velocity saturation, short channel effect (e.g., "V_{th} roll-off"), mobility degradation with the increased vertical channel field, and source and drain series resistance of LDD structures, $R_d = R_s$, were not considered in this long channel MOS equation.

An accurate I_{dsat} model for MOSFET has been developed [5]. The mobility degradation is considered by the models (1)–(5) discussed in part A above. To account for the short channel effect of "V_{th} roll-off," the V_{th} used for this study are the measured values for each respective L_{eff}. Now, the velocity saturation can be considered by the following equation [6]:

$$I_{dsato} = Wv_{sat}C_{ox}(V_{gs}-V_{th}-V_{dsat}) = Wv_{sat}C_{ox}\frac{(V_{gs}-V_{th})^2}{V_{gs}-V_{th}+E_{sat}L_{eff}} \quad (7)$$

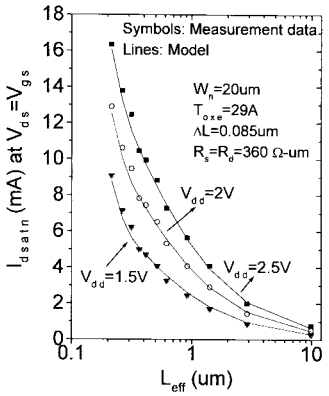

Fig. 3. Diving current prediction by new I_{dsat} model fits the measurement data well for NMOSFET's for wide range of L_{eff} with T_{ox} of 2.5 nm at V_{dd} from 1.5, 2, and 2.5 V.

where saturation velocity $v_{sat} = 8 \times 10^6$ cm/s, V_{dsat}, and $E_{sat} = 2v_{sat}/\mu_{eff}$ are the drain saturation voltage and the electric field corresponding to velocity saturation, respectively. The series resistance at the source end, R_s, are typically 300–500 Ω-μm. The effective gate bias $(V_{gs} - V_{th})$ is reduced by $I_{dsat}R_s \sim 0.2$ to 0.3 to $(V_{gs} - V_{th} - I_{dsat}R_s)$. Hence, the effect of R_s becomes more severe at lower V_{dd}. To account for the effect of R_s on I_{dsat}, (7) can be rewritten as

$$I_{dsat} = W v_{sat} C_{ox} \frac{(V_{gs} - V_{th} - I_{dsat}R_s)^2}{V_{gs} - V_{th} - I_{dsat}R_s + E_{sat}L_{eff}}. \quad (8)$$

Solving the quadratic equation for I_{dsat}, we obtain [5] as shown in (9), at the bottom of the page, where $V_1 = (V_{gs} - V_{th}) + E_{sat}L_{eff}$. The first order Taylor expansion of (9) leads to the following approximation for $I_{dsat}(R_s)$:

$$I_{dsat}(R_s) = \frac{I_{dsato}}{1 - \dfrac{2I_{dsato}R_s}{V_{gs} - V_{th}} + \dfrac{I_{dsato}R_s}{V_{gs} - V_{th} + E_{sat}L_{eff}}} \quad (10)$$

where $I_{dsato} \equiv I_{dsat}(R_s = 0)$ is given by (7). Fig. 3 shows that the model fits measurement data well for wide range of L_{eff} and V_{dd}. More verifications of this I_{dsat} model with measurement data can be found in [5] and [7].

It should be noted that to compare this new I_{dsat} model with the measurement data of thin gate oxide MOSFET's, electrical measured gate oxide thickness, T_{oxe}, instead of physical thickness normally monitored by optical measurement or other physical characterizations such as tunneling current method, should be used. Further discussions on this topic can be found in the Appendix of this paper and [7] and [8]. The L_{eff} is determined by the conventional textbook method as described in [9]. Threshold voltage V_{th} used here was defined and characterized as the linear extrapolation of I_{ds} versus V_{gs} for small V_{ds} for relatively long MOSFET's. For smaller MOSFET's, constant current V_{th} method is recommended in order to consider drain induced barrier lowering (DIBL) or other short channel effects on V_{th}.

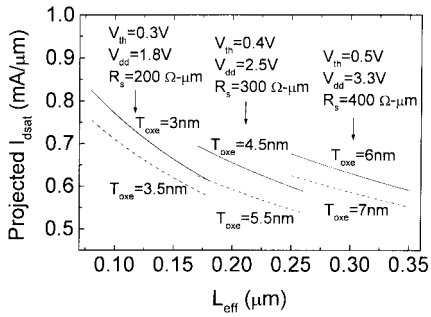

Fig. 4. Predicted I_{dsat} by the new model for current and future technologies as well as various power supply voltage, V_{dd}. The V_{th} is fixed as constant for the plot because for mature technologies, V_{th} does not change much for wide range of L_{eff}.

To get a general trend of future technology's driving capability, I_{dsat} versus L_{eff} for fixed V_{th} and R_s with different T_{oxe} and power supply voltage are plotted in Fig. 4. For even smaller $L_{eff} < 0.1$ μm, velocity overshoot may start kicking in and the accuracy of this model may need modification accordingly. Fig. 4 illustrates that due to mobility degradation and power supply voltage scaling, I_{dsat} will remain at the range of 0.6 to 0.8 mA/μm in the future.

To obtain a handy equation similar to (6) for quick evaluation of device and supply voltage scaling on I_{dsat}, numerous simulation for different conditions using the new accurate I_{dsat} model has been carried out. It was found that the following empirical equation is a good approximation for I_{dsat} projection for deep sub-micron MOSFETs:

$$I_{dsat} = k(R_s)L_{eff}^{-0.5}T_{ox}^{-0.8}(V_{gs} - V_{th})^{1.25}. \quad (11)$$

Comparing (6) and (11) for deep sub-micron and long channel MOSFET's, respectively, it shows that scale down L_{eff} or increase V_{gs} or V_{dd} for deep sub-micron MOSFET's will gain less than the long channel case due to velocity saturation. Similar conclusion can be drawn for scaling down T_{ox} due to mobility degradation. Equation (11) also indicates that if L_{eff}, T_{ox}, and $(V_{gs} - V_{th})$ all scale together with the same factor, I_{dsat} will remain about the same magnitude because the scaling factor of these three major parameters cancel out, as indicated by Fig. 4, to the first order.

In summary, MOSFET performance with device size and voltage scaling can be predicted by this new I_{dsat} model. The discussion has been limited for the NMOSFET case. Similar work can be done for PMOSFET's as well. A handy equation to conveniently evaluate scaling impact of L_{eff}, T_{ox}, V_{gs}, and V_{th} on I_{dsat} has been given in (11).

C. Experimental Characterization of Load Capacitance, C_L

To analytically evaluate CMOS gate propagation delay, it is necessary to develop a closed-form equation of load capacitance, C_L. This section proposes a simple C_L expression for CMOS ring oscillator propagation delay (t_{pd}) calculation as

$$I_{dsat} = \frac{[V_1 + 2(V_{gs} - V_{th})R_s W v_{sat} C_{ox}] - \sqrt{V_1^2 + 4(V_{gs} - V_{th})R_s E_{sat} L_{eff} W v_{sat} C_{ox}}}{2(R_s + W v_{sat} C_{ox} R_s^2)} \quad (9)$$

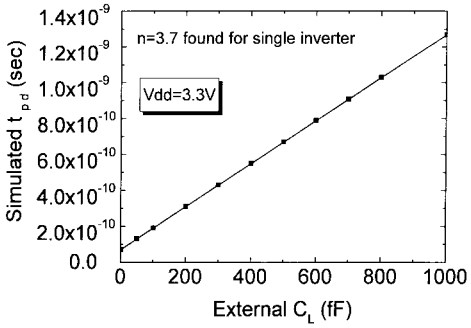

Fig. 5. SPICE simulation (device model is BSIM3v3) shows that $n = 3.7$.

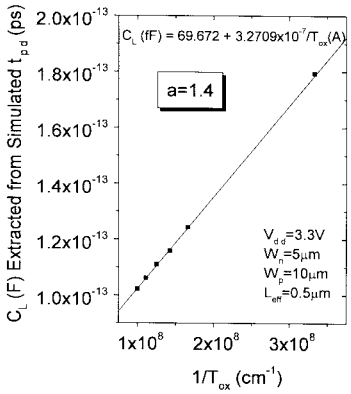

Fig. 6. SPICE simulation (device model is BSIM3v3) yields that $a = 1.5$.

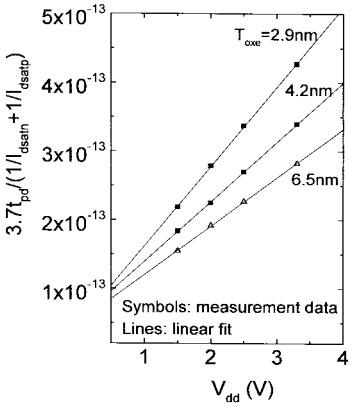

Fig. 7. The slope of each curve yields C_L corresponding to each T_{oxe}.

well as the experimental methods to characterize this proposed C_L expression.

The load capacitance of CMOS ring oscillator can be expressed as T_{ox}-related and non-T_{ox}-related capacitances as follows:

$$C_L = C_1 + \frac{A}{T_{oxeff}} + C_{int} = C_1 + a\frac{WL\varepsilon_{ox}}{T_{oxeff}} + C_{int} \quad (12)$$

where C_{int} represents interconnect capacitance, A/T_{oxeff} or $a(WL\varepsilon_{ox}/T_{oxeff})$ represents gate oxide related capacitance such as gate to channel oxide and gate-to-source or drain overlap capacitance where the factor "a" is a Miller-effect constant to be determined, T_{oxe} is the oxide thickness determined by C–V technique and is discussed in the Appendix at the end of this paper, C_1 represents the remainder of capacitance, presumably the junction capacitance, W and L are gate width and length of MOSFET, respectively. Value of constant "a" should be larger than unity due to Miller effect and fringing capacitance. Once the values of "a" and C_1 are known for the unloaded ring oscillator, (12) can be substituted into the following equation to calculate the propagation delay of CMOS ring oscillator, t_{pd} [10]:

$$t_{pd} = \frac{C_L V_{dd}}{n}\left(\frac{1}{I_{dsatn}} + \frac{1}{I_{dsatp}}\right) \quad (13)$$

where "n" is a constant which can be determined by SPICE simulations. Fig. 5 shows the simulation result of $n = 3.7$ for CMOS ring oscillator of fan-out equals to one. The device model of the simulation was BSIM3v3 with parameters extracted from a typical 0.35 μm technology wafer. The external load capacitance C_L for each stage of ring oscillator is varied to get t_{pd}. Simulations with different technologies and V_{dd} other than 3.3 V resulted in similar n value. So $n = 3.7$ is relatively V_{dd} and process independent.

To determine "a," again SPICE simulation can help. The T_{ox} in the model card of SPICE input deck was varied. For each gate oxide thickness, T_{ox}, simulation results of the following three quantities, t_{pd}, I_{dsatn}, and I_{dsatp} were read. Then $C_L = 3.7t_{pd}/[(1/I_{dsatn} + 1/I_{dsatp})V_{dd}]$ versus T_{ox} is plotted as shown in Fig. 6. The slope of this linear plots yields $a = 1.4$ for $n = 3.7$.

To experimentally find C_1 and "a," one method is that for each gate oxide thickness, measure ring oscillator's t_{pd}, I_{dsatn}, and I_{dsatp} at different V_{dd}. When the measurement data $nt_{pd}/(1/I_{dsatn} + 1/I_{dsatp})$ is plotted against V_{dd}, the slope of such a plot yields C_L, as shown in Fig. 7. The slope of each straight line gives the C_L for each T_{oxe}. Then plot C_L against $1/T_{oxe}$ as shown in Fig. 8 (the solid symbols). The slope and Y-intercept of Fig. 8 yield "a" and C_1, respectively, according to (12), if C_{int} can be ignored for the unloaded ring oscillators. The measurement data for the particular set of wafers presented in this paper yields that $C_1 = 18.3$ fF and $a = 1.37$ or $A = 283$ (fF-nm) [7].

The other experimental method to characterize C_1 and "a" to measure the dynamic current I_{dd} and oscillating frequency f of a ring oscillator at different supply voltages, V_{dd}. According to [11]:

$$I_{dd} = kfC_L V_{dd} = kf\left(C_1 + \frac{A}{T_{oxe}}\right)V_{dd} \quad (14)$$

(k is taken to be "1" here because normally the spike (short-circuit) current is negligible), C_L can be obtained from the slope of the plot I_{dd}/f versus V_{dd}. Then plot C_L against $1/T_{oxe}$. "C_1" and "a" can be determined from the Y-intercept and the slope of the straight line as shown in Fig. 8 (open symbol). So two independent methods yield almost the same C_1 and a in (12). Further discussions on the two experimental characterizations of C_L can be found in [12].

The value of "a" experimentally characterized is close to the value obtained by SPICE simulation. It should be noted that the wafers of above characterization of C_L is C_{ox} dominated. For junction dominated case, C_L is more V_{dd} and process dependent. More accurate characterization on C_L is an undergoing research.

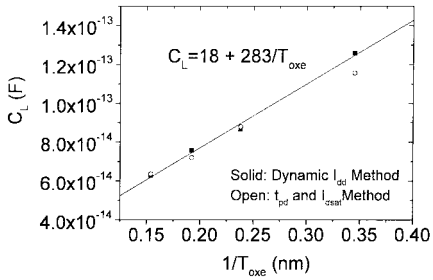

Fig. 8. Two independent ways to experimentally characterize C_L achieves close result: $C_L = 18 + 283/T_{oxe}$ (nm) for this particular set of wafers.

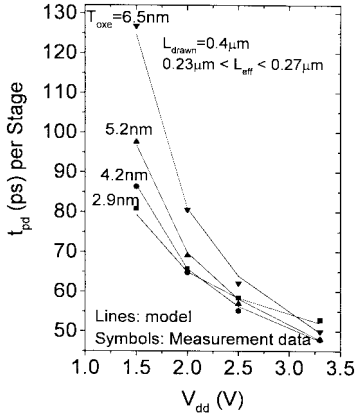

Fig. 9. The t_{pd} predicted by the new analytical gate delay model fits the measurement data well for physical gate oxide thicknesses from 2.5 to 5.9 nm at voltages of 1.5, 2, 2.5, and 3.3 V. The crossovers indicate the existence of optimal gate oxide as shown in Fig. 10.

D. Analytical Propagation Delay Equation and Its Experimental Confirmation

Substituting C_L calculated from (12) with $a = 1.37$ or $A = 283$ (fF-nm), $C_1 = 18.3$ (fF) and $C_{int} = 0$ into (13), t_{pd} can be predicted for various T_{oxe}, L_{eff}, V_{th}, R_s, and V_{dd}. The prediction made by this model and the measurement data are compared in Fig. 9. It shows that accurate prediction on CMOS speed can be made for a wide range of power supply voltages from 1.5–3.3 V for a wide variation of (physical) gate oxide thickness from 2.5–5.9 nm. The crossover of curves corresponding different T_{ox} indicates that there exists an optimal gate oxide thickness for given V_{th}, L_{eff}, and V_{dd}, as also experimentally observed in [13]. The existence of optimal T_{ox} can be understood because two competing processes are involved when T_{ox} is reduced: the increase in channel charge, $Q_{inv} \propto \varepsilon_{ox}/T_{ox}$, is insufficient to compensate for the decrease in mobility and the increase in gate oxide related capacitance, A/T_{oxe}.

III. Prediction and Interconnect Loading Effect

A. Projection of Future IC Speed With Gate Oxide and Voltage Scaling

The experimental confirmation of the gate delay model gives us confidence to make some projection of CMOS IC speed. In next two prediction plots, we use $W_{\rm p}/W_{\rm n} = 1.4/1$ although it is common to use $W_{\rm p}/W_{\rm n} = J_{\rm n}/J_{\rm p} \approx 2$, where $J_{\rm n}$ and $J_{\rm p}$ are the I_{dsat} per unit width for N- and P-MOSFET, respectively. It can be easily shown that for the given total

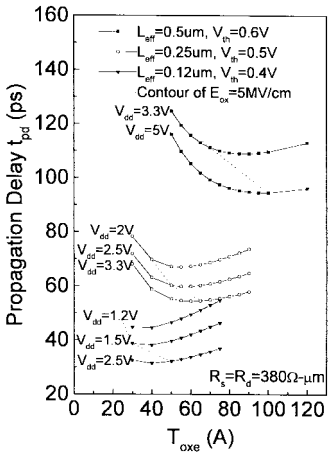

Fig. 10. Minimum t_{pd} may be reached at a thicker T_{ox} than that allowed by reliability requirement represented by the dashed contour of $E_{ox} = 5$ MV/cm in the figure.

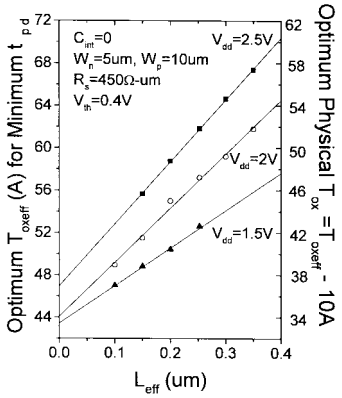

Fig. 11. Optimal T_{oxe} versus L_{eff} at power supply voltages of 1.5, 2, and 2.5 V for the CMOS ring oscillator with $C_{int} = 0$.

area of an inverter, i.e., given $(W_{\rm n} + W_{\rm p})$, a smaller width ratio $W_{\rm p}/W_{\rm n} \approx \sqrt{J_{\rm n}/J_{\rm p}} \approx 1.4$ minimizes t_{pd}.

Fig. 10 shows the prediction for various gate oxide thicknesses, supply voltages and threshold voltages, etc. If we assume that the oxide reliability limit is $E_{ox} < 5$ MV/cm (the dotted contour lines in Fig. 10), it can be seen that minimum t_{pd} may be reached at a thicker T_{ox} than reliability allows.

Fig. 11 shows how the optimal T_{oxe} for different effective channel length, L_{eff}, change with supply voltage. For lower supply voltage, optimal T_{ox} becomes less sensitive to channel length variation.

B. Interconnect Loading Effects on t_{pd}

So far, only propagation delay, t_{pd}, of unloaded CMOS ring oscillators has been studied. In real integrated circuits, particularly for the sub-quarter micron technologies, interconnect capacitance, C_{int}, can not always be ignored [14].

As mentioned in Section II-C above, for the CMOS ring oscillators fabricated for this particular study, $W_{\rm n} = 5$ μm, $W_{\rm p} = 10$ μm, drawn channel length $L = 0.4$ μm, physical gate oxide thickness, T_{ox}, of 2.5–5.9 nm and different V_{th} resulted from different T_{ox} (universal V_{th} ion implant was used for all T_{ox} wafer splits), the load capacitance C_{L0} has

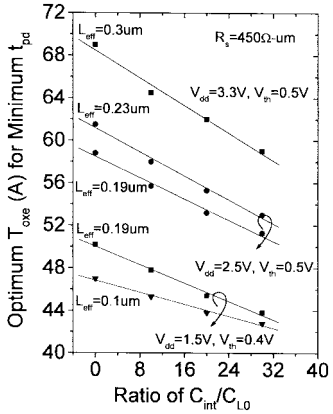

Fig. 12. Optimal T_{oxe} as a function of different C_{int}, V_{dd}, and L_{eff}.

been measured as:

$$C_{L0}\ (\text{fF}) = C_1 + \frac{A}{T_{oxe}} = 18.3 + \frac{283}{T_{oxe}}\ (\text{nm}). \quad (15)$$

For the loaded ring oscillators, the load capacitance for the loaded ring oscillator, C_L, can be expressed as as follows:

$$C_L\ (\text{fF}) = C_{L0} + C_{int}. \quad (16)$$

When T_{ox} is varied, the percentage of gate oxide thickness related capacitance, C_{ox}, over the whole load capacitance, C_L, is varied. For different interconnect loading, the optimal gate oxide thickness is different. For heavier interconnect loading, thinner gate oxide thickness should be chosen. This can be more effectively illustrated by Fig. 12. The window for optimal T_{ox} is also narrower for heavier interconnect loading case [15]. For higher power supply voltage, optimal T_{ox} changes faster (or is more sensitive) with L_{eff} than the lower V_{dd} case. For further discussion on the impact of interconnect loading effect on CMOS performance, please refer to [15].

IV. Conclusions

I_{dsat} and CMOS ring oscillator gate delay can be accurately modeled and predicted. This experimentally confirmed accurate model is a new tool to optimize I_{dsat}, CMOS gate speed, and to predict their dependencies on T_{ox}, V_{th}, V_{dd}, L, and capacitance loading.

Appendix

It has been found that electrical *C–V* measurement of T_{ox} in the accumulation regime (i.e., $V_{gs} = -3$ V, to avoid poly-gate depletion) is consistently larger than the thickness measured with the optical method. Such discrepancy between the electrical and optical thicknesses resulted from the finite thickness of the accumulation layer which is approximately the same as the thickness of the inversion layer. Fowler–Nordheim tunneling current method originally presented in [16] was employed in this study to characterize T_{ox} and the polysilicon gate depletion. The results are listed in Table I. The polysilicon depletion layer thickness and voltage may be modeled with a polysilicon doping concentration of 5×10^{19} cm^{-3}. It was concluded that the optical and tunneling-current methods give consistent physical thickness. The electrical oxide thickness, T_{oxe}, is always larger than the physical thickness by 0.4 nm at very large E_{ox} (thinner T_{ox}) to 0.7 nm at moderate E_{ox} (thicker T_{ox}). To discern the difference between the two T_{ox} thicknesses, T_{ox} and T_{oxe} are used to denote the physical and electrical oxide thickness, respectively. Thus, the electrical thickness should be used in all equations of I_{dsat}, C_L, and t_{pd}, in order to match the measured I_{dsat} as shown in Fig. 3. Furthermore, V_{gs} in (7)–(10) needs to be reduced by the polysilicon depletion voltage:

$$V_{polydep} = \frac{\varepsilon_{ox}^2 E_{ox}^2}{2q\varepsilon_{\text{Si}} N_{poly}} \quad (17)$$

where E_{ox}, $q = 1.6 \times 10^{-19}$ C, and N_{poly} are vertical electric field in the MOSFET gate oxide, Coulomb charge of electron, and effective doping concentration of polysilicon gate, respectively. The effective gate bias, V_{gseff}, becomes [17]:

$$V_{gseff} = V_{fb} + 2\phi_b + \frac{q\varepsilon_s N_{poly} T_{ox}^2}{\varepsilon_{ox}^2} \cdot \left[\sqrt{1 + \frac{2\varepsilon_{ox}^2 (V_{gs} - V_{fb} - 2\phi_b)}{q\varepsilon s_s N_{poly} T_{ox}^2}}\right] \quad (18)$$

where V_{fb} and ϕ_b are flat-band voltage and substrate body Fermi potential, respectively. This effectively increase the gate oxide thickness due to the polysilicon depletion effect, $T_{polydep}$. Alternatively, T_{oxe} used in previous equations is further increased by the polysilicon depletion layer thickness divided by 3,

$$T_{oxe} = T_{ox} + \frac{\text{inversion layer centroid}}{3} + \frac{T_{polydep}}{3}. \quad (19)$$

The factor "1/3" in the second and third term on the right hand side comes from the conversion from silicon thickness to silicon dioxide thickness because the ratio of their dielectric constant is 1/3.

TABLE I
Gate Oxide Thickness, T_{ox}, Characterized by Three Different Methods

	Wafer #1	Wafer #2	Wafer #3	Wafer #4
Process target (Å)	25	35	45	60
Optical (Å)	25	36	47	58
Physical (tunneling current) (Å)	25.8	36.5	45.5	57
Electrical (C-V) (Å)	29	42	52	65

Acknowledgment

The authors thank Prof. P. K. Ko of Hong Kong University of Science and Technology, Dr. C. H. Wann of IBM, Dr. J. Duster and Dr. D. Sinitsky of UC Berkeley, Dr. D. Pramanik and Dr. S. Nariani of VLSI, T. Tanaka of Fujitsu, and M. Yoshida of Hitachi for their helpful discussions and help in general.

References

[1] Y. Cheng, M.-C. Jeng, Z. Liu, M. Chan, K. Hui, J. Huang, K. Chen, P. K. Ko, and C. Hu, "A physical and scalable BSIM *I–V* model for analog/digital circuit simulation," *IEEE Trans. Electron Devices*, vol. 44, pp. 277–287, Feb. 1997.

[2] K. Chen, H. C. Wann, J. Duster, P. K. Ko, and C. Hu, "MOSFET carrier mobility model based on gate oxide thickness, threshold and gate voltages," *J. Solid-State Electron.*, vol. 39, no. 10, pp. 1515–1518, Oct. 1996.
[3] A. G. Sabnis and J. T. Clemens, "Characterization of electron velocity in the inverted ⟨100⟩ Si surface," in *IEDM Tech. Dig.*, 1979, pp. 18–21.
[4] M. S. Liang, J. Y. Choi, P. K. Ko, and C. Hu, "Inversion-layer capacitance and mobility of very thin gate-oxide MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-33, p. 409, 1986.
[5] K. Chen, H. C. Wann, J. Duster, D. Pramanik, S. Nariani, P. K. Ko, and C. Hu, "An accurate semi-empirical saturation drain current model for LDD N-MOSFET," *IEEE Electron Device Lett.*, vol. 17, pp. 145–147, Mar. 1996.
[6] P. K. Ko, "Approaches to scaling," in *VLSI Electronics: Microstructure Science.* New York: Academic, 1989, vol. 18, pp. 1–37.
[7] K. Chen, C. Hu, P. Fang, and A. Gupta, "Experimental confirmation of an accurate CMOS gate delay model for gate oxide and voltage scaling," *IEEE Electron Device Lett.* vol. 18, pp. 275–277, June 1997.
[8] A. Gupta, P. Fang, M. Song, K. Chen, and C. Hu, "Characterization of physical ultra-thin gate oxide thickness and effective polysilicon doping of CMOS devices," *IEEE Electron Device Lett.*, to be published.
[9] R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed. New York: Wiley, 1986, ch. 10.
[10] C. Hu, "Future CMOS scaling and reliability," *Proc. IEEE*, vol. 81, no. 5, May 1993, invited paper.
[11] ——, "Device and technology impact on low power electronics," *Low Power Design Methodologies*, J. M. Rabaey and M. Pedram, Eds. Boston, MA: Kluwer, 1995, ch. 2.
[12] K. Chen, C. Hu, P. Fang, M. R. Lin, and D. L. Wollesen, "Experimental characterization of load capacitance of CMOS gate," submitted for publication.
[13] T. Kuroi *et al.*, *1996 Symp. VLSI Technology*, Japan, pp. 210–211.
[14] M. T. Bohr, "Interconnect scaling—The real limiter to high performance ULSI," *IEDM*, Section 10.1, pp. 241–244, 1995.
[15] K. Chen, C. Hu, P. Fang, M. R. Lin, and D. L. Wollesen, "Optimizing quarter and sub-quarter micron CMOS circuit speed considering interconnect loading effects," *IEEE Trans. Electron Devices*, vol. 44, pp. 1556–1558, Sept. 1997.
[16] K. F. Schuegraf, C. C. King, and C. Hu, "Ultra-thin silicon dioxide leakage current and scaling limit," in *Symp. VLSI Technology, Dig. Tech. Papers*, 1992, Section 3-1, pp. 18–19.
[17] K. Chen, M. Chan, P. K. Ko, C. Hu, and J. H. Huang, "Polysilicon gate depletion effect on IC performance," *J. Solid-State Electron.*, vol. 38, no. 11, pp. 1975–1977, Nov. 1995.

Kai Chen (S'88–M'89) was born in Beijing, China. He received the B. Eng. (electrical engineering) degree with honors, the M.A. degree in physics, and the M.S.E.E. and Ph.D. degrees in electrical engineering from Tsinghua University, China, Wayne State University, Detroit, MI, Purdue University, West Lafayette, IN, and the University of California, Berkeley, in 1985, 1988, 1989, and 1997, respectively.

He is now with Advanced Logig and SRAM Devices, Semiconductor Research and Development Center (SRDC) of IBM, Hopewell Junction, NY, responsible for CMOS device design and compact modeling of IBM's new CMOS technology under development. From 1989 to 1992, he was a senior engineer with Fairchild Research Center of National Semiconductor Corporation, Santa Clara, CA, worked on process integration and yield enhancement of sub-micron CMOS technology development. He holds one U.S. patent and has co-authored 40 research papers published on international journals and conferences. He was one of the major contributors of the widely regarded "industry standard" compact model for circuit design simulation-BSIM3v3, which also received the "R&D 100 Award" in 1996.

Dr. Chen has been a senior member of Chinese Institute of Electronics (CIE) since 1995. He is a founding and board member of North American Chinese Semiconductor Association (NACSA) since 1996. He served as President of Berkeley Chinese Student and Scholars Association (BCSSA) for the 1995–1996 term.

Chenming Hu (S–71–M'76–SM'83–F'90) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, R.O.C., and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley in 1968, 1970, and 1973 respectively.

From 1973 to 1976 he was an assistant professor at Massachusetts Institute of Technology, Cambridge. Since 1976, he has been a professor of Electrical Engineering and Computer Sciences at UC Berkeley. While on leave in 1980–1981, he was manager of nonvolatile memory development at National Semiconductor. He has served as an advisor to many industry, government, and educational institutions. His present research areas include VLSI devices, silicon-on-insulator devices, hot electron effects, thin dielectrics, electromigration, circuit reliability simulation, and nonvolatile semiconductor memories. He has been awarded several patents on semiconductor devices and technology. He has authored or co-authored four books and over 500 research papers. He has delivered dozens of keynote addresses and invited papers at scientific conferences, and has received several best-paper awards.

Dr. Hu was elected a member of the National Academy of Engineering and received the Berkeley Distinguished Teaching Award in 1997. He is an Honorary Professor of Beijing University, China, and of the Chinese Academy of Science. He received the 1991 Grand Prize of Excellence in Design Award from Design News Award and the first Semiconductor Research Corporation Technical Excellence Award in 1991 for leading the development of IC simulation in 1995 and given an R&D 100 Award as one of the 100 most technologically significant new products of the year in 1996. The Board of Directors of the IEEE awarded him the 1997 Jack A. Morton Award for contributions to MOSFET reliability physics and modeling. He was guest editor of PROCEEDINGS OF THE IEEE and of the IEEE TRANSACTIONS ON ELECTRON DEVICES. He serves on editorial boards of *Semiconductor Science and Technology*, and *Journal of Microelectronics and Reliability*. He was Board Chairman of East San Francisco Bay Chinese School from 1988 to 1991.

Peng Fang (S'87–M'91) received the Ph.D degree in electrical engineering from the University of Minnesota, Minneapolis, in 1988.

From 1988 to 1990, he was a Research Associate at the Electronics Research Lab of the University of California, Berkeley and the University of South Florida, Tampa. He was a faculty member with University of Electronics Science and Technology of China from 1982 to 1985. His research topics were on IC devices and interconnects reliability. He joined Advanced Micro Devices (AMD), Sunnyvale, CA, in 1990. Since then he was with technology reliability and strategic technology reliability development department of technology development group of AMD. In charge of new technology qualification and reliability issues including hot carrier, electromigration and thin oxide.

Dr. Fang has been a committee member of IEEE international conferences such as IRPS and CICC since 1992. He is the Chairman of IEEE Electronic Device Society Silicon Valley Chapter for 1997–1998

Min Ren Lin, photograph and biography not available at the time of publication.

Donald L. Wollesen (A'93), photograph and biography not available at the time of publication.