Device and Technology Optimizations for Low Power Design in Deep Sub-micron Regime

Kai Chen and Chenming Hu

211-72 Cory Hall, E.E.C.S. Department, University of California at Berkeley, Berkeley, CA 94720-1772
E-mail: \{kai, hu\}@eecs.berkeley.edu

INTRODUCTION
Numerous studies on technology’s impact or optimization for low power design have been carried out [1]-[5]. Most analysis or trend projections [6] are often made by using a simple long-channel MOSFET I-V model because the accurate analytical device model for deep sub-micron MOSFET’s has been lacking. While the long channel MOSFET model is convenient to use, it may give false results for today’s deep sub-micron ICs. For example, Fig. 1 shows the CMOS performance predicted by the long-channel versus accurate deep sub-micron MOSFET models [7][8] for gate oxide and power supply scaling. Fig. 1 a) illustrates that excessive $T_{ox}$ scaling in deep sub-micron regime can degrade circuit performance, opposite to what the basic long-channel device model predicts. Fig. 1 b) shows that the $V_{dd}$ reduction is not as deleterious to speed as previously thought. It is therefore useful to update the theories that guide device and technology scaling.

This report, based on the most recent analytical and experimental studies of CMOS scaling and gate delay models [7]-[10], reexamines the fundamental design quantities such as driving current, $I_{d_{sat}}$, propagation delay, $t_{pd}$, and switching energy, $E$, and investigates device optimization issues in deep sub-micron regime. Empirical $I_{d_{sat}}$ equations and device optimization guidelines with gate oxide, channel length and power supply scaling as well as interconnect loading are extracted.

DESIGN GUIDELINES AND DISCUSSIONS
Recently, accurate analytical model [9][10] for deep sub-micron MOSFETs has been developed.

Fig. 2 shows that it can accurately predict the device driving current for wide ranges of gate oxide thickness and supply voltage as long as gate oxide thickness, threshold voltages, effective channel lengths and parasitic source and drain series resistances are known or correctly characterized. Based on this model, empirical $I_{d_{sat}}$ equations in terms of power supply and threshold voltage, channel length and gate oxide scaling can

![Fig. 1 a) Optimum $T_{ox}$ exists, opposite to that predicted by the basic long channel model.](image)

![Fig. 1 b) Reducing power supply voltage has less impact on $t_{pd}$ than that predicted by the outdated long channel $I_{d_{sat}}$ model.](image)
Fig. 2 a) Accurate analytical $I_{\text{dstat}}$ model fits the measurement data well for wide ranges of $L_{\text{eff}}$ and $V_{dd}$.

Fig. 2 b) Accurate analytical $I_{\text{dstat}}$ model fits the measurement data well for wide ranges of $T_{\text{oxe}}$ and $V_{dd}$.

Fig. 3 Accurate analytical model predicts that $I_{\text{dstat}} \sim T_{\text{oxe}}^{-0.8}$ due to mobility degradation.

Fig. 4 a) Measurement data of PMOS and NMOS $I_{\text{dstat}}$ with wide range of $T_{\text{oxe}}$ confirms that the current density $J_{\text{dstat}}/J_{\text{dstat}} \sim 2.2$.

be developed as follows:

$$I_{\text{dstat}} \sim L_{\text{eff}}^{-0.5} T_{\text{oxe}}^{-0.8} (V_{dd} - V_t)_{1.25}$$  \hspace{1cm} (1a)

where $T_{\text{oxe}}$ is the electrical measured gate oxide thickness. It is usually 0.4~6.0nm thicker than the physical $T_{\text{oxe}}$ due to the finite thickness of inversion layer. Fig. 3 shows an example of how the empirical relationship between $I_{\text{dstat}}$ and $T_{\text{oxe}}$ for deep submicron MOSFETs is determined. The $I_{\text{dstat}}$ and $T_{\text{oxe}}$ dependence of $I_{\text{dstat}}$ expressed in equation (1a) is difference from the well known long channel model:

$$I_{\text{dstat}} \sim L_{\text{eff}}^{-1} T_{\text{oxe}}^{-1} (V_{dd} - V_t)^2$$  \hspace{1cm} (1b)

due to velocity saturation and mobility degradation, respectively. The difference of $(V_{dd} - V_t)$ dependence between (1a) and (1b) is due to the effects of both.

As for PMOSFETs, Fig. 4 a) shows the measurement data of PMOS and NMOS $I_{\text{dstat}}$ for wide ranges of $T_{\text{oxe}}$ and $V_{dd}$. The ratio of drain current per unit width, $J_{\text{dstat}}/J_{\text{dstat}}$, remains unchanged at approximately 2.2. This ratio will remain a constant because the effective vertical field in the channel will remain at approximately 0.8MV/cm in future devices, making $\mu_d/\mu_p$ a constant as shown in Fig. 4 b).

It is easy to show that for the deep sub-micron MOSFETs,

$$t_{\text{pd}}(T_{\text{oxe}}) = \left[ C' + \alpha(W_n + W_p)L_{\text{oxe}}/T_{\text{oxe}} \right] I_{\text{dstat}}$$  \hspace{1cm} (2)
where \( C^* \) represent all the overlap/parasitic/depletion and interconnect capacitance, respectively. \( \varepsilon_{ox} \) is the dielectric coefficient of SiO\(_2\), \( \alpha \approx 1.5 \) is a constant larger than unity due to overlap capacitance and Miller effect. The optimum gate oxide \( T_{ox}^{opt} \) can be found from (2) as:

\[
T_{ox}^{opt} = \alpha (W_n + W_p) L e_{ox}/4(C_1 + C_{int})
\]

(3)

or

\[
C_{ox}^{opt} = 4C^*/\alpha
\]

(4)

![Graph showing effective electric field as a function of voltage and oxide thickness](image)

**Fig. 4 b)** The ratio of mobilities of electrons and holes will not change significantly with scaling as the \( E_{eff} \) will remain \( \sim 0.85 \text{MV/cm} \).

![Graph showing carrier mobility and effective electric field](image)

**Fig. 5** Analytical gate delay model can accurately predict \( t_{pd} \) for the wide range of \( V_{dd} \) and \( T_{ox} \).

**Fig. 6 a)** \( t_{pd} \) is insensitive to \( V_{dd} \) if \( V_{dd} > 5V_t \) or if \( V_t \) scales with \( V_{dd} \). Energy\(^*\)\( t_{pd} \) is sensitive to \( V_{dd} \).

![Graph showing \( V_{dd}/V_t \) versus \( t_{pd} \) and energy\(^*\)\( t_{pd} \)](image)

**Fig. 6 b)** Both \( t_{pd} \) and Energy\(^*\)\( t_{pd} \) remains relatively insensitive to \( V_{dd} \) scaling when \( V_t \) is scaled with \( V_{dd} \).

for given \( V_t \) and \( V_{dd} \). It can be shown that for given total area, i.e. given \( W_n + W_p \), both \( t_{pd} \) and Energy\(^*\)\( t_{pd} \) are minimized when \( W_n/W_p = (2.2)^{1/2} \sim 1.5 \) [4]. The minima are broad, however, and \( W_n/W_p = 2 \) is almost as good for \( t_{pd} \) and Energy\(^*\)\( t_{pd} \).

Fig. 5 shows that the analytical model can accurately predict CMOS ring oscillator \( t_{pd} \) for wide ranges of \( V_{dd} \) and \( T_{ox} \). This model also predicts that if \( V_t \) is scaled together with \( V_{dd} \), \( t_{pd} \) is not sensitive to \( V_t \) scaling and not very sensitive to \( V_{dd} \) as long as \( V_{dd} > 5V_t \), as shown in Fig. 6 a). However, the energy and \( t_{pd} \) product, Energy\(^*\)\( t_{pd} \), is a strong function of \( V_{dd} \). Fig. 6 b) shows that \( t_{pd} \) and Energy\(^*\)\( t_{pd} \), are independent of \( V_{dd} \) scaling.
if $V_{dd}/V_t$ is fixed. Fig. 7 and 8 show the sensitivities of CMOS gate delay to $V_{dd}$ and $V_t$. Fig. 5 through 8 offer design guidelines to trading speed for energy consumption when $V_t$ or $V_{dd}$ is used as the leverage.

For low power design, interconnect loading effects must be considered. Fig. 9 shows how CMOS speed is affected by $T_{ox}$ scaling for different capacitive loads. It is found that for deep sub-micron CMOS circuits, an optimum $T_{ox}$ exists, as experimentally indicated by the crossovers in Fig. 5 and the work reported in reference [11]. This is because that for increasing $T_{ox}$, junction/depletion and interconnect capacitances become more dominant while $I_{Sat}$ drops. This results in a decreased performance.

For decreasing $T_{ox}$, the gate capacitance tends to dominate while $I_{Sat}$ does not increase proportionally to $1/T_{ox}$ due to mobility degradation with higher vertical electrical field. Thus performance tends to decrease. Due to these two mechanisms, there exists an optimum $T_{ox}$ for best performance. Larger interconnect loading results in thinner optimum $T_{ox}$, as shown in Fig. 9. The process window of $T_{ox}$ Opt is also narrower for heavier loaded circuits.

**CONCLUSION**

Device and technology optimization guidelines for low power design in deep sub-micron regime based on original analytical and experimental studies have been presented. Difference between long channel and deep sub-micron devices is emphasized. Optimum $T_{ox}$ was demonstrated by experiment and discussed in theory.

**ACKNOWLEDGEMENT**

This work is supported by SRC grant 96-SJ-417, IDT, Rockwell International, and HP under MICRO program. Authors also want to thank AMD and VLSI Technology for the use of their experimental data.
REFERENCES:
10. K. Chen, H.C. Wann, J. Duster, M. Yoshida, P. Ko and C. Hu, “MOSFET Carrier Mobility Model Based on Gate Oxide Thickness, Threshold and Gate Voltages”, Journal of