

# Fully Depleted CMOS/SOI Device Design Guidelines for Low Power Applications

Srinivasa R. Banna, Philip C. H. Chan, Mansun Chan, Samuel K. H. Fung  
and Ping K. Ko

Department of Electrical and Electronic Engineering  
Hong Kong University of Science and Technology  
Clear Water Bay, Kowloon, Hong Kong

## ABSTRACT

In this paper we report the fully depleted CMOS/SOI device design guidelines for low power applications. Optimal technology, device and circuit parameters are discussed and compared with bulk CMOS based design. The differences and similarities are summarized. We believe this is the first such study to be reported.

## 1. INTRODUCTION

In CMOS ULSI design, the interplay between device technology and low-power applications using FDSOI technology needs careful study. A detailed device design and its impact on low power in bulk technology were considered in [1]. Device design guidelines drawn from bulk technology may not necessarily be optimized for low-power design on FDSOI technology. In this paper, device design guidelines for FDSOI technology are reported and compared with bulk technology guidelines for low power applications.

## 2. DYNAMIC POWER IN FDSOI CMOS

Consider a simple CMOS inverter circuit driving another load inverter as shown in Figure 1. Ignoring the short circuit power, dynamic power consumed in the inverter can be written as:

$$P_{\text{dynamic}} = C_L V_{\text{dd}}^2 \alpha f \quad (1)$$

where  $C_L$  is the load capacitance at the output of the inverter,  $V_{\text{dd}}$  is the supply voltage,  $\alpha$  is the activity factor and  $f$  is the clock frequency.  $C_L$  can also be written as

$$C_L = C_{\text{gate}} + C_{\text{junc}} + C_{\text{wire}} \quad (2)$$

where  $C_{\text{gate}}$ ,  $C_{\text{junc}}$  and  $C_{\text{wire}}$  are respectively, the gate capacitance of the load inverter, source/drain junction capacitance and wire capacitance at the driver inverter output node. In FDSOI technology, gate capacitance ( $C_{\text{gate}}$ ) is less than in bulk technology as shown in Figure 2. This reduction in  $C_{\text{gate}}$  is due to negligible gate-to-body capacitance, which is  $C_{\text{box}}$  in SOI technology compared with  $C_{\text{ox}}$  in bulk technology. An average gate capacitance can be obtained by integrating Figure 2 as shown below

$$\frac{1}{V_{\text{dd}}} \int C dV = C_{\text{gate}} \quad (3)$$

$$C_{\text{gate}} = (C_{\text{foxn}} + C_{\text{foxp}}) \left(1 - \frac{V_{\text{th}}}{V_{\text{dd}}}\right) + (C_{\text{boxn}} + C_{\text{boxp}}) \frac{V_{\text{th}}}{V_{\text{dd}}} \quad (4)$$

where  $C_{\text{foxn}}/C_{\text{boxn}}$ ,  $C_{\text{foxp}}/C_{\text{boxp}}$  are front/back gate capacitance of load inverter n and p-channel devices.  $C_{\text{fox}}$  and  $C_{\text{box}}$  are defined as  $\frac{WL\epsilon_{\text{ox}}}{T_{\text{fox}}}$  and  $\frac{WL\epsilon_{\text{ox}}}{T_{\text{box}}}$ . Junction capacitance,

$C_{\text{junc}}$ , in FDSOI is also reduced. It is a function of buried oxide thickness ( $T_{\text{box}}$ ) and it can be expressed as  $C_{\text{junc}} = \frac{L_{\text{sd}}W\epsilon_{\text{ox}}}{T_{\text{box}}}$  where  $L_{\text{sd}}$  is the length of the source/drain

region. However, in bulk technology,  $C_{\text{junc}}$  depends on the source/drain junction depletion widths that are a function of substrate/well doping. It can be expressed as  $C_{\text{junc}} =$

$$\frac{L_{\text{sd}}W\epsilon_{\text{ox}}}{X_{\text{d}}} \quad \text{where } X_{\text{d}} = \sqrt{\frac{2\epsilon_{\text{ox}}(V_{\text{bi}} - V_{\text{dd}})}{qN_{\text{sub}}}}, \quad V_{\text{bi}} \text{ is the}$$

junction built-in potential,  $\epsilon_{\text{si}}$  and  $\epsilon_{\text{ox}}$  are the silicon and silicon dioxide permittivity.  $C_{\text{junc}}$  in bulk technology is at least an order higher than  $C_{\text{junc}}$  in SOI technology. A constant  $C_{\text{wire}}$  is assumed in this study.

### 3. MODELING INVERTER DELAY

Scaling down  $V_{dd}$  reduces power dissipation. In fact, it is the most effective way of reducing CMOS power. However the speed of the inverter also slows with reducing  $V_{dd}$ . Hence optimizing the technology for lower  $V_{dd}$  could minimize the speed loss due to  $V_{dd}$  reduction. An accurate expression for inverter delay  $\tau$  is given by

$$\tau = \frac{C_L V_{dd}}{4I_{dsatn}} \left( 1 + 2.2 \frac{W_n}{W_p} \right) \quad (5)$$

where  $I_{dsatn}$  is the n-channel saturation drain current,  $W_n$  and  $W_p$  are the widths of the n and p-channel devices [1]. In (5)  $\tau$  is interpreted as the average of the time for the n-channel device to discharge  $C_L$  from  $V_{dd}$  to  $V_{dd}/2$  and the time for the p-channel device to charge  $C_L$  from zero to  $V_{dd}/2$ . It is also assumed that  $I_{dsatn} \approx 2.2 I_{dsatp}$  if  $W_n=W_p$  to account for the difference in electron and hole mobilities.

$I_{dsat}$  dependence on channel length ( $L$ ), mobility  $\mu_{eff}$ , velocity saturation  $v_{sat}$  and effective front gate bias ( $V_{gsf} - V_{th}$ ) can be simplified by neglecting  $R_{ds}$  effect as

$$I_{dsat} = C_{fox} W v_{sat} (V_{gsf} - V_{th} - a V_{dsat}) \quad (6a)$$

$$V_{dsat} = \frac{(V_{gsf} - V_{th}) E_{sat} L_{eff}}{V_{gsf} - V_{th} + a E_{sat} L_{eff}} \quad (6b)$$

$$E_{sat} = \frac{2v_{sat}}{\mu_{eff}}. \quad (6c)$$

Equation (6) relates drain saturation current to device parameters such as surface mobility and applied front gate bias ( $V_{gsf}$ ). Surface mobility is a function of front gate oxide thickness ( $T_{fox}$ ), effective gate bias ( $V_{gsf} - V_{th}$ ) and silicon film thickness ( $T_{si}$ ). However in FDSOI technology,  $I_{dsat}$  is a weak function of  $T_{box}$  and silicon film thickness ( $T_{si}$ ) [2].  $I_{dsat}$  dependence on  $T_{si}$  comes from threshold voltage ( $V_{th}$ ) and surface mobility. For a fixed  $V_{th}$ , its dependence on  $T_{si}$  comes from surface mobility only. Furthermore, in the deep submicron regime, velocity saturation is dominant and the surface mobility dependence on  $T_{si}$  can be ignored. Hence surface mobility dependence on ( $V_{gsf} - V_{th}$ ) and  $T_{fox}$  can be written as

$$\mu_{eff} = \frac{\mu_0}{1 + U_0 \left( \frac{V_{gsf} - V_{th}}{T_{fox}} \right) + U_1 \left( \frac{V_{gsf} - V_{th}}{T_{fox}} \right)^2}. \quad (7)$$

Substituting (7) in (6) one can obtain  $I_{dsat}$  as a function of  $T_{fox}$ ,  $L$  and  $V_{gsf} - V_{th}$ . An approximate dependence of  $I_{dsat}$

on  $T_{fox}$ ,  $L$  and  $V_{gsf} - V_{th}$  is obtained by fitting Equation (8) to experimental data. Experimental data are obtained from devices fabricated on SIMOX and BESOI substrates. Devices with different channel lengths and front/back gate thicknesses biased at different front gate voltages were included.

$$I_{dsat} \propto L^{-0.57} T_{fox}^{-1.23} (V_{gsf} - V_{th})^{1.6} \quad (8)$$

Equation (8) fits well with experimental data obtained from NMOSFET as shown in Figure 3 and Figure 4 for different effective gate biases, channel lengths and front oxide thicknesses. One can also include  $T_{si}$  dependence on  $I_{dsat}$  in the curve fitting, which is given by

$$I_{dsat} \propto L^{-0.57} T_{fox}^{-1.34} T_{si}^{0.13} (V_{gsf} - V_{th})^{1.6} \quad (9)$$

The weak dependence of  $I_{dsat}$  on  $T_{si}$  is re-confirmed in (9) as explained earlier. Hence  $I_{dsat}$  dependence on  $T_{si}$  is ignored. Bulk NMOSFET data suggested that  $I_{dsat}$  is a weak function of  $L$  and  $T_{ox}$  as presented by [1] and re-stated for convenience here:

$$I_{dsat} \propto L^{-0.5} T_{ox}^{-0.5} (V_{gs} - V_{th})^{1.3} \quad (10)$$

By comparing (8) and (10), one can state that  $I_{dsat}$  in SOI is a stronger function of front oxide thickness ( $T_{fox}$ ) and ( $V_{gsf} - V_{th}$ ) than  $I_{dsat}$  in bulk. Substituting (8) in (5) with  $V_{gsf} \approx 0.9V_{dd}$ , one obtains

$$\tau \propto \frac{C_L L^{0.57} T_{fox}^{1.23}}{(0.9 - V_{th} / V_{dd})^{1.6} V_{dd}^{0.6}} \left( \frac{1}{W_n} + \frac{2.2}{W_p} \right) \quad (11)$$

Switching energy in the CMOS inverter is a function of load capacitance and the supply voltage, which is given by  $E \propto C_L V_{dd}^2$ . Hence the  $E\tau$  product is given by

$$E\tau \propto \frac{C_L^2 V_{dd}^{1.4} L^{0.57} T_{fox}^{1.23}}{(0.9 - V_{th} / V_{dd})^{1.6}} \left( \frac{1}{W_n} + \frac{2.2}{W_p} \right) \quad (12)$$

Compared to bulk technology, our data suggests that inverter delay dependence on  $T_{fox}$  and  $V_{dd}$  is increased. Equation (11) was also verified this by device simulation. An FDSOI CMOS inverter driving another load inverter as shown in Figure 1 was simulated in a MEDICI 2-D device simulator with a circuit-analysis advanced- application module for different stage ratio ( $a$ ) and wire capacitance ( $C_{wire}$ ). PMOS and NMOS devices in the simulated driver and load inverters are  $L=0.1\mu m$ ,  $T_{fox}=5nm$ ,  $T_{si}=50nm$  and  $T_{box}=400nm$ . Figure 5 shows the comparison between Equation (11) and device simulation results. Good agreement between Equation (11) and simulation data were obtained for different threshold voltages, stage ratios and wire capacitances.

#### 4. OPTIMAL $V_{DD}/V_{TH}$ RATIO

To minimize  $P_{dynamic}$  without suffering excessive speed degradation, the normalized delay and energy-delay products obtained from Equations (11) and (12) are plotted for varying  $V_{dd}/V_{th}$  ratios in Figure 6. It appears that a  $V_{dd}$  between 3-4 times the  $V_{th}$  is the optimal without suffering large degradation in speed. The delay is independent of  $V_{dd}/V_{th}$  for  $V_{dd} > 3V_{th}$ . This is less than bulk CMOS technology [1]. Similarly, the energy-delay product is minimal at  $V_{dd}=1.5V_{th}$  instead of  $2V_{th}$  as in the case of bulk CMOS technology. Due to sharp turn-on characteristics of FDSOI devices,  $V_{th}$  scaling versus device leakage current trade-off is improved in FDSOI technology.

#### 5. TRANSISTOR SIZING

Sizing the transistor for optimal delay and energy-delay product is essentially the same as for bulk CMOS technology. This is due to the same functional dependence of the delay and the energy-delay product on  $W_n$  and  $W_p$ . The optimal ratio is independent of load capacitance and it is given by

$$\frac{W_p}{W_n} = \sqrt{2.2} = 1.5 .$$

Also the optimal  $W_n+W_p$  is obtained by plotting delay and energy-delay product versus the driver device gate capacitance as a fraction of total load capacitance  $C_L$ . Figure 7 shows that the delay reduces with increasing  $W_n+W_p$  and there is no minimum. However, the energy-delay product initially improves with reducing the delay and degrades thereafter due to an increase in the load capacitance. There is a broad minimum for the energy-delay product that occurs when the driving device gate capacitance ( $C_{gate}$ ) is around one third of the load capacitance ( $C_L$ ) for different  $C_{wire}$  with a unity fanout. The minimum for energy-delay product shifts to a lower value of  $C_{gate}/C_L$  for a higher fanout.

#### 6. OPTIMAL BURIED OXIDE THICKNESS

In SOI technology, reduced junction capacitance is due to the presence of thick buried oxide. However, thick buried oxide formation is difficult. Moreover the self-heating and floating-body effects will increase with increasing buried oxide thickness. An optimal buried oxide thickness that minimizes delay and energy-delay product is important to know. Figure 8 plots normalized buried oxide thickness versus the delay and the energy-delay product. The buried oxide thickness chosen is normalized by 1mm. A buried oxide thicknesses between 300-400nm is a good choice.

The delay and energy-delay products are independent when  $T_{box}$  is  $> 400nm$ . This means a  $T_{box}$  thicker than 400nm would not improve speed or power consumption significantly.

#### 7. OPTIMAL GATE OXIDE THICKNESS

Dependence of the delay and the energy-delay products on the gate oxide thickness in FDSOI technology is increased due to increased mobility dependence on the gate oxide thickness. Figure 9 plots the delay and energy-delay products versus front gate oxide thickness ( $T_{fox}$ ) for different fanouts. No minimum exists for delay with reducing the front gate oxide thickness. However there exists an optimal gate oxide thickness at which energy-delay product is minimized. The optimal gate oxide thickness increases with increasing fanouts. For the unity fanout condition, the optimal front gate oxide thickness is between 7 to 8 nm for low power applications. Interestingly, the general trend in front gate oxide scaling [3] falls in the vicinity of the energy-delay product minimum. Hence, high-speed and low-power conditions can be met without significantly sacrificing either speed, power or short channel effects in deep submicron FDSOI MOSFET technology.

#### 8. OPTIMAL STAGE RATIO

The optimal way to drive a large capacitance is to use a minimally sized inverter to drive a larger inverter. The next step is to use larger inverter to drive a still larger inverter until at some point the larger inverter is able to drive the load capacitance directly. If  $N$  such stages are used, each larger than the previous by stage ratio 'a', then the total delay of the inverter chain is given by [4] :

$$\tau_{total} = \log\left(\frac{C_L}{C_{gate}}\right) \left[ \frac{a}{\log(a)} \right] \tau \quad (13)$$

where  $\tau$  is the single inverter delay and 'a' is the stage ratio. A similar approach is used to compute the energy dissipated in the inverter chain, which is given by

$$E_{total} = \frac{a(a^N - 1)}{a - 1} E \quad (14)$$

where  $N = \log\left(\frac{C_L}{C_{gate}}\right) / \log(a)$ .  $E$  is the single inverter energy dissipation. Figure 10 plots the total delay, the total energy and the total energy-delay products versus the stage ratio used in driving large capacitive loads. The energy is minimized at the stage ratios that are at an optimum and are

suitable for performance. Hence the optimal stage ratio is between 2 to 4 for low-power designs.

## 9. CONCLUSION

The device design guidelines obtained from the above discussions are summarized in Table 1. By comparing this system with bulk CMOS [1], it appears that bulk CMOS designs can all be transferred to FDSOI with minimal perturbation. This saves design effort and time in meeting the SOI low-power application requirements. However, the optimal front gate oxide thickness that minimizes the delay and energy-delay products in FDSOI is different in bulk technology. An optimal buried oxide thickness is about 400nm for low-power applications without significant speed degradation.

Device design guidelines using devices with  $L=0.1\mu\text{m}$  for FDSOI low-power applications are presented using a simple drain saturation current model fitted to experimental results and 2-D numerical simulations. The optimum occurs at  $V_{dd}=3V_{th}$  for performance and  $V_{dd}=1.5V_{th}$  for low power. The optimal buried oxide thickness is between 300nm to 400nm. The optimal gate oxide thickness for low power and good performance is between 3nm to 8nm. The optimal transistor sizing is when the driver device gate capacitance is 0.3 time of total load capacitance. The optimal stage ratio

for low power is the same optimal stage ratio for high performance.

## ACKNOWLEDGMENT

This work is supported by Research Grant Council Earmarked Research Grant HKUST 547/94E. The authors would like to acknowledge the use of MEDICI device simulator from Technology Modeling Associates, Inc.

## REFERENCES

- [1] Jan M.Rabaey and Massoud Pedram, Low power design methodologies, Boston/Dordrecht/London: Kluwer Academic Publishers, 1996
- [2] Srinivasa R.Banna, Philip Chan, Mansun Chan and Ping K.Ko, "A Physically based compact device model for Fully Depleted and nearly Fully Depleted SOI MOSFET," IEEE Trans. on Electron Devices, Vol 43, pp1914-1923, Nov 1996.
- [3] C.Hu, "Gate Oxide Scaling Limits and Projections," in the Technical Digest of International Electron Device Meetings, pp. 319-322, 1996.
- [4] Carver Mead and Lynn Conway, Introduction to VLSI system, Addison-Wesley Publishing Company, 1980.

Parameter	SOI $\tau$ (Bulk $\tau$ )	SOI E. $\tau$ (Bulk E $\tau$ )
<b>L</b>	<b>min (min)</b>	<b>min (min)</b>
<b>V<sub>dd</sub></b>	<b>&gt;3V<sub>th</sub> (4V<sub>th</sub>)</b>	<b>&gt;1.5 (2V<sub>th</sub>)</b>
<b>T<sub>fox</sub></b>	<b>C<sub>tox</sub>=C/1.5 (C/2) (or) 7 - 8 nm</b>	<b>C<sub>tox</sub>=C/1.5 (C/4)</b>
<b>W<sub>p</sub>/W<sub>n</sub></b>	<b>1-3 (1-3)</b>	<b>1-3 (1-3)</b>
<b>W<sub>p</sub>+W<sub>n</sub></b>	<b>max (max)</b>	<b>Cd=C/3 (C/2)</b>
<b>T<sub>box</sub></b>	<b>300-400nm</b>	<b>300-400nm</b>
<b>Stage Ratio (a)</b>	<b>2-4 (2.8)</b>	<b>2-4</b>

Table 1. Comparison of Device design Guidelines for SOI and bulk technology. Bulk technology device design Guidelines are obtained from [1] and shown in parenthesis.

## LIST OF FIGURES

Figure 1 A simple two inverter circuit considered in the study. ‘a’ stands for stage ratio.

Figure 2 Comparison of gate capacitance in bulk and SOI technology.  $C_{dn}$  and  $C_{dp}$  are the depletion capacitance.

Figure 3 Comparison of measured n-channel FDSOI MOSFET drain saturation current and Eq. (8) for  $T_{fox}=7.5nm$ .

Figure 4. Comparison of measured n-channel FDSOI MOSFET drain saturation current and Eq(7.8) for  $T_{fox}=10.7nm$ .

Figure 5 The delay is compared between Eq. (8) and 2-D numeral simulation. The circuit configuration considered in a 2-D MEDICI simulator is the same as shown in Figure 1.

Figure 6 Effect of  $V_{dd}/V_{th}$  scaling on delay and energy-delay product in FDSOI CMOS inverter for unity stage ratio and  $C_{wire}=0$ .

Figure 7 Normalized delay and energy-delay products versus drive device gate capacitance as a fraction of total load capacitance for unity stage ratio,  $T_{fox}=5nm$ ,  $T_{box}=400nm$ ,  $V_{dd}=1.2V$ , and  $V_{th}=0.3V$ .

Figure 8 The delay and energy-delay products versus normalized buried oxide thickness are plotted. The buried oxide thickness chosen is normalized with  $1mm$ .

Figure 9 Normalized delay and energy-delay products are plotted for various gate oxide thicknesses.

Figure 10 The delay and energy-delay products versus stage ratio are plotted. Low power and optimal speed can be obtained at a stage ratio equal to 2.8.

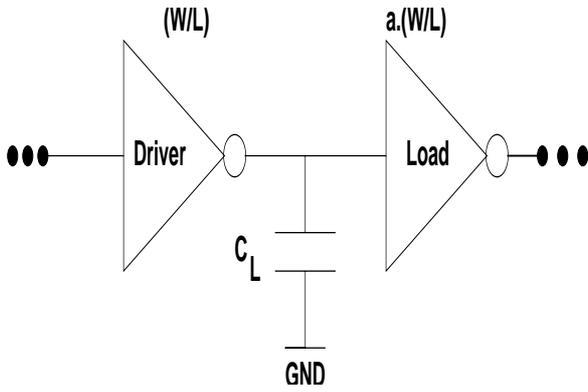


Figure 1 A simple two inverter circuit considered in the study. 'a' stands for stage ratio.

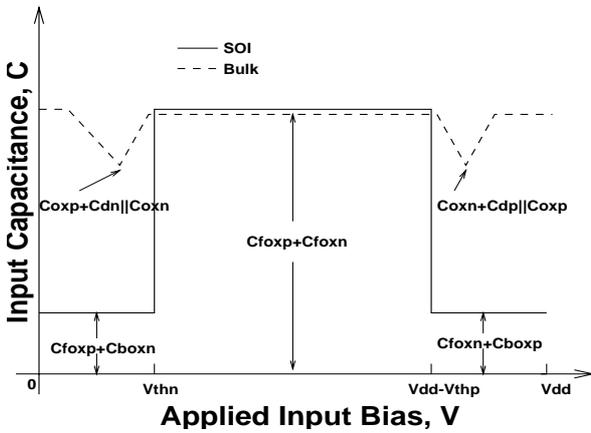


Figure 2 Comparison of gate capacitance in bulk and SOI technology.  $C_{dn}$  and  $C_{dp}$  are the depletion capacitance.

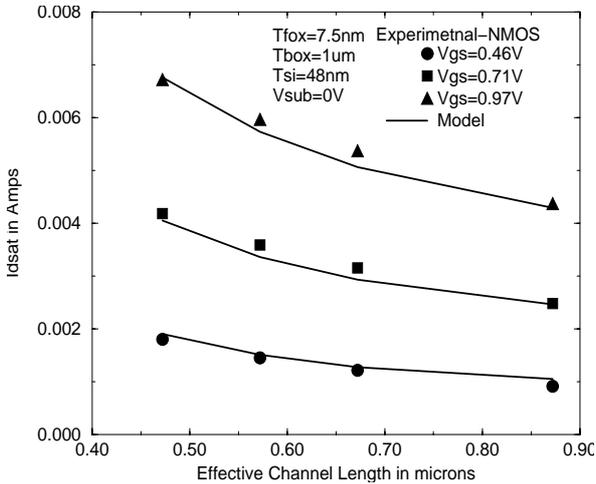


Figure 3 Comparison of measured n-channel FDSOI MOSFET drain saturation current and Eq. (8) for  $T_{fox}=7.5nm$ .

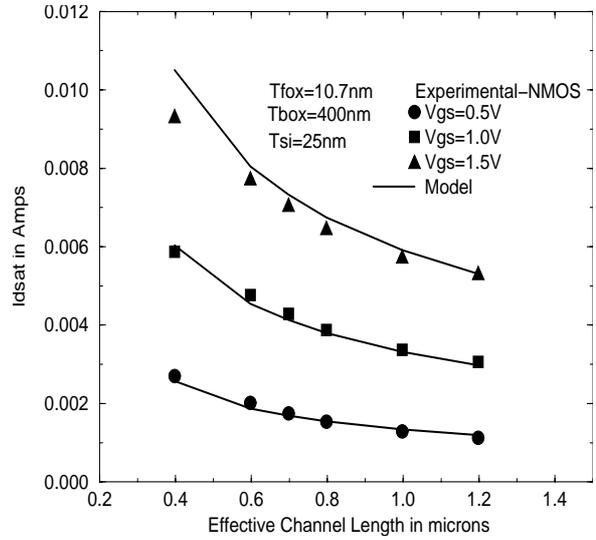


Figure 4 Comparison of measured n-channel FDSOI MOSFET drain saturation current and Eq(7.8) for  $T_{fox}=10.7nm$ .

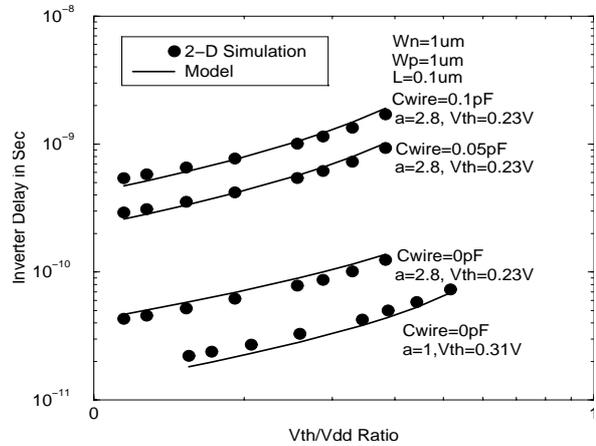


Figure 5 The delay is compared between Eq. (8) and 2-D numeral simulation. The circuit configuration considered in a 2-D MEDICI simulator is the same as shown in Figure 1.

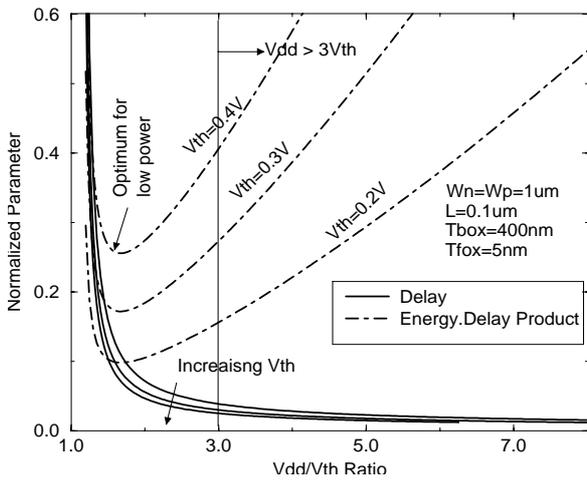


Figure 6 Effect of  $V_{dd}/V_{th}$  scaling on delay and energy-delay product in FDSOI CMOS inverter for unity stage ratio and  $C_{wire}=0$ .

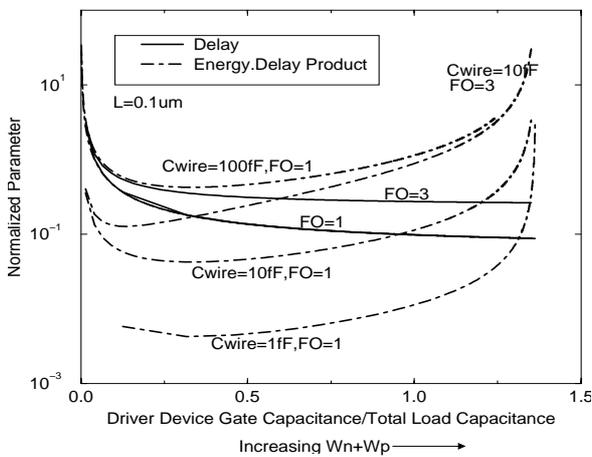


Figure 7 Normalized delay and energy-delay products versus drive device gate capacitance as a fraction of total load capacitance for unity stage ratio,  $T_{fox}=5nm$ ,  $T_{box}=400nm$ ,  $V_{dd}=1.2V$ , and  $V_{th}=0.3V$ .

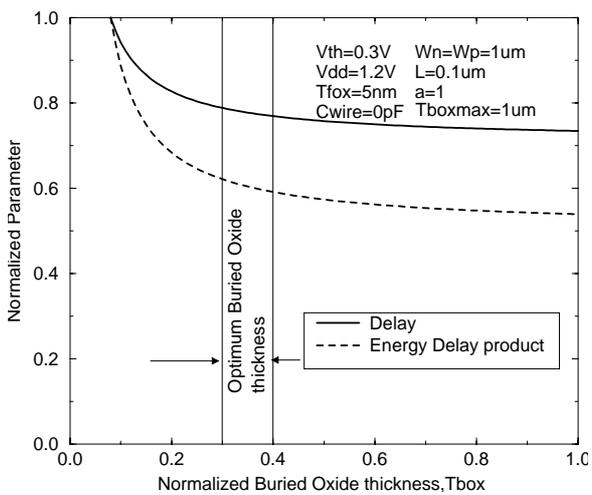


Figure 8 The delay and energy-delay products versus normalized buried oxide thickness are plotted. The buried oxide thickness chosen is normalized with 1mm.

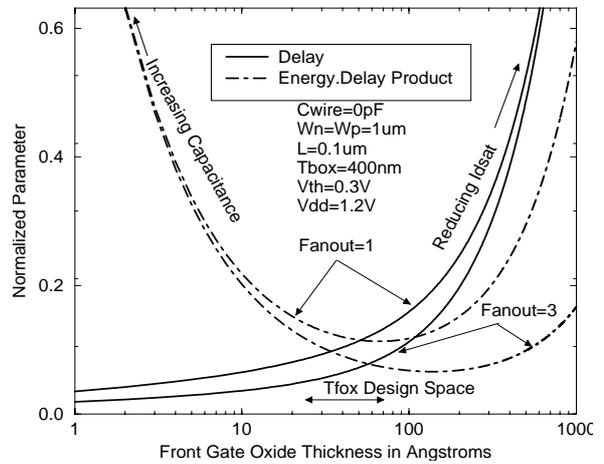


Figure 9 Normalized delay and energy-delay products are plotted for various gate oxide thicknesses.

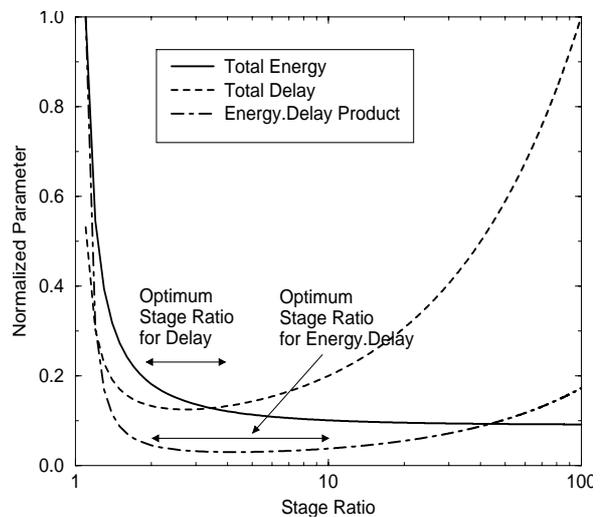


Figure 10 The delay and energy-delay products versus stage ratio are plotted. Low power and optimal speed can be obtained at a stage ratio equal to 2.8.