

# Power-Conscious Interconnect Buffer Optimization with Improved Modeling of Driver MOSFET and Its Implications to Bulk and SOI CMOS Technology

Koichi Nose and Takayasu Sakurai

Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan  
Phone: +81-3-5452-6253, FAX: +81-3-5452-6632, {nose, tsakurai}@iis.u-tokyo.ac.jp

## ABSTRACT

Closed-form formulas for optimum buffer insertion where the junction capacitance is taken into account are proposed. In order to use the derived formulas, an appropriate choice of the effective linear resistance of the driving transistor is also clarified. Using the proposed formulas, the optimum interconnect delay and power comparison between bulk and SOI CMOS technology are discussed. The calculation results show that both the optimum delay and power with SOI can be reduced by 15% compared with the bulk MOSFET whose junction capacitance is assumed to be equal to the gate capacitance.

## Categories & Subject / General Terms

B.7.1 Integrated circuits / Performance, design

## 1. Introduction

Interconnect delay optimization by buffer insertion is an indispensable technique for deep submicron VLSI's. RC models for MOSFET's have been used to optimize the buffered interconnect. As for the resistor, the transistor has been approximated as a linear resistor without detailed consideration on the non-linear feature of MOS I-V curves. As for the capacitance, the junction capacitance,  $C_J$ , has often been neglected [1] or even if  $C_J$  is taken into account, the delay formula including  $C_J$  is not sufficiently accurate. Moreover, the existing theories for buffered interconnect optimization are lacking in the trade-off between the delay and the power consumption although the power is one of the most important index in future giga-scale integration.

In order to overcome the shortcomings of the conventional approach, in this paper, approximation of MOSFET as a linear resistor is investigated and the delay formula including  $C_J$  is proposed.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ISLPED '02, August 12-14, 2002, Monterey, California, USA.  
Copyright 2002 ACM 1-58113-475-4/02/0008...\$5.00.

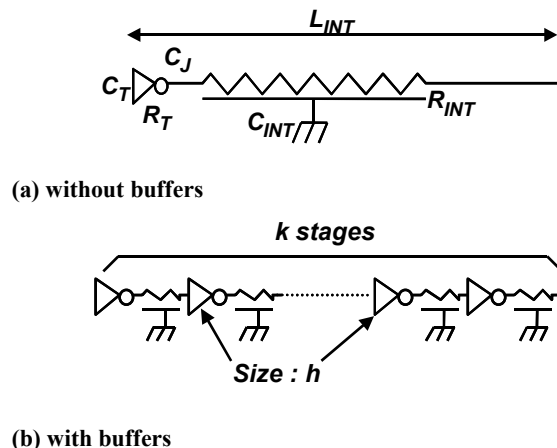


Figure 1 Distributed RC interconnect model

The paper also gives attention to the power consumption in the optimization process and derives closed-form formulas for optimum buffer insertion. The results have been applied to bulk and SOI technologies and implications of buffered interconnect on technologies are discussed.

## 2. Analytical Model for Buffer Optimization

Figure 1 shows a basic configuration of buffered interconnect. The inductive effect is neglected in this paper since the effect on optimum buffer inserted lines will diminish and become negligible for global interconnects in the future [2]. In order to minimize the delay, uniform buffers are inserted [3]. The delay formula without buffers can be approximated as (1). Suffix 0 signifies quantity per unit size or length.

This expression is newly derived and the relative error of the delay of (1) and the SPICE simulation results is shown in Table I. The relative error is within 3% when  $C_J=0$  and within 8% when  $C_J$  is equal to or less than  $C_T$ , which is the input capacitance of a transistor.

When the buffers are inserted like in Fig. 1(b), the optimum size of the buffers and the optimum number of the buffers can be derived analytically as

**Table I Relative error of  $t_d$  of (1) and SPICE simulation results**

(a)  $C_J/C_T=0$

	$R_T/R_{INT}$				
$C_T/C_{INT}$	0.1	0.5	1	5	10
0.1	2.4	2.7	2.1	0.7	0.3
0.5	2.7	3.0	2.5	0.9	0.5
1	2.1	2.5	2.2	0.8	0.4
5	0.7	0.9	0.8	0.4	0.2
10	0.3	0.5	0.4	0.2	0.1

(b)  $C_J/C_T=1$

	$R_T/R_{INT}$				
$C_T/C_{INT}$	0.1	0.5	1	5	10
0.1	2.9	3.5	2.7	0.8	0.4
0.5	4.2	5.7	4.6	1.4	0.7
1	4.3	6.5	5.4	1.6	0.8
5	3.9	7.4	6.3	1.9	1.0
10	3.8	7.5	6.4	1.9	1.0
100	3.6	7.6	6.5	2.0	1.0

\* Errors are in percent.

$$\frac{\partial t_d}{\partial h} = 0 \rightarrow h_{OPT} = \sqrt{\frac{C_{INT0} R_0}{R_{INT0} C_0}}, \quad (2)$$

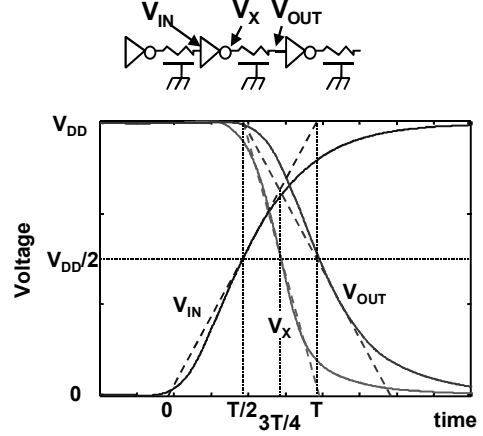
$$\frac{\partial t_d}{\partial k} = 0 \rightarrow k_{OPT} = L_{INT} \sqrt{\frac{p_1}{p_2}} \sqrt{\frac{R_{INT0} C_{INT0}}{R_0 (C_0 + C_{J0})}}. \quad (3)$$

$h_{OPT}$  is the optimum size of the buffers and  $k_{OPT}$  is the optimum number of the buffers.  $L_{INT}$  is the interconnect length.

Substituting  $h_{OPT}$  (2) and  $k_{OPT}$  (3) into (1), the optimum delay ( $t_{dOPT}$ ) can be expressed as

$$\begin{aligned} t_{dOPT} &= 2L_{INT} \left( \sqrt{p_1 p_2} + p_2 \sqrt{\frac{C_0}{C_0 + C_{J0}}} \right) \sqrt{\tau_{INT0} \tau_{MOS0}} \\ &\cong 2.4L_{INT} \sqrt{\tau_{INT0} \tau_{MOS0}} \quad (\text{when } C_{J0} = 0) \\ &\cong 2.0L_{INT} \sqrt{\tau_{INT0} \tau_{MOS0}} \quad (\text{when } C_{J0} = C_0) \end{aligned} \quad (4)$$

$\tau_{INT0}$  is the time constant of interconnect ( $=R_{INT0}C_{INT0}$ ) and  $\tau_{MOS0}$  is the time constant of a buffer ( $=R_0(C_0+C_{J0})$ ) which corresponds to the inverter delay with fanout of 1. The optimum delay is proportional to a geometric mean of the



**Figure 2 Signal waveforms when buffers are inserted**

interconnect delay ( $\tau_{INT0}$ ) and the gate delay ( $\tau_{MOS0}$ ). This means that the delay of optimally buffered interconnect is approximately scaled as  $\sqrt{s}$  where  $s$  is a scaling variable. It is also shown that the optimal condition is met when inserted buffer delay is approximately equal to the interconnect delay.

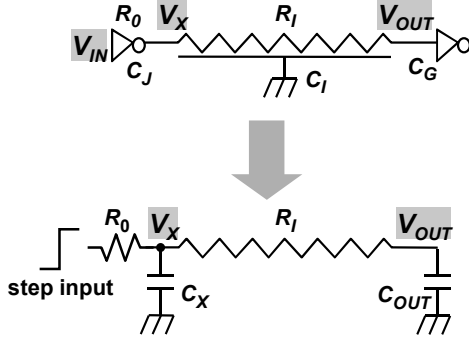
In order to use the derived formulas, the effective linear resistance of the unit-sized transistor ( $R_0$ ) has to be determined from device characteristics. Here, an appropriate choice of the effective constant resistance is to be discussed. The waveform of the input voltage ( $V_{IN}$ ), driver output voltage ( $V_X$ ) and interconnect output voltage ( $V_{OUT}$ ) are shown in Fig. 2. The waveforms can be considered as the ramp waveforms and  $\alpha$ -power model [4] is used as the drain current model.

In order to derive  $R_0$ , one section of buffered interconnect is approximated by one-step  $\pi$  RC circuit connected to  $R_0$  [5], depicted in Fig.3.  $R_I$  and  $C_I$  are the interconnect resistance and interconnect capacitance of one section, respectively.  $C_X$  is the sum of  $C_J$  and  $C_I/2$  and  $C_{OUT}$  is the sum of  $C_G$  and  $C_I/2$ . The expression for  $R_0$  is calculated first assuming the following points and then evaluated using rigorous simulations.

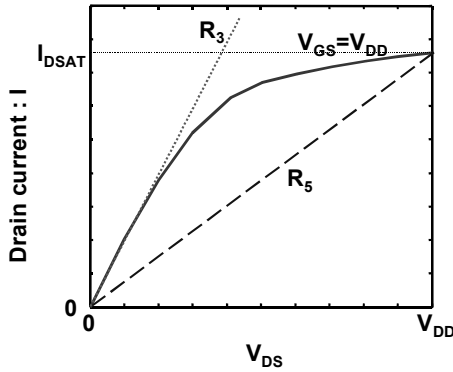
- (a) Fanout is set to 1, since sections are repeated.
- (b)  $V_X$  and  $V_{OUT}$  are start to fall at  $T/2$  simultaneously as in Fig.2.
- (c) The time constant of  $V_{OUT}$  ( $\tau_{OUT}$ ) is twice as large as that of  $V_X$  ( $\tau_X$ ), as in Fig.2.
- (d)  $C_X = C_{OUT}$

$$\begin{aligned} t_d &= p_1 R_{INT} C_{INT} + p_2 (R_T C_T + R_T C_J + R_T C_{INT} + R_{INT} C_T) \\ &= k \left[ p_1 \frac{R_{INT0} L_{INT}}{k} \frac{C_{INT0} L_{INT}}{k} + p_2 \left( \frac{R_0}{h} h(C_0 + C_{J0}) + \frac{R_0}{h} \frac{C_{INT0} L_{INT}}{k} + \frac{R_{INT0} L_{INT}}{k} h C_0 \right) \right] \end{aligned} \quad (1)$$

where  $p_1=0.377$  and  $p_2=0.693$



**Figure 3** Simple model for deviation of effective linear resistance



**Figure 4** Definition of  $R_5$  and  $R_3$

$\tau_X$  and  $\tau_{OUT}$  are described as

$$\tau_X = R_0(C_X + C_{OUT}) \quad (5)$$

$$\tau_{OUT} = R_0(C_X + C_{OUT}) + R_I C_{OUT} = 2\tau_X \quad (6)$$

$V_X$  is expressed as the function of  $\tau_X$ .

$$V_X = V_{DD} \left( 1 - e^{-\frac{t}{\tau_X}} \right) \quad (7)$$

$V_X$  is  $V_{DD}$  at  $T/2$  and falls to  $V_{DD}/2$  at  $3T/4$  as is shown in Fig.2. Then,  $T$  can be derived from Eq.7.

$$e^{-\frac{T/4}{\tau_X}} = \frac{1}{2} \quad (8)$$

$$\rightarrow T = (4 \ln 2) \cdot R_0(C_X + C_{OUT}) = (4 \ln 2) \cdot R_0 C$$

where  $C = (C_X + C_{OUT})$ .

The total charge which is discharged during  $T/2 \sim 3T/4$  is written as

$$\begin{aligned} \Delta Q &= \frac{1}{2} C_X V_{DD} + \frac{1}{4} C_I V_{DD} \\ &= \frac{1}{4} C V_{DD} + \frac{1}{8} C V_{DD} = \frac{3}{8} C V_{DD} \end{aligned} \quad (9)$$

From the viewpoint of the drain current which is expressed as

$$I = \beta(V_{GS} - V_{TH})^\alpha, \quad (10)$$

the total charge supplied from the input buffer between  $T/2$  and  $3T/2$  ( $\Delta Q$ ) can be calculated.

$$\begin{aligned} \Delta Q &= \int_{T/2}^{3T/4} \beta(V_{GS} - V_{TH})^\alpha dT \\ &= \int_{V_{DD}/2}^{3V_{DD}/4} \beta(V_{GS} - V_{TH})^\alpha \cdot \frac{T}{V_{DD}} dV_{GS} \\ &= \frac{\beta T V_{DD}^\alpha}{\alpha + 1} \left[ \left( \frac{3}{4} - v_T \right)^{\alpha+1} - \left( \frac{1}{2} - v_T \right)^{\alpha+1} \right] \end{aligned} \quad (11)$$

where  $v_T = V_{TH}/V_{DD}$ .

$R_5$ , which is the transistor resistance when  $V_{DS} = V_{GS} = V_{DD}$  (see Fig. 4), is written as

$$R_5 = \frac{V_{DD}}{\beta(V_{DD} - V_{TH})^\alpha} = \frac{1}{\beta V_{DD}^{\alpha-1} (1 - v_T)^\alpha} \quad (12)$$

Substituting (8), (9) and (12) into (11), following equation is derived.

$$\begin{aligned} \Delta Q &= \frac{T V_{DD}}{R_5} \frac{1}{(\alpha + 1)(1 - v_T)^\alpha} \left[ \left( \frac{3}{4} - v_T \right)^{\alpha+1} - \left( \frac{1}{2} - v_T \right)^{\alpha+1} \right] \\ &= \frac{R_0}{R_5} \cdot \frac{(4 \ln 2) C V_{DD}}{(\alpha + 1)(1 - v_T)^\alpha} \left[ \left( \frac{3}{4} - v_T \right)^{\alpha+1} - \left( \frac{1}{2} - v_T \right)^{\alpha+1} \right] \\ &= \frac{3}{8} C V_{DD} \end{aligned} \quad (13)$$

From (13), the effective linear resistance can be solved as the function of  $R_5$ .

$$\eta = \frac{R_0}{R_5} = \frac{3 \cdot (\alpha + 1)}{32 \cdot \ln 2} \cdot \frac{(1 - v_T)^\alpha}{(3/4 - v_T)^{\alpha+1} - (1/2 - v_T)^{\alpha+1}} \quad (14)$$

where  $v_T$  is  $V_{TH}/V_{DD}$ . In order to give insight into the parametric dependence of  $\eta$ , (14) is simplified as

$$\eta = \frac{R_0}{R_5} = \frac{R_T / h}{V_{DD} / I_{D0}} \cong 0.7\alpha + 1.5v_T \quad (15)$$

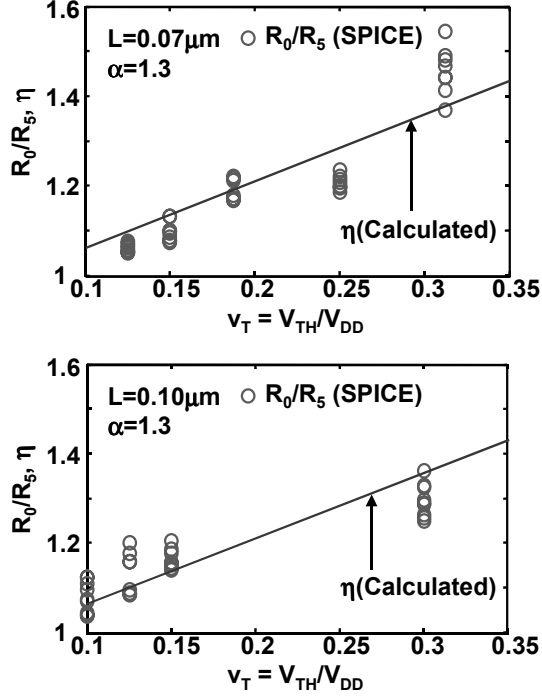


Figure 5  $R_0/R_5$  and  $\eta$  dependence on  $v_T$

This expression acts as a bridge between the effective transistor resistance and device characteristics. In Fig. 5, the SPICE simulation results are compared with (15). Different technology models and various interconnect width and height are used for this simulation and the validity of (15) is confirmed. Figure 6 shows the optimum delay comparison between the proposed method where the effective linear resistance ( $R_0$ ) is used and the conventional method in [5] where the linear resistance is chosen as the  $R_3$  ( $=1/(\text{maximum drain conductance})$ ) as is shown in Fig. 4). The discrepancy between the delay simulated by SPICE with real buffers and a distributed RC line and the calculated delay with the effective linear resistance ( $R_0$ ) is within 3%. On the other hand, the discrepancy between SPICE simulated delay and the delay calculated with the conventional  $R_3$  is more than 30%. On the other hand, the discrepancy in power between these methods is within 6% (see Fig.7). The optimum buffer size ( $h_{OPT}$ ) is proportional to  $\sqrt{R_0}$  and the optimum number of buffers ( $k_{OPT}$ ) is proportional to  $1/\sqrt{R_0}$ . This is why the total power with buffers, which is proportional to  $h_{OPT} \cdot k_{OPT}$ , is unchanged even if the effective linear resistance is changed.

Then, in order to confirm the validity of the proposed formulas for  $h_{OPT}$ ,  $k_{OPT}$  and  $t_{dOPT}$ , theoretical calculations and SPICE results are compared. The model parameter set for SPICE simulation and for proposed formulas are extracted from measured data with 0.25μm PD-SOI technology whose test chip is shown in Fig. 8. The SPICE model agrees well with the measured results as in Fig. 9. Figure 10 shows the  $h_{OPT}$ ,  $k_{OPT}$  and  $t_{dOPT}$  comparison between rigorous optimization results with SPICE and the

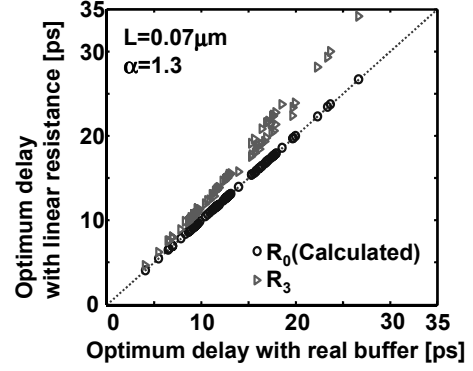


Figure 6 Optimum delay comparison between  $R_0$  and  $R_3$

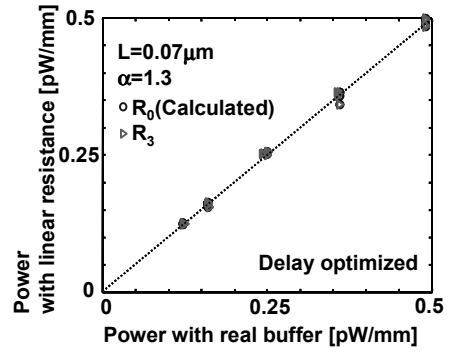


Figure 7 Power comparison between  $R_0$  and  $R_3$

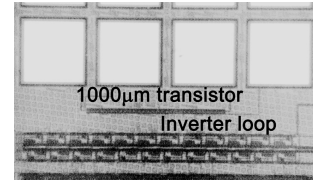


Figure 8 Microphotograph of test chip fabricated by 0.25μm PD SOI process

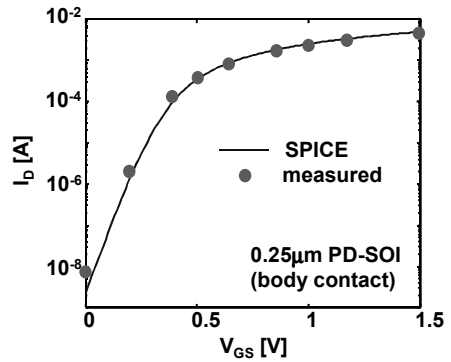


Figure 9 Drain current comparison between SPICE model and measured data

calculated results. Figure 11 shows the power dependence on the  $C_{J0}/C_0$ . When the junction capacitance is negligible, both the optimum delay and the power with buffers are suppressed by 15% compared with the MOSFET with  $C_{J0}=C_0$ . It is shown from (2), (3) and (4) that the 15% reduction on power and delay is independent from the technology node.

### 3. Interconnect Delay and Power Comparison Between Bulk and SOI Technology

Extending the analysis, the optimum interconnect delay comparison among bulk, PD-SOI, FD-SOI and double-gate structure [6] is discussed using the simple model. The characteristics of these models are listed in Table II. We set the leakage current of these structures equal to make the comparison fair. Then,  $V_{TH}$  of FD-SOI and double-gate can be lowered since the  $S$ -factor is smaller than other structures.  $C_{J0}/C_0$  and  $V_{TH}/V_{DD}$  are the measured data of five different technologies.  $C_{J0}/C_0$  of conventional bulk process are 0.7~1.3. This value does not change drastically over generations.

The calculated results are shown in Fig. 12. PD-SOI with body contact is 12% faster than bulk CMOS technology due to the small junction capacitance. It is often discussed that SOI technology does not give speed and power improvement over bulk CMOS technology in deep submicron designs, since speed and power are determined by interconnects and SOI technology does not change interconnect layers. It is not necessarily true because deep submicron interconnect systems need relatively large buffers and due to the improvement through buffers, SOI technology still enjoys advantage over bulk CMOS. The delay can be further decreased by using PD-SOI with a floating body or FD-SOI since the drain current is enhanced by the kink effect and the lower threshold voltage. If lower  $C_J$  is achievable with bulk CMOS technology, the bulk technology approaches SOI results.

In the optimally buffered interconnect, the power dissipation increases due to the buffers. Here, the trade-off between power and delay is discussed. Let us introduce the parameter,  $p$ , which is the ratio of the total power (buffers and interconnect),  $P_{TOTAL}$ , to the power consumed by pure interconnect,  $P_{INT}$ .

$$p = \frac{P_{TOTAL}}{P_{INT}} = \frac{C_{INT} + kh(C_0 + C_{J0})}{C_{INT}} \quad (16)$$

If  $p$  is fixed, the optimum buffer size,  $h$ , the number of the sections,  $k$ , and the delay,  $t_d$ , can be expressed as follows.

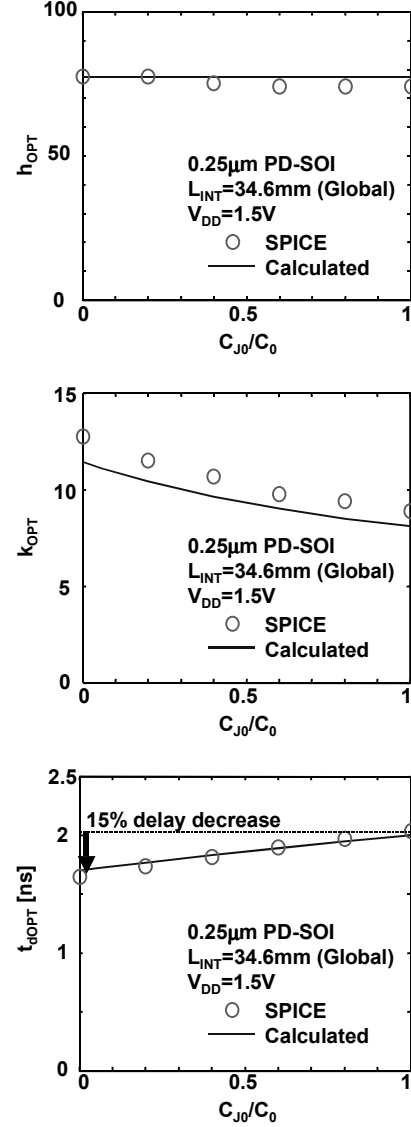


Figure 10  $h_{OPT}$ ,  $k_{OPT}$  and  $t_{dOPT}$  comparison between calculated results and SPICE simulations

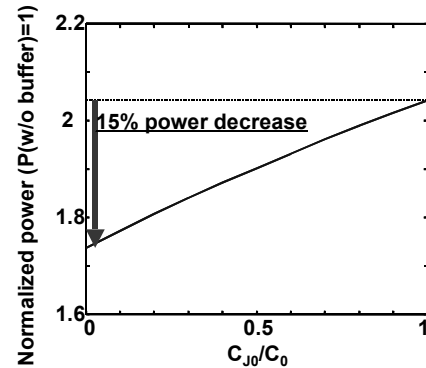
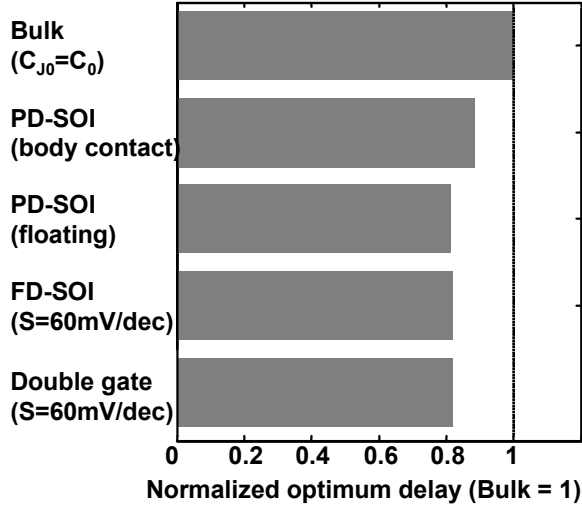


Figure 11 Power dependence on  $C_{J0}/C_0$

**Table II Bulk and SOI structure**

	$C_{J0}/C_0$	$V_{TH}/V_{DD}$	
Bulk	0.92	0.18	
PD-SOI (body contact)	0.13	0.18	
PD-SOI (floating)	0.13	0.18	$I_{ON} \times 1.15$ (kink)
FD-SOI	0.13	0.13	$S=60\text{mV/decade}$
Double-gate [5]	0.13	0.13	$S=60\text{mV/decade}$ $I_{ON} \times 2, C_0 \times 2$



**Figure 12 Delay comparison between bulk and SOI**

$$\begin{aligned} \frac{h}{h_{OPT}} &= \sqrt{\frac{p(p-1)p_2C_0}{C_P}} \\ \frac{k}{k_{OPT}} &= \sqrt{\frac{(p-1)C_P}{pp_1C_0}} \\ \frac{t_d}{t_{dOPT}} &= \frac{1}{\sqrt{p_1 + \sqrt{p_2}}} \sqrt{\frac{pC_P}{(p-1)C_0}} \end{aligned} \quad (17)$$

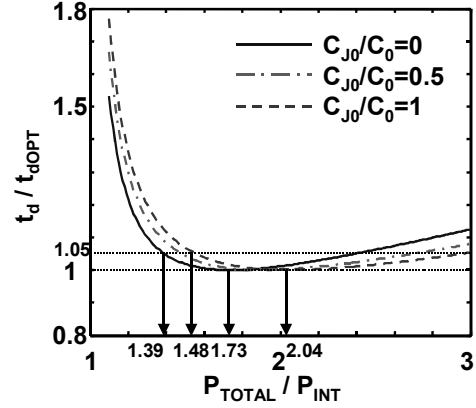
where

$$C_P = p_1(C_0 + C_{J0}) + p_2C_0(p-1) \quad (18)$$

The delay dependence on the total power is calculated using the proposed formulas. The result is shown in Fig. 13. It can be seen from the figure that the power can be reduced by 20% if delay is allowed to increase by 5%.

## 4. Conclusion

Closed-form formulas for optimum buffer insertion with the junction capacitance effect taken into account are proposed and an approximation of MOSFET as a linear resistor is investigated.



**Figure 13 Delay dependence on total power**

Using these formulas, the optimum interconnect delay comparison among bulk, PD-SOI, FD-SOI and double-gate structure is discussed. If the junction capacitance can be negligible, the optimum interconnect delay is 15% smaller than the delay when  $C_{J0}=C_0$ . MOSFET with small junction capacitance, like SOI, can suppress the interconnect delay by 15% compared with MOSFET with  $C_{J0}=C_0$ , like conventional bulk MOSFET.

## Acknowledgement

The work has been supported by Toshiba Corporation.

## References

- [1] Y. I. Ismail and E. G. Friedman, "Effects of inductance on the propagation delay and repeater insertion in VLSI circuits," *IEEE Trans. VLSI systems*, vol. 8, No. 2, pp.195-206, Apr. 2000.
- [2] K. Banerjee and A. Mehrotra, "Accurate analysis of on-chip inductance effects and implications for optimal repeater insertion and technology scaling," *Symposium on VLSI Circuits, Dig. of Tech. Papers*, pp.195-198, 2001.
- [3] V. Adler and E. G. Friedman, "Repeater design to reduce delay and power in resistive interconnect," *IEEE Trans. Circuits and Systems II*, vol. 45, pp.607-616, May, 1998.
- [4] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its application to CMOS inverter delay and other formulas," *IEEE Journal of Solid-State Circuits*, vol.25, pp.584-593, Apr., 1990.
- [5] T. Sakurai, "Approximation of wiring delay in MOSFET LSI," *IEEE Journal of Solid-State Circuits*, vol.SC-18, no.4, pp.418-426, Aug., 1983.
- [6] T. Tanaka, H. Horie, S. Ando and S. Hijiya, "Analysis of P<sup>+</sup> poly Si double-gate thin-film SOI MOSFETs," *IEEE International Electron Device Meeting (IEDM)*, pp.683-686, 1991.