# A SIMPLE MOSFET MODEL FOR CIRCUIT ANALYSIS AND ITS APPLICATION TO CMOS GATE DELAY ANALYSIS AND SERIES-CONNECTED MOSFET STRUCTURE

by

Takayasu Sakurai and A. Richard Newton

Memorandum No. UCB/ERL M90/19

16 March 1990

COVER PAGE

# A SIMPLE MOSFET MODEL FOR CIRCUIT ANALYSIS AND ITS APPLICATION TO CMOS GATE DELAY ANALYSIS AND SERIES-CONNECTED MOSFET STRUCTURE

by

Takayasu Sakurai and A. Richard Newton

Memorandum No. UCB/ERL M90/19

16 March 1990

**ELECTRONICS RESEARCH LABORATORY**

College of Engineering
University of California, Berkeley
94720

# A SIMPLE MOSFET MODEL FOR CIRCUIT ANALYSIS AND ITS APPLICATION TO CMOS GATE DELAY ANALYSIS AND SERIES-CONNECTED MOSFET STRUCTURE

by

Takayasu Sakurai and A. Richard Newton

Memorandum No. UCB/ERL M90/19

16 March 1990

**ELECTRONICS RESEARCH LABORATORY**

College of Engineering
University of California, Berkeley
94720

# A Simple MOSFET Model for Circuit Analysis and its Application to CMOS Gate Delay Analysis and Series-Connected MOSFET Structure

**Takayasu Sakurai* and A. Richard Newton**

**Department of Electrical Engineering and Computer Sciences**
**University of Califiifornia, Berkeley, CA94720, U.S.A.**
***) On leave from Semiconductor Device Engineering Lab.,**
**Toshiba Corporation, Kawasaki, 210, Japan**

## Contents

This report consists of two parts. The first half titled "A Simple MOSFET Model for Simulators and Circuit Analysis" describes a new simple yet realistic MOSFET model which is suitable for analytical treatment of MOSFET circuits. The second part is dedicated to a application of the MOS model to series-connected MOSFET structures. The title of the second half is "CMOS Gate Delay Analysis in the Submicron Region and its Application to Series-Connected MOSFETs".

### The first half



### The second half

# A Simple MOSFET Model for Simulators and Circuit Analysis

Takayasu Sakurai* and A. Richard Newton

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA94720, U.S.A.

* On leave from Semiconductor Device Engineering Lab.,
Toshiba Corporation, Kawasaki, 210 Japan.

**Abstract**

A simple, general and realistic MOSFET model is introduced. The model can express the current characteristics of short-channel MOSFETs at least down to 0.25μm channel-length, GaAs FET, and resistance inserted MOSFETs. The model evaluation time is about 1/3 of the evaluation time of the SPICE3 MOS LEVEL3 model. The model parameter extraction is done by solving single variable equations and thus can be done within a second, being different from the fitting procedure with expensive numerical iterations employed for the conventional models. The model also enables analytical treatments of circuits in short-channel region and makes up for a missing link between a complicated MOSFET current characteristics and circuit behaviors in the deep submicron region.

# 1. Introduction

The SPICE LEVEL1 model[1], which is based on the Shockley model, is widely used in circuit analysis However, the I-V characteristics reproduced by the model is far from accurate in the short-channel region as shown in Fig. 1. There is a discussion that even if it does not model the individual MOSFET characteristics accurately, it might predict the relative gate delay correctly. However, it has been shown that this is not true in the submicron region[2]. For example, in the submicron region, an 8-input NAND shows only 4~5 times longer delay compared with an inverter when the input is fast and an output capacitance is very large. This is because the series-connected MOSFETs structure used in the NAND structure mitigates the $V_{DS}$ and $V_{GS}$ of each MOSFET and this in turn reduces the severe velocity saturation effect observed in the submicron region. The simulation with the Shockley model always shows more than 8 times longer delay for the 8-input NAND gate compared with an inverter, because it does not include the velocity-saturation effects at all. So even for first-order approximation, SPICE LEVEL1 model is not sufficient.

On the other hand, there are more precise MOS models like SPICE LEVEL3 model[3], BSIM[4], table look-up models[5], and so on[6,7]. However, some of them are time-consuming in evaluating models[3] and some of them needs special system with hardware/software combination for extracting model parameters[4,5,6] and the number of parameters is huge. Moreover, most of the precise models[3,4,6,7] need a model parameter extraction procedure with expensive numerical iterations[8,9] and once the extracted model parameters happen not to give a satisfactory results, there is no way to know whether the problem lies in the model itself or in the extracting procedure. Sometimes it takes hours to struggle with the parameter set. In a word, they

lack of the easy-to-use feature.

In order to fill the gap between the too simple LEVEL1 model and the more precise models, a new model is proposed in this paper. The new model offers a fast, easy-to-use, general and realistic feature at a cost of rough approximation near and below the threshold voltage. The near- and sub- threshold region modeling is not important in calculating delay of most VLSIs. The modeling of the region is important in estimating the charge decay characteristic of charge storage nodes but in this case a statistical model should be used since it is very sensitive to process variation.

The fast model is suitable especially for timing simulators and electrical-logic simulators[17,18], where the model evaluation is the most time-consuming part of the program and the use of the complicated model greatly reduce their attractiveness.

The model is presented in Section 2. The model parameter set is compact and can be used for exchanging MOSFET information among institutions instead of LEVEL3 model parameter set. The model parameter extraction is done easily as is explained in Section 3. Section 4 is dedicated for the property of the proposed model implemented in SPICE3.

The proposed model also enables an analytical treatment of circuit operation, which helps the general understanding of circuit behavior in the submicron region. The analysis is described in Section 5 followed by the conclusion in Section 6.

## 2. Model Description

The proposed model equations are as follows. $I_D$ is the drain current.

$$V_{TH} = V_{T0} + \gamma \left(\sqrt{2\phi_F - V_{BS}} - \sqrt{2\phi_F}\right) \tag{2.1}$$

$$V_{DSAT} = K \left(V_{GS} - V_{TH}\right)^m \tag{2.2}$$

$$I_{DSAT} = \frac{W}{L_{eff}} B \left(V_{GS} - V_{TH}\right)^n \tag{2.3}$$

$$I_D = I_{D5} = I_{DSAT} \left(1 + \lambda V_{DS}\right) \quad , \lambda = \lambda_0 - \lambda_1 V_{BS} \quad \left(V_{DS} \geq V_{DSAT} : \text{saturated region}\right) \tag{2.4}$$

$$I_D = I_{D3} = I_{D5} \left[2 - \frac{V_{DS}}{V_{DSAT}}\right] \frac{V_{DS}}{V_{DSAT}} \quad \left(V_{DS} < V_{DSAT} : \text{linear region}\right) \tag{2.5}$$

where $V_{GS}$, $V_{DS}$ and $V_{BS}$ are gate-source, drain-source, bulk-source voltages, respectively. W is a channel width and $L_{eff}$ is an effective channel length. $V_{TH}$ denotes a threshold voltage, $V_{DSAT}$ a drain saturation voltage, and $I_{DSAT}$ a drain saturation current. $V_{T0}$, $\gamma$ and $2\phi_F$ are parameters which describe the threshold voltage. Parameters K and m control the linear region characteristics while B and n determine the saturated region characteristics. $\lambda_0$ and $\lambda_1$ are related to the finite drain conductance in the saturated region.

The model is reduced to the Shockley model if K=1, m=1, B=0.5Kp, and n=2. The model can also express an I-V characteristics where $V_{DSAT}$ is proportional to $(V_{GS} - V_{TH})^{0.5}$ and $I_{DSAT}$ is proportional to $(V_{GS} - V_{TH})$, which is predicted by a short-channel MOSFET theory[10]. An application of the model to 0.25μm MOSFETs[6] is shown in Figs. 2-7. The results are satisfactory. The model parameters are listed in TABLE 1. The primitive version of this model is seen in Ref.[11] but the linear

region modeling was very coarse.

In the submicron devices, the contact resistance drain/source diffusion resistance and hot-carrier induced drain resistance[12] are indispensable. So it is better for the contemporary MOS model to incorporate these resistance effects in current parameters. If the series resistance is modeled by resistors, it is practically difficult to separate these resistance component and the resistors increase number of nodes and hence calculation time. Since the present model is quite general, it can reproduce the I-V curves of resistor-inserted MOSFET only by changing parameters. An example is shown in Fig. 8, where resistors whose value is 10% of the effective MOSFET resistance are inserted in the drain and the source.

To demonstrate that the model is quite general, the model is applied to GaAs FET[13] in Fig. 9. The salient feature of the GaAs FET is that $V_{DSAT}$ is constant and not a function of $V_{GS}$.

## 3. Parameter Extraction

The model parameter extraction starts by selecting fitting points on the I-V curves like in Fig. 10. Then the following formulas give all the parameters. The subindex ",i" (i = 1-11) corresponds to the fitting point number in the figure.

$$\lambda_0 = \frac{I_{D,2} - I_{D,1}}{I_{D,1} V_{DS,2} - I_{D,2} V_{DS,1}} \tag{3.1}$$

$$I_{Z3} = I_{D,3} / (1 + \lambda_0 V_{DS,3}) \ , \ I_{Z4} = I_{D,4} / (1 + \lambda_0 V_{DS,4}) \ , \ I_{Z5} = I_{D,5} / (1 + \lambda_0 V_{DS,5}) \tag{3.2}$$

Then, $V_{T0}$ can be obtained by solving the following equation. The bisection method[14] is the best choice for the solution since it finds out the root without fail

within 10 iterations.

$$f_V(V_{T0}) = \log\left[\frac{I_{Z3}}{I_{Z4}}\right] \log\left[\frac{V_{GS,4}-V_{T0}}{V_{GS,5}-V_{T0}}\right] - \log\left[\frac{I_{Z4}}{I_{Z5}}\right] \log\left[\frac{V_{GS,3}-V_{T0}}{V_{GS,4}-V_{T0}}\right] = 0 \qquad (3.3)$$

$$n = \frac{\log\,(I_{Z3}/I_{Z4})}{\log\,(\,(V_{GS,3}-V_{T0})\,/\,(V_{GS,4}-V_{T0})\,)}\,,\quad B = \frac{I_{Z3}}{(V_{GS,3}-V_{T0})^n} \qquad (3.4)$$

$$E_6 = I_{D,6}\,/\,\{B\,(V_{GS,6}-V_{T0})^n\,(1+\lambda_0 V_{DS,6})\} \qquad (3.5)$$

$$E_7 = I_{D,7}\,/\,\{B\,(V_{GS,7}-V_{T0})^n\,(1+\lambda_0 V_{DS,6})\} \qquad (3.6)$$

$$V_{DSAT,6} = V_{DS,6}\,(1+\sqrt{1-E_6})/E_6\,,\quad V_{DSAT,7} = V_{DS,7}\,(1+\sqrt{1-E_7})/E_7 \qquad (3.7)$$

$$m = \frac{\log\,(V_{DSAT,6}/V_{DSAT,7})}{\log\,(\,(V_{GS,6}-V_{T0})\,/\,(V_{GS,7}-V_{T0})\,)}\,,\quad K = \frac{V_{DSAT,6}}{(V_{GS,6}-V_{T0})^m} \qquad (3.8)$$

$\lambda_1$ is obtained from the following equation.

$$\frac{I_{D,11}-I_{D,10}}{I_{D,10}V_{DS,11}-I_{D,11}V_{DS,10}} = \lambda_0 - \lambda_1 V_{BS,10} \qquad (3.9)$$

$$I_{D,8}\,/\,(1+\lambda_0 V_{DS,8}-\lambda_1 V_{BS,8} V_{DS,8}) = K\,(\,V_{GS,8}-V_{TH,8})^n \qquad (3.10)$$

$$I_{D,9}\,/\,(1+\lambda_0 V_{DS,9}-\lambda_1 V_{BS,9} V_{DS,9}) = K\,(\,V_{GS,9}-V_{TH,9})^n \qquad (3.11)$$

After obtaining $V_{TH,8}$ and $V_{TH,9}$ using the above equations, $2\phi_F$ is obtained by solving the following equation with the bisection method.

$$f_P(2\phi_F) = (\sqrt{2\phi_F - V_{BS,8}} - \sqrt{2\phi_F})\,(V_{TH,9}-V_{T0}) - (\sqrt{2\phi_F - V_{BS,9}} - \sqrt{2\phi_F})\,(V_{TH,8}-V_{T0}) = 0$$

(3.12)

Even if the fitting results are not satisfying at the first trial it is easy and fast to

try again with slightly different fitting points, since the model parameters are appearance-oriented, that is, they have a direct meaning in controlling I-V curve shapes and they are non-degenerate. It is different from the other models where parameters are physics-oriented and thus some of the parameters have no direct connection to the shape of the I-V curves like VMAX parameter of LEVEL3 model.

Usually, from two to four retries were enough for the satisfactory results for 2μm, 1.2μm, 0.8μm and 0.5μm MOSFETs. It is even possible to extract model parameters from an I-V plot without any on-line data.

The extracted parameter set is valid only for a narrow range of channel-length but usually the shortest channel-length is used for almost all the MOSFETs in a VLSI and two or three sets of parameters are enough in designing a whole VLSI. The separate parameter set is also required for a very narrow width device and a shallow $V_{TH}$ device and an i-type device if they are employed. Even for more precise models, it is a good practice to use them near the condition where the model parameters are extracted, otherwise the model prediction is not guaranteed. And then several parameter sets are required with the more precise models.

## 4. Application to Circuit Simulation

Some of the computational properties of the model are listed in TABLE 2. The coding is straight-forward and the model evaluation time is about 1/3 of the LEVEL3 model. The codes are extracted from SPICE3. If the precision is not so important, the use of approximated formulas for log and exp functions[15] is effective and 30% further reduction in time is possible.

The simulation time when implemented in SPICE3 is listed in TABLE 3. The capacitance model used is the same model as LEVEL1, LEVEL2 and LEVEL3 capacitance model based on the Meyer's model which can be improved further[16]. The present model usually shows faster total simulation time than the LEVEL3 model. The simulated waveforms are compared in Fig. 11.

The fast model is suitable especially for timing simulators and electrical-logic simulators[17,18], where the model evaluation is the most time-consuming part of the program and the use of the complicated model greatly reduce their attractiveness.

## 5. Application to Circuit Analysis

As an application of the model to the circuit analysis, CMOS inverter delay is analyzed here. The derivation begins by setting up the differential equation which governs gate operation. This equation is then solved for the very fast input case and for the very slow input case and the two solutions are connected smoothly. The ramp input waveform is approximated by Vin,ap in Fig. 12 using the logic threshold $V_{INV}$ (See Fig. 13). The detailed derivation can be found in Ref.[2]. First, define a critical input transition time $t_{T0}$.

$$t_{T0} = \tau \frac{(n+1)(1-\nu_T)^n}{(1-\nu_T)^{n+1} - (\nu_V - \nu_T)^{n+1}} \left[ \frac{1}{2} + \frac{\lambda'}{7} \right] \tag{5.1}$$

where $\nu_T = V_{T0}/V_{DD}$, $\nu_V = V_{INV}/V_{DD}$, $\tau = C_O V_{DD}/I_{D0}$, $\lambda' = \lambda_0 V_{DD}$. and $\nu_{D0} = V_{D0}/V_{DD}$. $C_O$ is an output capacitance and $V_{DD}$ is an applied power voltage. $I_{D0}$ is defined as the drain current observed when $V_{GS} = V_{DS} = V_{DD}$ and is good index of drivability of a MOSFET (See Fig. 1). $V_{D0}$ is defined as the drain saturation voltage when $V_{GS} = V_{DD}$.

These two quantities together with the velocity saturation index n play an essential role in determining circuit behavior.

Then the delay $t_d$, the delay from $0.5V_{DD}$ of input to $0.5V_{DD}$ of output, and the effective output transition time, $t_{TOUT}$, can be expressed as follows. In calculating $t_{TOUT}$, the output waveform slope is approximated by 70% of its derivative at the half $V_{DD}$ point[19]. $t_{TOUT}$ can be used as $t_T$ for the next logic gate.

($t_T \le t_{T0}$ : for faster input)

$$t_d = t_T \left\{ \frac{1}{2} - \frac{1-v_T}{n+1} + \frac{(v_V - v_T)^{n+1}}{(n+1)(1-v_T)^n} \right\} + \tau \left[ \frac{1}{2} + \frac{\lambda'}{7} \right] \tag{5.2}$$

$$t_{TOUT} = \frac{\tau}{0.7} \frac{4v_{D0}^2(1+\lambda')}{(4v_{D0}-1)(2+\lambda')} \tag{5.3}$$

($t_T > t_{T0}$ : for slower input)

$$t_d = t_T \left[ v_T - \frac{1}{2} + \left\{ (v_V - v_T)^{n+1} + \frac{(n+1)(1-v_T)^n}{t_T/\tau} \left[ \frac{1}{2} + \frac{\lambda'}{7} \right] \right\}^{\frac{1}{n+1}} \right] \tag{5.4}$$

$$t_{TOUT} = \frac{\tau}{0.7} \left[ \frac{1-v_T}{t_d/t_T + 1/2 - v_T} \right]^n \frac{2+2\lambda'}{2+\lambda'} \tag{5.5}$$

The formulas are valid in wide range of $t_T$ and the channel-width ratio of PMOS and NMOS (Wp / Wn) as shown in Fig. 14. The logic threshold voltage, $V_{INV}$, was calculated by the following expression.

$$v_V = \frac{V_{INV}}{V_{DD}} = \frac{I_{D0n}^{1/n} v_{Tn} + I_{D0p}^{1/n} (1-v_{Tn})}{I_{D0n}^{1/n} + I_{D0p}^{1/n} (1-v_{Tn})/(1-v_{Tp})}, \quad n = \frac{n_n + n_p}{2} \tag{5.6}$$

where subindex n and p denote NMOS and PMOS, respectively. The accuracy of the

formulae is shown in Fig. 15. An application of the model for the analysis of series-connected MOSFET structure is seen in Ref.[2].

## 6. Conclusion

A new MOS model is proposed. The model offers the following features:

- simple yet realistic
- fast in evaluation
- easy and fast to extract model parameters
- general
- good for analytical treatment

The feasibility and effectiveness of the model are demonstrated by using 0.25μm MOSFETs and SPICE3.

Using the model, analytical expression is derived for CMOS inverter delay, which includes the CMOS effects and the velocity saturation effects.

## Acknowledgment

The encouragement of R. Brayton, A. Sangiovanni-Vincentelli, Y. Unno, Y. Takeishi, Y. Fukuda, H. Yamada and T. Iizuka throughout the course of this work is appreciated. 0.25μm MOSFET data was provided by M. C. Jeng and is acknowledged together with the fruitful suggestions from him. Assistance provided by H. Ishiuchi, T. Quarles and R. Spickelmier concerning MOS physics, SPICE and computer environments is also appreciated. This work was supported by a grant from Toshiba Corporation.

## References

[1] T.Quarles, A.R.Newton, D.O.Pederson, A.Sangiovanni-Vincentelli, "SPICE 3B1 User's Guide," EECS, Univ. of Calif. Berkeley, 1988.

[2] T.Sakurai and A.R.Newton, "CMOS Gate Delay Analysis in the Submicron Region and its Application to Series-Connected MOSFETs," submitted, IEEE J. Solid State Circuits.

[3] A.Vladimirescu and S.Liu, "The Simulation of MOS Integrated Circuits Using SPICE2," U.C.Berkeley ERL Memo, No.M80/7, Oct.1980.

[4] B.J.Sheu, "MOS Transistor Modeling and Characterization for Circuit Simulation," UCB/ERL Memo M85/85, Oct.1985.

[5] T.Shima, H.Yamada and R.Dang, "Table Lookup MOSFET Modeling System Using a 2-D Device Simulator and Monotonic Piecewise Cubic Interpolation," IEEE Trans. on CAD., CAD-2, pp.121-125, Apr.1983.

[6] M.C.Jeng, P.K.Ko and C.Hu, "A Deep-Submicron MOSFET Model for Analog/Digital Circuit Simulations," IEDM'88, pp.114-117, Dec.1988.

[7] A.Chatterjee, C.F.Machala and P.Yang, "A Submicron MOSFET Model for Simulation of Analog Circuit," ICCAD'88, pp.120-123, Nov.1988.

[8] J.Tsujimoto and J.Miyamoto, "Automatic Parameter Extraction for MOS Transistors," Nat'l Convention of IECEJ, No.392, p.2-155, 1984.

[9] K.Doganis and D.L.Sharfetter, "General Optimization and Extraction of IC Device Model Parameters," IEEE Trans. on ED, ED-30, pp.1219-1228, 1983.

[10] R.S.Muller and T.I.Kamins, Device Electronics for Integrated Circuits, 2nd Ed., p.482, John Wiley & Sons, NY, 1986.

[11] T.Sakurai, "CMOS Inverter Delay and Other Formulas Using α-Power MOS

Model," ICCAD'88, pp.74-77, Santa Clara, Nov.1988.

[12] K.Nogami, K.Sawada, M.Kinugawa and T.Sakurai, "VLSI Circuit Reliability under AC Hot-Carrier Stress," Symp. on VLSI Circuits, pp.13-14, May 1987.

[13] H.Statz, P.Newman, I.R.Smith, R.A.Pucel, and H.A.Haus, "GaAs FET Device and Circuit Simulation in SPICE," IEEE Trans. on ED, ED-34, No.2, pp.160-169, Feb.1987.

[14] W.H.Press, B.P.Flannery, S.A.Teukolsky and W.T.Vettering, Numerical Recipes in C, p.261, Cambridge Univ. Press, 1988.

[15] C.Hastings, Jr., Approximations for Digital Computers, Princeton Univ. Press, 1955.

[16] K.A.Sakallah, Y.T.Yen, and S.S.Greenberg, "The Meyer Model Revisited: Explaining and Correcting the Charge Non-Conservation Problem," ICCAD'87, pp.204-207, Nov.1987.

[17] Y.H.Kim, J.Kleckner, R.Saleh and A.R.Newton, "Electrical-Logic Simulation," ICCAD'84, Nov.1984.

[18] Y.H.Kim, "Accurate Timing Verification for Digital VLSI Designs," UCB/ERL Memo M89/2, Jan.1989.

[19] N.Hedenstierna and K.O.Jeppson, "CMOS Circuit Speed and Buffer Optimization," IEEE Trans. on CAD, CAD-6, No.2, pp.270-280, Mar.1987.

## TABLE 1. Model Parameters for 0.25um MOSFETs

| Parameters | NMOSFET | PMOSFET |
|---|---|---|
| B | 4.9721e-05 | 1.1151e-05 |
| n | 1.0484 | 1.3649 |
| K | 0.83496 | 1.0541 |
| m | 0.6193 | 0.74003 |
| $\lambda_0$ | 0.066265 | 0.128 |
| $\lambda_1$ | 0.0038573 | 0.012923 |
| $V_{T0}$ | 0.85502 | -0.87241 |
| $\gamma$ | 0.29648 | 0.26074 |
| $2\phi_F$ | 0.20556 | 0.21691 |

## TABLE 2. Computational Characteristics

| Item | LEVEL1 | Proposed Model | LEVEL3 |
|---|---|---|---|
| Coding lines in C[1] | ~40 | ~60 | ~210 |
| Time required for current/derivative calculation[2] | 3.4s | 7.5s | 21.1s |

1) Including derivative calculation and excluding comment lines

2) loop for $-3V \le V_{BS} \le 0$ step 0.5V, $0 \le V_{GS} \le 5V$ step 0.05V and $0 \le V_{DS} \le 5V$ step 0.05V

## TABLE 3. SPICE3 Simulation Time

| Circuit# | # of MOS's | LEVEL1 | Proposed Model | LEVEL3 |
|---|---|---|---|---|
| 1 | 14 | 2.7s | 6.4s | 6.9s |
| 2 | 68 | 33.3s | 34.5s | 118s |
| 3 | 640 | 227s | 171s | 184s |
| 4 | 1060 | 303s | 372s | 808s |

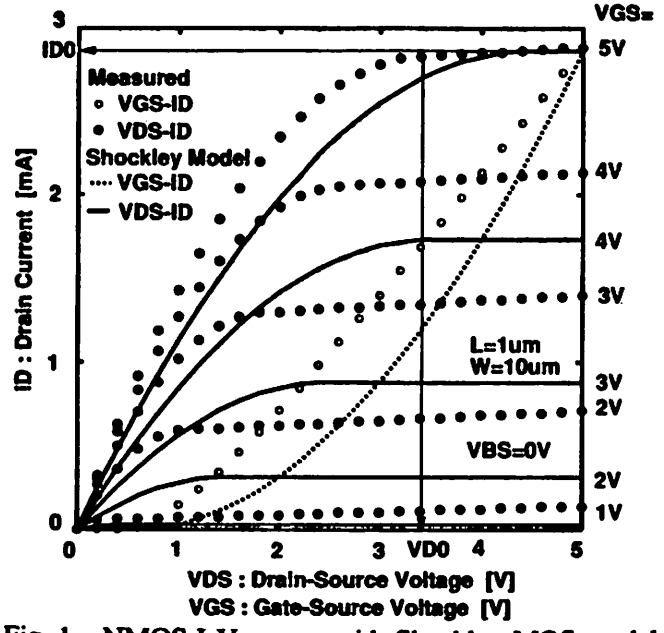


Fig. 1 NMOS I-V curves with Shockley MOS model

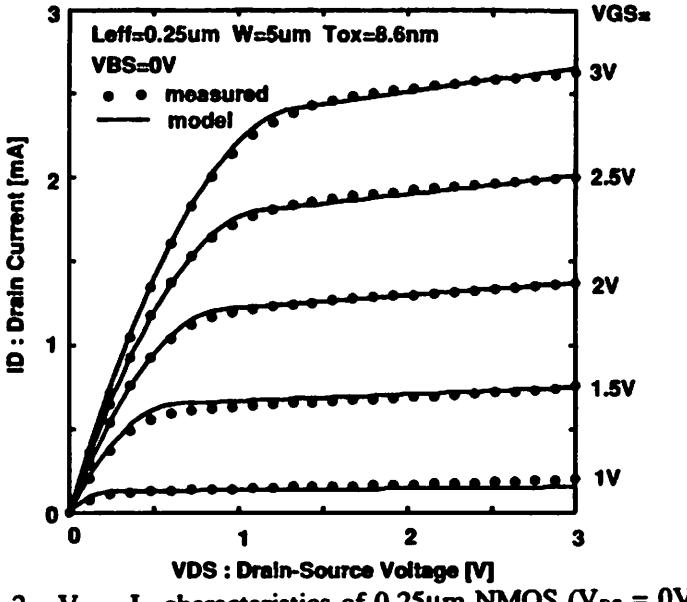


Fig. 2 $V_{DS} - I_D$ characteristics of 0.25μm NMOS ($V_{BS}$ = 0V)

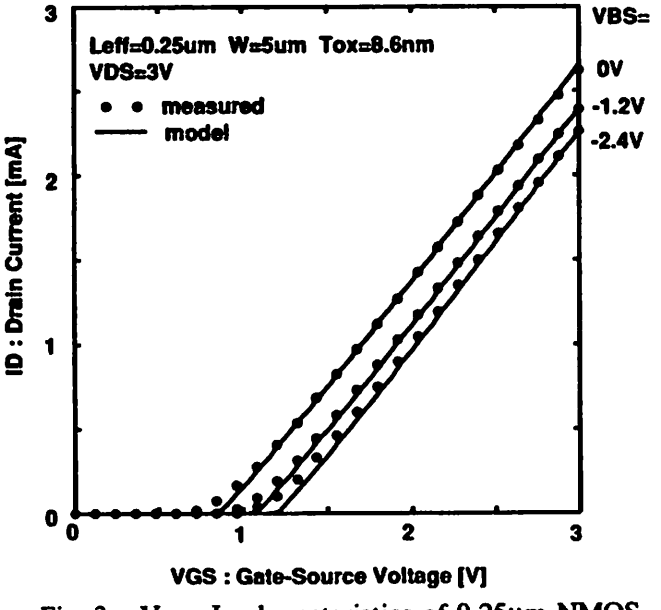


Fig. 3 $V_{GS} - I_D$ characteristics of 0.25μm NMOS

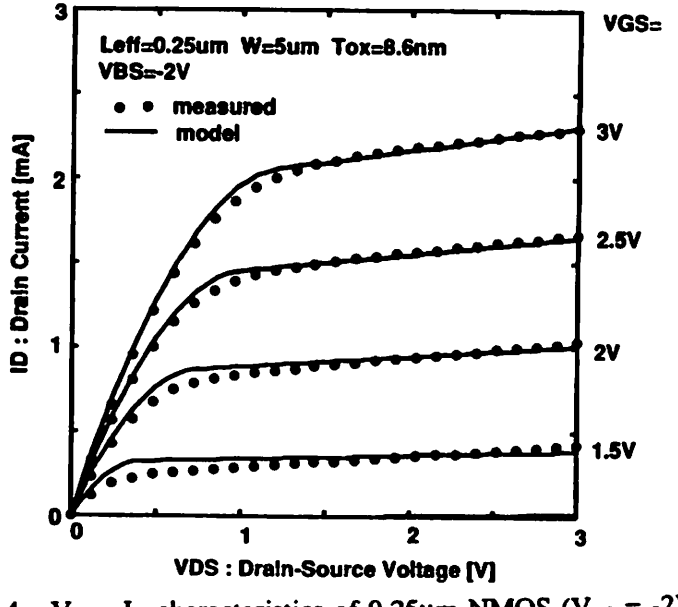


Fig. 4 $V_{DS} - I_D$ characteristics of 0.25μm NMOS ($V_{BS}$ = -2V)

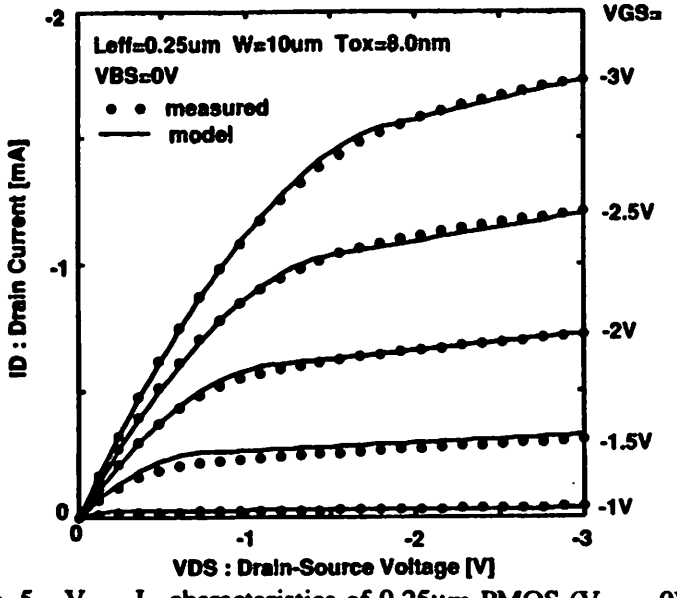


Fig. 5 $V_{DS}-I_D$ characteristics of 0.25μm PMOS ($V_{BS}$ = 0V)

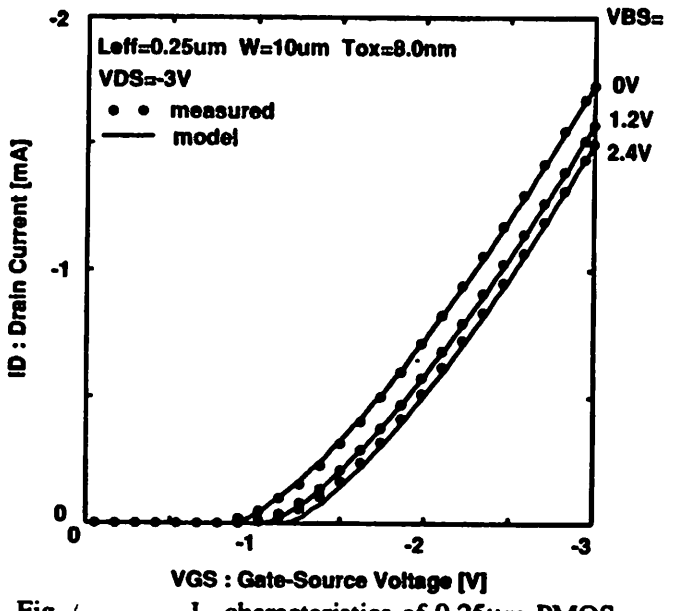


Fig. 6 $V_{DS}-I_D$ characteristics of 0.25μm PMOS

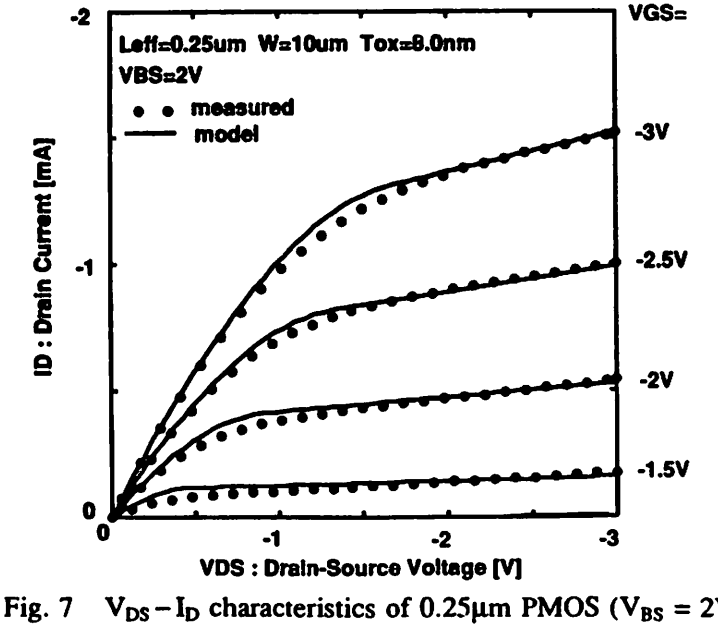


Fig. 7 $V_{DS}-I_D$ characteristics of 0.25μm PMOS ($V_{BS}$ = 2V)

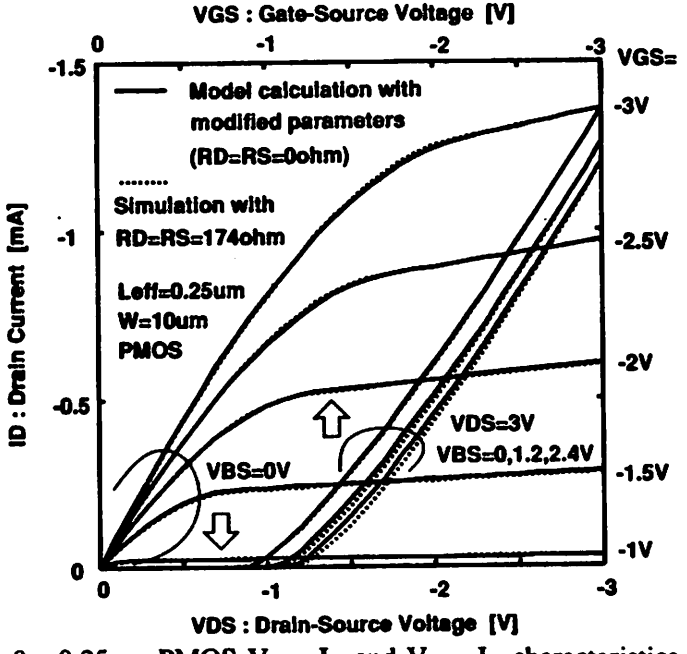


Fig. 8 0.25μm PMOS $V_{DS}-I_D$ and $V_{GS}-I_D$ characteristics with and without Source and Drain Resistance

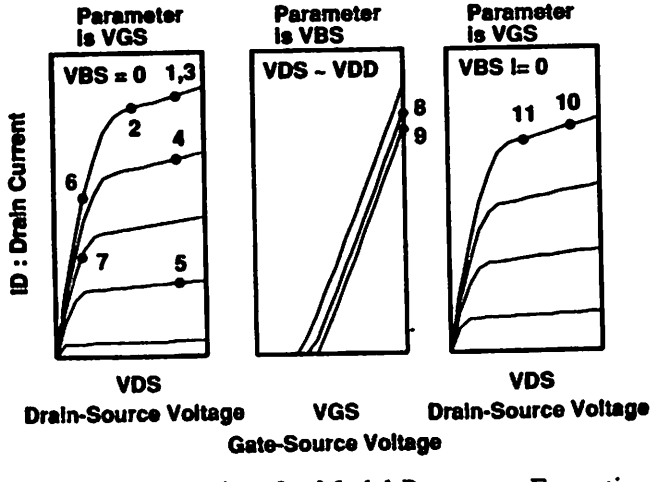


Fig. 10 Selected Points for Model Parameter Extraction

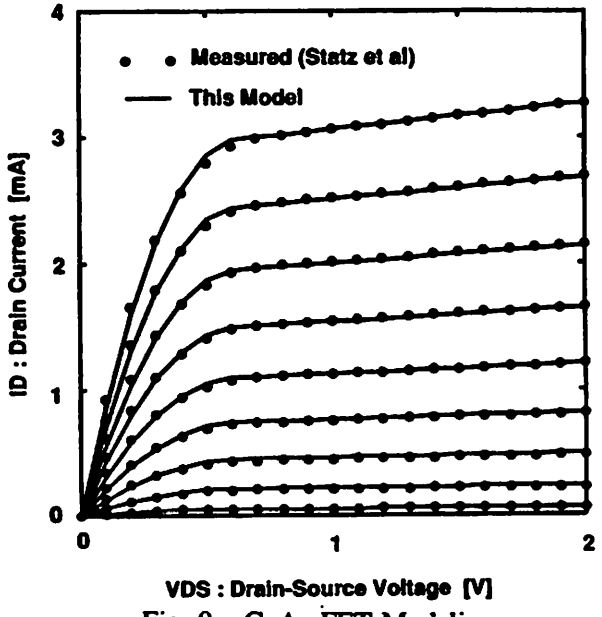


Fig. 9 GaAs FET Modeling

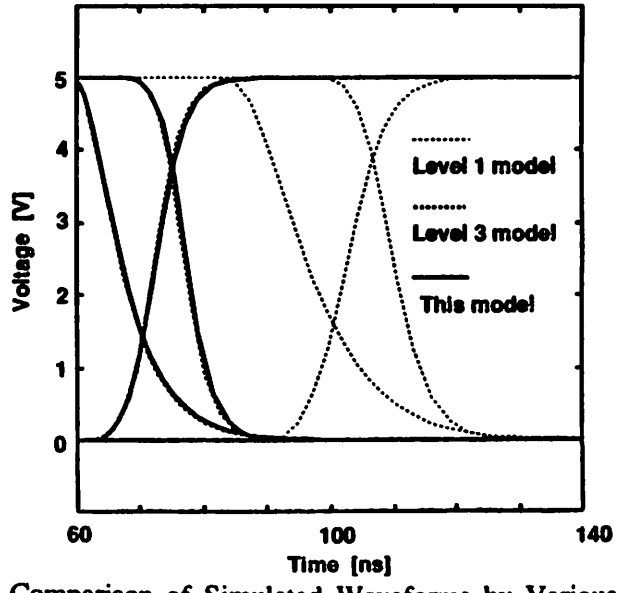


Fig. 11 Comparison of Simulated Waveforms by Various MOS Models

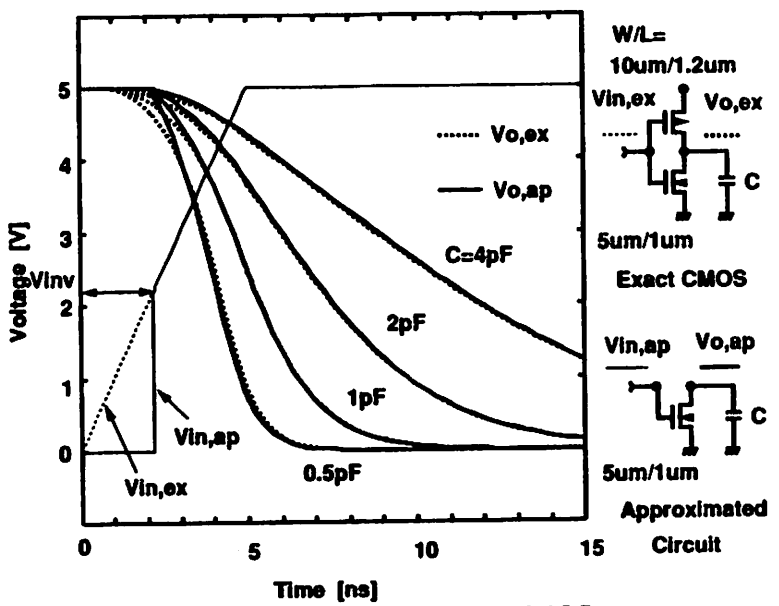


Fig. 12 Approximating CMOS by NMOS

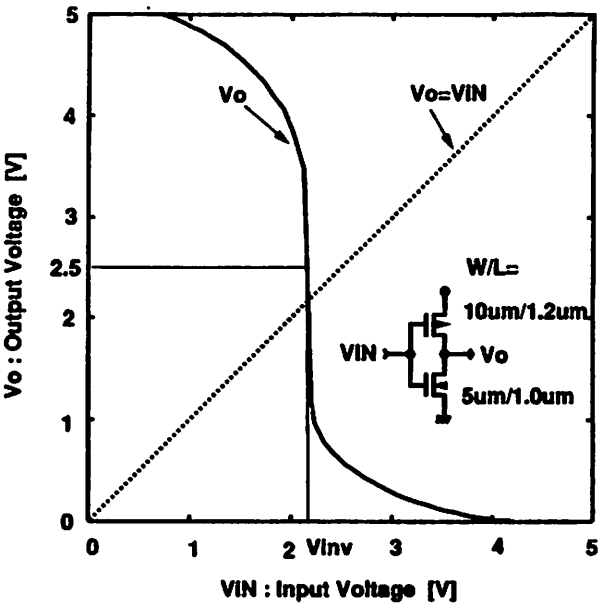


Fig. 13 Logic Threshold Voltage

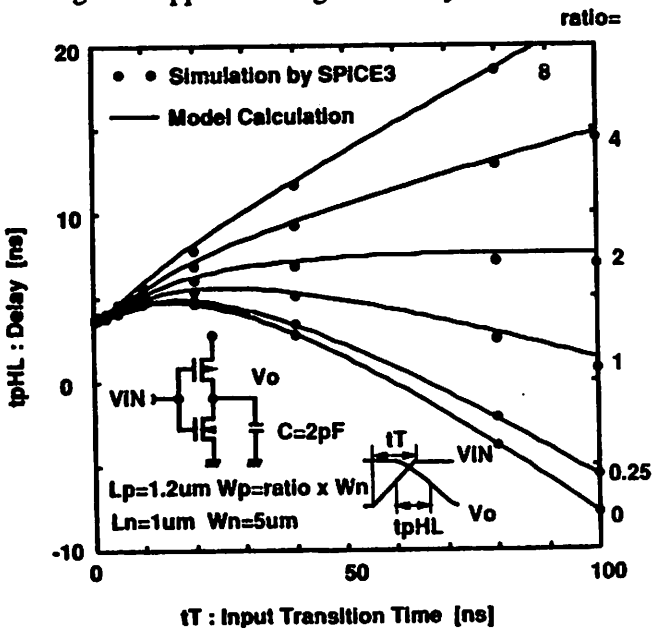


Fig. 14 Delay Dependence on Input Transition Time and Wp / Wn Ratio

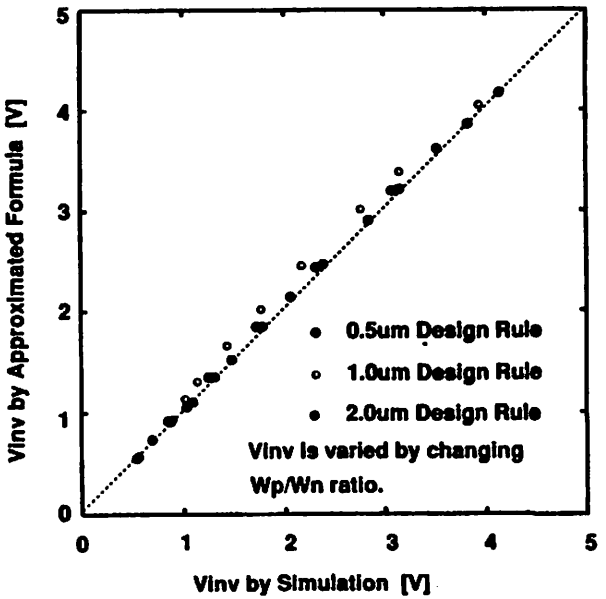


Fig. 15 Comparison of Approximated Formula and Simulation for CMOS Inverter Logic Threshold Voltage $V_{INV}$

# CMOS Gate Delay Analysis in the Submicron Region and its Application to Series-Connected MOSFETs

**Takayasu Sakurai* and A. Richard Newton**

**Department of Electrical Engineering and Computer Sciences**
**University of California, Berkeley, CA94720, U.S.A.**
***) On leave from Semiconductor Device Engineering Lab.,**
**Toshiba Corporation, Kawasaki, 210, Japan**

## Abstract

The CMOS gate delay is analyzed using a new realistic short-channel MOS model. Closed-form delay formulas are obtained which includes short-channel effects and CMOS effects. The analysis is extended to series-connected MOSFET structures (SCMS). It is shown that the ratio (delay of NAND/NOR) / (delay of inverter) becomes smaller in the submicron region. This is because the $V_{DS}$ and $V_{GS}$ of each MOSFET in the SCMS are smaller than those of an inverter MOSFET. The smaller voltages in turn mitigate and relax the severe carrier velocity saturation in miniaturized MOSFET's. This result encourages more extensive use of NAND/NOR/complex gates, cascode voltage switch logic[6] and hot-carrier resistant logic[1] in submicron circuit design. The result also prompts re-examination of circuit design/optimization in the submicron region. The delay dependence of input terminal position for SCMS structures are also described.

# 1. Introduction

The series-connected MOSFET structure (SCMS) appears in NAND/NOR gates, complex gates, PLAs, and cascode voltage switch logic[6] and is widely used in VLSI designs. However, little has been known about the behavior of the SCMS because of its relatively complicated nature and the non-linearity of MOSFETs. The naive understanding is that N connected MOSFET shows N times as long a delay as a single MOSFET. Is this correct? Main purpose of this paper is to shed light on the behavior of the SCMS and give an answer to this question. It is shown that the ratio (delay of NAND/NOR) / (delay of inverter) becomes smaller in the submicron region. There are cases where N series connected MOSFET's shows only N/2 times as long a delay as a single MOSFET.

In order to analyze the CMOS gate in the submicron region, a realistic yet simple MOS model is required. So the other objective of this paper is to give a simple short-channel MOS model suitable for analytical treatment and using the model to obtain closed form delay formulas for CMOS gate delay.

In section 2 and 3, the description of a simple, yet realistic, short-channel MOS model, delay expressions suitable for analyzing the SCMS are derived. The derived delay expression is then applied to a logic circuit and the limitations of RC-based models in the submicron region are identified in section 4. Section 5 describes the delay ratio of SCMS and an inverter for a simple case, and in the following chapter the more complex case is presented. In section 7, delay dependence on input terminal position is described. The final section is dedicated for conclusions.

## 2. A Short-Channel MOS Model

The Shockley MOS model has been used extensively for the analytical treatment of MOS circuit behavior. However, the model is not capable of reproducing short-channel I-V curves of the form shown in Fig.1.

There are two main discrepancies, both of which are induced by velocity saturation effects in the short-channel MOSFET. First, the measured $I_D - V_{GS}$ characteristic of a short-channel MOSFET is approximately linear while the Shockley model predicts a square-law dependence[3,4,11]. Second, the drain saturation voltage $V_{DSAT}$ is not $(V_{GS} - V_{TH})$ but is lowered in short-channel region[3]. Considering these issues, the following model is proposed and is used in this paper as a short-channel MOS model.

$$V_{TH} = V_{TO} - \gamma_1 V_{BS} \tag{1}$$

$$V_{DSAT} = K (V_{GS} - V_{TH})^{m} \tag{2}$$

$(V_{DS} \geq V_{DSAT}$ : saturated region)

$$I_D = I_{DSAT} = \frac{W}{L_{EFF}} B (V_{GS} - V_{TH})^{n} \tag{3}$$

$(V_{DS} < V_{DSAT}$ : linear region)

$$I_D = I_{DSAT}\left(2 - \frac{V_{DS}}{V_{DSAT}}\right)\frac{V_{DS}}{V_{DSAT}} \tag{4}$$

This simple model can reproduce the measured characteristics even in short-channel region, as shown in Fig.2. Comparisons have been made with a wide variety of short-channel devices. The model reduces to the Shockley model for K = 1, m = 1, B = 1/2•Kp, and n = 2. The body-effect is

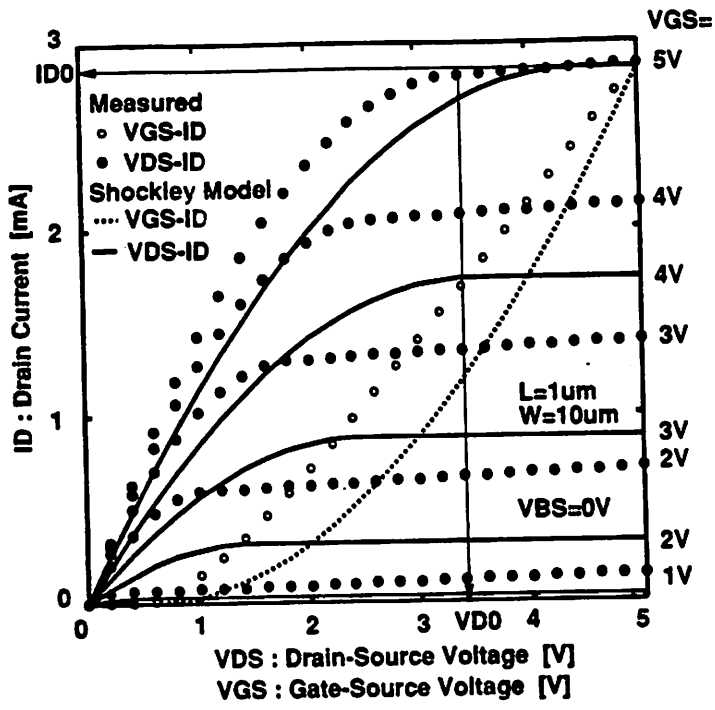


Fig.1 NMOS I-V curves with Shockley MOS model

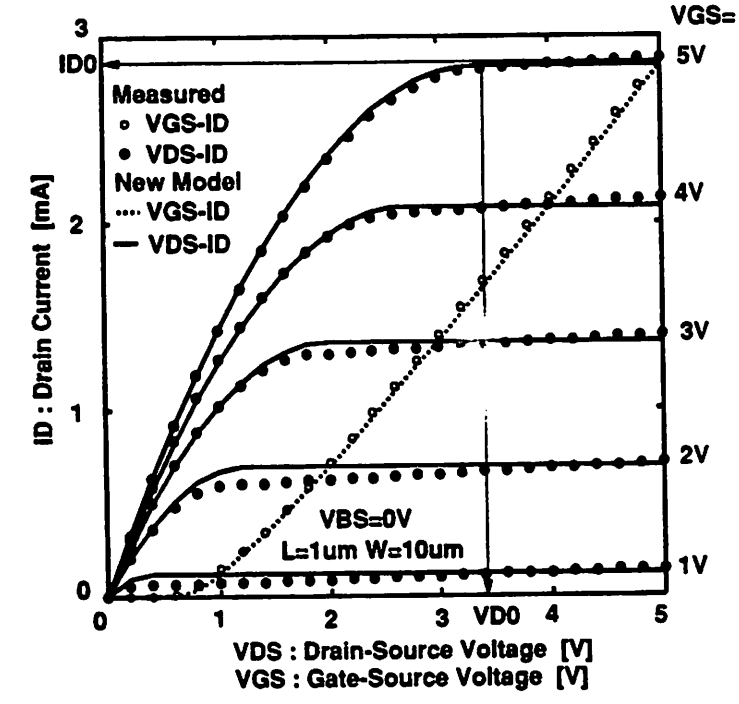


Fig.2 NMOS I-V curves with the new MOS model

approximated by a linear form, the meaning of which is illustrated in Fig.3. The back-gate bias is normally less than 2.5V in analyzing the SCMS.

In Figs. 1, 2, $I_{D0}$ is defined as the drain current observed when $V_{GS} = V_{DS} = V_{DD}$ and is good index of drivability of a MOSFET. $V_{D0}$ is defined as the drain saturation voltage when $V_{GS} = V_{DD}$. These two quantities together with the velocity saturation index n play an essential role in determining circuit behavior.

## 3. A Delay Expression for CMOS Gates

Using the above model, delay formulae for a CMOS inverter can now be derived. These formulae are also effective in analyzing the SCMS because the MOS model is general enough to express the equivalent I-V characteristics of the SCMS as shown in Fig.4.

The derivation begins by setting up the differential equation which governs gate operation[5]. This equation is then solved for the very fast input case and for the very slow input case and the two solutions are connected smoothly. The ramp input waveform is approximated by Vin,ap in Fig.5 using the logic threshold $V_{INV}$. The detailed derivation can be found in Appendix A. First, define a critical input transition time $t_{T0}$.

$$t_{T0} = \frac{C_O V_{DD}}{2I_{D0}} \frac{(n+1)(1-v_T)^n}{(1-v_T)^{n+1} - (v_V - v_T)^{n+1}}, \tag{5}$$

where $v_T = V_{T0}/V_{DD}$ and $v_V = V_{INV}/V_{DD}$. Then the delay $t_d$, the delay from $0.5V_{DD}$ of input to $0.5V_{DD}$ of output, and the effective output transition time, $t_{TOUT}$, can be expressed as follows. In calculating $t_{TOUT}$, the output waveform slope is approximated by 70% of its derivative at the half $V_{DD}$ point[7]. $t_{TOUT}$ can be used as $t_T$ for the next logic gate.

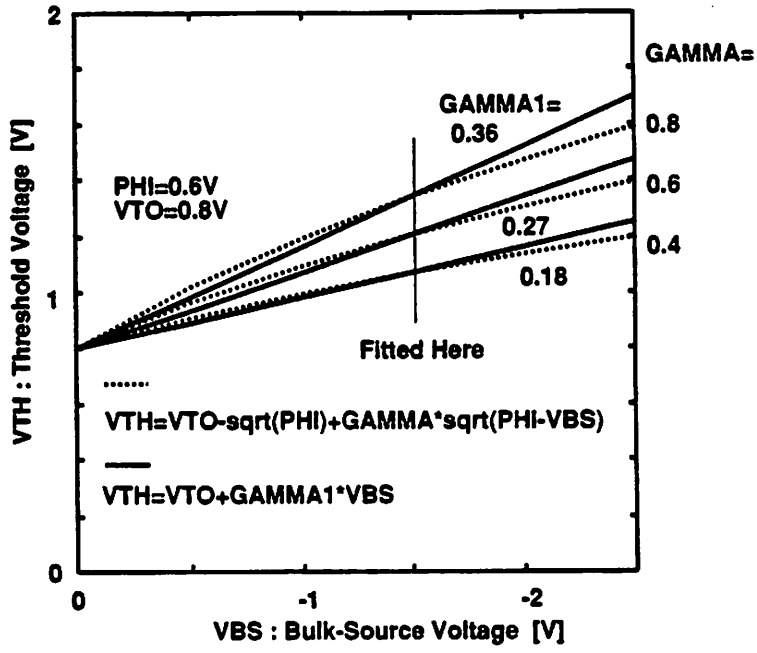


Fig.3 Linear approximation of body-effect

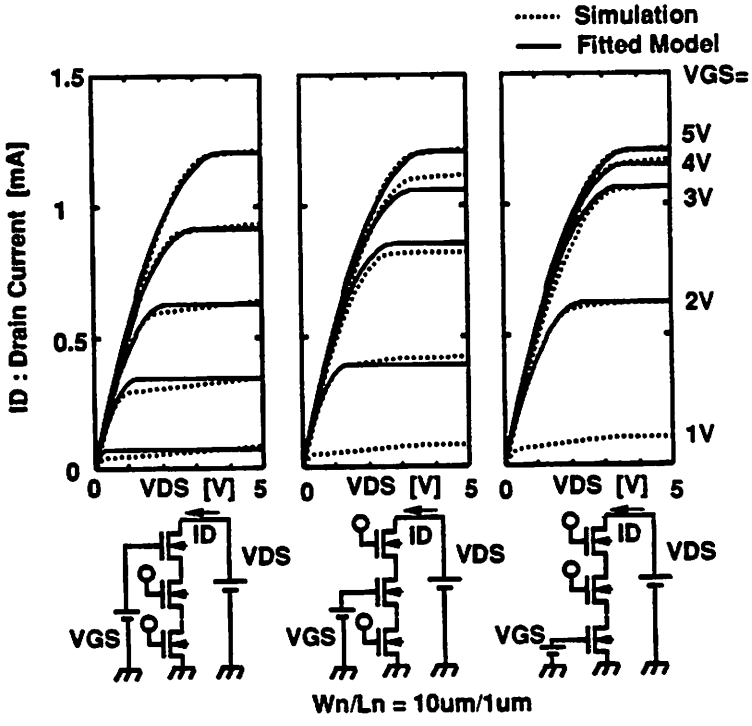


Fig.4 I-V curves of series-connected MOSFET's

($t_T \leq t_{T0}$ : for the faster input)

$$t_d = t_T \left\{ \frac{1}{2} - \frac{1-v_T}{n+1} + \frac{(v_V - v_T)^{n+1}}{(n+1)(1-v_T)^n} \right\} + \frac{1}{2} \frac{C_O V_{DD}}{I_{D0}} \tag{6}$$

$$t_{TOUT} = \frac{C_O V_{DD}}{0.7 I_{D0}} \frac{4 v_{D0}^2}{(4 v_{D0} - 1)} \tag{7}$$

($t_T > t_{T0}$ : for the slower input)

$$td = t_T \left[ v_T - \frac{1}{2} + \left\{ (v_V - v_T)^{n+1} + \frac{(n+1)(1-v_T)^n}{2 t_T I_{D0} / C_O V_{DD}} \right\}^{\frac{1}{n+1}} \right] \tag{8}$$

$$t_{TOUT} = \frac{C_O V_{DD}}{0.7 I_{D0}} \left( \frac{1-v_T}{t_d/t_T + 1/2 - v_T} \right)^n , \tag{9}$$

where $C_O$ is an output capacitance and $v_{D0} = V_{D0} / V_{DD}$.

## 4. Application of the Delay Expression and Assessment of an RC model

To apply the above-mentioned formulae to a circuit of the form of Fig.6, quantities including effective $I_{D0}$, n and $V_{D0}$, are required for the N series-connected MOSFET structure. One way of obtaining these values is by extracting them by fitting models to all possible compound I-V curves, as in Fig.4.

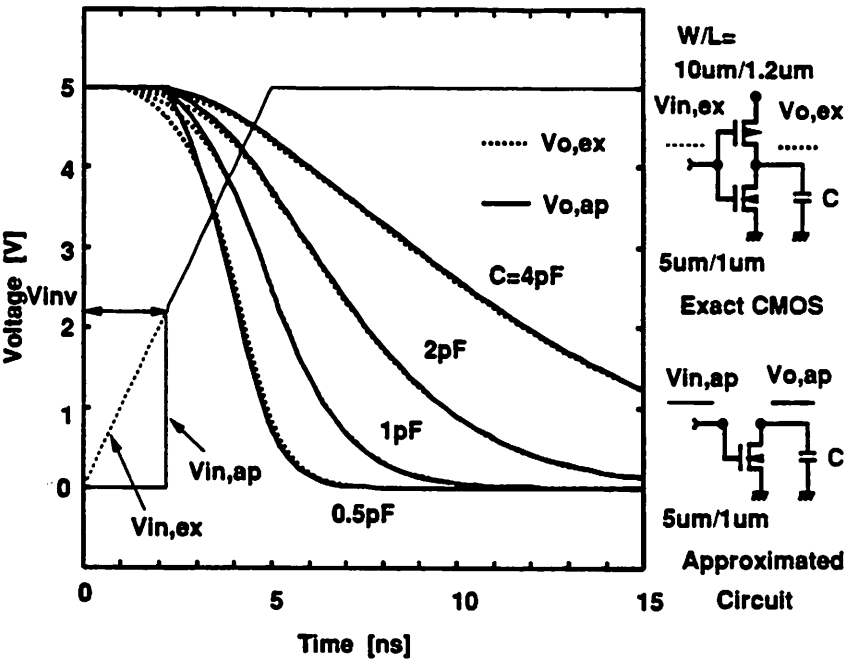


Fig.5 Approximating CMOS by NMOS

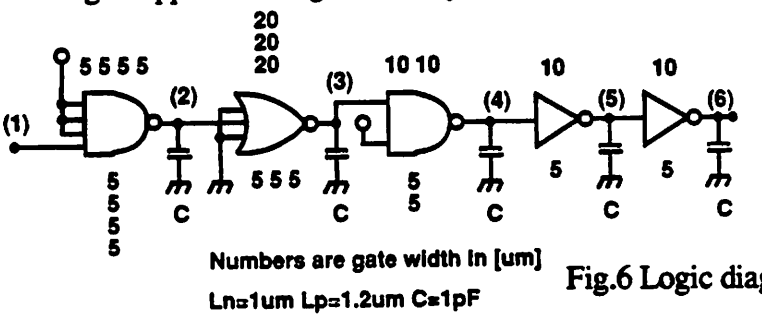


Fig.6 Logic diagram of example circuit

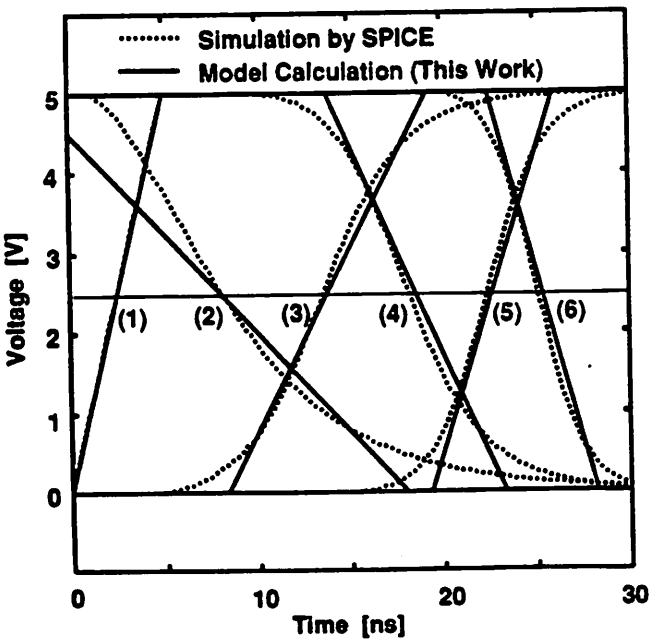


Fig.7 Waveform comparison between simulation and model calculation

A more approximate approach is also possible but since the description is lengthy (see Appendix C for full description) only the results are shown here. In Fig.7, the calculated waveforms using the approximate approach are compared with simulated waveforms from SPICE[8] for the circuit given in Fig.6.

Fig.8 shows a delay comparison for the same circuit, as well as the dual-slope RC model calculation of Crystal[9]. Such RC models of MOSFET networks[2] are used extensively in MOS timing simulators[9] and timing verifiers (the input slope effect is often included in such models). As seen from the figure, in long-channel region, the RC model predicts the delay fairly well. However, in the short-channel region, the RC model prediction is particularly poor for SCMS's such as NAND and NOR gates.

## 5. Step Input with Large Output Capacitance

For the simple case where the input is a step waveform and $C_O$ is large, a straightforward analytical treatment of the SCMS delay is possible. Such an analysis is helpful in understanding the essence of SCMS operation in the submicron region. A waveform example for this simple case is shown in Fig.9, where the SCMS delay normalized by an inverter delay becomes smaller as the channel gets shorter. Overall output waveforms are similar for both long and short channel MOS models; only the time constant is different.

To begin the analysis, let $I_{D0N}$ represent the effective $I_{D0}$ for N series-connected MOSFET's. When the output capacitance $C_O$ dominates the logic gate capacitance itself and $t_T = 0$, then $t_d$ is inversely proportional to $I_{D0N}$ as can be seen from Eqn.(6). In consequence, the *delay degradation factor* FD, defined as (delay of a SCMS) / (delay of a single MOSFET) can be calculated as $I_{D0}/I_{D0N}$.

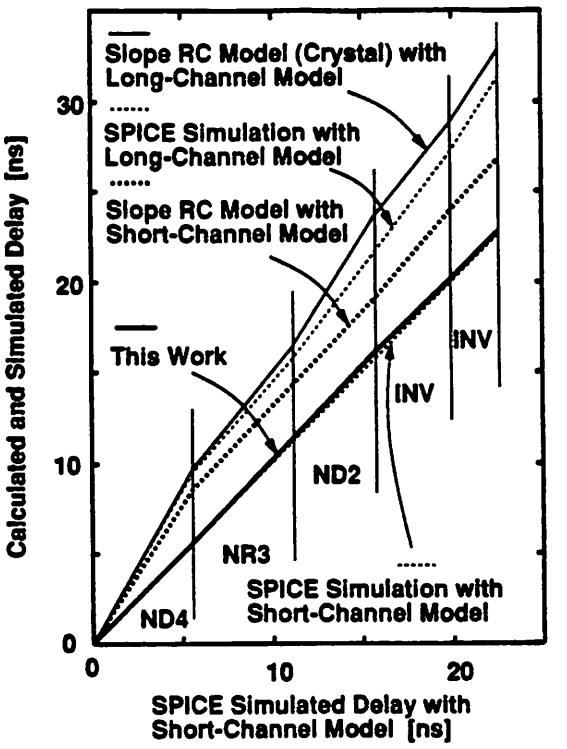


Fig.8 Comparison of calculated delay

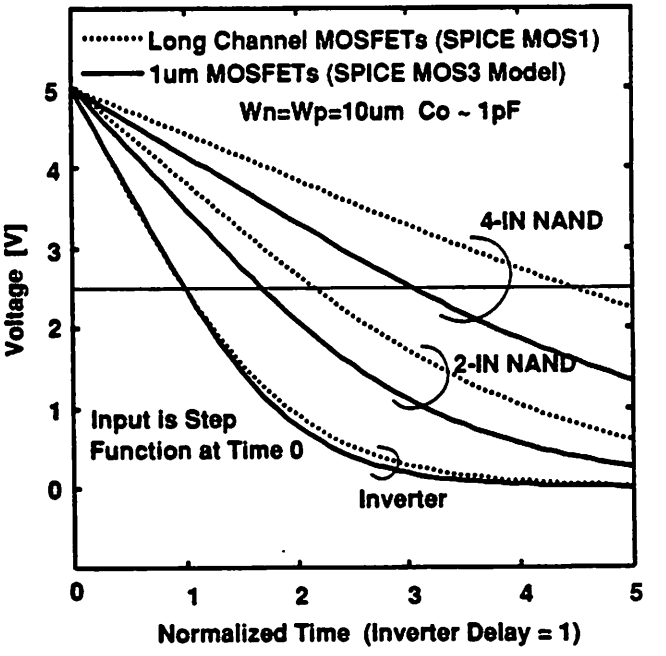


Fig.9 Waveforms of NAND gates

The expression for FD is obtained by using a linear approximation (lines LL and LU) of the actual MOS current curves (L and U respectively), as illustrated in Fig.10. It is clear by inspection that FD improves (becomes smaller) as channel length is reduced. That is, in shorter channel MOSFET's, both $V_{D0}$ and n decrease which leads to the larger $I_{D02}$. It is also worth noting that as n tends to 0 for the "ultimate" short-channel MOSFET, $I_{D0N}$ tends to $I_{D0}$ and hence FD approaches unity.

Although only the N = 2 case is described in this figure, the N-connected MOSFET case can be treated similarly by reducing the value of LL to LL / (N−1), since (N−1) of the MOSFET's are operating in the linear region and so can be added linearly. The following expression for FD can thus be derived (the detailed derivation is found in Appendix B):

$$FD = \frac{\text{delay(SCMS)}}{\text{delay(inverter)}} = 1 + \frac{1 - 2^{-1/2}}{1 - 2^{-1/n}} \frac{v_{D0}}{1-v_T} (1 + \gamma_1) (N-1) \qquad (10)$$

$$\approx 1 + \frac{1}{2} n \frac{v_{D0}}{1-v_T} (1 + \gamma_1) (N-1) \qquad (11)$$

An RC model would predict N series-connected MOSFET's would show approximately N times the delay compared of a single MOSFET, when $C_O$ is dominant. This is accurate for ideally long MOSFET's without body-effect, where n = 2, $v_{D0} = 1-v_T$, and $\gamma_1=0$, because FD becomes exactly equal to N. However, for shorter MOSFETs the approximation is far from accurate.

For a long-channel MOSFET without body-effect, the relation FD = N can be shown in a more rigorous way as follows. In this ideal case, the drain current $I_D$ can be decomposed into $f(V_D)-f(V_S)$ [10], where $V_D$ and $V_S$ are the drain and source potential respectively, and $f(V)=1- V^2$. Using the notation of Fig.11, the following equations hold:

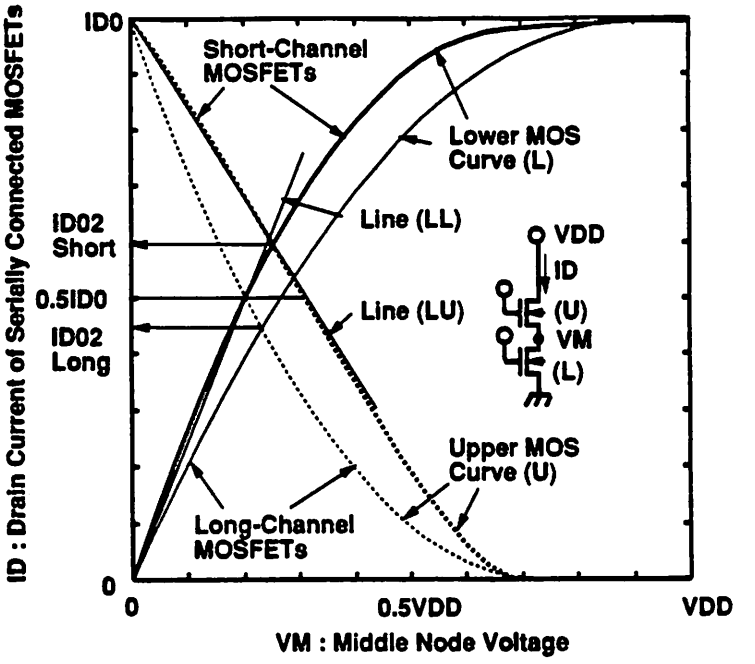


Fig.10 Drain current of SCMS in long-channel device and short-channel device

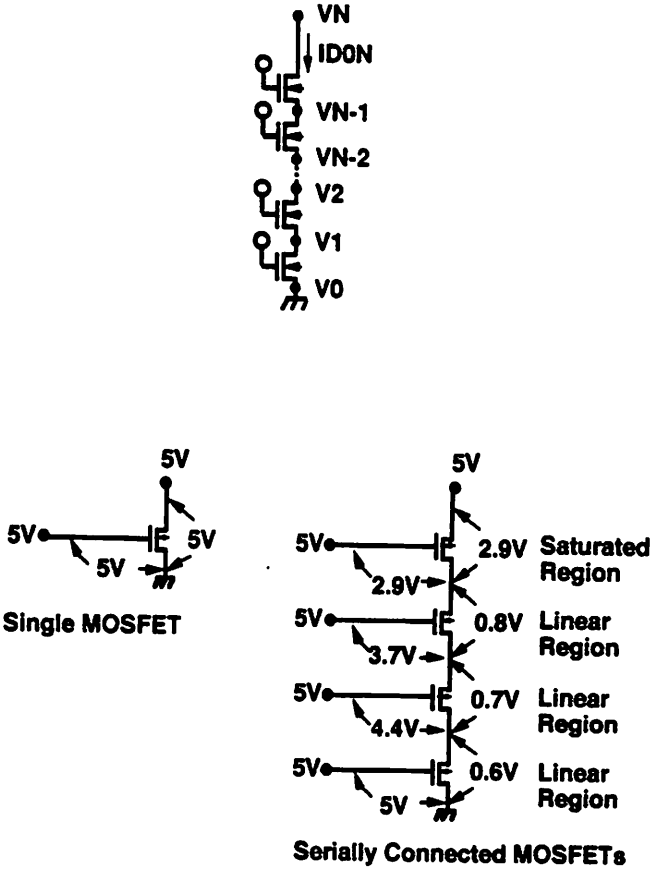


Fig.11 Notations for SCMS

$$I_{D0N} = f(V_N) - f(V_{N-1}) \quad (12)$$

• • •

$$I_{D0N} = f(V_2) - f V_1) \quad (13)$$

$$I_{D0N} = f(V_1) - f(V_0) \quad (14)$$

Summing these equations leads to:

$$N \cdot I_{D0N} = f(V_N) - f(V_0) = I_{D0}\,, \rightarrow \quad F_D = N \quad (15,16)$$

The relation $F_D = N$ is rather surprising considering the non-linear nature of MOSFET's.

In Fig.12, calculated results using Eqn.(10) are compared with simulated data for various generations of MOSFET's and excellent agreement can be seen. The figure clearly shows the improvement of $F_D$ in submicron region. Eqn.\ (11) can be used as a simple index to estimate the delay degradation of the SCMS over a single MOSFET and provide insight into SCMS operation in the submicron region.

## 6. The General Case and Physical Interpretation

In the general case, where $C_O$ may not dominate and the input waveform has a finite slope, the analysis becomes more complex. However, the claim that $F_D$ decreases in submicron region is still true, an example of which is shown in Fig.13. This is because the capacitance ratio of the SCMS and the single MOSFET is basically unchanged even if the feature size is changed, while the current ratio of the SCMS to a single MOSFET is improved in the submicron region.

The physical interpretation of the improvement in the current ratio is as follows. In the SCMS, $V_{DS}$ of each MOSFET is smaller than that of an inverter MOSFET since the output voltage is spread

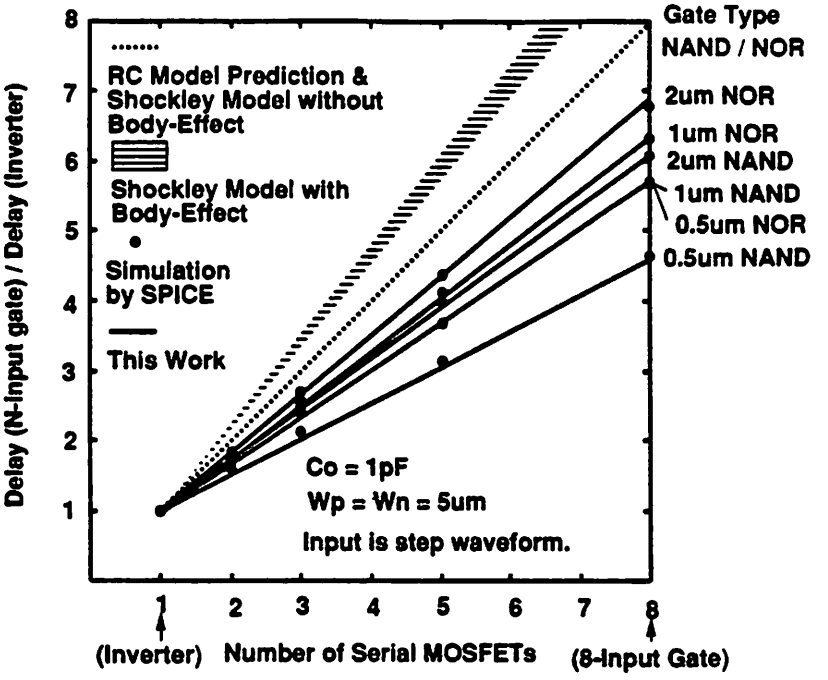


Fig.12 Delay degradation factor of SCMS

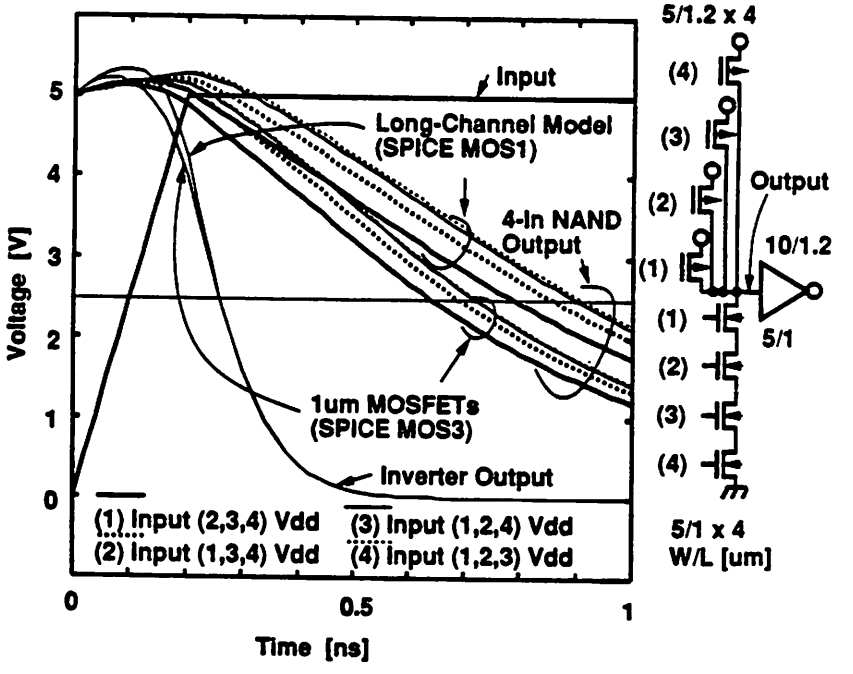


Fig.13 Inverter and NAND gate behavior with long-channel and short-channel MOS model

across multiple MOSFET's (see Fig.11 for an example in case of 5V power supply). $V_{GS}$ of each MOSFET is also smaller because the source voltage is raised from ground (or lowered from $V_{DD}$ in PMOS case).

Because of the reduced $V_{DS}$ and $V_{GS}$, the carriers feel less electric field both parallel to and perpendicular to the channel. Consequently, velocity saturation is mitigated in the SCMS compared with an inverter an a relatively large current flows in the SCMS in the submicron region.

The situation might change because the SCMS but not the inverter suffers from the body-effect. However, as shown in Fig.12, the current improvement induced by the mitigated velocity saturation dominates the current degradation induced by body-effect. Moreover, there are technologies like p-pocket which can suppress the body-effect while suppressing velocity saturation seems to be impossible.

The reduction of the electric field in the SCMS suppresses the hot-electron generation, too. This is a principle of HOt-carrier REsistant Logic family (HOREL)[1,12], where additional normally-on enhancement NMOS is inserted at the top of a NMOS logic tree to reduce VDS. In the submicron region, the HOREL shows good performance because the higher $V_{DD}$ can be applied while maintaining the same reliability and the speed degradation by the inserted MOSFET becomes small.

## 7. Delay Dependence on Input Terminal Position

Which input of a 4-input NAND has the shortest delay to the output? Consider the series NMOS case (i.e. NAND) since the series PMOS case follows from symmetry. When the output capacitance $C_O$ is very large compared with the capacitance of the logic gate itself, the lower (nearer to ground) terminal shows the shorter delay. This is because n becomes smaller for these lower terminals, as shown in Fig.4. This means that the drain current quickly approaches to its final value when changing $V_{GS}$ and enables the faster discharging of the output capacitance.

If the output capacitance $C_O$ is small, there are two cases to consider, depending on the value of $t_T$, as shown in Fig.14. When $t_T$ is large, the lower terminals show faster operation because the logic threshold is lowered and is achieved faster (n is smaller and only small $V_{GS}$ is needed to turn the device on hard). When $t_T$ is small, the lower MOSFET must discharge upper MOSFET's' capacitances so the upper terminal shows a faster delay. These situations are illustrated in Fig.15.

## 8. Conclusion

A new realistic short-channel MOS model is proposed and using the model closed-form delay formulas are obtained which includes short-channel effects and CMOS effects. In these formulas, the logic threshold, the velocity saturation index n, $I_{D0}$, and $V_{D0}$ play an essential role.

The analysis is extended to the SCMS. It has been shown that the ratio (delay of NAND/NOR) / (delay of inverter) becomes smaller in the submicron region. There are cases where N series connected MOSFET's shows only 1/2 times as long a delay as a single MOSFET. This result encourages more extensive use of NAND/NOR/complex gates, PLA's, CVSL[6] and hot-carrier resistant logic[1,12] in submicron circuit design.

The result also suggests the re-examination of the VLSI design/optimization in the submicron region. For example, the logic threshold of a NAND gate becomes much lower than $0.5V_{DD}$ in the submicron region, if Wp / Wn ratio is chosen the same as in a long-channel MOSFET generation. It has also been shown that the accuracy of an RC-based model is deteriorated in the submicron region.

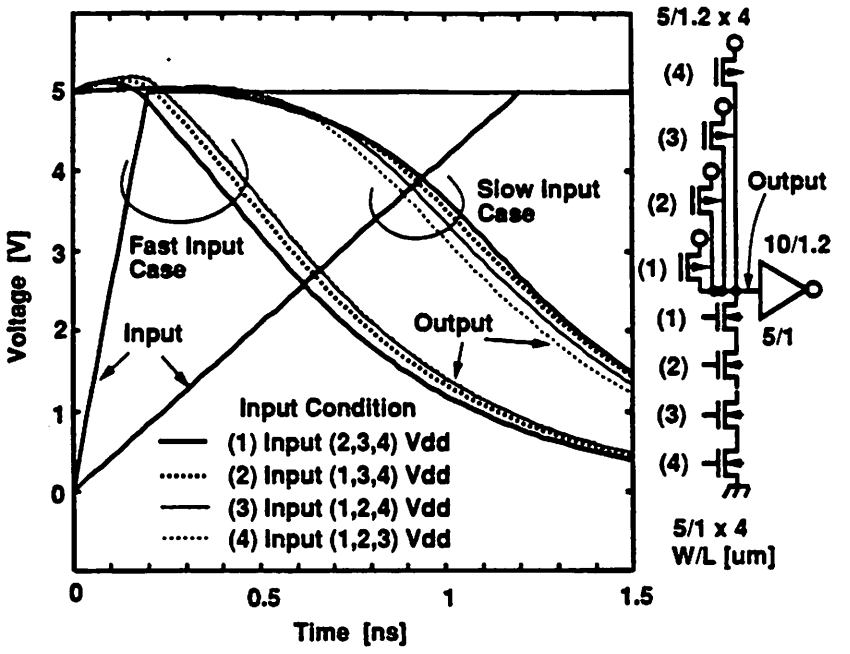


Fig.14 Delay comparison among various input terminals of 4-input NAND gate

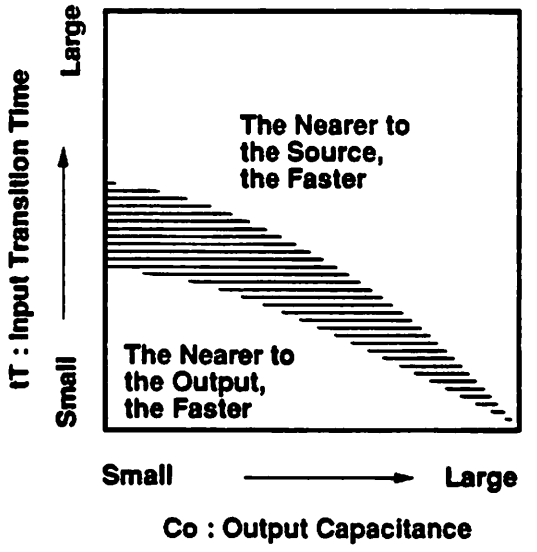


Fig.15 Delay dependence on input terminal in NAND/NOR/complex gates

## Acknowledgments

The encouragement of R. Brayton, A. Sangiovanni-Vincentelli, Y. Unno, Y. Takeishi, Y. Fukuda, H. Yamada and T. Iizuka throughout the course of this work is appreciated. Discussions with H. Ishiuchi and T. Fujii on MOS physics and circuit designs were inspiring and should be acknowledged. Assistance provided by T. Quarles and R. Spickelmier concerning SPICE and computer environments is also appreciated. This work was supported by a grant from Toshiba Corporation.

## Appendix A : Derivation of the Delay Expression

In most practical cases, the channel-length modulation effect is small because MOSFET's are usually engineered in such a way that the channel-length modulation is minimized. However, in some cases the channel-length modulation is eminent, the effect is included in the Appendix through the modification of Eqs.3,4 as follows.

$$I_D = I_{D5} = I_{DSAT}(1 + \lambda V_{DS}) \quad (V_{DS} \geq V_{DSAT} : \text{saturated region}) \tag{A1}$$

$$I_D = I_{D3} = I_{D5}\left(2 - \frac{V_{DS}}{V_{DSAT}}\right)\frac{V_{DS}}{V_{DSAT}} \quad (V_{DS} < V_{DSAT} : \text{linear region}), \tag{A2}$$

where $\lambda$ is a widely-used channel-length modulation parameter.

In the Appendix, the discharging of an output capacitance through NMOS's is explained since the discussion for the charging by PMOS's is just symmetric. As seen from Fig.5, a CMOS inverter with a ramp input can be approximated by an NMOS circuit with an input waveform like Vin,ap.

Vin,ap is the same as the real ramp input except that it remains zero until the input reaches the logic threshold voltage.

For the extreme cases, this approximation is exact. That is, for the ultimately fast input case, the ramp input becomes a step function and Vin,ap also becomes the step function and the current through PMOS can be completely neglected. For the extremely slow input, the output changes abruptly and comes down to $0.5V_{DD}$ when the input goes across the logic threshold. The approximated circuit shows the same delay. The intermediate case is shown in Fig.5 and This approximation greatly reduces the complexity of the system make it possible to treat the CMOS inverter delay analytically.

The key strategy for solving the differential equation which governs the discharging process is to solve it for the very fast input case and for the very slow input case separately as is mentioned in the text. The two solutions for the two extreme cases happen to be connected smoothly.

In the following, voltages are normalized by $V_{DD}$, currents by $I_{D0}$, and time by $\tau = C_O V_{DD}/I_{D0}$. The normalized voltage is denoted as v instead of V, the normalized current *i* instead of I, and the normalized time t' instead of t. First, consider the very fast input case as shown in Fig.A1 (see this figure for notations below). There are three regions: Region 1, the time before input reaches $V_{DD}$, Region 2, the time before output reaches $V_{D0}$, and Region 3, the time after output reached $V_{D0}$.

In Region 1, the differential equation which governs the discharging process can be written as

$$\frac{dv_O}{dt'} = -i_5 = -\left(\frac{t'/t'_T - v_T}{1 - v_T}\right)^n \frac{1 + \lambda' v_O}{1 + \lambda'}, \qquad \text{(A1)}$$

which should be solved with the initial condition of $v_O = 1$ at $t' = t'_V$. The solution is

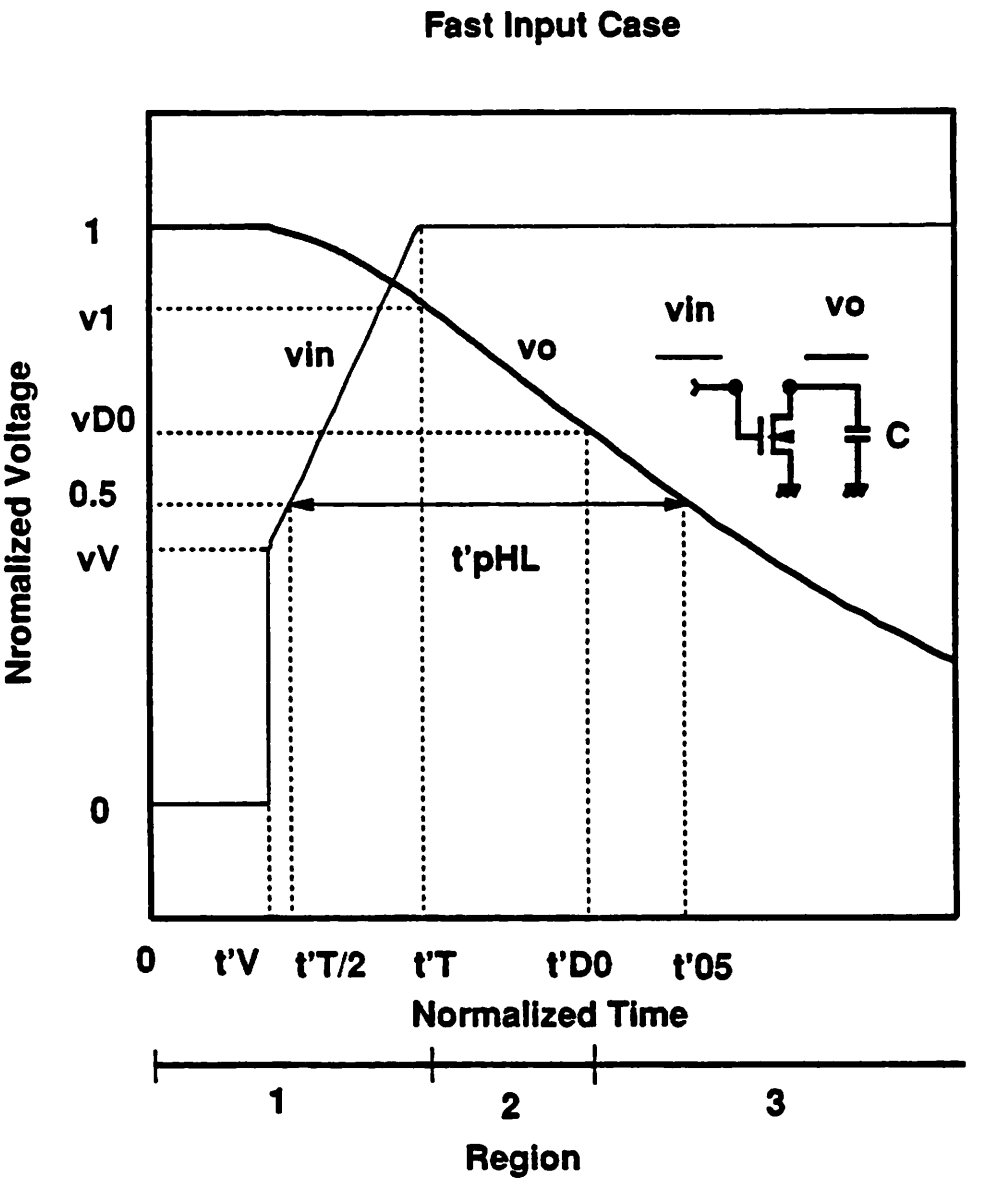


Fig.A1 Input/output waveforms of fast input case

$$\frac{1+\lambda'}{\lambda'} ln\frac{1+\lambda'v_O}{1+\lambda'} = \frac{t'_T}{(1-v_T)^n(n+1)}\left\{\left(\frac{t'}{t'_T} - v_T\right)^{n+1} - (v_V - v_T)^{n+1}\right\} \quad \text{(A2)}$$

$v_1$ is obtained by substituting t' by $t'_T$.

In Region 2, the differential equation is simple since the input is constant $V_{DD}$.

$$\frac{dv_O}{dt'} = -i_5 = \frac{1+\lambda'v_O}{1+\lambda'}. \quad \text{(A3)}$$

The initial condition is $v_O = v_1$ at t' = t'T and the solution is

$$\frac{1+\lambda'}{\lambda'} ln\frac{1+\lambda'v_O}{1+\lambda'} = -\frac{t'_T}{(1-v_T)^n(n+1)}\left\{\left(\frac{t'}{t'_T} - v_T\right)^{n+1} - (v_V - v_T)^{n+1}\right\} - (t' - t'_T) \quad \text{(A4)}$$

(A3)

$t'_{D0}$ is obtained by letting $v_O$ to $v_{D0}$ and is written as follows.

$$t_d = t_T\left\{1 - \frac{1-v_T}{n+1} + \frac{(v_V-v_T)^{n+1}}{(n+1)(1-v_T)^n}\right\} \frac{1+\lambda'}{\lambda'} ln\frac{1+\lambda'v_{D0}}{1+\lambda'} \quad \text{(A5)}$$

In Region 2, where MOSFET is operating in the linear region,

$$\frac{dv_O}{dt'} = -i_3 = -\left(2 - \frac{v_O}{v_{D0}}\right)\frac{v_O}{v_{D0}}\frac{1+\lambda'v_O}{1+\lambda'}. \quad \text{(A6)}$$

is the differential equation to be solved with the initial condition of $v_O = v_{D0}$ at $t' = t'_{D0}$. The solution is

$$t' = t'_{D0} + (1+\lambda')v_{D0}\left\{\frac{\lambda' v_{D0}}{1+2\lambda' v_{D0}}\, ln\, \frac{1+\lambda' v_O}{1+\lambda' v_{D0}} + \frac{1}{2(1+2\lambda' v_{D0})} ln\left(2-\frac{v_O}{v_{D0}}\right) - \frac{1}{2} ln\, \frac{v_O}{v_{D0}}\right\} \quad (A7)$$

Therefore, the delay $t'_d$ ( $= t'_{pHL}$) can be expressed as follows.

$$t'd = t'\left(v_O = \frac{1}{2}\right) - \frac{t'_T}{2} \approx t'_T\left\{\frac{1}{2} - \frac{1-v_T}{n+1} + \frac{(v_V-v_T)^{n+1}}{(n+1)(1-v_T)^n}\right\} + \frac{1}{2} + \frac{\lambda'}{7} . \quad (A8)$$

To derive this expression, the complicated term of $v_{D0}$ and $\lambda'$is approximated by (1/2 + $\lambda'$/7). The error of this approximation is less than 2% when $0.5 < v_{D0} < 0.7$, $0 <= \lambda' < 0.25$, and less than 4% when $0.4 < v_{D0} < 0.8$, $0 <= \lambda' < 0.4$, as shown in Fig.A3.

The transition time of the output waveform, $t'_{TOUT}$ is calculated as

$$t'_{TOUT} = \frac{1}{0.7}\left.\frac{dt'}{dv_O}\right]_{v_O=0.5} = \frac{8v_{D0}^2\,(1+\lambda')}{(4v_{D0}-1)(2+\lambda')} \quad (A9)$$

When the input is very slow, the output crosses $0.5V_{DD}$ in Region 1 as shown in Fig.A2. In this case, using Eq. A4, the delay $t'_d$ ($= t'_{05} - 1/2\, t_T$) is expressed as:

$$t'_d = t'_T\left[v_T - \frac{1}{2} + \left\{(v_V-v_T)^{n+1} + \frac{(1-v_T)^n(n+1)}{t'_T}\,\frac{1+\lambda'}{\lambda'}\, ln\, \frac{2+\lambda'}{2+2\lambda'}\right\}^{\frac{1}{n+1}}\right] \quad (A10)$$

$$\approx t'_T\left[v_T - \frac{1}{2} + \left\{(v_V-v_T)^{n+1} + \frac{(1-v_T)^n(n+1)}{t'_T}\,\frac{1+\lambda'}{\lambda'}\left(\frac{1}{2} + \frac{\lambda'}{7}\right)\right\}^{\frac{1}{n+1}}\right] \quad (A11)$$

The error of the approximated formulae is less than 2% when $0 <= \lambda' < 0.25$, and less than 4% when $0 <= \lambda' < 0.4$. $t'_{TOUT}$ is calculated as

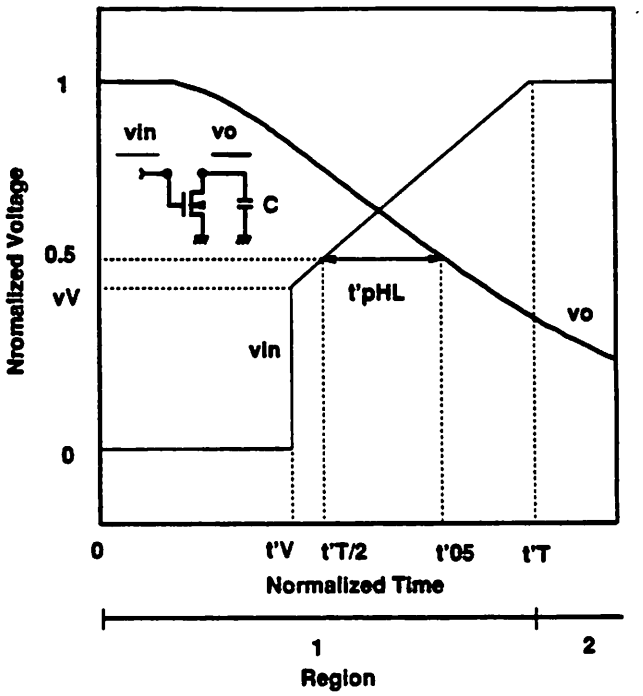


Fig.A2 Input/output waveforms of slow input case

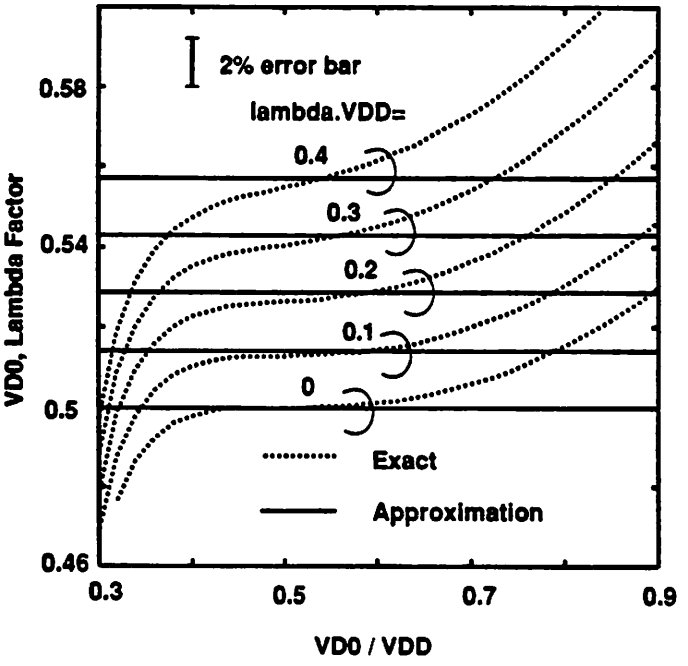


Fig.A3 Approximation of $V_{D0}$ and $\lambda$ term

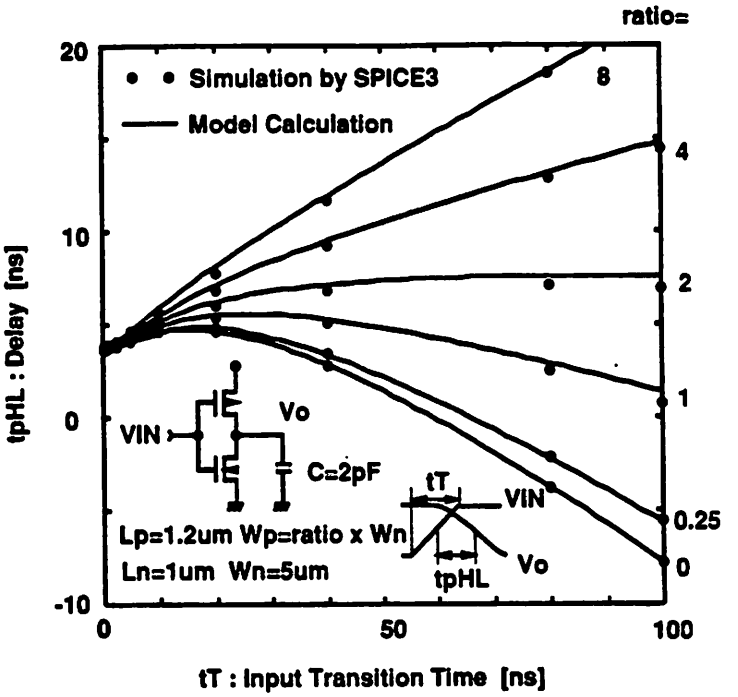


Fig.A4 Comparison of simulated and calculated delay for a CMOS inverter with various $t_T$ and Wp/Wn

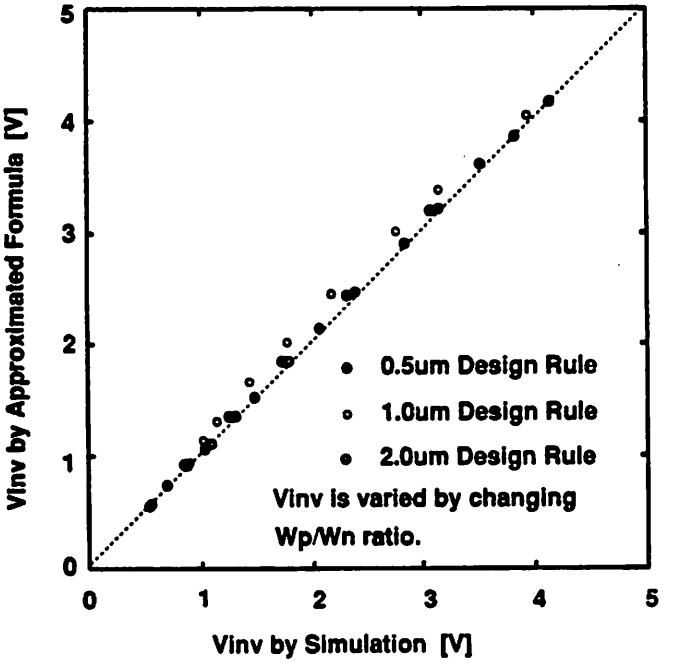


Fig.A5 Approximation of $V_{INV}$

$$t'_{TOUT} = \frac{1}{0.7}\left.\frac{dt'}{dv_O}\right]_{v_O=0.5\&t'=t'_{05}=t'_d+t'_T/2} = \left(\frac{1-v_T}{t'_d/t'_T+1/2-v_T}\right)^n \frac{2+2\lambda'}{2+\lambda'} \qquad \text{(A12)}$$

The solution for the fast input case Eq. A8 and that for the very slow input case Eq. A12 can be connected at the critical input transition time $t'_{T0}$ given below. The first derivative of the both equations also coincide at the critical time.

$$t_{T0} = \frac{1}{2}\,\frac{(n+1)(1-v_T)^n}{(1-v_T)^{n+1} - (v_V-v_T)^{n+1}}, \qquad \text{(A13)}$$

Equations A13, A8, A9, A11 and A12 correspond to Eqs. 5, 6, 7, 8 and 9 of the text, respectively. An example of a calculation using these formulas are demonstrated in Fig.A4. The formulas are valid in wide range of $t_T$ and the channel-width ratio of PMOS and NMOS. The logic threshold voltage, $V_{INV}$, was calculated by the following expression.

$$v_V = \frac{V_{INV}}{V_{DD}} = \frac{I_{D0N}^{1/n}\, v_{TN} + I_{D0P}^{1/n}(1-v_{TN})}{I_{D0N}^{1/n} + I_{D0P}^{1/n}(1-v_{TN}) / (1-v_{TP})}, \quad n = \frac{n_n + n_p}{2}, \qquad \text{(A14)}$$

where the subindex N and P denote NMOS and PMOS, respectively. The accuracy of this formula is shown in Fig.A5 for various generations of design rules.

## Appendix B : Expression for the Delay Degradation Factor

Here, the case where the channel-length modulation is not negligible, that is, $\lambda \neq 0$ See Fig.10 for the notation. For the upper MOSFET, the drain current $I_D$ is written as

$$I_D = I_{D0}\left(\frac{1-v_M-v_T-\gamma_1 v_M}{1-v_T}\right)^n \frac{1+\lambda'(1-v_M)}{1+\lambda'} . \qquad (B1)$$

This curve (curve U in Fig.10) goes through $(v_M, I_U)$ defined as

$$v_M = v_U = \frac{1-v_T}{1+\gamma_1}\left(1 - \frac{1}{2^{1/n}}\right), \quad I_U = \frac{1}{2} I_{D0} \frac{1+\lambda'(1-v_U)}{1+\lambda'} . \qquad (B2)$$

The line LU in Fig.10 is drawn to pass (0, ID0) and (vM, IU).

For the lower MOSFET, $I_D$ is expressed as

$$I_D = I_{D0} \frac{1+\lambda' v_M}{1+\lambda'} \left(2 - \frac{v_M}{v_{D0}}\right) \frac{v_M}{v_{D0}} . \qquad (B3)$$

Therefore, the curve L goes through $(v_L, I_L)$ defined by the following expressions.

$$v_M = v_L = \left(1 - \frac{1}{\sqrt{2}}\right) v_{D0} , \quad I_L = \frac{1+\lambda' v_L}{1+\lambda'} I_{D0} .$$

The line LL is chosen so as to pass (0, 0) and ($v_M$, $I_U$). the N-connected MOSFET case can be treated similarly by reducing the value of LL to LL / (N-1), since (N-1) of the MOSFET's are operating in the linear region and so can be added linearly.

By solving the intersection of the line LU and the line LL, $I_{D0N}$ can be obtained:

$$I_{D0} - \frac{I_{D0}-I_U}{v_U} v_M = \frac{I_L}{v_L} \frac{v_M}{N-1} = I_{D0N} . \tag{B4}$$

Elimination of $v_M$ leads

$$F_D = \frac{I_{D0}}{I_{D0N}} = 1 + \frac{I_{D0}-I_U}{v_U} \frac{v_L}{I_L} (N-1) . \tag{B5}$$

With the assumption that $\lambda$ is small and and $(v_U - v_L) << 1$, for $F_D$ is reduced to

$$F_D = 1 + \frac{1-2^{-1/2}}{1-2^{-1/n}} \frac{v_{D0}}{1-v_T} (1 + \gamma_1) (1 + \lambda') (N-1) \tag{B6}$$

This formulae corresponds to Eq. 10 in the text.

## Appendix C : Approximated Expressions for the SCMS

In order to use the delay formulas derived in Appendix A for the SCMS, $I_{D0N}$, $V_{INV}$, $V_{D0}$, and n for the SCMS are required. In this Appendix, an approximated way of obtaining these parameters is described. As for $I_{D0N}$ and $V_{INV}$, they have already been given by Eq.B6 and

Eq. A14, respectively. The effective drain saturation voltage $V_{D0}$ of the SCMS is essentially unchanged from the VD0 of the single MOSFET, as seen in Fig.C1.

The remaining quantity is n. Let $n_{NJ}$ denote the velocity saturation index n observed when the J-th input terminal counting from the output of the N series-connected MOSFET's is chosen as an input. The NMOS case is explained here, but the PMOS case is symmetric.

First, the case of N = 2 and J = 2 is discussed. Suppose $0.5V_{DD}$ is applied to the lower MOSFET gate, the drain current, $I_{DM22}$, is expressed as follows because the lower MOSFET is operating in the saturated region.

$$I_{DM22} = \frac{W}{L_{EFF}} B \left(\frac{1}{2}V_{DD} - V_{TH}\right)^{n}. \tag{C1}$$

Knowing that the drain current is $I_{D02}$ at $V_{GS} = V_{DD}$ and $I_{DM22}$ at $V_{GS} = 0.5V_{DD}$, $n_{22}$ is calculated as

$$n_{22} = \frac{ln\,(I_{D02} / I_{DM22})}{ln\,((1-v_T) / (0.5-v_T))}. \tag{C2}$$

The next case is N = 2 and J = 1. In this case, the drain current, $I_{DM21}$, flowed when the gate voltage is set to $0.5V_{DD}$ can be calculated using the similar technique used in Appendix B.

$$I_{DM21} = I_{DM22} / \left\{1 + \frac{1}{2} n \frac{v_{D0}}{0.5-v_T}(1+\gamma_1)\right\}. \tag{C3}$$

Then, the following expression holds.

$$n_{12} = \frac{ln\,(I_{D02} / I_{DM21})}{ln\,((1-v_T) / (0.5-v_T))}. \tag{C4}$$

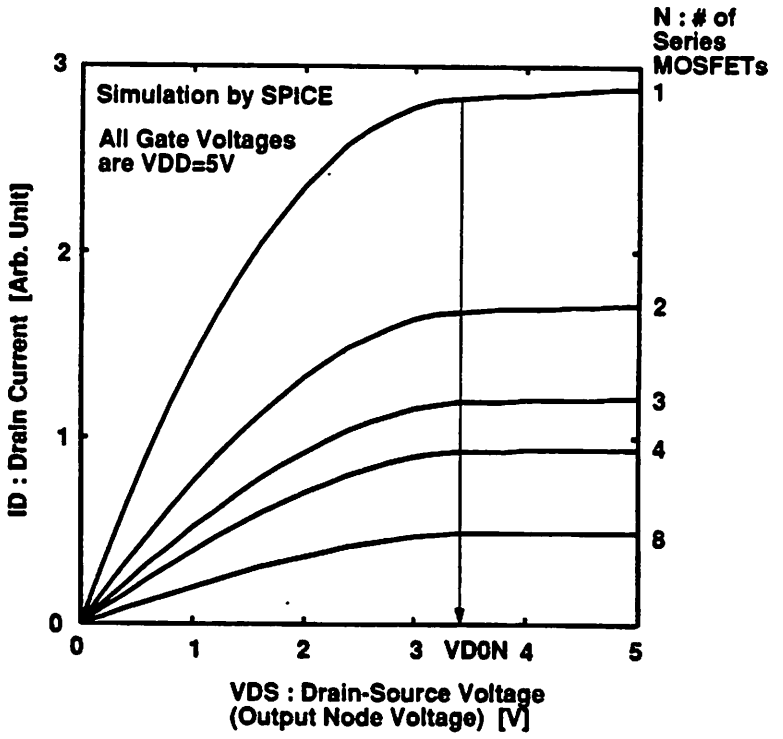


Fig.C1 Effective drain saturation voltage of SCMS

For the general N and J, the following empirical formulae can be employed.

$$n_{NJ} = \frac{n_{21}\, n_{22}}{(n_{21} - n_{22})\,(J-1) + n_{22}} \qquad \text{(C5)}$$

## References

[1] T.Sakurai, K.Nogami, M.Kakumu, and T.Iizuka, "Hot-Carrier Generation in Submicrometer VLSI Environment," IEEE J. Solid State Circ., SC-21, No.1, pp.187-192, Feb.1986.

[2] P.Penfield and J.Rubinstein, "Signal Delay in RC Tree Networks," 18th DAC, pp.613-617, June 1981.

[3] T.Sakurai, "CMOS Inverter Delay and Other Formulas Using alpha-Power MOS Model," ICCAD'88, pp.74-77, Santa Clara, Nov.1988.

[4] T.Sakurai, "Optimization of CMOS Arbiter and Synchronizer with Sub-micron MOSFET's," IEEE J. Solid State Circuits, SC-23, No.4, pp.901-906, Aug.1988.

[5] J.Mavor, M.Jack and P.Denyer, Introduction to MOS LSI Design, Addison-Wesley, 1983.

[6] L.Heller, W.Griffin, J.Davis and N.Thoma, "Cascode Voltage Switch Logic - A Differential Logic family," ISSCC'84, pp.16-17, Feb.1984.

[7] N.Hedenstierna and K.O.Jeppson, "CMOS Circuit Speed and Buffer Optimization," IEEE Trans., CAD-6, No.2, pp.270-280, Mar.1987.

[8] A.Vladimirescu, and S.Liu, "The Simulation of MOS Integrated Circuits Using SPICE2," UCB ERL Memo, No.M80/7, Oct.1980.

[9] J.K.Ousterhout, "A Switch-Level Timing Verifier for Digital MOS VLSI," IEEE Trans., CAD-4, No.3, pp.336-349, July 1985.

[10] M.Horowitz, "Timing Models for MOS Pass Networks," ISCAS'83, pp.198-201, 1983.

[11] T.Sakurai and A.R.Newton, "Analysis of Series-Connected MOSFET's in the Submicron Region", ESSCIRC'89, Sept.1989.

[12] K.Nogami, K.Sawada, M.Kinugawa and T.Sakurai, "VLSI Circuit Reliability under AC Hot-Carrier Stress," Symp. on VLSI Circuits, pp.13-14, Tokyo, May 1987.