Modeling the Overshooting Effect for CMOS Inverter Delay Analysis in Nanometer Technologies

Zhangcai Huang, Member, IEEE, Atsushi Kurokawa, Member, IEEE, Masanori Hashimoto, Member, IEEE, Takashi Sato, Member, IEEE, Minglu Jiang, and Yasuaki Inoue, Member, IEEE

Abstract-With the scaling of complementary metal-oxidesemiconductor (CMOS) technology into the nanometer regime, the overshooting effect due to the input-to-output coupling capacitance has more significant influence on CMOS gate analysis, especially on CMOS gate static timing analysis. In this paper, the overshooting effect is modeled for CMOS inverter delay analysis in nanometer technologies. The results produced by the proposed model are close to simulation program with integrated circuit emphasis (SPICE). Moreover, the influence of the overshooting effect on CMOS inverter analysis is discussed. An analytical model is presented to calculate the CMOS inverter delay time based on the proposed overshooting effect model, which is verified to be in good agreement with SPICE results. Furthermore, the proposed model is used to improve the accuracy of the switch-resistor model for approximating the inverter output waveform.

Index Terms—CMOS inverter, gate delay, nanometer technology, overshooting time, switch-resistor model, timing analysis.

I. INTRODUCTION

D UE TO THE input-to-output coupling capacitance, the complementary metal–oxide–semiconductor (CMOS) gate output voltage will be beyond the power supply range at the beginning of the transition. That is known as the overshooting effect and the time during which the output voltage is out of the supply voltage range is defined as the overshooting time. For the CMOS gate using conventional

Manuscript received January 27, 2009; revised July 5, 2009. Current version published January 22, 2010. This work was supported by the Waseda University Grant for Special Research Projects under Grant 2007A-968. This paper was recommended by Associate Editor D. Sylvester.

Z. Huang is with Fukuoka Industry, Science and Technology Foundation, Fukuoka-shi 814-0001, Japan. He is also with the Research Center of Information, Production, and Systems, Waseda University, Kitakyushu-shi 808-0135, Japan (e-mail: hzc 2002@asagi.waseda.jp).

A. Kurokawa is with Sanyo Electric Company, Ltd., Gifu 503-0195, Japan (e-mail: atsushi.kurokawa@sanyo.com).

M. Hashimoto is with the Department of Information Systems Engineering, Graduate School of Information Science and Technology, Osaka University, Suita-shi 565-0871, Japan (e-mail: hasimoto@ist.osaka-u.ac.jp).

T. Sato is with the Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (e-mail: takashi@i.kyoto-u.ac.jp).

M. Jiang and Y. Inoue are with the Graduate School of Information, Production, and Systems, Waseda University, Kitakyushu-shi, 808-0135 Japan (e-mail: jiangminglv1983@fuji.waseda.jp; inoue_yasuaki@waseda.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCAD.2009.2035539

long channel transistors, the overshooting effect is seldom of importance in digital circuits and is only of major importance in analog circuits [1]. Traditional gate delay models [2]–[6] associate a constant delay to an output-load related characteristic of its size and input signal transition time without taking into account the influence of the input-to-output coupling capacitance. Similarly, the CMOS gate switch-resistor model in [7] and the short-circuit power estimation model in [8] also ignored the influence of overshooting effect.

With the scaling of CMOS technology into submicronmeter features sizes, the overshooting effect becomes significant and thus begin to be considered in CMOS gate analysis. In [9]-[16], the nonlinearity induced by the input-to-output coupling capacitance is taken into account for gate delay models. In [17], [18], the power consumption estimation models of CMOS buffers were given accounting for the influence of the input-to-output coupling capacitance in submicronmeter technologies. Since the overshooting time becomes one of important parameters for CMOS gate timing analysis or power consumption estimation, several empirical models have been proposed to estimate overshooting time. In [12], an expression was derived to compute the CMOS gate delay time while the step input delay time was required. In [18], an empirical model was proposed to estimate CMOS gate power consumption. However, these models are still not accurate enough in various conditions.

As CMOS technology enters nanometer regime, the overshooting effect due to the input-to-output coupling capacitance becomes more significant and cannot be neglected anymore because it has much influence on CMOS gate analysis including CMOS circuit timing prediction, power estimation, and output waveform approximation. However, there has been still few researches to focus on the overshooting time estimation in nanometer technologies. It has been customarily assumed as constant values or zero. Therefore, it is necessary to develop an accurate model of overshooting time suitable for CMOS gate analysis. At the same time, it is also required to improve conventional CMOS gate delay models when the increasing influence of the input-to-output coupling capacitance is considered.

In this paper, the overshooting effect due to the inputto-output coupling capacitance is modeled for the CMOS

0278-0070/\$26.00 © 2010 IEEE

inverter delay analysis in nanometer technologies. By using the proposed overshooting time model, an analytical delay model is presented for CMOS inverters, and an improved switch-resistor model is given to approximate the inverter output waveform. The major contributions of this paper are as follows.

- The overshooting effect of CMOS inverter due to inputto-output coupling capacitance is demonstrated to have much influence on CMOS gate timing analysis.
- 2) The overshooting effect is modeled for the CMOS inverter analysis in nanometer technologies. The overshooting time predicted by the proposed model is verified with 32 nm predictive technology model (PTM) [21], [22] to be close to simulation program with integrated circuit emphasis (SPICE). Since the linear approximation for MOS transistor drain current is used, the proposed model can eliminate numerical iteration, and thus results in short calculation time.
- The proposed model is verified to be applicable to various device models such as 32 nm PTM high-k/metalgate model [21], [22].
- 4) An analytical model is proposed to calculate the CMOS inverter delay time based on the overshooting effect model, and is verified to be in good agreement with SPICE results.
- 5) The practical application of the proposed overshooting effect model is shown to improve the accuracy of the switch-resistor model. The simulation results show that the switch-resistor model using the proposed overshooting time model can accurately predict the CMOS inverter output waveform.

The rest of this paper is organized as follows. The problem formulation is described in Section II. The proposed analytical model of overshooting time is given in Section III. CMOS inverter delay model considering the overshooting effect is shown in Section IV, and the switch-resistor model considering the influence of the coupling capacitance is given in Section V. This paper is concluded in Section VI.

II. PRELIMINARIES

Fig. 1 shows a CMOS inverter with capacitive load C_L , where the input-to-output coupling capacitance C_M is considered. Fig. 2 shows the output voltage V_{out} of CMOS inverter for a falling input V_{in} using CMOS 32 nm PTM model [21], [22]. t_{50} is the time for the gate output voltage from the initial point to 50% V_{DD} point and t_{50} consists of the following two parts.

- 1) Overshooting time t_{ov} , which is caused by the input-tooutput coupling capacitance as shown in Fig. 2.
- 2) Gate output waveform rise time t_r from $t = t_{ov}$ to 50% V_{DD} point.

We thus have

$$t_{50} = t_{\rm ov} + t_r. (1)$$

The gate propagation delay t_D is defined as the time interval from $V_{in} = V_{DD}/2$ to $V_{out} = V_{DD}/2$. From Fig. 2, we can obtain



Fig. 1. CMOS inverter model and transistor leakage currents.



Fig. 2. Output voltage of CMOS inverter for a falling input.

the gate delay as

$$t_D = t_{50} - \frac{t_{\rm in}}{2}.$$
 (2)

It is known that CMOS gate output waveform t_{50} in nanometer technologies consists of two main parts. One is the output rise time t_r . The other is the overshooting time t_{ov} which is generally ignored in traditional research. The delay time t_D is thus expressed as

$$t_D = t_{\rm ov} + t_r - \frac{t_{\rm in}}{2}.$$
 (3)

Fig. 3 shows the rise time t_r and the overshooting time t_{ov} for various inverter sizes using the same 32 nm process model. The negative channel metal–oxide–semiconductor (NMOS) width W_n varies from 80 nm to 800 nm with $W_p = 2W_n$. Both positive channel metal–oxide–semiconductor (PMOS) and NMOS transistor length L is 40 nm. The capacitive load C_L is equal to 0.01 pF, and the input signal transition time t_{in} is 50 ps. From Fig. 3, it can be seen that the rise time decreases greatly as the inverter width increases. The overshooting time t_{ov} , in contrast, changes little with the increase of inverter size.



Fig. 3. Gate output waveform rise time t_r and overshooting time t_{ov} with various inverter sizes.



Fig. 4. Gate output waveform rise time t_r and overshooting time t_{ov} with various technologies.

Fig. 4 shows the delay time t_D and the overshooting time t_{ov} for various process technologies from 130 nm to 22 nm with typical supply voltages [23]. The transistor length *L* is the minimum feature size. The widths of all PMOS transistors W_p are ten times of *L* while W_p is two times of W_n . The load capacitance C_L is 0.01 pF. The input transition time t_{in} is also 50 ps. From Fig. 4, it is clearly shown that both times t_D and t_{ov} decrease with the scaling of the process technologies. However, the ratio of t_{ov}/t_D increases from 151% at 130 nm process to 202% at 22 nm process.

Therefore, it is inferred from the simulation results in Figs. 3 and 4 that the overshooting time t_{ov} becomes one of the most dominant parts of the output waveform of the CMOS inverter as the process technologies enter nanometer regime. Moreover, the overshooting time t_{ov} contributes much more to t_{50} when the inverter size increases. Consequently, the overshooting effect has more significant influence on CMOS inverter analysis, especially on static timing analysis in nanometer technology.

Currently, the use of delay models for performance cellbased delay calculation has become an industry standard in practical delay calculation tools. As cells are connected together by interconnect wires, the delay time of CMOS circuits between the input signal to output signal is just the sum of the delay time of each cell. Since the overshooting time t_{ov} is one of the most important part of the delay time, the total overshooting time of each cell is also accumulated in the delay time estimation.

To illustrate this, a chain of five inverters is used to observe the output waveform of each inverter as shown in Fig. 5. It can be seen that the overshooting time of the first gate propagates from stage to stage before it arrives the last one. The time when the output voltage V_{out5} begins to fall from V_{DD} to ground is



Fig. 5. Simulation results of the chain of inverters with 32 nm PTM model.

almost equal to the total overshooting time of all inverters. In Fig. 5, the delay time t_D of the inverter chain is 23.4 ps, while the overshooting time t_{ov} of all inverters is 25.32 ps which is larger than the delay time of the inverter chain.

It is clearly shown that the overshooting time is one of dominated parts in the inverter delay time estimation. Therefore, it is necessary to develop an accurate model for overshooting effect for the CMOS inverter analysis in nanometer technologies.

III. MODELING THE OVERSHOOTING EFFECT

The differential equation for the CMOS inverter of Fig. 1 loaded by a coupling capacitance C_M can be described as

$$C_L \frac{dV_{\text{out}}}{dt} = I_p - I_n + C_M \frac{d\left(V_{\text{in}} - V_{\text{out}}\right)}{dt}$$
(4)

where C_L is the output capacitive load including the inverter diffusion capacitance, the interconnect wire capacitance and the load gate input capacitance. Also, V_{out} and V_{in} are the gate output and input voltages, respectively. I_p and I_n are the PMOS and NMOS transistor currents, respectively. Here, the differential equation is solved only for the falling input ramp. Similar expressions can be derived for the rising input ramp. The input voltage for the falling input ramp is expressed as

$$V_{\rm in} = \begin{cases} V_{DD} & : t \le 0\\ \left(1 - \frac{t}{t_{\rm in}}\right) V_{DD} : 0 < t \le t_{\rm in} \\ 0 & : t > t_{\rm in} \end{cases}$$
(5)

where t_{in} is the input falling time. The value of C_M in input high state consists of the side-wall capacitance of both transistors drain and the gate to drain overlap capacitance of NMOS transistor in the linear region [18]

$$C_M = C_{\text{ox}} \left(\frac{W_{n \text{ eff}} L_{n \text{ eff}}}{2} + X_{Dp} W_{p \text{ eff}} + X_{Dn} W_{n \text{ eff}} \right)$$
(6)

where W_{peff} and W_{neff} are PMOS and NMOS effective channel widths, respectively, while L_{neff} is the NMOS effective channel length. Also, X_{Dp} and X_{Dn} are the gate-drain



Fig. 6. PMOS and NMOS transistor drain currents I_p and I_n during the overlapping period.

underdiffusion for PMOS and NMOS transistors, respectively. C_{ox} is the gate-oxide capacitance per unit area.

Leakage current in deep-submicrometer regimes is becoming a significant contributor to power dissipation of CMOS circuits as threshold voltage, channel length, and gate oxide thickness are reduced [24]. Fig. 1 shows an NMOS transistor considering leakage currents, which include the oxide tunneling current I_1 , the subthreshold leakage I_2 , the reverse-bias pn junction leakage I_3 , and so on. Commonly, the leakage components have very important influence for estimation and reduction of leakage power, especially for low-power applications. For the overshooting effect analysis, however, its influence is very small and can be ignored in the proposed method because the current components in (4) are much larger than the leakage currents.

In the proposed model, we firstly define two time points t_{TP} and $t_{v \text{ min}}$. As shown in Fig. 6, t_{TP} is the time when the PMOS transistor is on and $t_{TP} = (|V_{TP}|/V_{DD})t_{\text{in}}$, where V_{TP} is the threshold voltage of PMOS transistor. Also, $t_{v \text{ min}}$ is the time when the output voltage is at its minimum. Then the overshooting period can be divided into following two parts.

- 1) $t_{TP} \ge t > 0$: PMOS transistor is off. The voltage of the capacitance C_L decreases from zero due to a partial discharge through coupling capacitance.
- 2) $t_{ov} \ge t > t_{TP}$: PMOS transistor begins to conduct and the PMOS current I_p charges the load capacitance C_L . Then the voltage of the load capacitance increases. Once the output voltage goes up to zero, the NMOS current I_n becomes positive.

A. Proposed Model for t_{ov}

At the time $t = t_{ov}$, the charge of C_L is zero and the gate output waveform rises from zero. Therefore, we have

$$Q_L = Q_2 - Q_1 = 0 \tag{7}$$

where Q_1 is the charge flowing out of C_L , and Q_2 is the charge flowing into C_L during the overshooting time.

We rearrange (4) as

$$I_{C_L} = I_p - I_n + I_{CM} \tag{8}$$



Fig. 7. Approximation of the current I_{C_L} during t_{ov} .

where I_{C_L} is the load current and I_{CM} is the current flowing through the coupling capacitance C_M . When $t < t_{TP}$, $I_{C_L} = I_{CM} - I_n$. When $t = t_{TP}$, the PMOS transistor turns on. Then $I_{C_L} = I_p - I_n + I_{CM}$. Fig. 7 shows the load current I_{C_L} during the overlapping period. As shown in Fig. 7, the load current I_{C_L} can be assumed as linear during the charge time $t > t_v \min$, that is verified to be valid by SPICE simulations for various simulation conditions. Therefore, we can derive a simple expression for Q_2 as

$$Q_2 = \int_{t_{v \min}}^{t_{ov}} I_{C_L}(t) dt \approx \frac{I_{C_L}(t_{ov})}{2} [t_{ov} - t_{v \min}]$$
(9)

where $I_{C_L}(t_{ov})$ is the output load current when $t = t_{ov}$. Since $Q_1 = Q_2$, we obtain

$$t_{\rm ov} = \frac{2Q_1}{I_{C_L}(t_{\rm ov})} + t_{v\,\rm min} \tag{10}$$

where the second term $t_{v \text{ min}}$ corresponds to the discharge time and the first term $2Q_1/I_{C_L}(t_{ov})$ corresponds to the charge time.

When $t = t_{ov}$, the drain current I_n should be zero because the drain-source voltage V_{DS} of NMOS transistor is zero. Also, due to $dV_{in}/dt \gg dV_{out}/dt$ during overshooting period, $I_{C_L}(t_{ov})$ can be expressed as

$$I_{C_L}(t_{\rm ov}) = C_M \frac{dV_{\rm in}}{dt} + I_p(t_{\rm ov})$$
(11)

where $I_p(t_{ov})$ is the PMOS transistor current when $t = t_{ov}$.

From (10) and (11), it is seen that the overshooting time t_{ov} is determined by Q_1 , $t_{v \min}$ and $I_p(t_{ov})$. In the following, we will show the detailed description of these three terms.

When $t \leq t_{TP}$, the PMOS transistor is cut off. Since $V_{DS} \ll V_{GS}$, the NMOS transistor drain current can be simply expressed as [1]

$$I_n \approx K' \frac{W}{L} (V_{GS} - V_{TN}) V_{DS}$$
(12)

where K' is the transconductance parameter, V_{GS} is the gateto-source voltage and V_{TN} is the NMOS transistor threshold voltage. Therefore, we have the expression

$$(C_L + C_M)\frac{dV_{\text{out}}}{dt} + \beta_n (V_{GS} - V_{TN})V_{\text{out}} = C_M \frac{dV_{\text{in}}}{dt}$$
(13)

where $\beta_n = K'W/L$. When $t_{TP} < t < t_v \min$, the PMOS transistor is on and the expression for the output voltage is more complicated. It is also inferred from Figs. 6 and 7 that the output voltage V_{out} at $t_v \min$ is close to the one at t_{TP} because the load currents I_{C_L} at the two time points are almost zero. The reason can be expressed as follows. When $t < t_{TP}$, the current flowing out of the capacitive load begins to decrease to zero because of the NMOS drain current. Then from the time $t = t_{TP}$, the PMOS drain current starts to increase significantly, and the load current becomes zero in very short time. The solution of the differential equation when $t = t_{TP}$ is given as

$$V_{\text{out}}(t_{TP}) = \frac{C_M \frac{dV_{\text{in}}}{dt}}{\beta_n (V_{DD} - |V_{TP}| - V_{TN})} \left[1 - e^{-\frac{\beta_n (V_{DD} - |V_{TP}| - V_{TN})}{C_L + C_M} \frac{t_{\text{in}} |V_{TP}|}{V_{DD}}} \right].$$
(14)

Because of $V_{\text{out}}(t_{v \min}) \approx V_{\text{out}}(t_{TP})$, the charge Q_1 is obtained from the output voltage and expressed as

$$Q_1 \approx C_L V_{\text{out}}(t_{TP}). \tag{15}$$

When $t = t_{v \text{ min}}$, since $dV_{\text{out}}/dt = 0$, the current flowing into capacitance load is zero. Therefore, from (4) we can derive the PMOS transistor current when $t = t_{v \text{ min}}$ as

$$I_p(t_{v \min}) = I_n(t_{v \min}) - C_M \frac{dV_{\text{in}}}{dt}$$
(16)

where $I_n(t_v \min)$ is the NMOS transistor current when $t = t_v \min$. Since the value of $V_{out}(t_v \min)$ is close to $V_{out}(t_{TP})$, $I_n(t_v \min) \approx I_n(t_{TP})$. Then substituting (14) into (12), we obtain

$$I_n(t_{v \min}) \approx C_M \frac{dV_{\text{in}}}{dt} \left[1 - e^{-\frac{\beta_n (V_{DD} - |V_{TP}| - V_{TN})}{C_L + C_M} \frac{|V_{TP}|}{V_{DD}} t_{\text{in}}} \right].$$
(17)

The assumption in (17) is reasonable because its accuracy is determined mainly by the approximation for $V_{\text{out}}(t_{v \min})$.

As shown in Fig. 8, the α -power approximation is very close to SPICE simulation to predict the PMOS transistor current. Thus, the α -power metal–oxide–semiconductor field effect transistor (MOSFET) model [4] is used to calculate $t_{v \min}$, and is written as

$$I_p(t_{v \min}) = I_{D0} \left(\frac{|V_{GS}| - |V_{TP}|}{|V_{DD} - |V_{TP}|} \right)^{\alpha}$$
(18)

where I_{D0} is the drain current in saturation at $V_{GS} = V_{DS} = V_{DD}$, α is the velocity saturation index and V_{GS} is a function of $t_{v \text{ min}}$. Note that although α -power MOSFET model was proposed for submicronmeter transistor, it is still helpful to model the transistor drain current in nanometer technologies, especially when the drain-source voltage is almost constant such as during the overshooting effect period. In this paper, α is assumed as 1.57 for 32 nm PTM model, which is verified by SPICE simulation results. Thus, we obtain the expression of $t_{v \text{ min}}$

$$t_{v \min} = t_{TP} + (t_{in} - t_{TP}) \left[\frac{C_M V_{DD}}{I_{D0} t_{in}} e^{-\frac{\beta_n (V_{DD} - |V_{TP}| - V_{TN})}{C_L + C_M} \frac{|V_{TP}|}{V_{DD}} t_{in}} \right]^{\frac{1}{\alpha}}.$$
(19)

In (10), I_{C_L} is also determined by t_{ov} . Although the α -power MOSFET model or some other more accurate models can be used to obtain t_{ov} , the solution of such a nonlinear problem



Fig. 8. α-power approximation for PMOS transistor drain current.



Fig. 9. Linear approximation for PMOS transistor drain current.

will result in complicated computation. To avoid numerical procedures, an initial value for t_{ov} is first solved by using a linear transistor drain current model, as shown in Fig. 9. Note that V_G is the PMOS gate voltage when $t = t_{ov}$ and the transistor current is described as

$$I_p = I_{D0} \left(\frac{|V_{GS}| - |V_{TPL}|}{|V_{DD} - |V_{TPL}|} \right)$$
(20)

where V_{TPL} is the equivalent threshold voltages. The value is obtained by extending the line and intercept with *x*-axis. Also

$$|V_{GS}| = \frac{t_{\rm ov}}{t_{\rm in}} V_{DD}.$$
 (21)

Through (10), (11), and (20), we can obtain

$$t_{\rm ov} = \frac{2Q_1}{\frac{t_{\rm ov} - t_{TPL}}{t_{\rm in} - t_{TPL}} I_{D0} + C_M \frac{dV_{\rm in}}{dt}} + t_v \min$$
(22)

where $t_{TPL} = \frac{|V_{TPL}|}{V_{DD}} t_{in}$. By solving (22), we then can derive the final expression for t_{ov} as

$$t_{\rm ov} = t_{v\,\rm min} - t_{\gamma} + \sqrt{\frac{2Q_1}{I_{D0}}(t_{\rm in} - t_{TPL}) + t_{\gamma}^2}$$
(23)

where

$$t_{\gamma} = \frac{1}{2} \left(t_{v \min} - t_{TPL} - C_M \frac{V_{DD} t_{\rm in}}{t_{\rm in}} \frac{t_{\rm in} - t_{TPL}}{I_{D0}} \right).$$
(24)

Here, Q_1 and $t_{v \min}$ are obtained using (15) and (19).

Table I shows the calculation results for t_{ov} with (23). It is inferred that the model using linear assumption for PMOS drain current is still not very accurate to estimate the



TABLE I COMPARISON OF THE TWO CALCULATION APPROACHES FOR t_{ov}



Fig. 10. Overshooting time t_{ov} of the proposed method and SPICE simulation for various gate drivability.

overshooting time. Therefore, another step is given to improve the accuracy. Substituting the initial t_{ov} of (23) into α -power MOSFET model, the new drain current I_p is solved. Then using (10) and (11), the new value for t_{ov} is finally obtained. It is clearly demonstrated from Table I that the error to predict t_{ov} is reduced significantly by using the improved method.

It is possible to improve the accuracy further by using iterative computation based on new t_{ov} . However, the calculation time will increase because of numerical procedures. Moreover, the proposed method using only two steps is accurate enough to calculate the overshooting time.

B. Simulation Results

We have applied the proposed method to the CMOS inverters to further verify its appropriateness with the same CMOS 32 nm PTM model. Note that the length of all NMOS and PMOS transistors are set to be $L_n = L_p = 40$ nm. And the ratio of PMOS width to NMOS width is $W_p/W_n = 2/1$.

Fig. 10 plots the results of overshooting time versus transistor widths which vary from 40 nm to 400 nm. Results show that the overshooting time t_{ov} decreases a little when the gate size increases. Moreover, it is seen that the presented model matches very well with SPICE simulations within 1.6% error for various simulation conditions.

Fig. 11 shows the results of the proposed model and SPICE simulation with respect to the various capacitive loads. The capacitive load varies from 0.02 pF to 0.2 pF. Simulation results show that the overshooting time t_{ov} increases a little against the capacitive load. The proposed model provides a good estimation of the overshooting time for small and large capacitive loads within 3.01% error.

Fig. 12 compares the overshooting time of the proposed method, SPICE simulation and the model in [18] with respect



Fig. 11. Overshooting time t_{ov} of the proposed model and SPICE simulation for various capacitive loads.



Fig. 12. Overshooting time t_{ov} of the proposed model and SPICE simulation for various input signal transition times.

to various input signal transition times. Note that the model was proposed in [18] for CMOS gate power estimation, and was also used to predict CMOS gate delay time in [16]. In Fig. 12, the input signal transition time ranges from 20 ps to 300 ps. Also the width of NMOS transistor W_n is 80 nm while the capacitive load is 0.01 pF. It can be observed that the presented analytical model for the evaluation of the overshooting time gives results close to those derived from SPICE simulations with 2.06% average error. Moreover, the prediction result of the model in [18] is not accurate compared with SPICE simulation with 19.72% average error.

From the simulation results in Figs. 10–12, it is clearly demonstrated that the proposed model is accurate for predicting the overshooting time t_{ov} in various conditions.

C. High-k/Metal-Gate Model

Since the gate oxide leakage current is increasing with decreasing SiO_2 thickness, its scaling is becoming exceedingly difficult. With further scaling, the high-*k* gate dielectrics and metal gate electrodes will be required for high-performance and low-power CMOS applications. Much work has been done for high-*k*/metal-gate transistors in [25], [26] recently. In this paper the proposed overshooting model is also applied to the inverter constructed with high-*k*/metal-gate transistor, where the 32 nm PTM high-*k*/metal-gate models for high-performance applications (PTM HP) and for low-power ap-



Fig. 13. Overshooting time t_{ov} of the proposed model and SPICE simulation for various capacitive loads with 32 nm PTM HP model.



Fig. 14. Overshooting time t_{ov} of the proposed method and SPICE simulation for various gate drivability with 32 nm PTM LP model.

plications (PTM LP) [20], [21] are utilized in the SPICE simulation.

Fig. 13 shows the results of the proposed model and SPICE simulation with respect to the various capacitive loads, where 32 nm PTM HP model is used. The capacitive load varies from 0.02 pF to 0.2 pF. Simulation results show that the proposed model provides a good estimation of the overshooting time for various capacitive loads within 3.31% error.

Fig. 14 plots the results of overshooting time versus transistor widths which vary from 80 nm to 800 m with 32 nm PTM LP model. Results show that the proposed model reproduces the overshooting time accurately compared with SPICE simulations with 1.17% average error.

Figs. 13 and 14 demonstrate that the proposed model is applicable to predict the overshooting time of CMOS inverters for various device models such as the high-*k*/metal-gate model.

IV. CMOS INVERTER DELAY ANALYSIS

In this section, the CMOS inverter delay analysis in nanometer technologies is discussed with considering overshooting effect because the delay time is one of the most important parameters for very-large-scale integration (VLSI) circuit performance.

As shown in (3), the delay time t_D in nanometer technologies can be obtained if the overshooting time t_{ov} , the rise time t_r and the input signal transition time t_{in} are known.



Fig. 15. Linear approximation for the current through capacitive load C_L to delay analysis.

Generally, the transition time t_{in} is given in delay calculation. In previous section, we have proposed an analytical model for the overshooting time t_{ov} . In the following, we will obtain the analytical expression of the rise time t_r .

A. Delay Model

Fig. 15 shows the output voltage and output current through capacitive load C_L for CMOS inverter in Fig. 1. It is seen that the output currents I_{CL} can be approximated linearly in order to simplify the analysis. Since the area with fill in Fig. 15 is just the total charges stored in capacitor C_L , we can obtain

$$\frac{I_{CL}^{t_{\rm ov}} + I_{CL}^{t_{\rm in}}}{2} (t_{\rm in} - t_{\rm ov}) + I_{CL}^{t_{\rm in}} [t_r - (t_{\rm in} - t_{\rm ov})] = \frac{V_{DD}}{2} C_L.$$
 (25)

Therefore, we can derive the expression for t_r as

$$t_r = \frac{V_{DD}C_L + (I_{CL}^{i_{\rm in}} - I_{CL}^{i_{\rm ov}})(t_{\rm in} - t_{\rm ov})}{2I_{CL}^{t_{\rm in}}}$$
(26)

where $I_{CL}^{t_{ov}}$, $I_{CL}^{t_{in}}$ are the load currents flowing through C_L at the times t_{ov} , t_{in} in Fig. 15, respectively. Here, $I_{CL}^{t_{ov}}$ can be obtained using (11). Commonly the PMOS transistor works in saturated region because $|V_{GS} - V_{TP}| < V_{DS}$. Thus, the current $I_{CL}^{t_{in}}$ can be approximated as the current I_{D0} of PMOS transistor. Then substituting (23) and (26) into (3), we can obtain the delay time of CMOS inverter.

In some cases, for example, when the load capacitance becomes very small or the input signal transition time is very slow, the t_{50} obtained using (23) and (26) will be smaller than t_{in} . Then the load current is difficult to predict, and the above analysis is not accurate to calculate the rise time. However, as described in [1], the rise time t_r is linearly changed with the input transition time t_{in} . Fig. 16 shows the rise time t_r versus the input transition time t_{in} using the same 32 nm PTM model. It is seen that the rise time t_r almost increases linearly with t_{in} . Fig. 16 indicates that the linear approximation is close to the simulation results. Therefore, the rise time t_r for $t_{in} > t_{50}$ can be approximately expressed as

1

$$t_r = t_r^0 + \kappa (t_{\rm in} - t_{\rm in}^0)$$
 (27)



Fig. 16. Rise time t_r versus input signal transition time t_{in} .



Fig. 17. Comparison of the proposed method and SPICE simulation for various gate drive ability with 32 nm PTM LP model.

where t_r^0 is the rise time that the input transition time t_{in} is equal to t_{50} . Actually, the parameter κ is almost constant and can be obtained with two rise time points as

$$\kappa = \frac{t_r^0 - t_r^1}{t_{\rm in}^0 - t_{\rm in}^1}$$
(28)

 t_r^1 are another rise time point for an input signal t_{in}^1 which is smaller than t_{50} . Substituting (23) and (27) into (3), the delay time is then obtained for $t_{in} > t_{50}$.

B. Simulation Results

The proposed analytical delay model is first verified with 32 nm PTM high-*k*/metal-gate models for high-performance applications (PTM HP) and for low-power applications (PTM LP) [21], [22].

Fig. 17 shows the comparisons with various NMOS widths from 80 nm to 800 nm with 32 nm PTM LP model. Fig. 18 shows the comparison with respect to the various capacitive load from 0.02 pF to 0.24 pF with 32 nm PTM HP model. From Figs. 17 and 18, it is known that the proposed model is close to SPICE simulation to predict the inverter delay time for various simulation conditions.

The following shows the comparisons of the proposed model with the conventional method in [12] and SPICE simulation. Note that the step input delay times in [12] are obtained from SPICE simulation results. Also pseudoempirical coefficients for delay model are extracted from simulations.



Fig. 18. Comparison of the proposed method and SPICE simulation for various capacitive loads with 32 nm PTM HP model.



Fig. 19. Comparison of the proposed method, SPICE simulation and [12] for various gate drive ability.



Fig. 20. Comparison of the proposed method, SPICE simulation, and [12] for various capacitive loads.



Fig. 21. Comparison of the proposed method with SPICE, [12], and the simple model for various input signal transition times.



Fig. 22. CMOS inverter with capacitive load and its switch-resistor model.

Fig. 19 shows the comparison with various NMOS widths from $0.2 \,\mu\text{m}$ to $2.2 \,\mu\text{m}$. The capacitive load is $0.02 \,\text{pF}$ and the input transition time is 100 ps. From Fig. 19, it can be seen that the proposed model is close to the SPICE simulation. The average error of the proposed model to predict the delay time is 3.24% while the average error of the delay model in [12] is 6.52%.

Fig. 20 shows the comparison with respect to the various capacitive loads from 0.02 pF to 0.24 pF. The input signal transition time is 50 ps and the size of inverter is that $W_p/W_n = 4 \,\mu$ m/2 μ m with $L_n = L_p = 40$ nm. The error of the proposed model is within 1.71%. If the conventional method in [12] is used, the error is from 6.71% to 19.57%.

Fig. 21 shows the comparison with various input signal transition times from 20 ps to 220 ps. In Fig. 21, the sizes of inverter transistors are $W_p/W_n = 4 \,\mu \text{m}/2 \,\mu \text{m}$ and the capacitive load C_L is 0.05 pF. From Fig. 21, it can be inferred that the proposed method is close to the SPICE simulation with 3.66% average error. The method using the overshooting model in [12] has a 9.66% average error with the increase of the input signal transition time.

From Figs. 19–21, it can be seen that the delay time can be reduced greatly by using large inverter sizes. The delay time increases with respect to capacitive load and input transition time. Since the proposed overshooting effect model is close to SPICE simulation results very well, the proposed analytical model can accurately predict the delay time of CMOS inverter in nanometer technologies.

V. SWITCH-RESISTOR MODEL

In this section, the application of the proposed overshooting time model is discussed to improve the accuracy of the switchresistor model.

Fig. 22 shows a CMOS inverter driving capacitive loads. The inverter can be modeled as a linear resistor with ramp input voltage for CMOS VLSI circuit analysis, which is called as switch-resistor model. With the scaling of CMOS technology into the nanometer regime, the overshooting time increases and becomes one of the most important parts of CMOS gate output waveform. In [19], the authors proposed a generalized CMOS gate delay model where a shift time t_0 is



Fig. 23. Improved switch-resistor model.



Fig. 24. SPICE simulation and the predictions of the improved switch-resistor model.



Fig. 25. Simulation results of the improved switch-resistor model to predict the output waveform of an inverter with various capacitive loads.

added to ramp input signal to improve the switching-resistor model. Similar work was also done in [20]. However, the shift time t_0 is not given.

In this paper, the accuracy of the switch-resistor model is improved while the overshooting effect is considered as shown in Fig. 23. For a ramp input signal $V_{in}(t)$ with t_{in} applied to CMOS inverter, the input signal in the improved model is changed as $V'_{in}(t)$ composed of a shift time t_{ov} and the ramp input with the same t_{in} . For the improved model, firstly t_{ov} is obtained using the proposed model to calculate the overshooting time.

Fig. 24 shows the comparison of SPICE simulation and the improved switch-resistor model to predict the output waveform of an inverter with capacitive loads. Note that the linear resistor in the switch-resistor model is derived by using the SPICE simulation results. From Fig. 24, it can be seen that the improved switch-resistor model can accurately predict the inverter output waveform. Here, the same 32 nm PTM model is utilized in the simulation results.

Fig. 25 shows the simulation results of the inverter when capacitive loads are 0.08 pF, 0.14 pF and 0.25 pF, respectively. From Fig. 25, it can be seen that the switch-resistor model can capture the output waveform of the CMOS inverter with various capacitive loads.

Therefore, the switch-resistor model with the proposed overshooting effect model can be used to predict the output waveform of CMOS inverters. Moreover, the accuracy of switch-models can be enhanced with the assistance use of the transistor-level parameters.

VI. CONCLUSION

The overshooting effect becomes much larger and cannot be neglected for VLSI analysis with the scaling of CMOS technology into nanometer. In this paper, an effective analytical model has been proposed to estimate the overshooting time for the CMOS inverter delay analysis in nanometer technologies. The accuracy of the proposed model has been proved to greatly agree with SPICE simulation results. Moreover, based on the proposed overshooting time model, an analytical gate delay model has been presented, which is verified to be agreement with SPICE results very well. Furthermore, the accuracy of the switch-resistor model has been improved to predict the inverter output waveform.

ACKNOWLEDGMENT

The authors would like to thank to the anonymous reviewers for their comments and suggestions, which have helped to improve this paper.

REFERENCES

- N. H. E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design:* A Systems Perspective. Reading, MA: Addison-Wesley, 1993, ch. 2, p. 51 and ch. 4, pp. 216–218.
- [2] L. M. Brocco, S. P. McCormick, and J. Allen, "Macromodeling CMOS circuits for timing simulation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 7, no. 12, pp. 1237–1249, Dec. 1988.
- [3] D. Auvergne, N. Azemard, D. Deschacht, and M. Robert, "Input waveform slope effect in CMOS delays," *IEEE J. Solid-State Circuits*, vol. 25, no. 6, pp. 1588–1590, Apr. 1990.
- [4] T. Sakurai and A. R. Newton, "Alpha-power-law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, Apr. 1990.
- [5] S. Dutta, S. S. Shetti, and S. L. Lushy, "A comprehensive delay model for CMOS inverters," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 864–871, Aug. 1995.
- [6] H.-C. Chow and W.-S. Feng, "An analytical CMOS inverter delay model including channel-length modulations," *IEEE J. Solid-State Circuits*, vol. 27, no. 9, pp. 1303–1306, Sep. 1992.
- [7] R. E. Bryant, "A switch-level model and simulator for MOS digital systems," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 33, no. 2, pp. 160–177, Feb. 1984.
- [8] H. J. M Veendrick, "Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 33, no. 2, pp. 160–177, Feb. 1984.

- [9] K. O. Jeppson, "Modeling the influence of the transistor gain ratio and the input-to-output coupling capacitance of the CMOS inverter delay," *IEEE J. Solid-State Circuits*, vol. 29, no. 6, pp. 646–654, Jun. 1994.
- [10] A. Nabavi-Lishi and N. C. Rumin, "Inverter models of CMOS gates for supply current and delay evaluation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 13, no. 10, pp. 1271–1279, Oct. 1994.
- [11] L. Bisdounis, S. Nikolaidis, and O. Koufopavlou, "Analytical transient response and propagation delay evaluation of the CMOS inverter for short-channel devices," *IEEE J. Solid-State Circuits*, vol. 33, no. 2, pp. 302–306, Feb. 1998.
- [12] J. M. Daga and D. Auvergne, "A comprehensive delay macro modeling for submicrometer CMOS logics," *IEEE J. Solid-State Circuits*, vol. 34, no. 1, pp. 42–55, Jan. 1999.
- [13] D. Auvergne, J. M. Daga, and M. Rezzoug, "Signal transition time effect on CMOS delay evaluation," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 47, no. 9, pp. 1362–1369, Sep. 2000.
- [14] A. A. Hamoui and N. C. Rumin, "An analytical model for current, delay, and power analysis of submicron CMOS logic circuits," *IEEE Trans. Circuits Syst. II, Analog Digit Signal Process.*, vol. 47, no. 10, pp. 999– 1007, Oct. 2000.
- [15] P. Maurine, M. Rezzoug, N. Azemard, and D. Auvergne "Transition time modeling in deep submicron CMOS," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 21, no. 11, pp. 1352–1363, Nov. 2002.
- [16] J. L. Rossello and J. Segura, "An analytical charge-based compact delay model for submicron CMOS inverters," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 51, no. 7, pp. 1301–1311, Jul. 2004.
- [17] S. Turgis and D. Auvergne, "A novel macromodel for power estimation in CMOS structures," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 17, no. 11, pp. 1090–1098, Nov. 1998.
- [18] J. L. Rossello and J. Segura, "Charge-based analytical model for the evaluation of power consumption in submicron CMOS buffers," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 21, no. 4, pp. 433–448, Apr. 2002.
- [19] F. Dartu, N. Menezes, and L. T. Pileggi, "Performance computation for precharacterized CMOS gates with RC loads," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 15, no. 5, pp. 544–553, May 1996.
- [20] M. Shao, M. D. F. Wong, H. Gao, Y. Gao, L.-P Yuan, L.-D. Huang, and S. Lee "Explicit gate delay model for timing evaluation," in *Proc. Int. Symp. Phys. Des.*, Apr. 2003, pp. 32–38.
- [21] Predictive Technology Model (PTM) [Online]. Available: http://www.eas.asu.edu/~ptm/
- [22] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45 nm early design exploration," *IEEE Trans. Electron Devices*, vol. 53, no. 11, pp. 2816–2823, Nov. 2006.
- [23] Semiconductor Industry Association, "Process integration, devices, and structures," in *The International Technology Roadmap for Semi*conductors (ITRS), 2007 ed. [Online]. Available: http://www.itrs.net/ Links/2007ITRS/Home2007.htm
- [24] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deepsubmicrometer CMOS circuits," *Proc. IEEE*, vol. 91, no. 2, pp. 305–327, Feb. 2003.
- [25] B. Cheng, M. Cao, R. Rao, A. Inani, P. V. Voorde, W. M. Greene, J. M. C. Stork, Z. Yu, P. M. Zeitzoff, and J. C. S. Woo, "The impact of high-gate dielectrics and metal gate electrodes on sub-100 nm MOSFETs," *IEEE Trans. Electron Devices*, vol. 46, no. 7, pp. 1537–1544, Jul. 1999.
- [26] R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz, "High-k/metal-gate-stack and its MOSFET characteristics," *IEEE Elec. Dev. Lett.*, vol. 25, no. 6, pp. 408–410, Jun. 2004.



Zhangcai Huang (S'05–M'08) was born on March 1978, in Jiangxi Province, China. He received the B.Eng. degree in electrical engineering from East China JiaoTong University, Nanchang, China, in 2001, the M.Eng. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2004, and the Ph.D. degree in information, production and systems engineering from the Graduate School of Information, Production and Systems, Waseda University, Kitakyushu, Japan, in 2007.

He was a Research Associate at the Research Center of Information, Production and System, Waseda University from April 2007 to October 2007. He is currently a Research Fellow of Fukuoka Industry, Science and Technology Foundation, Fukuoka-shi, Japan, and he has also been a Visiting Research Fellow at Research Center of Information, Production, and System, Waseda University, since November 2007. His major research interests include very-large-scale integration (VLSI) timing analysis, lowpower and low-voltage analog circuit design, analog VLSI processing systems, and numerical analysis of nonlinear circuits and systems.

Dr. Huang is a Member of the Institute of Electronics, Information, and Communication Engineers.



Atsushi Kurokawa (M'98) received the B.Eng. degree in electrical engineering from Seikei University, Tokyo, Japan, in 1986, and the Ph.D. degree in information, production, and systems engineering from Waseda University, Kitakyushu, Japan, in 2005.

Since 1986, he has been with SANYO Electric Company, Ltd., Gifu, Japan. From 2002 to 2005, he was with the Semiconductor Technology Academic Research Center, and from 2006 to 2008, he was with SANYO Semiconductor Company, Ltd. In 2008 and 2009, he was also a Visiting Associate Professor

with the Department of Electronic Engineering, Gunma University. His current research interests focus on developing the design technology of advanced SoC designs.

Dr. Kurokawa is a Member of the Information Processing Society of Japan and the Institute of Electronics, Information, and Communication Engineers.



Masanori Hashimoto (S'00–A'01–M'03) received the B.Eng., M.Eng., and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 1997, 1999, and 2001, respectively.

From 2001 to 2004, he was an Instructor in the Department of Communications and Computer Engineering, Kyoto University. Since 2004, he has been an Associate Professor in the Department of Information Systems Engineering, Graduate School of Information Science and Technology, Osaka Uni-

versity, Suita-shi, Japan. His research interests include computer-aided-design for digital integrated circuits, and high-speed circuit design.

Dr. Hashimoto is a Member of the Institute of Electronics, Information, and Communication Engineers, the Association for Computing Machinery, and the Information Processing Society of Japan. He received the Best Paper Award at the Asia and South Pacific Design Automation Conference 2004. He has served on the technical program committees for several international conferences, including the Design Automation Conference, the International Conference on Computer Aided Design, the Asia and South Pacific Design Automation Conference, the International Conference on Computer Design, and the International Symposium on Quality Electronic Design.



Takashi Sato (M'98) received the B.Eng. and M.Eng. degrees from Waseda University, Tokyo, Japan, and the Ph.D. degree from Kyoto University, Kyoto, Japan.

He was with Hitachi, Ltd., Tokyo, Japan, from 1991 to 2003, with Renesas Technology Corp., Tokyo, Japan, from 2003 to 2006, and with Tokyo Institute of Technology, Yokohama, Japan, from 2006–2009. In 2009, he joined the Graduate School of Informatics, Kyoto University, Kyoto, Japan, where he is currently a Professor. He was a Visiting

Industrial Fellow at the University of California, Berkeley, from 1998 to 1999. His research interests include computer-aided design for nanometerscale large-scale integration design, fabrication-aware design methodology, and performance optimization for variation tolerance.

Dr. Sato is a Member of the Institute of Electronics, Information, and Communication Engineers. He received the Beatrice Winner Award at the International Solid-State Circuits Conference, 2000, and the Best Paper Award at the International Symposium on Quality Electronic Design, 2003.



Minglu Jiang was born on November 13, 1983, in Jiangsu Province, China. He received the B.Eng. degree in electrical engineering from Southeast University, Nanjing, China in 2006, and the M.Eng. degree in information, production and systems engineering from the Graduate School of Information, Production, and Systems, Waseda University, Kitakyushu, Japan, in 2008, where he is currently pursuing the Ph.D. degree in information, production, and systems engineering.

His research interests include static timing analy-

sis, statistical analysis, and low-power and low-voltage analog circuit design. Mr. Jiang is a Student Member of the Institute of Electronics, Information and Communication Engineers.



Yasuaki Inoue (M'96) was born on September 6, 1945, in Niigata, Japan. He received a Diploma from the Department of Electronics, Nagaoka Technical High School, Niigata, Japan, in 1964, and the D.E. degree in electronics and communication engineering from Waseda University, Tokyo, Japan, in 1996.

From 1964 to 2000, he was with Sanyo Electric Company, Ltd., Gunma, Japan, where he was engaged in research and development in analog integrated circuits and analog/digital computer-aided

design (CAD) systems. In Sanyo Semiconductor Company, he was a General Manager of the Department of CAD Engineering from 1993 to 1998 and the Department of Memory Development from 1998 to 2000. From 2000 to 2003, he was a Professor in the Department of Integrated Cultures and Humanities, and also a Professor at the Graduate School of Integrated Science and Art, University of East Asia, Shimonoseki, Japan. Since 2003, he has been a Professor with the Graduate School of Information, Production, and Systems, Waseda University, Kitakyushu, Japan. He holds over 40 patents. His research interests include numerical analysis of nonlinear circuits and systems, analog circuits, and large-scale integration CAD systems.

Prof. Inoue was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART II from 1997 to 1999. He received the Ishikawa Award from the Union of Japanese Scientists and Engineers in 1988, the Distinguished Service Award from the Science and Technology Agency, the Japanese Government in 1999, the TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2002, the Achievement Award from the Information Processing Society of Japan in 2003, and the Funai Information, and Communication Engineers, the Information Processing Society of Japan, and the Japan, the Institute of Electrical Engineers of Japan, and the Japan Society for Simulation Technology.