An Accurate Analytical Propagation Delay Model of Nano CMOS Circuits

Ping Liu, Yong-Bin Kim
Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Email: {pliu, ybk }@ece.neu.edu

Young Jun Lee
Nextchip Company
Seoul, South Korea
Email: yjlee@nextchip.com

Abstract—An accurate analytical transient response and propagation delay model of nano CMOS inverter is presented. A modified version of α-power law MOSFET current model is proposed. The proposed model overcomes the over-estimation of linear region current in α-power law current model, and takes into account the channel length modulation effects in nano devices. A new methodology to estimate propagation delay by solving non-homogeneous linear differential equation of CMOS inverter is developed based on this current model. An analytical transient output response and propagation delay expression is derived. The final results are in excellent agreement with HSPICE simulations within 3% error for a wide range of transistor sizes, capacitor loads and input transition time. This delay estimation model demonstrates robustness for a wide range of processes from 0.25um to 65nm technology.

I. INTRODUCTION

The estimation of critical path delay using HPSICE simulation has become more expensive in terms of CPU time as the complexity of modern VLSI system increases rapidly. A fast yet accurate analytical model is often desired. It not only saves the estimation time, but also provides valuable insights for designers by including dominant parameters in expression.

A number of methods have been proposed to derive the propagation delay of CMOS inverters [1]-[4]. Sakurai and Newton derived a simple close-form delay expression for series-connected MOSFET circuits in [1], based on the α-power law MOSFET current model. However, the gate-to-drain coupling capacitance and short-circuit current are ignored. In [2], Bisdounis et al. presented an analytical propagation delay model which overcomes the weakness of the work in [1], but it ignores the channel length modulation effect, which is important for modern deep submicron technologies. Hamoui et al. [3] developed a comprehensive model for current, delay and power by first computing definable reference points on the output waveforms and then using linear approximation. Since numerical methods are used, this increases the computation time. Rossello et al. proposed an analytical charge-based delay model for CMOS inverters in [4]. The charge transferred when input is rising and static high is calculated and then used to derive the delay expression.

In this paper, we propose a modified α-power law MOSFET model which overcomes the over-estimation of linear region current in previous model and takes into account the channel length modulation. A general solution for the non-homogeneous linear differential equation of CMOS inverter is developed based on this current model. Operation regions of CMOS inverter are defined, and output transient response expressions are derived for each region. Propagation delay is calculated by taking the time difference of the 50% transition points of the input and output waveforms. The model is compared to HPSICE simulation considering a large set of parameters for TSMC 0.25um, 0.18um, 90nm, 65nm technology to show the accuracy.

II. DEEP SUBMICRON MOSFET CURRENT MODEL

The α-power law MOSFET model proposed by Sakurai and Newton [5] is widely used to model short-channel devices.

However, as technology keeps scaling down, this model does not best fit the actual I-V curve. First, the $\alpha/2$ power has made the current in linear region being over-estimated., which causes a discontinuity from the saturation region to the linear region. Second, channel length modulation is more important as device dimension keeps shrinking, especially for the MOSFETs which have minimum channel length and relatively big channel width. A modified current model is given below:

$$I_D = \begin{cases} 0, & (V_{GS} \le V_{TH} : cutoff\,) \\ \beta_l\,(v - v_{th})^{\alpha} V_{DS}, & (V_{DS} < V_{DSAT} : linear\,) \\ \beta_s\,(v - v_{th})^{\alpha}\left[1 + \lambda(V_{DS} - V_{DD})\right], & (V_{DS} \ge V_{DAST} : saturation\,) \end{cases} \quad (1)$$

where $\beta_s = \dfrac{I_{D0}}{(1 - v_{th})^{\alpha}}$, $\quad \beta_l = \dfrac{\beta_s \cdot \left[1 + \lambda(V_{DSAT} - V_{DD})\right]}{V_{DSAT}}$

α is the velocity saturation index, V_{D0} is the drain saturation voltage at $V_{GS} = V_{DD}$, I_{D0} is the drain current at $V_{GS} = V_{DS} = V_{DD}$. V_{DSAT} is the drain saturation voltage at $V_{GS} = V_{DD}$, but for the purpose of delay calculation, it could be considered to be the saturation voltage for all V_{GS}. Because V_{DSAT} is small and will be smaller as technology keeps scaling down, different V_{GS} won't make much variation to the saturation voltage. $v = V_{GS} / V_{DD}$ is the normalized input, when input ramp is in transition, $v = t / tr$.

Similarly, $v_{th} = V_{th} / V_{DD}$ is the normalized threshold voltage. λ is the empirical channel length modulation factor.

III. TRANSIENT RESPONSE ANALYSIS

The following transient analysis will be concerning the dynamic discharging behavior of CMOS inverter when input ramp is rising with transition time tr. The charging case will be symmetric.

Fig. 1 illustrates the discharging behavior of CMOS inverter. C_M is the gate-to-drain coupling capacitance, which could be calculated as in [6]. C_L is the load capacitance.

The discharging behavior of CMOS inverter could be described as the differential equation below:

$$\left(C_L + C_M\right)\frac{dV_{out}}{dt} = C_M \frac{dV_{in}}{dt} + I_p - I_n \qquad (2)$$

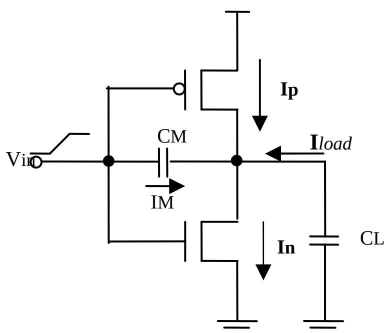

Fig. 1 Discharging Inverter Model

(A) Fast Input Case

Region 1: $0 \le v \le v_{thn}$. In this region, NMOS is in cutoff region, the output voltage keeps at V_{DD}.

Region 2: $v_{thn} < v \le 1 - v_{thp}$. In this region, NMOS is saturated and PMOS is in linear region. For fast input ramp, $V_{sgp} = V_{DD} - V_{in}$ is pulled down to a small value in a short time, which means only a small amount current is conducting. But for NMOS, $V_{gs} = V_{in}$ increases quickly, and results in relatively high drain current. Therefore, PMOS current could be neglected.

Also in this region, the following differential equation is set up for the case:

$$\frac{dV_{out}}{dv} + p(v) \cdot V_{out} = q(v) \qquad (3)$$

where $p(v) = \dfrac{\beta_{sn} \cdot tr \cdot \lambda_n}{C_L + C_M} \cdot (v - v_{thn})^{\alpha n}$

$$q(v) = cr \cdot V_{DD} - K_q \cdot (v - v_{thn})^{\alpha n}$$

$$cr = \frac{C_M}{C_L + C_M}, \; K_q = \frac{\beta_{sn} \cdot tr \cdot (1 - \lambda_n \cdot V_{DD})}{C_L + C_M}$$

The solution of this differential equation is obtained as following:

$$V_{out} = e^{-\int p(v)dv} \cdot \left(\int q(v) \cdot e^{\int p(v)dv} dv + C12\right) \qquad (4)$$

Let $y = \int p(v)dv = K_y \cdot (v - v_{thn})^{\alpha n + 1}$ (5)

where $K_y = \dfrac{\beta_{sn} \cdot tr \cdot \lambda_n}{(C_L + C_M) \cdot (\alpha_n + 1)}$

$$\int q(v) \cdot e^{\int p(v)dv} dv$$

$$= \int cr \cdot V_{DD} \cdot e^y \cdot dv - \int K_q \cdot (v - v_{thn})^{\alpha n} \cdot e^y \cdot dv$$

$$= \int cr \cdot V_{DD} \cdot e^y \cdot dv - \frac{\lambda_n \cdot V_{DD} - 1}{\lambda_n} \cdot e^y \qquad (6)$$

Taylor series expansion of e^y is used to make the first item on the right side integrable.

Finally, the output transient response in this region can be expressed as:

$$V_{out} = e^{-y}\{cr \cdot V_{DD} \cdot [v + \frac{K_y}{\alpha_n + 2}(v - v_{thn})^{\alpha n + 2}] + \frac{\lambda_n \cdot V_{DD} - 1}{\lambda_n} \cdot e^y + C12\} \qquad (7)$$

C12 could be calculated given initial condition that $v = v_{thn}$, $V_{out} = V_{DD}$, that gives:

$$C12 = \frac{1}{\lambda_n} - cr \cdot V_{DD} \cdot v_{thn} \qquad (8)$$

Region 3: $1 - v_{thp} < v \le 1$. In this region, NMOS is saturated and PMOS is in cutoff region. The output transient response expression is the same as region 2.

Region 4: $1 < v \le v_{linn}$. In this region, the input ramp has arrived at the static value V_{DD}, NMOS is still saturated, and PMOS is off.

$$V_{out} = (V_{DD} - \frac{1}{\lambda_n}) + C34 \cdot e^{-z} \qquad (9)$$

where $z = K_z \cdot v$, $K_z = \dfrac{\beta_{sn} \cdot tr \cdot \lambda_n}{C_L + C_M} \cdot (1 - v_{thn})^{\alpha n}$

C34 is calculated given the initial condition that $v = 1$, $V_{out3} = V_{out4}$ where V_{out3}, V_{out4} stand for the output transient response calculated using the expressions in region 3 and region 4 separately.

$$C34 = [cr \cdot V_{DD} \cdot (1 + \frac{K_y \cdot (1 - v_{thn})^{\alpha n+2}}{\alpha_n + 2}) + \frac{1}{\lambda_n}$$

$$- cr \cdot V_{DD} \cdot v_{thn}] \cdot \exp[K_z - K_y (1 - v_{thn})^{\alpha n+1}] \quad (10)$$

v_{linn} is calculated given the boundary condition that $v = v_{linn}$, $V_{out4} = V_{DSATn}$

$$v_{linn} = \frac{1}{K_z} \cdot \log(\frac{C34}{V_{DSATn} + \frac{1}{\lambda_n} - V_{DD}}) \quad (11)$$

Region 5: $v > v_{linn}$, NMOS enters the linear region.

$$V_{out} = \exp\{-\frac{tr \cdot I_{D0n}}{V_{DSATn} \cdot (C_L + C_M)} \cdot [1 + \lambda_n \cdot (V_{DSATn} - V_{DD})] \cdot v + C45\} \quad (12)$$

C45 is calculated given the boundary condition that $v = v_{linn}$, $V_{out5} = V_{DSATn}$

$$C45 = \frac{tr \cdot I_{D0n}}{V_{DSATn} \cdot (C_L + C_M)} \cdot [1 + \lambda_n \cdot (V_{DSATn} - V_{DD})] \cdot v_{linn} + \log(V_{DSATn}) \quad (13)$$

(B) Slow Input or Small Fanout Case

When the transition time of input ramp is large or the fanout is very small, short circuit current could not be ignored. Approximated expressions of output transient response are given below:

Region 1: $0 \le v \le v_{thn}$, NMOS is off, PMOS is in linear region.

$$V_{out} = \frac{1}{A} \cdot [(cr + A) \cdot V_{DD} - e^{-A \cdot v + C1}] \quad (14)$$

where

$$A = \frac{tr \cdot \beta_{lp}}{C_L + C_M} \cdot (1 - \frac{v_{thn}}{2} - v_{thp})^{\alpha p}, \; C1 = \log(cr \cdot V_{DD})$$

Region 2: $v_{thn} < v < v_{sp}$, NMOS is saturated, PMOS is in linear region.

$$V_{out} = V_{DD} \cdot (1 + \frac{cr}{K}) - C12 \cdot e^{-K \cdot v} + \frac{1}{2} \cdot K^2 \cdot \frac{(v - v_{thn})^{\alpha n+3}}{\alpha_n + 3}]$$

$$- b \cdot e^{K(v_{thn} - v)} \cdot [\frac{(v - v_{thn})^{\alpha n+1}}{\alpha_n + 1} + K \cdot \frac{(v - v_{thn})^{\alpha n+2}}{\alpha_n + 2} \quad (15)$$

where $b = \frac{\beta_{sn} \cdot tr}{C_L + C_M}$, $C12 = \frac{cr}{K} \cdot V_{DD} \cdot e^{K \cdot v_{thn}}$

$$K = \frac{\beta_{lp} \cdot tr}{C_L + C_M} \cdot (\frac{1 - v_{thp}}{2})^{\alpha p} + \lambda_n \cdot b \cdot (\frac{1 - v_{thp}}{2} - v_{thn})^{\alpha n}$$

v_{sp} is calculated by letting $V_{out2} = V_{DD} - V_{DSATp}$.

Region 3: $v_{sp} \le v \le 1 - v_{thp}$, NMOS is saturated, and PMOS is saturated.

$$V_{out} = [-b^2 \cdot \frac{\lambda_n (1 - \lambda_n \cdot V_{DD})}{2 \cdot (\alpha_n + 1)^2} \cdot (v - v_{thn})^{2(\alpha n+1)}$$

$$- b(1 - \lambda_n V_{DD}) \cdot \frac{1}{\alpha n + 1} \cdot (v - v_{thn})^{\alpha n+1} + C23]$$

$$\cdot \exp[\frac{b \cdot \lambda_p}{\alpha_p + 1} \cdot (1 - v - v_{thp})^{\alpha p+1} - \frac{b \cdot \lambda_n}{\alpha_n + 1} \cdot (v - v_{thn})^{\alpha n+1}] \quad (16)$$

C23 is calculated given that $v = v_{sp}$, $V_{out3} = V_{DD} - V_{DSATp}$. When input voltage exceeds $1 - v_{thp}$, PMOS is off, the solutions are the same as the fast input case.

IV. Propagation Delay Estimation

The propagation delay is estimated between the 50% transition points of the input and output waveforms, so $t_{pHL} = tr \cdot v_{0.5} - 0.5tr$, where $v_{0.5}$ is the value of v when $V_{out} = 0.5V_{DD}$.

For the fast input case, the point where output voltage equals to half V_{DD} usually falls in the region 4, which gives:

$$t_{pHL} = \frac{tr}{K_z} \cdot \log(\frac{C34}{\frac{1}{\lambda_n} - \frac{V_{DD}}{2}}) - \frac{tr}{2} \quad (17)$$

In case of half output point falls in other regions or slow input case, the general method is to let $V_{out} = 0.5V_{DD}$ and solve for v.

V. Results

We plot model predictions versus HSPICE simulations for a set of various processes considering channel width, load capacitance and input transition time variation.

In Figure 2, we show the output waveform of a CMOS inverter with input transition time tr = 500ps, load capacitance CL=100fF, transistor channel width Ln=Lp=0.25um. Wp is kept being equal to 2Wn as it is in real design situation, and Wn/Ln is swept from 2 to 6. The results show that the model predicted output transient response matches the simulated output waveform with very good accuracy.

In Figure 3, we plot the propagation delay tpHL versus the load capacitance CL for different Wn/Ln ratios. Ln=Lp=0.18um, tr = 500ps and CL is swept from 50fF to 200fF. The maximum error of the model predicted delay is 1.8% compared to the HSPICE simulated results

		90nm Process							65nm Process						
Wn/Ln	Tr (ps)	100	150	200	250	300	350	400	100	150	200	250	300	350	400
2	Spice(ps)	254	261	268	276	284	292	299	278	285	293	301	309	316	325
	Model(ps)	254.2	261.8	269.4	277	284.5	292.2	299.8	274	282.8	291.5	300.3	309.1	317.8	326.6
	%error	**0.08**	**0.3**	**0.52**	**0.36**	**0.18**	**0.07**	**0.27**	**1.44**	**0.77**	**0.51**	**0.23**	**0.03**	**0.57**	**0.49**
3	Spice(ps)	175	183	191	198	206	214	222	189	196	204	212	220	228	236
	Model(ps)	176.5	184.1	191.7	199.3	207	214.6	222.2	189.2	198	206.8	215.6	224.4	233.2	241.9
	%error	**0.86**	**0.6**	**0.37**	**0.66**	**0.49**	**0.28**	**0.09**	**0.1**	**1.02**	**1.37**	**1.7**	**2**	**2.28**	**2.5**
4	Spice(ps)	136	144	152	160	168	176	184	145	153	161	169	177	186	194
	Model(ps)	137.6	145.3	152.9	160.5	168.1	175.8	182.4	146.6	155.4	164.2	173	181.8	190.5	199
	%error	**1.18**	**0.9**	**0.59**	**0.31**	**0.06**	**0.11**	**0.87**	**1.1**	**1.57**	**1.99**	**2.37**	**2.71**	**2.42**	**2.58**
5	Spice(ps)	113	121	129	137	145	152	158	120	128	136	144	152	160	168
	Model(ps)	114.4	122	129.6	137.3	144.5	150.7	155.4	121.4	130.2	139	147.8	156.6	164.8	171.8
	%error	**1.24**	**0.83**	**0.47**	**0.22**	**0.34**	**0.86**	**1.65**	**1.17**	**1.72**	**2.2**	**2.64**	**3**	**3**	**2.26**

Table I: Propagation Delay vs. Input Rise Time – TSMC 90nm, 65nm

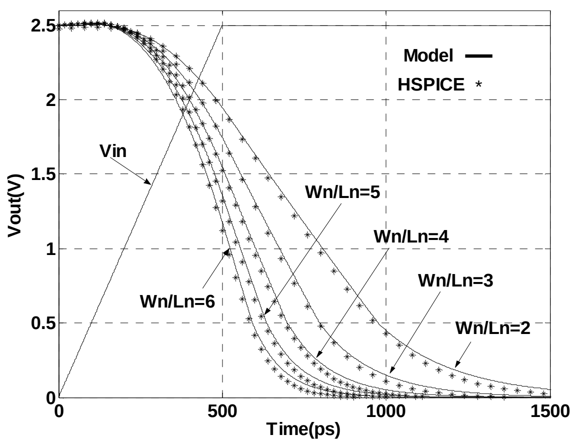

Fig. 2 Output Transient Responses with Different W/L Ratios-TSMC 0.25um

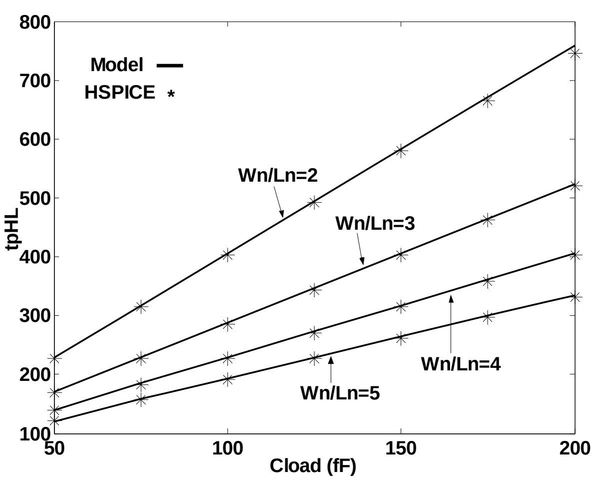

Fig. 3 Propagation Delay vs. Load Capacitance-TSMC 0.18um

Table I shows the computed delay along with HSPICE simulated results for various input rise time, using both TSMC 90nm and 65nm processes. CL is set to be 50fF and input rise time Tr is swept from 100ps to 400ps. Results for different Wn/Ln sets are shown. The maximum error between the computed delay and the HSPICE simulated results is 3%.

VI. CONCLUSION

An accurate analytical model to evaluate the propagation delay of Nano CMOS inverters based on output transient response has been presented. A modified α -power law MOSFET model is used to compute the propagation delay. This model overcomes the over-estimation of linear region current and is demonstrated to be accurate for variable processes. A new methodology to estimate propagation delay by solving non-homogeneous linear differential equation of CMOS inverter is developed. Our work provides a simple and direct analytical expression of output transient response and propagation delay, which avoids the time-consuming numerical procedures. High accuracy results compared to HPSICE simulation (within 3% error) have been demonstrated considering a wide range of transistor sizes, capacitor loads and input transition time, using various processes from TSMC 0.25um to 65nm technology.

Since more complex static gates could be mapped to equivalent inverters [7], this model is valuable in providing an accurate yet fast way to estimate critical path delay in real design.

REFERENCES

[1] T. Sakurai and A.R. Newton, “Delay analysis of series-connected MOSFET circuits,” IEEE J. Solid-State Circuits, vol. 26, pp. 122-131, Feb. 1991.

[2] 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, pp. 302-306, Feb. 1998.

[3] A.A. Hamoui and N.C. Rumin, “An analytical model for current, delay, and power analysis of submicron CMOS Logic circuits,” IEEE Trans. Circuits and Systems II, vol. 47, pp. 999-1007, Oct. 2000.

[4] J.L. Rossello, and J. Segura, “An analytical charge-based compact delay model for submicrometer CMOS inverters,” IEEE Trans. Circuits and Systems I, vol. 51, pp. 1301-1311, July 2004.

[5] T. Sakurai and A.R. Newton, “Alpha-Power law MOSFET model and its application to CMOS inverter delay and other formulas,” IEEE J. Solid-State Circuits, vol. 25, pp. 584-594, April 1990.

[6] J. M. Rabaey, Digital Integrated Circuits: A Design Perspective. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[7] A. Nabavi-Lishi and N.C. Rumin, “Inverter models of CMOS gates for supply current and delay evaluation,” ,” IEEE Trans. Computer-Aided Design, vol. 13, pp. 1271-1279, Oct. 1994.