CMOS Delay and Power Model Equations for Simultaneous Transistor and Interconnect Wire Analysis and Optimization

SANGHO LEE^{a,*} and EDWIN W. GREENEICH^b

^aAgere Systems, 7777 Center Avenue #300, Huntington Beach, CA 92647, USA; ^bElectrical Department, Arizona State University, Tempe, AZ 85287-5706, USA

(Received 15 March 2001; Revised 30 January 2002)

Efficient, generalized delay and power equations are proposed for large scale CMOS circuit analysis and optimization achieved by transistor and interconnect wire minimization. The proposed model equations are used to analyze the entire power-delay trade-off with less complexity and faster computation time. New equations can be adopted to perform the optimization of transistor and interconnect wire size concurrently. A single stage CMOS circuit and a clock generation block fabricated in 0.48 μm CMOS process are given as experimental examples.

Keywords: CMOS circuits; VLSI; Transistor sizing; Optimal interconnect wire sizing; Speed power optimization; CAD transistor model

INTRODUCTION

With CMOS digital circuits and digital systems having a few million transistors, the analysis at the transistor level using a circuit simulator requires a powerful computer, long computation time, and a special numerical method to handle convergence problems. If intensive design optimization is needed for CMOS circuits, it can be impossible, because the design optimization process requires the iteration of circuit analysis until the goal of optimization is reached. It is required to have the best of two simulation approaches: that is, the high accuracy of circuit simulation at the transistor level and the fast speed of logic simulation at the gate level. The unified equations for transistor and interconnect wire are needed to satisfy design requirements simultaneously using optimization tools such as [1–3] for transistor size, [4] only for interconnect wire sizing and [5] for gate sizing and buffer insertion. The model formulation section derives the delay and power equations using the distributed Elmore delay model [6], the multilevel empirical interconnect model [7] and input clock delay model [8]. In the model parameter extraction section, the method to extract key parameters from MOSFET BSIM model is introduced.

Two examples are presented in the experimental results section. Final results are discussed in the conclusion.

MODEL FORMULATION

Preliminaries

Two resistance parameters (PMOS and NMOS channel resistance) and four capacitance parameters (drain capacitance, source capacitance, PMOS gate capacitance, and NMOS gate capacitance) are used in the model equation formulation. Resistance and capacitance parameters are added into the delay and power model equations to consider the interconnect wire effect. The model parameters and definitions used in the derivation of the delay and power model equations are explained below. In this paper, we use $R_{p(p)}$ and $R_{n(p)}$ to denote PMOS and NMOS channel resistance. Let $C_{dp(p)}$, $C_{dn(p)}$, $C_{dps(p)}$, $C_{dns(p)}$, $C_{sp(p)}$, $C_{sn(p)}$, $C_{sps(p)}$, $C_{sns(p)}$, $C_{n(p)}$, $C_{gp(p)}$, $C_{gn(p)}$, and $C_{g(p)}$ denote PMOS drain capacitance, NMOS drain capacitance, PMOS drain sum capacitance, NMOS drain sum capacitance, PMOS source capacitance, NMOS source capacitance, PMOS source sum capacitance, NMOS source sum capacitance, sum of C_{dn} and C_{sn} ,

*Corresponding author.

FIGURE 1 Inverter delay and power model having one interconnect metal layer and output load.

PMOS gate capacitance, NMOS gate capacitance, gate sum capacitance of C_{gp} and C_{gn} , respectively, where (p) is the number of series transistors. C_A is expressed as $pC_{sp(p)} + C_{dn(p)} + C_{w(j)}/2$. For the interconnect wires, $R_{L(j)}$, $C_{w(j)}$, and $C_{N(j)}$ are used to denote wire routing resistance, wire routing capacitance, and branch capacitance in the metal routing path where $N_{(j)}$ is the number of branches in the metal routing path. $C_{L(j)}$ is the wire routing sum capacitance having the following expression, $C_{w(j)}/2 + C_{w(i+1)}/2 + C_{N(j)}$. The transistor channel resistance is expressed as $R_{n(or p)k} = A_{n(or p)k}/W_{n(or p)k}$ where $A_{n(or p)k}$ is NMOS (or PMOS) unit resistance. The sum $(C_{dp} + C_{sp})$ is $F_{1k}W_{n(or p)k}$ where F_{1k} is the unit sum $(C_{dp} + C_{sp})$ capacitance. The interconnect resistance and capacitance having the interconnect wire information are given as $R_{L(j)k} = D_{(j)k}/W_{w(j)k}$ where $D_{(j)}$ is the unit-width wire resistance, and $C_{w(j)k} = E_{(j)k}W_{w(j)k}$ where $E_{(j)}$ is the unit-width wire capacitance. The load capacitance at the end node is expressed as $C_{L(j)k} = E_{(j)k}/2W_{w(j)k} + C_{L(load)}$. When $C_{sp1k} = C_{sp2k} = \dots = C_{sp(p)k}$ and $C_{dn1k} = C_{dn2k} = \dots = C_{dn(p)k}$, the total capacitance at node A shown in

Fig. 2, denoted as $C_{An(or p)}$, is expressed as follows: $C_{An(or p)k} = J_k W_{n(or p)k} + K_k W_{p(or n)k} + 1/2 E_{1k} W_{w1k}$. The model parameters defined in this section are obtained by the method described in the parameter extraction section.

Delay Model Equations

Figure 1 shows the single input delay model including the interconnect wire. Rise and fall times are estimated by a useful method of RC chain approximation. RC approximation allows us to compute the delay time with a reasonable accuracy. This RC approximation is derived by the Elmore delay model. Modeling techniques by resistance and capacitance approximation are very useful in analyzing VLSI circuits with less computation time. It is extremely useful to optimize the design parameters where the model equations have non-linear functions with many variables. The average delay time is obtained by taking an average of the rise and fall delay times. The delay of a logic inverter can be expressed as

FIGURE 2 NAND delay and power models having N-inputs and J interconnect wires.

FIGURE 3 M-stage delay model including N-input gate delay.

$T_{\text{dav}} = (T_{\text{dr}} + T_{\text{df}})/2 = 0.5\{R_{p1}(C_A + C_{L1}) + R_{n1}(C_A + C_{L1}) + 2R_{L1}C_L\}$ where $C_A = C_{sp1} + C_{dn1} + C_{w1}/2$ and $C_{L1} = C_{w1}/2 + C_L$. The delay model equation for two inputs (NAND or NOR) is similar to the inverter delay model except that it has additional resistance and capacitance.

$$\begin{aligned} T_{\text{dav}} &= (T_{\text{dr}} + T_{\text{df}})/2 \\ &= 0.5\{R_{p1}(C_1 + C_A + C_{L1}) + R_{n2}(C_1 + C_A + C_{L1}) \\ &\quad + R_{n1}(C_1 + C_A + C_{L1}) + 2R_L C_{L1}\} \end{aligned}$$

where $C_A = 2C_{sp1} + C_{dn1} + C_{w1}/2$, $C_{L1} = C_{w1}/2 + C_L$.

The general rise time with p inputs and j interconnect wires is approximated using the rise model shown in Fig. 2,

$$\begin{aligned} T_{\text{dr}}(\text{NAND}) &= R_{p1}[C_{(p-1)} + \dots + C_2 + C_1 + C_A + C_{L1} \\ &\quad + C_{L2} + \dots + C_{L(j)}] + R_{n1}[C_{(p-1)} + \dots \\ &\quad + C_2 + C_1] + R_{n2}[C_{(p-1)} + \dots + C_3 \\ &\quad + C_2] + \dots + R_{n(p-1)}C_{(p-1)} + R_{L1}[C_{L1} \\ &\quad + C_{L2} + \dots + C_{L(j)}] + R_{L2}[C_{L2} + C_{L3} \\ &\quad + \dots + C_{L(j)}] + \dots + R_{L(j)}C_{L(j)} \end{aligned}$$

The general equation for fall delay time is expressed as

$$\begin{aligned} T_{\text{df}}(\text{NAND}) &= R_{n(p)}[C_{(p-1)} + \dots + C_2 + C_1 + C_A + C_{L1} \\ &\quad + \dots + C_{L(j)}] + R_{n(p-1)}[C_{(p-2)} + \dots \\ &\quad + C_2 + C_1 + C_A + C_{L1} + \dots + C_{L(j)}] \\ &\quad + R_{n(p-2)}[C_{(p-3)} + \dots + C_2 + C_1 + C_A \\ &\quad + C_{L1} + \dots + C_{L(j)}] + \dots + R_{n1}[C_A \\ &\quad + C_{L1} + \dots + C_{L(j)}] + R_{L1}[C_{L1} + C_{L2} \\ &\quad + \dots + C_{L(j)}] + R_{L2}[C_{L2} + C_{L3} + \dots \\ &\quad + C_{L(j)}] + \dots + R_{L(j)}C_{L(j)} \end{aligned}$$

By taking the average of rise and fall delay times the generalized average delay with p inputs and j interconnect wires is given by

$$\begin{aligned} T_{\text{dav_NAND(or NOR)}} &= 0.5\{R_{p1}(\text{or } n1)[C_{(p-1)} + \dots + C_2 \\ &\quad + C_1 + C_A + C_{L1} + C_{L2} + \dots + C_{L(j)}] \\ &\quad + 2R_{n(\text{or } p)}[C_{(p-1)} + \dots + C_2 + C_1] + 2R_{n(\text{or } p)}[C_{(p-1)} \\ &\quad + \dots + C_3 + C_2] + \dots + 2R_{n(\text{or } p)}C_{(p-1)} \\ &\quad + pR_{n(\text{or } p)}[C_A + C_{L1} + C_{L2} + \dots + C_{L(j)}] \\ &\quad + 2R_{L1}[C_{L1} + \dots + C_{L(j-1)} + C_{L(j)}] \\ &\quad + 2R_{L2}[C_{L2} + \dots + C_{L(j-1)} + C_{L(j)}] \\ &\quad + \dots + 2R_{L(j)}C_{L(j)}\} \end{aligned} \quad (1)$$

By substituting the model parameters defined in 2.1 Preliminaries into Eq. (1), the delay equation is expressed as a function of transistor and interconnect wire size. If one metal wire is used between drivers, $F_{(p-1)} = F_{(p-2)} = \dots = F_1$, and $C_{(p-1)} = C_{(p-2)} = \dots = C_1$, (1) is further simplified as

$$\begin{aligned} T_{\text{dav}} &= a_1X_1X_2^{-1} + a_2X_3X_2^{-1} + a_3X_5X_2^{-1} \\ &\quad + a_4X_4X_2^{-1} + a_5X_2X_1^{-1} + a_6X_3X_1^{-1} \\ &\quad + a_7X_5X_1^{-1} + a_8X_4X_1^{-1} + a_9X_5X_3^{-1} \\ &\quad + a_{10}X_4X_3^{-1} + a_{11} \end{aligned} \quad (2)$$

where $W_n = X_1$, $W_p = X_2$, $W_w = X_3$, $W_{n2} = X_4$, $W_{p2} = X_5$, $a_1 = (A_p(p-1)F + A_pK)/2$, $a_2 = A_pE_1/2$, $a_3 = A_pB_p/2$, $a_4 = A_pB_n/2$, $a_5 = pA_nJ/2$, $a_6 = pA_nE_1/2$, $a_7 = pA_nB_p/2$, $a_8 = pA_nB_n/2$, $a_9 = D_1B_p$, $a_{10} = D_1B_n$, $a_{11} = 1/2[A_pJ + A_nF(p-1)p + pA_nK + D_1E_1]$.

Next, the M-stage and N-input delay model equation including the interconnect wire effect is derived. The delay equation is a general equation, which is used for any large CMOS circuit. The model of the M-stage and N-input delay is shown in Fig. 3. When the effect of input

transition is incorporated, the total stage delay along the critical path including gate delay and interconnect wire delay is given as

$$\begin{aligned} T_{D(M_stage_delay)} &= SD_1 + SD_2 + SD_3 + SD_4 \cdots + SD_m \\ &= (1 + \alpha + \alpha^2 + \cdots + \alpha^{m-1})\Delta_c \\ &\quad + (1 + \alpha + \alpha^2 + \cdots + \alpha^{m-2})\Delta_2 \quad (3) \\ &\quad + (1 + \alpha + \alpha^2 + \cdots + \alpha^{m-3})\Delta_3 \\ &\quad + \cdots + \alpha^0 \Delta_m \end{aligned}$$

where $\alpha = (1 + 2|V_T|/V_{DD})/3$, T_r is the clock slope, and Δ_m is the stage delay at m stage.

$$SD_1 = \Delta_1 + \alpha_1 T_r = \Delta_c$$

$$SD_2 = \Delta_2 + \alpha \Delta_c$$

$$SD_3 = \Delta_3 + \alpha \Delta_2 + \alpha^2 \Delta_c$$

$$SD_4 = \Delta_4 + \alpha \Delta_3 + \alpha^2 \Delta_2 + \alpha^3 \Delta_c$$

$$SD_m = \Delta_m + \alpha(SD_{(m-1)} - SD_{(m-2)})$$

$$= \Delta_m + \alpha \Delta_{(m-1)} + \alpha^2 \Delta_{(m-2)} + \cdots + \alpha^{m-1} \Delta_c$$

Power Model Equations

Power consumption for CMOS circuits is expressed as the sum of the dynamic power component, the short-circuit power component, and the DC leakage component, given as

$$P_d = \kappa C_L f_p V_{DD}^2 + I_{SC} V_{DD} + I_{leakage} V_{DD} \quad (4)$$

where κ is the switching factor, C_L is the output load capacitance, V_{DD} is the supply voltage, f_p is the operating frequency, I_{SC} is the short-circuit current, and $I_{leakage}$ is the DC leakage current. Dynamic power consumption occurs when the current charges the transistor load capacitance from the supply voltage, and discharges to the ground. The switching factor is related to the transition probability of the input signals. In the case of static logic the switching factor depends on the previous logic state while in dynamic logic it is up to the input signal probability. Another source of power consumption is the direct current flow between the power supply and the ground, which exists for a short time. When the input voltage is in the range $V_m < V_{in} < V_{DD} - V_{tp}$, P and N_{type} transistors are on simultaneously, and then current flows through a direct path between the power supply and the ground. The short circuit current is dependent upon the rise and fall times of the input signal, the threshold voltage, and the ratio of channel-width to channel-length. The average short circuit current can be minimized when

input rise and fall times are much faster than the clock period. There are two types of DC leakage currents: one is the sub-threshold leakage current, and the other is the reverse bias leakage current. The sub-threshold leakage occurs due to carrier diffusion between the source and the drain. The source-drain current in the sub-threshold region is particularly important for the short channel low power circuits, because it affects the standby static power consumption. The second leakage component is the parasitic diode leakage current between the drain (or source) and the substrate. The short-circuit and leakage power can be incorporated in the power model by adding the process dependent parameters, η_1 and η_2 to the total power. In order to get the dynamic power, assume that the rise and fall times of the input signals are much faster than the clock period, and that there are V_t drops in the MOSFET channel resistance series as shown in Fig. 2. After integrating the equation, $P_{d(ave)} = 1/T \int_{t_1}^{t_2} v(t)i(t)dt$, the average dynamic power is given as

$$\begin{aligned} P_{d(ave)} &= 1/T \{ C_1 v_1^2 + \cdots + C_2 v_2^2 + \cdots + C_{(p-1)} v_{(p-1)}^2 \\ &\quad + C_1 v_1^2 + [C_A + C_{L1} + \cdots + C_{Ln}] V_{dd}^2 \} \end{aligned}$$

The transmission of logic one (V_{dd}) is degraded as it passes through the NMOS gate. If considering V_t drop in the transistor series, the average dynamic power is

$$P_{d(ave)} = 1/T V_{dd}^2 \{ C_1 K + C_A + C_{L1} + \cdots + C_{Ln} \}$$

where $K = \sum_{s=1}^p (1 - (p-s)V_t/V_{dd})^2$, $v_1 = V_{dd} - V_t$, $v_2 = V_{dd} - 2V_t, \dots, v_{(p-1)} = V_{dd} - (p-1)V_t$.

The dynamic power is expressed as a function of the transistor size and interconnect wire size.

$$\begin{aligned} P_{d(ave)} &= 1/T V_{dd} V_{dd} \{ (F_c K + K_c) X_1 + K_c X_2 + j E X_3 \\ &\quad + B_p X_5 + B_n X_4 \} \end{aligned}$$

Accordingly, the dynamic power for one stage is given as $P_d = b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5$. The total power model equation is obtained from the dynamic power and two parameters, η_1 and η_2 .

$$\begin{aligned} P_{d(ave)} &= \eta_1 (b_1 + \alpha b_1) X_1 + b_3 X_3 + b_4 X_4 \\ &\quad + b_5 X_5 + \eta_2 \end{aligned} \quad (5)$$

where $W_n = X_1$, $W_p = X_2$, $W_w = X_3$, $W_{n2} = X_4$, $W_{p2} = X_5$, $b_1 = V_{dd}^2 [F_c K + K_c]/T$, $b_2 = V_{dd}^2 J_c/T$, $b_3 = V_{dd}^2 J E_m/T$, $b_4 = V_{dd}^2 B_n/T$, $b_5 = V_{dd}^2 B_p/T$, $\alpha = X_2/X_1$. η_1 and η_2 are process dependent parameters that include leakage and short-circuit power components. The single stage dynamic power is a nonlinear function for the reason that the interconnect capacitance is a nonlinear function of the interconnect wire size. However, it can be approximated by a piece-wise linear function. The M-stage N-input total power is obtained by

summing the power consumption of each stage.

$$P_{\text{total}} = SP_1 + SP_2 + \dots + SP_{(k)}, \quad (6)$$

$$k = 1, 2, 3, \dots, m$$

where $SP_{(m)} = b_{1m}X_{1m} + b_{2m}X_{2m} + b_{3m}X_{3m} + b_{4m}X_{4m} + b_{5m}X_{5m}$ is the total power at m stage.

MODEL PARAMETER EXTRACTION

When the drain current is determined, the channel resistance is calculated by Ohm's law. However, due to the complexity of the current equation, the channel resistance is expressed as a non-linear function with many variables. The exact analytical and semi-analytical transistor models are used for high accuracy in the CMOS circuit simulations. However, when the number of transistors gets larger in CMOS VLSI circuit designs, it is necessary to find a stable solution without numerical convergence problems. Simulations with exact models usually require a long computation time. An empirical model for the channel resistance is obtained using the process based MOSFET models. One such model, useful in approximating the behavior of the voltage and current characteristic of the MOSFET, relates the channel resistance to the gate size. The transistor channel resistance is expressed as $R_{(\text{on})} = V_{\text{ds}}/I_{\text{ds}} \approx A/W$ where the unit resistance, A , is determined by the SPICE simulation. The unit channel resistance is shown in Fig. 4.

The gate capacitance is given by the sum of the oxide (C_{ox}) and semiconductor (C_s) capacitances, and is a non-linear equation. The empirical model parameters for the gate capacitance are obtained by $C = Q_{(t)}/\Delta V$ where the charge, Q , is calculated by the SPICE simulation. The total gate capacitance, $C_{g(\text{tot})}$, is

FIGURE 4 Unit channel resistance [ohm] versus transistor channel width [μm] with strong (ST) and weak (WK) BSIM model.

FIGURE 5 Gate unit capacitance [F] versus gate size [μm] with weak (wk) BSIM model.

comprised of $C_{gn} = B_n W_n$ and $C_{gp} = B_p W_p$ where B_n and B_p are the NMOS and PMOS unit capacitance, respectively. They are shown in Fig. 5.

The unit gate capacitance is used as a basic parameter in the delay and power model equations. The parameters for the drain and source capacitances are also given by the same model used for the gate. The drain and source capacitances are expressed in the transistor size terms, $C_{\text{dsn}} = K_c W_n$ and $C_{\text{dsp}} = J_c W_p$ where K_c and J_c are the drain/source unit capacitance for NMOS and PMOS transistors shown in Fig. 6.

The multi-level interconnect wire resistance, which is used to make the interconnection between the drivers, is expressed as $R_{(\text{tot})} = \rho_{s1}(L_1/W_1) + \rho_{s2}(L_2/W_2) + \dots + \rho_{sn}(L_n/W_n)$ where ρ_s is the sheet resistance. The interconnect resistance is simply obtained by multiplying the sheet resistance by the ratio of the metal length

FIGURE 6 Drain and source unit capacitance [F] vs. width [μm].

FIGURE 7 Interconnect wire capacitance components and geometrical parameters for three metal layers.

to metal width of the interconnect wires. The interconnect resistance can be expressed in terms of the metal wire width, W_w

$$R_{(\text{metal})} = \rho_{s(\text{metal})} L \frac{1}{W_w} \approx D_m \frac{1}{W_w} \quad (7)$$

Capacitance models for multilevel interconnect wiring are quite complex. Three dimensional field simulators can be used to calculate the metal capacitance. However, they are constrained to simulate a few physical configurations and are too complex to handle many different configurations. Empirical modeling

approaches overcome these problems with a reasonable accuracy. Multilevel empirical models [6] are used to calculate the parasitic capacitance of M1, M2 and M3. The approach uses three primitive structures to construct multilevel metal configurations. Three basic capacitances are line-to-line, line-to-ground, and cross-over capacitance.

Figure 7 shows the capacitance components and geometrical parameters. The interconnect capacitance is a non-linear function of the metal width. However, it is expressed as a linear function using a piece-wise linear approximation:

$$C_w \approx E_m W_w \quad (8)$$

where E_m is the unit metal capacitance. Figure 8 shows the M1, M2 and M3 interconnect capacitance per unit area.

EXPERIMENTAL RESULTS

As an example, consider the single stage CMOS circuit having an interconnect routing wire and an output load ($P/N = 40 \mu\text{m}/20 \mu\text{m}$). Delay, power, and power-delay product are plotted as a function of the transistor width shown in Figs. 9–11. Open circle symbols are SPICE simulation data, and open triangle symbols are calculation data from the delay and power equations.

Calculation results of the model equations closely match the SPICE simulations. There is about 3–8% mismatch between the model and SPICE, depending on the transistor size, $W_{n(\text{or } p)}$. The computation time of delay and power by the new model equations is much

FIGURE 8 M1, M2, and M3 interconnect capacitance per unit area versus transistor width.

FIGURE 9 Delay [S] comparison between model and SPICE for a single stage.

faster than the SPICE simulation. A practical clock generation block is presented as the other example to provide comparison data among the model, SPICE, and silicon. Figure 12 shows the critical path of the test clock circuit.

SPICE simulation data (~ 3.3 ns) has less delay than the measured silicon data (~ 3.7 ns). This is because the effect of routing wire RC delay is not included properly in the SPICE simulation model. The SPICE simulations are shown in Fig. 13. Clock delay data (~ 3.859 ns), calculated by the model, is close to the measured data because more accurate interconnect wire models are included in the model equations. A photomicrograph of the fabricated test clock circuit is shown in Fig. 14. The test clock circuit is fabricated in

FIGURE 10 Power [W] comparison between model and SPICE for a single stage.

FIGURE 11 Power-delay product [S·W] comparison between model and SPICE for a single stage.

0.48 μ m CMOS triple metal process technology. Figure 15 shows the waveforms generated from the XY data of a digital oscilloscope.

When the effect of the transition time is incorporated, the delay along the clock path is calculated by

$$\begin{aligned} T_{\text{tot}} &= T_{ad} + W_1 T_{bd} + W_2 T_{cd} + W_3 T_{dd} + W_4 T_{ed} \\ &\quad + W_5 T_{fd} + W_6 T_{gd} + W_7 T_{hd} + W_8 T_{id} + W_9 T_{jd} \\ &\quad + W_{10} T_{kd} + W_{10} T_{ld} \\ &= 3.859 \text{ ns} \end{aligned}$$

where $W_k = \sum_{i=0}^{n-k} [1/3(1 + 2|V_i|/V_{DD})]^i$

$$\begin{aligned} T_{ad} &= R_{am1}(C_{am1} + C_{am2} + 62.5B_p + 93.75B_n) \\ &\quad + R_{am2}(C_{am2} + 62.5B_p + 93.75B_n), \end{aligned}$$

$$\begin{aligned} T_{bd} &= A_n(82.5K_c + 60B_p + 95B_n)/20, \quad T_{cd} = A_p(55K_c + 126B_p + 18B_n)/40, \\ T_{dd} &= A_n(9K_c + 20B_p + 10B_n)/3, \\ T_{ed} &= A_p(30K_c + 50B_n)/20, \quad T_{fd} = A_n(130K_c + 305B_p + 125B_n)/50, \\ T_{gd} &= A_p(310K_c + 310B_p + 255B_n + C_{gm2} + C_{gm3} + 15B_p + 15B_p + 45B_n)/220 + R_{gm3} (C_{gm2} + C_{gm3}/2 + 15B_p + 45B_n) + R_{gm2} (C_{gm2}/2 + 115B_p + 45B_n), \\ T_{hd} &= A_n(90K_c + 70B_p + 50B_n)/45, \quad T_{id} = A_p(90K_c + C_{im2} + C_{im3} + 172B_p + 111B_n)/60 + R_{im3} (C_{im2} + C_{im3}/2 + 172B_p + 111B_n) + R_{im2} (C_{im2}/2 + 172B_p + 111B_n), \\ T_{jd} &= A_n(50K_c + 90K_c + 20B_p + 10B_n)/30 + A_n (45K_c + 20B_p + 10B_n)/60, \quad T_{kd} = A_p(30K_c + 30B_p + 25B_n)/20, \\ T_{ld} &= A_n(20K_c + 10B_p + 5B_n)/10, \quad V_t = 0.7, \quad V_{dd} = 2.9, \quad A_p = 1.73e + 4, \quad A_n = 9.56e + 3, \\ B_p &= 1.93e - 15, \quad B_n = 1.77e - 15, \quad K_c = 8.96e - 16, \quad R_{am1} = 46.2, \quad C_{am1} = 4.124e - 14, \quad R_{am2} = 8.0, \\ C_{am2} &= 2.852e - 14, \quad R_{gm2} = 28.0, \quad C_{gm2} = 9.937e - 14, \quad R_{gm3} = 69.3, \quad C_{gm3} = 2.729e - 13, \quad R_{im2} = 3.0, \quad C_{im2} = 1.081e - 14, \\ R_{im3} &= 277.8 \text{ and } C_{im3} = 1.093e - 2. \end{aligned}$$

FIGURE 12 Critical clock generation path.

FIGURE 13 SPICE simulation outputs.

CONCLUSION

Two experimental examples show that the calculated data using the new delay and power equations closely match the result of SPICE. However, a small discrepancy between the model equations and SPICE simulation results from a linear and piece-wise linear approximation of the model parameters extracted from the BSIM transistor models. This implies that the accuracy of the new equations can be improved by adding more model parameters, depending on the range of the transistor and interconnect wire size. The computation time of the delay-power equations is much faster than the SPICE simulation time, which depends on the size of circuits. The model equations can be used for optimizing the

FIGURE 14 Photomicrograph of the fabricated clock generation block using 0.48 μm CMOS triple metal process.

FIGURE 15 Measured outputs generated from digital oscilloscope.

TABLE I Summary of model parameters

Parameter	Model	Value	Remarks
A_n	Slow (-3σ) –Fast $(+3\sigma)$	$1.21e+04-9.32e+03$ [ohms/ μm]	NMOS unit resistance
A_p	Slow (-3σ) –Fast $(+3\sigma)$	$2.14e+04-1.73e+04$ [ohms/ μm]	PMOS unit resistance
B_n	Slow (-3σ) –Fast $(+3\sigma)$	$1.77e-15-1.65e-15$ [F/ μm]	N unit gate capacitance
B_p	Slow (-3σ) –Fast $(+3\sigma)$	$1.93e-15-1.90e-15$ [F/ μm]	P unit gate capacitance
F_c	Slow (-3σ) –Fast $(+3\sigma)$	$1.63e-15-2.07e-15$ [F/ μm]	Unit $C_s + C_d$
K_c	Slow (-3σ) –Fast $(+3\sigma)$	$8.96e-16-9.42e-16$ [F/ μm]	NMOS unit C_d
J_c	Slow (-3σ) –Fast $(+3\sigma)$	$7.07e-16-1.13e-15$ [F/ μm]	PMOS unit C_s
D_{m1}		$44.0e-02$ [ohms]	Unit M1 resistance
D_{m2}		$1.00e-01$ [ohms]	Unit M2 resistance
D_{m3}		$6.00e-02$ [ohms]	Unit M3 resistance
E_{m1}	$W \leq 1 \mu\text{m}$	$2.80e-16$ [F/ μm]	Unit M1 capacitance
	$2 \mu\text{m} \leq W \leq 5 \mu\text{m}$	$2.03e-16$ [F/ μm]	Unit M1 capacitance
	$6 \mu\text{m} \leq W \leq 8 \mu\text{m}$	$1.81e-16$ [F/ μm]	Unit M1 capacitance
	$9 \mu\text{m} \leq W$	$1.76e-16$ [F/ μm]	Unit M1 capacitance
E_{m2}	$W \leq 1 \mu\text{m}$	$2.60e-16$ [F/ μm]	Unit M2 capacitance
	$2 \mu\text{m} \leq W \leq 5 \mu\text{m}$	$1.86e-16$ [F/ μm]	Unit M2 capacitance
	$6 \mu\text{m} \leq W \leq 8 \mu\text{m}$	$1.65e-16$ [F/ μm]	Unit M2 capacitance
	$9 \mu\text{m} \leq W$	$1.60e-16$ [F/ μm]	Unit M2 capacitance
E_{m3}	$W \leq 1 \mu\text{m}$	$1.74e-16$ [F/ μm]	Unit M3 capacitance
	$2 \mu\text{m} \leq W \leq 5 \mu\text{m}$	$1.12e-16$ [F/ μm]	Unit M3 capacitance
	$6 \mu\text{m} \leq W \leq 8 \mu\text{m}$	$9.47e-17$ [F/ μm]	Unit M3 capacitance
	$9 \mu\text{m} \leq W$	$9.08e-17$ [F/ μm]	Unit M3 capacitance

power-delay trade-off. They are also used to analyze the static gate and the interconnect wire delay simultaneously. For future work, an optimization algorithm to handle large design parameters expressed by delay-power equations needs to be developed as well as the accuracy improvement of the model parameters (Table I).

References

- [1] Coudert, Olivier (1997) "Gate sizing for constrained delay, power, area optimization", *IEEE Trans. VLSI* 5(4), 465–472.
- [2] Berkelaar, Michel R.C.M. and Buurman, Pim H.W. (1996) "Computing the entire active area/power consumption versus delay tradeoff curve for gate sizing with a piece-wise linear simulator", *IEEE Trans. CAD* 15(11), 1424–1434.
- [3] Sapatnekar, S.S., Rao, V.B., Vidya, P.M. and Kang, S. (1993) "An exact solution to the transistor sizing problem for CMOS circuits using convex optimization", *IEEE Trans. CAD* 12(11), 1621–1634.
- [4] Sapatnekar, Sachin S. (1996) "Wire sizing as a convex optimization problem: exploring the area-delay tradeoff", *IEEE Trans. CAD* 15(8), 1001–1011.
- [5] Lowe, Kerry S. and Gulak, P. Glenn (1998) "A joint gate sizing and buffer insertion method for optimizing delay and power in CMOS and BiCMOS combination logic", *IEEE Trans CAD* 17(5), 419–434.
- [6] Elmore, W.C. (1948) "The transient response of damped linear networks with particular regard to wide-band amplifiers", *J. Appl. Phys.* 19-1(January), 55–63.
- [7] Chren, Jue-Hsien, Huang, Jean, Li, Ping-Chung and Yang, Ping (1992) "Multilevel metal capacitance models for CAD design synthesis systems", *IEEE Electron Device Lett.* 13(1), 32–34.

- [8] Hedenstierna, Nils and Jeppson, Kjell O. (1987) "CMOS circuit speed and buffer optimization", *IEEE Trans. CAD CAD-6*(2), 270–281.

Sangho Lee received the B.S. degree in electrical engineering from Hanyang University, Korea, and M.S. and Ph.D. degrees in electrical engineering from the Arizona State University, Tempe, Arizona, in 1991 and 1998, respectively. From 1995 to 1999, he was with Texas Instruments and Micron Technologies where he was engaged in Synchronous and DDR memory design. Since 1999, he has worked on a satellite CD radio project with Agere Systems, Huntington Beach, CA. His research interests include VLSI circuits in communications, low power design optimization, and mixed signal circuits in wireless communications.

Edwin W. Greeneich received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of California, Berkeley. Since 1982, he has been on the faculty at Arizona State University in the department of electrical engineering. Dr. Greeneich is the author of *Analog Integrated Circuits* (Chapman and Hall, 1996), a co-author of *Ultra Large Scale Integrated Microelectronics* (Prentice Hall, 1988), and a contributor to *The Circuits and Filters Handbook* (CRC Press, 1995). He has also published dozens of articles in technical journals. His current research interests are with high-frequency digital and analog integrated circuits using bipolar and MOSFET technologies.