DELAY AND POWER ESTIMATION FOR A CMOS INVERTER DRIVING RC INTERCONNECT LOADS

S. Nikolaidis, A. Chatzigeorgiou [1] and *E.D. Kyriakis-Bitzaros* [2]

Department of Physics, [1]Computer Science Department

Aristotle University of Thessaloniki

54006 Thessaloniki, Greece

[2]Institute of Microelectronics, NCSR "Demokritos",

15310 Agia Paraskevi, Greece

ABSTRACT

The resistive-capacitive behavior of long interconnects which are driven by CMOS gates is analyzed in this paper. The analysis is based on the π-model of an RC load and is developed for submicron devices. Accurate and analytical expressions for the output voltage waveform, the propagation delay and the short circuit power dissipation are derived by solving the system of differential equations which describe the behavior of the circuit. The effect of the coupling capacitance between input and output and that of short circuit current are also incorporated in the proposed model. The calculated propagation delay and short circuit power dissipation are in very good agreement with SPICE simulations.

1. INTRODUCTION

As the minimum feature sizes for integrated circuits scale downwards, the resistive component of the interconnect loads becomes comparable to the gate output impedance and a single capacitor is no longer a valid gate load model. More accurate load models have to be used for taking into account the increased role of the resistance in the determination of the load behavior and consequently the propagation delay of the driving CMOS gates.

Much research effort has been devoted and very powerful methods have been proposed during the last years for modeling CMOS gates driving simple capacitive loads [1], [2]. Expressions for the propagation delay of CMOS gates driving RC loads have also been derived [3], [4], [5] but they present significantly lower accuracy mainly because they are based on simplified assumptions for the transistor operation and use simple models for the representation of the interconnect loads. Recently, in [6] an analytical method with emphasis on the short-circuit power dissipation has been presented for an inverter driving an RC π load.

In order to find analytical expressions for the propagation delay and the output waveform shape, an interconnect load may be modeled in different ways [7]. Such an expression for the propagation delay of a load modeled simply by a resistor in series with a capacitor, was derived in [3]. However, the driving transistor was considered to operate always in linear mode, the influence of the short circuit current was ignored and the simplified case of step input was examined, thus resulting in limited accuracy.

In [4] an effective capacitance in order to replace the RC output load was calculated by an iteration procedure based on simplified assumptions for the shape of the output response. The real output waveform was approximated by the charging/discharging of the effective capacitance until the time point where the output voltage becomes equal to $V_{DD}/2$. Capturing of the remaining portion of the output response is achieved by a simple resistive model.

A time varying Thevenin equivalent model was proposed in [5] for the estimation of the gate delays. The gate was replaced by an equivalent circuit model composed of a linear voltage source and a linear resistor where their values were determined by using empirical factors thus reducing the accuracy, especially for submicron technologies.

A good approximation for an interconnect load is obtained with the π model, achieving an accuracy better than 3% in delay calculations [7]. In this way analytical expressions for the propagation delay and the output waveform can be found if the load is replaced by its π equivalent and the corresponding system equations are solved. This is the method followed in this paper in order to capture with higher accuracy the performance of CMOS gates driving RC interconnect loads.

2. TRANSIENT RESPONSE ANALYSIS

A circuit composed of an inverter driving an equivalent π-model is considered, where the gate-to-drain coupling capacitance, C_m, is taken into account (Fig. 1). The α-power law model [1], which considers the velocity saturation effect of short channel devices, is used for the transistor current representation :

$$I_D = \begin{cases} 0 & V_{GS} \le V_{TO} : \textit{cutoff region} \\ k_l \left(V_{GS} - V_{TO}\right)^{a/2} V_{DS} & V_{DS} < V_{D-SAT} : \textit{linear region} \\ k_s \left(V_{GS} - V_{TO}\right)^{a} & V_{DS} \ge V_{D-SAT} : \textit{saturation region} \end{cases} \quad (1)$$

where $V_{D\text{-}SAT}$ is the drain saturation voltage [1], k_l , k_s are the transconductance parameters, a is the velocity saturation index and V_{TO} is the zero bias threshold voltage.

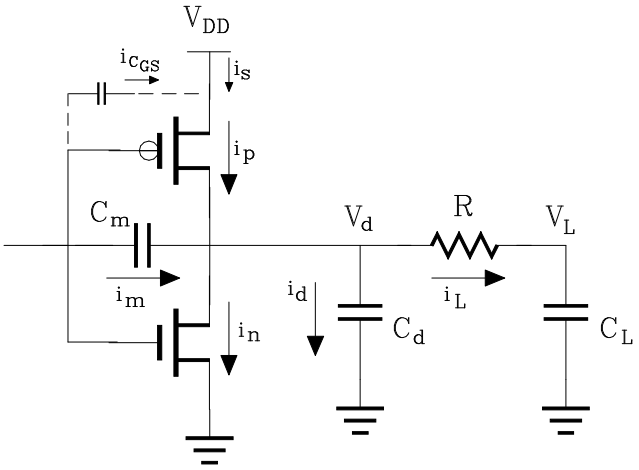

Fig. 1 Inverter driving the π-model of an RC load

A rising ramp input with transition time τ is applied to the transistor gates. The case for a falling ramp is symmetrical. The differential equations that describe the operation of the circuit in Fig. 1 are obtained by applying the Kirchhoff's voltage law in the loop of the π subcircuit :

$$V_d = V_L + V_R \Rightarrow V_d = V_L + RC_L \frac{dV_L}{dt} \quad (2)$$

and Kirchhoff's current law in the transistors drain node using eq. (2) :

$$i_n + i_d + i_L - i_m - i_p = 0 \Rightarrow$$

$$\frac{dV_L}{dt} + C_2 \frac{d^2V_L}{dt^2} - C_3 \frac{dV_{in}}{dt} + \frac{i_n}{C_1} - \frac{i_p}{C_1} = 0 \quad (3)$$

where :

$$C_1 = C_L + C_d + C_m, \quad C_2 = \frac{RC_L(C_d + C_m)}{C_1}, \quad C_3 = \frac{C_m}{C_1}$$

In order for the above differential equation to be solved analytically, the parasitic current through the pMOS transistor is initially considered negligible. At the end of the analysis its influence on the output response will be determined.

Two main cases for input ramps are considered : for fast (slow) inputs the nMOS device is in saturation (in the linear region) when the input voltage reaches its final value. In order to obtain the output voltage expression analytically, four regions of operation are considered.

Fast input ramps

Region 1 ($0<t<t_1$). The nMOS transistor is cut-off and differential equation (3) becomes :

$$\frac{dV_L}{dt} + C_2 \frac{d^2V_L}{dt^2} - C_4 = 0 \quad (4)$$

with initial conditions $V_L(0) = V_{DD}, \frac{dV_L}{dt}(0) = 0$ and $C_4 = C_3 V_{DD} / \tau$. The output waveform expression is given by :

$$V_L(t) = V_{DD} + C_4 t - C_2 C_4 \left(1 - e^{-\frac{t}{C_2}}\right) \quad (5)$$

This expression describes the small overshoot of the output waveform due to the coupling capacitance C_m. Generally for the case of driving long interconnection lines, since $C_m << C_d$ the overshoot value is almost negligible and V_{out} can be considered equal to V_{DD} without significant error in this region. This region extends until time $t_1 = \frac{V_{TO}\tau}{V_{DD}}$ where $V_{in}=V_{TO}$.

Region 2 ($t_1<t<\tau$). The nMOS device operates in saturation and the input signal is still in transition. Equation (3) becomes :

$$\frac{dV_L}{dt} + C_2 \frac{d^2V_L}{dt^2} - C_3 \frac{dV_{in}}{dt} + \frac{k_s}{C_1}\left(\frac{V_{DD}}{\tau} t - V_{TO}\right)^a = 0 \quad (6)$$

which can not be solved analytically. In order to obtain an analytical expression for the output waveform in this region, the current term is approximated by a second order Taylor series at $t=\tau/2$ where $V_{in}=V_{DD}/2$ with excellent accuracy (error < 1.5%) as $\frac{i_n}{C_1} = A_0 + A_1 t + A_2 t^2$. The differential equation is solved resulting in the following expression for the output waveform :

$$V_L(t) = C[1] + C_5 t + C_6 t^2 + C_7 t^3 + C[2] e^{-\frac{t}{C_2}} \quad (7)$$

where: $C_5 = C_4 - A_0 + 2C_2\left(\frac{A_1}{2} - C_2 A_2\right)$,

$C_6 = C_2 A_2 - \frac{A_1}{2}$, $C_7 = -\frac{A_2}{3}$

and C[1], C[2] are the integration constants.

Region 3 ($\tau<t<t_2$). The input has reached its final value and the nMOS transistor is still in saturation. Equation (3) becomes :

$$\frac{dV_L}{dt} + C_2 \frac{d^2V_L}{dt^2} + \frac{k_s}{C_1}(V_{DD} - V_{TO})^a = 0 \quad (8)$$

resulting in:

$$V_L(t) = C[3] - K_1 t + C[4] e^{-\frac{t}{C_2}} \quad (9)$$

where $K_1 = \frac{k_s}{C_1}(V_{DD} - V_{TO})^a$ and C[3], C[4] are the integration constants.

This region extends until time t_2 when the nMOS transistor exits saturation. The time point t_2 is calculated by the equation:

$$V_d(t_2) = V_L(t_2) + C_L R \frac{dV_L}{dt}(t_2) = V_{D-SATN}(t_2) \quad (10)$$

which can be solved without any approximation. V_{D-SATN} is the drain saturation voltage of the nMOS device.

Region 4 ($t>t_2$). The nMOS transistor operates in linear mode and the solution of equation (3) becomes :

$$V_L(t) = C[5] e^{-\frac{1+\sqrt{1-4C_8K_4}}{2C_8}t} + C[6] e^{-\frac{1-\sqrt{1-4C_8K_4}}{2C_8}t} \quad (11)$$

where $K_2 = \frac{k_1}{C_1}(V_{DD} - V_{TO})^{\frac{\alpha}{2}}$, $K_3 = 1 + K_2 C_L R$, $C_8 = \frac{C_2}{K_3}$,

$K_4 = K_2 / K_3$

Slow input ramps

The operating conditions of the structure in regions 1 and 2 are the same as for fast inputs, however region 2 extends from time t_1 to time t_2, where $t_2<\tau$.

Region 3 ($t_2<t<\tilde{t}$). The nMOS transistor operates in linear mode while the input is still a ramp. The differential equation describing the output evolution in this region is given by:

$$\frac{dV_L}{dt}+C_2\frac{d^2V_L}{dt^2}-C_4+\frac{k_l}{C_1}\left(V_{in}-V_{TO}\right)^{\frac{a}{2}}\left(V_L+RC_L\frac{dV_L}{dt}\right)=0 \quad (12)$$

which can not be solved analytically. For this reason, V_{in} is replaced by its average value $\tilde{V}_{in}=\frac{V_{in}(t_2)+V_{DD}}{2}$. This is a valid approximation since for most of the practical cases the duration of this region is very small and thus V_{in} takes values very close to that average value.

According to this, the solution of eq. (3) is:

$$V_L(t)=-\frac{C_{10}}{K_6}+C[7]\,e^{-\frac{1+\sqrt{1-4C_9K_6}}{2C_9}t}+C[8]\,e^{-\frac{1-\sqrt{1-4C_9K_6}}{2C_9}t} \quad (13)$$

where $K_5=\frac{k_l}{C_1}\left(\tilde{V}_{in}-V_{TO}\right)^{a/2}$, $K_6=\frac{K_5}{1+K_5RC_L}$,

$C_9=\frac{C_2}{1+K_5RC_L}$, $C_{10}=-\frac{C_4}{1+K_5RC_L}$

Region 4 is solved exacty as for fast inputs.

A comparison of the output response, V_L, calculated by the proposed method with that derived by SPICE simulations for fast and slow inputs is given in Fig. 2 for an HP 0.5 μm technology, W_n=30 μm and W_p=50 μm. The accuracy of the proposed analysis is obvious.

3. THE EFFECT OF THE SHORT-CIRCUIT CURRENT ON PROPAGATION DELAY

In the above analysis, the current through the pMOS transistor, called short circuit current, was considered negligible. Generally, this is a valid assumption because the capacitive load in long interconnection lines is large enough so that the output voltage doesn't change significantly until the time the pMOS transistor becomes off. This means that the drain-to-source voltage of the pMOS transistor remains small and its current also takes small values. Consequently, ignoring the short-circuit current, in order to simplify the mathematical analysis, does not have any significant effect on the accuracy of the presented analysis. However, a method for taking into account its influence in the estimation of the propagation delay of gates driving long interconnections is presented.

The short-circuit current through the pMOS transistor exists in the interval [t_{ov}, t_p] where t_{ov} is the time where the voltage overshoot at the output of the inverter ceases. That is because during the voltage overshoot, the pMOS current is flowing towards V_{DD} and thus no current path exists between V_{DD} and ground. Time t_{ov} can be calculated by setting the voltage expression for the inverter output, V_d, in region 2 equal to V_{DD}. t_p is the time when the pMOS transistor turns off (when $V_{in}=V_{DD}-|V_{TP}|$). The existence of the pMOS current results in a decrease of the discharging current and thus in an increase of the propagation delay. It acts like an amount of charge Q_e initially stored in the output node and which has to be removed through the nMOS transistor. Consequently, the equivalent charge can be calculated by integrating the current of the pMOS device from time t_{ov} to time t_p. Considering that the pMOS transistor operates for half of the interval [t_{ov}, t_p] in linear mode and that the current waveform is symmetrical around the middle of this interval [6], Q_e can be calculated as :

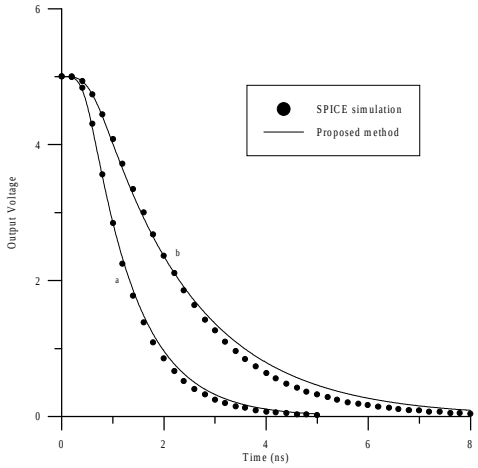

Fig. 2 Output waveform comparison between simulated and calculated values for (a) slow (τ=0.5 ns, R=400 Ω, C_d=C_L=1.5pF) and (b) fast (τ=0.5 ns, R=100 Ω, C_d=C_L=5 pF) cases.

$$Q_e=\int_{t_{ov}}^{t_p} i_p\,dt=2\int_{t_{ov}}^{\frac{t_{ov}+t_p}{2}} k_l\left(\frac{V_{DD}}{\tau}t-V_{TP}\right)^{\frac{a}{2}}\left|V_d-V_{DD}\right|dt \quad (14)$$

The pMOS drain-to-source voltage (V_d-V_{DD}) derived in the previous section is used in this integral.

In this way, the increase in the propagation delay is found as the time needed to remove the equivalent charge Q_e. An average value for the discharging current, I_{dis}, should be used. However, it can be approximated by the nMOS transistor current at time $t_p/2$, $I_{dis}=i_n[t_p/2]$,which is known from the previous analysis. Thus, the time needed to discharge this extra charge which causes the additional propagation delay can be calculated as $t_{ad}=\frac{Q_e}{I_{dis}}$.

4. ESTIMATION OF SHORT-CIRCUIT POWER DISSIPATION

The short-circuit power which is dissipated during the output switching is due to the current i_s (Fig. 1), which is drawn from V_{DD} towards the source of the pMOS transistor. Current i_s can be found by applying Kirchhoff's current law at the source of the pMOS transistor :

$$i_s=i_p-i_{C_{GS}} \quad (15)$$

where $i_{C_{GS}}=C_{GS}\frac{dV_{in}}{dt}$ is the current through the gate-to-source coupling capacitance.

Energy starts being dissipated at time t_s when i_s starts flowing towards the source of the pMOS transistor. Time t_s can be calculated by setting $i_s = 0$ using the linear region expression for the pMOS current. The pMOS transistor starts its operation in linear mode and then enters saturation at approximately $t_{sat} = \frac{t_{ov} + t_p}{2}$ [6], where i_p and consequently i_s reach their maximum value. Assuming that the pMOS current and consequently i_s is symmetrical around t_{sat}, the dissipated energy due to the short-circuit current is given by :

$$E_{sc} = 2 \cdot V_{DD} \int_{t_s}^{t_{sat}} i_s \, dt \qquad (16)$$

Consequently, the short circuit power dissipation for a symmetrical driver and for a system clock frequency f, is :

$$P_{sc} = 2 a f E_{SC} \qquad (17)$$

where α is the switching activity of the output node.

The logic stages following a large RC load will dissipate significant amounts of short-circuit power due to the degraded waveform which they receive as input. Connecting the 20% and 80% point of the output waveform, an effective ramp input for the following stages is obtained which can be used in the corresponding formulas [1] for the calculation of the short circuit power dissipation in these stages.

5. RESULTS AND CONCLUSIONS

Since the output waveform expression for each of the regions of operation is known, propagation delay can be calculated as the time from the half-V_{DD} point of the input to the half-V_{DD} point of the output. Using this definition, the propagation delay has been calculated for several output loads (Fig. 3). It has been observed that in all cases the calculated values are in very good agreement with the delay derived by SPICE simulations. The error was less than 3.5 % while in [3] the error in propagation delay that has been mentioned, for many cases exceeded 40 %. It has also been found that the effect of the input slope on the propagation delay is significant, proving that delay models which consider step input are inadequate.

A comparison between the short-circuit energy dissipation per output transition calculated using the proposed method and the energy which is measured by SPICE is given in Fig. 4. The calculated energy dissipation is very close to that computed by SPICE and it should be mentioned that it is always overestimated which is required since design specifications should always be met.

6. REFERENCES

[1] 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, April 1990.

[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, no. 2, pp. 302-306, February 1998.

[3] V. Adler and E.G. Friedman, "Delay and Power Expressions for a CMOS Inverter Driving a Resistive-Capacitive Load", *Proc. of IEEE Int. Symp. on Circuits and Systems* (*ISCAS*), pp.101-104, 1996.

[4] J. Qian, S. Pullela and L. Pillage, "Modeling the ''Effective Capacitance'' for the RC Interconnect of CMOS Gates", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 13, No. 12, pp. 1526-1535, Dec. 1994.

[5] F. Dartu, N. Menezes and L. T. Pileggi, "Performance Computation for Precharacterized CMOS Gates with RC Loads", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, no. 5, pp. 544-553, May 1996.

[6] A. Hirata, H. Onodera and K. Tamaru, "Estimation of Short-Circuit Power Dissipation for Static CMOS Gates Driving a CRC π Load", *Proc. International Workshop Power and Timing Modeling, Optimization and Simulation* (*PATMOS*), pp. 279-290, 1997.

[7] T. Sakurai, "Approximation of Wiring Delay in MOSFET LSI", *IEEE J. Solid-State Circuits*, vol. SC-18, pp. 418-426, Aug. 1983.

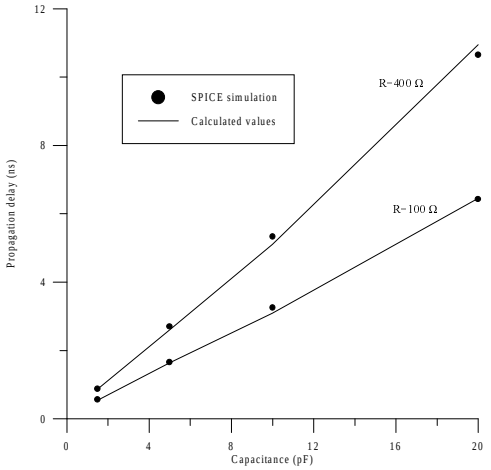

Fig. 3 Comparison between propagation delays measured with SPICE and calculated values using the proposed method for several output capacitances and two different resistance values. The input transition time is 0.5 ns

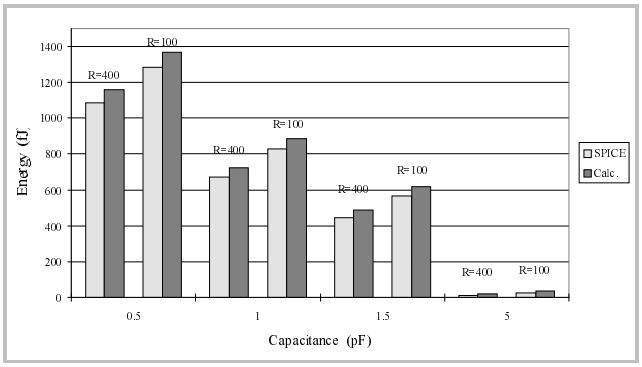

Fig. 4 Comparison between simulated and calculated values for short-circuit energy dissipation, for several capacitances and resistance values (Ω). The input transition time is 0.5 ns.