ACCURATE TIMING MODEL FOR THE CMOS INVERTER

L. Bisdounis, S. Nikolaidis[1], O. Koufopavlou, C. Goutis

VLSI Design Laboratory, Department of Electrical & Computer Engineering, University of Patras, 26500 Patras, GREECE

[1] Electronics & Computers Division, Department of Physics, University of Thessaloniki, 54006 Thessaloniki, GREECE

E-mail: bisdouni@ee.upatras.gr

ABSTRACT

This paper introduces an accurate, analytical timing model for the CMOS inverter. Analytical output waveform expressions for all the inverter operation regions and input waveform slopes are derived, which take into account the complete expression of the short-circuit current and the gate-to-drain coupling capacitance.

1. INTRODUCTION

In the recent years the circuit modeling research focused on the delay model development for CMOS circuits. The emphasis of this work is on evaluating analytically the propagation delay of the CMOS inverter.

Analytical expressions for the CMOS inverter output waveform and the propagation delay, including the effect of the input waveform slope, was presented in [1], [2], where the influence of the short-circuit current was neglected. More recently in [3], the differential equation describing the discharge of the load capacitor was solved for a rising input ramp considering the current through both transistors. However, the output waveform wasn't completely expressed analytically, because the integration constant between the region where the PMOS device is in the linear region and the region where the PMOS device is saturated, is not determined analytically. This results to a semi-empirical model for the propagation delay, using simulation results and numerical methods. Sakurai and Newton [4] presented a closed-form delay expression for the CMOS inverter, based on the α-power law MOS model which includes the carrier velocity saturation effect. However, this model requires the empirical velocity saturation index (α), and other model parameters to be recomputed for each transistor width. Moreover, the short-circuit current is neglected and the delay expression is valid only for fast input ramps.

In this paper, analytical expressions for the output waveform, which overcome the weaknesses of previous works, are derived. Based on these expressions, accurate, analytical formulas for the evaluation of the propagation delay of the CMOS inverter for all the cases of input ramps, are produced. To achieve better accuracy, the short-circuit current and the gate-to-drain capacitive coupling, are taken into account. The simplified bulk-charge MOS model [5] has been chosen, since short-channel models is very difficult to handle analytically, when the current through both transistors, and all the cases of input ramps, are considered. However, the experience derived from the results using a simple model could be expanded to more advanced models.

2. OUTPUT WAVEFORM ANALYSIS

The derivations presented in the following are for a rising input ramp,

$$V_{in} = \begin{cases} 0, & t \le 0 \\ V_{DD} \cdot (t/\tau), & 0 \le t \le \tau \\ V_{DD}, & t \ge \tau \end{cases}, \quad (1)$$

where τ is the input rise time. However, similar results can be obtained for falling input ramp. The differential equation which describes the discharge of the load capacitance C_L for the CMOS inverter of Fig.1, taking into account the gate-drain capacitive coupling (C_M) is,

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

The drain current of the devices is given by the following equations of the long-channel simplified bulk-charge MOS model [5],

$I_D = 0$, $V_{GS} < V_T$, Cutoff region

$$I_D = \beta \left[(V_{GS} - V_T) V_{DS} - \frac{(1+\delta)}{2} V_{DS}^2 \right],$$

$V_{DS} \le V_{SAT}$, Linear region

$$I_D = \frac{\beta}{2(1+\delta)} (V_{GS} - V_T)^2,$$

$V_{DS} > V_{SAT}$, Saturation region

where $V_{SAT} = \frac{V_{GS} - V_T}{1+\delta}$ is the saturation voltage of the device, V_T is the threshold voltage, β is the device gain factor, and δ is the slope of the first order term at the Taylor series expansion of the MOS bulk-charge equation.

In order to give a complete analysis, three cases of input ramp are considered. First, the case of very fast input ramps where the PMOS device is turned off after its linear region, without enters saturation, is studied. Since the input ramp will reach its final value with the NMOS device either

in saturation or in the linear region, two more cases of input ramps, are considered. For fast input ramps, the NMOS device is still saturated while for slow input ramps the NMOS is in its linear region, when the input voltage ramp reaches its final value. In the following, normalized voltages with respect to V_{DD}, i.e. $u_{in} = V_{in} / V_{DD} = t / \tau$, $u_{out} = V_{out} / V_{DD}$, $n = V_{TN} / V_{DD}$, $p = |V_{TP}| / V_{DD}$, and the variable $x = t / \tau$, are used.

Case A: As mentioned above, the first case to be studied is for very fast input ramps such that the PMOS transistor is turned off after its linear region, without enters saturation (Fig.2). Also, in this case the NMOS transistor is still saturated when the input voltage reaches its final value.

Region 1 ($0 \le x \le n$): The NMOS transistor is off, and the PMOS transistor is in the linear region. In this region the differential equation (2) becomes a non-linear Riccati equation [6] which cannot be solved analytically, if a particular solution is not known. Thus, a power-series expansion method [6] has been used, resulting to the following solution,

$$u_{out} = 1 - \sum_{k=1}^{\infty} f_k x^k , \qquad (3)$$

where $f_1 = -c_m$, $f_2 = -\frac{A_p}{2}(1-p) f_1$,

$$f_k = \frac{A_p}{k}\left\{\left[f_{k-2} - (1-p) f_{k-1}\right] + \frac{(1+\delta_p)}{2}\sum_{i=1}^{k-2} f_i f_{k-i-1}\right\},$$

for $k > 2$, $c_m = \frac{C_M}{C_L + C_M}$, $A_p = \frac{\beta_p V_{DD} \tau}{C_L + C_M}$.

The second term in the braces for k>2 corresponds to the influence of the quadratic current term of the PMOS device which was neglected in [3].

Region 2 ($n \le x \le 1-p$): The NMOS transistor is saturated and the PMOS transistor is in the linear region. In this case, the power-series method results to the following recursive expression,

$$u_{out} = u_{12} + 1 - \sum_{k=1}^{\infty} g_k (x-n)^k \qquad (4)$$

where $g_1 = -c_m$, $g_2 = -\frac{A_p}{2}(1-p-n)$,

$$g_3 = \frac{A_n}{6(1+\delta_n)} + \frac{A_p}{3}\left[g_1 - (1-p-n) g_2 + \frac{(1+\delta_p)}{2} g_1^2\right],$$

$$g_k = \frac{A_p}{k}\left\{\left[g_{k-2} - (1-p-n) g_{k-1}\right] + \frac{(1+\delta_p)}{2}\sum_{i=1}^{k-2} g_i g_{k-i-1}\right\},$$

for $k > 3$, $A_n = \frac{\beta_n V_{DD} \tau}{C_L + C_M}$, and $u_{12} = -\sum_{k=1}^{\infty} f_k n^k$ is the integration constant which is inserted to ensure continuity with respect to region 1.

Region 4 ($1-p \le x \le 1$): The NMOS transistor is saturated and the PMOS transistor is off. It can be observed in Fig.2 that for very fast input ramps (case A), the inverter doesn't pass from region 3, because the PMOS device is not saturated. The analytical solution of the differential equation (2) in this region is,

$$u_{out} = u_{24} + c_m x - \frac{A_n}{6(1+\delta_n)}(x-n)^3 , \qquad (5)$$

where the integration constant u_{24}, which is inserted to ensure continuity with respect to region 2, is given by,

$$u_{24} = u_{[1-p]} - c_m (1-p) + \frac{A_n}{6(1+\delta_n)}(1-p-n)^3$$

where $u_{[1-p]} = u_{12} + 1 - \sum_{k=1}^{\infty} g_k (1-p-n)^k$ is the value of the output voltage in which the PMOS device is turned off (when x=1-p).

Region 5A ($1 \le x \le x_{satn}$): The input ramp has reached its final value with the NMOS transistor still in saturation and the PMOS transistor off. x_{satn} is the normalized time value where the NMOS device leaves saturation, i.e. $V_{DS\text{-}NMOS} = V_{SAT\text{-}NMOS}$. The analytical solution of (2) becomes,

$$u_{out} = u_{24} + c_m - \frac{A_n (1-n)^3}{6(1+\delta_n)} - \frac{A_n (1-n)^2}{2(1+\delta_n)}(x-1) . \qquad (6)$$

Region 6 ($x \ge x_{satn}$): The NMOS device enters in its linear region, and the PMOS is off. In this region the analytical solution of (2) is,

$$u_{out} = \frac{2 u_{satn}}{1 + e^{A_n (x - x_{satn})(1-n)}} , \qquad (7)$$

where $u_{satn} = (1-n) / (1+\delta_n)$, and x_{satn} is calculated from equation (6) for $u_{out} = u_{satn}$.

Case B: In the second case we study fast input ramps such that the PMOS transistor enters saturation after the linear region, and the NMOS transistor is still saturated when the input ramp reaches its final value (Fig.2). The expressions of the output waveform for regions 1, 2 are the same with those of case A. Note, that the right limit of region 2 in this case is the normalized time value (x_{satp}), where the PMOS device enters the saturation region, i.e. $V_{DS\text{-}PMOS} = V_{SAT\text{-}PMOS}$.

Region 3 ($x_{satp} \le x \le 1-p$): Both transistors are saturated. The analytical solution of (2) is,

$$u_{out} = u_{23} + c_m x - \frac{A_n}{6(1+\delta_n)}(x-n)^3 + \frac{A_p}{6(1+\delta_p)}(x-1+p)^3 , \qquad (8)$$

where the integration constant u_{23}, which is inserted to ensure continuity with respect to region 2, is given by,

$$u_{23} = u_{satp} - c_m x_{satp} + \frac{A_n}{6(1+\delta_n)}(x_{satp} - n)^3 - \frac{A_p}{6(1+\delta_p)}(x_{satp} - 1 + p)^3 .$$

It is clear that for the derivation of an analytical expression for the output waveform the integration constant u_{23} must be determined. The values x_{satp}, u_{satp} satisfy the PMOS saturation condition,

$$u_{out} = 1 - \frac{1 - x - p}{1 + \delta_p}, \qquad (9)$$

and they can be found by solving the system of equations (4) and (9). Since, the order of equation (4) is high, the system of equations cannot be solved analytically. Hence, in the following an efficient method for the calculation of x_{satp}, u_{satp} is introduced, which is illustrated in Fig.3. Note, that the integration constant between the region where the PMOS device is in the linear region and the region where the PMOS is saturated, is not determined analytically in [3], where the propagation delay calculation is based on semi-empirical estimation of this integration constant using simulation results and numerical methods. The analytical solution of the differential equation (2) in region 2, if we neglect the PMOS current is,

$$u'_{out} = u'_{12} + c_m x - \frac{A_n}{6(1+\delta_n)}(x-n)^3, \qquad (10)$$

where $u'_{12} = 1 - c_m n - \sum_{k=1}^{\infty} f_k n^k$. From equations (9), (10) the normalized time x'_{satp} in which the inverter is entered region 3, with the assumption of negligible PMOS current, is derived. The next step of our method is to determine the tangent of the output waveform expressed by (4), at the point which corresponds to x'_{satp} (Fig.3). This tangent is expressed by the equation:

$$u''_{out} = a\,x + b, \qquad (11)$$

where $a = \left.\frac{du_{out}}{dx}\right|_{x=x'_{satp}} = -\sum_{k=1}^{\infty} k\,g_k\,(x'_{satp} - n)^{k-1}$,

and $b = 1 + u_{12} - a\,x'_{satp} - \sum_{k=1}^{\infty} g_k\,(x'_{satp} - n)^k$.

From equations (9) and (11) an accurate approximation for x_{satp} is derived,

$$x_{satp} = \frac{(1+\delta_p)\,b - \delta_p - p}{1 - a\,(1+\delta_p)}. \qquad (12)$$

By substituting x_{satp} in equation (4) the normalized output voltage u_{satp} is calculated. The expressions of the output waveform for the regions 4, 5A, 6 are the same with those of case A, if we substitute the integration constant u_{24} with the constant u_{23}.

Case C: In the third case slow input ramps, such that the NMOS device leaves saturation while the input voltage is still a ramp are studied (Fig.2). The output expressions for the regions 1, 2, 3 are the same with those of the previous case.

Region 4 ($1-p \le x \le x_{satn}$): The NMOS transistor is saturated and the PMOS transistor is off. The solution of the differential equation which describes the operation of this region, is given by equation (5), after the substitution of the constant u_{24} with the constant u_{23}. The normalized time value x_{satn} is calculated from this equation, for $u_{out} = (x-n)/(1+\delta_n)$, which corresponds to the NMOS device saturation line (Fig.2).

Region 5B ($x_{satn} \le x \le 1$): The NMOS transistor is in the linear region and the PMOS transistor is off. Neglecting the charging current through the coupling capacitance an approximated solution of the differential equation (2) in this region is,

$$u_{out} = \frac{(1+\delta_n)^{-1}\sqrt{\frac{2}{A_n}}\,e^{-y^2}}{\frac{1}{y_{satn}\,e^{y_{satn}^2}} - \frac{\sqrt{\pi}}{2}\left(\mathrm{erf}[y] - \mathrm{erf}[y_{satn}]\right)}, \qquad (13)$$

where $y = \sqrt{\frac{A_n}{2}}(x-n)$, $y_{satn} = \sqrt{\frac{A_n}{2}}(x_{satn} - n)$, and erf [y], erf [$y_{satn}$] are the error functions of y, y_{satn} respectively.

Region 6 ($x \ge 1$): The input ramp has reached its final value, the NMOS device is still in the linear region, and the PMOS device is off. The solution of the differential equation which describes the operation of the inverter in this region is,

$$u_{out} = \frac{2\,u_{satn}}{1 + \left(2\,u_{satn} - u_{[1]}\right)u_{[1]}^{-1}\,e^{A_n(x-1)(1-n)}}, \qquad (14)$$

where $u_{satn} = (1-n)/(1+\delta_n)$, and $u_{[1]}$ is the value of the normalized output voltage when the input ramp has its final value. $u_{[1]}$ is calculated if we set x=1 in equation (13).

3. PROPAGATION DELAY

The fall propagation delay at the 50% voltage level may be written as,

$$t_{PHL} = t_{0.5} - \frac{\tau}{2}, \qquad (15)$$

where $t_{0.5} = x_{0.5} \cdot \tau$ and $x_{0.5}$ is the normalized time value when $u_{out} = 0.5$.

Cases A & B: In the cases of very fast and fast input ramps the output voltage reaches the 50% level ($u_{out} = 0.5$), when the inverter operates in region 6. The normalized time value $x_{0.5}$ is evaluated from equation (7) for u_{out}=0.5,

$$x_{0.5} = x_{satn} + \frac{ln\,[4\,u_{satn} - 1]}{A_n\,(1-n)}. \qquad (16)$$

Case C: In the case of slow input ramps the condition $u_{out} = 0.5$ can occur in region 6 or in region 5B. For $6 \le A_n \le 8$ the output voltage reaches the 50% level when the inverter operates in region 6, while for $8 < A_n \le 14$ in region 5B. In the first case the normalized time value $x_{0.5}$ is evaluated from equation (14) for u_{out}=0.5,

$$x_{0.5} = 1 + \frac{\ell n\left[u_{[1]}(4u_{satn}-1)(2u_{satn}-u_{[1]})^{-1}\right]}{A_n(1-n)}. \quad (17)$$

For the evaluation of $x_{0.5}$ when the output voltage reaches the 50% level in region 5B, a linear approximation of the output voltage is used in the vicinity of $u_{out}=0.5$,

$$u_{out} = c\,x + d, \quad (18)$$

where $c = \left.\frac{du_{out}}{dx}\right|_{x=x_{satn}}$ and $d = u_{satn} - c\,x_{satn}$.

By setting u_{out}=0.5 in equation (18) an accurate approximation for $x_{0.5}$ is derived,

$$x_{0.5} = \frac{0.5-d}{c}, \quad (19)$$

By substituting $x_{0.5}$ from equations (16), (17), (19) in equation (15) the fall propagation delay of the inverter for all the cases of input ramps, is evaluated. The error which is introduced to the propagation delay due to the approximation in the calculation of x_{satp} for k=13 is up to 10^{-3} %, while in the calculation of $x_{0.5}$ (equat. (19)) up to 1.2 %. The theoretical fall delay of the inverter, is plotted as a function of A_n in Fig.4 (solid curve). These results have been produced for an inverter with $\beta_n=\beta_p=0.5mA/V^2$, $C_L=0.45pF$, $c_m=0.1$, n=p=0.17, and $\delta_n=\delta_p=0.2$, operating at $V_{DD}=5Volts$, with input rise times from 0.2nsec ($A_n=A_p=1$) to 2nsec ($A_n=A_p=10$). It can be observed, that the presented analytical method for the evaluation of the inverter propagation delay gives results very close to those derived from long-channel level 3 SPICE simulations (dashed curve).

4. CONCLUSION

In this paper an accurate, analytical method for the evaluation of the propagation delay in a CMOS inverter, has been presented. In order to achieve that, analytical expressions of the inverter output waveform which take into account the complete form of the short-circuit current, and the gate-to-drain coupling capacitance, have been derived.

REFERENCES

[1] N. Hedenstierna, K.O. Jeppson, "CMOS circuit speed and buffer optimization", *IEEE Trans. Computer-Aided Design*, vol. CAD-6, pp.270-281, March 1987.

[2] A.I. Kayssi, K.A. Sakallah, T.M. Burks, "Analytical transient response of CMOS inverters", *IEEE Trans. Circuits and Systems-I*, vol.39, pp. 42-45, Jan. 1992.

[3] K.O. Jeppson, "Modeling the influence of the transistor gain ratio, and the input-to-output coupling capacitance on the CMOS inverter delay", *IEEE J. Solid-State Circuits*, vol. 29, pp. 646-654, June 1994.

[4] T. Sakurai, 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, pp. 584-594, April 1990.

[5] J.P. Uyemura, *Circuit Design for CMOS VLSI*. Boston: Kluwer, 1992, pp. 33-34.

[6] E.D. Rainville, P. Bedient, *Elementary Differential Equations*. N.Y: MacMillan, 1969, pp. 275-289.

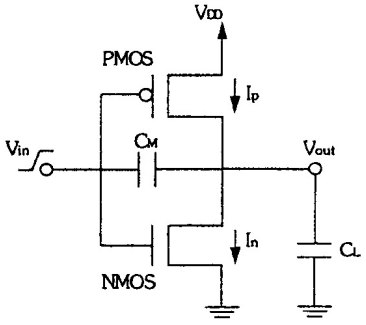

Fig.1: The CMOS Inverter

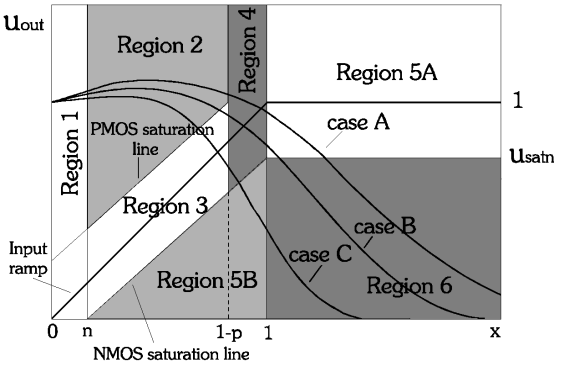

Fig.2: Operation regions of the inverter

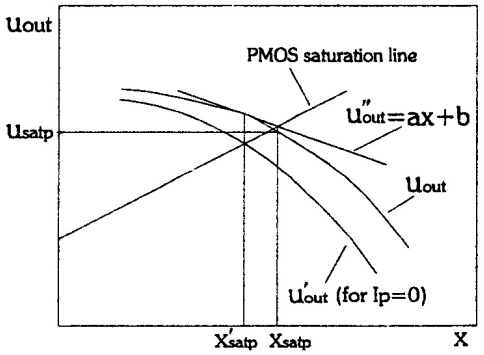

Fig.3: Approximation of the normalized time x_{satp}, in which inverter is entered region 3

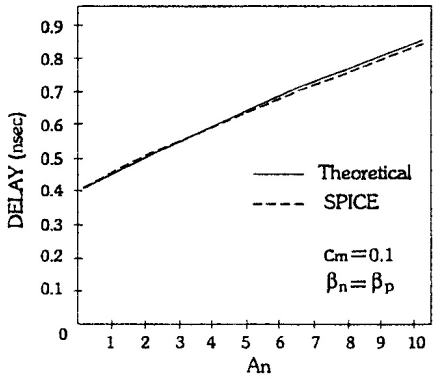

Fig.4: Inverter propagation delay