Switching Response Modeling of the CMOS Inverter for Sub-micron Devices

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,
Aristotle University of Thessaloniki,
54006 Thessaloniki, Greece

Abstract

In this paper an accurate, analytical model for the evaluation of the CMOS inverter delay in the sub-micron regime, is presented. A detailed analysis of the inverter operation is provided which results to accurate expressions describing the output waveform. These analytical expressions are valid for all the inverter operation regions and input waveform slopes. They take into account the influences of the short-circuit current during switching, and the gate-to-drain coupling capacitance. The presented model shows clearly the influence of the inverter design characteristics, the load capacitance, and the slope of the input waveform driving the inverter on the propagation delay. The results are in excellent agreement with SPICE simulations.

1. Introduction

Much effort has to be devoted for the extraction of accurate, analytical expressions for timing models of basic circuits, which can be incorporated in switch and logic simulators optimizing the design verification procedure. Using transistor level simulators with continuous-time modeling of the devices, like SPICE, is very expensive in terms of CPU time. Hence, much of past research has addressed the development of delay models for CMOS circuits. The emphasis of this paper is on the analytical evaluation of the propagation delay in a CMOS inverter. Analytical expressions of the output waveform are derived from the differential equation describing the temporal evolution of the inverter output. It is important to model accurately the CMOS inverter operation, since several fast methods for reducing a CMOS gate to an equivalent inverter have been proposed [1],[2].

The first closed-form delay expression based on the output response which was obtained directly from the differential equation describing the CMOS inverter operation, was derived in [3] for a step input. Analytical expressions for the output waveform and the propagation delay including the effect of the input waveform slope, was presented in [4] and [5]. In these the influence of the short-circuit current was neglected. These works are based on the Shichman-Hodges square-law MOS model [6] that ignores the carriers velocity saturation effect which becomes prominent in short-channel devices. In [7], the differential equation describing the discharge of the load capacitor was solved for a rising input ramp considering the current through both transistors and the gate-to-drain coupling capacitance. However, numerical and fitting methods are used, resulting in a semi-analytical model which is still based on the square-law MOS model. Nabavi-Lishi and Rumin [8] presented a method for the calculation of the CMOS inverter delay. They use a linear approximation of the output waveform based on empirical factors produced from SPICE simulations. Moreover, the method is based on an approximated version of the SPICE level-3 MOS model where the velocity saturation effect is neglected.

Sakurai and Newton [9],[10] presented closed-form delay expressions for the CMOS inverter, based on the α-power (n-power in [10]) law MOS model which includes the carriers velocity saturation effect. However, these models requires the extraction of the empirical velocity saturation index (α or n) from the static device characteristics for each transistor width. For the derivation of the output expression in [9] the short-circuit current is neglected and the delay expression is valid only for fast input ramps. In order to approximate the CMOS inverter by a NMOS circuit in [10], a fictitious input ramp is used which is clamped to ground for ramp voltages less than the switching voltage. An extension in the delay expression of [9] for the case of very lightly loaded inverter and/or slow input signals is presented in [11]. A table of coefficients produced from SPICE simulations is used, but still for negligible short-circuit current. The delay model presented in [12] uses the α-power MOS model taking into account the short-circuit current of the CMOS inverter, but the output voltage and the currents through the transistors are assumed to be piecewise linear.

In this paper, analytical expressions for the CMOS inverter output response to an input voltage ramp which overcome the weaknesses of previous works are derived. Based on these expressions, accurate, analytical formulae for the evaluation of the propagation delay for all the cases of input ramps are produced. The derived timing model takes into account the influences of the current through both transistors and the gate-to-drain coupling capacitance. It avoids numerical methods and empirical approaches based on pre-simulation data. A simple MOS model [13] which considers the carriers velocity saturation effects of short-channel devices has been chosen.

2. Inverter switching response analysis

The derivations presented in the following are for a rising input ramp: $V_{in} = V_{DD} \cdot (t/\tau)$ for $0 \le t \le \tau$, $V_{in} = 0$ for $t \le 0$ and $V_{in} = V_{DD}$ for $t \ge \tau$, where τ is the input rise time. The analysis for a falling input is symmetric, and similar results are obtained by appropriate substitutions in the derived equations. The differential equation, which describes the discharge of the load capacitance C_L for the CMOS inverter (Fig.1) taking into account the gate-drain capacitive coupling (C_M), is derived from the application of the Kirchoff's current law to the output node

$$I_{C_L} + I_{C_M} + I_p - I_n = 0,$$

$$C_L \frac{dV_{out}}{dt} = C_M \left(\frac{dV_{in}}{dt} - \frac{dV_{out}}{dt} \right) + I_p - I_n . \qquad (1)$$

For the input noted above, the equation (1) becomes

$$\frac{dV_{out}}{dt} = \begin{cases} \dfrac{I_p - I_n}{C_L + C_M}, & t \le 0 \text{ or } t > \tau \\ \dfrac{c_m V_{DD}}{\tau} + \dfrac{I_p - I_n}{C_L + C_M}, & 0 < t \le \tau \end{cases}, \qquad (2)$$

where $c_m = \dfrac{C_M}{C_L + C_M}$.

The output load consists of the inverter drain junction capacitances, the gate capacitances of fanout gates, and the interconnect capacitance. The equivalent gate-to-drain capacitance C_M is the sum of the gate-to-drain overlap capacitance and a part of the gate-to-channel capacitance of the transistors [14]. The overlap capacitance is voltage independent, and is given by

$$C_{gd-overlap} = W C_{gdo},$$

where W is the effective width of the transistor. C_{gdo} is the gate-to-drain overlap capacitance per micron which is

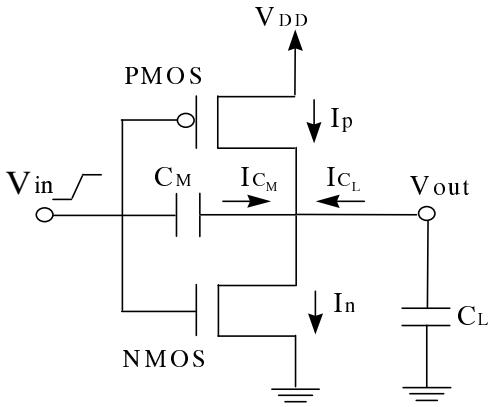

Fig.1: The CMOS inverter

determined by the process technology. In the cutoff region of the transistor there is no conducting channel, and in the saturation region the channel does not extend to the drain. Therefore, the gate-to-drain capacitance due to the channel charge is equal to zero. In the linear region the distributed gate-to-channel capacitance may be viewed as being shared equally between the source and the drain.

Thus in this case $C_{gd-channel} = \frac{1}{2} C_{ox} W L$,

where C_{ox} is the gate-oxide capacitance per unit area, and L is the effective length of the transistor.

Depending on the region of operation the drain current of the devices is given by the following equations of the used MOS model [13]

$$I_D = 0, \qquad V_{GS} < V_{TH}, \qquad \text{Cutoff,} \qquad (3)$$

$$I_D = \beta V_O \left(V_{GS} - V_{TH} \right), \quad V_{DS} > V_{DSAT}, \quad \text{Saturation,} \qquad (4)$$

$$I_D = \frac{\beta}{1 + V_O^{-1} V_{DS}} \left[\left(V_{GS} - V_{TH} \right) V_{DS} - \frac{V_{DS}^2}{2} \right], \qquad (5)$$

$$V_{DS} \le V_{DSAT}, \quad \text{Linear,}$$

where β is the device gain factor, and V_{TH} is the device threshold voltage. V_O is the voltage which specifies the effects of carriers velocity saturation and is extracted from the device static characteristics. V_{DSAT} is the device saturation voltage and is given by

$$V_{DSAT} = V_O \sqrt{1 + 2 V_O^{-1} \left(V_{GS} - V_{TH} \right)} - V_O .$$

In the following, normalized voltages with respect to V_{DD}, i.e. $u_{in} = V_{in} / V_{DD}$, $u_{out} = V_{out} / V_{DD}$, $n = V_{THN} / V_{DD}$, $p = |V_{THP}| / V_{DD}$, $v_{on} = V_{ON} / V_{DD}$, $v_{op} = V_{OP} / V_{DD}$, and the variable $x = t / \tau$ are used.

Since the input ramp will reach its final value with the NMOS device either in saturation or in the linear region, two main cases of input ramps must be considered, in order to give a complete analysis of the output waveform. For *fast* input ramps, the NMOS device is still

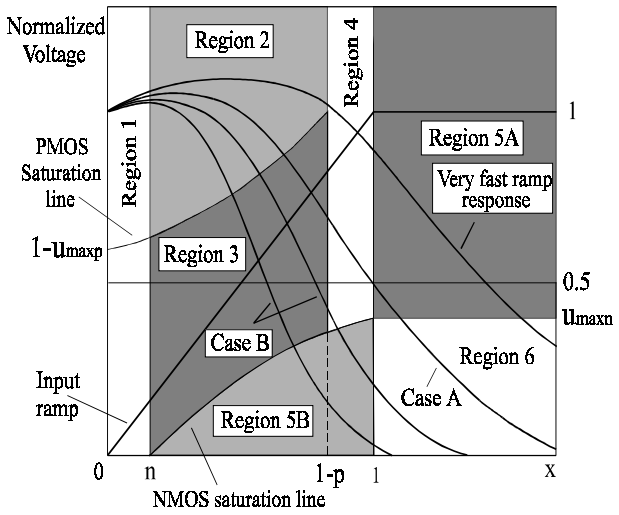

Fig.2: Operation regions of the inverter

saturated while for *slow* input ramps the NMOS is in its linear region, when the input voltage ramp reaches its final value.

Case A - Fast input ramps:

In the following, we analyze each region of the inverter operation for the case of fast input ramps (Fig.2).

Region 1, $0 \le x \le n$:

The NMOS transistor is off, and the PMOS transistor is in the linear region. The first term of the right part in equation (2) (for $0 < t \le \tau$) corresponds to the charging current through the gate-to-drain coupling capacitance (C_M). This causes the major influence on the output voltage waveform in this region. Part of the charge from the input which injected through this capacitance causes an overshoot at the early part of the output voltage waveform (Fig.2). During the overshoot the PMOS device operates in a reversed linear mode, because the output voltage is greater than the supply voltage. Thus, the PMOS device initially helps to discharging the load capacitance towards the supply voltage. The differential equation (2) using the current equations (3), (5) becomes a non-linear Riccati equation [15] which cannot be solved analytically, if a particular solution is not known. However, the quadratic term of the PMOS current can be neglected because the charge contributed by this term is negligible due to the small values of the drain-source voltage of the PMOS device in this region [7]. Also, in order to give a solution of the differential equation, an average value of u_{out} i.e. $u_1 = 1 + (c_m\, n / 2)$ is used at the denominator of the PMOS current expression. After that, the solution of equation (2) is

$$u_{out} = 1 + \frac{c_m}{\sqrt{G_1}} \sqrt{\frac{\pi}{2}}\; e^{y^2} \left(\text{erf}\,[y_1] - \text{erf}\,[y_0]\right), \quad (6)$$

where $G_1 = \dfrac{A_p}{1 + v_{op}^{-1}\left(1 - u_1\right)}$, $\quad A_p = \dfrac{\beta_p V_{DD}\,\tau}{C_L + C_M}$,

$$y_1 = \sqrt{\frac{G_1}{2}}\left(x + p - 1\right), \quad y_0 = \sqrt{\frac{G_1}{2}}\left(p - 1\right),$$

and erf $[y_1]$, erf $[y_0]$ are the error functions of y_1, y_0 respectively. Standard ways of evaluating the error function can be found in any mathematical handbook.

Region 2, $n \le x \le x_{satp}$:

The NMOS device is saturated and the PMOS device is in the linear region. During the output voltage overshoot the PMOS device still operates in a reversed linear mode. Note, that the right limit of this region (Fig.2) is the normalized time value x_{satp} where the PMOS device enters saturation, i.e. $V_{DD} - V_{out} = V_{DSATP}$. It is determined by the PMOS saturation condition

$$u_{satp} = 1 - v_{op}\left[\sqrt{1 + 2\, v_{op}^{-1}\left(1 - x_{satp} - p\right)} - 1\right],$$

where u_{satp} is the normalized output voltage value when PMOS device saturates. As in region 1 we neglect the quadratic current term of the PMOS device. Also, instead of u_{out} at the denominator of the PMOS current expression we use an average value of the normalized output voltage (u_2) i.e. $u_2 = \left(u_{[n]} + u'_{satp}\right) \big/ 2$. u'_{satp} is the value of the normalized output voltage at the end of region 2, if negligible PMOS current is assumed and is calculated below by equation (10). $u_{[n]}$ is the value of the normalized output voltage at the beginning of region 2 and is calculated from equation (6) for x = n. After the above approximations, the solution of the equation (2) is

$$u_{out} = 1 + \frac{A_n\, v_{on}}{G_2} + \left(u_{[n]} - \frac{A_n\, v_{on}}{G_2} - 1\right) \frac{e^{y^2}}{e^{y_n^2}} +$$

$$\sqrt{\pi}\; e^{y^2} \left(\frac{A_n\, y_n\, v_{on}}{G_2} + \sqrt{\frac{1}{2 G_2}}\, c_m\right) \left(\text{erf}[y_2] - \text{erf}[y_n]\right), \quad (7)$$

where $A_n = \dfrac{\beta_n V_{DD}\,\tau}{C_L + C_M}$, $\quad G_2 = \dfrac{A_p}{1 + v_{op}^{-1}\left(1 - u_2\right)}$,

$$y_2 = \sqrt{G_2 / 2}\,\left(x + p - 1\right), \quad y_2 = \sqrt{G_2 / 2}\,\left(n + p - 1\right).$$

The above equations (6), (7) give waveforms very close to those derived from SPICE simulations (as shown in the section 4), which indicates the validity of the above approximations. In order to continue the analysis for the next region the calculation of the values x_{satp}, u_{satp}, is required. As mentioned above, these values satisfy the PMOS saturation condition

$$u_{out} = 1 - v_{op}\left[\sqrt{1 + 2\, v_{op}^{-1}\left(1 - x - p\right)} - 1\right]. \quad (8)$$

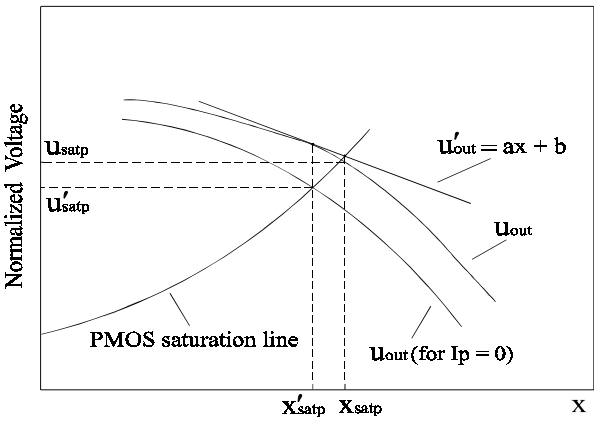

Fig.3: Evaluation of the normalized time x_{satp}

They can be found by solving the system of equations (7) and (8). Due to the error functions of equation (7), the system cannot be solved analytically. Hence, in the following an efficient method for the calculation of x_{satp}, u_{satp}, is used (Fig.3).

The analytical solution of the differential equation (2), if negligible PMOS current is assumed in region 2, is

$$u_{out} = u_{12} + c_m\, x - \frac{A_n v_{on}}{2}(x-n)^2, \qquad (9)$$

where $u_{12} = u_{[n]} - c_m\, n$ is the integration constant, which is inserted to ensure continuity with respect to region 1. By equating (8) and (9) the normalized time value x'_{satp} in which the inverter leaves region 2, with the assumption of negligible PMOS current, is derived. After the substitution of x'_{satp} in equation (9), the value of the normalized voltage u'_{satp}, is derived,

$$u'_{satp} = u_{12} + c_m\, x'_{satp} - \frac{A_n v_{on}}{2}\left(x'_{satp} - n\right)^2. \quad (10)$$

The next step of our method is to determine the tangent of the output waveform expressed by (7), 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 $a = \left.\frac{du_{out}}{dx}\right|_{x=x'_{satp}}$ and $b = u_{out}\Big|_{x=x'_{satp}} - a\,x'_{satp}$.

By equating (8) and (11), x_{satp} becomes the root of a simple quadratic equation. Then by substituting x_{satp} in (7) the normalized output voltage u_{satp} is evaluated. The error which is introduced in the calculation of x_{satp} due to the above method is up to 0.5%.

In the special case of very fast input ramps (Fig.2), the PMOS device is turned off after its linear region, without enters saturation. This occurs because the output voltage overshoot finishes when the PMOS is already off. Hence, the inverter does not enter in region 3, and the calculation of x_{satp} and u_{satp} is not required.

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

$$u_{out} = u_{23} + c_m\, x - \frac{A_n v_{on}}{2}(x-n)^2 - \frac{A_p v_{op}}{2}(1-x-p)^2, \qquad (12)$$

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

$$u_{23} = u_{satp} - c_m\, x_{satp} + \frac{A_n v_{on}}{2}\left(x_{satp} - n\right)^2 + \frac{A_p v_{op}}{2}\left(1 - x_{satp} - p\right)^2.$$

Region 4, $1\text{-}p \le x \le 1$:
The NMOS transistor is saturated and the PMOS transistor is off. The analytical solution of the differential equation (2) is

$$u_{out} = u_{23} + c_m\, x - \frac{A_n v_{on}}{2}(x-n)^2. \qquad (13)$$

As mentioned above for very fast input ramps the inverter does not pass from the region 3, because the PMOS device is not saturated. In this case the integration constant u_{23} is substituted by the following constant u_{24}

$$u_{24} = u_{[1-p]} - c_m(1-p) + \frac{A_n v_{on}}{2}(1-p-n)^2,$$

where $u_{[1-p]}$ is the value of the normalized output voltage in which the PMOS device is turned off. It is calculated from equation (7) for x = 1–p.

Region 5A, $1 \le x \le x_{satn}$:
The input ramp has reached its final value with the NMOS device still in saturation and the PMOS device off. x_{satn} is the normalized time value where the $V_{out} = V_{DSATN}$. In this region, the analytical solution of the differential equation (2) (for $t > \tau$) becomes

$$u_{out} = u_{23} + c_m - \frac{A_n v_{on}}{2}(1-n)^2 - A_n v_{on}(1-n)(x-1), \qquad (14)$$

where the constant u_{23} is substituted by u_{24} for very fast input ramps.

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

$$x - x_{satn} = \frac{1 + 2v_{on}^{-1}(1-n)}{A_n(1-n)}\,\ell n\left[\frac{2(1-n) - u_{out}}{2(1-n) - u_{maxn}}\right] - \frac{1}{A_n(1-n)}\,\ell n\left(\frac{u_{out}}{u_{maxn}}\right), \qquad (15)$$

where $u_{maxn} = v_{on}\left[\sqrt{1+2v_{on}^{-1}(1-n)}-1\right]$. x_{satn} is calculated from equation (14) for $u_{out} = u_{maxn}$.

Case B - Slow input ramps:
In the second case, slow input ramps are studied. The NMOS device leaves saturation while the input voltage is still a ramp. This occurs if the value of the normalized output voltage when the input ramp reaches its final value is lower than u_{maxn} (Fig.2). The output expressions for the regions 1, 2, 3 and 4 are the same with those of the previous case. The normalized time x_{satn} is calculated from equation (13) for

$$u_{out} = v_{on}\left[\sqrt{1+2v_{on}^{-1}(x-n)}-1\right],$$

which corresponds to the NMOS saturation line (Fig.2). In the case of slower input ramps the inverter doesn't enter in region 4. This occurs in the case where the PMOS transistor is turned off when the NMOS transistor is already in the linear region. In this case x_{satn} is calculated from equation (12).

Region 5B, $x_{satn} \le x \le 1$:
The NMOS transistor is in the linear region and the PMOS transistor is either off, or so poorly conducting that its influence can be neglected. Neglecting the charging current through the gate-to-drain coupling capacitance and using at the denominator of the NMOS current an average value for the output voltage (u_{satn} / 2), an approximated solution of (2) is

$$u_{out} = e^{-y^2}\left[\frac{1}{u_{satn}\, e^{y_{satn}^2}} - \sqrt{\frac{\pi A_n}{4\left(2+u_{satn}v_{on}^{-1}\right)}}\left(\mathrm{erf}[y_3]-\mathrm{erf}[y_{satn}]\right)\right]^{-1}, \quad (16)$$

where $y_3 = \sqrt{\frac{A_n}{2+u_{satn}v_{on}^{-1}}}(x-n)$,

$$y_{satn} = \sqrt{\frac{A_n}{2+u_{satn}v_{on}^{-1}}}(x_{satn}-n), \text{ and}$$

$$u_{satn} = v_{on}\left[\sqrt{1+2v_{on}^{-1}(x_{satn}-n)}-1\right].$$

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 output waveform is expressed as

$$x-1 = \frac{1+2v_{on}^{-1}(1-n)}{A_n(1-n)}\,\ell n\left[\frac{2(1-n)-u_{out}}{2(1-n)-u_{[1]}}\right] - \frac{1}{A_n(1-n)}\,\ell n\left(\frac{u_{out}}{u_{[1]}}\right), \quad (17)$$

where $u_{[1]}$ is the value of the normalized output voltage when the input ramp has its final value. It is calculated if we set x = 1 in equation (16).

3. Propagation delay analysis

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

$$T_D = t_{0.5} - \frac{\tau}{2} = x_{0.5}\,\tau - \frac{\tau}{2}, \quad (18)$$

where $x_{0.5}$ is the normalized time value when u_{out} = 0.5. Thus, for the evaluation of the propagation delay, the normalized time value $x_{0.5}$ must be determined for both cases of input ramps. A critical parameter in order to find in which region occurs the 50% level of the output voltage (u_{out} = 0.5), is the maximum drain saturation voltage of the NMOS device (u_{maxn} - see Fig.2). Hence, it is necessary to consider two possibilities in the delay calculation: $u_{maxn} \le 0.5$ and $u_{maxn} \ge 0.5$ where

$$u_{maxn} = v_{on}\left[\sqrt{1+2v_{on}^{-1}(1-n)}-1\right].$$

$u_{maxn} \le 0.5$:
In the case of fast input ramps the output voltage reaches the 50% level, when the inverter operates in region 5A if $u_{[1]} \ge 0.5$ and in region 4 if $u_{[1]} \le 0.5$. $u_{[1]}$ is the value of the normalized output voltage when the input ramp reaches its final value, and is calculated from equation (13) for x = 1. When u_{out} = 0.5 occurs in region 5A, $x_{0.5}$ is calculated from equation (14)

$$x_{0.5} = \frac{n+1}{2} + \frac{u_{23}+c_m-0.5}{A_n v_{on}(1-n)}. \quad (19)$$

In the case where u_{out} = 0.5 occurs in region 4, $x_{0.5}$ is calculated from equation (13)

$$x_{0.5} = \frac{A_n v_{on} n + c_m + \sqrt{c_m^2 + A_n v_{on}(2c_m n + 2u_{23} - 1)}}{A_n v_{on}}. \quad (20)$$

For slow input ramps the condition u_{out} = 0.5 occurs in region 4 if $u_{[1-p]} \ge 0.5$ and in region 3 if $u_{[1-p]} \le 0.5$. $u_{[1-p]}$ is the value of the normalized output voltage when the PMOS device enters the cutoff region. In the first case the normalized time value $x_{0.5}$ is given by equation (20), and in the second one is calculated from equation (12)

$$x_{0.5} = \frac{1}{E}\left(-2D + \sqrt{4D^2 - 2EK}\right), \quad (21)$$

where $D = A_p v_{op}(p-1) - A_n v_{on} n - c_m$,

$E = 2\left(A_n v_{on} + A_p v_{op}\right)$, and

$K = \left(1 - 2u_{23} + A_p v_{op}(p-1)^2 + A_n v_{on} n^2\right)$.

$u_{maxn} \geq 0.5$:
For fast input ramps $u_{out} = 0.5$ occurs in region 6, and $x_{0.5}$ is calculated from (15)

$$x_{0.5} = x_{satn} + \frac{1 + 2v_{on}^{-1}(1-n)}{A_n(1-n)} \ell n\left[\frac{2(1-n) - 0.5}{2(1-n) - u_{maxn}}\right] - \frac{1}{A_n(1-n)} \ell n\left(\frac{0.5}{u_{maxn}}\right). \quad (22)$$

In the case of slow input ramps the output voltage reaches the 50% level, when the inverter operates in region 6 if $u_{[1]} \geq 0.5$. If $u_{[1]} \leq 0.5$ there are two possibilities for the region in which $u_{out} = 0.5$. Either $u_{satn} \geq 0.5$ the output voltage reaches the 50% level in region 5B, or $u_{satn} \leq 0.5$ in region 3. In region 6 $x_{0.5}$ is calculated from (17)

$$x_{0.5} = 1 + \frac{1 + 2v_{on}^{-1}(1-n)}{A_n(1-n)} \ell n\left[\frac{2(1-n) - 0.5}{2(1-n) - u_{[1]}}\right] - \frac{1}{A_n(1-n)} \ell n\left(\frac{0.5}{u_{[1]}}\right), \quad (23)$$

and in region 3 $x_{0.5}$ is given by equation (21). Since, the expression of the output waveform in region 5B cannot be solved analytically, u_{out} can be approximated by a ramp in the vicinity of the 50% level, in this region. Then

$$x_{0.5} = x_{satn} + \frac{0.5 - u_{satn}}{d}, \quad (24)$$

where $d = \left.\frac{du_{out}}{dx}\right|_{x=x_{satn}}$ is the output waveform slope in region 5B and is calculated using equation (16).

In real CMOS datapaths, the input signal of a gate is not a ramp, but the output waveform of the preceding gate. In order the derived ramp delay model to be applicable to inverter chains, the real input waveform must be approximated by a ramp waveform to obtain an effective transition time. Some efficient approximations for the evaluation of the effective output transition time of the inverter can be found in [4], [9].

4. Results and discussion

In Fig.4 some typical output waveforms, produced from the above expressions, are shown. A sub-micron CMOS process technology of 0.5μm, has been used to validate the accuracy of the presented inverter output

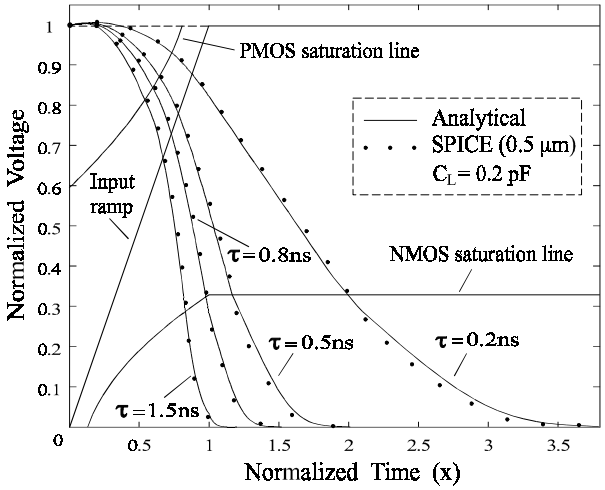

Fig.4: Inverter output waveforms

Parameter	NMOS	PMOS
L (μm)	0.5	0.5
W (μm)	3	6.45
V_O (Volts)	0.5	1
$\lvert V_{TH} \rvert$ (Volts)	0.657	0.921
C_{ox} (fF/μm^2)	3.56	3.56
C_{gdo} (fF/μm)	0.305	0.240

Table 1: Model parameters used in calculations

waveform expressions. The model parameters and the dimensions of both transistors, are listed in Table 1. The transistor widths have been selected in order to achieve equal drain currents at $V_{GS} = V_{DS} = V_{DD}$. The output waveforms produced by SPICE simulations are added for comparison. A supply voltage of 5Volts, and an output load of 0.2pF, were used. It can be observed, that the analytical waveforms are very close to those produced by SPICE simulations. In order to give output waveforms for several input rise times in the same diagram, the normalized output voltage is plotted as a function of the normalized time ($x = t / \tau$). The output waveforms for input times 0.2ns and 0.5ns correspond to case A, while those for input times 0.8ns and 1.5ns to case B. As we can see, the slope of the output waveforms in case A is smaller than the input slope, while in case B is greater than the input slope.

In Fig.5 the inverter propagation delay for a rising input ramp, is plotted as a function of $A_{no} = (\beta_n V_{DD} \tau)/C_L$. Since, A_{no} is a single lumped parameter which takes into account the input waveform slope, the drivability of the switching transistor and the load capacitance, determines the relation between the input and the output waveform. The results for $A_{no} < 15$ correspond to fast inputs (case A) compared to the output waveforms, and those for $A_{no} > 15$

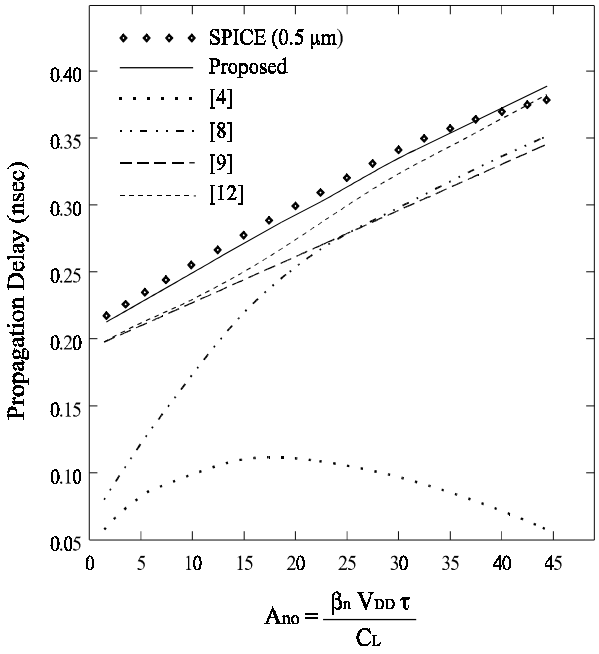

Fig.5: Inverter propagation delay

correspond to slow inputs (case B). Results using the approaches for the evaluation of the propagation delay presented in [4],[8],[9] and [12], are also given. It can be observed, that the presented model gives results closer to those derived from SPICE simulations than the other methods. The error is less than 3.5%. This occurs because our model includes the influences of the short-circuit current, and the gate-to-drain coupling capacitance on the expressions of the inverter output waveform. Another advantage of the previous analysis is the use of a simple MOS model which takes into account the velocity saturation effects of short-channel devices, without need of parameters extraction when the transistor width is changed.

The presented timing model can be used for more complex CMOS gates, since several fast methods [1],[2] have been proposed for reducing a CMOS gate to an equivalent inverter. The most critical issue in gate modeling is the reduction of serial connected MOSFETs, in order to reduce the drivability of the serial array to the drivability of an equivalent MOSFET. Using reduction techniques the propagation delay of a gate can be computed quickly and accurately using the timing model of the CMOS inverter and without the complications associated with trying to generalize the inverter model to complex gates.

5. Conclusion

In this paper an accurate, analytical method for the evaluation of the CMOS inverter propagation delay for sub-micron devices, has been presented. In order to achieve that, analytical expressions of the inverter output ramp response, for all the cases of input ramps, have been derived. These expressions take into account the influences of the short-circuit current and the gate-to-drain coupling capacitance, in all operation regions, without using empirical approaches based on pre-simulation results.

References

[1] A. Nabavi-Lishi, 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.

[2] J-T. Kong, D. Overhauser, “Methods to improve digital MOS macromodel accuracy”, *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 868-881, July 1995.

[3] J.R. Burns, “Switching response of complementary symmetry MOS transistor logic circuits”, *RCA Review*, vol. 25, pp. 627-661, Dec. 1964.

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

[5] 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.

[6] H. Shichman, D.A. Hodges, “Modeling and simulation of insulated-gate field-effect transistor switching circuits”, *IEEE J. Solid-State Circuits*, vol. SC-3, pp. 285-289, Sept. 1968.

[7] 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.

[8] A. Nabavi-Lishi, N.C. Rumin, “Simultaneous delay and maximum current calculation in CMOS gates”, *IEE Electronics Letters*, vol. 28, pp. 682-684, March 1992.

[9] 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.

[10] T. Sakurai, A.R. Newton, “A simple MOSFET model for circuit analysis”, *IEEE Trans. Electron Devices*, vol. 38, pp. 887-894, April 1991.

[11] S. Dutta, S.S. Mahant Shetti, S.L. Lusky, “A comprehensive delay model for CMOS inverters”, *IEEE J. Solid-State Circuits*, vol. 30, pp. 864-871, August 1995.

[12] S.H.K. Embabi, R. Damodaran, “Delay models for CMOS, BiCMOS, BiNMOS circuits and their applications for timing simulations”, *IEEE Trans. Computer-Aided Design*, vol. 13, pp. 1132-1142, Sept. 1994.

[13] M. Shoji, *CMOS Digital Circuit Technology*. N.J.: Prentice-Hall, 1988.

[14] N.H.E. Weste, K. Eshraghian, *Principles of CMOS VLSI Design: A Systems Perspective*. N.Y.: McGraw-Hill, 1993.

[15] E.D. Rainville, P.E. Bedient, *Elementary Differential Equations*. N.Y.: McGraw-Hill, 1969.