Modelling output waveform and propagation delay of a CMOS inverter in the submicron range

L. Bisdounis
O. Koufopavlou
S. Nikolaidis

Indexing terms: CMOS inverters, Submicron technology

Abstract: An accurate, analytical model is presented for the evaluation of the CMOS inverter delay in the submicron regime. Following an exhaustive analysis of the inverter operation, accurate expressions of the output response to an input ramp are derived, which result in the analytical calculation of the propagation delay. These expressions are valid for all the inverter operation regions and input waveform slopes, and take into account the influences of the short-circuit current and the gate–drain coupling capacitance. The effective output transition time of the inverter is determined, in order to map the real output waveform to a ramp waveform for the model to be applicable to CMOS gate chains. The results are in very good agreement with SPICE simulations.

1 Introduction

Much effort has to be devoted to the extraction of accurate timing models of basic circuits which can be incorporated in switch and logic simulators. Using transistor-level simulators with continuous-time modelling of the devices can be very expensive in terms of CPU time. Hence, much previous research has addressed the development of analytical delay models for CMOS circuits.

The emphasis of this work is on the analytical evaluation of the propagation delay in a CMOS inverter. It is important to have an accurate model for the CMOS inverter operation, since several fast methods have been proposed for reducing a CMOS gate to an equivalent inverter [1].

The first closed-form delay expression, based on the output response obtained directly from the differential equation describing the CMOS inverter operation, was derived [2] for a step input. Analytical expressions for the output waveform and the propagation delay, including the effect of the input waveform slope, have been presented [3, 4], but the influence of the short-circuit current was neglected. These previous works are based on the Shichman-Hodges square-law MOS model that ignores the carriers' velocity saturation effect, which becomes prominent in short-channel devices. The differential equation describing the discharge of the load capacitor has been solved [5] for a rising input ramp considering the current through both transistors and the gate–drain coupling capacitance. However, fitting methods based on simulation results are used, resulting in a semi-analytical delay model which is still based on the square-law MOS model. We have previously presented [6] an analytical timing model, also based on the square-law MOS model, without using pre-simulation or fitting methods.

Shih and Kang [7] proposed a solution for the non-linear case of the differential equation describing the temporal evolution of the output node. However, this solution is derived for simplified MOS models, where the current of the transistors can be expressed by quadratic equations. Nabavi-Lishi and Rumin [8] presented a method for the calculation of the inverter delay, using a linear approximation of the output waveform based on empirical factors produced from SPICE simulations.

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 require the empirical velocity saturation index (α or n) and other model parameters (such as I_{DO}) to be extracted from the I–V curves for each transistor width. For the derivation of the output expression [9], the short-circuit current is neglected and the delay expression is derived only for fast input ramps. They also used a fictitious input ramp [10] which is clamped to ground for ramp voltages less than the switching voltage, in order to approximate the CMOS inverter by an NMOS circuit. This approximation is exact only for extreme cases of input ramps (very fast and very slow). In addition, they neglected the influence of the gate–drain coupling capacitance.

An extension to the delay expression of Sakurai and Newton [9] for the case of very lightly loaded inverter and/or slow input signals is presented by Dutta *et al.* [11]. The delay is obtained by curve fitting between two extreme points corresponding to infinitely fast and infinitely slow inputs, using a table of coefficients produced from SPICE simulations. The delay model presented by Embabi and Damodaran [12] uses the

IEE Proceedings online no. 19982391

Paper first received 27th October 1997 and in revised form 22nd May 1998

L. Bisdounis and O. Koufopavlou are with VLSI Design Laboratory, Department of Electrical & Computer Engineering, University of Patras, 26500 Patras, Greece

S. Nikolaidis is with the Electronics & Computers Division, Department of Physics, Aristotle University of Thessaloniki, 54006 Thessaloniki, Greece

α-power MOS model and takes into account the short-circuit current of the CMOS inverter through a two-step iterative approach. In this approach, the output voltage and the currents through both transistors (which are nonlinear functions of time) are assumed to be piecewise linear.

An inverter timing macromodel has been presented [13], where the output voltage change over an interval of time is estimated through a computation of the average current over the interval, by using SPICE level-2 model evaluations. The approximations of this model include zero short-circuit current, resulting in inaccuracies for slow inputs. The short-circuit current waveform has been approximated [14] with a piecewise linear function of time, in order to estimate the short-circuit energy dissipation. Using the estimated short-circuit energy, a timing model is derived, in which the delay is calculated in a numerical and thus non-analytical way.

In this paper, analytical expressions are derived for the CMOS inverter output response to an input voltage ramp. Based on these expressions, accurate, analytical evaluation of the propagation delay for all the cases of input ramps is provided. The derived timing model takes into account the influences of the current through both transistors and the gate–drain coupling capacitance, without using empirical approaches based on simulation results. A simple MOS model [15], considering the carriers' velocity saturation effect that is prominent in submicron devices [9, 10, 15] has been chosen. Other second-order effects, such as channel-length modulation and mobility degradation with less influence on the device currents, can be added as empirical corrections to the simple model. However, the experience and results from this model could be expanded to more advanced models.

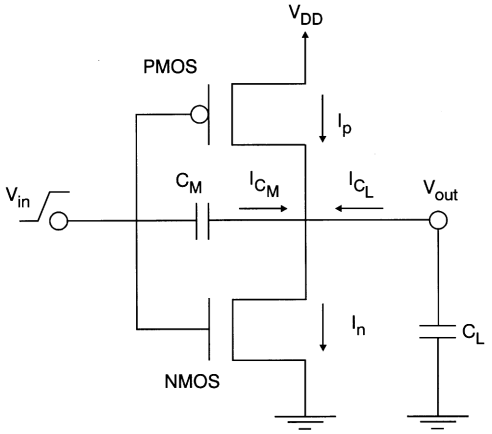

Fig.1 *CMOS inverter*

2 Inverter output waveform analysis

The derivations presented below 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 ramp is symmetrical. The differential equation that 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:

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

For the rising input ramp, eqn. 1 becomes

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

$$\text{where } c_m = \frac{C_M}{C_L + C_M} \quad (2)$$

The output load consists of the inverter drain junction capacitances, the gate capacitances of fanout gates and the interconnect capacitance. Although the first two capacitances have voltage dependency [16], the output load is well approximated by the equivalent voltage-independent capacitance C_L. The equivalent gate-drain capacitance C_M of the inverter is the sum of the gate–drain capacitances of both transistors, which consist of the gate–drain overlap capacitance and a part of the gate-to-channel capacitance. It is calculated using the simple model described by Weste and Eshraghian [16].

Depending on the mode of operation, the drain current of the devices are given by the following equations of the used MOS model [15]:

$$I_D = 0, \qquad V_{GS} < V_T, \text{ cut-off} \quad (3)$$

$$I_D = \beta V_O (V_{GS} - V_T), \; V_{DS} > V_{DSAT}, \text{ saturation} \quad (4)$$

$$I_D = \frac{\beta}{1 + V_O^{-1} V_{DS}} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right], \quad V_{DS} \le V_{DSAT}, \text{ linear} \quad (5)$$

where β is the device gain factor, V_T is the device threshold voltage, and V_O is the voltage that specifies the effects of the carriers' velocity saturation (see Appendix 6). V_{DSAT} is the device saturation voltage and is given by

$$V_{DSAT} = V_O \sqrt{1 + 2V_O^{-1}(V_{GS} - V_T)} - V_O \quad (6)$$

In the following, we used normalised voltages with respect to V_{DD}, i.e. $u_{in} = V_{in}/V_{DD}$, $u_{out} = V_{out}/V_{DD}$, $n = V_{TN}/V_{DD}$, $p = |V_{TP}|/V_{DD}$, $v_{on} = V_{ON}/V_{DD}$, $v_{op} = V_{OP}/V_{DD}$, and the variable $x = t/\tau$. 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 saturated, whereas for slow input ramps the NMOS is in its linear region when the input voltage ramp reaches its final value.

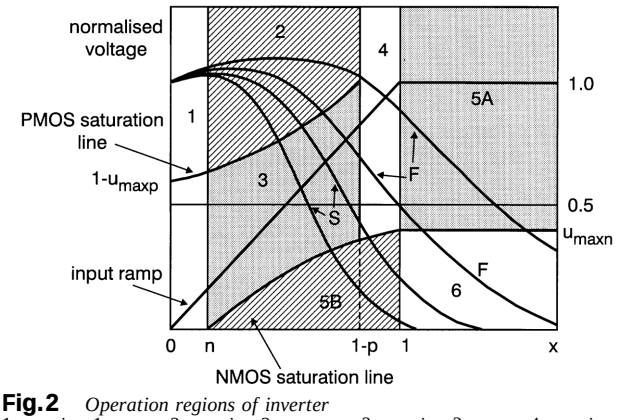

Fig.2 *Operation regions of inverter*
1 = region 1 2 = region 2 3 = region 3 4 = region 4
5A = region 5A 5B = region 5B 6 = region 6
F = fast input ramps S = slow input ramps

2.1 Fast input ramps

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

In region 1 ($0 \le x \le n$), the NMOS transistor is off and the PMOS is in the linear region. The first term on the right in eqn. 2 (for $0 < t \le \tau$) corresponds to the charging current through the gate–drain coupling capacitance (C_M), which 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. The differential equation (eqn. 2) using the current equations (eqns. 3 and 4) becomes a nonlinear Riccati equation [17], which cannot be solved analytically if a particular solution is not known. Thus, a power-series expansion method [15, 17] has been used, resulting in

$$u_{out} = 1 - \sum_{i=1}^{\infty} f_k x^k \tag{7}$$

where

$$f_1 = -c_m,\ f_2 = \frac{A_p}{2}(p-1)f_1,\ A_p = \frac{\beta_p V_{DD}\tau}{C_L + C_M},$$

and

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

A satisfactory limit to truncate the above series is obtained for $k = 8$.

In region 2 ($n \le x \le x_{satp}$), the NMOS transistor is saturated and the PMOS is in the linear region. Note that the right limit of this region (Fig. 2) is the normalised time value x_{satp} when the PMOS device is entering the saturation region, i.e. $V_{DD} - V_{out} = V_{DSATP}$. As in the previous region, the differential equation (eqn. 2) using the current equations (eqns. 4 and 5) becomes a nonlinear Riccati equation. In order to give a solution for eqn. 2, we neglect the quadratic current term of the PMOS device, because the charge contributed by this term is negligible due to the small values of the drain–source voltage of the PMOS device [5]. In addition, instead of u_{out} at the denominator of the PMOS current, we use an average value of the normalised output voltage (u_{av}) such that

$$u_{av} = \left(u_{[n]} + u'_{satp}\right)/2 \tag{8}$$

where u'_{satp} is the value of the normalised output voltage at the end of region 2, if negligible PMOS current is assumed (and is calculated using eqn. 11). $u_{[n]}$ is the value of the normalised output voltage at the beginning of region 2 and is calculated from eqn. 7 for $x = n$. After the above approximations, the solution of the differential equation (eqn. 2) is

$$u_{out} = 1 + \frac{A_n v_{on}}{G_p} + \left(u_{[n]} - \frac{A_n v_{on}}{G_p} - 1\right)\frac{e^{y^2}}{e^{y_n^2}} + \sqrt{\pi}e^{y^2}\left(\frac{A_n y_n}{v_{on}^{-1}G_p} + \sqrt{\frac{1}{2G_p}}c_m\right)(\text{erf}[y] - \text{erf}[y_n]), \tag{9}$$

where $A_n = \dfrac{\beta_n V_{DD}\tau}{C_L + C_M}, G_p = \dfrac{A_p}{1 + v_{op}^{-1}(1 - u_{av})}$,

$$y = \sqrt{\frac{G_p}{2}}(x - 1 + p),\ y_n = \sqrt{\frac{G_p}{2}}(n - 1 + p)$$

erf [y] and erf [y_n] are the error functions of y and y_n, respectively. Standard ways of evaluating the error function can be found in any mathematical textbook. Eqn. 9 yields waveforms very close to those derived from SPICE simulations (as shown in Section 4), which indicates the validity of the above approximations. To continue the analysis for the next region, calculation of x_{satp} and u_{satp} is required. These values satisfy the PMOS saturation condition

$$u_{out} = 1 - v_{op}\left[\sqrt{1 + 2v_{op}^{-1}(1 - x - p)} - 1\right] \tag{10}$$

and they can be found by solving the system of eqns. 9 and 10. Owing to the error functions of eqn. 9, the system cannot be solved analytically. Hence, in the following an efficient method for the calculation of x_{satp} and u_{satp} is used (Fig. 3).

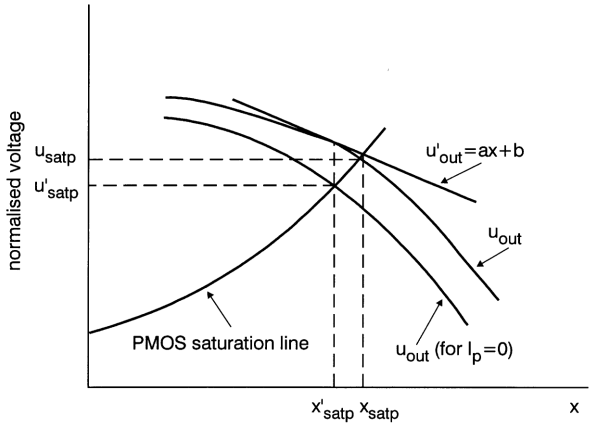

Fig.3 *Evaluation of normalised time x_{satp} where inverter is entering region 3*

If negligible PMOS current is assumed in region 2, the analytical solution of eqn. 2 is

$$u_{out} = u_{12} + c_m x - \frac{A_n v_{on}}{2}(x - n)^2 \tag{11}$$

where $u_{12} = u_{[n]} - c_m n$ is the integration constant, which is included to ensure continuity with respect to region 1. By equating eqns. 10 and 11, a quartic equation for the calculation of x'_{satp} is derived. x'_{satp} is the normalised time value when the inverter leaves region 2, with the assumption of negligible PMOS current.

The next step is to determine the tangent of the output waveform expressed by eqn. 9, at the point that corresponds to x'_{satp} (Fig. 3). This tangent is expressed by

$$u'_{out} = ax + b \tag{12}$$
$$\text{where } a = (du_{out}/dx)|_{x=x'_{satp}}$$
$$\text{and } b = (u_{out}|_{x=x'_{satp}}) - ax'_{satp}$$

By equating eqns. 10 and 12, x_{satp} becomes the root of a simple quadratic equation. By substituting x_{satp} in eqn. 9, the normalised output voltage u_{satp} is evaluated. The error introduced into the calculation of x_{satp} due to the above method is up to 0.5%.

As we can see in Fig. 2, in the special case of very fast input ramps, the PMOS device is turned off after

its linear region without entering the saturation region. This occurs because the output voltage overshoot finishes when the PMOS is already off. Hence, the inverter does not enter region 3 and the calculation of x_{satp} and u_{satp} is not required.

In region 3 ($x_{satp} \le x \le 1 - p$), both transistors are saturated. The analytical solution of eqn. 2 is

$$u_{out} = u_{23} + c_m x - \frac{A_n v_{on}}{2}(x-n)^2 - \frac{A_p v_{op}}{2}(1-x-p)^2 \quad (13)$$

where u_{23} is the integration constant, which is included to ensure continuity with respect to region 2.

In region 4 ($1 - p \le x \le 1$), the NMOS device is saturated and the PMOS is off. The analytical solution of eqn. 3 in this region is

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

In 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 normalised time, where $V_{out} = V_{DSATN}$. In this region, the analytical solution of the differential equation (eqn. 2) (for $t > \tau$) becomes

$$u_{out} = u_{23} + c_m - \frac{A_n v_{on}}{2}(1-n)^2 - A_n v_{on}(1-n)(x-1) \quad (15)$$

In region 6 ($x \ge x_{satn}$), the NMOS device is entering the linear region and the PMOS is off. The analytical solution of eqn. 2 is

$$x - x_{satn} = \frac{1 + 2v_{on}^{-1}(1-n)}{A_n(1-n)} \ln\left[\frac{2(1-n) - u_{out}}{2(1-n) - u_{maxn}}\right] - \frac{1}{A_n(1-n)} \ln\left[\frac{u_{out}}{u_{maxn}}\right] \quad (16)$$

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

and x_{satn} is calculated from eqn. 15 for $u_{out} = u_{maxn}$.

2.2 Slow input ramps

For slow input ramps, the NMOS device leaves saturation while the input voltage is still a ramp. The output expressions for regions 1, 2, 3 and 4 are the same as those for the fast inputs. In this case, the normalised time value x_{satn} is calculated from eqn. 14 for

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

which corresponds to the NMOS saturation line (Fig. 2). For slower input ramps, the inverter does not enter 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 eqn. 13.

In region 5B ($x_{satn} \le x \le 1$), the NMOS device is in the linear region and the PMOS transistor is either off or so poorly conducting that its influence can be neglected. SPICE simulations indicate that the PMOS transistor current in this region (for $x < 1 - p$) is up to 2-3% of the NMOS transistor current. Neglecting the charging current through the gate–drain coupling capacitance and using at the denominator of the NMOS current an average value of the normalised output voltage equal to (u_{satn}/2), an approximated solution of eqn. 2 is

$$u_{out} = e^{-y^2}\left[\frac{1}{u_{satn}e^{y_{satn}^2}} - \sqrt{\frac{\pi A_n}{4(2 + u_{satn}v_{on}^{-1})}} \times (\text{erf}[y_1] - \text{erf}[y_{satn}])\right]^{-1} \quad (18)$$

$$\text{where } y_1 = \sqrt{\frac{A_n}{2 + u_{satn}v_{on}^{-1}}}(x-n),$$

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

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

erf [y_1] and erf [y_{satn}] are the error functions of y_1 and y_{satn}, respectively.

In 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 (eqn. 2) is the given by eqn. 16 if we substitute x_{satn} by 1 and u_{maxn} by $u_{[1]}$. $u_{[1]}$ is the value of the normalised output voltage when the input ramp reaches its final value and is calculated if we set $x = 1$ in eqn. 18.

3 Propagation delay analysis

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

$$t_d = t_{0.5} - (\tau/2) = x_{0.5}\tau - (\tau/2) \quad (19)$$

where $x_{0.5}$ is the normalised time value when $u_{out} = 0.5$. Thus, for the evaluation of the propagation delay, the normalised time value $x_{0.5}$ must be determined from the derived output waveform expressions, for both cases of input ramps. As we can see in Fig. 2, a critical parameter for finding 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}). Hence, it is necessary to consider two possibilities in the delay calculation: $u_{maxn} \le 0.5$ and $u_{maxn} \ge 0.5$.

For $u_{maxn} \le 0.5$, 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 normalised output voltage when the input ramp reaches its final value. 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 normalised output voltage when the PMOS device is entering the cut-off region.

For $u_{maxn} \ge 0.5$, in the case of fast input ramps, $u_{out} = 0.5$ occurs in region 6. In the case of slow input ramps, the output voltage reaches the 50% level when the inverter operates in region 6 if $u_{[1]} \ge 0.5$. If $u_{[1]} \le 0.5$, there are two possibilities for the region in which $u_{out} = 0.5$: either $u_{satn} \le 0.5$ when the output voltage reaches the 50% level in region 3, or $u_{satn} \ge 0.5$ in region 5B.

In all the above cases (except the last one), $x_{0.5}$ becomes the root of simple first-order or quadratic equations. In the last case, 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

$$x_{0.5} = x_{satn} + \frac{0.5 - u_{satn}}{d} \quad (20)$$

where $d = (du_{out}/dx)|x = x_{satn}$ is the slope of the output waveform.

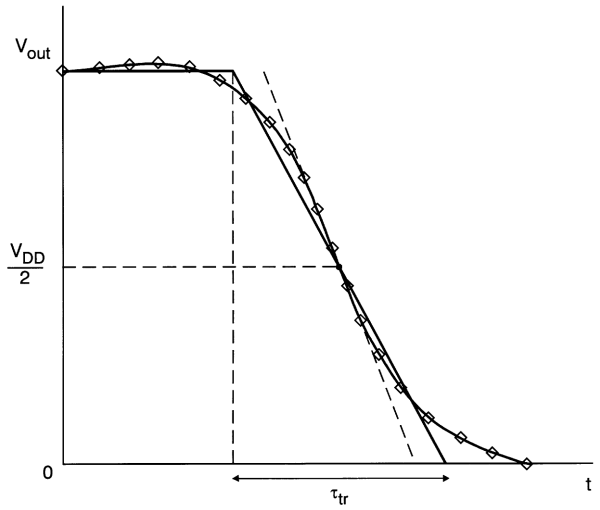

Fig. 4 *Evaluation of effective output transition time*
——— ramp approximation of V_{out}
◇ V_{out}

In real CMOS datapaths, the input signal of a gate is not a ramp but the output waveform of the preceding gate. For the derived ramp delay model to be applicable to a CMOS gate chain, we must approximate the real input waveform by a ramp waveform to obtain an effective transition time. A good approximation [3, 10] for the evaluation of the effective output transition time (τ_{tr}) of the inverter is achieved when the output waveform slope is approximated by 70% of its derivative at the point which corresponds to the half supply voltage level (Fig. 4). This percentage has been obtained from ring oscillator simulations. τ_{tr} can be used as τ for the succeeding inverter in the circuit. After that, the effective output transition time may be written as

$$\tau_{tr} = \frac{V_{DD}}{0.7\left|\frac{dV_{out}}{dt}\right|_{t=t_{0.5}}} = \frac{\tau}{0.7s} \qquad (21)$$

where $s = |du_{out}/dx|x = x_{0.5}|$ is calculated from the derived expressions of the inverter output.

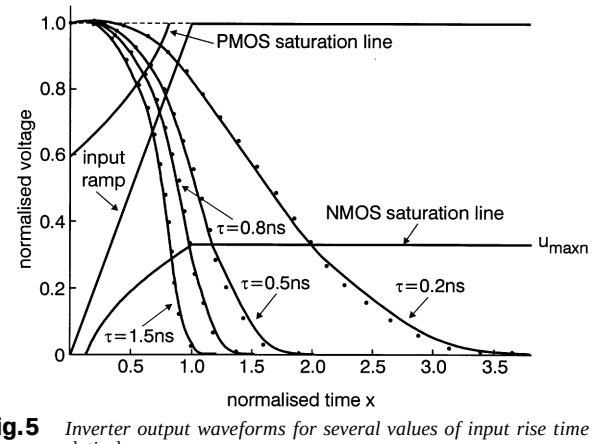

Fig. 5 *Inverter output waveforms for several values of input rise time*
——— analytical
····· SPICE (0.5μm)
C_L = 0.2pF

4 Results and conclusions

Fig. 5 show some typical output waveforms, produced from the derived expressions. A commercial submicron CMOS process technology of 0.5μm has been used to validate the accuracy of the presented inverter output 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 level-3 simulations are added for comparison. The basic parameters of the SPICE model are given in Table 2. A supply voltage of 5V 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 display output waveforms for several input rise times in the same Figure, the normalised output voltage is plotted as a function of the normalised time ($x = t/\tau$). The output waveforms for input rise times 0.2ns and 0.5ns correspond to the case for fast inputs (Section 2.1), whereas those for input times 0.8ns and 1.5ns correspond to the slow inputs (Section 2.2). As can be seen, the slope of the output waveforms for the fast inputs is less than the input slope, and for the slow inputs it is greater than the input slope.

Table 1: Basic MOSFET model parameters used in calculations

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

Table 2: Basic SPICE level-3 parameters used in simulations

Parameter	NMOS	PMOS
transconductance parameter K_P (A/V^2)	1.965×10^{-4}	4.874×10^{-5}
gate oxide thickness T_{OX} (m)	9.6×10^{-9}	9.6×10^{-9}
maximum drift velocity of carries V_{MAX} (m/s)	2.008×10^{5}	2.542×10^{5}
factor for channel-length modulation equation K (V^{-1})	0.015	0.055
zero-bias threshold voltage V_{TO} (V)	0.657	0.921
surface inversion potential ϕ (V)	0.7	0.7
body effect parameter γ (V$^{1/2}$)	0.5976	0.4673
surface mobility U_O (cm^2/V.s)	546.2	135.5
substrate doping N_{SUB} (cm^{-3})	1.392×10^{17}	8.512×10^{16}
gate-drain overlap capacitance C_{GDO} (F/m)	3.050×10^{-10}	2.400×10^{-10}
gate-source overlap capacitance C_{GSO} (F/m)	3.050×10^{-10}	2.400×10^{-10}
gate-bulk overlap capacitance C_{GBO} (F/m)	4.024×10^{-10}	3.758×10^{-10}
zero-bias junction capacitance C_J (F/m^2)	5.62×10^{-4}	9.35×10^{-4}
zero-bias perimeter capacitance C_{JSW} (F/m)	5×10^{-11}	2.89×10^{-10}

The inverter propagation delay for a rising input ramp is plotted as a function of $A_{no} = (\beta_n V_{DD}\tau)/C_L$ in

Fig. 6. 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 determine the relation between the input and the output waveform. The results for $A_{no} < 15$ correspond to fast inputs (Section 2.1) compared to the output waveforms and for $A_{no} > 15$ correspond to slow inputs (Section 2.2). Results using other approaches for the evaluation of the propagation delay [3, 8, 9, 12] are also shown. It can be observed that our model produces 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–drain coupling capacitance on the expressions of the inverter output waveform. Results with similar accuracy can be obtained for reduced values of the supply voltage (e.g. 3.3V).

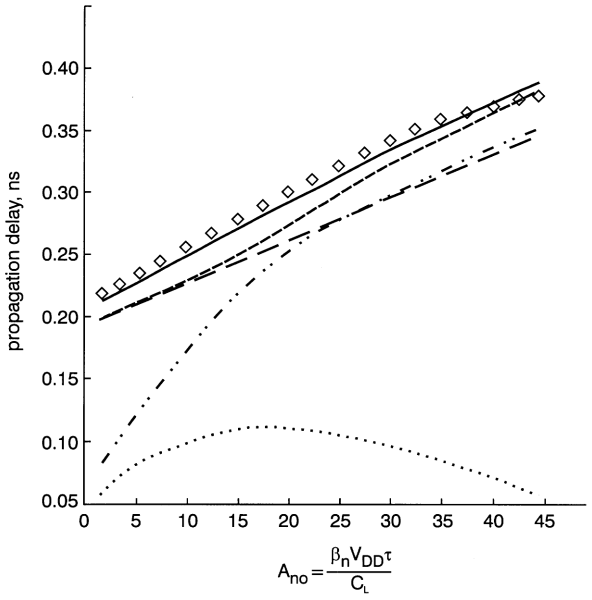

Fig. 6 *Inverter propagation delay results*
◇ SPICE (0.5) μm
——— proposed approach
..... [3]
·· — ·· [8]
– – – – [9]
- - - - [12]

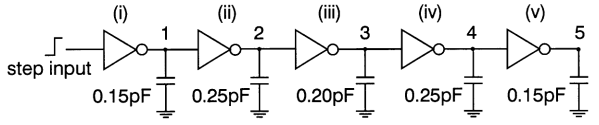

Fig. 7 *Example of CMOS inverter chain*
(i) W_n = 3μm, Wp = 6.5μm
(ii) W_n = 5μm, Wp = 10.5μm
(iii) W_n = 3μm, Wp = 6.5μm
(iv) W_n = 4μm, Wp = 8.5μm
(v) W_n = 2μm, Wp = 4.5μm

The error in the case of falling input is a little higher (< 5.5%) than that of rising input, mainly due to the higher channel-length modulation of the PMOS device. The accuracy of our model for both cases of rising and falling input is validated with the chain example of Fig. 7. The effective output transition time of each inverter is calculated as in Section 3. The inverters of the chain have different loads and drives. The results derived from the analytical model and those produced from SPICE simulations are compared in Fig. 8.

As the minimum feature sizes for CMOS circuits scale downward, the influence on the performance of the resistive load component is increased. The proposed timing model, which uses a purely capacitive load, can be interfaced with methodologies considering the effects of an RC interconnect load [18].

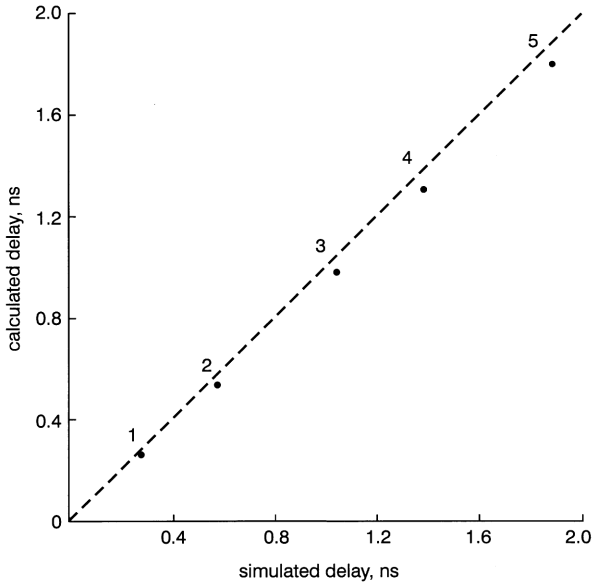

Fig. 8 *Comparison of calculated and simulated chain delay*

5 References

1 NABAVI-LISHI, A., and RUMIN, N.C.: 'Inverter models of CMOS gates for supply current and delay evaluation', *IEEE Trans. Computer-Aided Des.*, 1994, 13, (10), pp. 1271–1279
2 BURNS, J.R.: 'Switching response of complementary symmetry MOS transistor logic circuits', *RCA Rev.*, 1964, 25, pp. 627–661
3 HEDENSTIERNA, N., and JEPPSON, K.O.: 'CMOS circuit speed and buffer optimization', *IEEE Trans. Computer-Aided Des.*, 1987, CAD-6, (2), pp. 270–281
4 KAYSSI, A.I., SAKALLAH, K.A., and BURKS, T.M.: 'Analytical transient response of CMOS inverters', *IEEE Trans. Circuits Syst. I*, 1992, 39, (1), pp. 42–45
5 JEPPSON, K.O.: '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*, 1994, 29, (6), pp. 646–654
6 BISDOUNIS, L., NIKOLAIDIS, S., and KOUFOPAVLOU, O.: 'Propagation delay and short-circuit power dissipation modeling of the CMOS inverter', *IEEE Trans. Circuits Syst. I*, 1998, 45, (3), pp. 259–270
7 SHIH, Y.H., and KANG, S.M.: 'Analytic transient solution of general MOS circuit primitives', *IEEE Trans. Computer-Aided Des.*, 1992, 11, (6), pp. 719–731
8 NABAVI-LISHI, A., and RUMIN, N.C.: 'Simultaneous delay and maximum current calculation in CMOS gates', *Electron. Lett.*, 1992, 28, (7), pp. 682–684
9 SAKURAI, T., and NEWTON, A.R.: 'Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas', *IEEE J. Solid-State Circuits*, 1990, 25, (2), pp. 584–594
10 SAKURAI, T., and NEWTON, A.R.: 'Delay analysis of series-connected MOSFET circuits', *IEEE J. Solid-State Circuits*, 1991, 26, (2), pp. 122–131
11 DUTTA, S., MAHANT SHETTI, S.S., and LUSKY, S.L.: 'A comprehensive delay model for CMOS inverters', *IEEE J. Solid-State Circuits*, 1995, 30, (8), pp. 864–871
12 EMBABI, S.H.K., and DAMODARAN, R.: 'Delay models for CMOS, BiCMOS, BiNMOS circuits and their applications for timing simulations', *IEEE Trans. Computer-Aided Des.*, 1994, 13, (9), pp. 1132–1142
13 KONG, J.T., and OVERHAUSER, D.: 'Methods to improve digital MOS macromodel accuracy', *IEEE Trans. Computer-Aided Des.*, 1995, 14, (7), pp. 868–881
14 HIRATA, A., ONODERA, H., and TAMARU, K.: 'Estimation of short-circuit power dissipation and its influence on propagation delay for static CMOS gates'. IEEE international symposium on *Circuits and systems*, ISCAS'96, Atlanta, USA, May 1996, Vol. 4, pp. 751–754
15 SHOJI, M.: 'CMOS digital circuit technology' (Prentice–Hall, New Jersey, 1988), pp. 29–32, 121–123
16 WESTE, N.H.E., and ESHRAGHIAN, K.: 'Principles of CMOS VLSI design: a systems perspective' (McGraw–Hill, New York, 1993), pp. 183–191

17 NELSON, A.L., FOLLEY, K.W., and CORAL, M.: 'Differential equations' (D.C. Heath and Company, Boston, 1964), pp. 180–214
18 DARTU, F., MENEZES, N., and PILEGGI, L.T.: 'Performance computation for precharacterized CMOS gates with RC loads', *IEEE Trans. Computer-Aided Des.*, 1996, 15, (5), pp. 544–553

6 Appendix: Extraction of model parameter V_O

A simple method is described of extracting the voltage which specifies the effects of carriers' velocity saturation V_O. First, in the measured I–V curves (Fig. 9) of the transistor, select a point A at the saturation region for a particular gate-source voltage (V_{GS5}) and measure the drain current (I_{DA}). Point A is selected at the middle of the saturation region in order to minimise the effect of channel-length modulation on the model accuracy. From the simple expression of the saturation current model (eqn. 4) in the submicron range, the parameter V_O is then extracted

$$V_O = \frac{I_{DA}}{\beta(V_{GS5} - V_{TH})} \tag{22}$$

where β the device gain factor and V_{TH} is the device threshold voltage. Note that V_O is independent of the transistor width for a given process technology, and depends only on the device channel length, the effective mobility and the maximum saturation velocity of the carriers [15].

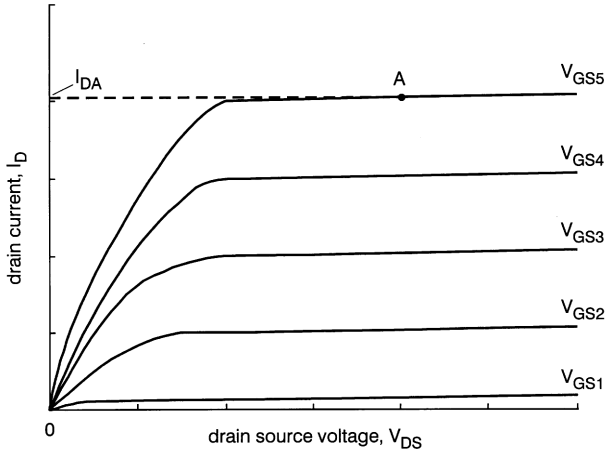

Fig. 9 *Extraction of model parameter V_O*