INTERNATIONAL JOURNAL OF CIRCUIT THEORY AND APPLICATIONS
Int. J. Circ. Theor. Appl. 2002; **30**:547–566 (DOI: 10.1002/cta.207)

Efficient output waveform evaluation of a CMOS inverter based on short-circuit current prediction

A. Chatzigeorgiou[1,*,†] and S. Nikolaidis[2,‡]

[1] *Department of Applied Informatics, University of Macedonia, 54006 Thessaloniki, Greece*
[2] *Department of Physics, Aristotle University of Thessaloniki, 54006 Thessaloniki, Greece*

SUMMARY

A novel approach for obtaining the output waveform, the propagation delay and the short-circuit power dissipation of a CMOS inverter is introduced. The output voltage is calculated by solving the circuit differential equation only for the conducting transistor while the effect of the short-circuit current is considered as an additional charge, which has to be discharged through the conducting transistor causing a shift to the output waveform. The short-circuit current as well as the corresponding discharging current are accurately predicted as functions of the required time shift of the output waveform. A program has been developed that implements the proposed method and the results prove that a significant speed improvement can be gained with a minor penalty in accuracy. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: propagation delay; short-circuit power; CMOS inverter; timing analysis

1. INTRODUCTION

The development of digital integrated circuits with short design cycles requires accurate and fast timing and power simulation. Unfortunately, simulators which employ numerical methods, such as SPICE are prohibitively slow for large designs. The need for analytical methods which can produce accurate results at short times is obvious and extended research has been conducted for the CMOS inverter [1–10] which forms a basic block to which all CMOS structures can be diminished. Since complex CMOS gates can be reduced to an equivalent inverter that has the same performance [11,12] it is sufficient to seek accurate and efficient models for the transient analysis of the inverter. Another important reason for focusing on the CMOS inverter is its extensive use in clock distribution networks and buses in most integrated circuits and the corresponding power that is consumed on these inverters [13].

Analytical modelling techniques at the inverter level are very accurate but this comes at the cost of increased complexity. The analysis of the inverter is complicated mainly due to

[*] Correspondence to: Dr. A. Chatzigeorgiou, Department of Applied Informatics, University of Macedonia, 156 Egnatia, St., 54006 Thessaloniki, Greece.
[†] E-mail: alec@ieee.org
[‡] E-mail: snikolaid@physics.auth.gr

Received 29 November 2001
Revised 1 March 2002

the presence of the short-circuit current which acts parasitically on the output evolution. The differential equation which describes the operation of the circuit has to be solved taking into account the current of both the pMOS and the nMOS transistor. This leads to complicated expressions for the output waveform increasing significantly the execution time.

Analytical techniques for the modelling of the CMOS inverter aim at obtaining an output waveform expression for a ramp input, for each region of operation which is determined by the mode of operation of each transistor and the status of the input. Several methodologies have been developed which can be categorized according to the transistor current model used (Shichman–Hodges square law [1], *n*th power law [2], alpha power law [3,4], bulk charge square law [5]) and according to the parameters and second-order effects which have or have not been taken into account (negligible short-circuit current in Reference [1], input-to-output coupling capacitance [6] and carrier velocity saturation effects [4]). An integration of all effects into one method has been presented in References [3,8]. A common drawback of all these techniques is that they result in complicated expressions for the output waveform thus requiring increased computational power and execution time in case of a simulator. Macromodelling techniques for the calculation of the delay and short-circuit power dissipation of CMOS structures have been presented in References [9,10], respectively, where parameters such as input transition time, input-to-output coupling and second-order effects of submicron transistors are incorporated in closed form macromodels.

In this paper a method which takes into account all of the above-mentioned parameters and second-order effects avoiding however the intricacy of the short-circuit current by solving the differential equation that describes the circuit operation considering only the conducting transistor is introduced. The effect of the short-circuit current is taken into account as an additional charge which has to be discharged through the conducting transistor. This charge causes a shift on the output waveform that has been obtained considering only the conducting transistor. The ‘translation’ of the effect of the short-circuit current to a time delay has also been discussed in a simplified manner in Reference [6] in order to calculate the propagation delay. However, the charge supplied by the short-circuiting transistor is calculated using fitting methods on SPICE simulation results resulting in a semi-empirical method, which is inaccurate for submicron technologies. In the approach proposed in References [7,13] the short-circuit current has been treated as an additional charge which however was calculated on an initial estimate of the output waveform, thereby resulting in large errors.

The method for the evaluation of the output response by shifting the initially obtained waveform is described in Section 2 while in Section 3 the proposed method is extended in order to calculate the short-circuit energy dissipation of an inverter for a single transition. Results of the proposed method in terms of accuracy and execution time and comparisons with SPICE and other approaches are given in Section 4. Finally, we conclude in Section 5.

2. OUTPUT WAVEFORM EVALUATION

Let us consider the inverter of Figure 1(a) whose operation is described for a rising input ramp with transition time τ by the following differential equation:

$$C_L \frac{\mathrm{d}V_{\mathrm{out}}}{\mathrm{d}t} = i_{C_M} + i_{\mathrm{p}} - i_n \tag{1}$$

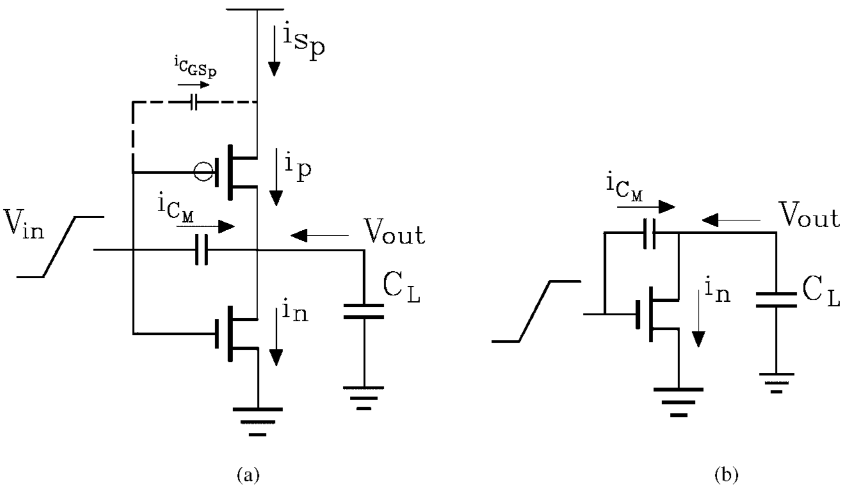

Figure 1. (a) Actual inverter, (b) conducting transistor.

where

$$i_{C_M} = C_M\left(\frac{\mathrm{d}V_{\mathrm{in}}}{\mathrm{d}t} - \frac{\mathrm{d}V_{\mathrm{out}}}{\mathrm{d}t}\right)$$

is the current through the coupling capacitance C_M between input and output [3,14]. The case of a falling input ramp is symmetrical. If the short-circuit current i_{p} is neglected (Figure 1(b)), then the output waveform can be obtained by solving the above equation according to the regions of operation of the conducting nMOS transistor. This waveform will be referred to as initial waveform, $V_{\mathrm{out}_{\mathrm{init}}}(t)$. For submicron devices the output waveform is obtained with excellent accuracy using the alpha-power law model for the transistor currents (here presented for an nMOS transistor) [4]:

$$I_D = \begin{cases} 0 & V_{\mathrm{GS}} \leqslant V_{\mathrm{TN}}\text{: cutoff region} \\ k_{\mathrm{l}}(V_{\mathrm{GS}} - V_{\mathrm{TN}})^{a/2} V_{\mathrm{DS}} & V_{\mathrm{DS}} < V_{\mathrm{DSAT}}\text{: linear region} \\ k_{\mathrm{s}}(V_{\mathrm{GS}} - V_{\mathrm{TN}})^{a} & V_{\mathrm{DS}} \geqslant V_{\mathrm{DSAT}}\text{: saturation region} \end{cases} \tag{2}$$

where V_{DSAT} is the drain saturation voltage, $k_{\mathrm{l}}, k_{\mathrm{s}}$ are the transconductance parameters, α is the carrier velocity saturation index and V_{TN} is the threshold voltage. It should be mentioned that in the solution of the differential equation for the single transistor, the coupling capacitance C_M should also take into account the pMOS transistor gate-to-drain capacitance. The expressions for $V_{\mathrm{out}_{\mathrm{init}}}(t)$ are given in Appendix A.

Calculating the output waveform considering only the conducting current leads not only to simpler expressions for the output waveform and thus to shorter execution time but is also substantially more simple in terms of program complexity. In Figure 2 the decision diagrams and the operating regions through which the output evolves for the fully analytical method [3] where both transistors are considered and for the proposed method are shown. Each

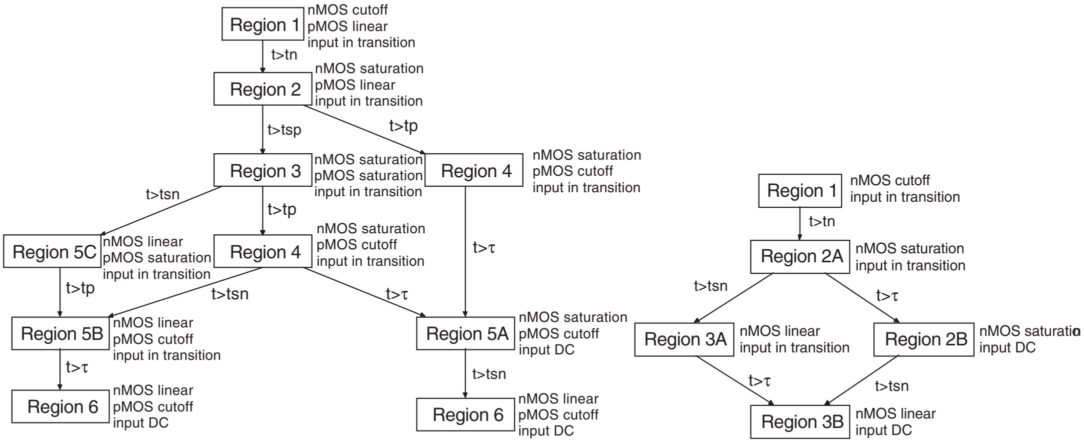

Figure 2. Decision diagrams for (a) fully analytical and (b) proposed method.

block represents a region of operation according to the mode of operation of the conducting and short-circuiting transistor (for the case of an inverter) and the status of the input. On each edge the condition which must be fulfilled in order for the corresponding transition from one operating region to another to take place is shown. A program for the simulation of each structure, inverter or single transistor, should be able to distinguish between the operating regions by branching according to the specified conditions. Obviously, the output voltage value at the region boundaries should be calculated for the transient simulation of each structure.

Timing analysis of a CMOS inverter can be performed either by calculating the output voltage at each time step thus obtaining the full output waveform or by simply calculating the propagation delay and the slope of the output at the half-V_{DD} point in order to define an equivalent ramp for the output waveform, that can be fed as input to the next stage [1]. The contribution of the program complexity in terms of region boundaries and branching between them on the total execution time is more intense when only the propagation delay is calculated rather than the full output waveform.

The discharging of the output when the short-circuit current is neglected is faster than in the actual inverter since the short-circuit current reduces the effective discharging current slowing down the output evolution. The effect of the pMOS short-circuit current can be considered as an additional charge, Q_{ad}, at the output load causing an additional delay, t_{ad}, to the output evolution. This proposition is based on the fact that the slope variation at the half-V_{DD} point between the output waveform of a single transistor driving an output load and that of an inverter driving the same load is very small, and thus it can be considered for modelling purposes that the output waveform is shifted. In Figure 3 the slope of the output waveform of an inverter at the half-V_{DD} point over the nominal slope, which is the slope of the output considering only the conducting transistor is displayed for two load capacitance values. As it can be observed, this ratio remains close to unity, validating the proposed approach.

The delay t_{ad} is calculated as the time required to discharge Q_{ad} through the conducting nMOS device.

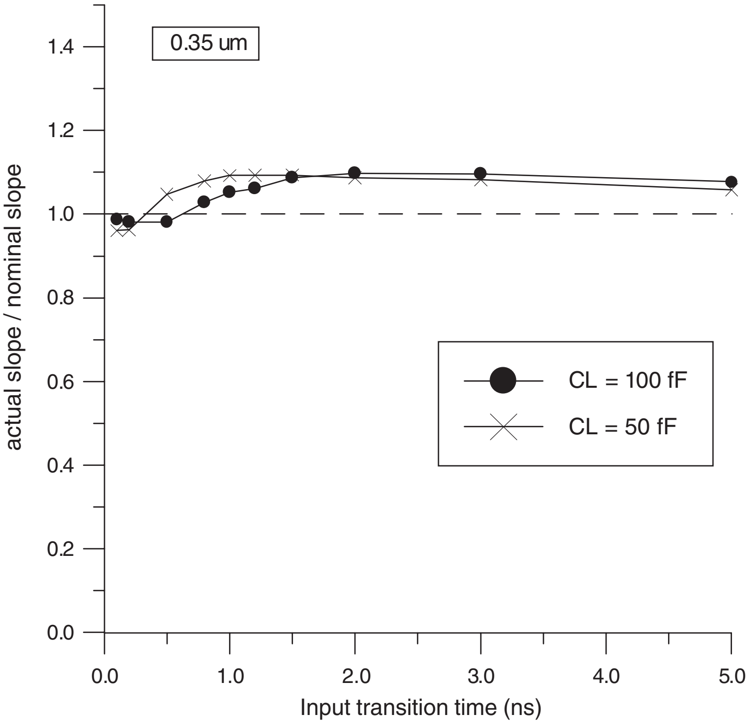

Figure 3. Slope variation between single transistor and inverter at $V_{DD}/2$ (HP 0.35 μm, $W_n = 1.4$ μm, $W_p = 2.1$ μm, $V_{DD} = 3.3$ V).

Consequently, the actual output waveform can be approximated by shifting the initial waveform by t_{ad}:

$$V_{out}(t) = V_{out_{init}}(t - t_{ad}) \tag{3}$$

However, the problem is that if the short-circuit current is calculated on the initial waveform, the charge would be overestimated while the corresponding nMOS discharging current, I_{dch}, would be underestimated leading to an overestimated additional delay, t_{ad_0} (Figure 4). That is because the initial waveform evolves faster than that of the actual inverter resulting in a larger V_{DS} value for the pMOS transistor and a smaller V_{DS} value for the nMOS transistor at each time point. To overcome this problem the proposed technique is employed: The pMOS short-circuit current and the corresponding discharging nMOS current are calculated on V_{out} and therefore result as functions of the additional delay. Thus, t_{ad}, can be obtained by solving

$$t_{ad} = \frac{Q_{ad}(t_{ad})}{I_{dch}(t_{ad})} \tag{4}$$

Consequently, the key point in the proposed approach is the accurate approximation of the pMOS and nMOS transistor currents as simplified functions of the additional delay, t_{ad}.

The actual pMOS short-circuit current can be approximated by a piece-wise linear function of time [2,15] (Figures 5 and 6). The charge that it contributes can be easily calculated as

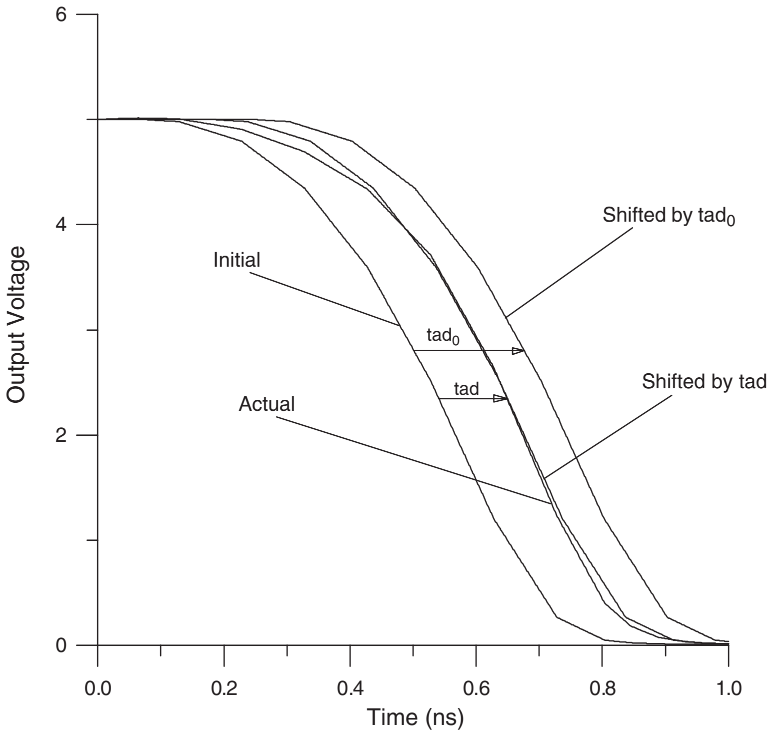

Figure 4. Initial, actual and shifted waveforms.

the area of the corresponding triangle which is equal to

$$Q_{\mathrm{ad}} = I_{\mathrm{p_{max}}}(t_{\mathrm{e}} - t_{\mathrm{s}})/2 \tag{5}$$

$I_{\mathrm{p_{max}}}$ is the maximum value of the pMOS current and occurs when the pMOS transistor enters saturation at $t = t_{\mathrm{sp}}$. The pMOS current initially presents an undershoot which is due to the coupling capacitance between input and ouput which in turn causes an overshoot in the output voltage (Figure 7) [2,3,6,9,10,16]. During the output voltage overshoot, current is flowing towards V_{DD} causing the undershoot of the pMOS current. The minimum value of the pMOS current occurs when the subthreshold current flowing through the nMOS transistor obtains a considerable value. The time at which the pMOS current is considered to start, t_{s}, is equal to the time point when the overshoot ceases, t_{ov} [9,10]. The calculation of the starting point of the pMOS current, taking into account the subthreshold current of the nMOS device, is given in Appendix B. The pMOS current ceases at time $t_{\mathrm{e}} = (V_{\mathrm{DD}} - |V_{\mathrm{TP}}|)\tau/V_{\mathrm{DD}}$, where V_{TP} is the threshold voltage of the pMOS transistor.

According to the above, both the conducting and the short-circuit current determine the time shift that is caused by the additional charge provided by the short-circuit current. The shape and the magnitude of the transistor currents in an inverter depend on the position of the time points where they change operating regions (i.e. the time when a transition from saturation to the linear region or *vice-versa* occurs). Since the current values have to be calculated on the actual output waveform, which is unknown, the corresponding time points are calculated on

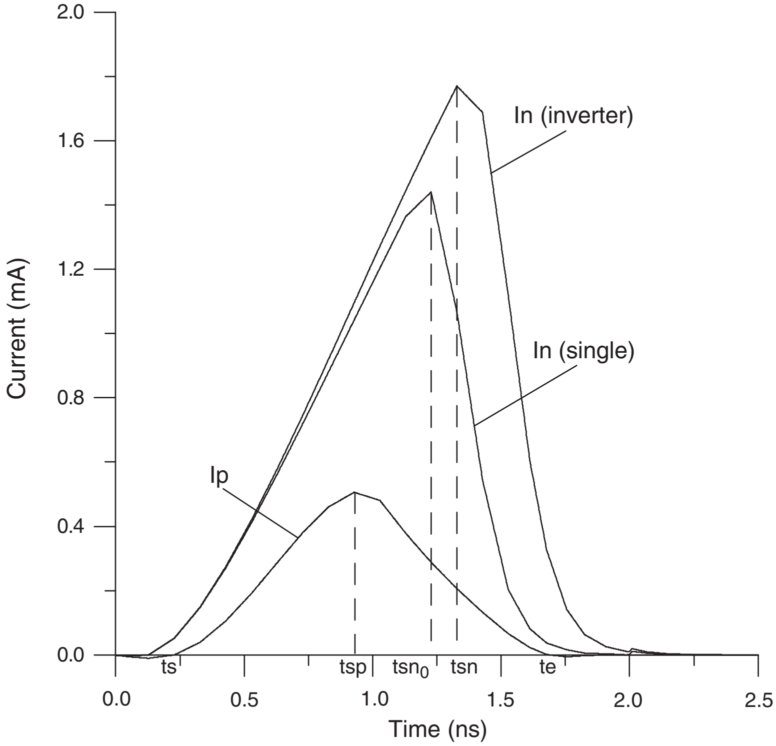

Figure 5. pMOS and nMOS currents.

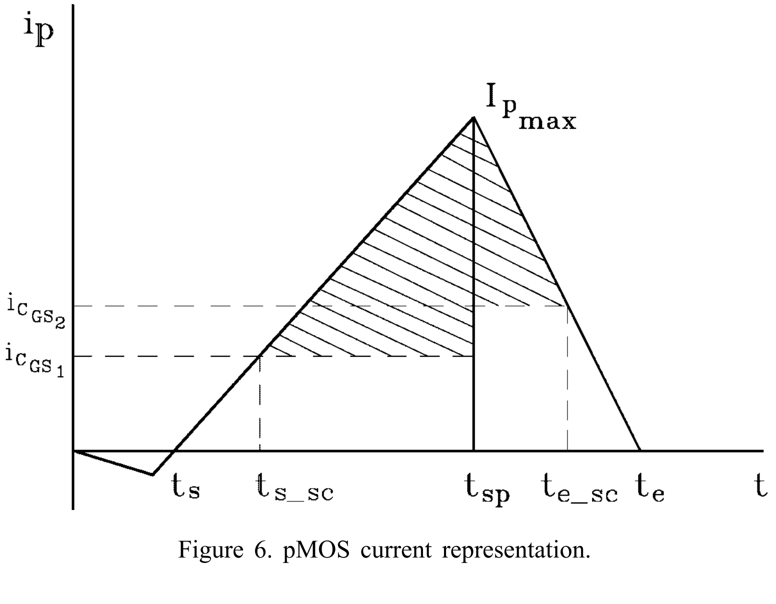

Figure 6. pMOS current representation.

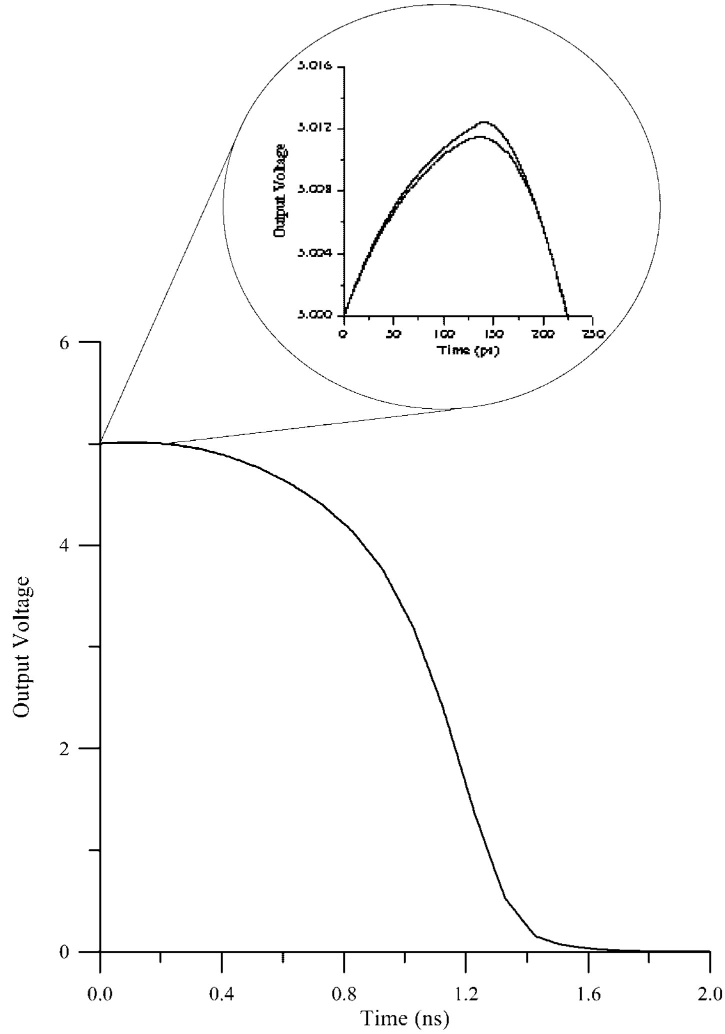

Figure 7. Output waveform comparison between simulated and calculated values during the initial voltage overshoot ($W_n = 4\,\mu m$, $W_p = 6\,\mu m$, $C_L = 100$ fF, $\tau = 2$ ns, HP 0.5 μm technology).

the initial waveform and then shifted appropriately. For this reason, linear approximations of the initial waveform and the drain saturation voltages for the regions of interest will be used next, in order to express the specific time points of the actual waveform as functions of t_{ad} and the corresponding time points in the initial waveform.

Time point t_{sp_0} when the drain saturation voltage line of the pMOS transistor crosses the initial output waveform (Figure 8) is found by solving $V_{DSAT_p}(t) = V_{DS_{p_{init}}}(t) = |V_{out_{init}}(t) - V_{DD}|$ where V_{DSAT_p} is the drain saturation voltage of the pMOS transistor [4] while for $V_{out_{init}}(t)$ the expressions derived in Appendix A are used. Approximating the drain saturation voltage by its first-order Taylor series around point t_{sp_0} as $V_{DSAT_p}(t) \approx s_1 - s_2 t$ and the output voltage by

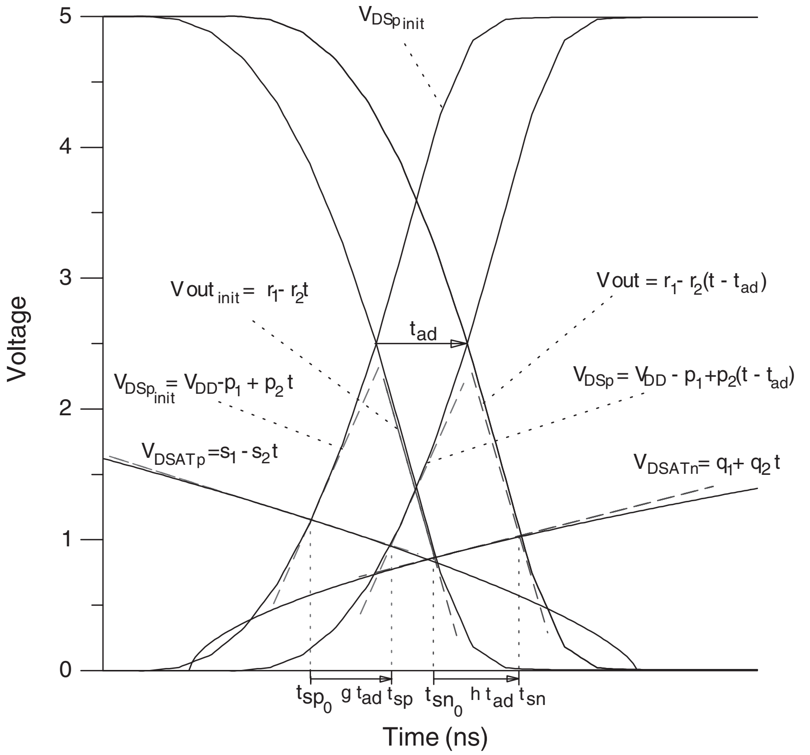

Figure 8. Actual and approximated drain-to-source voltages and drain saturation voltages for the pMOS and nMOS transistors.

its first-order Taylor series around the same point as $V_{\text{out}_{\text{init}}}(t) \approx p_1 - p_2 t$ (Figure 8), enables the calculation of t_{sp} (the time when the pMOS transistor enters saturation according to the shifted waveform) as a function of t_{sp_0} and t_{ad}. Solving $V_{\text{DSAT}_\text{p}}(t) = V_{\text{DS}_{\text{p}_\text{init}}}(t)$ using the linear approximations, t_{sp_0} can be expressed as

$$t_{\text{sp}_0} = \frac{p_1 + s_1 - V_{\text{DD}}}{p_2 + s_2} \tag{6}$$

The shifted output waveform according to the linear approximation will be $V_{\text{out}}(t) \approx p_1 - p_2(t - t_{\text{ad}})$ (Figure 8), and solving $V_{\text{DSAT}_\text{p}}(t) = V_{\text{DS}_\text{p}}(t) = V_{\text{DD}} - p_1 + p_2(t - t_{\text{ad}})$ for the shifted waveform and using the linear approximations results in:

$$t_{\text{sp}} = t_{\text{sp}_0} + g t_{\text{ad}} \tag{7}$$

where $g = p_2/(p_2 + s_2)$. It has to be noted that although the output waveform is shifted by t_{ad}, time point t_{sp} is shifted only by a fraction of t_{ad} due to the non-zero slope of V_{DSAT_p}. The possible mismatch between the actual time point when the pMOS transistor enters saturation according to the actual waveform and the calculated one is due to the fact that although the shifted waveform matches very well the actual waveform around the half-V_{DD} point, it presents a small deviation around the point where t_{sp} is calculated as it can be observed in

Table I. Accuracy and speed comparison between calculation and SPICE (HP 0.5 μm).

τ (ns)	C_L (fF)	Prop. delay SPICE (ns)	Prop. delay calc. (ns)	Error (%)	Exec. time SPICE (s)	Exec. time calc. (s)	Speedup
0.5	50	0.076	0.071	6.58	0.23	0.00164	140
0.5	100	0.129	0.127	1.55	0.16	0.00156	103
1	50	0.061	0.059	3.28	0.19	0.00181	105
1	100	0.138	0.135	2.17	0.16	0.00173	92
1.5	50	0.043	0.049	13.95	0.21	0.00198	106
1.5	100	0.127	0.136	7.09	0.19	0.00195	97
2	100	0.113	0.109	3.54	0.15	0.00208	72
2	200	0.263	0.268	1.90	0.27	0.00191	141
3	100	0.069	0.075	8.70	0.17	0.00239	71
3	200	0.244	0.248	1.64	0.16	0.00229	70

Figure 4. For the cases presented in Table I the average error in the calculation of t_{sp} due to this approximation is 5.5%. However, from the overall accuracy of the method it can be concluded that the error that is introduced by these approximations is limited and fully justifiable since it enables a very fast and accurate calculation of the output waveform. The maximum value of the pMOS short-circuit current of the inverter occurs at time point t_{sp} because after that point the pMOS transistor enters saturation and since its V_{GS} is decreasing the current will also be decreasing. Thus, the maximum value of the pMOS current can be approximated by substituting in the current expression for saturation time point t_{sp}

$$I_{p_{max}} = k_{s_p}(V_{DD} - V_{in}[t_{sp_0} + g t_{ad}] - |V_{TP}|)^{a_p} \tag{8}$$

where k_{s_p} is the transconductance of the pMOS transistor in saturation.

Similarly, the time when the nMOS transistor exits saturation according to the initial waveform, t_{sn_0}, can be found by solving $V_{out_{init}}(t) = V_{DSAT_n}(t)$. Approximating the drain saturation voltage of the nMOS transistor by its first-order Taylor series around the time point t_{sn_0} as $V_{DSAT_n}(t) \approx q_1 + q_2 t$ (Figure 8) and the initial output waveform by its first-order Taylor series around the same point as $V_{out_{init}}(t) \approx r_1 - r_2 t$, the time point when the nMOS transistor exits saturation according to the shifted waveform, t_{sn}, can be expressed as $t_{sn} = t_{sn_0} + h\, t_{ad} (h = r_2/(r_2 + q_2))$.

The effect of the pMOS current is to prolong the saturation region of the nMOS transistor resulting in an increase of its current (Figure 5). Therefore, the discharging current for the additional charge which determines the time shift of the initial output waveform is approximated by the average of the maximum nMOS current that corresponds to V_{out} and that corresponding to the initial waveform. Substituting in the current expression for saturation time points t_{sn_0} and t_{sn} results in:

$$I_{dch} = \frac{k_{s_n}(V_{in}[t_{sn_0} + h\, t_{ad}] - V_{TN})^{a_n} + k_{s_n}(V_{in}[t_{sn_0}] - V_{TN})^{a_n}}{2} \tag{9}$$

where k_{s_n} is the transconductance of the nMOS transistor in saturation.

Since currents are almost linear functions of time [2] Equation (4) can be solved for t_{ad} with sufficient accuracy using first-order Taylor series approximations for the pMOS and nMOS currents around the point $t_{ad} = t_{ad_0}/2$ as it will be verified by the results.

3. SHORT-CIRCUIT POWER ESTIMATION

Since the form and the magnitude of the pMOS current in the above analysis is obtained, the proposed method can also be directly applied in order to calculate the short-circuit energy which is dissipated when the output of the inverter switches state. The dissipated short-circuit energy during output discharging is calculated as

$$E_{\mathrm{SC}}^{\mathrm{d}} = V_{\mathrm{DD}} \cdot \int_{t_{\mathrm{s\_sc}}}^{t_{\mathrm{e\_sc}}} i_{\mathrm{s_p}}(t)\,\mathrm{d}t \tag{10}$$

where the current that is causing the short-circuit power dissipation, $i_{\mathrm{s_p}}$, is not the pMOS transistor current but the current that is flowing from V_{DD} towards the source of the pMOS transistor [16] (Figure 1(a)). The pMOS transistor current, i_{p}, carries both the actual short-circuit current $i_{\mathrm{s_p}}$ which causes the short-circuit power dissipation and the current $i_{C_{\mathrm{GS_p}}}$ that is flowing through the gate-to-source capacitance $C_{\mathrm{GS_p}}$ which is part of the dynamic power dissipation of the previous gate. Therefore, the actual short-circuit current is isolated by applying Kirchhoff's current law at the source node of the short-circuiting pMOS transistor:

$$i_{\mathrm{s_p}} = i_{\mathrm{p}} - i_{C_{\mathrm{GSp}}} \tag{11}$$

where $i_{C_{\mathrm{GSp}}} = C_{\mathrm{GS_p}}\mathrm{d}V_{\mathrm{in}}/\mathrm{d}t$. Since the gate-to-source capacitance has two different values, $C_{\mathrm{GS_1}} = \frac{1}{2}C_{\mathrm{ox}}WL + C_{\text{gs-overlap}}$ in the linear region and $C_{\mathrm{GS_2}} = \frac{2}{3}C_{\mathrm{ox}}WL + C_{\text{gs-overlap}}$ in saturation, where C_{ox} is the gate capacitance per unit area [14] and $C_{\text{gs-overlap}}$ is the gate-to-source overlap capacitance, two values will be used for $i_{C_{\mathrm{GS_p}}}$ according to the time point t_{sp} (Figure 6).

Energy starts being dissipated at time $t_{\mathrm{s\_sc}}$ when current $i_{\mathrm{s_p}}$ starts flowing towards the source of the pMOS transistor so that a current path between V_{DD} and ground exists. Time $t_{\mathrm{s\_sc}}$ can be calculated by setting $i_{\mathrm{s_p}} = 0$ in Equation (11) and using the linear approximation for the pMOS current, as shown in Figure 6. Thus

$$t_{\mathrm{s\_sc}} = t_{\mathrm{s}} + \frac{t_{\mathrm{sp}} - t_{\mathrm{s}}}{i_{\mathrm{p_{max}}}} i_{C_{\mathrm{GS_1}}}$$

The pMOS transistor starts its operation in linear mode and then enters saturation at time point t_{sp}, where i_{p} and consequently $i_{\mathrm{s_p}}$ reach their maximum value. Energy dissipation ceases at time point $t_{\mathrm{e\_sc}}$ when $i_{\mathrm{s_p}} = 0$, after time t_{sp}. $t_{\mathrm{e\_sc}}$ is again calculated using the linear approximation for the pMOS current resulting in

$$t_{\mathrm{e\_sc}} = t_{\mathrm{e}} - \frac{t_{\mathrm{e}} - t_{\mathrm{sp}}}{i_{\mathrm{p_{max}}}} i_{C_{\mathrm{GS_2}}}$$

Consequently, the short-circuit energy dissipation can be easily obtained as

$$\begin{aligned} E_{\mathrm{sc}}^{\mathrm{d}} &= V_{\mathrm{DD}}\left(\int_{t_{\mathrm{s\_sc}}}^{t_{\mathrm{sp}}} i_{\mathrm{s_p}}\,\mathrm{d}t + \int_{t_{\mathrm{sp}}}^{t_{\mathrm{e\_sc}}} i_{\mathrm{s_p}}\,\mathrm{d}t\right) \\ &= \frac{1}{2}V_{\mathrm{DD}}[(I_{\mathrm{p_{max}}} - i_{C_{\mathrm{GS_1}}})(t_{\mathrm{sp}} - t_{t_{\mathrm{s\_sc}}}) + (I_{\mathrm{p_{max}}} - i_{C_{\mathrm{GS_2}}})(t_{\mathrm{e\_sc}} - t_{\mathrm{sp}})] \end{aligned} \tag{12}$$

Table II. Accuracy and speed comparison between calculation and HSPICE (HP 0.35 μm).

τ (ns)	C_L (fF)	Prop. delay SPICE (ns)	Prop. delay calc. (ns)	Error (%)	Exec. time SPICE (s)	Exec. time calc. (s)	Speedup
0.2	50	0.098	0.094	4.08	0.80	0.00154	519
0.2	100	0.151	0.146	3.31	0.69	0.00143	483
0.5	50	0.128	0.126	1.56	0.74	0.00163	454
0.5	100	0.196	0.191	2.55	0.66	0.00149	443
1	50	0.154	0.164	6.49	0.72	0.00179	402
1	100	0.241	0.246	2.07	0.71	0.00178	399
1.5	50	0.165	0.185	12.12	0.77	0.00193	399
1.5	100	0.271	0.287	5.90	0.78	0.00181	431
2	100	0.292	0.321	9.93	0.80	0.00198	404
2	200	0.471	0.480	1.91	0.75	0.00181	414

The short-circuit energy dissipation during output charging, $E_{\rm SC}^{\rm c}$, can be obtained in a symmetrical way.

Therefore, the short-circuit energy dissipation during a complete transition at the output node [0 → 1 → 0] is $E_{\rm sc}=E_{\rm sc}^{\rm c}+E_{\rm sc}^{\rm d}$ and the corresponding power can be calculated simply by multiplying the calculated energy with the frequency of transitions at the output of the gate.

4. RESULTS

The presented technique has been implemented in a *C* program in order to verify its performance in terms of speed and accuracy. Since the output voltage expression is known for each region of operation, the time point when the output waveform crosses the half-$V_{\rm DD}$ point is obtained by solving $V_{\rm out}=V_{\rm DD}/2$. Propagation delay is calculated as the time from the half-$V_{\rm DD}$ point of the input to the half-$V_{\rm DD}$ point of the output. The slope of the output waveform at this point is obtained by calculating the derivative at $V_{\rm DD}/2$ of the corresponding output voltage expression in Appendix A. Accuracy comparisons with SPICE for a 0.5 μm HP technology and the corresponding execution times for the analysis of an inverter are shown in Table I for several input transition times and load capacitances ($W_{\rm n}=4$ μm, $W_{\rm p}=6$ μm, $V_{\rm DD}=5$ V). In Table II comparisons with HSPICE for a 0.35 μm HP technology are also given ($W_{\rm n}=2.8$ μm, $W_{\rm p}=4.2$ μm, $V_{\rm DD}=3.3$ V). The time step (0.01 ns) and the integration time interval (5 ns) is the same both for SPICE and the proposed method. It is observed that a significant speedup can be gained with only a minor penalty in accuracy. Parameters from the foundry-supplied model card are given in Table III. A second program has been developed in order to calculate the output waveform of a CMOS inverter based on a fully analytical solution [3]. In Table IV the accuracy and required execution time for the fully analytical method is compared to that of SPICE. It can be observed that the proposed method is on average 5 times faster than the fully analytical method while the average cost penalty (the difference between the average error of the proposed method and that of Reference [3]) is only 0.7%.

To display the applicability of the proposed method for real cases, the percentage error for propagation delay between the proposed method and SPICE is shown in Figure 9 for (a)

Table III. SPICE model parameters for two HP submicron technologies.

			0.5 μm (LEVEL = 3)		0.35 μm (LEVEL = 49)	
Name	Comment	Units	nMOS	pMOS	nMOS	pMOS
U0	Mobility	cm^2/(V s)	546.2	135.5	392.1	126.7
VMAX	Maximum drift velocity	m/s	2.008×10^5	2.542×10^5	—	—
VSAT	Saturation velocity of carrier	m/s	—	—	1.232×10^5	1.400×10^5
TOX	Gate oxide thickness	m	9.6×10^{-9}	9.6×10^{-9}	7.6×10^{-9}	7.6×10^{-9}
VTO	Zero bias threshold voltage	V	0.6566	−0.9213	0.5710	−0.6337

Table IV. Accuracy and speed comparison between SPICE and the method proposed in Reference [3] (0.5 μm).

τ (ns)	C_L (fF)	Prop. delay (ns)	Error (%)	Exec. time (s)	Speedup
0.5	50	0.072	5.26	0.00968	24
0.5	100	0.126	2.33	0.00963	17
1	50	0.062	1.64	0.00985	19
1	100	0.141	2.17	0.00963	17
1.5	50	0.047	9.30	0.01035	20
1.5	100	0.135	6.30	0.00980	19
2	100	0.117	3.54	0.01002	15
2	200	0.271	3.04	0.00990	27
3	100	0.073	5.80	0.01068	16
3	200	0.253	3.69	0.01007	16

several inverter sizes, (b) transistor ratios, (c) capacitive loads and (d) input transition times, respectively. As it can be observed, the proposed model agrees within 10% with SPICE for most cases enabling the simulation of any possible inverter configuration.

In order to compare the accuracy of the proposed approach to the technique presented in Reference [7] where the short-circuit current is treated as an additional charge, accuracy and speed results for Reference [7] are given in Table V. The large errors that can be observed, especially in the cases with considerable short-circuit current, are mainly due to the fact that the short-circuit current which determines the additional charge is calculated on an initial estimate of the output waveform which is obtained without taking into account the effect of the short-circuit current. A small error can be observed only in cases with a large load capacitance, where however almost no short-circuit current is present, as it can be observed in the bottom two rows of Table V. Similar errors for Reference [7] have also been mentioned in References [3,5]. Execution times are similar to that of the proposed approach, since both methods employ a common core where the initial output waveform is obtained and a small overhead to calculate the additional charge provided by the short-circuit current.

A widely accepted dynamic timing simulator is ILLIADS2 [17,18] which employs region-wise quadratic modelling (RWQ) for capturing of submicron MOS current models. Execution times for the simulation of an inverter by SPICE, ILLIADS2, the fully analytical method

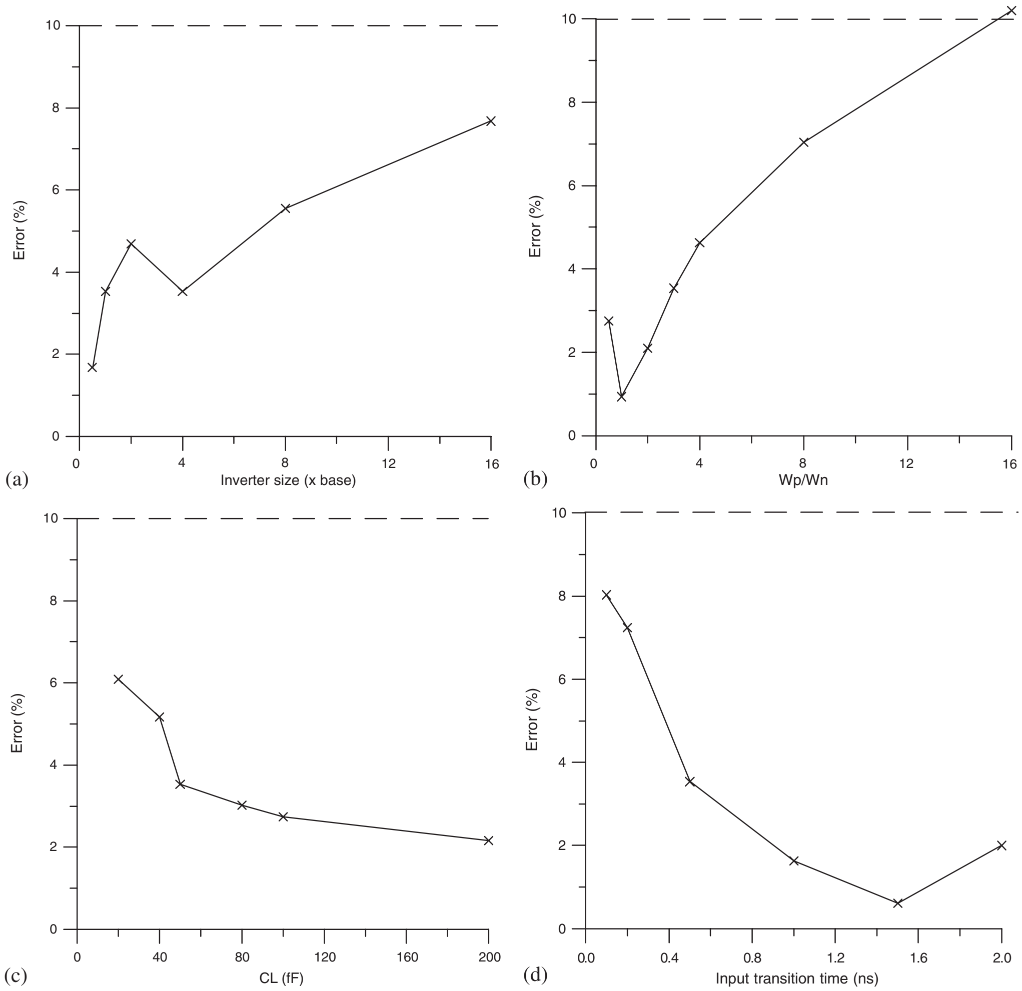

Figure 9. Accuracy of the proposed method in terms of percentage error (propagation delay) between HSPICE and the proposed method for (a) several inverter sizes (base $W_n = 1.4$ μm, $W_p = 2.1$ μm, $C_L = 50$ fF, $\tau = 0.5$ ns), (b) transistor ratios ($W_n = 1.4$ μm, $C_L = 50$ fF, $\tau = 0.5$ ns), (c) load capacitances ($W_n = 1.4$ μm, $W_p = 2.1$ μm, $\tau = 0.5$ ns) and (d) input transition times ($W_n = 1.4$ μm, $W_p = 2.1$ μm, $C_L = 50$ fF), (HP 0.35 μm technology).

[3] and the proposed method are given in Figure 10 for several input transition times/output loads. It is obvious that the proposed method is much faster than SPICE and significantly more efficient than ILLIADS2 and the fully analytical method.

It is important to mention that for timing purposes it is sufficient to know the time when the output waveform crosses $V_{DD}/2$ and its slope at this point, so that this waveform can be approximated by a ramp in order to feed it to the next stage [1]. The proposed method, although not very accurate at the beginning and the tail of the output waveform (as it can be observed in Figure 4), is very accurate around the 50% point and consequently is appropriate for use in timing simulators. Moreover, when only the propagation delay and the slope at

Table V. Accuracy and speed comparison between SPICE and the method proposed in Reference [7] (0.5 μm).

τ (ns)	C_L (fF)	Prop. delay (ns)	Error (%)	Exec. time (s)	Speedup
0.5	50	0.103	35.53	0.00162	142
0.5	100	0.138	6.97	0.00153	105
1	50	0.097	59.02	0.00180	106
1	100	0.221	60.14	0.00169	95
2	100	0.292	158.40	0.00203	74
2	200	0.448	70.34	0.00191	141
1	500	0.531	5.36	0.00158	101
1	1000	0.869	0.11	0.00150	133

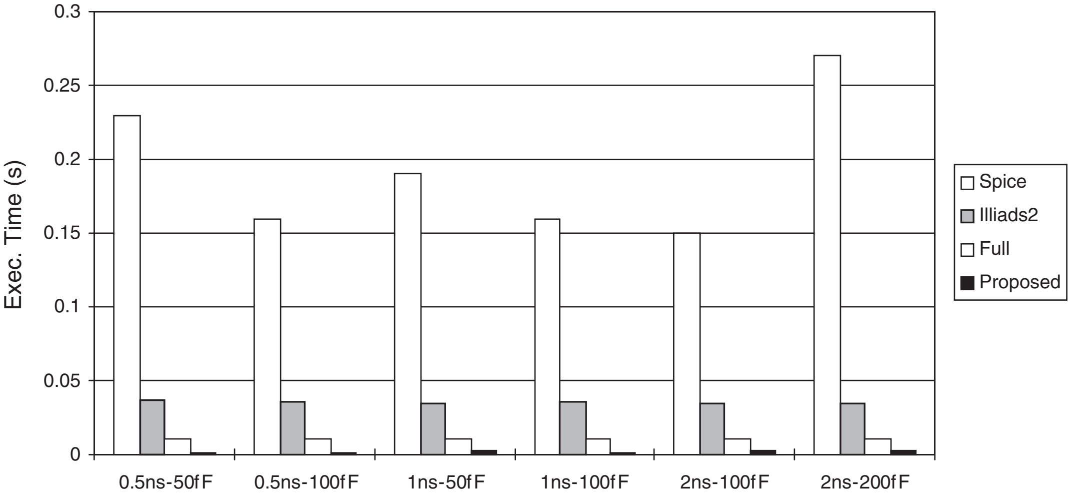

Figure 10. Execution times of several simulation tools (TRAN 0.01 ns, 5 ns–HP 0.5 μm).

$V_{DD}/2$ are calculated, a speedup of more than 2×10^3 compared to SPICE is achieved. However, if voltage levels other than the 50% point are required (e.g. 20–80% or 10–90%), they can be extracted from the voltage expressions which are known for the complete output waveform (Appendix A).

The proposed method can also be applied to complex CMOS gates since any gate can be reduced to an equivalent inverter [11,12]. Propagation delay results have been calculated for four complex CMOS gates and compared to SPICE simulation results in Table VI. A second way of applying the proposed method to complex gates is to obtain the output waveform considering only the conducting path and then to calculate the additional charge provided by the short-circuiting path [16], taking into account the shift of the output waveform.

The calculated energy for a single output transition lies very close to the energy which is measured from SPICE simulations. In Figure 11 a comparison of the calculated and simulated short-circuit energy values during output discharging is shown for an inverter and for several input transition times/output loads ($W_n = 4$ μm, $W_p = 6$ μm, $V_{DD} = 5$ V). The total energy

Table VI. Accuracy and speed comparison between calculation and SPICE (HP 0.5 μm) for complex gates.*

Gate	Prop.delay SPICE (ns)	Prop.delay calc. (ns)	Error (%)	Exec.time SPICE (s)	Exec.time calc. (s)	Speedup
NAND3	0.3224	0.3235	0.34	0.28	0.00238	118
NAND4	0.3387	0.3395	0.24	0.36	0.00268	134
MUX2 × 1	0.2025	0.2055	1.48	0.42	0.00320	131
AOI121	0.4501	0.4518	0.38	0.51	0.00299	171

*Input transition time is 0.5 ns in all cases, $C_L = 100$ fF; NAND3: $W_n = 3$ μm, $W_p = 1.5$ μm, all three inputs switching (rising); NAND4: $W_n = 4$ μm, $W_p = 1.5$ μm, all four inputs switching (rising); MUX2 × 1: $W_n = 4$ μm, $W_p = 1.5$ μm, *SELECT* falling, I_0 DC 0V, I_1 rising; AOI121: W_n(in series) = 1.33 μm, W_p(parallel) = 3 μm, W_n(two other) = 0.66 μm, W_p(two other) = 6 μm, all four inputs switching (falling).

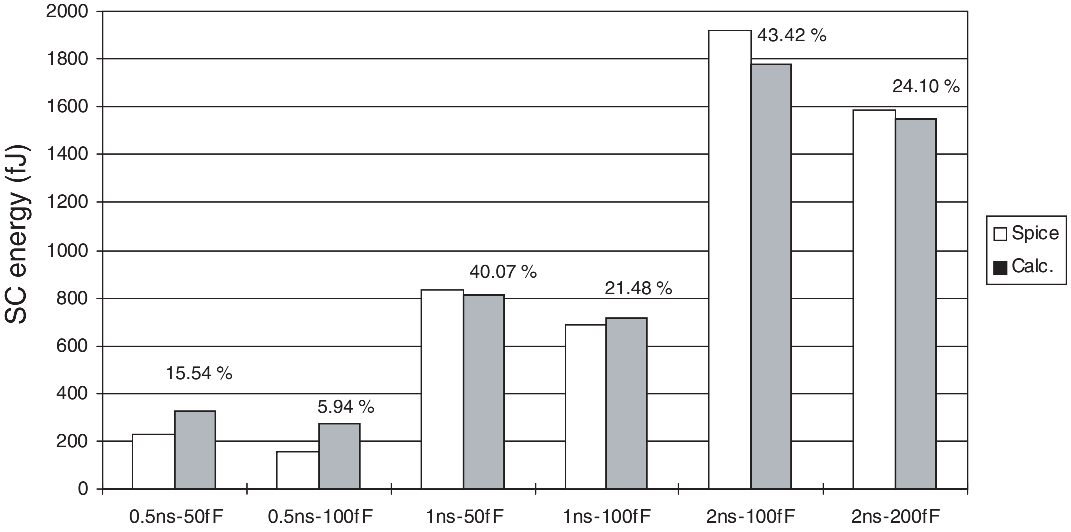

Figure 11. Simulated and calculated short-circuit energy dissipation (HP 0.5 μm). The percentage of the short-circuit power over the total power is also given.

dissipation for a CMOS inverter is the sum of the dynamic energy that is dissipated due to charging and discharging of the load capacitance and the energy that is dissipated due to the direct-path current from supply to ground. In Figure 11 the percentage of the short-circuit power over the total dissipated power is also given in order to indicate the fact that the short-circuit power can be a large portion of the total power and thus it has to be calculated accurately.

5. CONCLUSION

A method for improving the efficiency of the simulation of the CMOS inverter by avoiding to take into account the short-circuit current in the solution of the circuit differential equation has been introduced. The effect of the short-circuit current is considered as an additional charge which causes a shift to the output waveform. This additional charge as well as the corresponding discharging current of the conducting transistor are expressed as functions of the time shift of the output waveform. The calculated output waveform and the short-circuit power dissipation match SPICE simulation results with very good accuracy while a significant speedup compared to SPICE execution time has been achieved. The proposed method was also found to be significantly faster than the fully analytical method and the dynamic timing simulator ILLIADS2. Consequently, such a method is appropriate for integration in existing dynamic timing and power simulators in order to speed up the analysis of large integrated circuits.

APPENDIX A

The output waveform expressions for the circuit of Figure 1(b) according to the regions of operation of the nMOS transistor (Figure 2(b)) are given below:

A.1. Region 1 $(0 < t \leqslant t_{\mathrm{n}})$

The nMOS transistor before the input ramp reaches the nMOS transistor threshold voltage at time $t_{\mathrm{n}} = V_{\mathrm{TN}} \cdot \tau / V_{\mathrm{DD}}$ is cut-off and the output voltage presents a small overshoot due to the input-to-output coupling capacitance:

$$V_{\mathrm{out_{init}}}(t) = V_{\mathrm{DD}} + \frac{C_M s}{C_L + C_M}\, t \tag{A1}$$

where s is the slope of the input ramp ($s = V_{\mathrm{DD}}/\tau$).

A.2. Region 2A $(t_n < t \leqslant t_{sn_0})$

The nMOS transistor operates in saturation and the input is in transition:

$$V_{\mathrm{out_{init}}}(t) = k_1 t - k_2 (st - V_{\mathrm{TN}})^{a_{\mathrm{n}}+1} + C[1] \tag{A2}$$

where

$$k_1 = \frac{C_M}{C_L + C_M} s, \quad k_2 = \frac{k_{\mathrm{s_n}}}{(C_L + C_M) s (1 + a_{\mathrm{n}})}$$

and $C[1]$ is the integration constant. $t_{\mathrm{sn_0}}$ is the time point when the nMOS transistor enters the linear region when its drain-to-source voltage becomes equal to the drain saturation voltage [4] and it can be found by solving

$$V_{\mathrm{out_{init}}}(t) = V_{\mathrm{DSAT}_n}(t) = \frac{k_{\mathrm{s_n}}}{k_{\mathrm{l_n}}} (st - V_{\mathrm{TN}})^{a_{\mathrm{n}}/2} \tag{A3}$$

In case the nMOS transistor enters the linear region before the input ramp reaches its final value the input should be considered slow and in that case the next region is Region 3A. Otherwise it should be considered fast and the next region is Region 2B.

A.3. Region 3A $(t_{\mathrm{sn}_0} < t \leqslant \tau)$

The nMOS transistor operates in the linear region while the input is still in transition. Neglecting the influence of the input signal on the current through the coupling capacitance [3] the solution of the circuit differential equation becomes:

$$V_{\mathrm{out_{init}}}(t) = C[2]\mathrm{e}^{-k_3(st - V_{\mathrm{TN}})^{1+a_{\mathrm{n}}/2}} \tag{A4}$$

where

$$k_3 = \frac{2k_{ln}}{(C_L + C_M)s(2 + a_{\mathrm{n}})}$$

and $C[2]$ is the integration constant.

A.4. Region 3B $(t > \tau)$

The nMOS transistor operates in the linear region and the input has reached its final value:

$$V_{\mathrm{out_{init}}}(t) = C[3]\,\mathrm{e}^{-k_4 t} \tag{A5}$$

where

$$k_4 = \frac{k_{\mathrm{ln}}(V_{\mathrm{DD}} - V_{\mathrm{TN}})^{a_n/2}}{(C_L + C_M)}$$

and $C[3]$ is the integration constant.

A.5. Region 2B $(\tau < t \leqslant t_{\mathrm{sn}_0})$

In case the nMOS transistor enters the linear region after the input ramp reaches its final value, region 2A will be extended up to time point τ and will be followed by region 2B where the nMOS transistor is still in saturation and the input is V_{DD}. The solution in this case is

$$V_{\mathrm{out_{init}}}(t) = C[4] - k_5 t \tag{A6}$$

where

$$k_5 = \frac{k_{\mathrm{s_n}}(V_{\mathrm{DD}} - V_{\mathrm{TN}})^{a_{\mathrm{n}}}}{(C_L + C_M)}$$

and $C[4]$ is the integration constant.

In this case Region 2B is followed by Region 3B as shown in Figure 2.

APPENDIX B

The estimation of the starting point of the pMOS short-circuit current, which is of primary importance to the calculation of the additional charge that causes the time shift of the initial output waveform will be discussed next.

The coupling capacitance between the input and the output node, C_M, causes an overshoot on the output voltage (Figure 7). During the overshoot the pMOS current is flowing towards

V_{DD} as already mentioned. To perform an accurate modelling of the output evolution in this region the subthreshold current of the nMOS device should be taken into account.

The subthreshold current is given by [19]

$$I_{\mathrm{sub}} = I_{\mathrm{on}}\,\mathrm{e}^{(V_{\mathrm{GS}} - V_{\mathrm{TN}})(q/nkT)} \tag{B1}$$

where I_{on} is the current in strong inversion for $V_{\mathrm{GS}} = V_{\mathrm{on}}$ and the voltage V_{on} is found as $V_{\mathrm{on}} = V_{\mathrm{TN}} + nkT/q$ where $n = 1 + qN_{\mathrm{FS}}/C_{\mathrm{ox}} + C_{\mathrm{d}}/C_{\mathrm{ox}}$. The SPICE parameter N_{FS} is used as a fitting parameter that determines the slope of the subthreshold current–voltage characteristics [19]. C_{d} is the capacitance associated with the depletion region, q denotes the unit (electron) charge, k is the Boltzmann constant, and T is the temperature.

For the purpose of this analysis it is sufficient to approximate the subthreshold current by its first-order Taylor series approximation around $V_{\mathrm{GS}} = 0.75\,V_{\mathrm{TN}} \Rightarrow t = 0.75\,t_{\mathrm{n}}$:

$$I_{\mathrm{sub}} = I_{\mathrm{sub}}|_{t=0.75t_{\mathrm{n}}} + I'_{\mathrm{sub}}|_{t=0.75t_{\mathrm{n}}}(t - 0.75\,t_{\mathrm{n}}) = d_1 + d_2\,t \tag{B2}$$

The time when the subthreshold current starts ($I_{\mathrm{sub}} = 0$) will be referred to as t_{d}. Since the pMOS transistor operates initially in the linear region it can be modelled by a resistor with value

$$R_{\mathrm{p}} = \frac{1}{k_{l_{\mathrm{p}}}(V_{\mathrm{DD}} - V_{\mathrm{TP}})^{a_{\mathrm{p}}/2}}$$

When the input starts rising and until time point t_{d}, the differential equation at the output node of the inverter is

$$C_M\left(\frac{\mathrm{d}V_{\mathrm{in}}}{\mathrm{d}t} - \frac{\mathrm{d}V_{\mathrm{out}}}{\mathrm{d}t}\right) + \frac{V_{\mathrm{DD}} - V_{\mathrm{out}}}{R_{\mathrm{p}}} = C_L\,\frac{\mathrm{d}V_{\mathrm{out}}}{\mathrm{d}t} \tag{B3}$$

which has the solution:

$$V_{\mathrm{out}}(t) = V_{\mathrm{DD}} + C_M R_{\mathrm{p}} s(1 - \mathrm{e}^{-t/R_{\mathrm{p}}(C_L + C_M)}) \tag{B4}$$

After time point t_{d}, the subthreshold current should also be taken into account and the solution becomes

$$V_{\mathrm{out}}(t) = V_{\mathrm{DD}} + k_1 - R_{\mathrm{p}} d_2 t + k_2 \mathrm{e}^{(t_{\mathrm{d}} - t)/R_{\mathrm{p}}(C_L + C_M)} \tag{B5}$$

where $k_1 = R_{\mathrm{p}}^2 d_2(C_L + C_M) - d_1 R_{\mathrm{p}} + C_M R_{\mathrm{p}} s$ and $k_2 = V_{\mathrm{d}} - V_{\mathrm{DD}} - k_1 + R_{\mathrm{p}} d_2 t_{\mathrm{d}}$, and $V_{\mathrm{d}} = V_{\mathrm{out}}[t_{\mathrm{d}}]$ according to Equation (B4).

Solving $V_{\mathrm{out}} = V_{\mathrm{DD}}$ for Equation (B5) gives the exact time point when the overshoot ceases, t_{ov}, which in turn is the time point when the pMOS current starts flowing towards the output node, t_{s}.

In case the nMOS transistor starts conducting before the end of the overshoot ($t_{\mathrm{n}} < t_{\mathrm{ov}}$) the differential equation after time point t_{n}, should be solved by taking into account the expression of the nMOS current in saturation given by Equation (2). Since the resulting equation cannot be solved analytically the nMOS current is approximated by a first-order Taylor series around $t = 2t_{\mathrm{n}}$ resulting in an expression similar to (B5).

A comparison of the calculated and simulated output waveform for a CMOS inverter in this region is shown in the enlarged area of Figure 7 ($W_{\mathrm{n}} = 4\,\mu\mathrm{m}, W_{\mathrm{p}} = 6\,\mu\mathrm{m}, C_L = 100\,\mathrm{fF}, \tau = 2\,\mathrm{ns}$).

It should be mentioned that the pMOS current during the overshoot contributes to the discharging of the output load [3,6] and therefore the charge that it 'removes' from the output node should be deducted from Q_{ad} in the proposed analysis. However, this charge is negligible and does not affect the overall accuracy of the method.

REFERENCES

1. Hedenstierna N, Jeppson KO. CMOS circuit speed and buffer optimization. *IEEE Transactions on Computer-Aided Design* 1987; **CAD-6**(2):270–281.
2. Hirata A, Onodera H, Tamaru K. Estimation of short-circuit power dissipation for static CMOS gates. *IEICE Transactions on Fundamentals* 1996; **E79-A**(3):304–311.
3. Bisdounis L, Nikolaidis S, Koufopavlou O. Analytical transient response and propagation delay evaluation of the CMOS inverter for short-channel devices. *IEEE Journal of Solid-State Circuits* 1998; **33**(2):302–306.
4. Sakurai T, Newton AR. Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas. *IEEE Journal of Solid-State Circuits* 1990; **25**(2):584–594.
5. Bisdounis L, Nikolaidis S, Koufopavlou O. Propagation delay and short-circuit power dissipation modeling of the CMOS inverter. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications* 1998; **45**(3):259–270.
6. Jeppson KO. Modeling the influence of the transistor gain ratio and the input-to-output coupling capacitance on the CMOS inverter delay. *IEEE Journal of Solid-State Circuits* 1994; **29**(6):646–654.
7. Embabi SHK, Damodaran R. Delay models for CMOS, BiCMOS and BiNMOS circuits and their applications for timing simulations. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 1994; **13**(9):1132–1142.
8. Shih Y-H, Kang SM. Analytic transient solution of general MOS circuit primitives. *IEEE Transactions on Computer-Aided Design* 1992; **11**(6):719–731.
9. Daga JM, Auvergne D. A comprehensive delay macro modeling for submicrometer CMOS logics. *IEEE Journal of Solid-State Circuits* 1999; **34**(1):42–55.
10. Turgis S, Auvergne D. A novel macromodel for power estimation in CMOS structures. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 1998; **17**(11):1090–1098.
11. Chatzigeorgiou A, Nikolaidis S, Tsoukalas I. A modeling technique for CMOS gates. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 1999; **18**(5):557–575.
12. Kong J-T, Hussain SZ, Overhauser D. Performance estimation of complex MOS gates. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications* 1997; **44**(9):785–795.
13. Hamoui AA, Rumin NC. An analytical model for current, delay, and power analysis of submicron CMOS logic circuits. *IEEE Transactions on Circuits and Systems—II: Analog and Digital Signal Processing* 2000; **47**(10):999–1007.
14. Meyer JE. MOS models and circuit simulation. *RCA Review* 1971; **32**:42–63.
15. Hirata A, Onodera H, Tamaru K. Estimation of propagation delay considering short-circuit current for static CMOS gates. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications* 1998; **45**(11):1194–1198.
16. Nikolaidis S, Chatzigeorgiou A. Analytical estimation of propagation delay and short-circuit power dissipation in CMOS gates. *International Journal of Circuit Theory and Applications* 1999; **27**(4):375–392.
17. Shih Y-H, Leblebici Y, Kang SM. ILLIADS: a fast timing and reliability simulator for digital MOS circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 1993; **12**(9):1387–1402.
18. Dharchoudhury A, Kang SM, Kim KH, Lee SH. Fast and accurate timing simulation with regionwise quadratic models of MOS I–V characteristics. *Proceedings of the IEEE International Conference on Computer-Aided Design* (ICCAD), November 1994; 190–194.
19. Kang SM, Leblebici Y. *CMOS Digital Integrated Circuits: Analysis and Design*. McGraw Hill: New York, 1996.