A Compact Model to Identify Delay Faults due to Crosstalk

Jose L Rossello and Jaume Segura
Electronic Technology Group, Universitat de les Illes Balears, Palma de Mallorca, Spain.
emails:{j.rossello;jaume.segura}@uib.es

Abstract- In this work we present an analytical formulation to estimate quickly and accurately the impact of crosstalk induced delay in submicron CMOS ICs gates taking into account time skews. Crosstalk delay is computed from the additional charge injected from the aggressor gate on the victim gate during simultaneous switching. The model provides a very good agreement with HSPICE simulations for a 0.18μm technology.

I. INTRODUCTION

The constant scaling of CMOS technology feature sizes and supply voltage together with the increase in both operating frequency and signal rise/fall times has made digital designs more vulnerable to noise. In modern ICs, interconnect coupling noise (crosstalk) becomes a performance limiting factor that must be analyzed carefully during the design process. If not considered, crosstalk can cause extra delay, logic hazards and even logic malfunction. Fig.1 illustrates the well known crosstalk induced delay effect that appears when two lines (the aggressor and the victim) switch simultaneously. For a victim node falling transition, the gate delay (defined as \( t_{PHL,LO} \)) can be reduced (\( t_{PHL,ss} \)) or increased (\( t_{PHL,sd} \)) depending on if the aggressor makes a falling or a rising transition respectively. Since large circuits can handle tens of millions of interconnect lines on a single chip, simple and accurate analytical descriptions for crosstalk delay are of high interest to develop a fast timing verification at the chip level.

Crosstalk delay has been extensively analyzed during last years. In [1] Gross et. al. used a waveform iteration strategy to analyze crosstalk without considering the time skew between transitions; moreover no-closed form expression for the worst-case victim delay was provided. More recently crosstalk delay has been modeled analytically in [2] providing an excellent understanding of the main effects involved during the transition and deriving analytical expressions that quantify the severity of crosstalk delay. The dependence on circuit parameters like the rise/fall times of transitions and the skew between transitions is provided. The nMOS (pMOS) network is described through a pull-down (pull-up) resistance, while short-circuit currents are neglected. The main limitation of the work is the impossibility of deriving a closed-form expression for the propagation delay given the complex equations obtained for the victim voltage variation. Additionally, other effects like short-circuit currents (that have an important impact on propagation delay [3]) are not considered in their analysis. In [4] we developed an analytical modeling of crosstalk delay by considering the worst-case situation in which the aggressor and victim transitions are perfectly aligned. Although useful to compute the maximum delay due to crosstalk, the worst-case situation cannot be considered the general case in real ICs. A more realistic description requires the consideration of the time skew between the victim and aggressor transitions, and the consequent impact on the delay increase/decrease. In this work we propose a simple and accurate model to compute the crosstalk delay (\( t_{PHL,ss} \) and \( t_{PHL,sd} \)) in submicron CMOS gates that accounts for this effects. We modify a simple propagation delay model for CMOS inverters developed previously [3] to include more complex gates than inverters by using the collapsing technique developed in [5]. The model is compared to HSPICE simulations for a 0.18μm technology showing an excellent agreement.

II. CROSSTALK DELAY MODEL

Consider the circuit in Fig.1 where the aggressor and the victim gates drive output capacitances \( C_{La} \) and \( C_{Lb} \) respectively, and there is a coupling capacitance \( C_{c} \) between \( V_{out,a} \) and \( V_{out,v} \) (that causes the crosstalk between the two lines). The time skew \( t_{s} \) is defined as the time interval from \( V_{DD}/2 \) at the victim input to \( V_{DD}/2 \) at the aggressor input. For simplicity the output interconnect lines are modeled trough single capacitances as this has been shown to be a valid approach for medium size interconnects whose relative error is below 10% for a 2nm wire in a 0.18μm technology [6].
Figure 1 – An additional delay in the victim is induced by the switching of an aggressor gate.

For the circuit in Fig.1 we first present a compact charge-based propagation delay model based on the propagation delay model developed in [3], and then include crosstalk effects by describing the impact on delay of the additional charge to be transferred through the corresponding nMOS/pMOS block.

A. Propagation delay model

The propagation delay is usually computed as the time interval between the input and the output crossing $V_{DD}/2$. The crosstalk-free propagation delay ($t_{pHL0}$) can be described as a function of the charge transferred through the gate using the model in [3] based on the combination of the $n$th-power law MOSFET model [7], and an efficient and accurate transistor collapsing technique for complex gates developed in [5].

For a high to low output transition (for a low to high transition the analysis is equivalent) the propagation delay is given by:

$$t_{pHL0} = \begin{cases} 
  t_n + \left[\frac{Q_f(1+n)}{I_{D0n}} (t_{in} - t_n)\right]^\frac{1}{1+n} & \text{if } Q_f < Q_{f0} \\
  t_n + \frac{Q_f - Q_{f0}}{I_{D0n}} & \text{if } Q_f \geq Q_{f0} 
\end{cases}$$

where $t_{in}$ is the input transition time (i.e. the time during which the input is changing), parameter $n$ is the velocity saturation index of the nMOS transistors, $Q_f = C_t V_{DD}/2$ is the charge transferred through the nMOS block from the beginning of the transition until $V_{out} = V_{DD}/2$ (where $V_{DD}$ is the supply voltage), $C_t$ is the output load of the gate, $t_n = (V_{TN}/V_{DD}) t_{in}$ is the time when the nMOS block starts to conduct ($V_{TN}$ is the threshold voltage of nMOS), and $Q_{f0}$ is the charge transferred through the nMOS block when the input transition is finished, given by:

$$Q_{f0} = \frac{I_{D0n}(t_{in} - t_n)}{1 + n} \tag{2}$$

where parameter $I_{D0n}$ is the maximum current that the nMOS block can deliver (modeled as a chain of $N_n$ series-connected transistors) and is obtained from the transistor collapsing technique developed in [5]. For the simple case of a chain with $N_n$ identical transistors, the collapsing technique provides a simple expression for $I_{D0n}$:

$$I_{D0n} = \frac{3V_{DD} n (1 + \gamma_n)}{5(V_{DD} - V_{TN})} \tag{3}$$

where $V_{DD}$ is the saturation voltage of nMOS, $I_{D0n}$ is the maximum saturation current of each nMOS (drain current when $V_{GS} = V_{DS} = V_{DD}$), and the parameter $\gamma_n$ accounts for channel length modulation. The parameter $K_n$ is a technology-dependent parameter given by:

$$K_n = \frac{3V_{DD} n (1 + \gamma_n)}{5(V_{DD} - V_{TN})}$$

where $\gamma_n$ is the body effect parameter of nMOS [7].

The transition time at the output $t_{out0}$, required to evaluate the crosstalk delay, can also be given as a function of $Q_f$ as [3]:

$$t_{out0} = \begin{cases} 
  \frac{2Q_f}{I_{D0n}} \left[\frac{Q_f(1+n)}{I_{D0n}} (t_{in} - t_n)\right]^\frac{1}{1+n} & \text{if } Q_f < Q_{f0} \\
  \frac{2Q_f}{I_{D0n}} + \frac{Q_f - Q_{f0}}{I_{D0n}} & \text{if } Q_f \geq Q_{f0} 
\end{cases}$$

Equations (4) and (1) are valid for a high to low output transition. Equivalent expressions can be obtained for a low to high output transition.

B. Including short-circuit currents

The delay model in (1) is valid for CMOS gates when short-circuit currents are neglected. These currents are included in [3] as an additional charge to be transferred through the pull-down (pull-up) network for an output falling (rising) transition. For a high to low output transition, the charge transferred through the nMOS block is computed as $Q_f = C_t V_{DD}/2 + q_{sc}^f$, where $q_{sc}^f$ is defined as the short-circuit charge transferred during the falling output transition until $V_{out} = V_{DD}/2$. For simplicity we compute this charge as $q_{sc}^f = Q_{sc}^f/2$, where $Q_{sc}^f$ is the total short-circuit charge transferred. For the evaluation of $Q_{sc}^f$ we use a previously developed model described in [8].

$$Q_{sc}^f = \frac{2I_{D0n}(t_{in} - t_n)}{1 + n}$$

where $I_{D0n}$ is the maximum current that the nMOS block can deliver (modeled as a chain of $N_n$ series-connected transistors) and is obtained from the transistor collapsing technique developed in [5]. For the simple case of a chain with $N_n$ identical transistors, the collapsing technique provides a simple expression for $I_{D0n}$:
C. Crosstalk delay

The impact of crosstalk on the propagation delay is computed considering the additional charge injected from the aggressor driver to the victim output node through the coupling capacitance $C_c$. This additional contribution must be discharged by the victim pull-down network thus requiring an additional time equal to the crosstalk-induced delay. We consider the case of a falling transition at $V_{out,v}$ slower than the rising transition at $V_{out,a}$ (for other cases the analysis is similar).

Fig. 2 shows a high to low transition of the victim node $V_{out,v}$ when the aggressor output switches from low to high for six different skew times represented as a, b, ..., f. For case ‘a’, the aggressor transition ($A_g_a$) does not impact the victim propagation delay ($t_{pHL0}$) since the absolute value of the time skew between transitions is too large. The victim output (transition $V_{out,v}(a)$) rises until $V_{out,a} = V_{DD}$ and then falls back to $V_{DD}$ in a time defined as $t_D$ (see Fig. 3). This characteristic time ($t_D$) is computed assuming that the ON transistors of the victim gate pMOS block (that discharge the output) are in the linear region and that their drain-source voltage is small. Under these conditions, the current through the pMOS block is:

$$I_p = I_{D0P} \frac{V_{DD} - V_{out}}{V_{D0P}}$$

where $V_{D0P}$ is the saturation voltage of pMOS, and $I_{D0P}$ is computed from all the ON transistor chains that connect the supply and the output node. Each of these pMOS chains is collapsed to a single equivalent transistor with maximum saturation current $I_{D0P}^{(1,N_i)}$ (where $N_i$ is the number of transistors in each chain). Finally $I_{D0P}$ is expressed as the sum of these contributions as:

$$I_{D0P} = \sum_{i=1}^{m} I_{D0P}^{(1,N_i)}$$

where $m$ is the number of active chains connecting $V_{DD}$ and $V_{out,v}$. After the time point at which $V_{out,a} = V_{DD}$, the victim line output voltage evolution ($V_{out,v}$) is described as:

$$V_{out,v} = V_{DD} + (V_{max} - V_{DD}) e^{-\frac{(t-t_0)}{t_D}}$$

where $t_0$ is the time at which $V_{out,a} = V_{DD}$ and $V_{max}$ is the maximum voltage at $V_{out,v}$. The characteristic time $t_D$ (relaxation of $V_{out}$ back to $V_{DD}$) is obtained from (7) leading to:

$$t_D = 2ln\left(\frac{V_{DD}C_L}{I_{D0P}}\right)$$

It is well known that crosstalk may have an impact on delay only when the aggressor and victim transitions occur within a time window. Otherwise the effect of crosstalk is simply a glitch at the victim line. The time interval during which the coincidence of input transitions may give a crosstalk delay can be defined in terms of a limit time skew value ($t_{s1}$). This limit value will depend on the propagation delay of the aggressor driver, the characteristic time $t_D$, and the aggressor output transition time. Fig. 3 illustrates this time relationships graphically. The limit time $t_{s1}$ is obtained equating the time point at which the glitch at $V_{out,v}$ is finished to the beginning of the high to low transition at $V_{out,v}$, (at time 0):

$$t_{s1} + 2t_{pLH0,a} + t_D = 0$$

The parameter $t_{s1}$ can be easily extracted from (9). If $t_s < t_{s1}$ the effect of crosstalk is a glitch at the victim node and the voltage value is restored by victim

Figure 2 – Crosstalk delay is dependent on the time between transitions $V_{in,a}$ and $V_{in,v}$.

Figure 3 – The aggressor starts to affect the victim voltage variation when $t_s = t_{s1}$.

$$I_{D0P} = \sum_{i=1}^{m} I_{D0P}^{(1,N_i)}$$

$$V_{out,v} = V_{DD} + (V_{max} - V_{DD}) e^{-\frac{(t-t_0)}{t_D}}$$

$$t_D = 2ln\left(\frac{V_{DD}C_L}{I_{D0P}}\right)$$

$$t_{s1} + 2t_{pLH0,a} + t_D = 0$$
pMOS devices before the victim transition. When \( t_s > t_{s1} \), the crosstalk effect impacts the victim transition delay since an additional charge (defined as \( Q_c \)) must be drained by the nMOS devices during the transition. When \( V_{out,a} \) and \( V_{out,v} \) start switching at the same time (at \( t = 0 \)), the value of \( Q_c \) is maximum and consequently the delay is also maximum. Defining \( t_{s3} \) as the time skew at this time point we have that \( t_{s3} \simeq 0 \).

For \( t_s = t_{s2} \), the coupling charge can be taken as \( Q_c = C_v V_{DD} \) since the voltage variation at \( V_{out,a} \) occurs during the transition at \( V_{out,v} \). For the case where \( t_{s3} < t_s < t_{s2} \) we use a linear variation of \( Q_c \) with \( t_s \):

\[
Q_c = \frac{t_s - t_{s1}}{t_{s2} - t_{s1}} C_v V_{DD} \tag{10}
\]

The coupling charge \( Q_c \) is equal to \( C_v V_{DD} \) when the time skew is larger than \( t_{s2} \). This condition holds if the transition of \( V_{out,a} \) finishes before \( V_{out,v} \) reaches \( V_{DD}/2 \). This limit time skew value is defined as \( t_{s3} \) and is obtained solving:

\[
t_{s3} + 2 t_{pHL_{0,a}} = t_{pHL_{0,v}} \tag{11}
\]

If \( t_{s2} < t_s < t_{s3} \) (case d in Fig 2) \( Q_c \) reaches its maximum value (\( Q_c = C_v V_{DD} \)). When \( t_s > t_{s3} \) \( Q_c \) starts to decrease until \( Q_c = 0 \) at \( t_s = t_{s4} \), where \( t_{s4} \) is defined as the time at which the beginning of the transition at \( V_{out,a} \) and the time at which \( V_{out,v} = V_{DD}/2 \) are equal.

\[
t_{s4} = t_{pHL_{0,v}} \tag{12}
\]

In the interval \( t_{s3} < t_s < t_{s4} \) we use a linear variation of \( Q_c \) with \( t_s \) given by:

\[
Q_c = C_v V_{DD} \frac{t_s - t_{s4}}{t_{s3} - t_{s4}} \tag{13}
\]

Finally, for \( t_s > t_{s4} \) (transitions e and f in Fig 2) we use \( Q_c = 0 \). For the case where \( V_{out,a} \) makes a high to low transition, speeding up the propagation delay, a similar model can be obtained changing the sign of \( Q_c \). Once \( Q_c \) is obtained, the propagation delay (\( t_{pHL_{su}} \) and \( t_{pHL_{sd}} \)) at \( V_{out,v} \) are obtained using \( Q_f = C_i V_{DD}/2 + Q_{le}/2 + Q_c \) in (1).

### III. RESULTS

We compare the model results to HSPICE simulations for a 0.18\( \mu \)m technology. In Fig.4 we plot the propagation delay of a 3-NAND gate subjected to crosstalk (both speed-up and slow-down) from a buffer for different times of the skew values of the two gates between the two gate inputs \( V_{in,v} \) and \( V_{in,a} \), comparing the model results (solid lines) to HSPICE simulations (data points). Two channel width values of the nMOS transistors of the 3-NAND gate are selected (with the pMOS sized appropriately) and both the speed-up and slow-down cases are shown. A negative time skew value indicates that the aggressor transition is initiated before the victim transition. The Figure clearly shows that the maximum crosstalk induced delay corresponds to zero skew as expected since both the aggressor and the victim switch simultaneously. For negative values of the time skew the impact of crosstalk delay decreases progressively until it goes gradually to zero, while for positive values of this time the crosstalk-induced delay remains maximum until a certain time-point at which it decreases more rapidly. Both the relative increase of the crosstalk-delay with respect to the nominal delay, and the time window during which coupling impacts delay decrease when increasing the strength of the victim driver. Therefore two parameters need to be taken into account when characterizing crosstalk delay: the victim sensitivity to
crosstalk noise and the time window during which the victim can be impacted. Both parameters are design dependent and can be minimized by proper design. We define the crosstalk time window as the time skew values during which the victim delay is affected by the aggressor transition (this is the time difference $T_{CTK} = t_{s4} - t_{s1}$).

In Figs 5 and 6 we show both the crosstalk-induced propagation delay $\Delta t_{pHL} = t_{pHL} - t_{pHLO}$ and the crosstalk time window $T_{CTK}$ of the 3-NAND gate when varying different design parameters (victim and aggressor channel widths and coupling capacitance). $W_{nv}$ is the width of the nMOS victim, while $W_{na}$ is that of the aggressor. The channel widths are related to $W_{\min}$ that is defined as the minimum channel width allowed by the technology. Parameters $\Delta t_{pHL}$ and $T_{CTK}$ are related in percentage to the crosstalk-free propagation delay $t_{pHLO}$, and to the maximum circuit operating frequency respectively (in our case a clock period of $T = 1\text{ns}$). As can be appreciated, a coupling capacitance increment leads to an appreciable increment in both, propagation delay and the crosstalk time window. This increment in the crosstalk effects is also appreciated when decreasing the victim channel widths (see Fig. 5). When varying other design parameters as the aggressor width (see Fig. 6) we appreciate that the propagation delay is unaffected but that the crosstalk time window decreases considerably when increasing $W_{na}$.

Figs 5 and 6 demonstrate the validity of the proposed model to explore the impact of crosstalk depending on both design- and process-related parameters, and its utility in the development of advanced design and test tools to evaluate and detect the impact of crosstalk on delay.

IV. CONCLUSIONS AND FUTURE WORK

A simple description for the evaluation of crosstalk delay has been presented. The model is useful for a fast and accurate signal integrity simulation of large ICs. A very good agreement is achieved between model predictions and HSPICE simulations for a 0.18$\mu$m technology. The description accounts for short-circuit currents and short-channel effects and can be applied to multiple input gates. Experimental measurements for a more faithful validation of the proposed model are under development.

ACKNOWLEDGMENT

This work has been partially supported by the Spanish Ministry of Science and Technology, the Regional European Development Funds (FEDER) from EU project TEC2005-05712/MIC, and an Intel Laboratories-CRL research grant.

REFERENCES


