Delay Model for Static CMOS Gates Considering Single Input Multiple Transistor Switching

Felipe S. Marranghello, André I. Reis, Renato P. Ribas

PGMICRO, UFRGS, Av. Bento Gonalve 9500, Porto Alegre, Brazil

{fsmarranghello,andreis,rpribas}@inf.ufrgs.br

***Abstract*— This paper presents a delay for static CMOS gates presenting two transistors controlled by the same input in each network. Previous work have assumed that each input controls only one NMOS and one PMOS transistor and are therefore not applicable for gates presenting several transistors controlled by the same input. The proposed model requires only transistor parameters and does not need additional fitting parameters. Results are in very good agreement with electrical simulations based on BSIM4 model. The average error is near 3% with a worst case error smaller than 10%.**

***Keywords*— Gate delay, analytical delay model, multiple input switching, timing analysis**

I. INTRODUCTION

The standard cell methodology is often used in VLSI designs. One of the issues related to this methodology is the restrict set of cells available that have a major influence on circuit performance, power and area For this reason, the virtual library methodology appears as an alternative to overcome the limitations imposed by the traditional standard cell approach [1-2]. When a virtual library is used, cells can be generated on the fly. However, since the number of possible cells is huge, performing electrical simulation to characterize all cells becomes impracticable.

Analytical delay models have been widely discussed as a fast alternative to electrical simulation [3-15]. These models have been used to perform timing analysis and circuit optimization. Such models are also applicable to the virtual library methodology. Some delay models focus on the CMOS inverter since it is the simplest cell [3-6]. These models form the basis for the remaining models.

Deriving an accurate delay model for static CMOS gates is a hard task due to the many effects that impact the gate behavior. An accurate model must consider the influence of input transition time, parasitic capacitances, short circuit current, body effect, channel length modulation and drain induced barrier lowering (DIBL). Therefore, a pure mathematical approach is unfeasible for delay models aiming general CMOS gates.

An additional challenge for gate delay modeling is the possibility of multiple input switching [7]. Models that handle several inputs switching with different start and transition times require fitting parameters. Usually, a weight for each input is given based on the number of stacked transistors, on the input position on the stack and on which inputs are switching. This kind of models suffers from the fact that all possible stacks must be characterized.

As an alternative to obtain simpler models, several proposals assume that only one PMOS and one NMOS transistor on the gate switch [8-15]. In this work, this situation is named single input single transistor switching (SISTS). This approach is justified because timing analysis is usually performed assuming that only one input switches. Due to the smaller number of conditions that must be considered, these models are usually simpler than those that consider independent inputs switching. Nevertheless, in several proposals, fitting parameters are required to model slow inputs [11,14,15] or some effects are neglected [10,12,13].

Even under the assumption of single input switching the models in [8-15] are not always applicable because it is possible that one input controls more than one transistor on each plane. In this work, this situation is named single input multiple transistor switching (SIMTS). An example of such gate is shown in Fig 1a. The implementation shown in Fig 1a uses a bridge connection to diminish the number of transistors required. For comparison, a series-parallel implementation is shown in Fig 1b. Due to the possible reduction on the number of transistors, several works have studied the generation of gates using non-series parallel arrangements [16-17]. It must be noticed that the models in [8-15] can be used with SIMTS when only one switching transistor is important to the output switching process. As example, the gate shown in Fig 1b can be analysed considering SISTS even if the input controls more than one transistor.

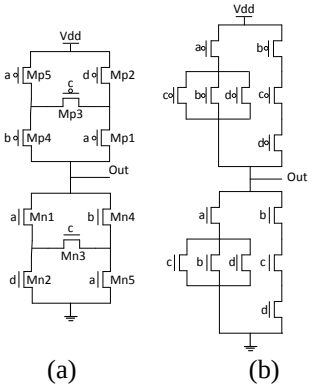

Figure 1: *Logically equivalent gates (a) non-series parallel gate and (b) series-parallel gate*

In this work, a novel delay model to consider SIMTS is proposed. Furthermore, several improvements to SISTS models are proposed. The model considers all important effects and the only empirical parameters used are those required to calibrate the transistor model. Still, the utilization of empirical parameters for the transistor model is a common characteristic to most delay models [3-15].

The rest of the paper is organized as follows. Section II presents a SIST delay model. Section III extends the model to the SIMTS scenario. Section IV presents the conclusions.

II. SIST MODEL OVERVIEW

This section discusses the main concepts in deriving a delay model for SIST gates. Most of the ideas presented herein are similar to those in previous work [3-15]. This discussion is important as the SIST delay model is extended to the SIMTS delay model in Section III. The α-power transistor model is used [3]. The equations for the NMOS drain current (*Ids*) device are:

$$Ids = \begin{cases} 0 & \text{cutoff} \\ K \ln Wn(Vgs - Vtn)^{\alpha/2} Vds & \text{linear} \\ KsnWn(Vgs - Vtn)^{\alpha}(1 + \lambda Vds) & \text{saturation} \end{cases} \quad (1)$$

The parameter α is the velocity saturation index, λ is the well-known channel length modulation parameter, *Wn* is the effective transistor width, *Kln*, *Ksn* and *Pvn* are empirical parameters. *Vds* is the drain-to-source voltage, *Vgs* is the gate-to-source voltage, *Vsat* is the saturation voltage. Similar parameters are defined for the PMOS transistor.

The transistor threshold voltage (*Vtn*) depends on *Vds* due to the DIBL effect, and on the source-to-bulk voltage (*Vsb*) due to the body effect. These effects should not be neglected in delay modeling. In this work, the threshold voltage considering DIBL and body effect is written as:

$$Vtn = Vt0 + \delta Vsb - \eta Vds \quad (2)$$

where δ is the body effect coefficient and η is the DIBL coefficient. Notice that a linear approximation for the body effect is used.

The *Vdd*/2 delay (*Td*) is defined as the time interval between the moment when the input and the output reach half of the supply voltage value:

$$Td = Tout50 - (Tin/2) \quad (3)$$

where *Tout50* is the time when the output signal reaches *Vdd*/2 and *Tin* is the input transition time. Since *Tin* is known, the model accuracy lies on the estimation of *Tout50*. Hereafter, a rising input (i.e falling output) is considered. The analysis for falling input is symmetrical.

Existing SIST delay models mostly rely on a charge based approach. In this approach, the model must estimate the total charge that must flow through the discharging path and a discharge current [9-14]. The total charge depends on drain and source capacitances, the output load, I/O coupling capacitances and short circuit current. These are influenced by the input transition time. The discharge current is influenced by the number of transistors on the discharge path and also by the input transition time.

A. Charge estimation

The total charge (*Qtot*) has several components. The charge stored in the output node (*Qout*), charge stored in internal nodes of the discharging network (i.e NMOS network) (*Qint*), charge stored in internal nodes of the parasitic network (i.e PMOS network) (*Qpar*) and due to short circuit (*Qsc*). *Qtot* is the sum of all components:

$$Qtot = Qout + Q\text{int} + Qpar + Qsc + Cm*Vdd \quad (4)$$

Where *Cm*Vdd* represents the extra charge due to I/O coupling capacitance.

Qout is obtained from the output load and from the diffusion capacitances that are attached to the output node. This component is the only that does not depend neither on the input transition nor on the input switching transition. If *Cout* represents the total capacitance attached to the output node, then *Qout* is given by:

$$Qout = \frac{Vdd}{2} * Cout \quad (5)$$

Qint is only important when the switching transistor is not the top one. In this case, the capacitances between the source of the top transistor and drain of the switching transistor are partially charged. To estimate *Qint*, the initial (*Vinit*) and final voltages values at Tout50 (*Vend*) of the internal nodes of the NMOS network must be estimated. *Vinit* is written as:

$$Vinit = Vdd - Vdrop \quad (6)$$

Usually, *Vdrop* is taken to be equal to threshold voltage of the top transistor [10-14]. This approximation is, however, inaccurate. A better value is obtained assuming that the same current flows in all stacked transistors.

As evidence for the claim that *Vdrop* cannot be the top transistor threshold voltage, the circuit in Fig. 2 is used. The top NMOS transistor has a high gate voltage and *N* bottom transistors are placed in parallel. Table 1 shows *Vdrop* for different values of *N* using a 65 nm technology along with the top transistor threshold voltage.

Despite the dependence of circuit topology on *Vdrop*, the absolute change is relatively small and one single value can be used as a technology constant. To determine the value, equate the subthreshold current of both top and bottom transistor and solve for the source voltage (*Vinit*):

$$Vinit = \frac{Vdd(1+\eta) + nUt * \ln(Wtop/Wbot)}{1 + \delta + 2\eta} \quad (7)$$

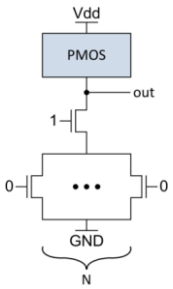

Figure 2: Test circuit to determine *Vdrop*

TABLE 1: *VDROP* COMPARISON TO THRESHOLD VOLTAGE

N	*Vdrop*	*Vthn_top*
1	0.252	0.571
2	0.268	0.566
3	0.278	0.563
4	0.285	0.561

Where *n* is the subthreshold slope, *Ut* is the thermal voltage, *Wtop* and *Wbot* are the effective widths of the top and bottom transistors, respectively.

To determine the final value for the internal voltages, the maximum stack current when the output voltage is *Vdd/2* is calculated. To determine the maximum stack current, set the gate voltages of all transistors in the discharge path to *Vdd*, and then replace all transistors below the top by one equivalent transistor using the simple conductance association [8-14]. Once the current is known, the voltages are directly determined.

Qpar estimation is often omitted [8,14,11]. In this work, the following approach is used. The extra charge due to internal capacitances only considers the nodes that are between the output node and the drain node of the switching transistor. These nodes are initially charged to *Vdd* and discharge to a minimum value of *Vdrop*. If the output transition to *Vdd/2* is considered, the voltage variation is smaller and a good approximation is *(Vdd-|Vdrop|)/2* for the source node of the transistors connected to the output, although the actual value depends on the input transition time. The final voltage of the other nodes is estimated considering a voltage divider. Each internal capacitance (*Ci*) is added to the output node as an equivalent capacitance (*Ceqi*) given by:

$$Ceqi = \frac{Ci*(Vdd - Vi)}{Vdd} \tag{8}$$

where the term *Vi* represents the node voltage when the output crosses Vdd/2. The short circuit component *Qsc* is calculated as in [6].

B. Current estimation

To estimate the discharge current, consider the switching transistor as the top transistor of a chain. The idea is to calculate the maximum stack current when the drain voltage of the switching transistor at *t=0* and *t=tout50*. For that, set all gate voltages to *Vdd* and equate the current of the switching transistor to the current of the equivalent bottom transistor. The effective maximum current (*Isist*) is the average of the stack current with both drain voltages.

For a fast input, *Tin* < *Tout50*, *Tout50* is:

$$Tout50 = \frac{Qtot}{Isist} + \frac{Tin*(\alpha + \frac{Vtn}{Vdd})}{\alpha + 1} \tag{9}$$

For a slow input *Tin* > *Tout50*, *Tout50* is:

$$Tout50 = \left(\frac{(\alpha+1)*Qtot*Tin^{\alpha}*(Vdd - Vtn)}{Isist*Vdd^{\alpha}}\right)^{1/(\alpha+1)} + Tin*Vtn/Vdd \tag{10}$$

III. SIMTS EXTENSION

The main influence of SIMTS is to modify the discharge current due to a different behaviour of internal voltages. Fig 3 compares the source voltage of the top transistor for the gate in Fig 1a in the SISTS and SIMTS scenarios. The SISTS scenario is obtained when *Mn4* switches and the discharge path is {*Mn4*,*Mn3*, *Mn2*}, the SIMTS scenario occurs when transistors *Mn1* and *Mn5* switch and the path is {*Mn1*,*Mn3*, *Mn5*}. Notice that a fast (slow) output corresponds to a slow (fast) input. It is observable that the initial source voltage in the SIMTS scenario is higher because there is no low resistance path between this node and ground. Additionally, the different source voltage behaviour is more prominent when the output is fast. Therefore, it can be expected that the impact of SIMTS on delay to vary with *Tin* and to be maximum for slow inputs. In order to estimate the voltage initial value (*Vsinit*), a similar approach to the used in Section II.A is used. The maximum source (*Vsmax*) is obtained summing *Vsinit* to the maximum source voltage with SIST. *Imin* is the stack current when the source voltage of the switching transistor is *Vsmax*. The effective stack current with SIMT is given by:

$$Isimt = Isist + (\mathrm{Im}\,in - Isist)*\sqrt{Tin/Tref} \tag{11}$$

Where *Tref* is:

$$Tref = \frac{(\alpha+1)*Qout*Vdd}{\mathrm{Im}\,in*(1 - Vt/Vdd)} \tag{12}$$

To calculate *Tout50* use *Isimt* instead of *Isist* in (10) and (9). The model is validated using an industrial 65 nm technology. Table 2 presents the average (*AVG*) and worst case (*WC*) relative errors for different output loads. For each output load, *Tin* varies from 1 to 500 ps with a time step of 1 ps. The load value is normalized to the NMOS gate capacitance. The worst case error of 8.7% is similar to those reported errors for SISTS models [12-16]. Table 3 summarizes the average and worst case errors for previous SIST works. The data are as provided by the authors (a NA means that the information is not explicitly indicated). The average error of the proposed is close to 3% which is a similar result to [11-15] that present

average error in the range 3%-5%. The observed worst case error for the new model is 8.70% which is also close to the worst case error for previous SISTS works that range from 7.5% to 10%.

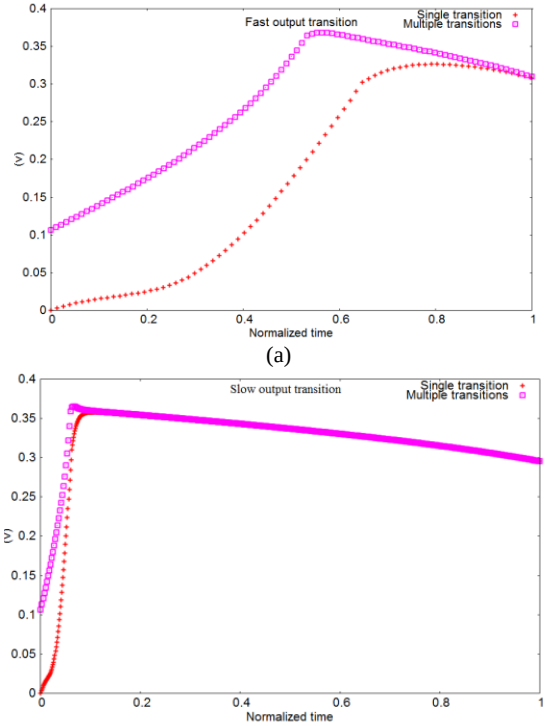

Figure 3: Comparison of SIST and SIMT top transistor source voltage (a) slow input and (b) fast input

TABLE 2: AVERAGE (AVG), WORST CASE (WC) ERRORS FOR DIFFERENT OUTPUT LOADS, CONSIDERING A 65 NM CMOS PROCESS PARAMETERS.

Load	*AVG* (%)	*WC* (%)
0.25	2.56	8.70
1	2.52	8.24
4	3.34	7.61
16	4.23	7.50
64	3.04	7.59
256	1.85	2.46

TABLE 3:PREVIOUS WORK AVERAGE (AVG) AND WORST CASE (WC) ERRORS

Work	*AVG* (%)	*WC* (%)
[15]	< 5	NA
[14]	4.5	7.5
[13]	3	9
[12]	3	NA
[11]	NA	<10

IV. CONCLUSIONS

An analytical delay model considering single input – multiple transistors switching is presented. Previous models are unable to consider this situation. The model proposed herein relies solely on transistor parameters. Comparisons of calculated results to electrical simulations based on BSIM4 model validate the proposal. The average error is near 3% with worst-case error less than 10% which is a similar result to previous work with the main addition that more gates can be analysed.

ACKNOWLEDGMENTS

Research funded by CAPES, CNPq and FAPERGS Brazilian funding agencies, and by the European Community's Seventh Framework Programme under grant 248538-Synaptic.

REFERENCES

[1] Pullerits, M.; Kabbani, A.; "Area minimization for library-free synthesis," in *Proc. of Circuits and Systems and TAISA Conference, NEWCAS-TAISA*, 2009, pp.1-4.

[2] F. S. Marques, L. S. Rosa, Jr., R. P. Ribas, S. S. Sapatnekar, and A. I. Reis. 2007. DAG based library-free technology mapping. In *Proceedings of the 17th ACM Great Lakes symposium on VLSI* (GLSVLSI '07). ACM, New York, NY, USA, 293-298.

[3] Sakurai, T.; Newton, A.R.; "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J.l of Solid-State Circuits*, , vol. 25, no. 2, Apr. 1990. pp.584-594,..

[4] Rossello, J. L.; Segura, J.; "An analytical charge-based compact delay model for submicrometer CMOS inverters," *IEEE Trans. on Circuits and Systems I: Regular Papers*, vol. 51, no. 7, July 2004. pp.1301-1311.

[5] Zhangcai H.; *et al.;* "Modeling the overshooting effect for CMOS inverter delay analysis in nanometer technologies," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 29, no. 2, Feb. 2010. pp.250-260.

[6] Marranghello, F.S.; Reis, A.I.; Ribas, R.P., "Design-oriented delay model for CMOS inverter,"Symp. *Integrated Circuits and Systems Design (SBCCI)*, vol., no., pp.1,6, Aug. 30 2012-Sept. 2 2012

[7] Chatzigeorgiou, A.; Nikolaidis, S.; Tsoukalas, I.; "A modeling technique for CMOS gates," *IEEE Trans.. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, no. 5, May 1999. pp.557-575.

[8] Elmore, W. C.; "The transient analysis of damped linear networks with particular regard to wideband amplifiers," *J. Applied Physics*, vol. 19, no.1, 1948.

[9] Sutherland, I.; Sproull, B.; Harries, D.; *Logical effort: Design Fast CMOS Circuits*, 1999, Morgan Kaufmann

[10] Sakurai, T.; Newton, A. R.; "Delay analysis of series-connected MOSFET circuits," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 2, Feb. 1991. pp.122-131.

[11] Daga, J. M.; Auvergne, D.; "A comprehensive delay macro modeling for submicrometer CMOS logics ," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 1, Jan. 1999. pp. 42-55.

[12] Rossello, J. L.; de Benito, C.; Segura, J.; "A compact gate-level energy and delay model of dynamic CMOS gates," *IEEE Trans. on Circuits and Systems II: Express Briefs*, vol. 52, no. 10, Oct. 2005. pp.685- 689..

[13] Lasbouygues, B.; *et al*.; "Logical effort model extension to propagation delay representation," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol.25, no.9, Sep. 2006. pp.1677-1684.

[14] Kabbani, A.; Al-Khalili, D.; Al-Khalili, A. J.; "Delay analysis of CMOS gates using modified logical effort model," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 6, June 2005. pp.937- 947.

[15] Wang, C. C.; Markovic, D.; "Delay estimation and sizing of CMOS ;ogic using logical effort with slope correction," *IEEE Trans. on Circuits and Systems II:Express Briefs*, vol. 56, no. 8, Aug. 2009. pp.634-638.

[16] V. Callegaro, F. S. Marques, C. E. Klock, L. S. da Rosa, Jr., R. P. Ribas, and A.; I. Reis. "SwitchCraft: a framework for transistor network design". In *Proceedings of the 23rd symposium on Integrated circuits and system design* (SBCCI '10)

[17] D. Kagaris,and T. Haniotakis, "A Methodology for Transistor-Efficient Supergate Design,"I*EEE Trans on VLSI* vol.15, no.4, pp.488,492, 2007