Accurate and Efficient Macromodel of Submicron Digital Standard Cells

Cristiano Forzan, Bruno Franzini and Carlo Guardiani

SGS-THOMSON Microelectronics,

via C. Olivetti, 2,

20041 Agrate Brianza (MI), ITALY

Abstract - In this paper a new analytic gate delay modeling technique is presented that allows to accurately reproduce the timing behavior of deep submicron digital standard cells for a large range of operating conditions. The proposed technique sensibly improves the accuracy of the existing analytic delay models and it usually requires less simulations for the cell characterization. Moreover it is compatible with the most advanced interconnect delay models that have been recently proposed in the literature.

I - INTRODUCTION

In order to analyze the timing behavior of modern CMOS circuits, the proper gate and interconnect delay models must be derived that allow to obtain both efficiency and accuracy at the same time. The gate model should be simple enough to reduce the computational time and the memory occupation and it should be as accurate as possible in order to enable the timing verification of multi-million gate, deep submicron logic circuits. Therefore, in general, the proper tradeoff between speed and accuracy must be found.

Moreover, the gate delay model must be consistent with the algorithm used to compute the interconnect delay, e. g. AWE [1], [2], [3] and PVL [4].

Usually, circuit delays are expressed as functions of the input signal transition time (T_{IN}) and of the load capacitance (C_L), often in the form of look-up table models or analytical expressions, the so called k-factor equations, in which the delay is expressed by means of a polynomial function of (T_{IN}, C_L) [5], [6]. The limitation of a purely capacitive load on the output has been addressed and solved [7], to account for generic RC trees, using AWE based simulation and reduction. More recently, the limitation due to the assumption of a perfectly linear input ramp has been overrun by introducing a piecewise linear input model [8].

In this paper, we present an analytical approach, based on the modeling of the output current waveform, similar to that used to derive the "*three region model*" described in [9]. The proposed model allows to take advantage of these recent results by capturing the actual wave shape of the output transitions.

In order to improve the accuracy of the model a new region has been added. Moreover, a methodology based on the use of Design of Experiments has been succesfully introduced in order to optimally sample the space of the V, T, T_{IN}, C_L operating conditions. By doing so the caracterization effort required to achieve a given accuracy level has been minimized. The delay model thus obtained is able to represent accurately the gate behavior in the specified range of operating conditions, for a large class of cells. Moreover, the proposed model requires a minimal characterization effort, as demonstrated by the experimental results presented in this paper, and it is consistent with an AWE model of the interconnect delay.

II - MOTIVATIONS

State of the art CAD tools for cell-based design analysis and synthesis require to describe the timing behavior of the gates using a look-up table model representing the input to output transition delay (tpd) and the output transition time (tt) as a function of T_{IN} and C_L, and additionally a global linear derating equation to account for different circuit supply voltage (V) and operating temperature (T) conditions. In pratical cases, the error due to the derating function approximation may represent a significant source of inaccuracy, thus leading to overlay pessimistic delay estimations and to undue buffer oversizing. Therefore a basic motivation for the work was the need of generating accurate look-up table models at a user specified voltage supply and temperature. Obviously this can be done by running a large number of SPICE simulations to fill in the tables, but this "brute force" approach is unacceptably too much computationally intensive. As an alternative we propose a methodology based on the intermediate creation of a suitable delay macromodel that can be eventually evaluated to generate the look-up table model at any user specified voltage and temperature.

III - SCOPE OF THE WORK

The library cells can be classified as single-stage or multiple-stage. Unbuffered pass gates are out of the scope of this work. A single-stage cell, or better a single-stage delay path, is defined as an I/O delay path going across a single stack of channel connected CMOS devices (i.e. INV, NOR, NAND). In general multiple-stage cells (or delay paths) can be represented (figure 1) by a combinatorial network followed by the single-stage cell suitable to drive the output load with the desired signal dynamics.
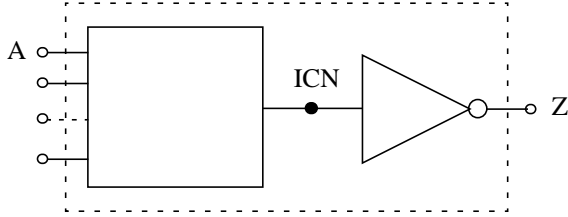

Fig. 1: Multiple-stage cell can be regarded as a logical stage (A-ICN) followed by a buffer stage (ICN-Z).

A lot of effort has been done in the past to accurately describe the timing behavior of the inverter, and to reduce a generic single-stage network to an equivalent inverter [10]. The approach presented in this paper allows to characterize the delay macromodel of any single-stage cell starting from an extracted SPICE netlist (including parasitics) by running electrical simulations without any preliminary reduction to an equivalent inverter. The detailed single-stage macromodel description is presented in Sections IV and V.

For multiple-stage cells, an additional effort is needed to characterize the propagation delay from input pin (A) to the internal controlling node (ICN):

$$T_0(ICN) = f_1(T_{IN}((A), V, T))$$ (1)

and the transition time at the ICN:

$$T_{CIN}(ICN) = f_2(T_{IN}((A), V, T))$$ (2)

These values depend on the input transition time on pin (A), on the voltage supply and on the operating temperature. The propagation delay from input pin (A) to output pin (Z) can be computed as the composition of the delay of the logical stage and that of the buffer stage.

IV - THE THREE REGION MODEL

In this section we will introduce the *three region* model, that is the basis of the work presented in this paper. The model will be described for the basic inverter, which is quite difficult to obtain because of the direct relationship between input and output voltage. Supposing a purely capacitive load, C_L, the output voltage V_{out} is related to the output current I_{out} by the following differential equation:

$$\frac{d}{dt}V_{out}(t) = -\frac{I_{out}(t)}{C_L}$$ (3)

An analytical expression for the delay can be obtained by choosing a suitable approximation for the output current. The derivation of the *three region* model proposed by Sakurai [9] can be easily understood from the waveform shown in figure 2, showing the SPICE simulation of a generic inverter for a rising input transition.
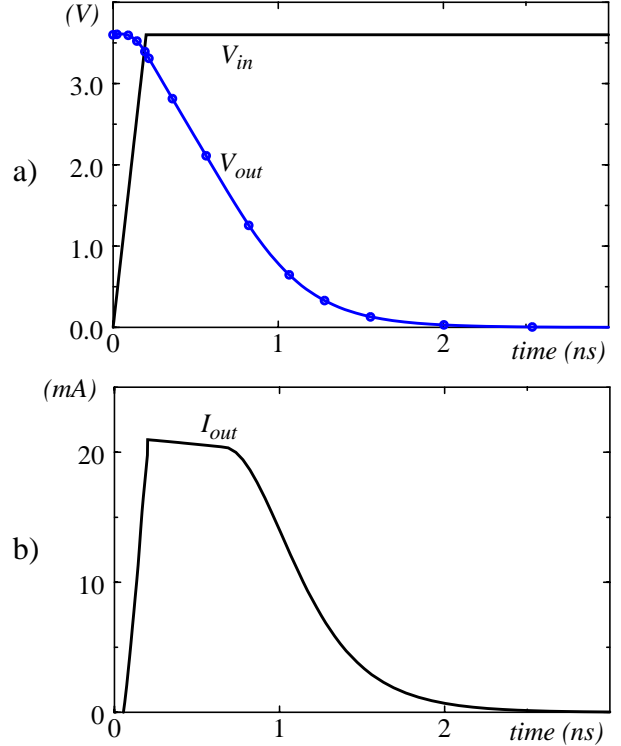

Fig. 2: SPICE simulation of a switching event for a generic inverter; a) input and output voltages; b) corresponding output current waveform.

By looking to the output current waveform, it can be observed that, initially, the current increases as the input voltage increases (first region). Then, the output current reaches a maximum value and then approximatively it remains constant while the input waveform has already completed the transition (second region). Finally, the current begins to decay roughly with the same rate of change as the output voltage and hence its behaviour is almost exponential (third region). These considerations suggest the following expression for the current:

$$I_{out}(t) = min\left(\frac{V_{in}(t) - V_T}{R_M}, \frac{V_{out}(t)}{R_F}\right)$$ (4)

where V_T, R_M, R_F are the fitting parameters of the model and V_{in} is the V_{DD} normalized input ramp signal, swinging

from 0 to 1 in a time T_{IN}; hence $V_{in}(t)$ can be written as follows:

$$V_{in}(t) = min\left(\frac{t}{T_{IN}}, 1\right) \quad (5)$$

By substituting equation (4) in (3) and solving for V_{out}, the expression for the output voltage as a function of time for the *three region* model can be found. In the first region, approximatively corresponding to the region where the switching device is in the saturation regime, V_{out} is a quadratic function, whereas in the second region it is a linear function of time. In the third region, which corresponds to the case when the switching device is in the linear regime, V_{out} can be described by a decaying exponential.

A similar expressions can be derived for falling input transitions.

V - THE FOUR REGION MODEL

When the input transition time T_{IN} is sufficiently large and the load capacitance C_L is sufficiently small, the *three region* model is no longer valid.
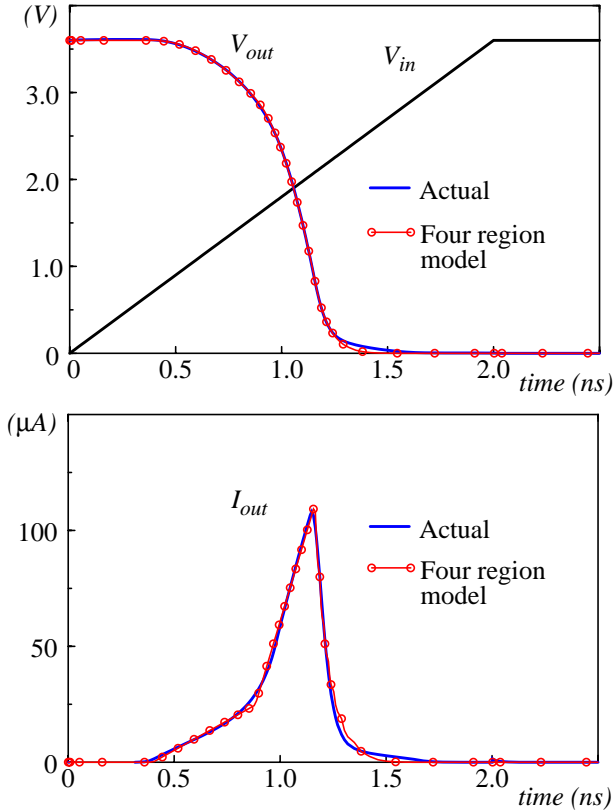
In fact, because of the effect of the short-circuit current, the output current presents an anomalous behavior in the linear region. This leads to a loss of accuracy, because the parameters R_M and V_T used to model the gate behavior in this region, cannot be adjusted to fit the actual current waveform. In order to solve this problem we introduced a new model region requiring two extra fitting parameters. The current in this region is fitted by using two line segments (pwl). Therefore the modified analytical expression for the output current is the following:

$$I_{out}(t) = min$$
$$\left(max\left(\frac{V_{in}(t) - V_{T0}}{R_{M0}}, \frac{V_{in}(t) - V_{T1}}{R_{M1}}\right), \frac{V_{out}(t)}{R_F}\right) \quad (6)$$

The model in (6) is able to accurately reproduce the current waveshape in every operating conditions, as shown in figure 3.

A further improvement is obtained by introducing a parameter, I_{MAX}, accounting for the current limiting effect that is due to the finite driving capability of the gate transistors. The value of this parameter is obtained during a pre-characterization phase as a function of the power supply and temperature. The accuracy improvement obtained with the introduction of I_{MAX} is shown in figure 4.

Fig. 3: Comparison between SPICE simulation and the four region model for large input transition time and small capacitive load.
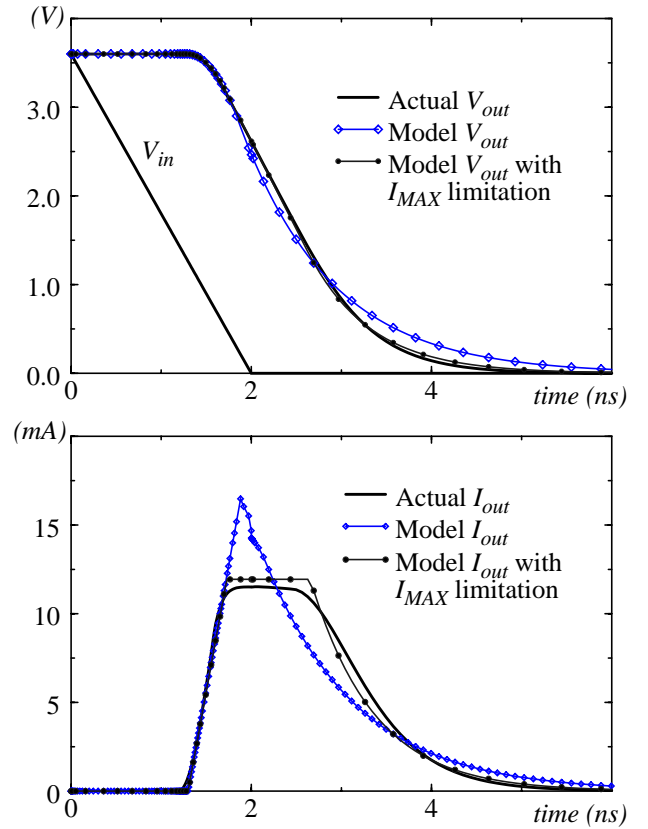

Fig. 4: Comparison of the *four region* model with and without current limiting factor (I_{MAX}) with SPICE simulation results.

VI - MODEL CHARACTERIZATION METHODOLOGY

In order to characterize the parameters of the model a "Design Of Experiments" technique [11] has been applied. After the range of variation of the operating conditions (i. e. T_{IN}, C_L, process, temperature and power supply) has been specified a Central Composite Design [11] is used to generate an optimal set of sampling points to be simulated. Then the five parameters of the *four region* model are obtained by using the Gauss-Newton algorithm [12] in order to fit V_{out} in a predefined time interval (e. g. from 10% to 90% of V_{DD}), for every point of the CCD. Finally, a second order polynomial approximation for V_{T0}, R_{M0}, V_{T1}, R_{M1}, R_F as function of the operating conditions is derived by using least squares.

The current limiting factor I_{MAX} is preliminarily determined as a function of V_{DD} and T by using the same procedure.

VII - RESULTS

The application of the proposed technique to a 0.35 μm CMOS library is presented in this section. A wide range of variation for the operating conditions has been specified:

- the input transition time T_{IN}:

  $(T_{INmin}, 10 \times T_{INmin})$

  where T_{INmin} is the smallest transition time that can be used in the library.
- the output load capacitance C_L:

  $(C_{OUTmin} \times DRIVE, 20 \times C_{OUTmin} \times DRIVE)$

  where C_{OUTmin} is the minimum capacitance of of the input pins in the library and DRIVE is the driving capability of the cell.
- the operating temperature T: (0 , 100) °C.
- the voltage supply V: (3.0 , 3.6) V.

With this setup, the Central Composite Design generates 25 simulations for both falling and rising transition. Other 9 simulations are necessary in order to obtain the value of I_{MAX}, for a total of 34 simulations. This has to be compared with an average value of 16 characterization points for every operating condition corner that is typical of a look-up table model.

The accuracy of the delay model with respect to SPICE, for different cells, evaluated over the characterization grid points is shown in table 1. As expected, the inverter is the most critical cell. In fact the maximum percent error is the largest one for this cell. However it has to be noted that the apparently large 16% error, actually represents only a delay error of the order of few ps (i.e. less than 20 ps), which is

almost comparable with the precision of the simulator. In order to show the predictive capability of the macromodel, a look-up table is generated at V=3.3V and T=25°C and the results are compared with measures from SPICE simulations. The maximum percent error that one obtains in this case is generally less than 9% as shown on table 2. Finally, figure 5 compares the accuracy obtained by replacing the *three region* model with the *four region* model for the propagation delay of the INV x 32 cell.

TABLE 1: ON-GRID MAX PERCENT ERROR OF THE MODEL FOR THE PROPAGATION DELAY AND TRANSITION TIME OF DIFFERENT CELLS

CELL TYPE	Prop. time (tpd) err.		Trans. time (tt) err.	
	Max	Std dev	Max	Std dev
INV x 1	10.6 %	3.4 %	6.7 %	3.5 %
INV x 8	12.2 %	4.0 %	9.1 %	3.7 %
INV x 32	16.0 %	4.3 %	9.7 %	3.9 %
BUF x 32	4.8 %	2.1 %	10.7 %	5.5 %
NAND x 1	10.4 %	3.4 %	11.4 %	4.0 %
NOR x 1	10.8 %	3.6 %	10.6 %	4.4 %
OR x 4	5.4 %	1.9 %	15.2 %	6.2 %

TABLE 2: MAX PERCENT ERROR ON LOOK-UP TABLE VALUES AT 3.3 V, 25 ° C GENERATED FROM THE MODEL FOR DIFFERENT CELLS

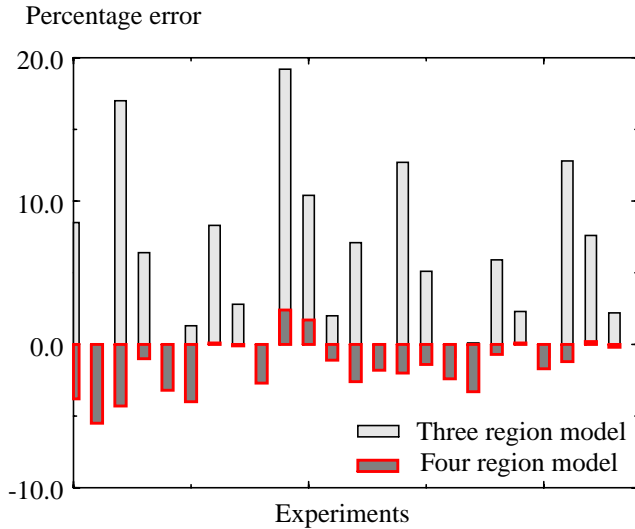| CELL TYPE	Prop. time (tpd) err.		Trans. time (tt) err.	
	Max	Std dev	Max	Std dev
INV x 1	9.1 %	2.8 %	6.6 %	3.1 %
INV x 8	7.2 %	2.6 %	7.9 %	3.3 %
INV x 32	5.4 %	2.4 %	7.4 %	3.2 %
BUF x 32	5.1 %	2.9 %	10.0 %	7.0 %
NAND x 1	8.7 %	3.4 %	7.1 %	3.7 %
NOR x 1	7.6 %	3.8 %	5.9 %	3.1 %
OR x 4	4.9 %	2.3 %	9.6 %	6.9 %

Fig. 5: Accuracy comparison between the *three region* and the *four region* model on propagation delay of INV x 32.

VIII - CONCLUSIONS AND FUTURE WORK

A new gate delay modeling methodology has been presented in this paper. The most important features that have been demonstrated are: improved accuracy with respect to the current state of the art, good predictive capability, reduced characterization effort. The application of the proposed methodology to a 0.35μm CMOS digital standard cell library has been presented, showing considerably good results. The model can be easily extended to deal with non purely capacitive load and to account for non linear input waveforms. The integration of the proposed gate delay model with an AWE based interconnect delay algorithm will be addressed as future work.

IX - REFERENCES

[1] C. L. Ratzlaff, L. T. Pillage, "RICE: Rapid interconnect circuit evaluating using AWE", *IEEE Trans. Computer-Aided Design*, 1994, vol. 13, pp. 763-776.

[2] B. Tutuianu, F. Dartu, L. T. Pileggi, "An Explicit RC-Circuit Delay Approximation Based on the First Three Moments of the Impulse Response", *33st ACM/IEEE Design Automation Conf.*, 1996, pp. 611-616.

[3] F. Dartu, B. Tutuianu, L. T. Pileggi, "RC-Interconnect Macromodels for Timing Simulation", *3st ACM/IEEE Design Automation Conf.*, 1996, pp. 544-547.

[4] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by Pade' approximation via the Lanczos process", *IEEE Trans. Computer-Aided Design*, 1995, vol. 14, N. 5, pp. 639-649.

[5] N. H. E. West and K. Eshraghian, "Principle of CMOS VLSI Design", *Empirical Delay Models, 2nd ed. Reading, MA: Addison-Wesley*, 1992, pp. 213.

[6] M. Horowitz, "Timing models for MOS Circuits", *Stanford University Dissertation*, Chapter 5, 1985.

[7] F. Dartu, N. Menezes, J. Qian and L. T. Pillage, "A gate-delay model for high speed CMOS circuits", *31st ACM/IEEE Design Automation Conference*, 1994, pp. 576-580.

[8] F. Dartu, L. T. Pileggi, "Modeling Signal Waveshapes for Empirical CMOS Gate Delay Models", *PATMOS '96*, p. 57.

[9] T. Sakurai, "Alpha-Power Law MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas", *IEEE Journal of Solid-State Circuits*, 1990, p. 584.

[10] A. Nabavi-Lishi, N. C. Rumin, "Inverter Models of CMOS Gates for Supply Current and Delay Evaluation", *IEEE Trans. Computer-Aided Design*, 1994, vol. 13, N. 10, pp. 1271-1279.

[11] G. E. P. Box and N. R. Draper, "Empirical Model Building and Response Surface", *J. Wiley and sons*, 1987.

[12] D. G. Luenberger, "Linear and Nonlinear Programming", *2nd ed. Reading, MA: Addison-Wesley*, 1984, p. 261.