Structure Independent Representation of Output Transition Time for CMOS Library

Philippe Maurine, Nadine Azemard, Daniel Auvergne

Structure Independent Representation of Output Transition Time for CMOS Library

P. Maurine, N. Azemard, and D. Auvergne

LIRMM UMR 5506 Univ. de Montpellier II 161 Rue Ada 34392 Montpellier France

Abstract. Non zero signal rise and fall times significantly contribute to the gate propagation delay. Designers must accurately consider them when defining timing library format. Based on a design oriented macro-model of the timing performance of CMOS structures, we present in this paper a general representation of transition times allowing fast and accurate cell performance evaluation. This general representation is then exploited to define a robust characterization protocol of the output transition time of standard cells. Both the representation and the protocol are finally validated comparing calculated gate input-output transition time values with standard look-up representation obtained from Hspice simulations (Bsim3v.3, level 69, 0.25µm process).

1 Introduction

In deep submicron technologies the propagation delay of any CMOS cell (i) is strongly dependent on the input ramp duration $\tau_{IN}(i)$ applied to its gate, which is the output transition time $\tau_{OUT}(i-1)$ of the preceding gate. As a consequence the accurate characterization of the output transition time of the different cells in their design environment is of prime importance in speed performance verification or optimization steps.

In the standard industrial approach the tabular method is used. The performance of a predefined set of cells is obtained from electrical simulations performed for a limited number of design conditions, such as load and input transition time values [1]. The resulting data are then listed in tables containing typically 25, 49, or 81 operating conditions (number of loading conditions 5,7 or 9 $\times$ number of input ramp conditions 5,7 or 9). Intermediate conditions are then directly obtained from a linear interpolation between these predefined points. Due to the non-linear variation, in submicron process, of the propagation delay and transition time with the loading and controlling conditions, this method may induce significant errors when interpolating in the non-linear part of the variation. As an example, in Fig.1 we illustrate the evolution of the output transition time of an inverter designed in a 0.25µm process. Here, this inverter is controlled by a rising linear input ramp of duration τ_{IN} and loaded by 5 times its input gate capacitance. As shown the values of the output transition time, τ_{OUT} interpolated from the look up table, may be underestimated by nearly 15%, compared to the values obtained from Hspice simulations.

B. Hochet et al. (Eds.): PATMOS 2002, LNCS 2451, pp. 247–257, 2002.

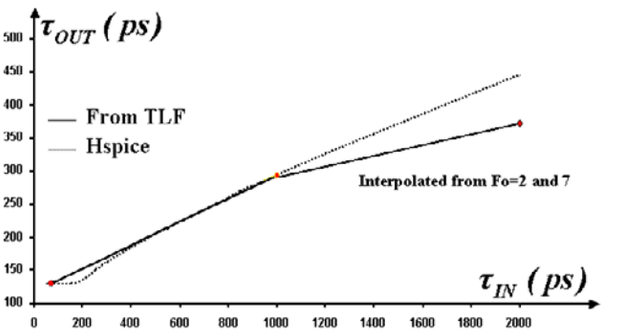

Fig. 1. Comparison of the output transition time values interpolated from the TLF and the simulated ones (Hspice).

It is clear that, for a given cell, the relative accuracy obtained (using Hspice simulation as a reference) with a tabular method strongly depends on the size or the granularity level of the table. Typically the size of the table is limited to 5 to 10 controlling and loading conditions by element, in order to reasonably limit the characterization time to few months.

As a result, the definition of a robust protocol of characterization, based on a uniform representation of the performance of a library is of great interest for cell designers. Indeed it must allow to increase the number of cells in a typical library, and/or to manage the trade off between the accuracy of the performance estimation and the time necessary to characterize a complete library.

Using a design oriented modeling of the CMOS cell timing performance, we propose in section 2, a unified representation of the CMOS cell output transition time, allowing a complete design space representation. Then, in section 3, we deduced from this unified representation a robust characterization protocol of the output transition time of typical CMOS structures. Conclusion is given in section 4.

2 Output Transition Time Modeling

2.1 General Expression

A lot of work has been devoted to the modeling of the output transition time [4-14]. It has been clearly shown that, for CMOS structure, τ_{OUT} can be obtained from the modeling of the charging (discharging) current that flows during the structure switching process.

Moreover it has been demonstrated [2,12] that the evaluation of both the maximum current I_{MAX} that can provide a structure, and the amount of charge ($C \cdot V_{DD}$) to be removed from the output node is sufficient to model the output transition time. More precisely, considering a linear variation of the output voltage, it has been shown that the driving element can be model as a current generator supplying a constant current

to the output loading capacitance. Consequently, a simple first order expression of the output transition time τ_{OUT} can then be obtained from

$$\boldsymbol{\tau_{OUT}} = \frac{\boldsymbol{C \cdot V_{DD}}}{\boldsymbol{I_{MAX}}}, \tag{1}$$

where τ_{out} represents the time spent by the output voltage to swing over the full supply voltage value V_{DD}, C is the output loading capacitance, and I_{MAX} the maximum value of the discharging (charging) current.

2.2 Inverter Switching Current and Definition of the Load

As shown in [2] the evaluation of the maximum current available in the structure imposes to consider two controlling conditions: the *Fast* and the *Slow* input ramp domains. The *Fast* input control range is obtained when the input signal reaches its maximum (minimum) value before the output begins to vary. In this case, the switching current exhibits an almost constant value. In the *Slow* input control range the cell output voltage varies in the same time interval than the input one. In this situation, for which a short circuit occurs between the N and P transistors, the switching current available in the cell presents a maximum value smaller than in the *Fast* input case. Moreover this maximum value depends on the value of the input transition time τ_{IN}. Using, for deep submicron process, the Sakurai's representation of the drain-source current with α=1 [3] the evaluation of the current in the *Fast* input range is straightforward for an inverter. Considering the maximum value of the input control voltage we obtain

$$\boldsymbol{I_{MAX}^{Fast}} = \boldsymbol{K_{N,P} \cdot W_{N,P} \cdot \left(V_{DD} - V_{TN,P}\right)} \tag{2}$$

for an output falling or rising edge, respectively.

The evaluation of the current value in the *Slow* input range is quite more difficult. However taking advantages of the symmetry properties of the current wave shape [2], the maximum current value can be evaluated from

$$\boldsymbol{I_{MAX}^{Slow}} = \sqrt{\frac{\boldsymbol{K_{N,P} \cdot W_{N,P} \cdot V_{DD}^2 \cdot C}}{\boldsymbol{\tau_{IN}}}} \tag{3}$$

where τ_{IN} is the transition time of the cell controlling signal, and C is the output load seen by the inverter defined as:

$$\boldsymbol{C = C_L + C_{PAR} + C_M} \tag{4}$$

where C_L is the sum of the input capacitance of the output loading gates and of the interconnect capacitance, C_{PAR} is the contribution of the cell parasitic capacitance and finally C_M is the contribution of the coupling capacitance [15]. Equation (4) leads to the following definition of the usual fan out factor

$$F_O = \frac{C}{C_{IN}} = \frac{C_L}{C_{IN}} + \frac{C_{PAR}}{C_{IN}} + \frac{C_M}{C_{IN}} = F_O^L + F_O^{PAR} + F_O^M \tag{5}$$

Thus Fo is the sum of three contributions.

- The first one is, as defined earlier, mainly due to the logic following the inverter and to the eventual routing capacitance. Thus it is entirely independent of the considered inverter. We call it the logic contribution.
- The second contribution is mainly due to the diffusion capacitance and the gate internal interconnect; its evaluation gives a good indicator of the quality of the cell design.
- The last contribution is due to the accumulation of charge in the channel of the P (N) transistor that must be removed during the switching process. These charges are usually modeled by an equivalent capacitance as proposed by Meyer [15]. Note, that neglecting this contribution may induce an underestimation of 20% of the load for small value of the fan out factor (F_O^L=1).

2.3 Inverter Output Transition Time Model

Finally, with such a definition of the load, the inverter output transition time is directly obtained by replacing I_{MAX} in (1) by its appropriate expression (2 or 3). This gives respectively for an input rising and falling edge

$$\tau_{OUT-HL}(Inv.) = MAX\left\{\tau_{OUT-HL}^{Fast}\sqrt{\frac{(V_{DD}-V_{TN})}{V_{DD}}}\cdot\sqrt{\tau_{OUT-HL}^{Fast}\cdot\tau_{IN-HL}}\right\} \tag{6}$$

$$\tau_{OUT-LH}(Inv.) = MAX\left\{\tau_{OUT-LH}^{Fast}\sqrt{\frac{(V_{DD}-V_{TP})}{V_{DD}}}\cdot\sqrt{\tau_{OUT-LH}^{Fast}\cdot\tau_{IN-LH}}\right\}. \tag{7}$$

Where τ_{OUT}^{Fast} is called the step response of the inverter. Its value can be directly obtained from (1) and (2) as

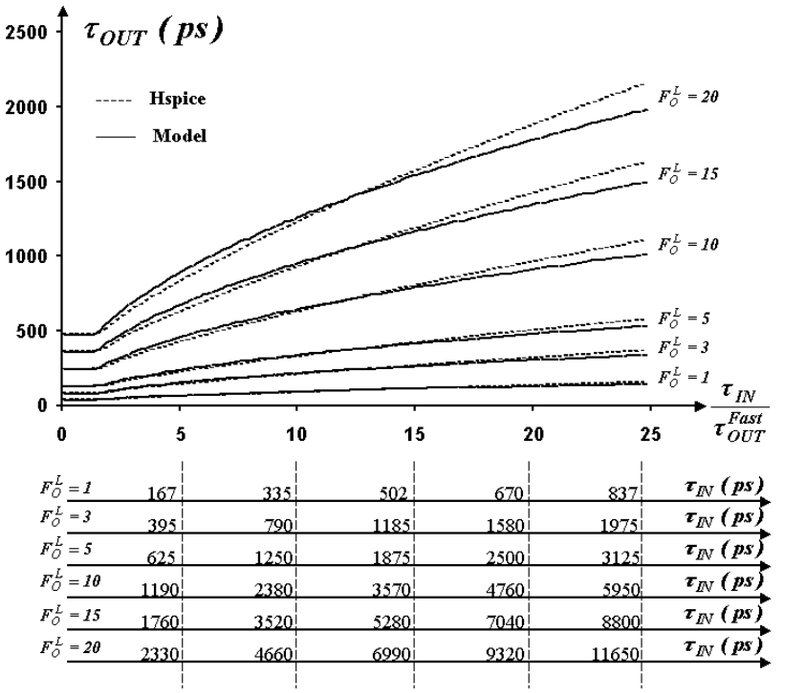

Fig. 2. Comparison between simulated and calculated (— eq.5) values of the output transition time of an inverter (W_N=0.72µm, L=0.25µm, k=1) for various loading and controlling conditions.

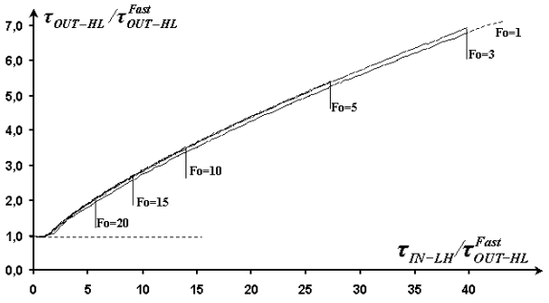

Fig. 3. Illustration of the general representation of the inverter output transition time

$$\tau_{OUT-HL}^{Fast} = \frac{V_{DD} \cdot L \cdot C_{OX}}{(V_{DD} - V_{TN}) \cdot K_N} \cdot \frac{C}{C_N} = \tau_{ST} \cdot \frac{C}{C_{IN}} \cdot (1+k) = \tau_{ST} \cdot \left(F_O^L + F_O^{PAR} + F_O^M\right) \cdot (1+k) \quad (8)$$

$$\tau_{OUT-LH}^{Fast} = \frac{V_{DD} \cdot L \cdot C_{OX}}{(V_{DD} - V_{TP}) \cdot K_P} \cdot \frac{C}{C_P} = \tau_{ST} \cdot R_\mu \cdot \frac{C}{C_{IN}} \cdot \frac{(1+k)}{k} = \tau_{ST} \cdot R_\mu \cdot \left(F_O^L + F_O^{PAR} + F_O^M\right) \cdot \frac{(1+k)}{k}$$

where C_N and C_p represent the gate capacitance of the N and P transistors, $C_{IN}=C_N+C_p$, $k=C_p/C_N$ is the cell internal configuration ratio and τ_{ST} appears as a technology dependent parameter characteristic of the process speed.
As shown in (6), for a well-defined inverter structure, in the *Fast* input range, the output transition time only depends on the ratio (load/ inverter input capacitance). In the *Slow* range, the output transition time exhibits an extra input duration time dependency that reflects the complete inter-stage interaction in an array of inverters or gates.
This is illustrated in Fig.2 where we represent the transition time for an output falling edge, for different loading factors, versus the input transition time value τ_{IN}. As shown the values of the transition time calculated with (6) are in good agreement with the simulated one.

2.4 Gate Output Transition Time Model

Similar expression can be obtained for gates, considering in (1) the ratio of maximum available current between an inverter and a gate with identically sized transistors such as

$$Red_{TOP,BOT,MID}^{Fast,Slow} = \frac{I_{MAX}^{Fast,Slow}(Inv)}{I_{MAX-Top,Bot,Mid}^{Fast,Slow}(gate)} \quad (9)$$

Where the Top, Bot, Mid are subscripts associated to the input used to control the serial array of transistors. The definition of this reduction factor, introduced as logical weight in [16], is detailed in [2], and allows to express the output transition time of the gates as

$$\tau_{OUT-HL}(Gate) = MAX\left\{ Red_{HL-Top,Bot,Mid}^{Fast} \cdot \tau_{OUT-HL}^{Fast} \;;\; Red_{HL-Top,Bot,Mid}^{Slow} \cdot \sqrt{\frac{(V_{DD} - V_{TN})}{V_{DD}}} \cdot \sqrt{\tau_{OUT-HL}^{Fast} \cdot \tau_{IN-HL}} \right\} \quad (10)$$

for a rising edge applied on one of the input (Top, Bot, Mid). For an input falling edge this becomes

$$\tau_{OUT-LH}\,(Gate) = MAX\left\{\begin{array}{l} Red^{\,Fast}_{LH-Top,Bot,Mid}\cdot\tau^{\,Fast}_{OUT-LH} \\ Red^{\,Slow}_{LH-Top,Bot,Mid}\cdot\sqrt{\frac{(V_{DD}-V_{TP})}{V_{DD}}}\cdot\sqrt{\tau^{\,Fast}_{OUT-LH}\cdot\tau_{IN-LH}} \end{array}\right\} \tag{11}$$

3 Output Transition Time Representation

3.1 Inverters

For simplicity, let us now only consider the case of rising edges applied to the input of gates. As it can be deduced from (6), in the *Fast* input range, τ_{OUT}^{FAST} is characteristic of the inverter structure (gate) and of its load. Considering the sensitivity to the input slope, τ_{OUT}^{FAST} can be used as an internal reference of the output transition time of the considered structure. In this case, (6) becomes

$$\frac{\tau_{OUT-HL}}{\tau^{\,Fast}_{OUT-HL}}(Inv.) = MAX\left\{\begin{array}{l} 1 \\ \sqrt{\frac{(V_{DD}-V_{TN})}{V_{DD}}}\cdot\sqrt{\frac{\tau_{IN-HL}}{\tau^{\,Fast}_{OUT-HL}}} \end{array}\right\} \tag{12}$$

We clearly observe in this equation that, using τ_{OUT}^{FAST} as a reference, the normalized inverter output transition time only depends on the input transition time. It has the same value for inverters with different configuration ratio value or loading conditions. This is illustrated in Fig.3 where we represent, using τ_{OUT}^{FAST} as a reference, the output transition time variations displayed in Fig.1. As expected all the curves pile up on the same one, representing the output transition time sensitivity to the input transition time. This is obtained for the complete family of inverters with different values of the configuration ratio and the load. The final value for specific cells is then directly obtained from the evaluation of τ_{OUT}^{FAST}, in (8), that contains the structure and load dependency.

3.2 Gates

As formerly mentioned, the extension to gates is straightforward, multiplying the right part of (6) by a reduction factor representing the ratio of current available in an inverter and a gate implemented with identically sized transistors [2,15].

$$\frac{\tau_{OUT-HL}}{\tau^{\,Fast}_{OUT-HL}}(Gates) = MAX\left\{\begin{array}{l} Red^{\,Fast}_{HL-Top,Bot,Mid} \\ Red^{\,Slow}_{HL-Top,Bot,Mid}\cdot\sqrt{\frac{(V_{DD}-V_{TN})}{V_{DD}}}\cdot\sqrt{\frac{\tau_{IN-HL}}{\tau^{\,Fast}_{OUT-HL}}} \end{array}\right\} \tag{13}$$

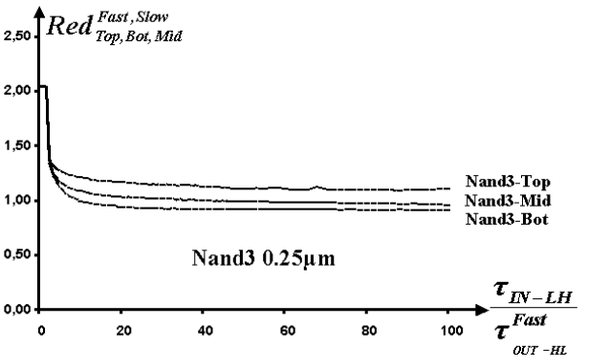

Fig. 4. Input transition time sensitivity of the reduction factor value associated to Nand3 gates designed in the 0.25µm process.

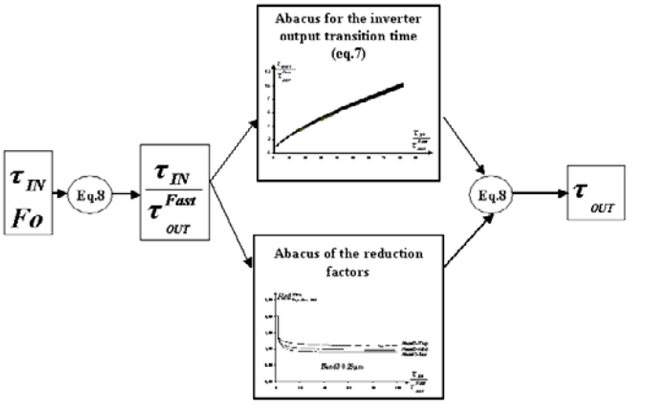

Fig. 5. Synoptic of the evaluation method of the output transition time.

3.3 Output Transition Time Characterization Protocol

From (12) and (13), it appears that it is possible to characterize the output transition time of all the gates of the library, with a reduced set of electrical simulations organized as follows.

- The extraction of τ_{ST} and Rµ values can be done trough the simulation of the step response of an heavily loaded inverter using (8) (the step response being extrapolated from the time spent by the output voltage to switch between 60% and 40% of V_{DD} or inversely).
- The simulation of the output transition time of any inverter can then provide the graph associated to (12).

- The simulation of the maximum discharging current available in typical Nand and Nor gates and their equivalent inverters, supply the unique representation (Fig.4) of the reduction factor value sensitivity to the input transition time.

Then, the evaluation of the output transition time of any gate of a given library can be process as illustrated by Fig.5. For a specified gate, fan out factor, and duration time value of the controlling input ramp, we determine from (8) the step response of this gate, and the value of the ratio $\tau_{IN}/\tau_{OUT}^{FAST}$. Then we deduce the value of the corresponding reduction factor and ratio $\tau_{OUT}/\tau_{OUT}^{FAST}$ to finally get the corresponding value of the output transition time.

3.4 Validation

In order to validate our approach, we use this protocol to fill the TLF associated to a 0.25µm technology. We then compare the results obtained to the TLF given by the foundry. The relative discrepancies obtained were below 6% for inverters, 10% for Nand2 and Nor2 gates, and 13% for Nand3 and Nor3 gates, validating the proposed unified representation.

However, this validation does not give evidence of the efficiency of the protocol. Indeed, as it needs a specific and reduced set of simulations, we can increase without a great time penalty the number of operating conditions reported in the calculated TLF, obtaining an improved resolution of the output transition time estimation. This means that the time necessary to apply this protocol exhibits a weak sensibility to the size of the table. As an illustration of this fact, we compare, in Fig.6 the value of the output transition time obtained using this method, the usual TLF, and Hspice simulations. This has been performed on an inverter designed in 0.25µm process. As shown, the improvement in accuracy is significant.

Moreover, another great advantage of this protocol is that the time spent to calibrate a library is almost independent of the number of gates it contains. This could be of great benefit in order to define a quasi-continuous sizing of cells or with the advent of on-the-fly-synthesized gate library.

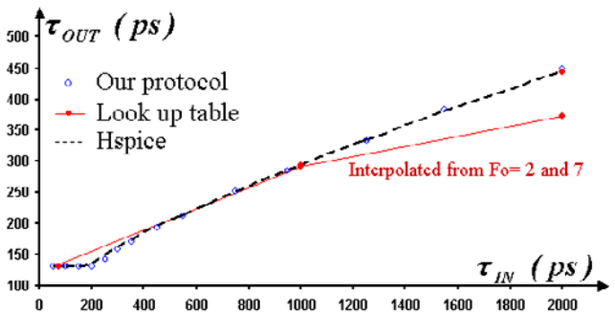

Fig. 6. Comparison of the output transition time value obtained using our protocol, deduced from the usual TLF, and simulated (Hspice).

4 Conclusion

Using an analytical model to evaluate the maximum switching current value we have obtained a simple but accurate design oriented representation of the output transition time of CMOS structures. We have shown its sensitivity to the design parameters, loading factor and input duration time value. Defining *Fast* and *Slow* input ramp controlling domain we have defined a reference τ_{OUT}^{FAST} for the input-output duration time that characterizes the switching cell. It can be used to obtain a unique representation of the timing performances for each category of library cell, independently of their configuration ratio or load. A protocol of characterization of the output transition time of CMOS structure, using only a reduced set of electrical simulations has been developed and validated on a 0.25µm process. This representation appears to be of great help in defining timing library format (TLF) since only one set of simulation by family of cell is necessary to characterize all the gates of different configuration ratio, size and loading conditions. Moreover the full representation obtained in Fig.2 gives a clear idea of the design range to be explored and mostly of the non linear part of the variation where the use of standard look up tables may induce large uncertainties in design performance estimation. The extension of this work to the definition of a characterization protocol for the propagation delay is under progress.

References

[1] Cadence open book "Timing Library Format References" User guide, v. IC445, 2000.
[2] P. Maurine, M. Rezzoug, D. Auvergne “Output transition time modeling of CMOS structures” pp. V-363-V-366, ISCAS 01, Sydney, Australia
[3] T. Sakurai and A.R. Newton, "Alpha-power model, and its application to CMOS inverter delay and other formulas", J. of Solid State Circuits vol.25, pp.584-594, April 1990.
[4] J.R. Burns, "Switching response of complementary symmetry MOS transistor logic circuits" RCA Review, vol. 25, pp627-661, 1964.
[5] N. Hedenstierna and K.O. Jepson, " CMOS circuit speed and buffer optimization", IEEE Trans. Computer-Aided Design, vol. 6, pp. 270-281, March 1987.
[6] I. Kayssi Ayman, A. Sakallah Karem, M. Burks Timothy, " Analytical transient response of CMOS inverters" IEEE Trans. on circuits and Syst. Vol. 39, pp. 42-45, 1992
[7] T. Sakurai and A.R. Newton, "Alpha-power model, and its application to CMOS inverter delay and other formulas", J. of Solid State Circuits vol. 25, pp. 584-594, April 1990.
[8] Santanu Dutta, Shivaling S. Mahant Shetti, and Stephen L. Lusky, "A Comprehensive Delay Model for CMOS Inverters" J. of Solid State Circuits, vol. 30, no. 8, pp. 864-871, 1995.
[9] T. Sakurai, A. R. Newton “A simple MOSFET model for circuit analysis” IEEE Trans. On electron devices, vol.38, n°4, pp. 887-894, April 1991.
[10] L. Bisdounis, S. Nikolaidis, O. Koufopavlou "Analytical transient response of propagation delay evaluation of the CMOS inverter for short channel devices” J. of Solid State Circuits vol. 33, n°2, pp. 302-306, Feb.1998.
[11] A. Hirata, H. Onodera, K. Tamaru "Proposal of a Timing Model for CMOS Logic Gates Driving a CRC π load" in proc. of the Int. Conf. On CAD 1998 (San Jose), pp 537-544.
[12] D. Auvergne, J. M. Daga, M. Rezzoug, “Signal transition time effect on CMOS delay evaluation ”IEEE Trans. on Circuit and Systems-1, vol.47, n°9, pp.1362-1369, sept.2000

[13] J. M. Daga, D. Auvergne "A comprehensive delay macromodeling for submicron CMOS logics" IEEE J. of Solid State Circuits Vol.34, n°1, pp.42-55, 1999.
[14] A. Nabavi-Lishi "Inverter Models of CMOS Gates for Supply Current and Delay Evaluation" IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems, Vol. 13, N° 10, October 1994.
[15] J. Meyer "Semiconductor Device Modeling for CAD" Ch. 5, Herskowitz and Schilling ed. mc Graw Hill, 1972.
[16] I. Sutherland, B. Sproull, D. Harris "Logical effort : designing fast CMOS circuits" Morgan Kaufmann Publishers.