



HAL
open science

Continuous Representation of the Performance of a CMOS Library

Benoit Lasbouygues, J. Schindler, Sylvain Engels, Philippe Maurine, Xavier Michel, Nadine Azemard, Daniel Auvergne

► **To cite this version:**

Benoit Lasbouygues, J. Schindler, Sylvain Engels, Philippe Maurine, Xavier Michel, et al.. Continuous Representation of the Performance of a CMOS Library. ESSCIRC: European Solid-State Circuits Conference, Sep 2003, Estoril, Portugal. pp.595-598, 10.1109/ESSCIRC.2003.1257205 . lirmm-00239459

HAL Id: lirmm-00239459

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00239459>

Submitted on 11 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Continuous representation of the performance of a CMOS library

B. Lasbouygues¹, J. Schindler², S. Engels², P. Maurine³, N. Azémard³, D. Auvergne³

¹ ISIM, Université de Montpellier II, Place E. Bataillon, 34095 Montpellier France

² STMicroelectronics Design Department 850 rue J. Monnet, 38926, Crolles, France

³ LIRMM, UMR CNRS/Université de Montpellier II, (C5506), 161 rue Ada, 34392 Montpellier, France
pmaurine, azemard, auvergne@lirmm.fr

Abstract:

In a standard industrial approach the timing performance verification is obtained using a tabular method that necessitates a great amount of simulations. They must specify, for each drive in each logic family: the load, the input slope, the temperature and the supply voltage sensitivity, for each edge, of the transition time and propagation delay. Extending a logical effort like based model of timing performance of CMOS structures we show in this paper that it is possible to define a specific performance representation allowing a continuous representation of the performance sensitivity of a complete family. We describe the parameter calibration procedure and validate the proposed representation on a 0.13 μm process by comparing the performance sensitivity deduced from this representation to sensitivity values obtained from electrical simulations performed with the full process model of the foundry.

1. Introduction

The complete characterization of the timing performance of a library involves a huge number of simulations. The performance of each gate on a path, for each loading and control condition, is obtained from an interpolation between a set of predefined values. These values are determined from electrical simulations performed for a limited number of design conditions, such as load, input transition time, supply voltage value and operating temperature [1]. Characterizing each edge of the transition time and the propagation delay of each library cell, for typically five loading and input ramp conditions, involves 100 simulations. Then considering the process corners, with different supply voltage and temperature values imposes to simulate thousand of points to characterize a logic function designed with different drives. This huge number of simulations just allows to represent the design space with few loading and control conditions. Intermediate conditions must then be interpolated using a linear characteristic equation (Synopsys) :

$$f(\tau_{\text{IN}}, C_L) = A\tau_{\text{IN}} + BC_L + C\tau_{\text{IN}}(CL + D).$$

It is well recognized that the transition time and the propagation delay exhibit a nonlinear variation with respect to the control and loading conditions. This imposes variable simulation steps because interpolating in this range may induce significant errors. Usually the largest error occurs for the smallest values of the load and the input transition time. It is clear that the relative accuracy, obtained with a tabular method, is strongly dependent on the granularity of the table that is not necessarily constant but must cover a significant part of the design range. For that, indications for defining the granularity and the coverage of the design space must be available.

In this paper we propose a continuous representation of the timing performance of a CMOS library. This representation defines the propagation delay time and output transition time sensitivities of the cells to the design space parameters, such as the load, the input transition time, the supply voltage and the temperature values. The performance model is presented in part 2. In part 3 we detail the calibration method to be used on a specific process. Application is given in part 4 to a 0.13 μm STM process, before to conclude in part 5.

2. Model

2.1 Gate switching delay An accurate delay model must be input slope dependent and must distinguish between falling and rising signals. Considering the input-to-output coupling effect [2], the input slope effect can be introduced in the propagation delay as:

$$\begin{aligned} t_{\text{HL}}(i) &= \frac{v_{\text{TN}}}{2} \tau_{\text{INLH}}(i-1) + \left(1 + \frac{2C_M}{C_M + C_L}\right) t_{\text{HLstep}}(i) \\ t_{\text{LH}}(i) &= \frac{v_{\text{TP}}}{2} \tau_{\text{INHL}}(i-1) + \left(1 + \frac{2C_M}{C_M + C_L}\right) t_{\text{LHstep}}(i) \end{aligned} \quad (1)$$

$\tau_{\text{INHL,LH}}$ is the duration time of the input signal, generated by the controlling gate, C_M the coupling capacitance between the input and output nodes [3], that can be evaluated or directly calibrated from SPICE simulation. Indexes (i), (i-1) refers to the location of the cell in the array.

2.2 *Inverter transition time.* The output transition time (defining the input transition time of the following cell) of a CMOS structure [4] can be obtained from the modeling of the charging (discharging) current that flows during the switching process of the structure and from the amount of charge ($C_L \cdot V_{DD}$) to be exchanged with the output node as:

$$\begin{aligned}\tau_{\text{outHL}} &= \frac{C_L \cdot V_{DD}}{I_{N\text{Max}}} \\ \tau_{\text{outLH}} &= \frac{C_L \cdot V_{DD}}{I_{P\text{Max}}}\end{aligned}\quad (2)$$

C_L is the output load and V_{DD} the supply voltage value. In (2) the output voltage variation has been supposed linear and the driving element considered as a constant current generator delivering the maximum current available in the structure. To evaluate the maximum current available in the structure two controlling conditions must be considered, the fast and slow input control ranges.

In the fast input range the switching current exhibits a constant and maximum value:

$$I_{\text{MAX}}^{\text{Fast}} = K_{N,P} \cdot W_{N,P} \cdot (V_{DD} - V_{TN,P}) \quad (3)$$

obtained for deep submicron from the Sakurai's representation [5] with $\alpha=1$. Here $K_{N,P}$ is an equivalent conduction coefficient to be calibrated on the process. From (2,3) we obtain the expression of the transition time for a fast input control condition as:

$$\begin{aligned}\tau_{\text{outHL}}^{\text{Fast}} &= \tau \cdot (1+k) \cdot \frac{C_L}{C_{IN}} \\ \tau_{\text{outLH}}^{\text{Fast}} &= \tau \cdot \frac{(1+k)}{k} \cdot R \cdot \frac{C_L}{C_{IN}}\end{aligned}\quad (4)$$

(4) represents an extension of the logical effort model [6] for non symmetrical inverters. τ is a unit delay characterizing the process, $k=W_p/W_n$ the inverter configuration ratio and R the speed ratio between electron and holes.

In the slow input range the input and output voltages vary simultaneously. The resulting short circuit between N and P transistors decreases the current value available at the output. Considering the symmetry property of the current wave shape, this new value of the maximum current available can easily be obtained as:

$$I_{\text{MAX}}^{\text{Slow}} = \sqrt{\frac{K_{N,P} \cdot W_{N,P} \cdot V_{DD}^2 \cdot C_L}{\tau_{INHL,LH}}}\quad (5)$$

Developing (2) with (5) and (4) gives for a falling output edge:

$$\tau_{\text{outHL}}^{\text{Slow}} = \sqrt{\frac{(V_{DD} - V_{TN})}{V_{DD}}} \cdot \sqrt{\tau_{\text{outHL}}^{\text{Fast}} \cdot \tau_{INLH}} \quad (6)$$

with an equivalent expression for a rising edge.

As shown in (4), 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.1 where we represent the transition time for output falling and rising edges, for different values of the configuration ratio, versus the input transition time value τ_{IN} . As shown the values of the transition time, calculated from the maximum value of (4) and (6), are in good agreement with the simulated one.

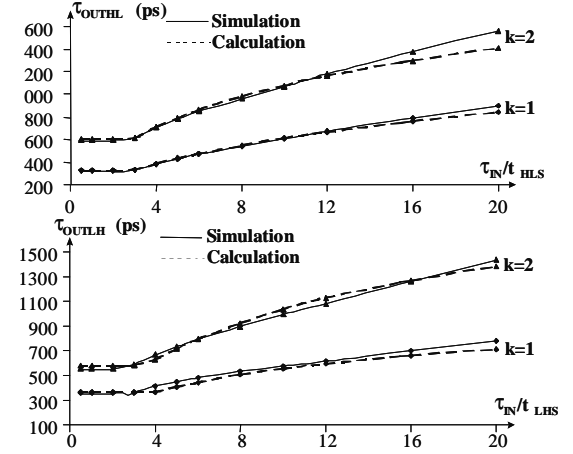


Fig.1. Illustration of the output transition time sensibility to the input transition time τ_{IN} .

2.3 *Extension to gates.* It is obtained [6] by reducing each gate to an equivalent inverter. The current available from the N (P) parallel array of transistors is evaluated as the maximum current of an inverter with identically sized transistors. The array of N (P) series-connected transistors is modeled as an input voltage controlled current generator with a current capability reduced by a factor $DW_{HL,LH}$. This reduction factor (DW) is defined as the ratio of the current available in an inverter to that of a series-connected array of transistors of identical size.

$$DW_{HL,LH}^{\text{Fast,Slow}} = \frac{I_{N,P}^{\text{Fast,Slow}}(\text{Inv})}{I_{N,P}^{\text{Fast,Slow}}(\text{Gate})} \cdot \frac{W_{N,P}(\text{Inv})}{W_{N,P}(\text{Gate})} \quad (7)$$

DW corresponds to the explicit form of the logical effort [6]. Note that depending on the input range and on the rank of the controlled input, for a given gate, it may have different values..

Considering eq.7, the output transition time of a gate is easily obtained under the generic expression of [6] as:

$$\begin{aligned}\tau_{\text{outHL}}^{\text{Fast,Slow}} &= \tau_{\text{HL}}^{\text{Fast,Slow}}(\text{Inv}) \cdot DW_{\text{HL}}^{\text{Fast,Slow}} \\ \tau_{\text{outLH}}^{\text{Fast,Slow}} &= \tau_{\text{LH}}^{\text{Fast,Slow}}(\text{Inv}) \cdot DW_{\text{LH}}^{\text{Fast,Slow}}\end{aligned}\quad (8)$$

2.4 *Supply voltage sensitivity.* As shown in (4), τ represents the output transition time of an ideal inverter (without parasitic capacitance) designed with

identical transistors and loaded by an identical one. Developing (4) from (2) and (3) gives:

$$\tau = \frac{C_{ox} \cdot L}{K_N} \cdot \left(\frac{V_{DD}}{V_{DD} - V_{TN}} \right) \quad (9)$$

$$R = \frac{K_N}{K_P} \cdot \left(\frac{V_{DD} - V_{TN}}{V_{DD} - V_{TP}} \right)$$

Eq.9 gives the supply voltage sensitivity of the transition time and of the propagation delay (1).

2.5 Temperature sensitivity. Can be included in the model considering both the mobility and the threshold voltage variations described in [7,8]:

$$K_{\theta} = K_{nom} \cdot \left(\frac{\theta_{nom}}{\theta} \right)^{XT} \quad (10)$$

$$V_T(\theta) = V_{Tnom} - \delta(\theta - \theta_{nom})$$

K and V_T are respectively the conductivity factor and the threshold voltage, θ_{nom} , θ , represent the reference and the targeted temperature, XT and δ are the temperature coefficients of the mobility and threshold voltage.

Coupling (9) and (10) a general expression for the τ supply voltage and temperature sensitivity can be obtained:

$$\frac{\tau(V_{DD}, \theta)}{\tau_{nom}} = \left(\frac{\theta}{\theta_{nom}} \right)^{XT} \cdot \left(\frac{V_{DD}}{V_{DDnom}} \right) \cdot \frac{1}{\frac{V_{DD} - V_{Tnom} + \delta(\theta - \theta_{nom})}{V_{DDnom} - V_{Tnom}}} \quad (11)$$

The different parameters of (11) can directly be determined from specific simulation conditions that we will define in the next part.

3. Continuous representation and Calibration

3.1 Uniform representation. Let us now consider the case of rising edges applied to the input of gates. As it can be deduced from (4,6,8), in the *Fast* input range, τ_{OUT}^{FAST} is characteristic of the inverter structure (gate) and of its load. Considering the sensitivity of the different expressions 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 condition we can write:

$$\frac{\tau_{outHL}}{\tau_{outHL}^{Fast}} (Inv) = \text{Max} \left[\frac{1}{\sqrt{\frac{V_{DD} - V_{TN}}{V_{DD}}}} \cdot \sqrt{\frac{\tau_{INLH}}{\tau_{outHL}^{Fast}}} \right] \quad (12)$$

Normalizing the output transition time with respect to τ_{outHL}^{Fast} , used as a reference, the resulting expression only depends on the input transition time. It is configuration ratio and load independent. Similar results are obtained for gates and the representation of propagation delay.

This is illustrated in Fig.2 where we represent, using τ_{OUT}^{FAST} as a reference, the output transition time variations for the complete family of inverters of a 0.25 μ m library. As expected all the curves pile up on the same one, representing the output transition time sensitivity to the input transition time. The final value for each specific cell is then directly obtained from the evaluation of τ_{OUT}^{FAST} , in (4), that contains the structure and load dependency.

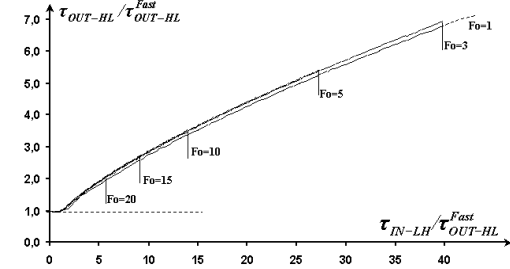


Fig.2. Full representation of the output transition time variation of the 7 inverters of a 0.25 μ m library.

3.2 Parameter extraction method. As shown in Fig.2 and (1,8,12) it appears that the output transition time and the propagation delay of all the gates of a library can be characterized with a reduced set of electrical simulations.

- The τ value is obtained from the output transition τ_{HL} (falling edge) of an heavily loaded inverter (with a known configuration ratio k) controlled by a fast input ramp ($\tau_{IN} < \tau_{OUT}$).
- R is obtained from the ratio τ_{LH}/τ_{HL} .
- For small load the variation of the apparent τ value determines the value of Cpar and C_M .
- In the slow input range ($\tau_{IN} > \tau_{OUT}$), varying τ_{IN} determines the input slope sensitivity (6).
- Using the inverter as a reference the parameters k and DW are determined from τ_{Gate}/τ_{Inv} .
- Eq.9 determines the supply voltage sensitivity.
- The temperature parameters, XT and δ are obtained from the preceding steps realized at different temperature values.

4. Application to a 0.13 μ m library

We have applied this method to a 0.13 μ m library from STM. In this work we just considered the simple gates, an Inverter with 7 different drives, a Nand and a Nor with two and three inputs and 5 different drives. Initially the timing performance (transition time and propagation delay) of all these elements has been characterized from electrical simulations. They are available in tables (TLF, STF) that give, for each edge of the transition time and the propagation delay of each element, for 3 temperature and supply voltage values, the corresponding performance for 5 different values of the load and the input transition time.

We have determined on the different tables the values of the technology parameters and plot in Fig3 to 5, examples of the variation of the transition time and the propagation delay of different logic families. As shown, the performance variation of all the elements of each family can be represented by one curve described by one of the preceding equation.

In Table 1 we compare the τ supply voltage and temperature sensitivity calculated from (9,12) to the value deduced from the simulated values on the look up tables. As shown we obtain an excellent agreement between calculated and simulated values for all the considered supply voltage and temperature range. These variations can then be completely represented by (12) as:

$$\frac{\tau(V_{DD}, \theta)}{\tau} = \left(\frac{\theta}{298} \right)^{1.65} \cdot \left(\frac{V_{DD}}{1.2} \right) \cdot \frac{1}{V_{DD} - 0.62 + 2 \cdot 10^{-3} (\theta - 298)} \cdot \frac{1}{0.58} \quad (13)$$

Where the different coefficients have been directly determined following the procedure given in part 3.

5. Conclusion

Extending of the logical effort model, from physical considerations, we have defined a simple but accurate representation of the timing performance of the simple CMOS structures and their temperature and supply voltage sensitivity. We have introduced a new way of a continuous representation of these performances, allowing to model the complete load and input ramp sensitivity by one curve. We gave a method to calibrate the parameters of this representation that has been completely validated on a 0.13 μ m process for different temperature and supply voltage conditions. The extension of this representation to complex gates is under progress.

References

[1] Cadence open book "Timing Library Format References" User guide, v. IC445, 2000
 [2] K. O. Jeppson, "Modeling the influence of the transistor gain ratio and the input-to-output coupling capacitance on the CMOS inverter delay", IEEE J. Solid State Circuits, vol.29, pp.646-654, 1994.
 [3] J. Meyer "Semiconductor Device Modeling for CAD" Ch. 5, Herskowitz and Schilling ed. mc Graw Hill, 1972.
 [4] C. Mead and L. Conway, "Introduction to VLSI systems", Reading MA: Addison Wesley 1980.
 [5] 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
 [6] I. Sutherland, B. Sproull, D. Harris, "Logical Effort: Designing Fast CMOS Circuits", Morgan

Kaufmann Publishers, INC., San Francisco, California, 1999.

[7] S. M. Sze, 'Physics of Semiconductor Devices', New York, Wiley, 1983.
 [8] J. A. Power et al, 'An Investigation of MOSFET Statistical and Temperature Effects', Proc. IEEE 1992 Int. Conference on Microelectronic Test Structures, Vol. 5, March 1992.

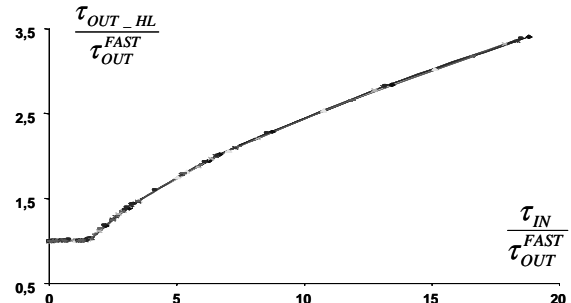


Fig.3. Output transition time representation of the 7 inverters of a 0.13 μ m process

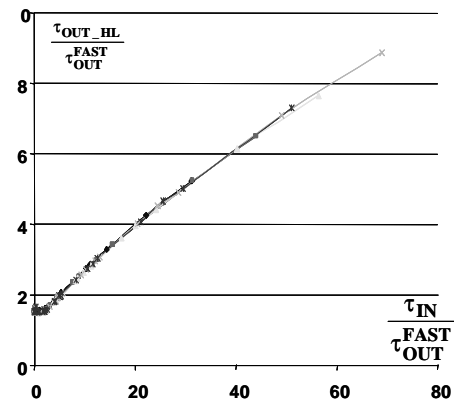


Fig.4. Output transition time representation of the 5 Nand2 of a 0.13 μ m process.

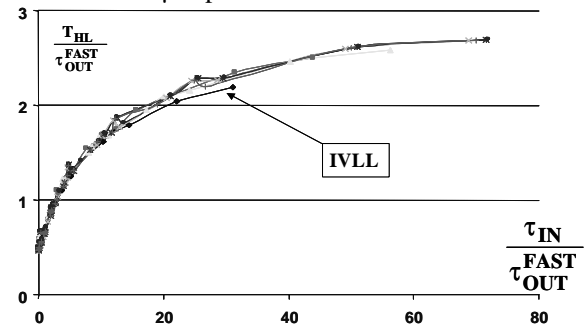


Fig.5. Propagation delay representation of the 7 inverters of a 0.13 μ m process.

τ	V_{DD} (V)						
	1.08		1.2		1.32		
	Model	Simul.	Model	Simul.	Model	Simul.	
Temp. (°K)	233	4.26	3.92	3.86	3.56	3.02	3.3
	298	4.59	4.56	4.05	4.05	3.69	3.66
	398	5.16	5.45	4.85	4.93	4.62	4.58

Table 1. Voltage and temperature sensitivity of the τ parameter, $XT=1.65$, $\delta = 2 \cdot 10^{-3}$.