



HAL
open science

Feasible delay Bound Definition

Nadine Azemard, Michel Aline, Philippe Maurine, Daniel Auvergne

► **To cite this version:**

Nadine Azemard, Michel Aline, Philippe Maurine, Daniel Auvergne. Feasible delay Bound Definition. SOC Design Methodologies, 90, Kluwer Academic Publishers, pp.325-335, 2002, IFIP - The International Federation for Information Processing, 978-1-4757-6530-4. 10.1007/978-0-387-35597-9_40. lirmm-00239363

HAL Id: lirmm-00239363

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00239363>

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.

Feasible delay bound definition

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

*LIRMM : Laboratoire d'Informatique ,de Robotique et de Microélectronique de Montpellier,
UMR 5506 Université Montpellier II //CNRS,
161 rue ADA , 34392 Montpellier cedex 5, France
Phone: (33) 4 67 41 85 21, Fax: (33) 4 67 41 85 00, E-mail: azemard@lirmm.fr*

Abstract: Minimizing the number of iterations when satisfying performance constraints in IC design is of fundamental importance to limit the design iterations. We present a method to determine the feasibility of delay constraint imposed on circuit path. From a layout oriented study of the path delay distribution, we show how to obtain the upper and lower bounds of the delay of combinatorial paths. Then we characterise these bounds and present a method to determine, , the average weighted loading factor allowing to satisfy the delay constraint. Example of application is given on different ISCAS circuits.

Key words: timing closure, delay bounds, fan out

1. INTRODUCTION

Great interest [1-4] has been given to the research of optimal solutions to the problem of transistor sizing under delay constraint. But very few information is available on the direct determination of bounds on delay [5] allowing to evaluate the feasibility of constraints and/or the efficiency of an implementation. This evaluation is of great importance, for all stages of the design flow, in evaluating the quality of any synthesis or implementation style alternative.

In Fig.1 we illustrate the sensitivity of the delay of a path to the width of the transistors. For simplicity we have considered a uniform sizing (W) of the transistors along the path but the trend observed is easily shown to be

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35597-9_40](https://doi.org/10.1007/978-0-387-35597-9_40)

M. Robert et al. (eds.), *SOC Design Methodologies*

© IFIP International Federation for Information Processing 2002

conserved for more general irregular sizing conditions. As shown in this figure, for any circuit topology, a delay-area-power tradeoff can be defined [6-9].

This characteristic can be easily obtained for any path by varying the transistor sizes or, equivalently, the fan out factor [5,10,11], defined as the ratio of the output to input capacitance ($F_{out} = C_{Load}/C_{IN}$) of the gates.

Considering a minimum transistor width implementation, the "maximum" delay value is obtained. We have to note here that this is a correctly designed maximum value, considering that any non justified extra loading on a path may lead to a greater value of the delay. This is also the minimum area solution but, due to controlling ramp effect, not necessarily the minimum power alternative [12,13]. On the other hand the value of the delay obtained for very large (and unpractical) transistor widths, gives the minimum delay that can be satisfied. If the plot in Fig.1 is determined from a post layout extraction of the path, this will give accurate bounds for the implementation from which optimization alternatives for delay-area-power may be selected. Obtained at the logical synthesis step, this plot will need to be updated after considering the interconnect parasitic capacitance introduced by the place and route step. This may induce numerous iterations implying prohibitive CPU time.

From first inspection, we can easily deduce that the difference $\Theta_{Max} - \Theta_{Min}$ obeys a "b/W" law that will be justified later.

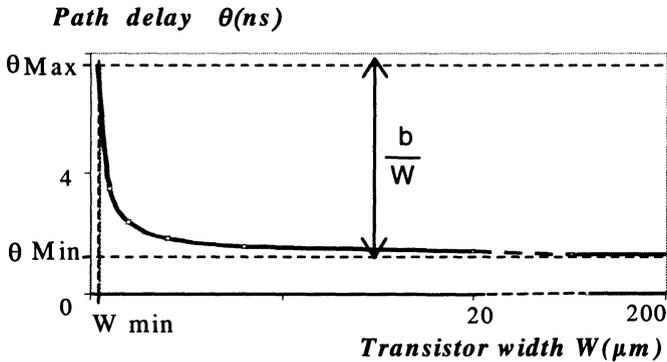


Figure 1. Path delay sensitivity to the transistor sizing.

We intend to show in this paper that the bounds of the delay can be easily determined and used to satisfy the delay constraint on critical paths, or, equivalently, to satisfy timing closure conditions at minimum iteration number.

This paper is organized as follows. In Section 2 we illustrate the path delay evolution of a standard circuit and demonstrate the existence of bounds

for delay. The upper and lower bounds of path delay are justified in Section 3. In Section 4, we present a new method for optimizing a circuit under delay constraint and we apply this new optimization method on ISCAS benchmarks. Conclusion is given in Section 5.

2. EVIDENCE OF DELAY BOUNDS

We have evaluated the delay profile of the 17000 paths of the C880 circuit (ISCAS benchmark) for different transistor sizing conditions, implemented in a $0.25\mu\text{m}$ CMOS process. The resulting path distribution obtained with a path analyzer [14] is represented in Fig.2. Each curve corresponds to a specific uniform transistor size, varying from $0.35\mu\text{m}$ to $200\mu\text{m}$. The layout of the circuit has been obtained from an automatic layout generator [15] and the timing analysis performed on a post layout extraction of the circuit. For a given transistor size (w) the right limit of each curve determines the value of the critical path.

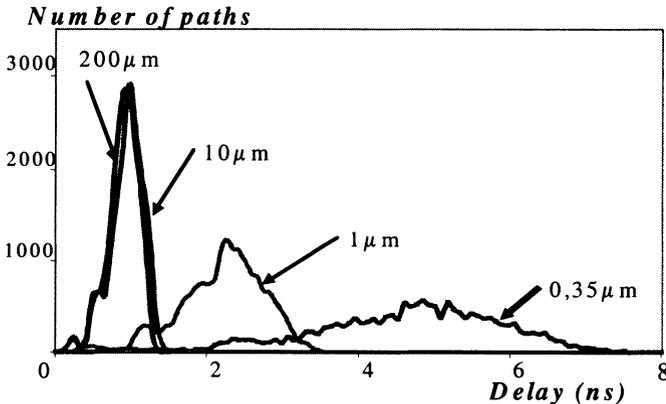


Figure 2. Distribution of delays on the paths of the C880 circuit implemented in a $0.25\mu\text{m}$ CMOS process.

As expected from Fig.1 the distribution corresponding to the minimum transistor width results in the worse path delay with a widely spread distribution. The distribution obtained with very large transistor widths (nearly infinite for the process under consideration) identifies the minimum achievable delay on the circuit critical path. As predicted from Fig.1 the difference of distribution for large transistor sizing alternatives (10 and $200\mu\text{m}$) is very small. The width of the distribution is also much narrower than

for the minimum size solution. This is a direct indication of the selected alternative for the path logical mapping.

Let us now only consider the parasitic diffusion capacitance. We obtain the delay distribution given in Fig.3 where all the distributions corresponding to different sizing conditions are superimposed. The weak difference of distribution (insert of Fig.3) will be identified later as due to the constant parasitic diffusion capacitance. We can observe, too, that the $200\mu\text{m}$ sized distribution of Fig.2 is identical to that of Fig.3. The difference in distributions with smaller sizing is a direct illustration of the interconnect limited character of this minimum sized transistor implementation. This clearly shows that depending on the selected sizing alternative, the robustness of the circuit to design or process parameters may be exhausted at the expense (see Fig.1) of area and power consumption.

The distribution shown in Fig.4 has been obtained on the same circuit for two extreme sizing alternatives ($200\mu\text{m}$ and $0.35\mu\text{m}$) reducing the effect of the circuit parasitic capacitance to the parasitic diffusion part proportional to the transistor width. As shown, the two distributions are rigorously identical and directly give the limit of minimum delay feasible on this circuit.

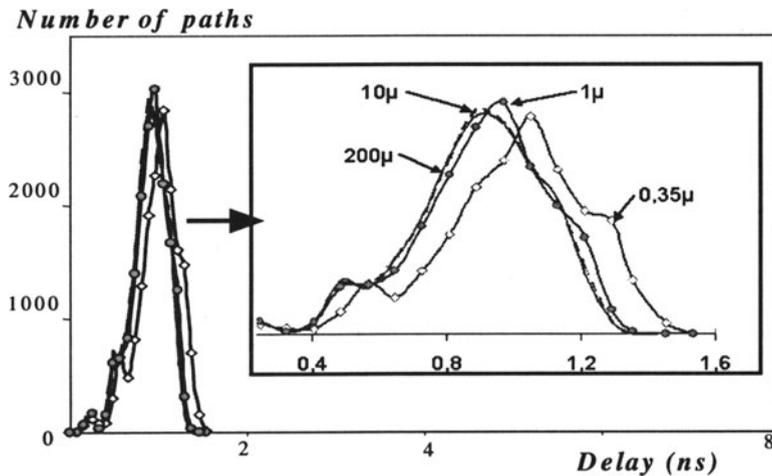


Figure 3. Distribution of delays on the paths of the C880 circuit implemented in a $0.25\mu\text{m}$ CMOS process, considering only the diffusion parasitic capacitance.

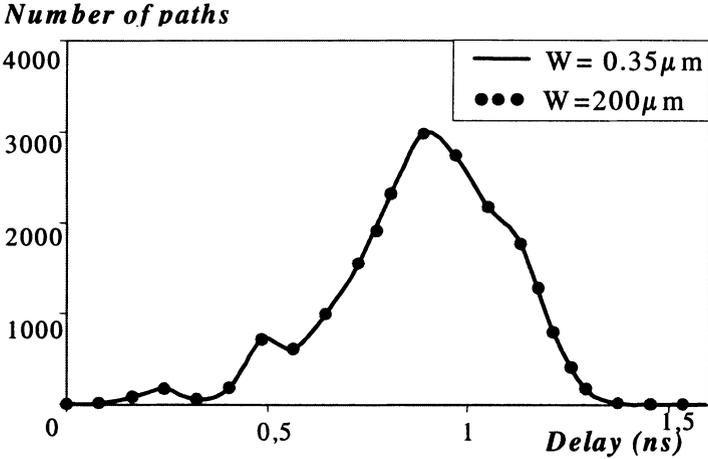


Figure 4. Distribution of delays on the paths of the C880 circuit implemented in a 0.25μm CMOS process, considering only the transistor width dependent diffusion parasitic capacitance.

These figures illustrate an easy way to determine the reasonable bounds of delay in a circuit. It is just necessary to perform a delay profiling of this circuit, with minimum sized transistors, considering the complete parasitic capacitance or its value reduced to the transistor width-dependent diffusion part. If the full parasitic content is only known after the final circuit place and route, the reduced evaluation can be predicted much sooner. This supplies a very instructive indication on the efficiency of an implementation with respect to the imposed delay constraint. Or, alternatively, it ascertains the feasibility of the constraint that is of prime importance in evaluating architectural alternatives.

3. MODELING

As previously discussed, the path distributions in Fig.2-4 have been obtained from a timing evaluation based on a design oriented modeling of delays [16] allowing a good understanding of the performance sensitivity to the design and environmental parameters. In this representation the delay of the gate (i) is evaluated from

$$t_{HL,LH}(i) = A\tau_{IN} + t_{HLS,LHS}(i)(1 - Cor) \tag{1}$$

where τ_{IN} is the input control ramp duration, $t_{HL,LHS}$ the fall or rise step response of the switching gate, Cor a correcting term associated to the carrier

speed saturation induced non linear variation of submicron process and A, a process dependent constant term [16].

τ_{IN} is directly associated to the step response of the controlling structure [16]. As a result, the switching delay of a gate can be expressed as a linear combination of step responses of the controlling and switching structures. Each step response can be evaluated from the ratio:

$$t_{HLS,LHS} = \frac{C_{Load} \cdot \Delta V}{I} \quad (2)$$

where ΔV is the output voltage variation used to evaluate the step response and I the maximum current available in the structure to charge or discharge the output.

Applied to an inverter this results in:

$$t_{HLS,LHS} (INV) = \tau_{ST} \frac{C_{Load}}{C_{IN}} \quad (3)$$

where τ_{ST} represents a metric for the process maximum speed, defined as the step response of an ideal inverter constituted of identically sized transistors, loaded by an identical one, and C_{Load}/C_{IN} is the fan out factor.

Extension to general inverter configurations and gates can be easily obtained by correcting eq.3 with a factor $DW_{HL,LH}$ representing the ratio of current available, for each considered edge, between a gate and an inverter with identically sized transistors

This results in:

$$t_{HLS,LHS} = DW_{HL,LH} \cdot \tau_{ST} \frac{C_{Load}}{C_{IN}} \quad (4)$$

where $DW_{HL,LH}$ introduced by [17] is a “digital weight” that allows to treat any structure of gate as an equivalent inverter. Its value has been explicitly defined in [18].

This demonstrates that with a very good approximation the delay on a path can be decomposed as a weighted sum of individual products: digital weight . fanout factor (DW .Fout).

Let us now consider the content of the terms in eq.4. C_{IN} , the gate input capacitance, is directly proportional to the transistor width. C_{Load} is generally constituted of active and parasitic capacitance components. The active capacitance defines the logic fan out (F_{Logic}) of the switching gate (i). It is given by the sum of the input gate capacitance (i +1) connected to the

output of gate (i), including the path and branch loading. The parasitic capacitance is constituted of the diffusion capacitance associated to the drains of the gate output (i) and of the interconnect metal capacitance. From direct inspection of a gate layout, the diffusion capacitance can be obtained from :

$$C_{\text{Diff}} = d1.W + d2 \quad (5)$$

where d1 and d2 are constants defined from the layout style and the design rules, and w is the transistor width of gate (i). The interconnect capacitance, known after the circuit place and route, is transistor width independent. Eq.5 can then be developed as:

$$DW(i).F_{\text{out}}(i) = DW(i) \cdot \frac{\sum_j C_{\text{IN}j} + C_{\text{Diff}}(i) + C_{\text{intercon}}(i,j)}{C_{\text{IN}}(i)} \quad (6)$$

$$= DW.F_{\text{Logic}} + a + \frac{b}{W} \quad (7)$$

The total path delay is then obtained from the contribution of all the nodes (eq.4,7). This explains the evolution of the distribution given in Fig.2-4: for minimum sized transistor the first and third terms of eq.7 have a maximum value, this defines the value of the upper bound of delay. For large transistor sizes the third term is negligible. So F_{Logic} resumes to the average number of logical fan out and defines, with the transistor width dependent parasitic capacitance ("a" term), the minimum value of delay achievable on the path. Note as illustrated in Fig.3 and 4 that this minimum can also be reached considering only the "a" term for an implementation with equally sized transistors. This is illustrated in Fig.1 where the difference between upper and lower bounds for the delay is given by the parasitic contribution independent of the transistor width (b/W term).

4. SIZING METHODOLOGY

Any critical path can be characterised by delay bounds. Comparing the delay constraint imposed on the path to these limits gives indication of:

- the feasibility of the delay constraint,
- the cost of this constraint in terms of area (transistor width).

This can be a significant help in defining the average fan out factor to be imposed on the path nodes to satisfy timing constraints. This is illustrated in

Fig.5 where we plot for the C 880 circuit the variation of the delay with the average value of the product : "digital weight x fanout factor" of the path ($DW.F_{out}$).

The delay value exhibits a linear variation between the min-max values previously defined. Indication is also given of the corresponding average transistor sizes for which, as illustrated in Fig.1, the delay is non linear. This gives the way to satisfy timing constraint on the path. To each constraint value corresponds an average value of the product $DW.F_{out}$ which determines the sizing conditions of the different gates of the path.

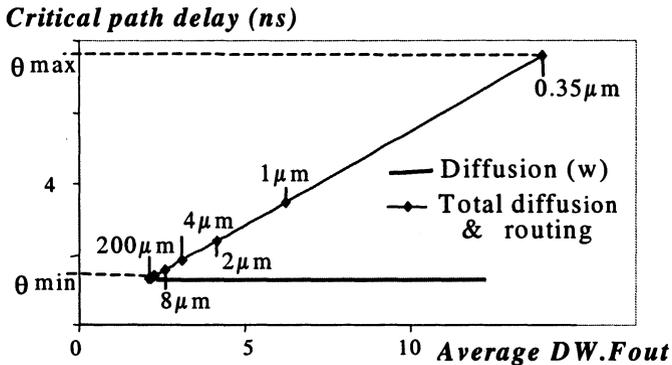


Figure 5. Illustration of the sensitivity of the critical path delay to the average value of the product "digital weight . fanout factor" of each gate on the path ($DW.F_{out}$).

This fan out factor can be determined as follows. Let us specify by $\theta_{max,min}$, the preceding bounds for delays and by $F_{out,max,min}$ the corresponding fan out factors. From the preceding discussion the limits of delay variation shown in Fig.5 satisfy:

$$\begin{aligned}
 \theta_{max} &= DW .(cF_{out max} + d) \\
 \theta_{min} &= DW .(cF_{out min} + d) \\
 \theta_c &= DW .(cF_{out} + d)
 \end{aligned} \tag{8}$$

where θ_c is the delay constraint to be satisfied on each node of the circuit, DW and F_{out} is the fan out factor value to be imposed in order to satisfy the constraint. Solving for the product $DW.F_{out}$ gives:

$$DW.F_{out} = \frac{1}{\Delta\theta} [\theta_c \Delta F + \theta_{max} F_{out min} - \theta_{min} F_{out max}] \tag{9}$$

where $\Delta\theta$ and ΔF represent the $\theta_{\max}-\theta_{\min}$ and $F_{\max}-F_{\min}$ differences as previously defined. c and d are the coefficients of the width-dependent and constant term of the parasitic contribution to the delay as given from eq.7 and illustrated in Fig.5.

In Table 1 we summarise the values of these parameters for different ISCAS benchmarks. $\theta_{\min,\max}$ correspond to the previously defined bounds for delay evaluated on the post lay out description of the critical path of the circuits implemented in a $0.25\mu\text{m}$ process. $F_{\text{out}\min,\max}$ are the corresponding average fan out factors. The last column gives the number of gates on the considered critical path.

Considering a delay constraint θ_c on these paths, it is then easy, using eq. 7 and 9, to determine the value of the average $DW.F_{\text{out}}$ factor to be imposed on the path.

Table 2 summarizes the results obtained on the preceding circuits, considering for simplicity equally sized transistors.

	θ_{\min} (ns)	θ_{\max} (ns)	$F_{\text{out}\min}$	$F_{\text{out}\max}$	Gate nb.
C1908	1.9	12.3	2.14	16.5	41
adder 2x16	3.6	17.5	1.7	8.8	100
C499	1.3	7.9	2.1	14.7	29
C1355	1.4	10	2.3	18.5	28
FPD	0.44	1.82	1.36	6.32	14
C880	1.3	7.6	2.1	14	29

Table 1. Example of values of delay constraints and fan out parameters for different ISCAS benchmarks.

	θ_c (ns)	$DW.F_{\text{out}c}$	W_c (μm)	θ_{sim} (ns)
C1908	3	3.7	3.1	2.97
Adder 2x16	6	2.9	2	5.73
C499	2	3.44	3.2	1.96
C1355	2.5	4.37	2.7	2.4
FPD	0.8	2.64	1.35	0.75
C880	2	3.42	3.18	1.96

Table 2. Delay values simulated on paths sized under the θ_c delay constraint.

Considering aggressive values for the delay constraint (θ_c) (of the order of two times the minimum available delay) we deduced the values of the

loading factor (F_{outc}) and transistor width (W_c) to be imposed on the different paths. The last column gives the value of the delay (θ_{sim}) simulated on the critical path after sizing all the transistors at the width defined from the delay constraint. As shown the resulting θ_{sim} values are in good agreement with the θ_c imposed constraint, demonstrating the validity of this approach.

We have to note that if the column W_c of Table 2 gives the average transistor sizing value. However eq.9 supplies an easy way to obtain a specific sizing of each gate that is digital weight dependent.

5. CONCLUSION

Minimizing the number of iterations in satisfying timing constraints imposes as well a good prediction of the place and route interconnect capacitance than a good knowledge of the feasibility of the constraint imposed on the different circuit parts. We have first defined and determined the upper and lower bounds of delay on combinatorial paths. Then we characterized these bounds from which min max average values of the product (digital weight.loading factor) have been deduced. To each delay constraint we have associated an average value of the digital weight.loading factor product that we defined. Validations on ISCAS benchmarks ascertain the validity of the proposed methodology. Work in development considers gate level distribution of the delay constraint allowing specific gate sizing.

6. REFERENCES

- [1] S.S. Sapatnekar, W.Chuang, "Power vs Delay in Gate Sizing: Conflicting Objectives", Proceeding of the IEEE/ACM ICCAD, pp. 463-466, 1995.
- [2] H.C. Chen, D.H.C. Du and L.R. Liu, "Critical Path Selection for Performance Optimization", IEEE trans. On CAD of Integrated Circuits and Systems, vol. 12, n°2, pp. 185-195, February 1995
- [3] N. Azemard, M. Aline, D. Auvergne, "Local Gate Resizing for Critical Path Optimization" DCIS'99, Palma de Majorqua, Espagne, 16-19 November 1999.
- [4] E. Macii, M. Pedram, F. Somenzi, " High Level Power Modeling , Estimation and Optimization", IEEE trans. on CAD of Integrated Circuits and Systems, vol.17, n°11 pp. 1061-1079, Nov. 1998.
- [5] P. Rezvani, A.H. Ajami, M. Pedram, H. Savoj, "Fanout Optimization using a gain-based Delay model", In IWLS, Granlibakken, USA, 1999.
- [6] P.M.Vaidya, "A new algorithm for minimizing Convex Functions Over Convex sets", proceedings of IEEE Foundations of Computer Science, pp. 332-337, Oct 1989

- [7] M. Berkelaar, P. Buurman, J. Jess, "Computing the Entire Active Area/Power Consumption versus Delay Tradeoff Curve for Gate Sizing with a Piecewise Linear Simulator", IEEE trans. on CAD of Integrated Circuits and Systems, vol.15, n°11 pp. 1424-1434, Nov. 1996.
- [8] I.E. Sutherland, R.F. Sproull, "Designing for Speed on the Back of an Envelops", Advanced Research in VLSI, Univ. Of Calif., Santa Cruz, 1991.
- [9] S.S. Sapatnekar, V.B. Rao, P.M. Vaidya, S.M. Kang, "An exact solution to the transistor sizing problem for CMOS circuits using convex functions", IEEE trans. CAD, pp. 1621-1634, Nov. 1993.
- [10] R. Murgai, "On the Global Fanout Optimization Problem", In IWLS, Granlibakken, USA, 1999.
- [11] C.L. Berman, J.L. Carter, K.F. Day, "The Fanout Problem: From Theory to Practice", In C.L. Seitz editor, Advanced Research in VLSI: Proceedings of the 1989 Decennial Caltech Conferences, pp. 69-99, MIT Press, March 1989.
- [12] D. Singh, J.M. Rabaey, M. Pedram, F. Catthoor, S. Rajgopal, N. Sehgal, T.J. Mozdzen, "Power Conscious CAD tools and Methodologies: a Perspective", Proc. IEEE, vol.83, n°4, pp.570-593, April 1995.
- [13] S.Turgis, D. Auvergne, "A Novel Macromodel for power estimation in CMOS structures", IEEE trans. On CAD of Integrated Circuits and Systems, vol. 17, n°11, pp. 1090-1098, November 1998.
- [14] S.Cremoux, N.Azémard, D.Auvergne, "Path Selection based on Incremental Technique", Mixed Design of Integrated Circuits and Systems, Kluwer Academic Publishers, 1998.
- [15] JF. Moraes, M. Robert, D. Auvergne, "A virtual CMOS library approach for fast layout synthesis", VLSI'99, Lisboa, Portugal, pp. 415-426, December 1999.
- [16] 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, January 1999.
- [17] I. Sutherland, B. Sproull, D. Harris, "Logical Effort: Designing Fast CMOS Circuits", Morgan Kaufmann Publishers, INC., San Francisco, California, 1999.
- [18] P. Maurine, M. Rezzoug, D. Auvergne " Output transition time modeling of CMOS structures" pp. V-363-V-366, ISCAS 01, Sydney, Australia