INTERNAL POWER MODELLING AND MINIMIZATION IN CMOS INVERTERS

S.Turgis, J.M. Daga, J.M.Portal, D.Auvergne

LIRMM UMR CNRS 5506 Un de Montpellier II 161 Rue ADA 34392 Montpellier FRANCE

Abstract

We present in this paper an alternative for the internal (short-circuit and overshoot) power dissipation estimation of CMOS structures. Using a first order macro-modelling, we consider submicronic additionnal effects such as: input slew dependency of short-circuit currents and input-to-output coupling. Considering an equivalent capacitance concept we directly compare the different power components. Validations are presented by comparing simulated values (HSPICE level 6, foundry model 0.7µm) to calculated ones. Application to buffer design enlightens the importance of the internal power component and clearly shows that common sizing alternatives for powerand delay minimization can be considered.

I- Introduction

Usual trade-offs for low power design interchange speed with power by reducing the average active capacitance necessary to propagate signals. However carefull study of power dissipation in CMOS structures gives evidence of the importance of internal components which can have higher contributions than the external capacitive one.

In order for designers to effectively reduce power dissipation, an analytical but accurate power dissipation model, considering the complete contribution of power components, is needed. This model must be design oriented to allow the estimation of power consumption for evaluating power induced design alternatives or identifying the specific parts that need to be optimized for low power. Usually three terms of power are distinguished in CMOS integrated circuits.

- A dynamic component due to charge and discharge of the different capacitances involved in the design: $P_d=\eta f V_{dd}^2 C_L$, where C_L includes active diffusion and interconnect capacitances.
- An internal component usually called the short-circuit component, that appears when both N and P blocks are conducting resulting in a direct path from supply to ground.
- A static component due to leakage currents usually neglected as long as $V_{dd} > V_{tn}+|V_{tp}|$.

First term is dominant and is mostly addressed by all authors when considering design techniques for low power [1,2,3]. Internal transistor power dissipation can also significantly contributes to the overall dynamic dissipation in certain design conditions as it has been shown by some authors[4,5]. If the purely capacitive dynamic component can easily be estimated, the internal one meets difficulties for an accurate evaluation in the sense that it strongly depends on the gate design [4]. Many techniques have been proposed to estimate this component using simulation [11,12] as well as closed form models [4,5,6,9,10] that accuracy depends on the considered design parameters and control signals. With the evolution of the submicronic range, it appears further impossible to use simple Shockley MOS models to reproduce the voltage-current characteristics of recent short-channel MOSFET's. Since CMOS circuits do not dissipate power if they are not switching, the goal of this paper will be to study the different internal dissipation sources due to input transitions depending on the design parameters. Section II relates the general switching problem of CMOS structures and describes the considered operating regions that will be used to derive closed form short-circuit and overshoot power macro-models. In section III we introduce our power dissipation macro-modelling and we present the different calculation steps. The main validations about the accuracy of the proposed model are then described and discussed in section IV. In section V we apply these results for optimizing circuit design. Finally in section VI we draw a conclusion.

II- Problem description

The general inverter behavior for an output high-to-low transition is given in figure 1 (HSPICE simulation waveforms) for both voltages and currents during the switching step. Switching operations in CMOS correspond to a typical charge transfer control which complete treatment constitutes a very complex task to be analyzed[14]. Currents in 1.b correspond to the different input slope durations considered in 1.a. In that configuration the P-Channel device is the short-circuiting transistor while the N-Channel device is the discharging one (for an output low-to-high transition the PMOS becomes the charging transistor and the NMOS the short-circuiting one).

As can be observed, when the input slope duration increases the N-transistorcurrent decreases and the short-

ED&TC '97 on CD-ROM

circuit current (Ip) increases. This trend which becomes very important for slow input slew results in an effective decrease of the average current available in the cell to unload the output. When considering now the left hand side curves on 1.a & 1.b, it can be observed that in the case of fast input transition times the P transistor current is first negative while the output voltage is greater than Vdd.

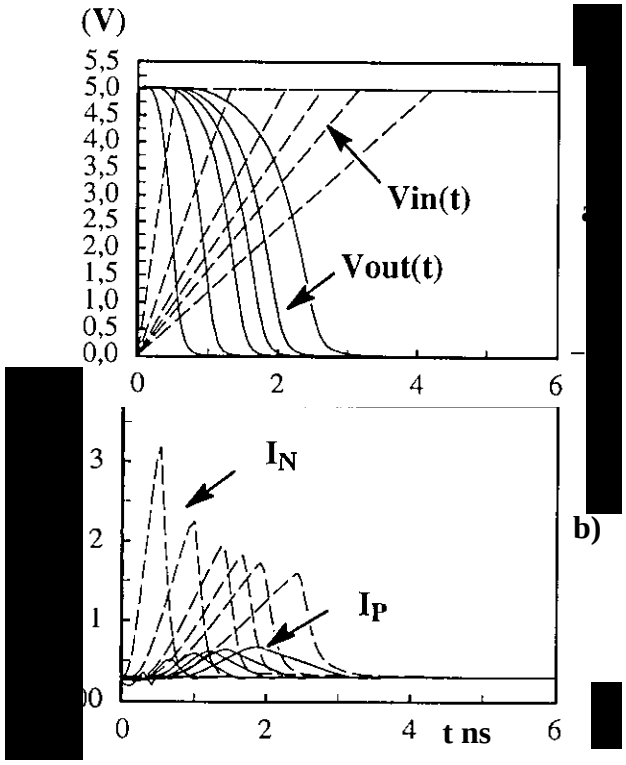

Figure 1: Input/output voltage (1.a) and current (1.b) waveforms of a symmetrical inverter ($k=W_p/W_n=1$) loaded by a capacitance Cload=10Cn=147.2fF, for different input slope durations ranging from $\tau_{in}=2t_{HLs}$ to $\tau_{in}=16t_{HLs}$ (t_{HLs} is the step response for the considered load).

This phenomenon corresponds to the overshoot produced by a direct transmission to the output of the input variation through the input-to-output coupling capacitance of the switching device. Figure 2 shows the different operating regions and inverter operating process to be considered in each region. In order to clarify the following discussion, the timing parameters are defined below:

- t_{Vtn}, $t_{(Vdd-|Vtp|)}$ are the time spent by the input slope to turn ON or OFF the N and P transistors respectively,

- t_{ov} is the time necessary to evacuate the extra output voltage due to the input-to-output coupling.

For each one of the considered region will be associated a power dissipation component as follows: the usual short-circuit considered component (internal dissipation term) will simply be split into 2 terms depending on the operating region.

- Region I: Although from t_{Vtn} both the NMOS and PMOS transistors are simultaneously conducting, no direct short-circuit current is flowing from supply to ground due to the reversed linear operating mode of the PMOS: this is the usual starting point for the short-circuit operating region that we now define as the overshoot operating region ending at t_{ov}.
- Region II: at t_{ov} the PMOS recovers the direct linear operating mode while the NMOS is still saturated and a direct path occurs between the supply rails resulting in a direct short-circuit current: this is the so called short-circuit operating region
- Region III: at $t_{(Vdd-Vtp)}$ the PMOS turns off and the only output load discharge occurs through the NMOS.

From those considerations 3 terms of dynamic power dissipation have to be discussed: a purely overshoot one: P_{ov} relative to region I, a purely short-circuit one: P_{sc} relative to region II and a purely capacitive one: P_{Cl} relative to regions II and III. Components associated to Pov and Psc are referred hereafter as internal power dissipation components.

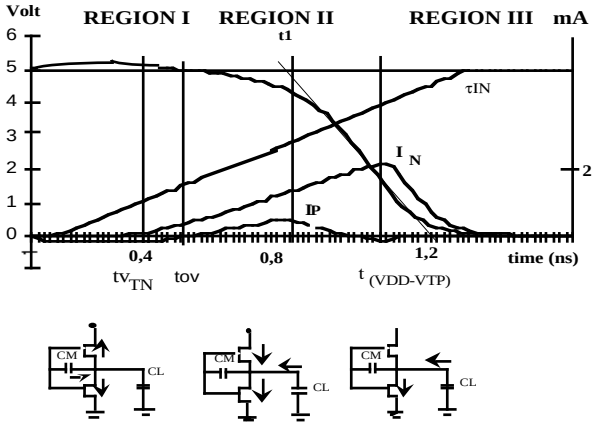

Figure 2: Illustration of the different operating regions for the output voltage and current evolution of a CMOS inverter with an internal configuration ratio k=3 loaded by Cl=2Cin=8Cn=117.76fF (the configuration ratio and load values have been chosen to exhaust the coupling and short-circuit current effects of region I and II respectively).

III-Definition of a macro-model for the internal power component

As previously discussed, internal power appears as the second significant component to be taken into account when evaluating power in CMOS ICs. In order to compare the different power components in terms of design conditions, it appears necessary to define an equivalent parameter allowing a direct investigation of the different design or mapping alternatives on the power dissipated in a

circuit. It has been previously shown [15] that alternatives for speed improvement on a circuit could be easily defined by evaluating each node in terms of load to drive capacitance ratio values (fanout factors). In the same way the comparison of power contributions could be easily obtained at circuit or design level if we were able to define for each power component an equivalent capacitance.

The macro-model used in our approach is based on the equivalent short-circuit and overshoot capacitance concept that allows us to write the internal power term in a similar way than the capacitive one through the capacitances Csc and Cov as: Psc=ηfVdd^{2}Csc and Pov=ηfVdd^{2}Cov [7]. With this definition it is then easy to compare the different power components in terms of capacitances such as:

$$P = \left(C_{ov} + C_{SC} + C_L\right) f V_{DD}^2 \quad P = \left(\frac{C_{ov} + C_{SC} + C_L}{C_{REF}}\right) C_{REF}\, f V_{DD}^2 \quad [1]$$

Where Cref is a reference capacitance used as a measurement unit for load, drive, delay and power.

Equivalent capacitance concept:

The different capacitance components are obtained from the mean accumulated charge calculation during the time intervals specified on figure 2 for the overshoot and short-circuit regions from:

$$C_{ov,SC} = \frac{1}{V_{DD}} \int_{t_{begin}}^{t_{end}} I(t)dt \qquad [2]$$

where I(t) corresponds, for example, to the current in the N and P transistors during the overshoot and short circuit regions, respectively, for an output falling edge.

In order to get quite simple expressions for both equivalent overshoot and short-circuit capacitances in terms of design parameters, the following simplifying assumptions have been considered:

- the waveshape of the short circuit current is considered symmetrical with respect to its maximum value [4],

- the slope of the input and output voltage variations are defined through the step responses of the different structures [13].

Using the same concept of delay macromodel, this step response is completely defined by the delay obtained when discharging the output load with the maximum available current. Taking account of the linear variation of saturated currents for submicronic processes ($I_N=K_N W_N(V_{GS}-V_{tn})$), the step response can be easily calculated as:

$$t_{HLS} = \frac{C_L(i)}{2K_N W_N\left(V_{DD} - V_{tn}\right)} \quad t_{LHS} = \frac{C_L(i)}{2K_P W_N\left(V_{DD} - |V_{tp}|\right)} \qquad [3]$$

where V_{DD}, V_{tn}, C_L and W_N have the usual signification of supply, threshold voltage, output load capacitance and transistor width respectively. K_N (K_P) is the drivability factor of the saturated transistor evaluated by transistor width unit, $K_N=Cox.v_S$, where Cox denotes the gate oxide capacitance and vs is the effective carrier speed limit of the process. The τ_{st} concept previously developed for micronic processes is still valid [13] and the step response for a general load can be given in terms of capacitance ratio as :

$$t_{HLS}(i) = \tau_{ST} \frac{C_L(i)}{2C_N(i)} \qquad t_{LHS}(i) = \tau_{ST} R^* \frac{C_L(i)}{2C_P(i)} \qquad [4]$$

where R* represents the equivalent ratio of current possibilities between N and P transistors.

Considering now the step responses, the input slope duration τ_{in}(i) applied to the inverter (i) is easily defined from:

$$V_{IN}(t) = \frac{V_{DD}}{\tau_{INLH}} t, \qquad V_{IN}(t) = V_{DD}\left(1 - \frac{t}{\tau_{INHL}}\right), \qquad \tau_{INLH} = 2t_{LHS}(i-1) \qquad \tau_{INHL} = 2t_{HLS}(i-1) \qquad [5]$$

for rising and falling input controlling ramps respectively.

After some tedious but not complicated calculations, the evaluation of eq.2 supplies the different expressions for the internal equivalent capacitances such as:

$$C_{ov\,LH} = \frac{K_N W_N \tau_{INLH}}{2}\left[\frac{t_{ovLH}}{\tau_{INLH}} - v_{tn}\right]^2 \qquad [6]$$

$$C_{SC_{LH}} = K_P W_p \tau_{INLH}\left[\left(1 - v_{tp}\right)^2 - \frac{\left(\frac{1}{2} - v_S + \frac{\tau_{INLH}}{4t_{HL}(i)} + \frac{t_{HL}(i)}{2t_{HLS}(i)}\right)}{\left(1 + \frac{\tau_{INLH}}{2t_{HL}(i)}\right)^2} * B\right]$$

$$B = \left[\left(\frac{3}{2} - 2v_{tp} + v_S\right) + \left(3 - 4v_{tp}\right)\frac{\tau_{INLH}}{4t_{HL}(i)} - \frac{t_{HL}(i)}{2t_{HLS}(i)}\right]$$

with equivalent expressions for falling input ramps.

Here K_N (K_P) is the drivability factor corresponding to the transistor saturated operating mode, t_{HL} is the real propagation delay defined by the time interval between input and output voltage values at half supply voltage.

Equations (6) directly illustrate the influence of the design and topological parameters on the overshoot and short-circuit power components respectively. As shown Cov is strongly input slope dependent, and Csc depends on the short-circuiting transistor sizes, the input slew and also on the ratio τ_{in}/t_{HLi} which is a good indicator of the load to drive ratio.

IV- Experimental results and discussion

The validations performed here concentrate on demonstrating the accuracy of the proposed macromodelling over a wide range of inverter configurations. Comparisons are given between simulated values of internal power and calculated ones using the proposed technique, and some other previously proposed formulas[4,5,6,9,10]. Table 1 summarizes the technology parameters for a .7µm process. K_N, K_P are directly extracted from I(V) static characteristics.

a) Power macromodelling validations

We use a validation cell constituted of an inverter (i) with identically sized N and P transistors, loaded by a capacitance C_L, and controlled by a linear voltage ramp of different durations. In table 2 we compare for this cell the simulated and calculated values of internal power, P_{int}.

PMOS	NMOS
Kp=0.038 (mA/V.µm)	K_N=0.08 (mA/V.µm)
$\|V_{tp}\|$=1V	V_{tn}=0.87V
Leff=0.65µm	Leff=0.65µm
Vs=1.5V	Vs=1.5V

Table 1: Technology parameters for the .7µm process used to validate the internal power macromodel.

The total dynamic power dissipated in inverter (i), for each input ramp condition is also given. Using the proposed macromodel it is easily calculated by adding all the capacitances as follows:
$Ptot=(Cin(i)+Cl(i)+Csc_{LH}+Csc_{HL}+Cov_{LH}+Cov_{HL}).f.Vdd^2$ where for one stage Cin(i) got to be added in order to get agreement with the simulations in which all the circuit capacitances are considered. To further compute the total dissipation on a chain, this formulae will only be used on the first gate and eq.(1) will be used for each following gate.The different dissipation terms are given in µW.

τ_{inhl}(ps)	111.2	278	444.8	556	667.2	1112
τ_{inlh}(ps)	233.5	583.8	934	1167	1401	2335
P_{sc} calc	1.12	4.03	7.77	10.55	13.5	26.4
P_{ov} calc	4.22	2.84	2.44	2.3	2.2	1.97
P_{int} calc	5.34	6.87	10.21	12.85	15.7	28.37
P_{int} sim	7.6	6.63	7.42	9.32	11.56	24.08
Ptot calc	167.3	168.8	172.1	174.7	177.6	190.3
Pt sim	163.1	162.6	164.5	166.9	169.4	182.4

Table 2: Internal power components in inverter (i) loaded by CL=10Cin (t_{HLS}=278ps) for different input slope durations, Pint sim, Pt sim are obtained from HSPICE level 6 simulations, Psc, Pov and Ptot correspond to the components calculated using eq. 1,6.

As it can be observed we obtain a quite good agreement between simulated and calculated values, all the most as validations were performed on a wide range of input control values.We can notice here that internal power dissipation contribution to the total power dissipation increases with the driver fanout increase. For relatively small values of this ratio, the overshoot power dissipation is dominant to compare to the short-circuit one.

b) Comparison with previously published works

In most of the published works on power dissipation the zero load short-circuit formulae from Veendrick [4] has been considered. This first proposed model to compute internal power does not take account of the load. When considering our macro-modelling, it has been clearly shown that this component was load dependent and hence results obtained using a zero-load model should overestimate the short-circuit component for a real structure. A more complete validation is given in table 3 where comparisons are made between different previously proposed formulas [4],[5],[9],[10] and our macromodel.

Equation 10 of ref.[4] denotes the zero-load formulae proposed by Veendrick: it uses the simple Schichman & Hodges MOSFET model. Equations 1 and 2 of ref.[5] are the formulas proposed by Hedenstierna and Jeppson and use also a Schichman and Hodges model. Equation 11 of ref.[9] presented by Sakurai and Newton is based on the α-power model and uses the zero-load assumption. Finally, equation 12 of ref.[10] is given by Vemuru and Scheinberg using the α-power model.

					P_{SC} µW			
C_L pf	τ_{in} ns	W/L	Spice Simu.	This Work eq.1, 6	Calc. eq.10 ref [4]	Calc. eq.1, 2 ref [5]	Calc. eq.11 ref [9]	Calc. eq.12 ref [10]
0.15	5	4	47.19	42.2	84.70	68.65	222.8	65.1
0.15	2	4	13.06	13.9	33.88	20	89.13	14.6
0.30	5	4	40	36.3	84.7	54.44	42.87	42.9
0.15	5	8	118.8	93.3	169.4	164.8	180.5	150
0.30	2	8	27.01	27.9	67.76	39.99	29.77	29.8

Table 3: Comparison of short-circuit power dissipation formulae with Spice simulation results for different load and input slope conditions ($\tau in_{HL}=\tau in_{LH}$, L=O.8µm, $W_P=W_N$).

As shown in this table the proposed (eq.1,6 including both overshoot and short-circuit components) macromodel is still accurate on a wide range of considered load and input slew. Moreover it is still accurate with respect to Spice simulation results while in some cases (lines 1 and 5), all other proposed formulas overestimate the internal component.

V APPLICATION TO POWER MINIMIZATION

As illustrated in table 2, if it is possible to detail the relative importance of the internal power dissipation

components, the resulting information is hard to be used in terms of design.

Usual design alternatives are chosen with respect to load or fanout factors evaluated on each node. It appears then interesting to be able to analyze power contribution in terms of capacitive loading referred to the driving strength of each structure (input capacitances). This is the reason why we introduced this equivalent capacitance concept. As previously mentioned, the control of any inverter input (gate) referred herein as an active load, is power consuming. The external power dissipated in the input capacitance is imposed by the cell selection, while the internal component (overshoot and short circuit) is input slope and load dependent and can be minimized. The problem to be considered is then to find the device able to control a fixed active load with a total minimum internal power. This is one of the important problem to be solved in designing buffer arrays or optimizing combinatorial paths for power.

For that we calculated and compared internal and external power components on an array of 3 inverters (INV1-3 of fig.3). INV3 constitutes the purely active load, controlled by INV2. INV1 is a reference of constant size used to control INV2. Varying the size of INV2 from that of INV1 to that of INV3 we explore all the design space in terms of controlling slopes for INV2 and INV3.

In figure 3 we illustrate the respective normalized variations of the internal $((C_{SC}+C_{OV})_{INV2,INV3} / C_{REF})$ and external $((C_{IN2}+C_{IN3})/C_{REF})$ power components (in terms of input and equivalent capacitances) with respect to the size of INV2.

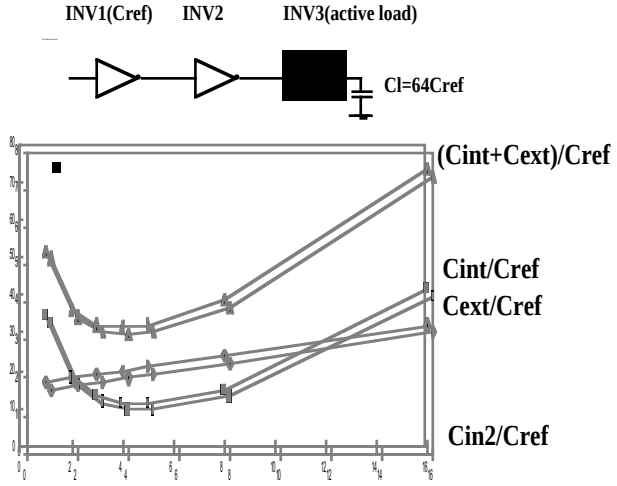

Figure 3: Illustration of the variations, with respect to the size of the drive INV2, of the internal, external and total power components necessary to drive an active load; the active load is fixed at Cin3=16Cref.

When C_{IN2} increases the corresponding external power increases monotoneously and as expected the internal component presents a minimum due to the opposite variations of overshoot and short circuit power components in INV2 and the active load (INV3). Indeed small values of C_{IN2} result in slow controlling ramps for INV3 (high value of the short circuit component) and fast controlling ramps for INV2 (high value of the overshoot component), while for large values of C_{IN2} the trend is completely reversed. The minimum observed results from a nearly balanced internal power in INV2 and its load. Note here that for each value of CIN2 the sum of the internal and external power components represents the power necessary to control the active load. This power depends on the selection of INV2 and can be minimized for appropriate values of this cell input capacitance. One can also observe that sizing INV2 with the minimum value (CREF) does not minimize the power, as it could be inferred neglecting the internal component, but results in an important increase of the power dissipation.

These results clearly show that internal power components cannot be neglected when evaluating power in a chain. Furthermore they can be minimized with an appropriate control of the load of the cells resulting in specific sizing conditions quite different from the trivial one: minimizing the power in selecting minimum sized cells.

Next point is to compare delay and power on the same array, this is given in figure 4 for the preceding array. As shown, varying Cin2 trades propagation delay between INV1 and INV2. A flat minimum of propagation delay can be observed for values of Cin2 ranging from 2 to 5 Cref.

Also given on this figure are the specific values of Cin2 corresponding to the definition of local minimum of delay and buffer insertion limit [15]. As shown these specific sizing conditions result in implementations with low internal power consumption, as it has been demonstrated before, only considering capacitive dissipation in evaluating power-delay product.

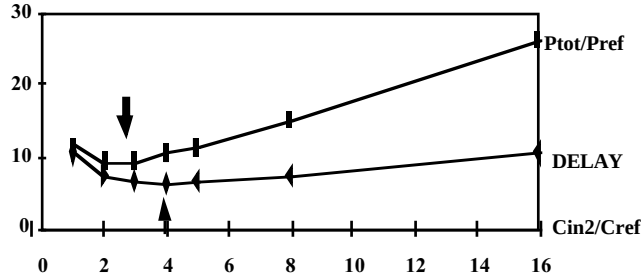

Figure 4: Illustration of the variations, with respect to the size of the drive INV2, of the total power component and propagation delay of the inverter array given in fig.3.

VI Conclusion

Based on a simple equivalent concept, we presented here a method for computing the internal power dissipation of submicronic CMOS structures. Though the closed form derived macro-models for both short-circuit and overshoot component takes more computations than the formulae used by Veendrick, it includes additional effects like input slope dependency, output load contribution, internal configuration, short-channel considerations that were formerly ignored. Moreover the proposed analytical

expressions contain all the relevant design and process parameters which can be determined on design files and foundry specifications respectively.

The equivalent capacitance concept used here allows the direct (frequency independent) comparison of the different power dissipation components in terms of fanout factors that can be evaluated at the circuit level to drive the different optimization sites of the physical design process. Validations have been performed by calculating the internal power dissipation on a wide range of inverter configurations and input control slews. Results obtained have been proved correct with respect to HSPICE simulations. Evidence of the internal power dissipation variation with the input-to-output slew ratio is given.

Applications to low power oriented optimization of buffer arrays give evidence of common delay and power optimal sizing alternatives.

Extension of this model to general gates is under progress for applications to low power oriented performance optimization of combinational paths, considering both complete delay and power components.

REFERENCES

[1] A.P. Chandrasakan, S.Sheng, R.W. Brodersen : *"Low power CMOS digital design*", IEEE J. Solid State Circuits, vol. 27, pp. 473-484, april 1992.

[2] D.Liu, C.Svensson : *"Trading speed for low power by choice of supply and threshold voltages*", IEEE J. Solid State Circuits, vol 28, pp.10-17, jan. 1993.

[3] C.Y.Tsui, M.Pedram, A.M. Despain : *"Power efficient technology decomposition and mapping under an extended power consumption model*", IEEE trans. on CAD, vol. 13, n°9, pp. 1110-1122, sept.1994.

[4] H.J.M. Veendrick : *"Short circuit power dissipation of static CMOS circuitry and its impact on the design of buffer circuits*", IEEEJ. Solid State Circuits, vol SC-19, pp. 468-473, aug.1984.

[5] N.Hedenstierna, K.Jeppson : *"Comments on 'A Module Generator for optimized CMOS Buffers',"* IEEE trans. on CAD, vol.12, pp.180-181, Jan.1993.

[6] N.Hedenstierna, K.Jeppson : *"CMOS circuit speed and buffer optimization*", IEEE trans. on CAD, vol.6, pp.270-281, mar.1987.

[7] S.Turgis, N.Azemard & D.Auvergne *:"Explicit Evaluation of Short-circuit Power Dissipation for CMOS logic Structures", proceedings of ISLPD'*95,Dana Point Resort, April 23-26 1995, pp-129-134.

[8] J-T Kong & D.Overhauser :*"Methods to improve Digital MOS Macromodel Accuracy"*,IEEE transaction on Computer-Aided-Design of Integrated Circuits & Systems,vool14,N.7, July 1995, pp-868-881.

[9] T.Sakurai & R.Newton *"Alpha-Power Law MOSFET Model and its Applications to CMOS Inverter Delay and other Formulas"*, IEEE Journal of Solid State Circuits, vol.25, N.2, April 1990, pp.584-593.

[10] S.R.Vemuru, N.Scheinberg : *"Short circuit power dissipation for CMOS logic gates*", IEEE Trans. on circuits and systems (fundamental theory and applications), vol. 41, n°11, pp 762-764, nov. 1994.

[11] S.M.Kang : *"Accurate simulation of power dissipation in VLSI circuits*", IEEE J. Solid State Circuits, vol. SC-21, pp. 889-891, oct.1986.

[12] G.Y.Yacoub, W.H.Ku : *"An enhanced technique for simulating short-circuit power dissipation*", IEEE J. Solid State Circuits, vol.24, pp. 844-847, june 1989

[13] D.Auvergne, N.Azemard, D.Deschacht, M.Robert : *"Input waveform slope effects in CMOS delays*", IEEE J. Solid State Circuits, vol. 26, n° 6, pp. 1588-1590, dec.1990.

[14] Kjell O.Jeppson : *"Modelling the influence of the transistor gain ratio and the input-to-output coupling capacitance on the CMOS inverter delay", IEEE J. of Solid State Circuits, vol.29,* n° 6, pp.646-654, june 1994.

[15] S.Turgis, N.Azemard & D.Auvergne *"Design and Selection of buffers for minimum Power-Delay Product"* proceedings of ED&TC'96, pp.224-228, March 96.