Gate Sizing for Low Power Design

Philippe Maurine, Nadine Azemard, Daniel Auvergne

Gate sizing for low power design

Philippe Maurine, Nadine Azemard, Daniel Auvergne
LIRMM, 161 Rue Ada, 34392 Montpellier, France

Abstract: Low power design based on minimal size gate implementation induces great speed penalty. We present a new gate sizing method for improving the speed performance of static logic paths designed in submicron CMOS technologies without increasing the power dissipation obtained with a minimal surface implantation. This methodology is based on the definition of local gate sizing criterion. It has been deduced from analytical models of the output transition time and of the short circuit power dissipation which are briefly introduced. Validations are given, on a 0.18 μm process using Hspice simulations(Bsim3v3 level69).

Key words: gate sizing, short circuit, power dissipation

1.   INTRODUCTION:

Lowering the power consumption under speed constraint has emerged as a critical issue for VLSI designers. This requires an accurate estimation and a very good control of the different power components. Various analytical models [1-5] and methods [6-10] allowing to handle the speed-power-surface trade-off have been developed at all level of the design flow. On the whole, these heuristics aim at reducing the external power dissipation (eq.1) resulting from the voltage variations across the capacitance of the different nodes of the circuit as

$$P_{EXT} = \eta \cdot f \cdot C_{eff} \cdot V_{DD}^2 \qquad (1)$$

where η is the activity rate, f the clock frequency, C_{eff} the effective capacitance and V_{DD} the supply voltage.

However the contribution of the power dissipation associated to the short circuit component, that may represent up to 20%-30% of the total power, is often neglected. In this work we show that considering the short circuit power component during the optimisation process may result in significant improvement of the speed and power performances of combinational paths.

To reach this goal, analytical models of the output transition time and short circuit power dissipation have been developed. They are briefly described in section 2 and 3, respectively (more detailed description can be found in [11-13]). In section 4, we deduced from these models a local sizing criterion that allows to minimize the power dissipated by a two inverter chain. Based on it, a gate sizing heuristic dedicated to the minimization of inverter tree power dissipation is developed in section 5. In section 6, some results obtained on really simple examples are presented and discussed, before to conclude in section 7.

2.   OUTPUT RAMP MODEL:

The output transition time of CMOS gates and more precisely of CMOS inverter depends strongly on its current capability. For a falling output edge, modeling the N transistor as a current generator, allows to express the inverter output transition time as the ratio between the charge to be evacuated from the output node and the maximum discharge current that can provide the N transistor as

$$\tau_{OUT} = \frac{C_L \cdot V_{DD}}{I_{MAX}} \qquad (2)$$

where C_L is the output load including parasitic capacitances. Obviously the determination of the maximum value of the discharge (charge) current in the structure is of prime importance in modeling the output transition time. This value depends of course on the N (P) transistor width, but also on the input ramp duration as illustrated in Fig.1 in which two switching characteristics corresponding to different input ramp duration domains, labelled ① and ②, are displayed.

In region 1, the set up of the current of the N transistor follows the input ramp variation to finally exhibits a constant maximum value (eq.3) during all the discharge process, this defines the fast input range.

$$I_{MAX}^{fast} = K_N \cdot W_N \cdot (V_{DD} - V_{TN}) \qquad (3)$$

In region 2, the maximum current is obtained before the input ramp reaches its maximum value, resulting in a smaller value of the charge evacuated by time unit. This defines the slow input range where the maximum value of the discharging current decreases when the input transition time increases.
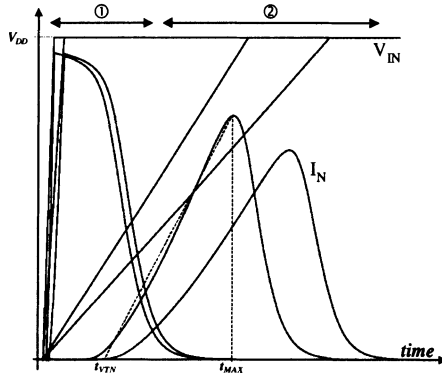

Figure 1. Sensitivity of the inverter discharging current to fast ① and slow ② input ramps.

In order to evaluate the value of I_{MAX}, we assumed that the discharge current $I_N(t)$ varies linearly from the time at which it begins to rise ($t=t_{VTN}$) and the time at which it reaches its maximum value I_{MAX} ($t=t_{MAX}$). Under this assumption we obtain

$$I_{MAX}^{Slow} = \sqrt{\frac{K_N \cdot W_N \cdot V_{DD}^2 \cdot C_L}{\tau_{IN}}} \qquad (4)$$

Combining eq.(2) with eq. (3) and (4), we finally get the output transition time expression for fast and slow input ranges as

$$\tau_{OUT}^{fast} = \tau_{ST} \cdot \frac{C_L}{C_N} = 2 \cdot T_{HLS} \qquad (5)$$

and:

$$\tau_{OUT}^{Slow} = \sqrt{\frac{C_{OX} \cdot L_{GEO}}{K_N}} \cdot \sqrt{\frac{C_L \cdot \tau_{IN}}{C_N}} \qquad (6)$$

where C_N is the thin oxide capacitance of the N transistor and T_{HLS} the time spent by the inverter to discharge the output voltage from V_{DD} to $V_{DD}/2$. τ_{ST} is a parameter characteristic of the process speed and defined from

$$\tau_{ST} = \frac{V_{DD} \cdot L_{GEO} \cdot C_{OX}}{(V_{DD} - V_{TN}) \cdot K_N} \qquad (7)$$

The extension to input falling edge can be easily obtained by exchanging p,n subscripts. Consideration of logic gates can be done replacing each gate by an equivalent inverter with identical current possibilities [11-13].

To validate the expressions (5) and (6), various comparisons between the model predictions an Hspice simulations (Most9, Bsim3v3 lvl 69) have been done on different processes ranging from $0.35\mu m$ to $0.18\mu m$. The relative observed discrepancy is always lower than 10%. As an example Fig.2 illustrates the output transition time evolution with respect to the input ramp duration value for an inverter defined by $W_N = 1\mu m$ $W_P = 2\mu m$ and $L = 0.18\mu m$.
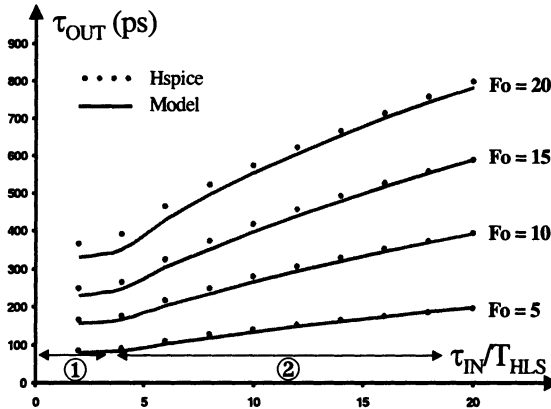

Figure 2. Output transition time values for an inverter loaded by 5 to 20 C_{IN}; ① and ② specify the fast and slow input ramp condition respectively.

As shown we obtain a very good agreement between simulated and calculated values of the output transition time.

3.   SHORT CIRCUIT MODEL:

To be useful for designers, an analytical model must allow quick and direct comparison between the various components of the power dissipation. We adopted the equivalent capacitance concept developed in [3] that represents any power contribution in the form of an equivalent capacitance. Thus, considering that the external power dissipation (the most important) is directly proportional to the load C_L on the output node we get

$$P_{EXT} = \eta \cdot f \cdot C_L \cdot V_{DD}^2 . \qquad (8)$$

We then express the short circuit component by an equivalent capacitance (C_{SC}) according to

$$P_{SC} = \eta \cdot f \cdot C_{SC} \cdot V_{DD}^2 \qquad (9)$$

where $C_{SC} \cdot V_{DD}$ is the amount of charge transferred from the supply rail to the ground during the short circuit process.

Then, assuming that the maximum short circuit current is reached while the P transistor operates in the linear mode, and that the short circuit current shape is symmetrical with respect to its maximum I_{MAX}, we can show [18-19] that the short circuit equivalent capacitance component value can be obtained from

$$C_{SC} = \frac{(1 - v_{THN} - v_{THP}) \cdot \tau_{IN}}{2 \cdot V_{DD}} \cdot \left(\psi_1 \cdot \frac{\tau_{IN}}{\tau_{OUT}} + \psi_2 \right) \cdot C_P \qquad (10)$$

where v_{THP} and V_{THN} are the normalized threshold voltages values of P and N transistors and ψ_1 ψ_2 are process parameters. These parameters must be calibrated on the process under consideration and are independent of the inverter configuration. Our approach has been validated by comparing C_{SC} calculated and simulated values on a wide design range, for various controlling and loading conditions. As illustrated in Fig.3 the relative observed discrepancy is always lower than 10%.
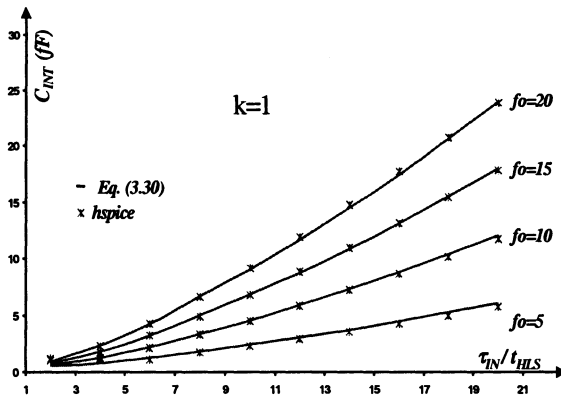
Figure 3. Comparison between simulated and calculated values of the short circuit equivalent capacitance for an inverter (W_N=1μm W_P=2μm L=0.18μm) loaded by 5 to 20 times its input gate capacitance.

4.   CRITERION FOR LOW POWER ASSIGNMENT

We want to demonstrates in this part that the minimal surface implantation does not minimize the total power dissipation but only the external power dissipation (eq.8). Let us consider the structure represented in Fig.4. The input inverter is controlled by a step input, C_{i+1} can either be a single inverter or a stack of inverters loaded by the same capacitance.
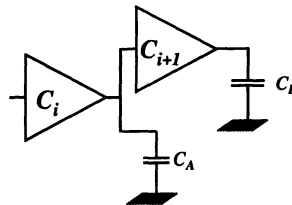

Figure 4. Inverter chain under consideration.

The existence of an optimum design that minimizes the total power dissipation can be justified by the following considerations:

–   if the inverter (i) is sized too small then it provides a really slow ramp at its output that induces an important short circuit power consumption in the next stage,

–   if (i) is chosen too large, of course the short circuit power dissipation in (i+1) is strongly reduced, but the external power dissipated in (i) is then much greater.

This clearly gives evidence of the existence of a sizing solution that results in a good trade off between short circuit and capacitive power components.

The total power dissipated by this inverter chain can be evaluated from

$$P = \eta f \left(C_i + C_{i+1} + C_{SC}^i + C_{SC}^{i+1} + C_A + C_L \right) . V_{DD}^2 \qquad (11)$$

Where C_k, C_{SC}^k are respectively the gate and the short circuit capacitances of the stage k, and C_A models the parasitic capacitance, including drain and interconnect components.

Using the output transition and short circuit power models it is easy to express the total power in terms of C_i and C_{i+1}. Cancelling the derivative of eq.11 with respect to C_i, we get a six order polynomial which can be solved numerically. Although this solution accurately predicts (<10%) the optimal sizing of the stage (i), we decided to found an approximated but analytical solution based on

$$C_{SC} = \frac{(1 - v_{THN} - v_{THP}) . \psi_1 . C_P}{2 . V_{DD}} . \frac{\tau_{IN}^2}{\tau_{OUT}} \qquad (12)$$

Cancelling the derivative of eq.11 with respect to C_i we get finally the following approximated value of optimum sizing of the stage (i):

$$C_{i\text{-}OPT}^{5/2} \approx \frac{3}{2} . A . \frac{C_{i+1}^{3/2} . (C_{i+1} + C_A)^{3/2}}{C_L^{1/2}} . \qquad (13)$$

A is a process parameter defined by

$$A = \frac{(1 - v_{THN} - v_{TP}) . (\psi_1^{HL} + R_\mu . \psi_1^{LH}) . \tau_{ST}}{2 . V_{DD} . C_{OX} . L_{GEO}} . \qquad (14)$$

In order to validate this approach, we compared the total power dissipated by the structure represented in Fig.4, for different sizing conditions, $C_i = C_{MIN}$ and $C_i = C_{i\text{-}OPT}$. Let us define by P_T^{MS} et P_T^{OS} the total power dissipated by a minimal surface implantation, and that following our proposal (eq.13). We define the gain of this sizing solution by

$$Gain = \frac{P_T^{MS} - P_T^{OS}}{P_T^{MS}} . \tag{15}$$

In Fig.5 we represent the evolution of this gain for different values of the active load W_{i+1}. As shown the improvement in power dissipation, with respect to a minimal size implementation may become significant and as large as 60% of the total dissipation for an important value of the terminal load.
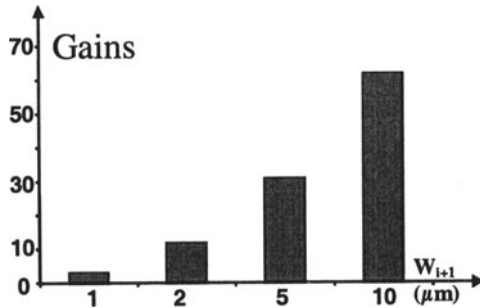

Figure 5. Gains obtained for the structure represented on the fig.4

5.   SIZING METHODOLOGY

The application of the sizing criterion (13) to an inverter tree is almost straightforward, processing backward from the output to the input of the tree. Two problems have still to be solved:

– firstly, it is necessary to determine the size of the output drivers that allows to minimize the total power on the whole tree,

– secondly, we have to manage the divergences, and more precisely to find the optimal sizing of the (i-1) controlling stage as shown in Fig.7.
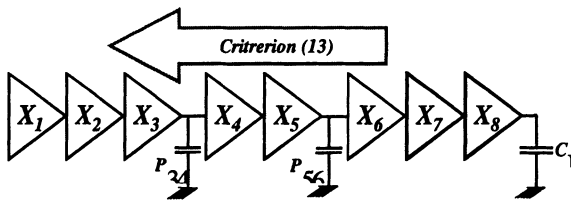

Figure 6. Illustration of how to apply the sizing criterion to an inverter chain

5.1   Output drivers:

In minimizing the power dissipated in an inverter tree, it appears that the optimal sizing of the output drivers depends strongly on the load content.

For example, in optimising the logic that drives a register or any sequential gates, we can consider that the output load is an active load or the sum of and active and passive load. Therefore, the sizing of the output driver has to be done using eq.13.

In the other hand, if the output driver controls a passive load, there is no short circuit power dissipation in the load and in this case the driver must be sized at the minimum value satisfying the delay constraint.

5.2   Divergence branches:

The case of divergence branches presents a difficulty because the sizing criterion developed in the preceding section does not allow to predict the optimal sizing of the stage (i-1).

The solution we adopted is based on the fact that the power is an additive characteristic of the structure. To justify our approach, let us consider the structure represented in Fig.7.
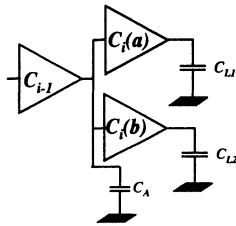

Figure 7. An example of divergence

The sizing criterion (eq.13) allows to predict the optimal value of C_{i-1}, only if $C_{L1}=C_{L2}$ in which case the two inverters can be lumped in an unique inverter with an input gate capacitance equal to $C_i(a)+ C_i(b)$. However in a general configuration C_{L1} and C_{L2} have different values.

Nevertheless, as the short circuit power dissipation is a decreasing function of C_L, we model the two inverter (a) and (b) by an unique inverter (see Fig.8) loaded by $C_L=MAX(C_{L1},C_{L2})$ to avoid any overestimation of the short circuit power dissipated by (a) and (b).
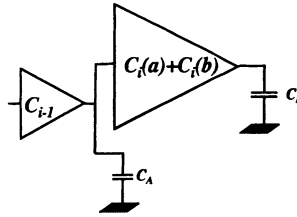
Figure 8. Equivalent structure to that of Fig.7.

6.   EXPERIMENTAL VALIDATION

This sizing heuristic based on the sizing criterion defined by eq.13 has been applied to an inverter tree represented in the Fig.9. The total power dissipated in the different implementations has been obtained from Hspice simulations.
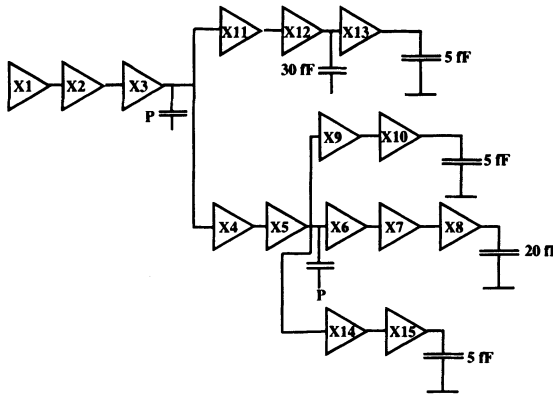

Figure 9. representation of the inverter configuration used to validate the sizing criterion (13)

In Fig.10 we compare the power gain and loss values (eq.15) obtained when comparing the sizing solution proposed to a minimal surface implementations. We considered different values the parasitic routing capacitance P to illustrate the sensitivity of the result to the parasitic content of the load.

As shown, depending on the value of P, the gains in power and speed are ranging from 3% to 15% and 13% to 45%, respectively.

The speed increase can be easily justified after a detailed analysis of the simulation results. For our example, the application the sizing criterion increases the size of the stages X12, X5 et X3. This induces a reduction of the ramp duration applied at the input of the stage X14, X11, X9, X6 et X4 reducing their switching delays.
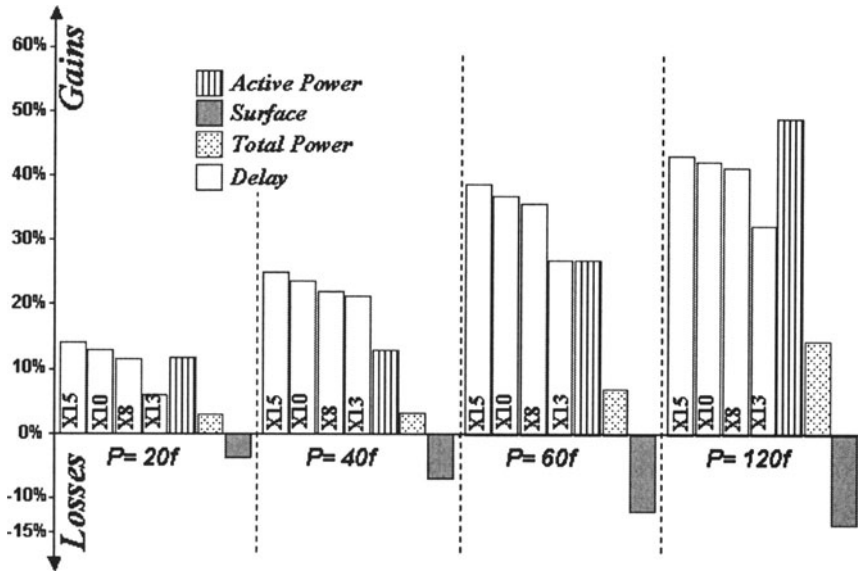
Figure 10. Gains and losses obtained for on the inverter tree plotted in Fig.9 with respect to the parasitic capacitance $P=P_{3,4}=P_{5,6}$

7. CONCLUSION

Considering the power dissipation as a critical design parameter we have presented a sizing criterion for minimising the switching power dissipation component. This has been obtained by lowering the short circuit component through a control of the gate input transition time. Using an analytical model of the short circuit power dissipation and of the output transition time we showed that a sizing condition, that minimises the short circuit component, can be defined. Application has been given to general inverter configurations in various loading conditions. Comparison to minimal size implementations clearly shows that gain in power and speed as large as 15 and 45% can be obtained.

8.   REFERENCES

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

[2] A. Hirata, H. Onodera, K. Tamaru "Estimation of Short-Circuit Power Dissipation for Static CMOS Gates" IEICE Trans. Fundamentals, vol. E79-A, N°3 March 1996

[3] S. Turgis, D. Auvergne "A novel macromodel for power estimation for CMOS structures" IEEE Trans. On CAD of integrated circuits and systems vol.17, n°11, nov.98.

[4] T. Sakurai, R. Newton "Alpha-Power Law MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas" IEEE J. of Solid State Circuits, Vol. 25, N°2, April 1990

[5] L. Bisdounis, S. Nikolaidis, O. Koufopavlou "Propagation Delay and Short Circuit Power Dissipation Modeling of the CMOS Inverter" IEEE Trans. On Circuits And Systems-I: Fund. Theory And Applications, Vol.45, N°3, March 1998

[6] M. R.C. Berkelaar, P. H. W. Buurman, J. Jess "Computing The Entire Area /Power Comsumption Versus Delay Tradeoff Curve For Gate Sizing With A Piecewise Linear Simulator" IEEE Trans. On CAD Of I.C. And Sys., Vol. 15, N° 11, Nov. 1996.

[7] S. Sapatnekar, V. B. Rao, P. Vaidya, S. M. Kang "An Exact Solution To The Transistor Sizing Problem For CMOS Circuits Using Convex Optimization" IEEE Trans. On CAD Of Integrated Circuits And Systems, Vol. 12, N° 11, Nov. 1993.

[8] M. Borah , R. Owens, M. Irwin "Transistor Sizing Power Consumption Of CMOS Circuits Under Delay Constraint" Int. Symp. On Low Power Design 95, P.167

[9] A. Chandrakasan, S. Sheng, R. Brodersen "Low-Power CMOS Digital Design" IEEE J. Of Solid State Circuits, Vol. 27, N° 4, April 1992

[10] K. Usami M. Horowitz "Clustered voltage Scaling Technique for Low-Power Design", in proc. of Int. Symp. on Low Power Design 95, pp 3-8

[11] P. Maurine , M. Rezzoug, D. Auvergne "Output transition time modeling of CMOS structures" To be published in Mai 2001 in the proc. of the IEEE Int. Symp. on Circuits And Systems, Sydney, Australia

[12] P. Maurine , M. Rezzoug, D. Auvergne "Internal power dissipation modeling and minimization for submicronic CMOS design" PATMOS'2000: Gottingen, Germany. Sept 13-15, 2000, pp.531-536

[14] J. Daga, D. Auvergne " A Comprehensive Delay Macro-Model of Submicron CMOS Logic" IEEE J. of Solid States Circuits, vol 34, n°1, pp.42-55, January 1999.