Temperature Dependence in Low Power CMOS UDSM Process

Benoit Lasbouygues, Robin M. Wilson, Philippe Maurine, Nadine Azemard, Daniel Auvergne

Temperature Dependence in Low Power CMOS UDSM Process

B. Lasbouygues[1], R. Wilson[1], P. Maurine[2], N. Azémard[2], and D. Auvergne[2]

[1] STMicroelectronics Design Department, 850 rue J. Monnet, 38926, Crolles, France
{benoit.lasbouygues,robin.wilson}@st.com
[2] LIRMM, UMR CNRS/Université de Montpellier II, (C5506),
161 rue Ada, 34392 Montpellier, France
{pmaurine,azemard,auvergne}@lirmm.fr

Abstract. In low power UDSM process the combined use of reduced value of the supply voltage and high threshold voltage value may greatly modify the temperature sensitivity of designs, which becomes structure and transition edge dependent. In this paper we propose a model for determining the temperature coefficient of CMOS structures and defining the worst Process, Voltage and Temperature condition to be verified for qualifying a design. This model is validated on two 0.13µm processes by comparing the calculated values of the temperature coefficient of the performance parameters to values deduced from electrical simulations (Eldo). Application to combinatorial path gives evidence of the occurrence of temperature inversion that is structure and control condition dependent and must carefully be considered for robust design validation.

1 Introduction

With the shrinking of technologies, the contribution of the leakage current to the total power consumption has increased dramatically, and is becoming a critical problem in CMOS UDSM technologies. In low power applications, one possible solution to reduce the static to dynamic power ratio is to use higher threshold voltage (Vt) devices. As expected this approach results in a design speed reduction. Moreover, the temperature sensitivity of the mobility and the threshold voltage [1,2] may reverse the temperature coefficient of the N and P transistor current. This completely modifies the critical characterization corners of a design (temperature inversion phenomenon). This effect is expected when the threshold voltage value approaches half Vdd. Unfortunately this configuration is becoming a standard for low power design in UDSM technologies (130nm and beyond).

In order to validate a design in a static timing analysis (STA) flow, it is necessary to validate the worst and the best case timing conditions. Without any temperature inversion phenomenon, it is well known that the worst case timings are observed for the worst case **P**rocess, the lowest **V**oltage value and the highest **T**emperature condition (PVT conditions). In low power design, due to the use of high threshold voltage devices, the worst case timing conditions are no more guaranteed at the highest temperature operating point.

E. Macii et al. (Eds.): PATMOS 2004, LNCS 3254, pp. 110–118, 2004.

Consequently to guarantee the correct behavior of a design it is necessary to verify various PVT conditions. This implies a standard cell library characterization for several (Three worst case for three different temperature values) PVT conditions, resulting in a huge amount of simulations for determining the standard tabular or polynomial representation of performance [3].

The main contribution of this paper is to propose a new representation of the timing performance of a CMOS library [4], to characterize the PVT conditions to be used for validating a design. The target of the proposed method is to deduce from a calibration of the model on one PVT condition, the evolution of each library cell performance at all the other PVT corners. This gives facility to the designer in predicting and estimating the evolution of the design performance with a minimum of simulation time.

The rest of the paper is organized as the following. In section 2, the analytical model generating the new representation of the timing performances is briefly introduced. Section 3 characterizes the temperature inversion phenomenon and defines the corresponding model. Section 4 applies this model to a data path of a circuit, designed in two different 0.13µm technologies, in order to demonstrate the influence of the process conditions to the temperature inversion phenomenon.

2 Physical Timing Model

A. Transition time modeling

Considering the transistor as a current generator [5], the output transition time of CMOS structures can be directly 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:

$$\tau_{outHL} = \frac{C_L \cdot V_{DD}}{I_{NMax}} \qquad \tau_{outLH} = \frac{C_L \cdot V_{DD}}{I_{PMax}} \tag{1}$$

where C_L represents the total output load (parasitic and active capacitance), Vdd the supply voltage value, and $I_{N/PMax}$ the maximum current available in the structure. The key point here is to evaluate this maximum current, which depends strongly on the input controlling condition. For that, two main domains have to be considered: the fast input and the slow input range.

In the fast input range, the driving condition imposes a constant and maximum current value in the structure. The current expression can then be directly obtained from the Sakurai's representation [6]:

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

Where $K_{N/P}$ is an equivalent conduction coefficient, $V_{TN/P}$ are the threshold voltages of N / P transistors, and α the velocity saturation index to be calibrated on the process.

Combining eq. 1 and 2 finally leads to the following expressions of the output transition time of a CMOS structure in the fast input range:

$$\tau_{outHL}^{Fast} = \tau_N \cdot S_{HL} \frac{C_L}{C_{IN}} \qquad \tau_{outLH}^{Fast} = \tau_P \cdot S_{LH} \frac{C_L}{C_{IN}} \tag{3}$$

where $S_{HL/LH}$ are the symmetry factors of the considered CMOS structure, C_L and C_{IN} its load and input capacitance respectively. The parameters:

$$\tau_N = \frac{Cox \cdot L \cdot V_{DD}}{K_N \cdot (V_{DD} - V_{TN})^{\alpha_N}}$$
$$\tau_P = \frac{Cox \cdot L \cdot V_{DD}}{K_P \cdot (V_{DD} - V_{TP})^{\alpha_P}} \quad (4)$$

are the process metrics for the transition time, since these parameters capture the sensitivity of the output transition time to both the supply and threshold voltage values.

B. Propagation delay modeling

The delay of a CMOS logic gate is load, gate size and input slew dependent. Extending the work of [7] by using a velocity saturation index different from unity, the input slope and the input-to-output coupling effects can be introduced in the propagation delay as:

$$t_{HL} = \frac{\tau_{in}}{\alpha_N + 1}\left(\frac{\alpha_N - 1}{2} + v_{TN}\right) + \left(1 + \frac{2C_M}{C_M + C_L}\right)\frac{\tau_{outHL}}{2} \quad (5)$$

3 Temperature Inversion: Characterization

The threshold voltage and the carrier mobility values are temperature dependent [1,2]. If both threshold voltage and carrier mobility values monotonically decrease when the temperature increases, the resulting impact on the timing performance of a design depends on the range of operating supply voltage values.

Considering (3,4), for an increase of temperature, it is obvious that a decrease of the $V_{TN/P}$ values results in a decrease of both the output transition time and the propagation delay, while a decrease of the carrier mobility induces an opposite variation of the timing performance.

It has been experimentally observed [8] that for a specific range of supply voltage value the temperature coefficient of the transition time becomes supply voltage dependent and may have negative values for $Vdd < 2V_T$. This is the temperature inversion phenomenon that could appear in low power processes, using high threshold voltage and reduced Vdd values.

Due to the resulting non-monotonic temperature impact on the design performance, it is clear, from eq. 3,5,6, that the STA flow must be performed with great care, in order to properly capture the worst performance of a design

Table 1. Threshold Voltage value at nominal conditions.

V_{DD}=1.2V	A Process	B Process
V_{TN} (V)	0.56	0.64
V_{TP}(V)	0.47	0.56

In order to illustrate this effect, let us study two different 0.13µm processes dedicated to general purpose design (A) and to low power applications (B). The only difference between these two processes is on the threshold voltage values, as given in Table 1.

A. Transistor current temperature dependence.

Let us analyze the drain source current temperature sensitivity of an NMOS transistor under different temperature conditions. In Fig 1 we report the NMOS drain source current (process B) evolution, with the Vdd supply voltage value, for three different temperature values. We consider the maximum current available by imposing V_{GS}=Vdd.

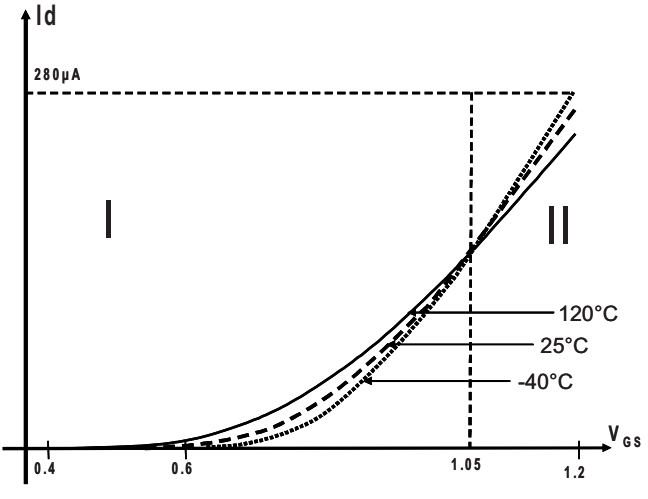

Fig. 1. Id = f (V_{GS}=Vdd) for NMOS (B process)

As shown, it is possible to define two supply voltage domains: I and II defined by the crossing points (V_{GS}=1.08V) of the three I_d versus V_{GS} curves.

In domain I, the NMOS drain source current has a greater value at high temperature. In the domain II, the situation is reversed, the greatest value of I_d occurs at low temperature (-40°C). This demonstrates that depending on the supply voltage value the worst performance operating point can be obtained for high or low temperature values. The temperature coefficient of the current is supply voltage dependent and is cancelled, in this case, for the crossing point value Vdd=1.05v.

This evolution of the worst case PVT condition from high to low temperatures can be explained by the temperature variation of the V_{TN} and K_N values. Both parameters have a negative temperature coefficient with an opposite effect on the evolution of the transistor current value. At high supply voltage value the sensitivity of the K term (2) dominates, but for low Vdd value the variation of the Vdd-VT term becomes preponderant, reversing the global I_D current temperature coefficient.

One interesting point here is to note that at a supply voltage equal to the crossing point value the effect of V_{TN} variation on the current is balanced by the effect of K_N variation, meaning that around this voltage, the performance of a design is quasi temperature independent.

As a result, the inversion voltage value of N and P transistors are the key parameters for characterizing design temperature sensitivity or predicting possible temperature inversion. As an example, in Table 2 we give the inversion voltage obtained from simulated values of the N, P transistor current, on both A and B processes. As shown, considering the voltage corners defined for B process (1.08V and 1.32V), it appears that the NMOS device is not subject to T inversion phenomenon (negative current temperature coefficient), while the PMOS is always in inversion (positive current temperature coefficient).

Table 2. Inversion voltage and nominal threshold voltages for A and B process.

	A Process		B Process	
	NMOS	**PMOS**	**NMOS**	**PMOS**
Inversion point	0.92 V	1.08 V	1.05 V	1.38 V
Vt nom.	0.56 V	0.47 V	0.64 V	0.56 V

B. Transition time temperature coefficient

In order to model and characterize a process with respect to the temperature inversion phenomenon, the δ threshold voltage and Xk conduction factor temperature coefficients must be considered [9, 10] for determining, respectively, the sensitivities of V_T and K to the temperature:

$$Vt = Vtnom - \delta \cdot (\theta - \theta nom)$$
$$K = Knom \cdot \left(\frac{\theta nom}{\theta}\right)^{Xk} \tag{6}$$

where the nominal values are chosen at 25°C and at (1.20V). Including these parameters into (2), supplies a derating coefficient *Der* that allows estimating the timing performance (4,5) of any cell operating under any PVT condition.

$$\mathrm{Der} = \frac{\tau(V_{DD}, \theta)}{\tau\mathrm{nom}} = \left(\frac{\theta}{\theta\mathrm{nom}}\right)^{Xk} \cdot \left(\frac{V_{DD}}{V_{DDnom}}\right) \cdot \left(\frac{V_{DDnom} - V_{Tnom}}{V_{DD} - V_{Tnom} + \delta \cdot (\theta - \theta\mathrm{nom})}\right)^{\alpha} \tag{7}$$

In fact, as previously discussed this temperature sensitivity is directly obtained from the $\tau_{N/P}$ parameters which capture all the sensitivity of the output transition time and propagation delay to both the supply and threshold voltage values.

In order to validate this approach, we have made an application to the PMOS and NMOS transistor of B process. We first have extracted $\tau_{N/P}$ values from Eldo simulations for all PVT corners. Then, using the determined $\tau_{N/P}$ values, at the nominal point, as a reference we have calculated the $\tau_{N,P}$ values at the different PVT corners by applying the derating coefficient (7) with δ_N=1.3mV/C and Xk_N=1.6, δ_p=1.6mV/C and Xk_p=1.2 to the τ expression (4). Results are reported in Table 3 and 4.

Table 3. τ_P for each PVT corners and errors between calculated and simulated values.

B	τ_P (ps) Simulation			Error (%) Model vs. Simulation		
	1.08V	**1.20V**	**1.32V**	**1.08V**	**1.20V**	**1.32V**
233°K	36.6	26.9	21.9	-0.7	-2.2	-6.1
298°K	33.2	26.2	21.8	0.3	0	-1.1
398°K	31.2	25.6	21.9	-3.2	-0.7	0.9

Table 4. τ_N for all PVT corners and errors between calculated and simulated values.

B	τ_N (ps) Simulation			Error (%) Model vs. Simulation		
	1.08V	**1.20V**	**1.32V**	**1.08V**	**1.20V**	**1.32V**
233°K	13.9	9.6	7.4	0.1	-4.4	-7.2
298°K	14.2	10.2	8.0	-1.0	0	0.1
398°K	14.5	11.1	9.1	-2.5	1.4	3.1

Table 5. Tau_N and Tau_P for A process.

A	Tau_P (ps) Simulation			Tau_N (ps) Simulation		
	1.08V	**1.20V**	**1.32V**	**1.08V**	**1.20V**	**1.32V**
233°K	27.3	22.0	18.6	9.5	7.6	6.5
298°K	27.6	23.1	20.0	10.9	8.2	7.0
398°K	27.9	24.1	21.4	11.1	9.3	8.2

As shown the calculated values are in good agreement with the simulated ones. The PMOS transistor of the B process is inverted over the whole supply voltage range and is temperature independent at 1.32V. At the contrary, the NMOS transistor is not impacted by the temperature inversion phenomenon as expected from Table 2 that gives its inversion voltage (1.05V) smaller than the worst-case supply voltage.

Finally, following the transition time and the delay equations (4,5), we can conclude that timing values (rising edge only) will suffer from temperature inversion. The worst case PVT must occur at the worst-case process defined by small supply voltage and low temperature values.

The main result obtained on this process, for the τ_N values sensitivity, is that the falling edge is not impacted by the temperature inversion. At 1.08V the sensitivity to the temperature is very small but the worst case operating mode corresponds to the standard definition with high temperature and low voltage values.

As a summary, for the B process, both edges vary in an opposite way. The temperature coefficient of the transition time is negative (-0.15ps/°C) for the rising edge. The falling edge exhibits a small but positive temperature coefficient (+0.04ps/°C). In that condition the critical operating mode will be defined by the rising edge.

Considering the A process with its small threshold voltage values, this technology is supposed to be less influenced by the temperature inversion phenomenon.

As shown in Table 5, the worst case for the current is clearly at high temperature and low supply voltage for both edges. The PMOS study gives an inversion point around 1.08V. At 1.08V τ_P exhibits a small variation between –40°C and 125°C (+ 0.021ps/°C). Following these results, the worst case for the drain current is at high temperature.

The main conclusion of this part is that the $\tau_{N,P}$ coefficients can be used to characterize the drain current variations taking account all the PVT parameters. We have shown that the model differentiates both edges separately to take care of the N and P MOS dissymmetry. With respect to the supply voltage value, the threshold voltage appears to be the most sensitive parameter to control the temperature coefficient. A 5% variation of V_T, can completely reverse the transistor temperature sensitivity.

4 Application to Path Timing Performance Sensitivity

In this part we study the temperature inversion phenomenon for cells and for a path designed in the two A and B technologies. We highlight the behavior of the propagation delay and the transition time.

A. Cell level

It is clear from (4) that the temperature sensitivity of the transition time (3) is completely defined from that of the $\tau_{N,P}$ parameter. Following this conclusion, only the rising edge for process B will exhibit a temperature inversion (Table 3). In Table 6 we have verified this result with Eldo for inverters implemented in A and B processes.

Table 6. Transition time at –40°C and 125°C.

Vdd 1.08V	INV Process A		INV Process B	
	Rising	Falling	Rising	Falling
-40°C	52ps	48ps	78ps	55ps
125°C	65ps	58ps	70ps	58ps

Considering the delay expression (5), two separate coefficients have to be considered. The first parameter is input slope and threshold voltage dependent. The second one is the output transition time, which is directly connected to $\tau_{N,P}$.

For the A process and the rising edge, τ_P is not in temperature inversion, so the second term of the delay is not impacted. But if we consider the first part of the formula, and the fact that the threshold voltage increases when the temperature decreases (Table 7), for a given input slope, this part of the delay has a greater contribution for low temperature and can cause temperature inversion. And the greater the input transition time value is, the more important the phenomenon is. The main result is that the delay can be in temperature inversion even if the drain current and transition time sensitivity are not modified.

Table 7. Vt for PMOS in A process.

Temp.	Vt_P(V)
-40°C	0.61
25°C	0.56
125°C	0.48

B. Path level

Validation is obtained by evaluating the temperature sensitivity of a data path designed in the two A and B technologies, and comparing the results with electrical

simulations (Eldo's). The data path is constituted of inverters, nand, nor, AO, buffer and multiplexer for a total of 10 cells and an input slope of 800ps. The characterizations are done for a worst-case process at 1.08V for –40°C, 25°C and 125°C.

Table 8. Path delay at several temperatures.

Temp. (°C)	Delay (ns) Simulation	Delay (ns) & error model
-40	2.178	2.261 (+4%)
25	2.263	2.231 (-2%)
125	2.178	2.073 (-5%)

From the simulated values, given in Table 8, we observe on this path an almost temperature independent delay, the worst case occurring around 25°C, with a very small temperature coefficient (0.01ps/°C). This is a consequence of the reverse temperature coefficient of the edges resulting in an overall compensation.

Let us focus on the first cell (inverter) of the path (controlled by the slower input ramp). We can see in Table 9 that, as generally expected, the transition time has a positive temperature coefficient. However, due to the threshold voltage temperature sensitivity (Vt,τ_{IN} ,eq.5), the delay exhibits a negative coefficient.

Table 9. Timing of the first cell, controlled by a 800ps input ramp.

Temp. (°C)	Delay (rising)	Slope (rising)
-40	378 ps	316 ps
25	362 ps	341 ps
125	333 ps	372 ps

This highlight the rule defines at the cell level study.

Considering the second process (B) with a greater threshold voltage value, on the same path (same cells, same drives), this technology appears deeper inverted. The transistor current study demonstrates that the rising edge is fully inverted (delay and transition time) and the falling edge is almost temperature independent. On a path, where falling and rising delays add up, the worst case is for a worst process, at low supply voltage and temperature values. The results are summarized in the Table 10. As shown the calculated values of delay (5) are in good agreement with the simulated one. This demonstrates the ability of the model in characterizing the temperature inversion phenomenon.

Table 10. Path delay at several temperatures.

Temp. (°C)	Delay (ns) Simulation	Delay (ns) & error model
-40	3.197	3.381 (+5.7%)
25	3.032	3.129 (+3.2%)
125	2.868	2.794 (-2.6%)

To summarize, for the B process that is fully inverted, the worst case is obtained for a PVT defined at the lowest temperature. For the A process, the delay temperature

dependence is input slope dependent. Despite the nearly temperature independence of the process, the path temperature sensitivity depends on its structure and must be carefully considered by controlling slope for defining the critical configuration.

5 Conclusion

The determination of robust PVT corners is of fundamental importance in actual low power UDSM process, for validating the critical configuration of designs. Using a physical and analytical model of the timing performance of CMOS structures, we have demonstrated that it was possible to accurately model the temperature sensitivity of the transition time and delay. We have clearly identified a process parameter, $\tau_{N,P}$, that can be used as a robust metric to define the temperature sensitivity of CMOS structures.

Validations have been given on two specific 0.13µm processes by comparing the temperature sensitivity of timing performance, calculated with the model to values obtained from electrical simulations.

We have shown the possibility on a specific process to get transition edges with opposite temperature coefficient (temperature inversion) resulting in a complete modification of the definition of the design validation critical conditions. This temperature inversion sensitivity has been shown to be input slope and path structure dependent.

This model appears as a powerful help for designers in characterizing, with few simulations, the timing performance of designs for any PVT corner.

References

[1] Changhae Park et all, "Reversal of temperature dependence of integrated circuits operating at very low voltages", Proc. IEDM conference, pp.71-74, 1995.
[2] S.M. Sze,"Physics of semiconductor devices", Wiley ed. 1983.
[3] Synopsys Inc.,"Scalable Polynomial Delay And Power Model", Rev 5, October 2002.
[4] B. Lasbouygues, J. Schindler, S. Engels, P. Maurine, N. Azemard, D. Auvergne, "Continuous representation of the performance of a CMOS library", European Solid-State Circuits, ESSIRC'03 Conf. on 16-18 Sept 2003.
[5] P. Maurine, M. Rezzoug, N. Azemard, D. Auvergne, "Transition time modeling in deep submicron CMOS" IEEE Trans. on Computer Aided Design, vol.21, n11, pp.1352-1363, nov.2002.
[6] T.Sakurai and A.R. Newton,"Alpha-power model, and its application to CMOS inverter delay and other formulas", J. Solid State Circuits vol. 25,pp.584-594, April 1990.
[7] K.O. Jeppson, "Modeling the Influence of the Transistor Gain Ratio and the Input-to-Output Coupling Capacitance on the CMOS Inverter Delay", IEEE JSSC, Vol. 29, pp. 646-654, 1994.
[8] Daga JM, Ottaviano E.; Auvergne D, "Temperature effect on delay for low voltage applications", Design, Automation and Test in Europe, Proc. , pp. 680-685, 23-26 Feb. 1998.
[9] J.A. Power et all, "An Investigation of MOSFET Statistical and Temperature Effects", Proc. IEEE 1992 Int. Conference on Microelectronic Test Structures, Vol. 5, March 1992.
[10] A. Osman et all, "An Extended Tanh Law MOSFET Model for High Temperature Circuit Simulation", IEEE JSSC, Vol. 30, No2, Feb. 1995.