Temperature Effect on Delay for Low Voltage Applications

J.M. Daga*; E. Ottaviano** and D. Auvergne
LIRMM, UMR CNRS/Un. de Montpellier II, (C55060),161 rue Ada, 34392 Montpellier, France
*Atmel-ES2 Z.I. 13106 Roussset France
**on leaves from DEIS, Un. Bologna Viale Risorgimento2, 401367 Bologna, Italy

Abstract
This paper presents one of the first analysis of the temperature dependence of CMOS integrated circuit delay at low voltage. Based on a low voltage extended Sakurai’s $\alpha$-power current law, a detail analysis of the temperature and voltage sensitivity of CMOS structure delay is given. Coupling effects between temperature and voltage are clearly demonstrated. Specific derating factors are defined for the low voltage range (1-3$V_{th}$). Experimental validations are obtained on specific ring oscillators integrated on a 0.7$\mu$m process by comparing the temperature and voltage evolution of the measured oscillation period to the calculated ones. A low temperature sensitivity operating region has been clearly identified and appears in excellent agreement with the expected calculated values.

1: Introduction
To satisfy low power dissipation constraints imposed by the exploding market of portable applications, designers explore all the useful ways in reducing energy dissipation in today technology circuits. Among the techniques generally used such as the reduction of switched capacitances, designing for low voltage appears as the most efficient way to trade speed and power. For that the $V_{DD}$ supply and the $V_{th}$ threshold voltages are mutually sized to respect a conservative value of the order of 5 for the ratio $V_{DD}/V_{th}$ [1]. However with submicronic processes the carrier operation in speed saturation limits the speed performance degradation, allowing $V_{DD}/V_{th}$ ratio values ranging between 2 and 3.

Moreover for special circuits used in medical or domestic applications, which impose very low power constraints, the use of supply voltage values not too different from $V_{th}$ is of current practice [2]. Here too the lower limit in reducing threshold values is defined from leakage power dissipation considerations [3]. So designing with very low supply voltage values (below 3$V_{th}$) is becoming more and more attractive.

If the great sensitivity of design performances to $V_{DD}$ and $V_{th}$ value fluctuations has been identified as one of the major limitations [4,5] of $V_{DD}$ scaling, few attention has been given to characterize their temperature sensitivity in the low voltage domain. If some results on simulations [12] and measurements [6] of speed and power performances of standard cell library and circuits operating at low $V_{DD}$ have been recently available, no modeling of these parameters and of their sensitivity has been proposed.

In this paper we analyze the combined effects of voltage reduction and temperature variation on speed performances. Based on an extended $\alpha$-power current representation [7], calibrated for low voltage, we propose a simple model allowing the accurate investigation of the speed performance temperature sensitivity of low voltage designs. This temperature sensitivity is experimentally investigated in the second part. The third part is devoted to the presentation of the extended current model. Application to the speed performance modeling is given in the fourth part. In part five we present validations on specific test structures implemented in a 0.7 $\mu$m process. Finally a conclusion is drawn in the last part.

2: Evidence of the temperature effect on speed performance
The complete characterization of CMOS standard cell libraries implies an accurate modeling of the effect of the supply voltage and temperature variations Usually library suppliers represent the cell performance variations to supply voltage and temperature through independent derating factors ($\text{Der}(V_{DD})$, $\text{Der}(\theta)$) which allow quick cell performance characterization for any deviation from...
well characterized nominal operating conditions. This can be described from:

\[
\text{Delay} = \text{Delay(nominal)} + \text{Der}(V_{DD}), \text{Der}(\theta) \tag{1}
\]

where Delay(nominal) = A+B.C.load, A and B coefficients being considered as delay calibration parameters, defined for nominal values of the supply voltage and the temperature. For usual supply voltage standard (V_{TH} value greater than 3 V_T0 ) this model results in a quite good accurate representation of the variations of the circuit operating conditions. In this voltage range this is due to an apparent independence of the temperature sensitivity to the supply voltage. This can be understood easily, considering that in speed saturation operating mode, the current variation is more dominated by the mobility variation than by the threshold voltage one. The relative variation of the measured oscillation period between 298°K and 398°K, for V_{TH} values ranging from 2.5v to 5v, is given in fig.1, considering specific ring oscillators implemented in a 0.7\mu m process. As shown the temperature sensitivity of the oscillation period appears quite independent of the supply voltage.

The low voltage evolution of the oscillation period in the same temperature range is given in fig. 2. This time the temperature sensitivity of the oscillation period is strongly V_{TH} dependent and exhibits an interesting reversal dependency. As discussed in [6] this evolution is the result, at low V_{TH}, of the increased current sensitivity to the threshold voltage. Both threshold voltage and carrier mobility decrease with increasing temperature. Lower threshold voltage increases the transistor current and lower mobility decreases its value. This results in an opposite variation of the current compensating and then dominating (when V_{TH} approaches V_{TH}) the mobility induced variations. These observations give a clear evidence of the coupled influence at low voltage of the temperature and V_{TH} variations on the current and speed performances of CMOS circuits. As a result, the equation 1 with independent derating factors is no more sufficient to describe these variations with reasonable accuracy. Because speed performances depend on the current available in the different structures we first introduce the current model we used to represent temperature effects at low voltage.

3: Current modeling

3.1: Submicronic modeling

Short channel effects increase considerably the complexity of models for submicronic processes. A very good synthesis together with a nice modeling has been presented in [8] where the current expression in strong inversion is given in terms of design and operating condition parameters such as:

\[
I_D = \frac{W}{R} \frac{C}{L} \frac{V_{GS} - V_{TH}}{1 + \frac{1}{2} \frac{V_{TH}}{V_{PD}}} \tag{2}
\]

where R includes the mobility degradation effects as given in [8] and the other parameters have the usual signification. Temperature sensitivity can be included easily considering mobility and threshold voltage variation such as [9,10]:

\[
\mu_\theta(\theta) = \mu_\theta(\theta_{\text{nom}}) \cdot \left( \frac{\theta_{\text{nom}}}{\theta} \right)^{\delta_\theta}
\]

\[
V_{TH}(\theta) = V_{TH}(\theta_{\text{nom}}) - \delta (\theta - \theta_{\text{nom}})
\]

where \( \theta_{\text{nom}} \) and \( \delta \) represent pseudo empirical coefficients which account for the lattice and impurity scattering effects and the temperature evolution of the intrinsic carrier concentration respectively. They are calibrated on the different processes and have values ranging between 1 - 2 and 10^{-7} - 4.10^{-3} V/\circ C for \( \theta_{\text{nom}} \) and \( \delta \) respectively.

If the accuracy of this model is satisfactory for analog applications it is still too complicated for the analytical modeling of performances of digital circuits. The simplified \( \alpha \)-power model introduced by Sakurai [7] may present a sufficient accuracy if calibrated in a supply voltage range reduced to low

**Figure 2:** Illustration of the oscillation period temperature sensitivity in the 2.5-5v range.

**Figure 3:** Illustration of the oscillation period temperature sensitivity in the 1.2-2.5v range.
voltage (1 to 3\( V_{T0} \)) and standard voltage (3 to 5\( V_{Tn} \)) domains.

3.2: Extended \( \alpha \)-power model

The simplified model we propose, extends the Sakurai’s \( \alpha \)-power law including the temperature effects in the mobility (limit speed) and the threshold voltage. As given in [7] MOS transistor current evolution in submicron process can be well represented from:

\[ I_{DS} = K \cdot W \cdot (V_{DD} - V_{T0})^\alpha \]

where \( \alpha \) and K are parameters calibrated on experimental current variations.

\( \alpha \) represents the velocity saturation index and has a value ranging from 2 (long channel) to 1 for deep saturation. K represents the process drivability factor, for \( \alpha = 1 \) (standard voltage range for a 0.7\( \mu m \) process) it can be easily shown to be: \( K = C_{ox} \cdot v_s \)

where \( v_s = \mu \cdot E_c \) is the carrier limit speed directly connected to the low field mobility [13]. For the low voltage range previously defined these parameters will be calibrated on the \( I_{DS}(V_{GS}) \) curve for \( V_{GS} \) values ranging between 1-3 \( V_{Tn} \).

Considering now that the temperature sensitivity of the current is completely defined through the threshold voltage and the mobility evolution we obtain:

\[ K(\theta) = K(\theta_{nom}) \left( \frac{\theta_{nom}}{\theta} \right)^\theta_k \]

\[ V_{T0}(\theta) = V_{T0}(\theta_{nom}) - \delta \cdot (\theta - \theta_{nom}) \]

As we will show later the definition of this temperature insensitive cross point voltage is of great importance for the speed of the structures which will exhibit the same temperature evolution. As an example let us consider a submicronic process with the following values of the parameters: \( \alpha = 1.5, \delta = 2.10^{-3}, \theta_k = 1 \) with \( \theta_{nom} = 298^\circ C \), the value of \( V_{GS} \) (\( V_{T0} \)) for which the current temperature coefficient cancels is \( V_{Tn} + 0.89 \nu \) which is nearly 2 \( V_{Tn} \).

3.3: Experimental validations

A first bench of validations has been obtained from HSPICE simulated values of the N and P transistor currents for a 0.7\( \mu m \) process using the ATMEL-ES2 foundry supplied model card (level 6). This level is sufficiently accurate to reproduce correctly the current evolution in a large voltage range including subthreshold domain for analog applications. \( \alpha, \delta, \theta_k, V_{T0} \) and K parameters have been calibrated in the low voltage domain (1 to 3 \( V_{Tn} \)), their corresponding values are given in Table 1.

<table>
<thead>
<tr>
<th>( V_{T0} ) (V)</th>
<th>K (( \mu A/\mu mV ))</th>
<th>( \alpha )</th>
<th>( \delta ) (mV/( \circ C ))</th>
<th>( \theta_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMOS</td>
<td>0.7</td>
<td>80.2</td>
<td>1.49</td>
<td>2 10^{-3}</td>
</tr>
<tr>
<td>PMOS</td>
<td>1.05</td>
<td>30.3</td>
<td>1.74</td>
<td>1.5 10^{-3}</td>
</tr>
</tbody>
</table>

Table 1: characteristic parameters of the 0.7\( \mu m \) process.

In figure 4 we compare the simulated current values (HSPICE level 6) to the calculated ones (NMOS), for different temperature conditions ranging from 298\(^\circ\)C to 398\(^\circ\)C. As observed the agreement between simulated and calculated values with the proposed model is excellent in the full voltage range, the maximum discrepancy is lower than 5%.

![Figure 4: Comparison between the calculated (\( \alpha \)-power model) and simulated (HSPICE foundry level 6) current values for the NMOS transistor. As observed in the figure the crossing point voltage value is obtained at: \( V_{GSNCROSS} = 1.3v \) which is not too different from 2 \( V_{Tn} \) as we calculated directly in the preceding part.](image-url)
We consider a simple cell such as an inverter and then extend the results to more complex cells. Delay modeling [14] can be develop into two steps: the study of the step response which constitutes the intrinsic response of the considered element and the extension to the real response to account for environmental phenomena such as input slope effects. Let us first consider the step response of a simple inverter.

4.1: Temperature effect in the step response

Defining the delay at half supply voltage, the step response is associated to the charge/discharge time of the cell output load under the maximum available current [14] resulting in:

\[ t_{HLs,LHs} = \frac{C_n \cdot L \cdot V_{DD}}{K_{NP}(\theta)(V_{DD} - V_{TNP}) \cdot V_{DD} \cdot 2CN} \]  

where \( L \) is the effective length of the switching transistor, \( C_i \) the total output load of the cell, \( C_{in} \) represent the driving input capacitance of the N,P transistor, respectively, the other parameters being defined in the preceding part.

As shown in this equation the temperature effect is easily entered through the \( K \) and \( V_{T0} \) temperature dependence previously discussed. The sensitivity analysis of this response can be studied directly from the partial derivatives of this equation with respect to these two key parameters, resulting in eq. 7 where the parameters \( \alpha, \delta \) and \( \theta \) which represent the saturation index and the temperature coefficient of the threshold voltage and the mobility, respectively, are calibrated on the process.

\[ t_{HLs,LHs} = \frac{C_n \cdot L \cdot V_{DD}}{K_{NP}(\theta)(V_{DD} - V_{TNP}) \cdot V_{DD} \cdot 2CN} \]  

where \( L \) is the effective length of the switching transistor, \( C_i \) the total output load of the cell, \( C_{in} \) represent the driving input capacitance of the N,P transistor, respectively, the other parameters being defined in the preceding part.

As shown in this equation the temperature effect is easily entered through the \( K \) and \( V_{T0} \) temperature dependence previously discussed. The sensitivity analysis of this response can be studied directly from the partial derivatives of this equation with respect to these two key parameters, resulting in eq. 7 where the parameters \( \alpha, \delta \) and \( \theta \) which represent the saturation index and the temperature coefficient of the threshold voltage and the mobility, respectively, are calibrated on the process.

\[ t_{HLs,LHs} = \frac{C_n \cdot L \cdot V_{DD}}{K_{NP}(\theta)(V_{DD} - V_{TNP}) \cdot V_{DD} \cdot 2CN} \]  

where \( L \) is the effective length of the switching transistor, \( C_i \) the total output load of the cell, \( C_{in} \) represent the driving input capacitance of the N,P transistor, respectively, the other parameters being defined in the preceding part.

As shown in this equation the temperature effect is easily entered through the \( K \) and \( V_{T0} \) temperature dependence previously discussed. The sensitivity analysis of this response can be studied directly from the partial derivatives of this equation with respect to these two key parameters, resulting in eq. 7 where the parameters \( \alpha, \delta \) and \( \theta \) which represent the saturation index and the temperature coefficient of the threshold voltage and the mobility, respectively, are calibrated on the process.

\[ t_{HLs,LHs} = \frac{C_n \cdot L \cdot V_{DD}}{K_{NP}(\theta)(V_{DD} - V_{TNP}) \cdot V_{DD} \cdot 2CN} \]  

where \( L \) is the effective length of the switching transistor, \( C_i \) the total output load of the cell, \( C_{in} \) represent the driving input capacitance of the N,P transistor, respectively, the other parameters being defined in the preceding part.

As shown in this equation the temperature effect is easily entered through the \( K \) and \( V_{T0} \) temperature dependence previously discussed. The sensitivity analysis of this response can be studied directly from the partial derivatives of this equation with respect to these two key parameters, resulting in eq. 7 where the parameters \( \alpha, \delta \) and \( \theta \) which represent the saturation index and the temperature coefficient of the threshold voltage and the mobility, respectively, are calibrated on the process.
Using the derating coefficients previously defined (eq.9) the inclusion of temperature and voltage effects in the real response (eq.10) is straightforward, we obtain:

$$t_{V,T}^{HL,LS} = [1 - R] \frac{A_N}{A_{Nom}} \frac{D_{RLS,LS}}{D_{RLS,LOS}} + RD_{RLS,LOS}$$ [11]

where $R = \frac{t_{V,T}^{HL,LS,LOS}}{t_{V,T}^{HL,LS,LOM}}$ is the ratio between the structure step and real responses for nominal operating conditions.

It can easily be verified that the value of this coefficient belongs to the 0 - 1 interval, the value 1 corresponding to a pure step response and 0 to a theoretical infinite contribution of the input ramp. From this equation it is possible to extract a simple evolution criterion for the real response: the derating factor belongs always to the interval $A_{DerNLOM} = \frac{A_N}{A_{Nom}} \frac{D_{RLS,LOM}}{D_{RLS,LS}}$ for the output falling edge and $A_{DerPLOM} = \frac{A_P}{A_{Nom}} \frac{D_{RLS,LOM}}{D_{RLS,LS}}$ for the rising one.

Considering now the delay performance of a full circuit as a sum of real rise and fall delay times it appears then possible to determine a derating factor for the circuit delay performance, around the nominal operating conditions. The total delay derating factor value of the complete circuit will always belong to the interval defined by:

$$A_{Der} = \min \left( \frac{A_N}{A_{Nom}} \frac{D_{RLS,LOM}}{D_{RLS,LS}}, \frac{A_P}{A_{Nom}} \frac{D_{RLS,LOM}}{D_{RLS,LS}} \right)$$

and

$$A_{Der} = \max \left( \frac{A_N}{A_{Nom}} \frac{D_{RLS,LOM}}{D_{RLS,LS}}, \frac{A_P}{A_{Nom}} \frac{D_{RLS,LOM}}{D_{RLS,LS}} \right)$$

As a result, the complete circuit delay performance will obey the law:

$$t_{total}(V_{DD}, T) = \left[ (1 - b)A_{Der} \text{ min} + bA_{Der} \text{ max} \right] t_{total_{Nom}}$$ [12]

where the b coefficient (values ranging between 0-1) characterizes the quality of the design in terms of performance dominated by N or P transistors as it can be found in specific circuits such as SRAM (N dominated) [6].

These equations allow not only an accurate determination at the cell level of the evolution of the delay performance but give also a good indication of the evolution interval of a complete circuit implemented in the considered process.

5: Experimental validations

Validation of these results has been obtained on a set of 4 ring oscillators constituted of arrays of 109 inverter stages with different configuration ratios ($W_P/W_N = 1$ for oscinv1, $W_P/W_N = 3$ for oscinv2), arrays of 54, 4 input AND, OR (oscand4, oscor4). We select these different configurations in order to exhaust the expected N or P dominated sensitivity. In fig.7 the derating factors measured on the 4 ring oscillators at 298°K has been plotted for $V_{DD}$ values ranging from 2.5 to 1.2v. Has expected, each ring oscillator has it's low voltage behavior included in the calculated min-max interval.

The comparison between the calculated and the measured derating factor on oscinv2 in the same range of $V_{DD}$ values, at 298°K and 398°K is given in fig.8. As we can observe on this plot the agreement is very good and confirms the validity of the proposed approach in modeling the supply voltage and temperature effects on performances in the low voltage range.

To conclude on these results we represent in fig.9 the evolution of both $V_{DD}$ derating (298°K) and $\theta_s$ sensitivity in the two $V_{DD}$ domains we considered, using the values obtained on oscinv2 as a reference:
Comparison between the calculated and measured

Illustration of the different voltage domains to be

for \( V_{DD} > 3V_{T0} \)

for \( V_{DD} < 3V_{T0} \) where

must not be considered of fundamental importance.

low voltage applications where the delay performance
to the temperature is reversed. This defines a range of very
sensitivity of CMOS structures in the low voltage

We proposed here a detailed analysis of the temperature

sensitivity to \( V_{DD} \) becomes exponential and the sensitivity
to the temperature decreases

zone, in good agreement with the value \( 2V_{T0} \) deduced
from the model. The definition of this biasing domain is of
great interest for temperature insensitive low power
applications. This approach constitutes one of the first
attempts to model temperature effects in low voltage
applications showing up the insufficiency of the usual
standard modeling.

References:
Exhibit, sept. 21-25, 1992
Design' Kluwer AP 1995
and Optimization of Deep Submicron CMOS Digital Circuits', IEEE
[4] Shih-Wei Sun and Paul G. Y. Tsui, 'Limitation of CMOS Supply-
Voltage Scaling by MOSFET Threshold-Voltage Variation', IEEE Journal
Improvement For Low Voltage Applications' Proc. Euro DAC 95, pp 216-
221, Sept.1995.
[6] Changhae Park et all, 'Reversal of Temperature dependence of
Integrated Circuits Operating at Very Low Voltages' Proc. IEDM
[7] T. Sakurai and A.R. Newton, '\( \alpha \)-power model, and its application to
CMOS inverter delay and other formulas', IEEE JSSC vol. 25, pp. 584-
[8] Amitava Chatterjee, Charles F. Machala and Ping Yang, 'A
Submicronic DC MOSFET Model for Simulation of Analog Circuits',
IEEE Transaction on Computer Aided Design of Integrated Circuits and
[9] J. A. Power et all, 'An Investigation of MOSFET Statistical and
the input-to-output coupling capacitance on the CMOS inverter
Submicrometre CMOS process', IEEE Electronic Letters, Vol. 32 No. 22,
October 1996.

6: Conclusion

We proposed here a detailed analysis of the temperature
sensitivity of CMOS structures in the low voltage
application range. We show that in the low voltage range
temperature and \( V_{dd} \) effects can not be separated as it can
be done in the standard voltage domain. A realistic model
of the delay performance of CMOS structures based on an
\( \alpha \)-power current law calibrated for low voltage \( (V_{DD} < 3V_{ns}) \) is used to define derating factors allowing direct
evaluation of the circuit delay evolution around the
nominal operating point. This has been validated by
comparing the calculated and measured oscillation period
evolution of specific ring oscillators. We show clearly that
for supply voltage values between 2 and \( 3V_{ns} \) the
sensitivity of the delay to the temperature decreases
strongly and cancels at the compensation point of the
temperature effects on mobility and threshold voltage. For
a 0.7\( \mu \)m process we verified that the voltage range 1.5-
1.7V corresponds to the temperature insensitive operating
zone, in good agreement with the value \( 2V_{T0} \) deduced
from the model. The definition of this biasing domain is of
great interest for temperature insensitive low power
applications. This approach constitutes one of the first
attempts to model temperature effects in low voltage
applications showing up the insufficiency of the usual
standard modeling.

References:
Exhibit, sept. 21-25, 1992
Design' Kluwer AP 1995
and Optimization of Deep Submicron CMOS Digital Circuits', IEEE
[4] Shih-Wei Sun and Paul G. Y. Tsui, 'Limitation of CMOS Supply-
Voltage Scaling by MOSFET Threshold-Voltage Variation', IEEE Journal
Improvement For Low Voltage Applications' Proc. Euro DAC 95, pp 216-
221, Sept.1995.
[6] Changhae Park et all, 'Reversal of Temperature dependence of
Integrated Circuits Operating at Very Low Voltages' Proc. IEDM
[7] T. Sakurai and A.R. Newton, '\( \alpha \)-power model, and its application to
CMOS inverter delay and other formulas', IEEE JSSC vol. 25, pp. 584-
[8] Amitava Chatterjee, Charles F. Machala and Ping Yang, 'A
Submicronic DC MOSFET Model for Simulation of Analog Circuits',
IEEE Transaction on Computer Aided Design of Integrated Circuits and
[9] J. A. Power et all, 'An Investigation of MOSFET Statistical and
the input-to-output coupling capacitance on the CMOS inverter
Submicrometre CMOS process', IEEE Electronic Letters, Vol. 32 No. 22,
October 1996.