Impact of Process Variability considering Transistor Networks Delay
Jerson Paulo Guex, Cristina Meinhardt, Ricardo Reis
{jpguex, cmeinhardt, reis}@inf.ufrgs.br
PGMICRO, PPGC, Instituto de Informática
UFRGS

Abstract

This paper presents an analysis of process variability considering the use of extracted layouts with their respective capacitance and parasitic resistance. The effects on PullDown and PullUp networks of transistors are verified separately. The performed experiments aimed to analyze the variability when using serial and/or parallel transistors. Also it analyzes the influence of process variability in fall delay, rise delay. Preliminary results demonstrated that both networks could be less sensitive to process variability if the network is composed by more than 2 transistors in parallel. The use of larger transistors than the minimum possible size for the used technology node also reduces the effects of variability [10].

1. Introduction

The continuous shrinking of devices adds new challenges to integrated circuit design due to variability of transistor’s technology parameters and dimensions. Variability can provide power consumption outside the design specifications, which can allow faster circuit degradation. In some cases, it can occur erroneous circuit functionality, that makes it unsuitable or with use restrictions. Prediction of such variability in the design phase can increase the success rate in the circuit development.

In nanometer scale, decreasing parasitic capacitances increase the transistor speed. On the other hand, there is an increase of the current mismatch, especially if device geometries vary in the manufacturing process [1]. The mismatch of parameters occurs in various steps of manufacturing process and the main parameters affected are transistor dimensions, mobility of electrons or holes and oxide thickness of the transistor gate [2].

According to [4], the variability may be classified by environment factor, commonly understood by variations in power supply and temperature; reliability factor, related to aging of the transistors and their effects like negative bias instability temperature (NBTI), Hot Carries and electromigration; and physical factors related to variations in parameters that induce a mismatch in transistor performance.

These parameters variations turn the circuits less predictable and demand a great effort to create suitable techniques to deal with this uncertain performance. Design for Manufacturability (DFM) is an effort to deal with this issue. DFM has a set of techniques that can be used to improve the manufacturability. According to [5], a regular layout is one of DFM techniques that can improve predictability in the physical implementation of a circuit. Thus, it can be obtained more accuracy in delay estimates.

Parameters like oxide thickness (Tox), width (W), length (L), threshold voltage (Vth), effective channel length (Leff), and change in resistance values of metals after CMP and mobility of transistor’s charges (electrons or holes) are the most important parameters affected by process variability [7]. In Fig. 1, it is possible to see some trends on parameter variability.

![Fig.1-Estimates on Parameters Variation](image)

Hierarchical or spatial mechanism can classify the variability mechanisms in physical factor. These mechanisms can be systematic or random [3] or design dependent according to [4]. The systematic mechanism can be found in many dies or wafers like a rapid annealing. The random mechanisms are related to doping the transistor channel and correction of threshold voltage (Vth).
This paper is focused on physical factors that are imposed in each stage of manufacturing. In this experiment, variability effects over extracted layouts are analyzed considering the parasitic capacitances and resistances. All layouts explore regular patterns on poly layers.

The remainder of this paper is as follows. In section 2 we describe the experimental work and the applied methodology. Results are explained in section 3 and conclusions are presented in section 4.

2. Methodology of the Experiments

The main objective of this work is to analyze the behavior of circuits designed using a 65nm technology considering the effects of process variability in rise and fall signal transitions. It was taken into account the effects in PullUp and PullDown networks. This allows evaluating the influence on delay of stacked transistors, parallel transistors and mixed arrangements. The layouts of the experiments consider the logical effort, method used to estimate delay in CMOS circuits [11], to determine the transistor sizing for each layout. The experiments were done as following:
1. Choosing of 18 circuits to test;
2. Layout Design of selected circuits in a 65nm technology;
3. Verification of layouts versus schematics and extraction;
4. Modification the technology model files to reflect the variability effects;
5. Running of Monte Carlo simulations using HSPICE;
6. Analysis of the results.

Each transistor network was considered separately. For PullDown network was designed 9 layout versions: three layouts to represent schematic diagrams with 2, 3 and 4 transistors in series; three layouts that represent schematic diagrams with 2, 3 and 4 transistors in parallel; and three layouts that represent the union of 2, 3 and 4 transistors in series/parallel. For PullUp network 9 layouts were also designed in the same conditions and specifications that PullDown.

Fig. 2 and 3 show the schematic diagram for some of the layouts. All layouts designed for this experiment were built aiming to be as regular as possible. The layouts were implemented using Cadence Virtuoso. After design rule check (DRC) of the layouts, they were compared with the schematic (LVS) to ensure the correctness of the layout. Following, the layout was passed through the extraction tool to obtain the parasitic capacitances and parasitic resistances.

In order to simulate the process variability, transistors parameters like width (W), length (L), oxide thickness (Tox), effective channel length (Leff) and threshold voltage (Vth) were varied by 10% of their nominal values. The correlations between these parameters is considered. All parameters are changed in the model and this model employs these basic parameters to determine all the others parameters values. The variability adopts a Gaussian distribution with three sigmas around the default values.

The Monte Carlo technique was adopted with a Gaussian distribution to deal with simulation. Even though this method is very expensive in terms of computing power and is not suitable for complex circuits with over one hundred transistors [6], in this work the method works efficiently because only small circuits are explored. These small circuits reflect blocks often found in complex circuits. Each of the extracted layouts was simulated using Hspice with the Monte Carlo method and doing 10,000 iterations per simulation.

The transistor model used for those experiments is a Low Power Typical model from the same foundry that provides the design kit of 65nm. A load capacitance of 10fF was placed in each extracted circuit output to avoid floating terminals to emulate a real situation of operation. Each input was fed with the same signal with a 1ns of period.

Fig. 2 - Schematic diagram of pulldown network with: (a) 2 transistors in series, (b) 2 in parallel and (c) 2 transistor in parallel and 1 in series

Fig. 3 - PullUP network with: (a) 2 transistors in series, (b) 2 transistors in parallel and (c) 2 transistors in parallel with 1 in series.
3. Experimental Results

The simulation experiments run over HSpice, performing Monte Carlo method with 10000 iterations using the BSIM4 model. For PullDown networks (series/parallel and mixed with 2, 3 and 4 transistors) were observed the mean values for fall delay, fall time of the analyzed network and their respective standard deviations. In the same way, the results for mean value and standard deviations of rise delay, rise time and power estimation of PullUP network were acquired.

The results for the PullDown network show that the normalized fall delay deviation tends to decrease in function of the number of stacked transistors and their respective width. The normalized fall delay deviation decay approximately by 9% comparing the scenarios with 2 and 4 stacked transistors.

Looking to the result of normalized fall delay deviations it is possible to conclude that parallel arrangements with 3 or 4 transistors could be a good strategy for the design of circuits more tolerant to process variability. The junction of series and parallel transistor shows that this combination could be more sensitive to the process variability. This experiment shows a linear trend in increasing of normalized fall delay deviation when both parallel and series arrangement are put together with 3 or 4 transistors. According to the obtained data, the implementation with 2 transistors in both parallel and stacked seems to be a good strategy to cope with process and to deal with fall delay constraints. Tab. I show the results of normalized fall delay deviation during the simulations. The results are normalized dividing the standard deviation by its mean value. This normalized process was made to provide a fare comparison between the experiments.

<table>
<thead>
<tr>
<th># Transistors</th>
<th>Series</th>
<th>Parallel</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.03263</td>
<td>0.03140</td>
<td>0.02917</td>
</tr>
<tr>
<td>3</td>
<td>0.03016</td>
<td>0.02358</td>
<td>0.03244</td>
</tr>
<tr>
<td>4</td>
<td>0.02987</td>
<td>0.02016</td>
<td>0.03458</td>
</tr>
</tbody>
</table>

The PullUP network results show that the rise delay deviation tends to decay according to the number of stacked transistors. This effect is noticed because of the continuous increment of transistors width according to the number of transistors in series observing the relation of \( W_p/W_n=2 \) mentioned in Section II and the position of transistors according to the Figure 2. Parallel transistors arrangements have also the rise delay deviation decay in all the situations, which include 2, 3 and 4 transistors.

With the arrangements to perform this experiment for the PullUP network, it is possible to compare the decrease of the rise delay deviation in function of the number of transistors in series or parallel and with both topologies. The decay of rise delay deviation in parallel and series arrangements is a direct product of the area of each transistor, as showed in Tab 2. As greater is the transistor, less is the rise delay deviations adopting a mixed arrangement, in the booth column, the experiments show a tendency of stability in rise delay deviation with 2, 3 and 4 transistors. Taking in account all the aspects related above, the best solution to implement a PullUP network regarding the effects of process variability is by using a parallel structure that guarantee a minor effect on rise delay deviation. This consideration is only valid if the transistors are not using minimal dimensions for the technology used.

<table>
<thead>
<tr>
<th># Transistors</th>
<th>Series</th>
<th>Parallel</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.03039</td>
<td>0.0254</td>
<td>0.02805</td>
</tr>
<tr>
<td>3</td>
<td>0.02642</td>
<td>0.01965</td>
<td>0.02761</td>
</tr>
<tr>
<td>4</td>
<td>0.02164</td>
<td>0.01563</td>
<td>0.02887</td>
</tr>
</tbody>
</table>

4. Conclusion

The experiments performed over the PullDown and PullUp networks show the dependence of variability in relation to the size of transistors. It can be verified in Figure 4, that the normalized fall and rise delays have the same behavior for variability decay when the number of transistors and their sizes in serial and parallel arrangements is increased. This indicates that the use of the minimum possible dimensions of the technology could be avoided to decrease the effects of process variability. Other conclusion highlights regarding the use of more parallel arrangements in PullDown and PullUp networks when the target is minimize the fall and rise delay deviation fluctuations under influence of process variability.
Fig. 4 – Comparison between rise and fall delay deviation.

5. Acknowledgment
The authors would like to thank the CAPES, Brazilian Funding Agency and also to the Walter E. C. Bartra for his help and collaboration.

6. References