CMOS Logic Gate Performance Variability Related to Transistor Network Arrangements

Digeorgia N. da Silva, André I. Reis, Renato P. Ribas
PGMicro - Federal University of 'Rio Grande do Sul', Av. Bento Gonçalves 9500, CEP 91501-970, Porto Alegre, RS, Brazil.

Corresponding author. Tel.: +55 51 3308 6810; fax: +55 51 3308 7308.
E-mail address: dnsilva@inf.ufrgs.br (Digeorgia da Silva).

ABSTRACT

The rapid scaling of CMOS technology has resulted in drastic variations of process parameters. Since different transistor arrangements present different electrical characteristics, this work analyzes the impact of process variability in performance of logic gates, according to their topology and the relative position of the switching device in network. Results have been obtained through Monte Carlo simulations and design guidelines for parametric yield improvement have been derived from.

Keywords
Performance variability, parametric yield, transistor networks, CMOS logic gates.

1. Introduction

Manufacturing variations may lead to significant discrepancies between designed and fabricated integrated circuits. Due to the shrinking of device size, the relative impact of critical dimension variations tends to increase at each new technology generation, since the process tolerances do not scale at the same rate [1]. Many studies about effects of intrinsic processes on the functionality and reliability of circuits have been done in recent years [2]-[6]. Since process variations become a more critical issue due to aggressive technology scaling, the migration from deterministic to statistical analysis of circuit designs may reduce conservatism and failure risk compared with applying the traditional worst-case corner approach. The traditional corner-case static timing analysis (STA) technique seems as a reasonable way to handle global variations on a wafer but not local ones [7][8]. In terms of circuit performance, a logic gate may become slower for a certain variation and faster for another one, and that might depend on its location on a die. The importance of intra-die variations has grown as well, and the number of process parameters which present considerable variations has also increased. Such situation requires some changes in STA in order to find alternatives to their deterministic nature. In nanoscale CMOS devices, the reduced average number of dopant atoms in the channel of a transistor increases the effect of random dopant fluctuations on its threshold voltage to increase [9].

Increasing levels of process variations have a major impact on power consumption and performance of a design. This impact may result in parametric yield loss [10]. Parametric yield improvement may be achieved by reducing the variability of performance and power consumption of a cell. A high sensitivity of a device to variations in its parameters means that the yield window, limited by frequency and power constraints, is narrower than when a device is more immune to variability. A narrow yield window means that a high quantity of manufactured chips may not satisfy operational specification, leading to a higher cost of fabrication, since many chip may become useless.

It is important to analyze circuit performance under process variation for yield prediction as well as for circuit optimization. By performing a full-scale transistor-level Monte-Carlo simulation on a circuit, one gets the most accurate way of incorporating the process variation effects into timing analysis. It generates samples for a given delay distribution and runs a static timing analyzer at each point. The results are put together to form the delay distribution [11].

On the other hand, different logic styles result in transistor networks with different electrical and physical characteristics, and there is more than one type of circuit that can be used to represent a certain logic function [12]. The impact of parameters variation of a cell on its metrics is not the same in different logic styles. Also, different topologies may even for the same logic style result in different behavior under process variations.

The purpose of this work is to evaluate the impact of variation of transistor threshold voltage on CMOS logic gate behavior, according to (i) network topology (transistor arrangement) and (ii) the relative position of the switching transistor in relation to the power supply and output terminals. These data may lead to the development of design guidelines for parametric yield improvement. This paper presents some timing analysis performed on different gates by using electrical simulations.

The paper is organized as follows. Section 2 outline the methodology applied. Simulation results and analysis are presented in Section 3. In Section 4, the conclusions are discussed.

2. Methodology

Transistor threshold voltage (V_{th}) was varied and timing measurements (delay propagation) were taken. The mean delay and standard deviation of the logic gates were then compared, and emphasize the relation of these values to the transistor network arrangements. Timing data were extracted by using Monte Carlo Spice simulations. CMOS Inverter, 2- to 4-input NAND and NOR, and more complex gate (AOI_21 and AOI_32) were used as case studies. Results were obtained from simulations for 3σ deviation of 10% from nominal V_{th}. Correlation between transistors, i.e. that a PMOS may change its parameters when placed in the vicinity of a NMOS was not taken into account. The technology node used in this work was 45 nm and the model file is the Predictive Technology Model (PTM) [13] based on BSIM4. Simulations were carried out by using HSPICE tool.

3. Simulation Results and Analysis

3.1 CMOS Inverter

In a first set of simulations, CMOS inverters were evaluated for different drive strengths while keeping fixed P/N ratio. Results are presented in Fig. 1. As can be seen, the increasing in the drive strength (X1, X2,…,X5) of the inverter results in different behaviors of its metrics and variability. It can be observed in Fig. 1 that the timing behavior directly related to the PMOS transistor (the rise delay deviation) is less impacted by variations in V_{th} of the transistor than the metric depending directly on the NMOS (the fall delay deviation). The larger the size of the inverter, the smaller the rise delay deviation and the larger the fall delay deviation. It means that falling transitions at the output node of inverters which are placed on critical paths of circuits are more critical for parametric yield and timing stability. Such information is quite useful for buffer insertion task, for instance.

3.2 NAND and NOR Gates

NAND and NOR static CMOS logic gates were also considered for such an investigation since they allows the evaluation of series transistors impact, for pull-up PMOS and pull-down NMOS transistor stacks in NOR and NAND cells, respectively. Usually, timing arcs are taken into account for each input signal transition. Fig. 2a shows rise and fall delay deviations according to the position of switching device in relation to the output node of NAND gates with different number of inputs. Two extreme situations can be identified: (i) when the switching transistor is connected to the cell output terminal ('close' switching) and (ii) when it is connected to the power supply terminal (Vdd or ground) in a stack arrangement ('far' switching). Transitions close to the logic gate output node result in lower mean rise delay and its deviation than transitions far from such node. In this case, the rise delay deviations obtained are similar for different numbers of inputs. For a signal applied close to the output, fall delay deviation decreases as the number of inputs of the NAND gate increases. The fall and rise delay deviations increase as the number of inputs increases for a transient signal applied far from the output node. Regarding the mean value of delay, there is an increase with the number of inputs, especially when the transient signal is applied far from the output.

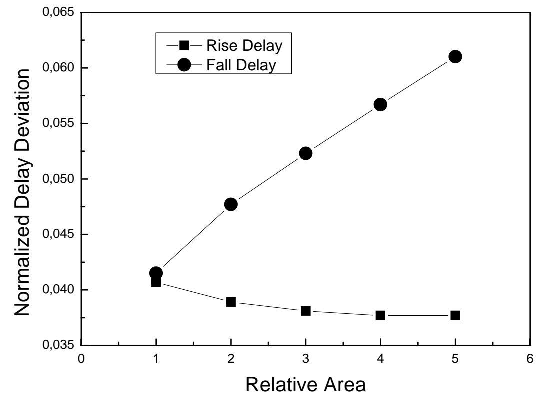


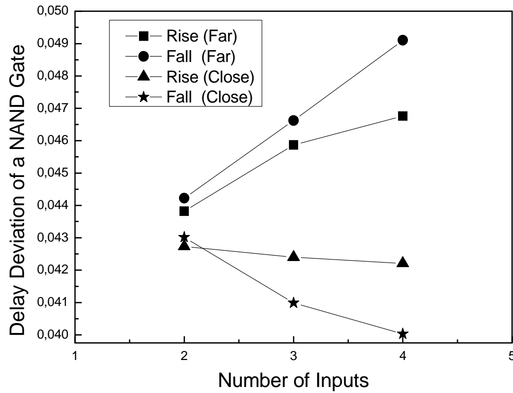
Fig. 1. Normalized rise and fall propagation delay deviations of CMOS inverter by varying drive strength (cell sizing).

In the particular case of NAND gates, lower delay values and delay deviations may be achieved when transient input signals are applied close to the output node. In a transistor stack there are differences in the potential of similar areas of devices, resulting in different gate-to-source (V_{gs}) and drain-to-source (V_{ds}) voltages. Therefore, variations in the threshold voltage may lead to different impact on the drive strength of devices. In NAND gates, the amount of charge that needs to go through a switching transistor far from the output is larger than when it is close to the output, considering other devices in 'on-state'. It helps to explain the dependence of performance variation of the logic gate on the position of the switching transistor.

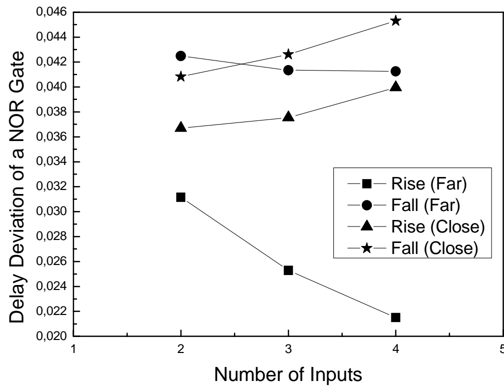
Fig. 2b shows rise and fall delay deviations for transitions far and close to the output node of a NOR gate. In the case of a switching transistor close to the output node, NOR presents rise delay deviations that increase with the number of inputs. The opposite happens for a switching transistor far from the output node, where the deviation decreases as the number of inputs increases. Rise delay is less affected by variations in the threshold voltage of transistors than fall delay, as observed in CMOS inverter.

Where a series PMOS close to the output is switched the situation is similar to that one where a NMOS far from the output is applied a transient signal, in the sense

that other intrinsic capacitances in the arrangement are already or still charged.



(a)



(b)

Fig. 2. Normalized rise and fall delay deviations in relation to the number of inputs: (a) NAND and (b) NOR gates.

The analysis of series transistor configuration in NAND and NOR arrangements showed that the position of the switching transistor in relation to the output node influences the sensitivity of the gate to performance variations. In NOR gat, the best situation (higher robustness) happens when the switching transistor is as far as possible from the output and less robustness is observed when the closest-to-the-output transistor switches. In the case of NAND gate, in turn, higher robustness is achieved by applying the transient signal close to the output node. The results for variations of delay are not the same as the results for the absolute delay value. It is well known that a better timing (lower delay) is achieved when a critical path signal is crossing through the switching device closer to the logic gate output node. A trade-off is required since it is not interesting to have the timing of the cell with a high mean value even though it presents low variability.

Rising- and falling-edge output signals go through essentially different paths in NAND and NOR gates. In the former, series transistors are in pull-down NMOS network and they are responsible for a falling-edge output signal. On the other hand, in the latter, series transistors are in pull-up PMOS network and are responsible for a rising-edge output. The comparison between the influences of the variations in the parameters on NOR and NAND delays are physically more appropriate by considering equivalent array of transistors: series-to-series or parallel-to-parallel. In this case, the fall delay variations of NAND gate may be compared to the rise delay variations of NOR gate, and vice-versa. Fig. 3 shows delay deviations for transitions far and close to the output node for NAND and NOR gates in stacked transistors. For both situations, NAND gates are more sensitive to variations in transistor threshold voltage than NOR gates.

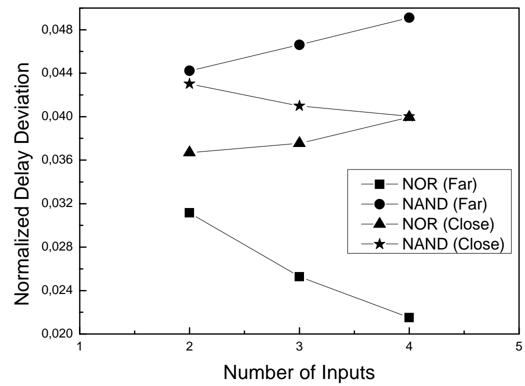


Fig. 3. Comparison of NMOS and PMOS transistor stacking in NAND and NOR gates, respectively, for different positions of the switching device ('close' to and 'far' from output node).

By analyzing the sensitivity of basic gates to V_{th} variations, some tendencies were observed in the deviations of their delay due to transistor network structure and the position of the switching transistor in relation to the output node. Such analysis cannot conclude that NAND and NOR gates with fewer inputs would be the best or the worst choice, once opposite behavior of delay deviation is observed according to the position of the switching device in the network. On the other hand, in critical paths optimization, by switching transistors closer to the gate output node tends to provide better performance in terms of absolute delay as well as parametric yield improvement.

3.3 NAND: Single Gate Versus Mapped Circuit

While evaluating topologies with different number of inputs a question arose: would it be better, in terms of variability, to replace a single complex gate with large number of inputs by a circuit mapped to basic gates with

fewer inputs, to implement the same logic function? Fig. 4 illustrates how it could be done in the case of a 3-input NAND gate. Table 1 presents the results obtained for single 3-input NAND gate (‘NAND3’) and a version composed by two 2-input NAND gates (‘2xNAND2’). This case was investigated considering only one transistor switching at a time, and the fastest and the slowest paths were identified.

Fig. 4. Illustration of single 3-input NAND gate implemented by using two 2-input NAND gates (‘2xNAND2’).

Table 1
Delay deviation of the shortest and the longest paths in Fig. 4.

	Best-case delay		**Worst-case delay**	
	2xNAND2	**NAND3**	**2xNAND2**	**NAND3**
Mean Rise Delay (ps)	87.99	82.39	153.93	141.30
Norm. Rise Delay Dev.	0.0446	0.0424	0.0293	0.0457
Mean Fall Delay (ps)	83.42	48.75	163.18	67,85
Norm. Fall Delay Dev.	0.0339	0.0466	0.0243	0.0410

The single 3-input NAND gate is more sensitive to variations of transistor threshold voltage than the version composed by two 2-input NAND gates for the slowest signal propagation (worst-case), since variations in the threshold voltage of a 3-input NAND resulted in higher delay deviation than in the case when two 2-input NAND gates (with additional inverter) were used. For the fastest propagation (best-case) it is not completely so. Though fall delay deviation is higher for NAND3, rise delay deviation is almost the same for both configurations. Also, the NAND3 is much faster than the implementation with 2-input NAND gates for a falling-edge at the output node. The results shown in Table 1 agree well with Fig. 2a for rise delay deviation, once it is not really affected by the number of input signals in the logic gate.

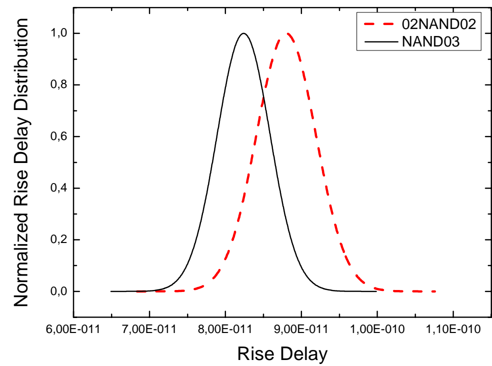
A more complete analysis is possible by the probability density functions of delay for both topologies, as presented in Fig. 5. Though NAND03 is more sensitive to variations in V_{th}, the probability density functions of rise and fall delays for the longest and the shortest paths show that this gate guarantees faster signal propagation for almost every variation in V_{th}.

In terms of design guidelines derived from, it could be concluded that perform the technology mapping task using preferentially small (basic) logic gates instead of complex ones leads to a significant parametric yield improvement. It is probably true for the worst-case rise delay in Table 1, whose mean delay is similar for both approaches. In the case of the fall delay values shown in the same table, such analysis must be continued by considering circuit sizing optimization, once the mean fall delays are quite different.

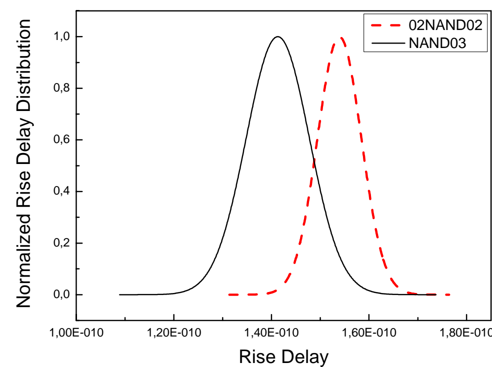
3.4 And-Or-Inverter (AOI) Logic Gates

Previous analysis, taking into account separately pull-down NMOS and pull-up PMOS logic networks, has demonstrated that the fewer device count is present in transistor arrangements, the less sensitive it is to performance variations. And-Or-Inverter configurations (AOI_21 and AOI_32) were implemented in two versions: (i) as a single CMOS complex gate and (ii) by using basic cells (2-input NAND and NOR gates). These topologies provide mixed arrangements of series and parallel transistors in the pull-up PMOS and pull-down NMOS networks. The goal is to evaluate if such implementation becomes more susceptible to variations than the same logic function mapped with basic gates.

The implementation of AOI_21 by considering basic gates presented lower delay deviations, but higher mean fall delay in comparison to the single complex gate approach. Similar situation has been observed for AOI_32 implementation. Results are summarized in Table 2.



(a)



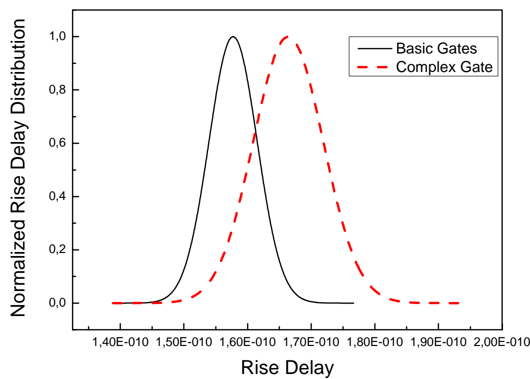
(b)

Fig. 5. Probability Density Functions of rise delay for a 3-input NAND gate and a circuit performing the same logic function implemented by using two 2-input NAND gates for the best (a) and the worst (b) delay propagations.

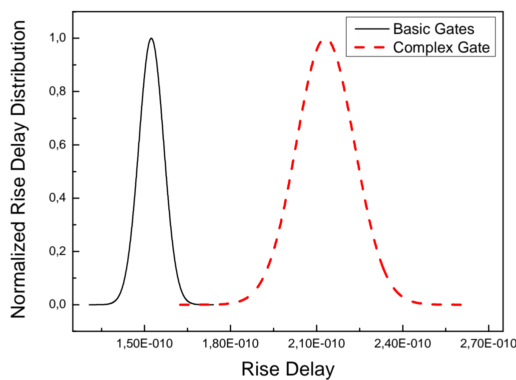
Table 2
Delay deviation for AOI_21 and AOI_32 logic gates.

	AOI_21		AOI_32	
	complex gate	**basic gates**	**complex gate**	**basic gates**
Mean Rise Delay (ps)	166.39	157.73	213.05	152.38
Norm. Rise Delay Dev.	0.0331	0.0240	0.0475	0.0284
Mean Fall Delay (ps)	52.64	126.97	103.75	172.45
Norm. Fall Delay Dev.	0.0465	0.0188	0.0468	0.0303

Fig. 6 illustrates the rise delay distributions for both topologies. The implementation with basic gates was able to reduce the overall delay of AOI_32 configuration and guaranteed more reliability for changes in transistor Threshold voltage. It suggests that complex implementations presenting a larger number of series and parallel transistors in the cell topology may reduce the mean delay value but at expense of increasing the performance variability. Circuit sizing was not considered for performance optimization, being all gate sized for similar drive strength.



(a)



(b)

Fig. 6. Probability Density Functions of rise delay for AOI_21 (a) and AOI_32 (b) gates implemented by using basic CMOS cells and as a single complex gate.

These last experiments by considering AOI logic gates, the results and analysis are similar to the ones discussed in the previous section. The mapped circuits, based on small cells, provide better performance in terms of delay variability. Even if gate sizing could influence such results, the values presented in Table 2 suggest lower normalized delay deviations for 'basic gates' approach in both cases when the mean delay does not present the same tendency, as observed in the AOI_32 results. It reinforce the design guideline that suggests the use of small (basic) gates as preferential choice in the technology mapping task when parametric yield improvement in targeted.

4.Conclusions

Results obtained in this work, about performance variability in CMOS logic gates submitted to transistor threshold voltage variation, demonstrated the strong dependency in relation to gate topology, number of stacked transistors, and the relative position of switching device in transistor network arrangements. Such analysis suggests the preferential use of basic CMOS gates instead of complex ones (AOI, for instance) in the technology mapping task of combinational circuits. Moreover, in terms of critical delay paths optimization, switching transistors placed close on the gate outputs are preferable for absolute delay propagation as well as delay variability resulted from V_{th} variation.

Acknowledgement

This work has been developed in cooperation with Nangate Inc., including financial support. The author Digeorgia N. da Silva was supported by a doctoral scholarship from CNPq Brazilian agency.

References

[1] Borkar S. Designing reliable systems from unreliable components: the challenges of transistor variability and degradation. IEEE Micro 2005; 6:10-6.

[2] Orshansky M, Milor L, Hu C. Characterization of spatial intrafield gate CD variability, its impact on circuit performance, and spatial mask-level correction. IEEE Trans. Semiconductor Manufacturing 2004; 1:02-11.

[3] Kishor M, Pineda de Gyvez J. Threshold voltage and power-supply tolerance of CMOS logic design families. In: IEEE Symp. on Defect and Fault Tolerance in VLSI Systems 2000; 349-57.

[4] Olivieri M, Scotti G, Trifiletti A. A novel yield optimization technique for digital CMOS circuits design by means of process parameters run-time estimation and body bias active control. IEEE Trans. Very Large Scale Integration (VLSI) Systems 2005; 5:630-8.

[5] Argawal K, Rao R, Sylvester D, Brown R. Parametric yield analysis and optimization in leakage dominated technologies. IEEE Trans. Very Large Scale Integration (VLSI) Systems 2007; 6:613-23.

[6] Okada K., Yamaoka K., Onodera H. A statistical gate-delay model considering intra-gate variability. In: Int. Conf. on Computer Aided Design (ICCAD) 2003; 908-13.
[7] Sapatnekar S. Timing. 1st ed. Boston: Kluwer Academic Publishers, 2004.
[8] Blaauw D, Zolotov V, Sundareswaran S. Slope propagation in static timing analysis. IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems 2002; 10:1180-95.
[9] Mahmoodi H., Mukhopadhyay S., Roy K. Estimation of delay variations due to random-dopant fluctuations in nanoscale CMOS circuits. IEEE Journal of Solid-State Circuits 2005; 9:1787-96.
[10] Seung H. C., Paul B.C., Roy K. Novel sizing algorithm for yield improvement under process variation in nanometer technology. In: Design Automation Conference (DAC) 2004; 454 -9.
[11] Orshansky M, Nassif S, Boning D. Design for manufacturability and statistical design: a constructive approach. 1st ed. Springer, 2008.
[12] Rosa Jr. L. Automatic generation and evaluation of transistor networks in different logic styles. PhD thesis. UFRGS, 2008.
[13] Zhao W, Cao Y. New generation of Predictive Technology Model for sub-45nm early design exploration. IEEE Trans. Electron Devices 2006; 11:2816-23.