Analysis of Propagation Delay Deviation under Process Induced Threshold Voltage Variation

K.G. Verma
School of Electronics,
Shobhit University,
Meerut, Uttar Pradesh, INDIA

Dr. Brajesh Kumar Kaushik
Department of Electronics and Computer Engineering,
Indian Institute of Technology Roorkee, Uttarakhand, INDIA

Dr. Raghuvir Singh
School of Electronics,
Shobhit University,
Meerut, Uttar Pradesh, INDIA

ABSTRACT
Process variation has become a major concern in the design of many nanometer circuits, including interconnect pipelines. The primary sources of manufacturing variation include Deposition, Chemical Mechanical Planarization (CMP), Etching, Resolution Enhancement Technology (RET). Process variations manifest themselves as the uncertainties of circuit performance, such as delay, noise and power consumption. The performance of VLSI/ULSI chip is becoming less predictable as device dimensions shrinks below the sub-100-nm scale. The reduced predictability can be attributed to poor control of the physical features of devices and interconnects during the manufacturing process. Variations in these quantities maps to variations in the electrical behavior of circuits. Threshold voltage of a MOSFET varies due to changes in oxide thickness; substrate, polysilicon and implant impurity level; and surface charge. This paper provides a comprehensive analysis of the effect of threshold variation on the propagation delay through driver-interconnect-load (DIL) system. The impact of process induced threshold variations on circuit delay is discussed for three different technologies i.e 130nm, 70nm and 45nm. The comparison of results between these three technologies shows that as device size shrinks, the process variation issues becomes dominant during design cycle and subsequently increases the uncertainty of the delays.

Keywords
Process variation, interconnects, VLSI, systematic variation, propagation delay.

1. INTRODUCTION
The semiconductor industry has been fueled by enhancements in integrated circuit (IC) density and performance, resulting in information revolution for over four decades and is expected to continue in future. The periodic improvement in density (as per Moore's Law) and performance has been mainly achieved through aggressive device scaling and/or increase in chip size. As far as MOS transistor scaling is concerned, device performance improves as gate length, gate dielectric thickness, and junction depth are scaled. In sharp contrast to this, scaled chip wiring (interconnect) suffers from increased resistance due to decrease in conductor cross-sectional area and may also suffer from increased capacitance if metal height is not reduced with conductor spacing.

Variability in modern nanometer circuits has not scaled down in proportion to the scaling down of their feature sizes. Manufacturing process variations (e.g. threshold voltage, effective channel length), environmental variations (e.g., supply voltage, temperature), and device fatigue phenomenon contribute to uncertainties. Uncertainty due to parametric variations deeply impacts the timing characteristics of a circuit and makes timing verification extremely difficult. This necessitates the consideration of the parametric variations in timing analysis for accurate timing estimation.

The feature size of integrated circuits has been aggressively reduced in the pursuit of improved speed, power, silicon area and cost characteristics [1]. Semiconductor technologies with feature sizes of several tens of nanometers are currently in development. As per, International Technology Roadmap for Semiconductors (ITRS), the future nanometer scale circuits will contain more than a billion transistors and will operate at clock speeds well over 10GHz. Distributing robust and reliable power and ground lines; clock; data and address; and other control signals through interconnects in such a high-speed, high-complexity environment, is a challenging task [2, 3] as every system implemented either through ASIC design or on FPGA are prone to the effects of the parasitic components of interconnect impedance.

The function of interconnects or wiring systems is to distribute clock and other signals and to provide power/ground to and among the various circuits/systems functions on the chip [2]. The performance *s.a.* time delay and power dissipation of a high-speed chip is highly dependent on the interconnects, which connect different macro cells within a VLSI chip. To escape prohibitively large delays, designers scale down global wire dimensions more sluggishly than the transistor dimensions. As technology advances, interconnects have turned out to be more and more important than the transistor resource, and it is essential to use global interconnects optimally. For high-density high-speed submicron-geometry chips, it is mostly the interconnection rather than the device performance that determines the chip performance.

Distribution of the clock and signal functions is accomplished on three types of wiring (local, intermediate, and global). An interconnect depending on its length, can be classified as local, semi-global and global [1]. Local wiring, consisting of very thin lines, connects gates and transistors within an execution unit or a functional block (such as embedded logic, cache memory, or address adder) on the chip. Local wires usually span a few gates and occupy first and sometimes second metal layers in a multi-level system. The length of a local interconnect wire approximately scales with scaling of technology, as the increased packing density of the devices make it possible to similarly reduce the wire lengths. Intermediate wiring provides clock and signal distribution within a functional block with typical lengths up to 3–4 mm. Intermediate wires are wider and taller than local wires to provide lower resistance signal/clock paths. Global wiring provides clock and signal distribution between the functional

blocks, and it delivers power/ground to all functions on a chip. Global wires, which occupy the top one or two layers, are longer than 4mm and can be as long as half of the chip perimeter. The length of global interconnect wires grow proportionally to the die size. The length of semi-global interconnect behaves intermediately. The global interconnects are much wider than local and semi-global interconnects. Thus resistance of global interconnects is small and therefore their behavior resembles that of lossless transmission lines.

The performance of a high-speed chip is highly dependent on the interconnects, which connect different macro cells within a VLSI/ULSI chip. With ever-growing length of interconnects and clock frequency on a chip, the effects of interconnects cannot be restricted to *RC* models [2, 3]. The importance of on-chip inductance is continuously increasing with faster on-chip rise times, wider wires, and the introduction of new materials for low resistance interconnects. It has become well accepted that interconnect delay dominates gate delay in current deep sub micrometer VLSI circuits [1-3]. With the continuous scaling of technology and increased die area, this behavior is expected to continue.

Wide wires are frequently encountered in clock distribution networks, power and ground lines, and other global interconnects such as data bus and control lines in upper metal layers. These wires are low resistive lines that can exhibit significant inductive effects. Due to presence of these inductive effects, the new generation VLSI designers have been forced to model the interconnects as distributed *RLC* transmission lines, rather than simple *RC*–ones. Modeling interconnects as distributed *RLC* transmission line, has posed many challenges in terms of accurately determining the signal propagation delay; power dissipation through an interconnect; crosstalk between co-planar interconnects and interconnects on different planes due to capacitive and inductive coupling; and optimal repeater insertion [1-3].

On-chip global interconnects is among the top challenges in CMOS technology scaling due to rapidly increasing operating frequencies and growing chip size. The clock signal has already been brought into the multi-gigahertz range where inductance and other transmission line effects of on-chip long lines become important. For higher operating frequencies, dispersion and skin effects are among the new concerns. The use of reverse scaling methodology will decrease the line resistance, but the line inductance effects will become more prominent. The global clock network, which was already power hungry, is likely to consume more power and hence become even more difficult to design. Particularly, the delay induced by word lines, bit lines, clock lines, and bus lines in memory or logic VLSI will remain the key concerns while designing the interconnects.

The performance of VLSI/ULSI chip is becoming less predictable as device dimensions shrink below the sub-100-nm scale [4-6]. The reduced predictability can be attributed to poor control of the physical features of devices and interconnects during the manufacturing process. Variations in these quantities result in variations in the electrical behavior of circuits. These variations have interdie and intradie components, as well as layout pattern dependencies. The device material variations in geometry (t_{ox}, L_{eff}, W), and variations in doping levels and profiles have a direct impact on the behavior of a MOSFET. Variations in the linewidth affect the resistance and the interlayer capacitance. Variations in the interwire spacing may cause a significant degradation in the signal integrity. Layout pattern dependent variations within the interlayer oxide and the chip multiprocessing process also have a significant impact on the interconnect parasitics. The dissimilar sources of variations in the IC fabrication process lead to both random and systematic effects on circuit performance. All of these make it increasingly difficult to accurately predict the performance of a circuit at the design stage, which ultimately translates to a parametric yield loss. The recent trends in VLSI chip exhibit significant variations within a chip and between chips, due to the high complexity of design and the presence of large number of correlated parameters. Therefore, fast and efficient methods are required to compute an accurate statistical description of the response.

The threshold voltage of a MOSFET is usually defined as the gate voltage where an inversion layer forms at the interface between the insulating layer (oxide) and the substrate (body) of the transistor. In an n-MOSFET the substrate of the transistor is composed of p-type silicon, which has positively charged mobile holes as carriers. When a positive voltage is applied on the gate, an electric field causes the holes to be repelled from the interface, creating a depletion region containing immobile negatively charged acceptor ions. A further increase in the gate voltage eventually causes electrons to appear at the interface, in what is called an inversion layer, or channel. Historically the gate voltage at which the electron density at the interface is the same as the hole density in the neutral bulk material is called the threshold voltage. Practically the threshold voltage is the voltage at which there are sufficient electrons in the inversion layer to make a low resistance conducting path between the MOSFET source and drain.

Process variations are not completely random. It can be divided into deterministic part and nondeterministic part [6, 7]. Random variations are intrinsic fluctuations in process parameters such as dopant fluctuations from wafer to wafer, lot to lot. On the other hand, systematic variations depend on the layout pattern and are therefore predictable for the systematic part, the variations need to be experimentally modeled and calibrated, in order to either compensate hiring the design phase or captured in the analysis phase. These effects, which include optical proximity correction (OPC), residual error and chemical mechanical planarization (CMP) dishing [8, 9], have a substantial but deterministic impact on the critical dimension (CD) of a transistor gate or the width and thickness of an interconnect wire. By accounting for systematic part of process variation in timing analysis, uncertainty can be reduced, thereby achieving closer bound for circuit performance. With the shrinking feature size in VLSI technology, the impact of process variation is increasingly felt. To address the effect, great amount of research has been done recently, such as the clock skew analysis under process variation [4-10], statistical performance analysis [9, 10], worst case performance analysis [11, 12], parametric yield estimation [12, 13], impact analysis on micro architecture [12, 13] and delay fault [14, 15] test under process variation [14-17]. As the technology reaches deep submicron or nanometer regime, the errors due to process variations becomes prominent [17-19]. Threshold voltage of a MOSFET varies due to (1) Changes in oxide thickness; (2) Substrate, polysilicon and implant impurity level; (3) Surface charge.

This paper analyzes the effect of threshold voltage variation due to process variation on the propagation delay of Driver-Interconnect-Load (DIL) system as shown in Figure 1. The propagation delay variations through DIL system are observed

due to process variations in driver individually for different technologies *i.e* 130nm, 70nm and 45nm.

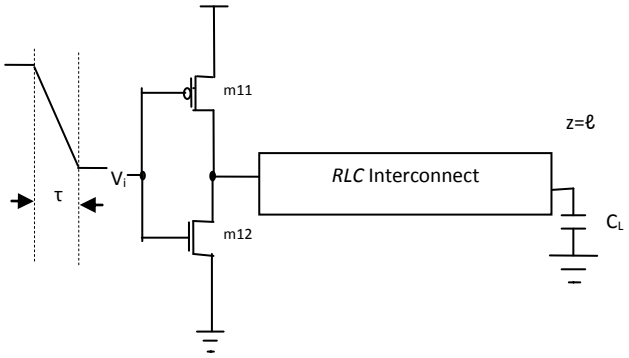

Figure 1 Driver Interconnect Load (DIL) System

2. MONTE CARLO ANALYSIS

The term "Monte Carlo method" was coined in the 1940s by physicists working on nuclear weapon projects in the Los Alamos National Laboratory, after the famous Monte Carlo casino, a gambling venue based on random-number generation.

Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to compute their results. Monte Carlo methods are often used in simulating physical and mathematical systems. Because of their reliance on repeated computation of random or pseudo-random numbers, these methods are most suited to calculation by a computer and tend to be used when it is infeasible or impossible to compute an exact result with a deterministic algorithm. This method is also used to complement the theoretical derivations.

Monte Carlo simulation methods are especially useful in studying systems with a large number of coupled degrees of freedom. More broadly, Monte Carlo methods are useful for modeling phenomena with significant uncertainty in inputs. These methods are also widely used in mathematics: a classic use is for the evaluation of definite integrals, particularly multidimensional integrals with complicated boundary conditions. It is a widely successful method in risk analysis when compared with alternative methods or human intuition.

The analysis carried out in this work takes into account a Driver-Interconnect-Load (DIL) system as shown in Figure 1. The driver is an inverter gate driving the interconnect. The threshold voltage of the transistor in the driver is described by the following equation

$$V_T = V_{TO} + \gamma\sqrt{2\Phi_f + V_{SB}} - \sqrt{2\Phi_f} \quad (1)$$

where in equation (1)

V_{TO} = Threshold voltage for $V_{SB} = 0$V

$$\Phi_f = -\frac{kT}{q} ln \frac{N_A}{n_i}$$

γ = Fabrication –process parameter and is given as

$$\gamma = \frac{\sqrt{2qN_A\varepsilon_s}}{C_{ox}}$$

N_A =doping concentration of p-type substrate.

C_{ox} = Gate oxide capacitance

The threshold voltage of a device is dependent on various physical parameters which are prone to process variation. In this analysis, the driver is subjected to process variations in reference to threshold voltage for three different technologies of 130nm, 70nm and 45nm. To obtain statistical information on how much the characteristics of a circuit can be expected to scatter over the process, Monte Carlo analysis is applied. Monte Carlo analysis performs numerous simulations with different boundary conditions. It chooses randomly different process parameters within the worst case deviations from the nominal conditions for each run and allows statistical interpretation of the results. In addition to the process parameter variations, mismatch can be taken into account as well, providing a more sophisticated estimation of the overall stability of the performance with respect to variations in the processing steps. In most cases the parameters on which the assumptions for the mismatch are based are worst case parameters. A proper layout and choice of devices can significantly improve scatter due to mismatch. In order to obtain reasonable statistical results, a large number of simulations are needed, leading to quite long simulation times.

3. EFFECT OF THRESHOLD VOLTAGE VARIATION ON DELAY OF DIL SYSTEM

Monte Carlo simulations are run for threshold voltage variations in 130nm, 70nm and 45nm fabrication technology. Figure 2 shows the SPICE input and output voltage for a variation of 30% in threshold voltage in NMOS and PMOS transistors in 130nm technology. It is observed that the output varies significantly due to the process variation parameter. Table-I accounts for NMOS threshold voltage (*Vtn*); PMOS threshold voltage (*Vtp*); the delay due to driver and interconnect line; the percentage variation in NMOS and PMOS threshold voltage and percentage variation in delay of driver and line. It is observed that the variation in delay ranges from -2.39% to 4.60% for 130nm technology [20].

Similarly, Monte Carlo simulations are run for threshold voltage variations in 70nm fabrication technology also. Figure 3 shows the SPICE input and output voltage variations for variation in threshold voltage for NMOS and PMOS transistors of the driver in 70nm technology. It is observed that the output varies appreciably higher than the results observed for 130nm technology due to the process variation parameter.

Table-II accounts for NMOS threshold voltage (*Vtn*); PMOS threshold voltage (*Vtp*); the delay due to driver and interconnect line; the percentage variation in NMOS and PMOS threshold voltage and percentage variation in delay of driver and line. It is observed that the variation in delay ranges from -9.13% to 7.49% for 70nm technology [20].

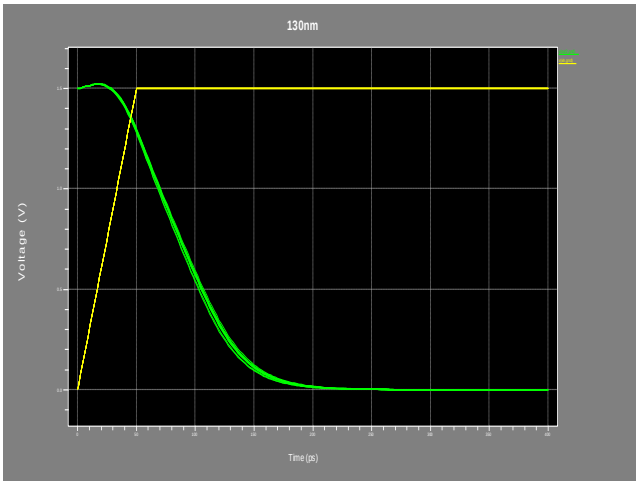

Figure 2 SPICE input and output waveform through DIL for 130nm technology Driver

Table I Variation in delay due to change in threshold voltage of NMOS & PMOS for 130nm process technology

Vtn (V)	*Vtp* (V)	Driver and Line Delay (ps)	Variation in *Vtn* (%)	Variation in *Vtp* (%)	Variation in Delay of Driver and line (%)
0.044	-0.218	59.88	-34.15	2.39	-2.39
0.049	-0.201	60.36	-26.11	-5.68	-2.23
0.064	-0.205	61.60	-4.06	-3.89	-0.47
0.066	-0.302	61.64	-0.48	41.53	-0.62
0.067	-0.213	61.82	0.00	0.00	0.00
0.070	-0.217	62.08	4.33	1.99	0.61
0.071	-0.173	62.28	6.21	-19.00	1.30
0.073	-0.191	62.40	8.81	-10.14	1.54
0.074	-0.145	62.61	10.54	-31.83	2.31
0.075	-0.233	62.56	12.72	9.22	1.87
0.085	-0.230	63.43	26.69	7.98	4.60

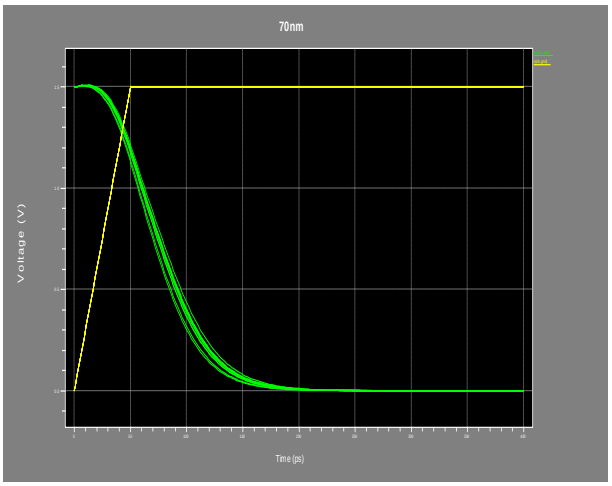

Figure 3 SPICE input and output waveform through DIL for 70nm technology Driver

Table II Variation in delay due to change in threshold voltage of NMOS & PMOS for 70nm process technology

Vtn (V)	*Vtp* (V)	Driver and Line Delay (ps)	Variation in *Vtn* (%)	Variation in *Vtp* (%)	Variation in Delay of Driver and line (%)
0.132	-0.225	44.91	-34.15	2.39	-9.13
0.148	-0.208	45.97	-26.11	-5.68	-6.97
0.192	-0.211	48.88	-4.06	-3.89	-1.10
0.199	-0.311	49.05	-0.48	41.53	0.75
0.200	-0.220	49.42	0.00	0.00	0
0.209	-0.224	50.01	4.33	1.99	1.20
0.212	-0.178	50.46	6.21	-19.00	2.09
0.218	-0.198	50.74	8.81	-10.14	2.66
0.221	-0.150	51.19	10.54	-31.83	3.57
0.225	-0.240	51.14	12.72	9.22	3.47
0.253	-0.238	53.12	26.69	7.98	7.49

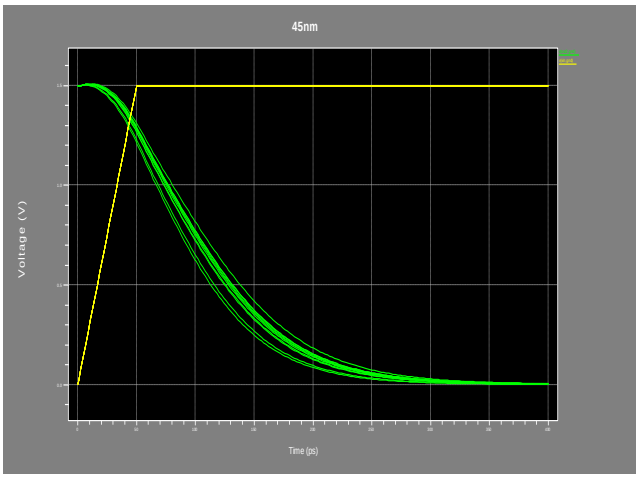

Figure 4 SPICE input and output waveform through DIL for 45nm technology Driver

Table III Variation in delay due to change in threshold voltage of NMOS & PMOS for 45nm process technology

Vtn (V)	*Vtp* (V)	Driver and Line Delay (ps)	Variation in *Vtn* (%)	Variation in *Vtp* (%)	Variation in Delay of Driver and line (%)
0.145	-0.225	63.06	-34.15	2.39	-13.90
0.163	-0.208	65.43	-26.11	-5.68	-10.60
0.211	-0.211	71.98	-4.06	-3.89	-1.70
0.219	-0.311	72.61	-0.48	41.53	-0.83
0.220	-0.220	73.22	0.00	0.00	0
0.230	-0.224	74.66	4.33	1.99	1.96
0.234	-0.178	75.6	6.21	-19.00	3.25
0.239	-0.198	76.33	8.81	-10.14	4.26
0.243	-0.150	77.26	10.54	-31.83	5.52
0.248	-0.240	77.41	12.72	9.22	5.73
0.279	-0.238	82.39	26.69	7.98	12.5

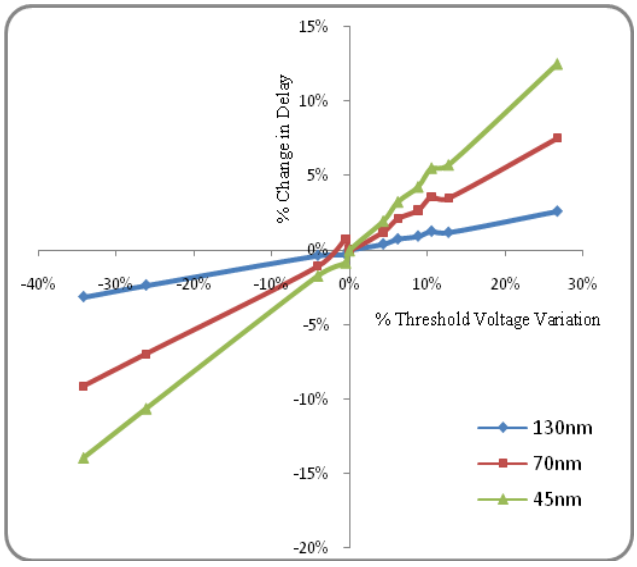

Figure 5 Comparison of percentage change in delay due to variations in threshold voltage for 130nm, 70nm and 45nm technologies.

Figure 4 demonstrates the Monte Carlo SPICE simulation input and output voltage variations due to variation in threshold voltage of NMOS and PMOS transistors of the driver in 45nm technology. It is observed that the output varies drastically due to the process variation parameter in 45nm technology compared to 130nm and 70nm technologies.

Table III accounts for NMOS threshold voltage (*Vtn*); PMOS threshold voltage (*Vtp*); the delay due to driver and interconnect line; the percentage variation in NMOS and PMOS threshold voltage and percentage variation in delay of driver and line. It is observed that the variation in delay ranges from -13.9% to 12.5% for 45nm technology [20].

The comparison between three technologies shows that as device size shrinks, the process variation becomes dominant and subsequently gives rise in variation of delays. Figure 5 demonstrates this claim by comparing the percentage change in delay due to variations in threshold voltage for 130nm, 70nm and 45nm technologies. It is observed that as feature reduces the variation in delay performance increases due to change in threshold voltage. Thus these simulation results reveals that process variation has large effect on the driver delay due to variation in threshold voltage.

4. CONCLUSION

Process variation represents a major challenge to design system-on-chip using nanometer technologies. In this paper, we have evaluated process variation effects on the delay of Driver-interconnect-load system due to threshold voltage variations. Variations in the driver and interconnect geometry of nanoscale chips deciphers to variations in their performance. The resulting diminished accuracy in the estimates of performance at the design stage can lead to a significant reduction in the parametric yield. Thus, determining an accurate statistical description of the DIL response is critical for designers. The random or systematic part of variations plays an important role in deviating electrical parameter. In the presence of significant variations of device model parameters the variations in performance parameter such as delay is severely affected. The comparison between three technologies shows that as device size shrinks the process variation becomes a dominant factor and subsequently raises the variation in delays.

5. REFERENCES

[1] Sai-Halasz G.A., "Performance trends in high-end processors", *Proceed. of IEEE*, vol.83, issue 1, pp. 20-36, Jan.1995.

[2] Sylvester D. and Wu C., "Analytical modeling and characterization of deep-submicrometer interconnect," *Proceed. of IEEE*, vol.89, issue 5, pp. 634-664, May 2001.

[3] Kaushik B.K. and Sarkar S., "Crosstalk Analysis for a CMOS-Gate-Driven Coupled Interconnects" IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems, vol. 27, no. 6, pp. 1150-1154, June, 2008.

[4] Orshansky, M. Milor, L., Chen, P., Keutzer K., and Hu, C., (2000) "Impact of systematic spatial intra-chip gate length variability on performance of high-speed digital circuits," ICCAD 2000, pp.62–67.

[5] Liu, Y., Nassif, S. R., Pileggi L. T. and Strojwas, A. J. (2000) "Impact of interconnect variations on the clock skew of a gigahertz microprocessor," DAC 2000, pp. 168–171.

[6] Mehrotra, V., Sam, S., Boning, L. D., Chandrakasan, A., Vallishayee, R. and Nassif, S. (2000) "A methodology for modeling the effects of systematic within-die interconnect and device variation on circuit performance," DAC 2000, pp. 172–175.

[7] Malavasi, E., Zanella, S., Min J. Uschersohn, C., Misheloff M. and Guardiani, C. (2002) "Impact analysis of process variability on clock skew," ISQED 2002, pp. 129–132.

[8] Brawhear, R. B., Menezes, N., Oh, C., Pillage, L. T., and Mercer, M. R., (1994) "Predicting circuit performance using circuit-level statistical timing analysis," DATE 1994, pp. 332–337.

[9] Chang, H., and Sapatnekar, S. S., (2003) "Statistical timing analysis considering spatial correlations using a single PERT-like traversal," ICCAD 2003, pp. 621–625.

[10] Agarwal, A., Blaauw, D., and Zolotov, V., (2003) "Statistical timing analysis for intra-die process variations with spatial correlations," ICCAD 2003, pp. 271–276.

[11] Acar, E., Nassif, S. N., Ying, L., and Pileggi, L. T., (2001) "Assessment of true worst case circuit performance under interconnect parameter variations," ISQED 2001, pp. 431–436.

[12] Borkar, S., Kamik, T., Narendra, S., Tschanz, J., Keshavarzi, A., and De, V., (2003) "Parameter variations and impact on circuits and microarchitecture," DAC 2003, pp. 338–342.

[13] Gattiker, A., Nassif, S., Dinakar, R., and Long, C. (2001) "Timing yield estimation from static timing analysis," ISQED 2001, pp. 437–442.

[14] Luong, G. M., and Walker, D. M. H., "Test generation for global delay faults," ITC 1996, pp. 433–442.

[15] Liou, J. J., Krstic, A., Wang, L. C., and Cheng, K. T., (2002) "False path-aware statistical timing analysis and efficient path selection for delay testing and timing validation," DAC 2002, pp. 566–569.

[16] Krstic, A., Wang, L. C., Cheng, K. T., and Liou, J. J., (2003) "Diagnosis of delay defects using statistical timing models," VTS 2003, pp. 339–344.

[17] Lu, X., Li, Z., Qiu, W., Walker, D. M. H., and Shi, W. (2004) "Longest path selection for delay test under process variation," ASP-DAC 2004, pp 99-103.

[18] Fabbro, A.D., Franzini, B., Croce, L., and Guardiani, C., (1995) "An assigned probability technique to derive realistic worst-case timing models of digital standard cells," DAC 1995, pp. 702–706.

[19] Vrudhula, S., Wang, J. M., and Ghanta, P., (2006) "Hermite Polynomial Based Interconnect Analysis in the Presence of Process Variations" IEEE Trans. on CAD of Integrated Circuits And Systems, vol. 25, No. 10, October 2006, pp. 2001-2011.

[20] Krishan Gopal Verma, Brajesh Kumar Kaushik and Raghuvir Singh, "Propagation Delay Variation due to Process Induced Threshold Voltage Variation" Communications in Computer and Information Science, 1, Information and Communication Technologies-2010, Springer, vol.101, Part 3, pp. 520-524.