Correlation analysis of CD-variation and circuit performance under multiple sources of variability

Amir Borna[a], Chris Progler[b], David Blaauw[a]
[a]University of Michigan, Ann Arbor, EECS Department, ACAL lab
[b]Photronics Inc., 901 Millennium Drive, Allen, TX 75013

ABSTRACT

Variability of digital integrated circuits is becoming an increasing concern with shrinking transistor geometries due to process scaling. As a result, the electrical properties of MOS devices can exhibit significant deviation from their design specifications, causing substantial variation in the performance of high-end designs. Lithography perturbations can affect a number of layout geometries, although the most critical parameter for circuit performance is the transistor channel length or Critical Dimension (CD). Key sources of CD variation include dose, focus, lens aberration and mask errors. In this paper, we compare the impact of above sources of CD variation on circuit performance. We present a new design analysis methodology which models the CD variation from each individual source in static timing analysis for different circuit blocks. Using this analysis capability, we study the impact of lithographic perturbations on block-level circuit performance for two adders. Furthermore, we study the correlation between the CD variability resulting from a lithographic perturbation source, and the resulting circuit performance variability. Through this analysis we determine the suitability of CD variability as an accurate predictor for circuit performance.

Keywords: Circuit Performance, Critical Paths, CD variation, Static Timing Analyzer

1. INTRODUCTION

The goal of semiconductor fabrication is to produce devices which meet their constraints while maintaining process complexity and keeping the overall process costs as low as possible. These devices are the smallest components of the ICs and their specifications are defined on the circuit design side in terms of performance, power and reliability. Manufacturing-induced perturbations impact both interconnects and devices and cause the circuit performance to shift from its intended value causing performance variation and consequently yield degradation.

A cross section of an NMOS device is depicted in Fig. 1; all the device parameters are prone to variability including L_{eff}, T_{ox}, V_{th}, W and the doping profile. These device parameters variations change the device properties and therefore affect the circuit performance [2]. Among these components, L_{eff} variations have the largest impact on circuit performance variations as depicted in Table 1 [1]. Since L_{eff} is the smallest feature implemented on the silicon, it has the largest intra-die variation due to OPE (Optical Proximity Errors) and other process-related effects. When considering the CMOS gate delay formula, $C_L V_{DD}/2I_D$, and the fact that I_D is strongly dependent on L_{eff} for Short Channel (SC) devices, it can be concluded that variations in L_{eff} result in substantial circuit performance fluctuations. Another effect depicted in Table 1 is the reduction in CD-variation-caused performance fluctuations in lower k regimes; with shrinking L_{eff} the introduced CD variability increases however according to short channel length effects I_D (drain current) dependency on L_{eff} will decrease [3], resulting in less performance variation from CD variability.

So far, the semiconductor fabrication has focused on maintaining the CD (Critical Dimension or MOS transistor channel length) variability statistics within an acceptable range to meet the design requirements and this has become a challenging issue with the new low k regimes. As depicted in Fig. 2 there are two types of CD variations: 1) Intra-die: Considers CD variations in a die and 2) Inter-die variation which models with CD variations across dies. Until now, considering the Intra-die variation as random component in corner-based performance- and yield-evaluations was adequate but in the nanometer era (low lithography k regime) the Intra-die variation is more dominant and treating this variability as a random component leads to pessimistic designs.

Design and Process Integration for Microelectronic Manufacturing III, edited by Lars W. Liebmann, Proceedings of SPIE Vol. 5756
(SPIE, Bellingham, WA, 2005) · 0277-786X/05/$15 · doi: 10.1117/12.604606

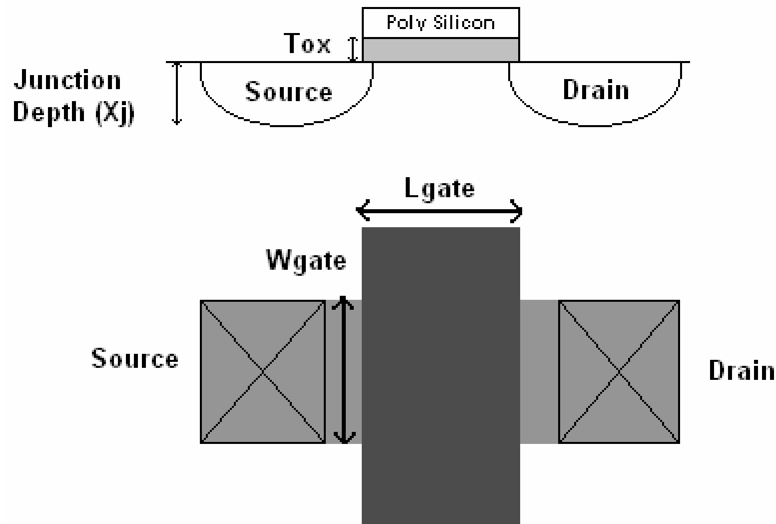

Fig. 1: CMOS a) Cross Section b) Top View

	1997	**1999**	**2002**	**2005**	**2006**
V_{dd}	9.5	10.8	10.0	9.5	8.9
Device					
T_{ox}	1.3	2.5	3.2	3.9	4.9
V_t	3.8	5.3	5.5	6.5	7.2
L_{eff}	32.4	28.3	25.5	24.6	23.8
Wire					
W	13.3	12.0	11.7	11.4	10.5
S	9.3	9.4	9.9	9.5	9.4
T	6.8	7.0	8.0	8.2	8.2
H	7.8	8.0	8.1	8.3	7.1
ρ	16.0	16.6	17.9	18.4	20.1

Table 1: Impacts of circuit parameter variations on circuit performance variation [1]

Inter-die variations are independent of designs and considering them as random components is still the main method to model them. Finally, systematic Intra-die variations are directly related to designs layouts and therefore predictable. The systematic component of Intra-die CD variation is caused by stepper-induced illumination which is getting worse with shrinking transistor sizes. This type of CD variation is systematic and predictable once the layout data is available [8]. CD variability comes from several sources which can be categorized into three distinct types: mask errors, optical errors (lithography errors) and process errors. To keep the CD within the acceptable range in presence of these multiple error sources, the process should be controlled and the process window computation methods serve for this purpose.

Assessing the impact of process variation induced CD perturbations on circuit performance involves two main tasks: Perturbing the circuit and Performance evaluation. There are two approaches to perturb the circuit: 1) Lithography simulation: Off-the-Shelf lithography simulation tools are used to predict the fabricated transistor gates on the silicon as a perturbed layout layer; the perturbed layout layer is later used in netlist extraction. This approach is the most accurate one but is computationally expensive 2) Categorizing transistor gates: in this method the transistor gates are tagged based on the neighboring distances and new values are assigned to tagged gates in the original netlist to represent the perturbed netlist. Also, it is possible to categorize the standard library cells based on their neighboring distance to other cells and characterize the timing of the tagged cells. Static Timing Analyzers can then be used in performance evaluation provided the library cells are characterized [12]. If the perturbed layout is represented as a netlist, a device level circuit simulator can be used in performance evaluation which is computationally expensive. Stine and Boning used aerial image simulation and device level circuit simulator to study the performance fluctuations [9]; M. Orshansky et.al.

approached the problem by categorizing the CDs based on their neighboring distances and modifying the circuit netlist by the new CD values [8].

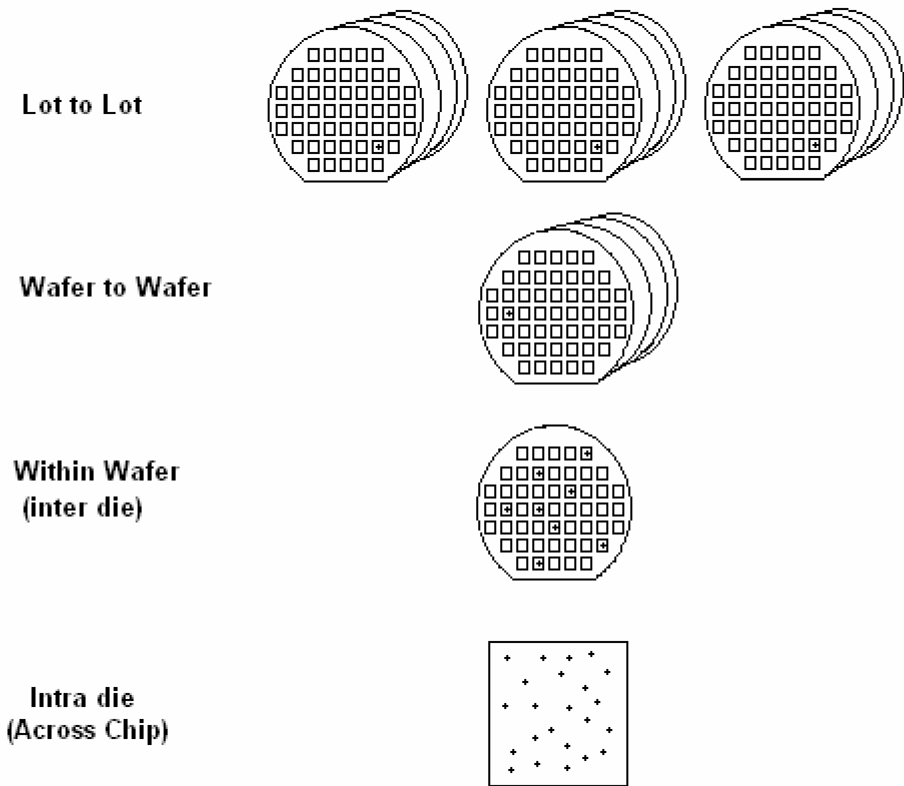

Fig 2: Inter-die and Intra-die CD variations

In Section 2, a methodology is proposed to consider the CD variability from different lithography (optical) error sources on performance evaluation at early design phase; through this methodology the impact of multiple variability sources on performance for two designs is evaluated and the results are presented in Section 3. Section 4, concludes the results and includes suggestions for future works.

2. METHODOLOGY

As shown in Fig. 3 the analysis flow has three components:

1) Perturbed Layout Generator
2) Perturbed Layout Netlist Extractor
3) Timing Analyzer

We describe each component in more detail below.

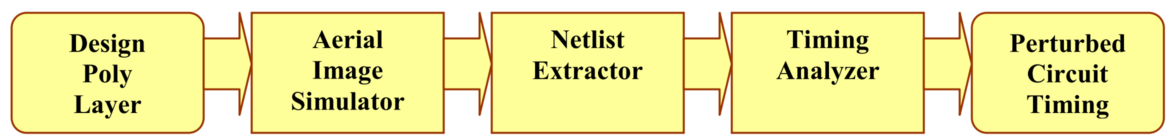

Fig 3: The main components of analysis flow

2.1. Perturbed Layer Generator
This component involves two tasks:

1) Generating the perturbed poly layer
2) Replacing the original poly layer with the perturbed poly layer

Caliber RET (Mentor Graphics) is used for aerial image simulation and Virtuoso layout editor (cadence) are the tools for merging the layouts. For each of the process variability sources, the associated variable in the aerial image simulator was varied across a certain range to yield CD variation bounded by sigma of $10\%CD_{nominal}$ and the silicon image was simulated. The silicon image in gds2 format was used as an input to layout merger and SKILL language was used to merge the layouts.

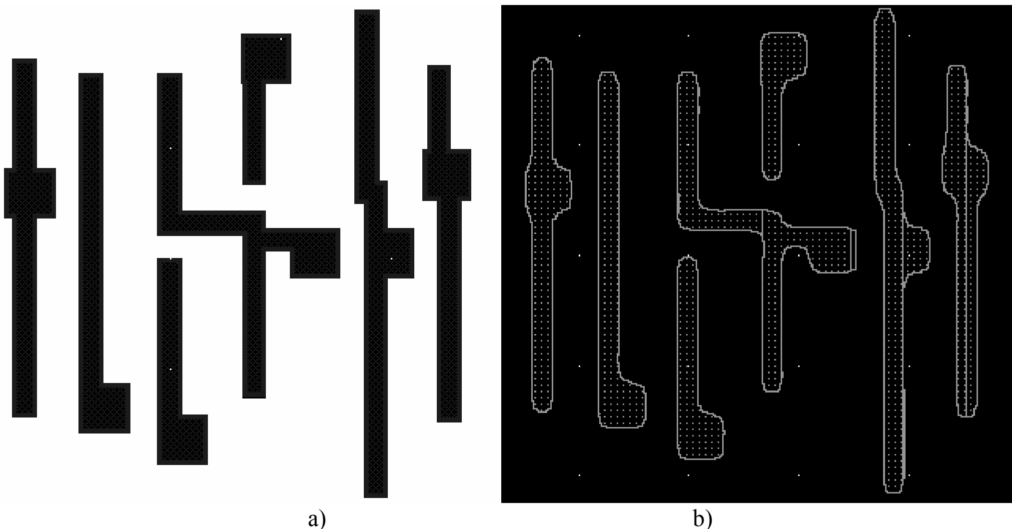

Fig 4: a) The original poly layer b) The simulated Aerial Image

2.2. Netlist Extraction
Diva Extraction (cadence) was used for netlist extraction. However, it takes the average of the gate end-points as the channel length (or equivalently CD) which is sufficient for the ideal layout. However, if there is CD non-uniformity across the channel this method can result in up to 8% CD difference when comparing to the gate of the transistor composed of parallel transistors (the ideal but expensive method to compute the transistor channel length). For our purpose we modified the device extraction subroutine of the extraction script to compute the average channel length by considering the area of the poly gate on active area instead of taking the average of the gate end-points and our method was within 1% accuracy of the ideal method of device extraction.

2.3 Timing Analyzer
Commercially available CAD tools are used to develop two timing analyzers (TA1 and TA2) for performance evaluation of the perturbed netlist. The main difference between TA1 and TA2 is the delay model which they employ; TA1 uses a gate delay model while TA2 employs a path delay model.
A block diagram of TA1 is depicted in Fig 5. It has the technology file and the circuit device-level netlist as the inputs and employs a device-level circuit simulator as its main engine (spectre from cadence). TA1 has three phases: In first phase the flat device-level netlist is broken into individual CMOS gates. In the second phase the timing of CMOS gates generated in first phase is computed. To characterize the timing behavior of gates, all the CMOS gates are simulated by the device-level circuit simulator once their input waveforms are known. The third phase divides the design into timing paths by the algorithm presented in [4] and reports all the path delays.

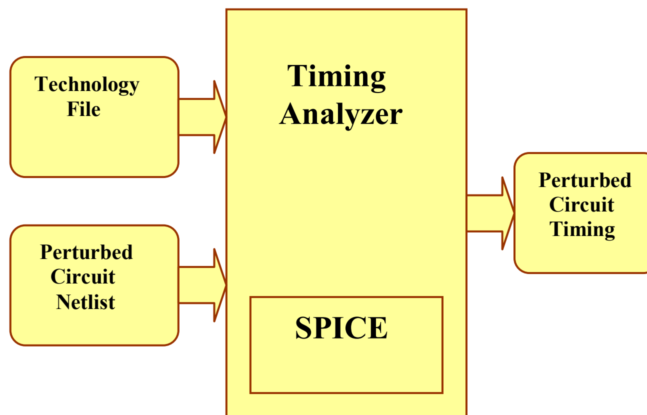

Fig. 5: TA1 block diagram

TA2 is shown in Fig 6; a set of primary input vectors for sensitizing the true critical paths is computed by a commercial Static Timing Analyzer (STA) [5] for a gate-level design netlist. This forms the input to the timing analyzer; the other inputs are the perturbed netlist and the technology file. By running the device-level simulator for all the input vectors, the delays for each of the main critical paths are computed and the maximum of these delays represents the circuit performance. Unlike TA1 which reports both true and false paths TA2 only considers the true critical paths.

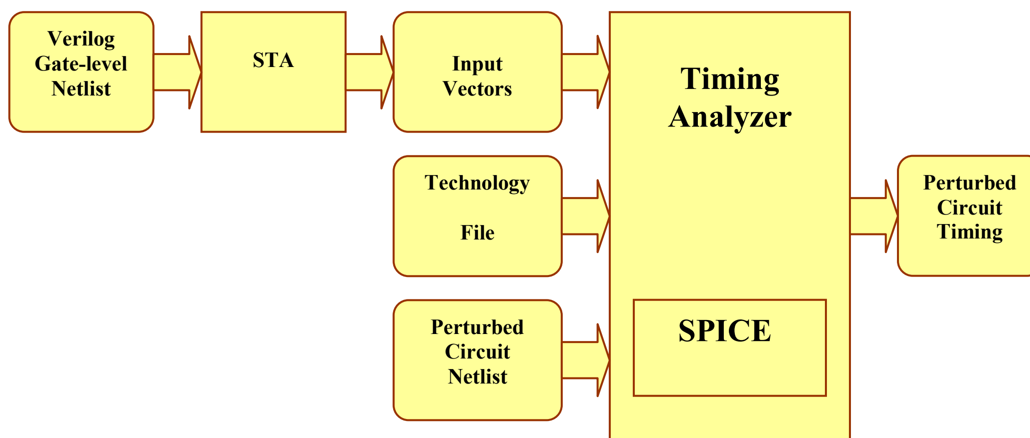

Fig. 6: TA2 block diagram

The upper bound on the number of input vectors is limited by run-time of the timing analyzer which is of critical importance, since it is used in Monte-Carlo simulation. The lower bound on this number is the desired accuracy; one might argue that the delay of the circuit might be from a path not considered in the input vector set, but our results show that there is a set of critical paths which can accurately present the circuit performance over a wide range of CD variation.

To compare these two timing analyzers, a 16-bit adder poly layer was perturbed by dose induced CD variation and the modified circuit was analyzed by both TA1 and TA2. There are two distinct differences. First as depicted in Fig. 7, the circuit delays computed by TA1 are larger than those computed by TA2. This mismatch stems from the fact that TA1 considers both true and false critical paths and false critical paths have larger delays than true critical paths. The second difference between the circuit timings computed by TA1 and TA2 was in the paths representing the circuit delay. Unlike TA2, in TA1-computed circuit delays, the path which is the worst critical path of the design is not fixed in rank and therefore a set of critical paths is required to compute the circuit performance; the input vector set size depends on the induced CD variation and by increasing the CD variation the set size should increase as well to maintain the same timing

precision. In the extreme case when all the CDs are the same as $CD_{nominal}$ the CD variation is zero and only the worst critical path is sufficient to represent the circuit performance. For TA2-computed circuit delays the worst critical path is always fixed and TA2 is used to obtain the results presented in the next section.

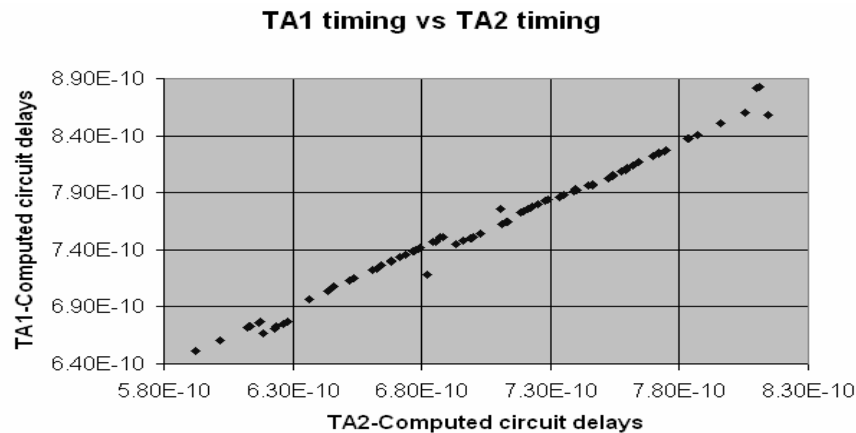

Fig. 7: Comparison between circuit delays computed by TA1 and TA2

3. RESULTS

A 16-bit adder and a 4-bit adder are the Designs Under Test (DUT) studied under multiple sources of variability in k-regimes of 0.31 and 0.46. Variability sources considered in this work are from the optical part of the lithography and include: dose, focus and lens aberration (astigmatism and coma). Dose or Exposure Dose are the amount of light source energy generated by the laser pulse; the precision of this parameter can be found in the stepper specifications and varies from stepper to stepper. Dose can vary from chip-to-chip or within chip; there are two reasons for within-chip dose variation: 1) Due to laser performance the dose along the exposure slot can vary which results in within-chip dose errors 2) Scan errors which happen along the scan direction can also result in within-chip dose errors. Chip-to-chip dose errors are due to the fact that the laser pulse generator is not ideal and there are laser pulse-to-pulse variations. Within-chip focus error components are field tilt, level sensor precision, total focal plane deviation (TFPD), reticle non-flatness, chuck/wafer non-flatness, and underlying wafer topography. Chip-to-chip focus error components are auto-focus precision, wafer stage non-flatness [6]. Lens aberration stems from different optical path lengths due to non-perfect lenses; this type of non-uniformity diminishes over time with improving lens manufacturing. CD 3sigma variation doesn’t reflect circuit timing accurately [13] therefore to study the correlation between circuit performances and die CD statistics two populations of CDs were considered: CD statistics, e.g. average and standard deviation, of the critical paths and CD statistics of the whole design; further for each CD population statistics, the correlation with circuit performance were computed and the correlation coefficients are presented in Table 2.

For 16-bit adder, critical path has 7% of design transistors whereas for 4-bit adder critical path is composed of 86% of total transistors; since the circuit performance is dictated by critical path devices, in most cases, the 16-bit adder critical path CD statistics correlate with performance slightly better than global CD statistics. However, this improvement is insignificant. In the 4-bit adder, the difference between critical path CD statistics and that of total transistors are negligible. Also, as can be seen among the three CD statistics metrics in table 2, mean CD has always the best correlation with performance (close to one).

For k regime of 0.31 and dose variation, the 16-bit adder performance fluctuation is depicted in Fig. 8; in this figure the reason behind different correlation coefficients of performance with CD statistics is depicted.

As mentioned in the previous section, considering merely the true critical paths in performance evaluation has the advantage of having only or equivalently input vector) representing the circuit delay; therefore in all of the charts of Fig. 8 the circuit performance of the 16-bit adder is represented by only one particular path; the same holds true for 4-bit adder.

			Mean CD		StdDev_mean		StdDev_nominal	
Perturbation Type	**K regime**	**Design**	**Critical Path**	**Global**	**Critical Path**	**Global**	**Critical Path**	**Global**
dose	**0.46**	Adder16	0.9912	0.9911	0.8253	-0.9629	0.3518	0.2588
		Adder4	0.9999	0.9998	0.7893	0.7881	0.8957	0.8847
	0.31	Adder16	0.9936	0.9938	0.9900	-0.4446	0.1939	-0.6720
		Adder4	0.9981	0.9987	-0.9334	-0.9395	-0.6435	-0.6055
focus	**0.46**	Adder16	0.9797	0.9814	-0.9660	-0.9630	-0.9396	-0.9613
		Adder4	0.9942	0.9989	-0.9432	-0.9496	-0.9828	-0.9849
	0.31	Adder16	0.9959	0.9969	0.4873	0.3308	0.8750	0.1209
		Adder4	0.9891	0.9971	-0.8837	-0.8865	-0.9946	-0.9966
Astigmatism	**0.46**	Adder16	0.9969	0.9967	-0.9106	-0.8475	-0.9959	-0.9961
		Adder4	0.9947	0.9962	0.3854	0.3895	-0.9663	-0.9645
	0.31	Adder16	0.9948	0.9971	0.6777	0.7196	0.8220	0.4405
		Adder4	0.9745	0.9781	0.3389	0.3334	0.6963	0.6957
Coma	**0.46**	Adder16	0.9952	0.9957	-0.7154	-0.7177	-0.9833	-0.9891
		Adder4	0.9936	0.9944	-0.5599	-0.5656	-0.1342	-0.1353
	0.31	Adder16	0.9906	0.9959	0.2511	0.2662	0.8569	0.8355
		Adder4	0.9918	0.9935	0.0894	0.0812	0.4709	0.3463

Table 2: Correlation coefficients of CD statistics with circuit performance; global contains the CD statistics of all the gates and critical path points to that of critical path transistors; Mean CD is the average of the CD populations; StdDev_mean is the standard deviation computed with respect to average CD populations and StdDev_nominal is the standard deviation of the CD populations computed with respect to nominal (ideal) CD value.

As the process develops, by applying RET tricks, the image contrast for small apertures and the photoresist chemical reactions improve. The residuals of exposure dose/focus and other variable process parameters are taken care of, by applying Optical Proximity Correction (OPC) techniques to design layout. The current OPC techniques are mostly model based and are optimized for a set of exposure dose/focus and etching conditions. The objective of these model-based OPC techniques is to modify the layout geometries to match the simulated printed features on silicon with their original design layout counterparts. The standard-deviation of total CD population is used to guide the process development and OPC; according to data in Table 2 the timing is weakly correlated with CD standard deviation but has a strong correlation with mean of the total CD population. Having this in mind we developed a simplified OPC technique which does not improve Edge Placement Error (EPE) of each feature individually but attempts to match the mean length of all the transistors channels with the nominal transistor channel length of the technology. The OPC algorithm is depicted in Fig. 9.

In this diagram, the original layout poly silicon layer is given to the aerial image simulator; based on the given process parameters the perturbed poly silicon layer is simulated and the netlist extractor uses this perturbed poly silicon layer to generate the perturbed netlist. From the perturbed netlist CD statistics are computed and compared against the nominal technology CD; if they match the iterative flow stops otherwise from the mismatch of the mean CDs computed in the current and previous iterations a step size (negative or positive) is computed which all the gate polygons should be modified according to the step size. For some cases, especially low k regimes, the mean CD might diverge with each iteration; the adaptive step size calculator can detect these cases and compute the step size accordingly. The flow iterates until the mean CD of the perturbed netlist is within an acceptable distance from the nominal technology defined CD. Also for each netlist the circuit performance is computed which is not a part of the flow and drawn with thin line.

This OPC is applied to a 16-bit adder in lithography regime of 0.46 with defocus induced perturbation. The results are shown in Fig 10; as it can be seen in Fig 10-a, the mean CD iterates until it converges to the nominal CD. Also the standard deviation (computed with respect to nominal CD) of the design reduces with each iteration (Fig 10-b). For the first 2 iterations the mean CD is off by 44% and therefore these first two iterations computed timings can be excluded

from the results; also it is worthwhile to note that for these first two iterations the critical path of the design was not fixed as the other cases. The circuit performance excluding first two iterations is depicted in Fig. 11.

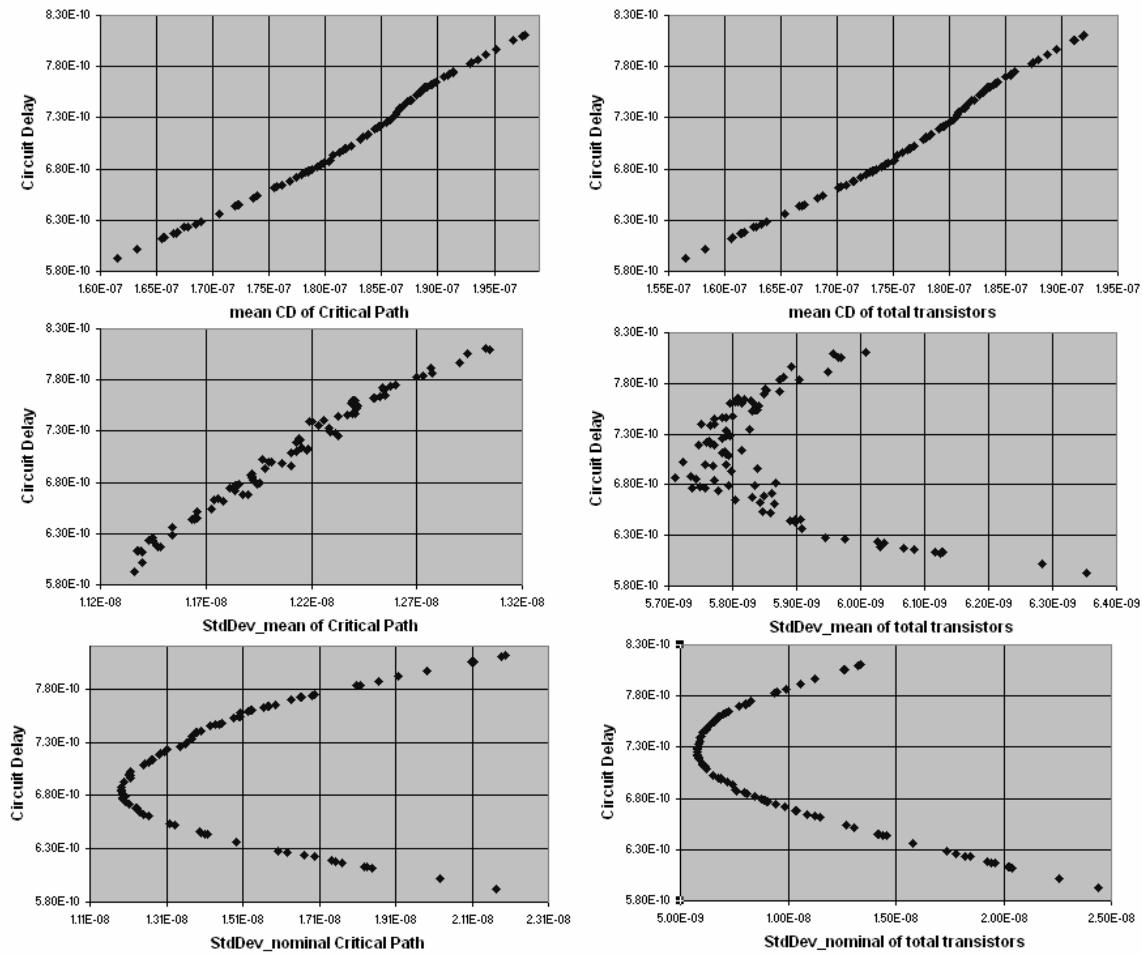

Fig. 8: 16-bit adder performance fluctuations with dose variations and k regime of 0.31

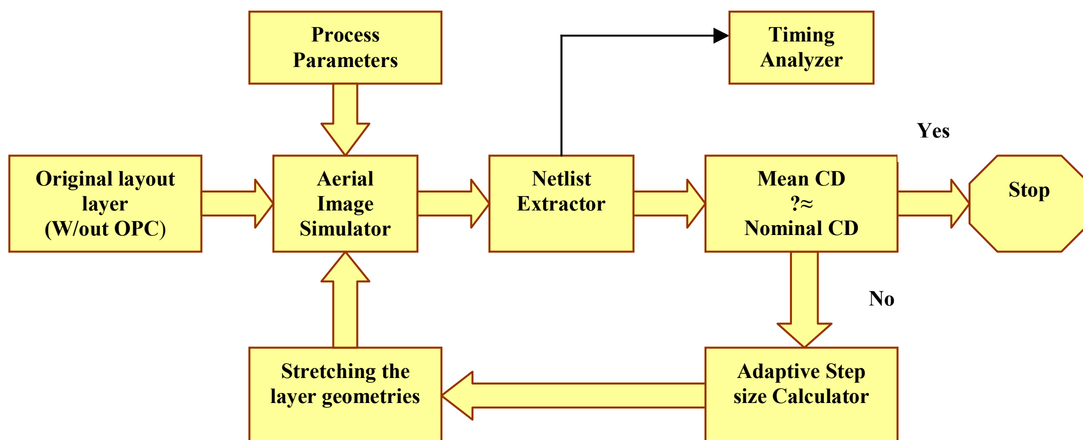

Fig. 9: The mean CD guided simplified OPC diagram

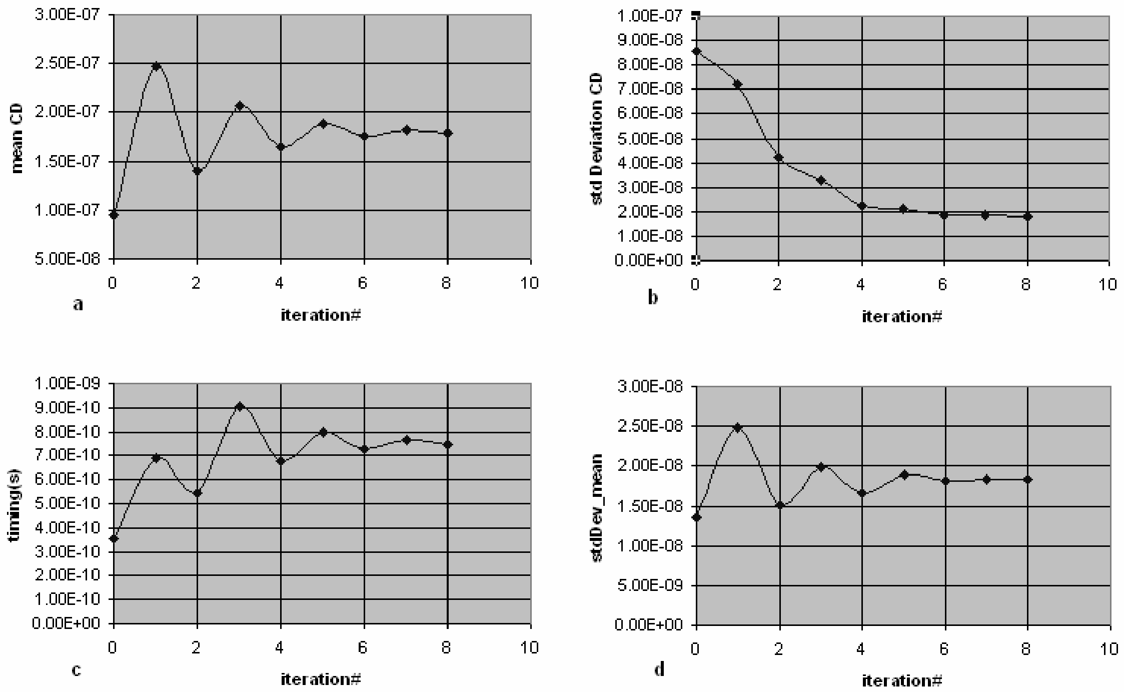

Fig 10: The results of moving the mean channel length for adder16, k=0.46 and defocus induced perturbation. a) The mean length of CDs b) standard deviation of CDs w/ respect to nominal CD c) circuit performance d) standard deviation of CDs w/ respect to mean Length

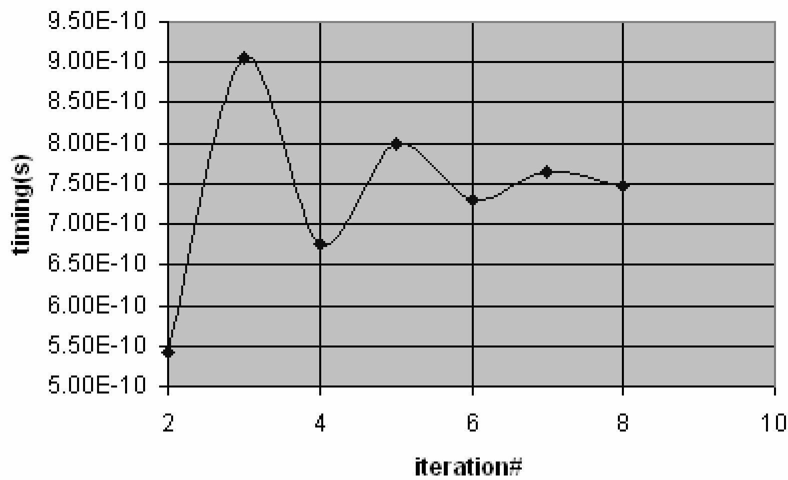

Fig. 11: Circuit performance excluding iterations 0 and 1. adder16, k=0.46, defocus

The correlation coefficients of timing and CD statistics excluding and including the first two iterations are shown in Table 3.

	Corr(performance, mean CD)	Corr(performance, stdDev CD)
Including iterations 0 and 1	0.76	-0.70
Excluding iterations 0 and 1	0.99	-0.37

Table 3: Correlation coefficients of timing and CD statistics of design modified with simplified OPC

Therefore it can be concluded that independent of the OPC algorithm used, the circuit performance follows the mean CD of the total transistors population; this fact can be used in OPC algorithms since from a circuit performance point of view it might not be necessary to have all the features printed perfectly on the silicon, therefore potentially reducing mask cost.

4. CONCLUSION

A new methodology for performance evaluations of digital ICs in presence of process variation is proposed; employing the introduced methodology the Circuit-Performance CD-Statistics correlation coefficients, for two K regime and two different designs, are computed. According to the results, circuit performance is best correlated with mean CD of the total transistors and the critical path CD statistics have only slightly improved correlation with performance which is negligible. It is shown that for induced-perturbation circuit delay can be represented by only one true critical path; the same does not hold true if both true and false paths are considered for performance evaluation. A simple OPC-like algorithm is introduced which attempts to match the mean length of the transistors channels with nominal CD; though individual transistors gates polygons are not perfectly printed on the silicon (which is the object of the current model-based OPC techniques), the mean length of the total transistors channels is close to the nominal CD and consequently the circuit performance is in the vicinity of unperturbed design performance. It reveals that to improve the circuit performance it might not be necessary to have all the polygons printed as designed on the wafer, which results in lower OPC-run time and mask costs.

Acknowledgment

The authors would like to thank Dr. Andres Torres from Mentor Graphics Inc. for his valuable comments throughout this work.

REFRENCES

[1] Sani R. Nassif, “Design for Variability in DSM Technologies,” ISQED Proceedings, pp. 451-454, March 2000

[2] Vikas Mehrotra, “Modeling the effects if systematic process variation on circuit performance”, PhD thesis May 2001, MIT.

[3] D. Sylvester, K. Keutzer, “Getting to the Bottom of Deep Submicron,” ICCAD, pp. 203-211, Nov 1998

[4] S. H. Yen, D. H. Du, S. Ghanta, “Efficient Algorithm for Extracting the K Most Critical Paths in Timing Analysis,” DAC Proceedings, pp. 649-654, 1989

[5] PrimeTime User Manual, Synopsys Inc., 2003

[6] Anatoly Bourov, Sergei V. Postnikov, and Kevin Lucas, “Lithographic process window analysis by statistical means,” Proc. SPIE Int. Soc. Opt. Eng., Vol 4689, pp. 484-491, July 2002

[7] A. Chandrakasan (Editor), W. J. Bowhill (Editor), F. Fox (Editor), “Design of High-Performance Microprocessor Circuits,” Wiley-IEEE Press (September 26, 2000)

[8] M. Orshansky, L. Milor, P. Chen, K. Keutzer, C. Hu, “Impact of systematic spatial Intra-chip gate length variability on performance of high speed digital circuits,” Proceedings of the 2000 IEEE/ACM international conference on Computer-aided design

[9] B. Stine, D. S. Boning, J. E. Chung, D. J. Ciplickas, J. K. Kibarian, “Simulating the impact of pattern dependent poly-CD variation on circuit performance”, IEEE transactions on semiconductor manufacturing, VOL. 11, NO. 4, Nov. 1998.

[10] calibre RET user manual, Mentor Graphics Inc.

[11] SKILL user manual, cadence Inc.

[12] Puneet Gupta, Fook-Luen Heng, “Toward a systematic-variation aware timing methodology,” DAC Proceedings, 2004

[13] C. Progler, A. Borna, D. Blaauw, P. Sixt, “Impact of lithography variability on statistical timing behavior”, SPIE Proceedings, Feb. 2004