



**HAL**  
open science

# Statistical Timing Analysis using Levelized Covariance Propagation

Kunhyuk Kang, Bipul C. Paul, Kaushik Roy

► **To cite this version:**

Kunhyuk Kang, Bipul C. Paul, Kaushik Roy. Statistical Timing Analysis using Levelized Covariance Propagation. DATE'05, Mar 2005, Munich, Germany. pp.764-769. hal-00181203

**HAL Id: hal-00181203**

**<https://hal.science/hal-00181203>**

Submitted on 23 Oct 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Statistical Timing Analysis using Levelized Covariance Propagation \*

Kunhyuk Kang, Bipul C. Paul and Kaushik Roy

Dept. of Electrical and Computer Engineering, Purdue University, West Lafayette, IN-47907, USA  
{kang18, paulb, kaushik}@ecn.purdue.edu

## Abstract

*Variability in process parameters is making accurate timing analysis of nano-scale integrated circuits an extremely challenging task. In this paper, we propose a new algorithm for statistical timing analysis using Levelized Covariance Propagation (LCP). The algorithm simultaneously considers the impact of random placement of dopants (which makes every transistor in a die independent in terms of threshold voltage) and the spatial correlation of the process parameters such as channel length, transistor width and oxide thickness due to the intra-die variations. It also considers the signal correlation due to reconvergent paths in the circuit. Results on several benchmark circuits in 70nm technology show an average of 0.21% and 1.07% errors in mean and the standard deviation, respectively, in timing analysis using the proposed technique compared to the Monte-Carlo analysis.*

## 1. Introduction

Timing verification is necessary for efficient integrated circuit design to achieve a desired performance. Conventionally, *static* timing analysis [9] is used for the timing verification of digital integrated circuits. However, as the scaling of technology continues, *static* timing analysis fails to provide a realistic timing information [14]. This is mainly due to the ever-increasing variability in the process parameters such as channel length, transistor width, oxide thickness and the random placement of dopants in the channel.

Process variations can be classified into two main categories - intra- and inter-die variations. Due to inter-die variations, the same device on a die can have different characteristics across different dies (i.e., dies from one wafer, from wafer to wafer, and from wafer lot to wafer lot). Intra-die variations, on the other hand, are the variations of transistor characteristics within a single die. As we are moving towards sub-100nm technology regime, intra-die process variations are becoming increasingly significant [4]. Intra-die variations in length, width and oxide thickness, are expected to exhibit some spatial correlations among devices located close to each other (systematic variation). For example, transistors close to each other are expected to have similar parameter variation than the ones that are far apart in a die. The effect of random placement of dopants (threshold voltage variation), on the other hand, is different in every transistor irrespective of their spatial location [13]. In addition to this, signal correlation due to the reconvergent paths under process variation also plays a significant role in circuit timing.

An accurate timing analysis should therefore, simultaneously consider the three issues, namely; (1) spatial correlation of the process parameters such as channel length, transistor width and the oxide

thickness, (2) random placement of dopants, and (3) the signal correlation due to the reconvergent paths. Several efforts have been made to independently model these effects to provide algorithms called *Statistical Timing Analysis*, to obtain the timing information of circuits under process variation. However, to the best of our knowledge, no algorithm is available for timing analysis considering all the above three effects simultaneously. For example, in [15, 8, 3], intra-die variations of process parameters are modeled as independent random variables (without spatial correlation). Chang et.al. [6], proposed a PERT-like traversal algorithm for statistical timing analysis considering the spatial correlation of process parameters as well as the signal correlation due to reconvergent paths. In this technique, the correlated process parameters are transformed into a set of uncorrelated principle components to compute the circuit delay. A block-based statistical timing analysis method is also proposed in [10] to obtain circuit delay under process variation considering both the spatial correlation and the signal correlation due to reconvergent paths. However, none of the above methods consider the effect of random placement of dopants, spatial correlation of process parameters and the signal correlation due to reconvergent paths simultaneously. Neglecting any of these effects may result in unrealistic timing information of the circuit [5].

In this paper, we propose a new statistical timing algorithm that simultaneously considers the above three effects. We employ the grid based model proposed in [2] to incorporate the spatial correlation of process parameters such as length, width and oxide thickness. The effect of random placement of dopants is considered by representing the threshold voltage of each transistor as an independent random variable. The circuit delay is calculated considering the above effects using *Levelized Covariance Propagation (LCP)*. In this technique, the *timing information* at the output of all gates in a single logic level is calculated and propagated to the next logic level. The *timing information* includes the arrival time (mean and variance) at the output of each gate in a particular logic level and the covariance among different outputs. The signal correlation due to reconvergent paths is also considered while calculating the *timing information* at a particular logic level. We use the proposed algorithm to calculate the circuit delays of several ISCAS benchmark circuits under process variation and compared the results with the Monte-Carlo analysis.

The rest of the paper is organized as follows. In section 2, we discuss the modeling of process variation. Section 3 describes gate delay model considering the effect of process variation. In section 4, the proposed algorithm is discussed in detail. Section 5 presents the experimental results on several ISCAS benchmark circuits.

## 2. Modeling the Process Variation

In this work, we consider channel length, transistor width, oxide thickness and threshold voltage as random process parameters which

\*The work is sponsored in part by Marco Gigascale Systems Research Center (GSRC), Intel Corp. and Semiconductor Research Corp. (SRC).

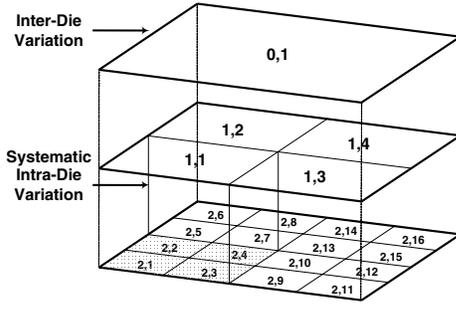


Figure 1. Rectangular Grid Model [2]

can be represented by a normal distribution [11]. We also assume that the variation in the above parameters are mutually independent. The process parameters can be divided into two groups based on their statistical behavior. The parameters such as the length, width and the oxide thickness will exhibit spatial correlations among different transistors close to each other. Therefore, for a group of transistors depending on their proximity, a single random variable can be used to represent the variation in the above process parameters. The threshold voltage, on the other hand, will be different for each transistor in a die due to the effect of random placement of dopants. Hence, separate random variables are to be used to represent the threshold voltages of all the transistors.

The following subsections explain how we model the spatial correlation and the random placement of dopants.

### 2.1. Modeling Spatial Correlation

To capture the effect of spatial correlation of process parameters among different transistor locations, we use the rectangular grid model proposed in [2]. In this model, the die area is divided by a multi-level quad-tree partitioning as shown in Fig. 1. All the transistors in a particular grid are assumed to experience the same parameter variation. The tree structure forms a hierarchical relationship among all the grids. A grid is defined to be a *parent* of a lower level grid if they share a common region. All the transistors in a grid at the bottom most level experience the same parameter variations. The spatial correlation of parameter variation among transistors in different grids is governed by their *parent* grids.

For every grid (at all levels), we maintain a single random variable for each process parameter. The parameter variation of a transistor in any grid at the bottom most level is represented by the sum of the variation in that particular grid and the variations in all its *parent* grids. For example, the variation of channel length of transistors in grid [2,1] (at the bottom most level (Fig. 1)) is represented as,

$$L(2, 1) = L_0 + \Delta L_{2,1} + \Delta L_{1,1} + \Delta L_{0,1} \quad (1)$$

where  $L(2, 1)$  is the random variable which represents the variation in the channel length in grid [2,1].  $L_0$  is the nominal value of the channel length.  $\Delta L(2, 1)$ ,  $\Delta L(1, 1)$  and  $\Delta L(0, 1)$  represent the deviation in channel length in grids (2, 1), (1, 1) and (0, 1) respectively.

The covariance of the process parameters between different grids is obtained depending on their common *parent* grids. For example, the random variables representing the channel length in grid [2,1], [2,2] and [2,16] are expressed as,

$$\begin{aligned} L(2, 1) &= L_0 + \Delta L_{2,1} + \Delta L_{1,1} + \Delta L_{0,1} \\ L(2, 2) &= L_0 + \Delta L_{2,2} + \Delta L_{1,1} + \Delta L_{0,1} \\ L(2, 16) &= L_0 + \Delta L_{2,16} + \Delta L_{1,4} + \Delta L_{0,1} \end{aligned}$$

In this example,  $L(2, 1)$  and  $L(2, 2)$  share two parent grids  $\Delta L_{1,1}$  and  $\Delta L_{0,1}$ , while  $L(2, 1)$  and  $L(2, 16)$  share only  $\Delta L_{0,1}$ . Thus, the covariance between  $L(2, 1)$  and  $L(2, 2)$  will be two times that of the covariance between  $L(2, 1)$  and  $L(2, 16)$ . For the case of 3 level grid structure shown in Fig. 1, a  $16 \times 16$  covariance matrix is necessary for each process parameter to represent the correlation between any two grids in the bottom most level. Also note that, since the top-most level is the *parent* grid of all the grids, it represents the inter-die variation. While the intra-die variation of the process parameters is represented by all other levels of grids. Hence, by using this model, intra- and inter-die variations can be considered simultaneously.

### 2.2. Random Placement of Dopants

In scaled CMOS devices, there exists a statistical fluctuation in the number of dopants in the channel, which can be translated into a threshold-voltage variation [13]. This discrete dopant effect on threshold voltage variation is incorporated by employing the following equation [13] into the simulator while computing the variation of threshold voltage of each transistor.

$$\sigma_{V_{th}} = \frac{q}{C_{ox}} \sqrt{\frac{N_a W_{dm}^0}{3LW}} \quad (2)$$

$\sigma_{V_{th}}$  represents the standard deviation of threshold variation due to random placement of dopants,  $q$  is electron charge,  $C_{ox}$  is the oxide capacitance,  $N_a$  is the substrate doping concentration, and  $W_{dm}^0$  represents the maximum depletion layer width.  $L$  and  $W$  are the channel length and width of the transistor, respectively.

Considering the effect of random placement of dopants, threshold voltages of all the transistors are represented by independent random variables with normal distribution. The variance of each random variable is obtained from Eq. (2).

### 3. Modeling the gate delay

Usually gate delay is a complicated non-linear function of the process parameters. In this work, we use first order Taylor's series expansions [6, 15] to approximate the delay distribution as follows,

$$D = D_0 + \sum_i^n s_i X_i \quad (3)$$

$D_0$  represents the nominal value of the gate delay ( $\mu_D$ ) without variation. We use Sakurai's gate delay model [12] to obtain the nominal delay of each gate in the circuit.  $X_i$ 's are the sources of variation, which are modeled as normal random variables.  $X_i$  can be any process parameter such as channel length, width, oxide thickness and the threshold voltage, that affects the delay of a specific gate.  $s_i$  is the sensitivity of the gate delay to the variation in process parameter  $X_i$ . All  $s_i$ 's are extracted through HSPICE simulation by varying the process parameters from their nominal value.

The variance of the delay of a gate can be obtained as,

$$\sigma_D^2 = \sum_i^n s_i^2 \sigma_{X_i}^2 \quad (4)$$

where  $\sigma_{X_i}$  is the standard deviation of the random variable  $X_i$ .

### 4. Statistical Timing Analysis

The major challenge in statistical timing analysis is to incorporate the effect of signal correlation due to process variation. The signal

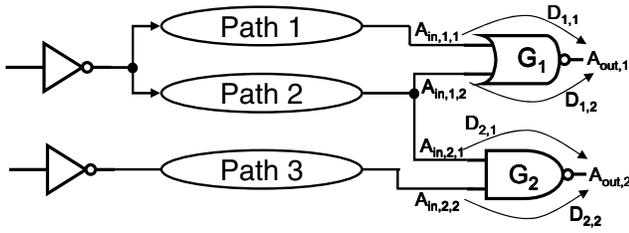


Figure 2. Example of signal correlation

correlation occurs due to both spatial correlation of process parameters and the random placement of dopants (affects the signal correlation on reconvergent paths). Hence, estimating the correlation between any two signals in a circuit will require a computation involving a large number of random variables which represent the process variation. While, the number of variables representing the variation in length, width and oxide thickness with spatial correlation can be limited by using the rectangular grid model explained in section 2.1, variation in threshold voltage (due to random placement of dopants) has to be represented by a large number of random variables (equal to the total number of transistors in a circuit).

To reduce the computational complexity of the above problem we propose a new algorithm for *statistical* timing analysis of circuits under process variation. The algorithm computes the signal arrival time at the output edge of each gate in a particular logic level. The arrival time information is then propagated to calculate the timing information of the next logic level. Finally, at the primary output, the circuit delay is obtained by applying the *MAX* function on all primary output arrival times. This technique is named as *Levelized Covariance Propagation (LCP)*. Since the timing information (signal arrival times and their correlation) is always confined to a single logic level, in this technique the problem of signal correlation can be handled easily. Furthermore, we need only a few random variables (equal to the number of transistors in a gate) to effectively incorporate the effect of random placement of dopants.

#### 4.1. Levelized Covariance Propagation

In this technique, the timing information at the output edges of a particular logic level is calculated using the signal information at its input edges and the delays of all gates in that logic level. The calculated timing information is then propagated to the next logic level. The statistical timing information contains;

1. The mean and the variance of signal arrival time at each output edge.
2. The covariances among the output signal arrival times.
3. The covariance between the output arrival time and the process parameters.

The above information is required to calculate the arrival times at the output of any gate considering the effects of process variation and the reconvergent paths. Consider the example circuit shown in Fig. 2. In this figure, the signal arrival time at the output edge of any gate (e.g.,  $G_1$ ) depends on the gate delay, the signal arrival times at its inputs and their correlation information. The signal correlation at the inputs occurs due to (1) spatial correlation of process parameters and (2) the reconvergent paths (e.g., paths 1 and 2 in Fig. 2). Similarly, the correlation between the arrival times at the outputs of gates  $G_1$  and  $G_2$  is necessary to calculate the signal information at the next logic level.

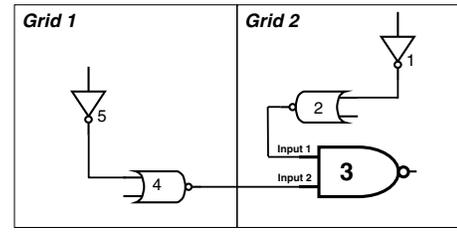


Figure 3. Dependency of gate delay on the input arrival time

Furthermore, the delay of a gate depends on the signal arrival time at its corresponding input due to the spatial correlation of process parameters. For example, consider the signal path shown in Fig. 3. In the figure, signal to the input 1 of gate 3 propagates through gates 1 and 2. Since gates 1, 2 and 3 are closely located (within the same grid), the delays of these gates are highly correlated due to the spatial correlation of process parameters. Consequently, the delay of gate 3 (input 1 to output) will be strongly dependent on the signal arrival time at input 1, which is a function of the delays of gates 1 and 2. On the other hand, the signal to the input 2 of gate 3 propagates through gates 4 and 5, which are not geometrically close to gate 3 (located at different grids). Hence, the delay of gate 3 (input 2 to output) will be weakly correlated to the signal arrival time at input 2. Therefore, the geometric information of the preceding gates along with the input arrival time is necessary to accurately calculate the delay of a gate.

We explain below the methodology to calculate the statistical timing information at the outputs of a logic level.

#### Mean and variance

The mean and the variance of signal arrival time at the output edge of a gate is calculated using the input arrival time and the corresponding gate delay. To show how we calculate the signal information at the output edges, let us consider the example of a 2-input gate. The mean and the variance of the signal arrival time at the output of the gate can be obtained using the *MAX* function as follows,

$$A_{out} = MAX(A_{in,1} + D_1, A_{in,2} + D_2) \quad (5)$$

where  $A_{in,1}$ ,  $A_{in,2}$  are respectively, the signal arrival times at input 1 and 2.  $D_1$  and  $D_2$  are the corresponding pin-to-pin delays of the gate. The mean ( $\mu_{A_{out}}$ ) and the sigma ( $\sigma_{A_{out}}$ ) are obtained using the technique described in [7] as follows,

$$\begin{aligned} \mu_{A_{out}} &= \mu_{S_1} \Phi(\alpha) + \mu_{S_2} \Phi(-\alpha) + \beta \phi(\alpha) \\ \sigma_{A_{out}}^2 &= (\mu_{S_1}^2 + \sigma_{S_1}^2) \Phi(\alpha) + (\mu_{S_2}^2 + \sigma_{S_2}^2) \Phi(-\alpha) \\ &\quad + (\mu_{S_1} + \mu_{S_2}) \beta \phi(\alpha) - \mu_{A_{out}}^2 \end{aligned} \quad (6)$$

where,

$$\begin{aligned} S_i &= A_{in,i} + D_i; \quad i = 1, 2 \\ \mu_{S_i} &= \mu_{A_{in,i}} + \mu_{D_i} \\ \sigma_{S_i}^2 &= \sigma_{A_{in,i}}^2 + \sigma_{D_i}^2 + 2 \cdot cov(A_{in,i}, D_i) \end{aligned} \quad (7)$$

$\mu_{A_{in,i}}$ ,  $\sigma_{A_{in,i}}$ ,  $\mu_{D_i}$  and  $\sigma_{D_i}$  are the mean and the standard deviation of input arrival time and the gate delay, respectively.  $cov(A_{in,i}, D_i)$  are the covariance between the input arrival time and the corresponding gate delay.  $\alpha$  and  $\beta$  in Eq. (6) are given by,

$$\alpha = (\mu_{S_1} - \mu_{S_2}) / \beta \quad (8)$$

$$\beta^2 = \sigma_{S_1}^2 + \sigma_{S_2}^2 - 2\sigma_{S_1}\sigma_{S_2}\rho \quad (9)$$

where,

$$\begin{aligned}\rho &= \text{cov}(S_1, S_2) / \sigma_{S_1} \sigma_{S_2} \\ \text{cov}(S_1, S_2) &= \text{cov}(A_{in,1}, D_2) + \text{cov}(A_{in,2}, D_1) \\ &+ \text{cov}(A_{in,1}, A_{in,2}) + \text{cov}(D_1, D_2)\end{aligned}\quad (10)$$

$\Phi(\alpha)$  and  $\phi(\alpha)$  in Eq. (6) represent the CDF (cumulative distribution function) and PDF (probability density function) of standard normal distribution, respectively.

The mean ( $\mu_{A_{out}}$ ) and variance ( $\sigma_{A_{out}}^2$ ) of the arrival time at the output of a 2-input gate thus, can be obtained using the signal information of the input edges such as input arrival time information ( $\mu_{A_{in,i}}$  and  $\sigma_{A_{in,i}}$ ,  $i = 1, 2$ ), gate delay information ( $\mu_{D_i}$  and  $\sigma_{D_i}$ ) and the covariance information ( $\text{cov}(A_{in,1}, A_{in,2})$ ,  $\text{cov}(A_{in}, D)$  and  $\text{cov}(D_1, D_2)$ ).  $\mu_{A_{in,i}}$ ,  $\sigma_{A_{in,i}}$  and  $\text{cov}(A_{in,1}, A_{in,2})$  are obtained from the previous logic level.  $\mu_{D_i}$ ,  $\sigma_{D_i}$  are obtained using Eq. (3) and 4.  $\text{cov}(D_1, D_2)$  can be calculated using Eq. (3) and (4) as follows,

$$\begin{aligned}\text{cov}(D_1, D_2) &= \text{cov}(D_{1,0} + \sum_i^n s_{1,i} X_i, D_{2,0} + \sum_i^n s_{2,i} X_i) \\ &= \sum_i^n \sum_j^n s_{1,i} s_{2,j} \text{cov}(X_{1,i}, X_{2,j})\end{aligned}\quad (11)$$

where  $X_i$ 's are the random variables representing the parameter variation in length, width, oxide thickness and the threshold voltage. Since  $D_1$  and  $D_2$  represent the delay of the same gate, they are correlated by the parameter variations of all the transistors in that gate. While due to the spatial correlation, a single random variable can be used to represent the variation in a process parameter (length, width, or oxide thickness) for all transistors in the gate, theoretically separate variables should be used to represent the variation in threshold voltage (due to the random placement of dopants) of individual transistor. However, the covariances among the threshold voltage variations of different transistors will be zero ( $\text{cov}(X_{V_{th_i}}, X_{V_{th_j}}) = 0$ ;  $i \neq j$ , where  $X_{V_{th}}$  represents the variation in threshold voltage). Hence, the Eq. (11) is simplified to,

$$\text{cov}(D_1, D_2) = \sum_i^n s_{1,i} s_{2,i} \sigma_{X_i}^2 \quad (12)$$

where  $\sigma_{X_i}$ 's are the standard deviation of the process parameter  $X_i$ .  $\sigma_{V_{th}}$  of each transistor in the circuit is pre-calculated and hence, no additional random variable is required in the algorithm to incorporate the threshold voltage variation due to random placement of dopants.  $\text{cov}(D_1, D_2)$  carries the information regarding the  $V_{th}$  variation of all transistors in a gate. This information is used to calculate the correlation between the signals in reconvergent paths.

Finally,  $\text{cov}(A_{in}, D)$  is obtained as follows,

$$\begin{aligned}\text{cov}(A_{in}, D) &= \text{cov}(A_{in}, D_0 + \sum_i s_i X_i) \\ &= \sum_i s_i \text{cov}(A_{in}, X_i)\end{aligned}\quad (13)$$

where  $s_i$ 's are the sensitivity of the gate delay to parameter variations.  $\text{cov}(A_{in}, X_i)$  is the covariance between the input signal arrival time and the process parameter variation, and is obtained from the previous logic level.  $X_i$ 's are the random variables representing the variation in length, width and the oxide thickness. The correlation between the input arrival time and the threshold voltage variation

will be zero due to the random placement of dopants.

### Covariance among the output signals

We have mentioned above that in order to calculate the signal arrival time ( $\mu_{A_{out}}$  and  $\sigma_{A_{out}}$ ) at the output of a gate in a particular logic level, covariance between the input signals are obtained from the previous logic level. In other words, the signal correlation information among all output edges of a particular logic level needs to be propagated to the next logic level. Consider the example circuit shown in Fig. 2. The covariance between the arrival times at the outputs of  $G_1$  and  $G_2$  can be expressed as,

$$\begin{aligned}\text{cov}(A_{out,1}, A_{out,2}) \\ = \text{cov}(MAX(a_{1,1}, a_{1,2}), MAX(a_{2,1}, a_{2,2}))\end{aligned}\quad (14)$$

where  $A_{out,1}$  and  $A_{out,2}$  are the signal arrival times at the output of gates  $G_1$  and  $G_2$ , respectively.  $a_{i,j}$ 's are given by,

$$a_{i,j} = A_{in,i,j} + D_{i,j}; \quad i, j = 1, 2 \quad (15)$$

where  $A_{in,i,j}$  represents signal arrival time at the  $j$ 'th input of the  $i$ 'th gate and  $D_{i,j}$  is the corresponding pin-to-pin gate delay. In order to solve Eq. (14), we used the following relationship given in [7].

$$\begin{aligned}\text{cov}(X, MAX(Y_1, Y_2)) \\ = \text{cov}(X, Y_1) \Phi(\alpha) + \text{cov}(X, Y_2) \Phi(-\alpha)\end{aligned}\quad (16)$$

where,  $\alpha$  can be obtained using Eq. (8). Using the above relationship, Eq. (14) can be rewritten as,

$$\begin{aligned}\text{cov}(A_{out,1}, A_{out,2}) \\ = \text{cov}(a_{1,1}, \mathbf{a}_{2,12}) \Phi(\alpha_a) + \text{cov}(a_{1,2}, \mathbf{a}_{2,12}) \Phi(-\alpha_a) \\ = \text{cov}(a_{1,1}, a_{2,1}) t_{11} + \text{cov}(a_{1,1}, a_{2,2}) t_{12} \\ + \text{cov}(a_{1,2}, a_{2,1}) t_{21} + \text{cov}(a_{1,2}, a_{2,2}) t_{22} \\ t_{11} = \Phi(\alpha_{a,1}) \Phi(\alpha_{a,2}), \quad t_{12} = \Phi(\alpha_{a,1}) \Phi(-\alpha_{a,2}) \\ t_{21} = \Phi(-\alpha_{a,1}) \Phi(\alpha_{a,2}), \quad t_{22} = \Phi(-\alpha_{a,1}) \Phi(-\alpha_{a,2}) \\ (\mathbf{a}_{2,12} = MAX(a_{2,1}, a_{2,2}))\end{aligned}\quad (17)$$

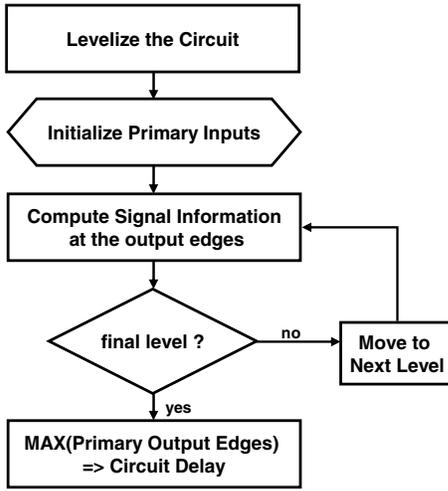
$\text{cov}(a_{1,i}, a_{2,j})$  are calculated as follows,

$$\begin{aligned}\text{cov}(a_{1,i}, a_{2,j}) &= \text{cov}(A_{in,1,i}, A_{in,2,j}) + \text{cov}(A_{in,1,i}, D_{2,j}) \\ &+ \text{cov}(A_{in,2,j}, D_{1,i}) + \text{cov}(D_{1,i}, D_{2,j})\end{aligned}\quad (18)$$

where the first term in the right hand side of Eq. (18) is obtained from the previous logic level. The second and third terms are calculated using Eq. (13). The last term in Eq. (18) is obtained as follows,

$$\text{cov}(D_{1,i}, D_{2,j}) = \sum_k \sum_l s_{1,i,k} s_{2,j,l} \text{cov}(X_{1,k}, X_{2,l}) \quad (19)$$

where  $X_1$  and  $X_2$  are the random variables representing the process parameter variations in gate 1 and 2, respectively.  $s_{1,j,k}$  is the sensitivity of  $k$ 'th parameter to the  $j$ 'th delay of gate 1 (similarly  $s_{2,j,l}$  is the sensitivity of  $l$ 'th parameter to the  $j$ 'th delay of gate 2). Following the above method, the covariances among the arrival times at all output edges can be calculated. The covariance information is stored in an  $N \times N$  matrix and propagated to the next logic level. Note that, the use of the above correlation information in timing analysis automatically takes care of the signal correlation due to the reconvergent paths.



**Figure 4. Levelized Covariance Propagation Algorithm**

#### Covariance between the signal arrival time and the process parameters

The covariance between signal arrival time at the output of a 2-input gate and the process parameters can be expressed as,

$$\begin{aligned}
 cov(A_{out}, X) &= cov(MAX(A_{in,1} + D_1, A_{in,2} + D_2), X) \\
 &= cov(A_{in,1} + D_1, X)\Phi(\alpha) + cov(A_{in,2} + D_2, X)\Phi(-\alpha) \\
 &= cov(A_{in,1}, X)\Phi(\alpha) + cov(A_{in,2}, X)\Phi(-\alpha) \\
 &+ cov(D_1, X)\Phi(\alpha) + cov(D_2, X)\Phi(-\alpha)
 \end{aligned} \quad (20)$$

where  $X$  is set of the random variables representing the variation in length, width and the oxide thickness.  $A_{in}$ 's and  $D$ 's are respectively, the arrival times at the inputs and the corresponding pin-to-pin delays of the gate.  $cov(A_{in,1}, X)$  and  $cov(A_{in,2}, X)$  are obtained from the previous logic level.  $cov(D_i, X)$  is calculated using Eq. (3) as,

$$\begin{aligned}
 cov(D_i, X) &= cov(D_{i,0} + \sum_j^n s_j X_j, X) \\
 &= \sum_j^n s_j cov(X_j, X)
 \end{aligned} \quad (21)$$

where  $X_j$ 's are random variables representing the spatial correlation of the process parameters.  $cov(X_j, X)$ 's are obtained using the covariance matrix described in section 2.1. Similarly, the covariance between all output signals and the process parameters are obtained and the covariance information is stored in an  $N \times M$  matrix ( $M$  is the number of random variables used to model the spatial correlation of process parameters) and propagated to the next logic level.

#### 4.2 Statistical Timing Analysis Algorithm

Fig. 4 shows the flow diagram of the proposed algorithm. The algorithm starts with levelizing the logic gates in a given circuit. In the second step, the algorithm is initialized by feeding the signal information to the primary inputs of the circuit. The signal information contains;

- The mean ( $\mu_{A_{in}}$ ) and the variance ( $\sigma_{A_{in}}^2$ ) of signal arrival time at each primary input.

- The covariances among the input signals ( $cov(A_{in,i}, A_{in,j})$ ;  $i, j = 1, 2, \dots, n$ ;, where  $n$  is the number of primary inputs).
- The covariance between the input arrival time and the process parameters ( $cov(A_{in}, X)$ ).

Based on the signal information at the primary input, the signal information at the output edges of logic level 1 is calculated using the method described in the previous section. The signal information is then propagated to the next logic level. Finally, at the primary output we apply the  $MAX$  function to calculate the circuit delay.

$$D_{circuit} = MAX(A_{PO,out,1}, A_{PO,out,2}, \dots, A_{PO,out,n}) \quad (22)$$

where  $A_{PO,out,i}$  is the arrival at the  $i$ 'th primary output, and  $n$  represents the total number of primary outputs.

#### 4.3. Complexity of the Algorithm

In this section, we present a run-time complexity analysis of the proposed algorithm. Let us assume there are  $N$  number of gates and  $m$  logic levels in a given circuit. The computational complexity of the algorithm involves the calculation of (1) Means and variances, (2) covariances among the signal arrival times and (3) the covariances between signal arrival time and the process parameters. The computation time for (1) and (3) has linear dependency on the number of gates in a particular level. In addition to that, the computation time of (3) also depends on the number of random variables representing the spatial correlation of process parameters. On the other hand, the computation time for (2) depends quadratically on the number of gates in a particular logic level. This is because the covariances between all possible pairs of output signals are computed. Hence, the overall run-time complexity of the proposed algorithm can be expressed as,

$$\sum_i^m O(n_i) + O(n_i \times k) + O(n_i^2) \quad (23)$$

where  $k$  is number of random variables representing the spatial correlation and  $n_i$  is the number of gates in the  $i$ 'th logic level. It can be seen that the complexity of the proposed algorithm depends mostly on the number of gates in a logic level. If  $n_i$  is large, the overall complexity will be dominated by the third term ( $O(n_i^2)$ ) in Eq. (23) and will quadratically depend on  $n_i$ .

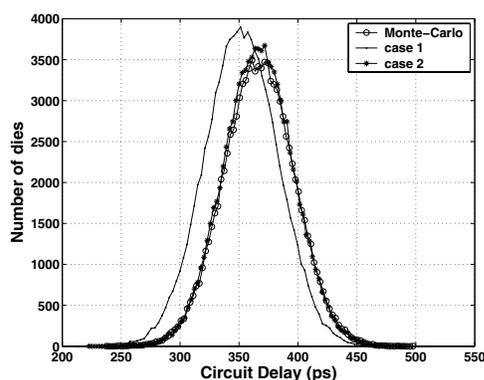
#### 5. Simulation Results

The proposed statistical timing algorithm was implemented in C. We used this algorithm to calculate the circuit delay of several ISCAS benchmark circuits. All the circuits were synthesized using BPTM 70nm technology [1]. We assumed 15% ( $3\sigma$ ) variation in channel length, width, oxide thickness and the threshold voltage for all our experiments. A  $16 \times 16$  rectangular grid structure is used to model the spatial correlation of parameters such as length, width and oxide thickness.

In order to verify the proposed algorithm, we also compared the simulation results with Monte-Carlo simulation. For a considerable accuracy in the analysis, we chose to run 10000 iterations for Monte-Carlo simulation. Fig. 5 shows the circuit delay distribution of ISCAS c499 benchmark circuit. It can be observed from the figure that the circuit delay distribution obtained using the proposed algorithm (considering both spatial correlation and the effect of random placement of dopants) closely matches with the Monte-Carlo simulation result. However, the circuit delay analysis considering only the spatial correlation (case 2) provides optimistic results compared to the

**Table 1. Simulation results for ISCAS benchmark circuits**

Circuit	No. of TR	logic depth	Mean delay(ps)			delay sigma(ps)			CPU-time (s)	
			MC	LCP	error (%)	MC	LCP	error (%)	MC	LCP
c74L85	148	11	118.78	119.01	0.20	12.22	12.12	0.81	1.86	0.28
c74181	372	17	197.26	197.35	0.05	18.44	18.50	0.37	2.87	0.39
c432	590	28	531.06	530.91	0.03	26.74	26.73	0.06	4.03	0.55
c1908	1582	38	531.30	533.13	0.34	38.54	38.05	1.28	11.81	0.63
c880	1642	30	364.48	364.00	0.13	31.11	31.17	0.18	12.87	0.86
c499	1816	29	380.43	381.46	0.27	29.91	28.93	3.28	14.48	0.67
c3540	3638	43	609.63	612.54	0.48	40.64	40.24	0.99	24.55	1.19
c5315	6284	45	558.80	557.73	0.19	45.81	46.82	2.19	46.10	2.78
c6588	9472	122	1646.40	1648.52	0.13	123.82	123.78	0.03	63.19	3.70



**Figure 5. Circuit delay distribution of c499; case 1: with considering the spatial correlation, case 2: with considering both the spatial correlation and the random placement of dopants**

Monte-Carlo simulation. More specifically, statistical timing analysis considering only spatial correlation under-estimates the mean of the delay, while considering only the effect of random placement of dopants under-estimates the variance of the delay [5].

Table. 1 shows the circuit delays of several ISCAS benchmark circuits considering both the spatial correlation and the effect of random placement of dopants (LCP). It also shows the error in mean and standard deviation of the circuit delay compared to the results obtained through Monte-Carlo simulation (MC). Results show that the average errors in mean and standard deviation were 0.21% and 1.07%, respectively.

## 6. Conclusion

In this paper, we proposed a new statistical timing algorithm using *Levelized Covariance Propagation*. This algorithm incorporates both the spatial correlation and the effect of random placement of dopants in the channel in calculating the circuit delay distribution. The simulation results on several ISCAS benchmark circuits show that the proposed algorithm matches with Monte-Carlo simulation. We have also shown that the complexity of the algorithm depends quadratically on the number of gates in a logic level.

## References

- [1] *Berkeley Predictive Transistor Model*. <http://www-device.eecs.berkeley.edu/ptm>.
- [2] A. Agarwal, D. Blaauw, V. Zolotov, S. Sundareswaran, M. Zhao, K. Gala, and R. Panda. Path-Based Statistical Timing Analysis Considering Inter- and Intra-Die Correlations. In *Proc. of TAU*, 2002.
- [3] A. Agarwal, D. Blaauw, V. Zolotov, and S. Vrudhula. Computation and refinement of statistical bounds on circuit delay. In *DAC*, 2003.
- [4] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De. Parameter variations and impact on circuits and microarchitecture. In *Design Automation Conference*, pages 338–342, 2003.
- [5] K. A. Bowman, S. G. Duvall, and J. D. Meindl. Impact of Die-to-Die and Within-Die Parameter Fluctuations on the Maximum Clock Frequency Distribution for Gigascale Integration. *IEEE Journal of Solid-State Circuits*, 37(2):183–190, 2002.
- [6] H. Chang and S. S. Sapatnekar. Statistical Timing Analysis Considering Spatial Correlations using a Single PERT-like Traversal. In *ICCAD*, pages 621–625, 2003.
- [7] C. E. Clark. *The Greatest of a Finite Set of Random Variables*, volume 9. Operations Research, 1961.
- [8] A. Devgan and C. Kashyap. Block-based Static Timing Analysis with Uncertainty. In *ICCAD*, pages 607–614, 2003.
- [9] R. B. Hitchcock. Timing verification and the timing analysis problem. In *ACM Design Automation Conference*, pages 594–604, 1982.
- [10] J. Le, X. Li, and L. T. Pileggi. STAC: Statistical Timing Analysis with Correlation. In *DAC*, pages 343–348, 2004.
- [11] S. R. Nassif. Modeling and Analysis of Manufacturing Variations. In *Custom Integrated Circuits Conference*, pages 223–228, 2001.
- [12] T. Sakurai and R. Newton. Delay analysis of series-connected MOSFET circuits. *IEEE JSSC*, pages 122–131, 1991.
- [13] Y. Taur and T. H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 1998.
- [14] C. Visweswariah. Death, taxes and failing chips. In *Design Automation Conference*, pages 343–347, 2003.
- [15] C. Visweswariah, K. Ravindran, K. Kalafala, S. G. Walker, and S. Narayan. First-order Incremental Block-Based Statistical Timing Analysis. In *Design Automation Conference*, pages 331–336, 2004.