# Distribution of Critical Path Delays in a Combinatorial Circuit

Daigo Yanagi and Shuji Tsukiyama
Dept. of EECE, Chuo University
Tokyo 112-8551, Japan
{daigo@tsuki., tsuki@}elect.chuo-u.ac.jp

**Abstract**: In this paper, we consider how to treat delay-time uncertainties caused by inter-die and intra-die variabilities in evaluating the distribution of the critical delay of a CMOS combinatorial circuit, and formulate inter-die variability as a correlation of delays. Then, we propose an algorithm to evaluate the distribution of the critical delay based on the algorithm in [1] which takes correlations into account. We also show some experimental results to see the effect of the formulation.

## 1. Introduction

Since variability of design parameters increases in deep sub-micron era, the statistical static timing analysis[1], [2], [3], [4], [5], which analyzes the distribution of the maximum delay (critical delay) of a given combinatorial circuit, becomes important to design high speed and low power VLSIs. Because, designers often set excessive margins derived from the worst-case analysis in order to avoid the effect of delay time uncertainty, and such excessive margins bring over-design of circuits. If designers can estimate the distribution of critical path delay exactly, over-design of circuits may be eliminated so that high density and high performance VLSIs can be produced with high yield[6], [7], [8].

Several researches have been done on this topic[1], [2], [3], [4], [5], but all of them treat mainly random-within-die variability, and did not treat inter-die variability explicitly. Since inter-die variability can be considered to affect each element (transistor and interconnect) in a die equally[9], [10], we may be able to treat its effect as a constant. But, if we do so, we misunderstand the expected value of the critical delay, and cannot obtain the precise distribution of the critical delay of the circuit. Moreover, if the effect to the elements in a die is not equal, we may misidentify the critical path. Therefore, we must consider how to treat the effect of the inter-die variability, in order to make the statistical static timing analysis more effective.

The methods proposed so far for the statistical static timing analysis can be divided into three types. The first type uses Monte Carlo simulation[2], [3]. This type is flexible, but time consuming, especially if we introduce correlations of delays. Because, a large number of random values with correlations are to be generated for simulation. The second type selects candidates of the critical paths and evaluates the distributions of those paths[4]. This type is efficient if the number of candidate paths is small. However, the number of paths may explode exponentially with respect to the number of gates, and the number of paths with a similar long delay tends to increase in the practical circuit. Therefore, if the number of candidate-paths is large, this type becomes inefficient, and if only a limited number of candidate paths are considered, the results may be inaccurate. The third type uses a systematic method to calculate the maximum delay to the output of each logic gate[1], [5]. This type is most efficient among these three types, but all of the algorithms of this type other than [1] assume that the distributions of all signal delays are independent each other, which does not hold in combinatorial circuits with re-convergent paths. Therefore, we proposed an algorithm which can treat correlations between delays in [1]. We use this algorithm to treat inter-die variability.

In this paper, we consider how to treat delay-time uncertainties caused by inter-die and intra-die variabilities in the statistical static timing analysis for a CMOS combinatorial circuit, and formulate inter-die variability as a correlation of delays. Then, we propose an algorithm based on the algorithm in [1] which takes correlations into account. We also show some experimental results to see the effect of the formulation.

## 2. Preliminaries

The factors causing delay-time uncertainty can be divided into two types; intrinsic physical factors and dynamic environmental factors[9], [10]. The former is caused by fabrication process and mask imperfections, whose effects increase in nano-technologies. The latter contains temporal factor and is divided into two sub-types. For example, IR drop, cross-talk noise, and the effect of temperature change are short-term factors. On the other hand, hot-carrier degradation and electromigration are examples of long-term factors. As for the dynamic environmental factors, we should consider the worst-case, in order to guarantee the behavior and the lifetime of a circuit. Therefore, the intrinsic physical factors are targets in this paper.

There are several factors classified in intrinsic physical factors, which bring delay-time uncertainty. For example, as for the transistor, there are threshold voltage $V_{th}$, effective channel length $L_{eff}$, drain-source resistance $R_{ds}$, and gate oxide thickness $t_{ox}$, and as for the interconnect, there are width $w$, separation $s$, thickness $t$, via resistance $R_{via}$, interlayer dielectric thickness $ILD_t$, and dielectric constant $\epsilon$. Due to the variability of these parameters, saturated current $I_{dsat}$ between source and drain, load capacitance, interconnect capacitance and resistance and so on vary, so do the delay.

Such variability of the intrinsic physical factors can be divided into two sub-types; inter-die and intra-die

variabilities[10]. The effect of the former is equal within die, and that of the latter varies within die. The intra-die variability is composed of systematic factor and random factor, and the systematic intra-die variability depends on spatial factor and proximity factor[11]. Some proximity-dependent factors can be decreased by optical proximity corrections, phase shift mask, and etc.

Based on the classifications stated above, we denote the delay $\tau$ of each element (gate or interconnect) by $\tau = \mu + \Delta + \nu + \delta$, where $\mu$ is the nominal delay, $\Delta$ is the inter-die variation, $\nu$ is the spatial intra-die variation, and $\delta$ is the random intra-die variation. In this notation, we assume that $\mu$ and $\nu$ are constant, and $\Delta$ and $\delta$ are stochastic variables denoted by normal distribution $N(0, \sigma^2)$ with the mean equal to 0. The image of the distribution of delay $\tau$ is shown in Fig.1.

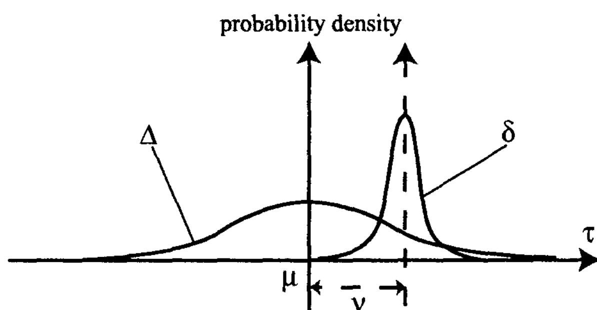


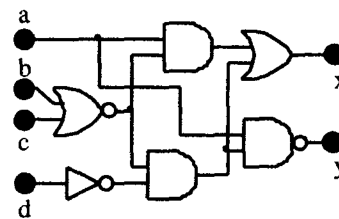

Figure 1. Delay model

The amount of spatial intra-die variation is given as $\nu$, whose value depends on the position of the element in a die. The proximity-dependent systematic variation is assumed to be represented by a correlation coefficient between $\delta_1$ and $\delta_2$ of delays of elements 1 and 2, that is, if the delays of elements 1 and 2 vary dependently on the distance between the elements, then we represent such an effect by a correlation coefficient between $\delta_1$ and $\delta_2$.

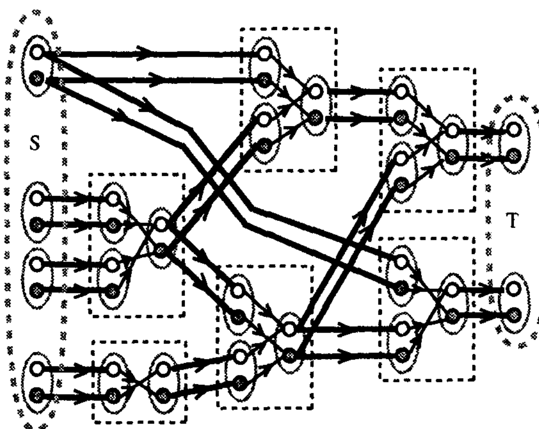
In order to show inter-die variations, we introduce a strong correlation, such as correlation coefficient almost equal to 1, between $\Delta_1$ and $\Delta_2$ of elements 1 and 2 such that the devise structures of the elements are the same in a die. For example, if elements 1 and 2 are interconnects on metal 1 layer, then the inter-die variability factors cause similar effects on the delays of these interconnects, and hence we assign a correlation coefficient between $\Delta_1$ and $\Delta_2$.

## 3. Algorithm

In order to find the distribution of the maximum delay of a CMOS combinatorial circuit, we represent the circuit by an acyclic graph $G = (V, E)$, as shown in Fig.2. In $G$, each terminal $v$ of a circuit is represented by a pair of vertices $v0$ and $v1$, which are denoted in the figure by a white circle and a gray circle contained in an ellipse, respectively. Each primary input corresponds to a pair of sources in an ellipse, into which no edge comes, and each primary output corresponds to a pair of sinks in an ellipse, from which no edge goes out. We denote the set of sources by $S$ and that of sinks by $T$. Each logic gate is represented by a box, and each left and right ellipse in a box corresponds to an input and output terminals of the logic gate corresponding to the box, respectively.




(a) A given combinatorial circuit.




(b) An acyclic graph G=(V,E).

Figure 2. The graph representing a given circuit

Vertex $v0$ ($v1$) is called 0-vertex (1-vertex) of terminal $v$, and represents logic value 0 (1) transmitted to terminal $v$. Namely, we will represent the maximum delay $d(v,0)$ ($d(v,1)$) spent for transmitting logic value 0 (1) from a primary input to terminal $v$ by the longest path length $d(v0)$ ($d(v1)$) from a source to 0-vertex (1-vertex) of terminal $v$. In order to do so, we generate an edge $e = (v, w)$ and assign delay $t(e)$ to the edge, if a logic value of $w$ is determined by the logic value of $v$. It is easy to generate edge(s) corresponding to a net.

For a given logic gate, an input logic value is called a control signal, if the value of output terminal of the logic gate is determined by the input logic value, whatever values other input terminals have. Otherwise, it is called a non-control signal. For AND and NAND (OR and NOR) gates, logic value 0 (1) is the control signal and 1 (0) is the non-control signal. For an inverter, both 0 and 1 are control signals, and for an XOR gate, both 0 and 1 are non-control signals.

Let us consider the case when a non-control signal $b$ is transmitted to all inputs $v_i$ ($1 \le i \le k$) of a logic gate, and let $b'$ be the logic value of the output $w$ of the gate determined by $b$. In this case, the maximum delay

$d(w,b')$ is calculated by taking the maximum as shown below,

$$d(w,b') = max[\ d(v_i,b) + t(v_i,b) \mid 1 \le i \le k\ ]$$

where $t(v_i,b)$ is the gate switching delay for setting $w$ to be $b'$ by the signal $b$ of input $v_i$. Therefore, if we generate an edge $e_ib = (v_ib, wb')$ from each vertex $v_ib$ to vertex $wb'$, we can represent $d(w,b')$ by the longest path length $d(wb')$ to $wb'$.

On the other hand, consider the case when a control signal $c$ is transmitted to an input terminal $v_i$ of a gate. In this case, the logic value $c'$ of the output $w$ of the gate is determined as soon as $v_i$ becomes $c$. However, if all inputs other than $v_i$ are non-control signal, $w$ is set to $c'$ after the delay equal to $d(v_i,c) + t(v_i,c)$, where $t(v_i,c)$ is the gate switching delay for setting $w$ to be $c'$ by the signal $c$ of input $v_i$. Hence, the maximum delay $d(w,c')$ for $w$ to be $c'$ is obtained by taking the maximum among these delays $d(v_i,c)+t(v_i,c)$, $(1 \le i \le k)$. Thus, we can calculate $d(w,c')$ by the longest path length $d(wc')$, if we generate an edge $e_ic = (v_ic, wc')$ from each vertex $v_ic$ to $wc'$.

Delay $t(e)$ of an edge $e$ thus generated is a stochastic variable, and is modeled by $t(e) = \tau(e) = \mu(e) + \Delta(e) + \nu(e) + \delta(e)$, as described in the previous section. Since $\Delta(e)$ and $\delta(e)$ are independent, the mean $Exp[t(e)]$ and the variance $Var[t(e)]$ of $t(e)$ are given by

$$\begin{aligned} Exp[t(e)] &= \mu(e) + \nu(e) \qquad ,and \\ Var[t(e)] &= Var[\Delta(e)] + Var[\delta(e)] \end{aligned}$$

respectively. If delays $t(e')$ and $t(e")$ of edges $e'$ and $e"$ are not independent, the correlation coefficient $\rho(e', e") \ne 0$ between $t(e')$ and $t(e")$ can be calculated from the following equation,

$$\begin{aligned} &\sqrt{Var[t(e')] \cdot Var[t(e")]} \cdot \rho(e', e") \\ &= \sqrt{Var[\Delta(e')] \cdot Var[\Delta t(e")]} \cdot \rho_\Delta(e', e") \\ &\quad + \sqrt{Var[\delta(e')] \cdot Var[\delta(e")]} \cdot \rho_\delta(e', e") \end{aligned}$$

where $\rho_\Delta(e', e")$ is the correlation coefficient between $\Delta(e')$ and $\Delta(e")$, and $\rho_\delta(e', e")$ is that between $\delta(e')$ and $\delta(e")$.

Our algorithm proposed in [1] can compute the distribution of the longest delay of an acyclic graph $G = (V, E)$ even if the edge-delays have correlations. The algorithm computes the mean $Exp[d(w)]$ and the variance $Var[d(w)]$ of the longest path length $d(w)$ to each vertex $w$ in topological order, together with the necessary correlation coefficients between $d(w)$ and $d(v)$ of other vertex $v$ and between $d(w)$ and $t(e)$ of an edge $e$. Finally, it computes the mean $Exp[MaxD]$ and variance $Var[MaxD]$ of the longest delay $MaxD = max[\ d(w) \mid w \in T\ ]$ from a source to a sink. The worst-case time complexity of the algorithm is $O(|E|^2)$, where $|E|$ is the number of edges.

# 4. Experimental Results

In the algorithm, the maximum among more than two stochastic variables is calculated by repeating an operation of finding the maximum among two variables, and hence the order of the operations may be important to improve the accuracy. Because, in [12] authors claim that if two variables with larger mean values are selected first and the distribution of the maximum is computed from these selected variables, then the result is more accurate than random selection and than selecting two variables with smaller mean values first, although [12] does not take correlations into account. Therefore, in the case when correlations are taken into consideration, there may exist another method to improve accuracy. For example, in the computation of $t = max[\ x, y, x'\ ]$, if $\rho(x, x') = 1$, then the calculation of $t = max[\ y,\ max[x, x']\ ]$ is more accurate than the calculation of $t = max[\ max[x,y],\ x'\ ]$. Hence, selecting two variables with a stronger correlation first may give a better solution. We checked this by several experiments in the computation of finding the maximum of 5 variables, and compare the results obtained by our algorithm with the results obtained by 1,000 times Monte Carlo simulation. Then, we found that there exists no specific method to improve accuracy. Therefore, we may say that our algorithm is pretty robust for the order of taking the maximum.

In order to see the effect of correlations (inter-die variability), we applied our algorithm to the graph $G$ obtained by cascading the graph $G_C$ shown in Fig.3 three times. In the graph $G$, the mean $\mu(e)$ of each edge $e$ is $\mu(e) = 30$, and the standard deviation $\sigma(e)$ of the delay of edges $e_i$ $(1 \le i \le 8)$ is $\sigma(e) = 0.6\mu(e)$ and that of the remaining edge is $\sigma(e) = 0.1\mu(e)$. As correlations of edge-delays, we introduced correlations represented by correlation coefficients $\rho_{net}$ and $\rho_{gate}$ to the pair of edge-delays of outgoing edges from a vertex and of incoming edges to a vertex, respectively. The pair of these edges are depicted in Fig.3. Moreover, we introduced a correlation represented by correlation coefficient $\rho_{D2D}$ to the pair of edge $e_i$ $(1 \le i \le 4)$ and the corresponding edge $e_i$ in the other component of the graph. This correlation models the inter-die variability. Since there are 3 components in the graph, there are 3 edges corresponding to $e_i$ $(1 \le i \le 4)$. In the experiments, we assign the same correlation coefficient to all pairs of these edges.

Table1 shows the results of the effects of these correlations. In the table, "no-correlation" indicates results obtained by ignoring all these correlations as well as the correlations between the delays of paths. The percentage written in parenthesis is the relative error of these results to the result obtained by 1,000 times Monte Carlo simulation.

In the same graph $G$, we changed the mean of the delays of those edges, which are located on the top and drawn in bold lines in the figure, from 30 to 60. The relation between the mean and the standard deviation is unchanged. The results are shown in Table2, where

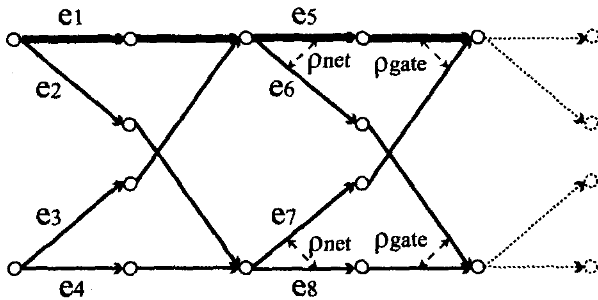



Figure 3. Component $G_c$ of a tested graph $G$

Table 1. Effect of correlations 1

| correlations | | | critical delay | |
|---|---|---|---|---|
| $\rho_{net}$ | $\rho_{gate}$ | $\rho_{D2D}$ | mean | s.d. |
| no-Correlation | | | 467.44 (4.97[%]) | 21.72 (-34.68[%]) |
| 0.0 | 0.0 | 0.0 | 452.19 (1.54[%]) | 30.33 (-8.81[%]) |
| 0.3 | 0.3 | 0.0 | 449.81 (2.26[%]) | 31.30 (0.03[%]) |
| 0.3 | 0.3 | 0.9 | 449.81 (1.74[%]) | 37.64 (-7.50[%]) |

$\rho_{D2Dcp}$ is the correlation coefficients between the pair of edges $e_1$ contained in different components of the graph, and $\rho_{D2D}$ is the correlation coefficients between the pair of edge $e_i$ $(2 \leq i \leq 4)$ contained in the different components. In this experiment, both the correlation coefficients $\rho_{net}$ and $\rho_{gate}$ are set to be 0.

From the results shown in Tables1 and 2, we can see that the correlations represented by $\rho_{net}$ and $\rho_{gate}$ decrease the mean of the critical delay. Moreover, since we may be able to assume $\rho_{D2D} > 0$, the correlation representing the inter-die variability increases the distribution of the critical delay. Therefore, we can say that if the correlations including inter-die variability are ignored, we cannot obtain precise distribution of the critical delay.

Table 2. Effect of correlations 2

| correlations | | critical delay | |
|---|---|---|---|
| $\rho_{D2Dcp}$ | $\rho_{D2D}$ | mean | s.d. |
| no-Cor | | 757.98 (5.54[%]) | 59.04 (-32.97[%]) |
| 0.0 | 0.0 | 726.00 (2.36[%]) | 87.35 (1.22[%]) |
| 0.9 | 0.0 | 726.26 (2.13[%]) | 116.56 (0.17[%]) |
| 0.9 | 0.9 | 726.25 (2.25[%]) | 116.59 (-0.60[%]) |

## 5. Conclusions

In this paper, we proposed an algorithm to calculate he distribution of the critical delay of a given CMOS ombinatorial circuit, which can treat delay-time uncer-ainties caused by inter-die variability as well as intra-ie variability. And we showed some experimental re-ılts, which indicate the effect of the inter-die variabil-y. Since the algorithm treated the inter-die variability, e can compute correctly the probability for a circuit to ork in a given clock frequency.

Many works are to be done in this research. We are now studying problems to calculate the probability for a path to be critical and the probability for an edge to be included in the critical path, which will be useful in the circuit design considering uncertainty.

## References


[1] S. Tsukiyama, M. Tanaka, and M. Fukui, "An algorithm for statistical static timing analysis considering correlations between delays," *IEICE Trans. Fundamentals*, vol.E84-A, no.11, pp.2746-2754, 2001.

[2] H.-F. Jyu, S. Malik, S. Devadas, and K. Keutzer. Statistical timing analysis of combinatorial logic circuits. *IEEE Trans. VLSI Systems*, 1(2):126–137, 1993.

[3] J. Rung-Bin and W. Meng-Chiou. A new statistical approach to timing analysis of VLSI circuits. In *Proc. 11th Int. Conf. on VLSI Design*, pages 507–513, 1997.

[4] H. Matsunaga, D. Mizoguchi, and H. Yasuura. Estimation of combinatorial circuit delays using the distribution of CMOS gate delays. In *Proc. DA Symp. '99*, pages 77–82, JIPS, 1997 (in Japanese).

[5] M. Berkelaar. Statistical delay calculation, a linear time method. In *Proc. Int. Workshop on Timing Issues in the Specification and Synthesis of Digital Systems (TAU97)*, pages 15–24. ACM, 1997.

[6] M. Hashimoto and H. Onodera. A performance optimization method by gate resizing based on statistical static timing analysis. In *Proc. Workshop on Synthesis And System Integration of MIxed Technology (SASIMI 2000)*, pages 77–82, 1999.

[7] E.T.A.F.Jacobs and M.R.C.M.Berkelaar. Gate sizing using a statistical delay model. In *Proc. Design Automation and Test in Europe (DATE2000)*, pages 283–290, 2000.
timing analysis,"

[8] S. Nishimoto, S. Tsukiyama, M. Tanaka, and M. Fukui, "A macrocell layout system considering manufacturing fluctuations," *Tech. Report of IEICE, VLD 2000-130*, pp.21–26, 2001 (in Japanese).

[9] S.R. Nassif, "Delay variability: Sources, impacts and trends", *Digest of ISSCC* , pp.368–369, 2000.

[10] K. Bernstein, K.M. Carrig, C.M. Durham, P.R. Hansen, D. Hogenmiller, E.J. Nowak, and N.J. Rohrer. *High Speed CMOS Design Styles*, Chapt.1, 1–50, 1999.

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

[12] E. Jacobs and M. Berkelaar. Improving the Accuracy of Statistical Delay Calculation. In *Proc. Int. Workshop on Logic Synthesis*, pages 297–301, 2000.
path problem,"