Defining Statistical Timing Sensitivity for Logic Circuits With Large-Scale Process and Environmental Variations

Xin Li, *Member, IEEE*, Jiayong Le, *Member, IEEE*, Mustafa Celik, *Member, IEEE*, and Lawrence T. Pileggi, *Fellow, IEEE*

***Abstract*—The large-scale process and environmental variations for today's nanoscale ICs require statistical approaches for timing analysis and optimization. In this paper, we demonstrate why the traditional concept of slack and critical path becomes ineffective under large-scale variations and propose a novel sensitivity framework to assess the "criticality" of every path, arc, and node in a statistical timing graph. We theoretically prove that the path sensitivity is exactly equal to the probability that a path is critical and that the arc (or node) sensitivity is exactly equal to the probability that an arc (or a node) sits on the critical path. An efficient algorithm with incremental analysis capability is developed for fast sensitivity computation that has linear runtime complexity in circuit size. The efficacy of the proposed sensitivity analysis is demonstrated on both standard benchmark circuits and large industrial examples.**

***Index Terms*—Process variations, sensitivity, statistical static timing analysis.**

I. Introduction

As IC technologies are scaled to finer feature sizes, the increasing fluctuations in manufacturing processes introduce various uncertainties in circuit behavior, thereby significantly impacting product yield. Further exacerbating the problem is the increasing impact of environmental fluctuations, such as those due to temperature and power supply variations. Addressing the nanoscale manufacturing and design realities requires a paradigm shift in the current design methodology such that large-scale variations are considered at all levels of design hierarchy.

Toward this goal, various algorithms have been proposed for statistical timing analysis with the consideration of both process and environmental variations [3]–[20]. Most of the proposed solutions fall into one of two broad categories: path-based approaches [3]–[9] and block-based approaches [10]–[20]. The path-based approaches can take into account the correlations from both path sharing and global process parameters; however, a set of critical paths must be preselected based on their nominal delay values. In contrast, the block-based techniques are more general yet are limited by various delay modeling assumptions. For example, many block-based statistical timing analysis algorithms [13]–[17] assume that delay variations can be approximated as normal distributions in order to efficiently handle both spatial correlations and reconvergent fan-outs.

Whereas statistical timing analysis has been intensively studied, how to interpret and utilize its results remains an open question. Most importantly, a new methodology of using timing-analysis results to guide timing optimization and explore the tradeoffs between performance, yield, and cost is required in the statistical domain. In nominal timing analysis, critical path and slack are two important metrics that have been widely applied to timing optimization, but the inclusion of large-scale variations renders these traditional methodologies obsolete.

First, the delay of each path is a random variable, instead of a deterministic value, in statistical timing analysis. As such, every path can be critical (i.e., have the maximal delay) with certain probability. Second, the slacks at all nodes are random variables that are statistically correlated. The parametric timing yield is determined by the probability distributions of all these slacks as well as their correlations. It, in turn, implies that an individual slack at a single node is not a sufficiently good metric that can be utilized to guide timing optimization. For these reasons, a new timing performance criterion must be proposed to accommodate the special properties of statistical timing analysis/optimization.

In this paper, we propose a new concept of statistical timing sensitivity to guide the timing optimization of logic circuits with large-scale parameter variations. We define statistical timing sensitivities for paths, arcs, and nodes. The novelty of this paper is the creation of a link between probability and sensitivity. We prove that the path sensitivity is exactly equal to the probability that a path is critical and that the arc (or node) sensitivity is exactly equal to the probability that an arc (or a node) sits on the critical path.

An important contribution of this paper is to propose a novel algorithm for fast sensitivity computation and demonstrate how one can evaluate the timing sensitivities by a single breadth-first graph traversal. The computational complexity of the proposed

Manuscript received May 4, 2007; revised September 11, 2007. This paper was presented in part at the IEEE/ACM International Conference on Computer Aided Design, San Jose, CA, 2005. This paper was recommended by Associate Editor F. N. Najm.

X. Li was with Extreme DA, Inc., Santa Clara, CA 95054 USA. He is now with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: xinli@ece.cmu.edu).

J. Le and M. Celik are with Extreme DA, Inc., Santa Clara, CA 95054 USA (e-mail: kelvin@extreme-da.com; mustafa@extreme-da.com).

L. T. Pileggi is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: pileggi@ece.cmu.edu).

Digital Object Identifier 10.1109/TCAD.2008.923241

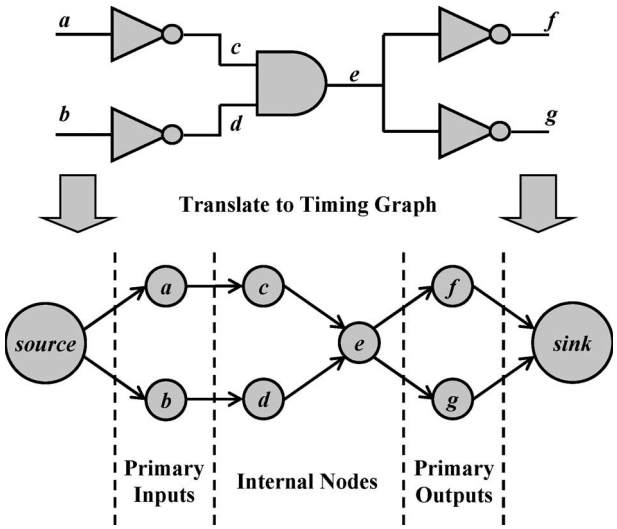


Fig. 1. Simple timing graph example.

sensitivity analysis is linear in circuit size. In addition, an incremental analysis capability is provided to quickly update the statistical timing and sensitivity information after changes to a circuit are made.

Our proposed path and arc sensitivities are theoretically equivalent to the path and edge criticalities defined in [16] and [21], respectively. Namely, the path (or arc) sensitivity value is identical to the path (or arc) criticality value except for numerical errors. However, the proposed sensitivity framework is ready to handle several extensions, including (but not limited to) the high-order sensitivities that will be discussed in Section IV-E.

This paper is organized as follows. In Section II, we review the background of statistical static timing analysis and then discuss the statistical properties of slack and critical path in Section III. We define various statistical timing sensitivities in Section IV and develop the algorithm for sensitivity computation in Section V. The efficacy of the proposed sensitivity analysis is demonstrated by several numerical examples in Section VI. Finally, we conclude in Section VII.

II. Background

A. Nominal Static Timing Analysis

Given a gate-level netlist, static timing analysis translates the netlist into a timing graph, i.e., a weighted directed graph $G = (V, E)$ where each node $V_i \in V$ denotes a primary input, output, or internal net, each edge $E_i = \langle V_m, V_n \rangle \in E$ denotes a timing arc, and the weight $D(V_m, V_n)$ of E_i stands for the delay value from the nodes V_m to V_n. In addition, a source/sink node is conceptually added before/after the primary inputs/outputs so that the timing graph can be analyzed as a single-input single-output network. Fig. 1 shows a simple timing graph example.

There are several key concepts in nominal static timing analysis, which are briefly summarized as follows. More details on static timing analysis can be found in [31].

1) The arrival time (AT) at a node V_i is the latest time that the signal becomes stable at V_i. It is determined by the longest path from the source node to V_i.

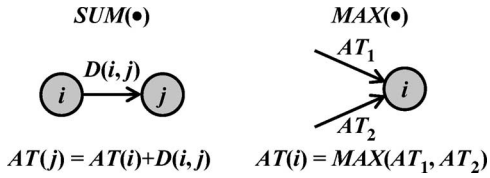


Fig. 2. Two atomic operations for static timing analysis.

2) The required time (RT) at a node V_i is the latest time that the signal is allowed to become stable at V_i. It is determined by the longest path from V_i to the sink node.
3) Slack is the difference between the required time and the arrival time, i.e., RT $-$ AT. Therefore, a positive (or negative) slack means that the timing constraint is satisfied (or failed).
4) Critical path is the longest path between the source and sink nodes. In nominal timing analysis, all nodes along the critical path have the same (i.e., smallest) slack.

The purpose of static timing analysis is to compute the arrival time, the required time, and the slack at each node and then identify the critical path. Taking the arrival time as an example, static timing analysis starts from the source node, propagates the arrival times through all timing arcs by a breadth-first graph traversal and eventually reaches the sink node. Two atomic operations, i.e., SUM($\bullet$) and MAX($\bullet$), as shown in Fig. 2, are repeatedly applied during such a traversal.

After the static timing analysis is complete, the critical path and the slack provide the necessary information that is required for timing optimization. Roughly speaking, the gates and interconnects along the critical path (where the slacks are small) should be upsized to reduce delay, whereas those along the noncritical paths (where the slacks are large) should be downsized to save area and/or power.

B. Process Variation Modeling

According to the geometrical scale of their occurrence, process variations can be classified into two broad categories: interdie and intradie variations. Interdie variations model the common/average variations across the die, whereas intradie variations model the individual, but spatially correlated, local variations within the same die.

In most practical applications, both interdie and intradie variations are modeled as the random variables that are jointly normal. In such cases, principal component analysis (PCA) can be applied to find a set of independent factors to represent the original correlated random variables [29].

Given N process parameters $\eta = [\eta_1, \eta_2, \ldots, \eta_N]^{\mathrm{T}}$, the process variations $\Delta\eta = \eta - \eta_0$, where η_0 contains the mean values of η, are modeled as zero-mean random variables. The correlation of $\Delta\eta$ can be represented by a symmetric, positive-semidefinite covariance matrix R [29]. PCA decomposes R as follows:

$$R = V \cdot \Sigma \cdot V^{\mathrm{T}} \tag{1}$$

where $\Sigma = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$ contains the eigenvalues of R and $V = [V_1, V_2, \ldots, V_N]$ contains the corresponding eigenvectors that are orthonormal, i.e., $V^{\mathrm{T}}V = I$ (I is an identity

matrix). Based on Σ and V, PCA defines a set of new random variables

$$\Delta\varepsilon = \Sigma^{-0.5} \cdot V^{\mathrm{T}} \cdot \Delta\eta. \tag{2}$$

These new random variables $\Delta\varepsilon = [\Delta\varepsilon_1, \Delta\varepsilon_2, \ldots, \Delta\varepsilon_N]^{\mathrm{T}}$ are called the principal components or factors. It is easy to verify that $\{\Delta\varepsilon_i; i = 1, 2, \ldots, N\}$ are independent and standard normal (i.e., zero mean and unit variance) [29].

The essence of PCA can be interpreted as a coordinate rotation of the space defined by the original random variables. In addition, if the magnitude of the eigenvalues $\{\lambda_i\}$ decreases quickly, it is possible to use a small number of principal components to approximate the original N-dimensional space. More details on PCA can be found in [29].

C. Statistical Static Timing Analysis

Unlike nominal timing analysis, the gate/interconnect delays in statistical timing analysis are all modeled as random variables to account for large-scale process variations. It means that the weight $D(V_m, V_n)$ associated with each timing arc is a random variable instead of a deterministic value. Therefore, the two atomic operations, SUM($\bullet$) and MAX($\bullet$), must handle statistical distributions.

Many statistical timing analysis algorithms [13]–[17] approximate the gate/interconnect delays and the arrival times as linear models

$$x = B_x^{\mathrm{T}} \cdot \Delta\varepsilon + C_x = \sum_{i=1}^{N} B_{xi} \cdot \Delta\varepsilon_i + C_x \tag{3}$$

$$y = B_y^{\mathrm{T}} \cdot \Delta\varepsilon + C_y = \sum_{i=1}^{N} B_{yi} \cdot \Delta\varepsilon_i + C_y \tag{4}$$

where x and y denote two gate/interconnect delays or arrival times, $C_x,\ C_y \in R$ are the constant terms, $B_x,\ B_y \in R^N$ contain the linear coefficients, $\{\Delta\varepsilon_i; i = 1, 2, \ldots, N\}$ is a set of random variables to model process variations, and N is the total number of these random variables. We assume that $\{\Delta\varepsilon_i; i = 1, 2, \ldots, N\}$ are independent standard normal distributions and that they are extracted by the PCA in Section II-B. The random variables x and y in (3) and (4) are the linear combinations of multiple normal distributions and, therefore, are also normal [30].

Given the linear models in (3) and (4), the SUM($\bullet$) operation can be easily handled by the following:

$$x + y = (B_x + B_y)^{\mathrm{T}} \cdot \Delta\varepsilon + (C_x + C_y). \tag{5}$$

The MAX($\bullet$) operation, however, is nonlinear. In other words, the maximum of two normal distributions is not necessarily normal. However, it is possible to approximate MAX($\bullet$) by a linear model [13]–[17]

$$\mathrm{MAX}(x, y) = \alpha \cdot x + \beta \cdot y + \gamma \tag{6}$$

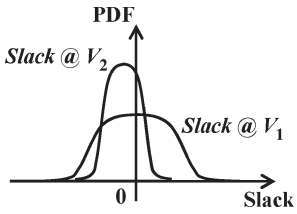


Fig. 3. Slack distribution in statistical timing analysis.

where the constant term γ is determined by matching the mean value

$$\gamma = E\left[\mathrm{MAX}(x, y)\right] - \alpha \cdot E[x] - \beta \cdot E[y] \tag{7}$$

and the linear coefficients α and β are determined by either matching the moments [13], [14], [17] or calculating the tightness probabilities [16]

$$\alpha = P(x \geq y) \tag{8}$$

$$\beta = P(y \geq x). \tag{9}$$

In (7)–(9), $E(\bullet)$ stands for the expected value and $P(\bullet)$ represents the probability.

III. Statistics of Slack and Critical Path

In this section, we highlight the reasons why the traditional concept of slack and critical path becomes ineffective under process variations.

A. Slack

In nominal timing analysis, slack is utilized as a metric to measure how tightly the timing constraint is satisfied. A negative slack means that the timing constraint is not met, whereas a (small) positive slack means that the timing constraint is (marginally) satisfied. In the statistical case, however, it is difficult to make such a straightforward judgment, because all slacks are random variables instead of deterministic values. For instance, Fig. 3 shows two slack distributions computed from statistical timing analysis. The node V_1 presents a larger probability that the slack is positive than the node V_2. However, the worst case (smallest) slack at V_1 is more negative than that at V_2. Hence, it is hard to conclude which slack distribution is better using a simple criterion.

More importantly, the slacks throughout a timing graph are statistically correlated in statistical timing analysis and must be concurrently considered to determine the parametric timing yield. In nominal timing analysis, it is well known that the timing constraint is satisfied if and only if all slacks in the timing graph are positive. In the statistical case, this condition can be stated as follows: The probability that the timing constraint is satisfied (i.e., the parametric timing yield) is equal to the probability that all slacks are positive.

$$\text{Yield} = P[\text{Slack}_{V1} \geq 0 \quad \& \quad \text{Slack}_{V2} \geq 0 \quad \cdots]. \tag{10}$$

Studying (10), we notice that such a probability depends on all slack distributions as well as their correlations. Unlike nominal timing analysis where slacks are deterministic values

without correlations, knowing individual slack distributions in statistical timing analysis is not sufficient to determine the parametric timing yield. The probability in (10) cannot be accurately evaluated if the correlations are ignored. The aforementioned analysis implies an important fact that an individual slack distribution at a single node may not be meaningful in statistical timing analysis.

However, there exist some "important" nodes for which the slacks have special meanings. Given a timing graph, we define the node $V_{\rm IN}$ as an important node if all paths in the timing graph pass $V_{\rm IN}$. Based on this definition, the source and sink nodes are two important nodes, because all paths start from the source node and terminate at the sink node. In some special timing graphs, it is possible to find other important nodes. For example, the node e in Fig. 1 is an important node by this definition. The importance of the node is that, if $V_{\rm IN}$ is an important node, the parametric timing yield in (10) can be uniquely determined by the slack at $V_{\rm IN}$

$$\text{Yield} = P[\text{Slack}_{\rm VIN} \geq 0]. \tag{11}$$

The physical meaning of (11) can be intuitively explained by the concept of Monte Carlo simulation. When a timing graph is simulated by Monte Carlo analysis, a delay sample (i.e., a set of deterministic delay values for all timing arcs) is drawn from the random variable space in each Monte Carlo run. The parametric timing yield is equal to Num_1 (the number of samples for which the timing constraint is satisfied) divided by Num (the total number of Monte Carlo runs). Similarly, the probability $\text{Slack}_{\rm VIN} \geq 0$ is equal to Num_2 (the number of samples for which the slack at $V_{\rm IN}$ is positive) divided by Num. In each Monte Carlo run, the timing constraint is failed if and only if there is a path P whose delay is larger than the given specification. In this case, the slack at $V_{\rm IN}$ must be negative because all paths pass the important node $V_{\rm IN}$, and therefore, $V_{\rm IN}$ must be on the path P. The aforementioned analysis proves that Num_1 is equal to Num_2, yielding (11).

Equations (10) and (11) indicate another difference between nominal and statistical timing analyses. In nominal timing analysis, the slack at any node along the critical path uniquely determines the timing performance. In statistical timing analysis, however, only the slack at an important node uniquely determines the timing performance. Compared with those nodes on the critical path, important nodes belong to a much smaller subset, because they must be shared by all paths in the timing graph.

The aforementioned concept of important node can be extended to a node set. If we remove a set of nodes and cut the entire timing graph into two disconnected subgraphs, where one subgraph contains the source node and the other subgraph contains the sink node, we refer to the set of the removed nodes as a separating node set. It is easy to verify that all paths from the source node to the sink node pass through the separating node set. Therefore, following the same reasoning of the Monte Carlo simulation, it can be proven that the parametric timing yield is uniquely determined by the slacks of all nodes in a separating node set

$$\text{Yield} = P[\text{Slack}_{W1} \geq 0 \quad \& \quad \text{Slack}_{W2} \geq 0 \quad \cdots] \tag{12}$$

where Slack_{Wi} represents the slack at the ith node of the separating node set.

B. Critical Path

Similar to slack, there are several major differences between nominal and statistical timing analyses on critical path. First, given a timing graph, the maximal delay from the source node to the sink node can be expressed as

$$D = \text{MAX}(D_{P1}, D_{P2}, \ldots) \tag{13}$$

where D_{Pi} is the delay of the ith path. In nominal timing analysis, $D = D_{Pi}$ if and only if the path P_i is critical. In statistical timing analysis, however, every path can be critical (i.e., have the maximal delay) with certain probability. Although it is possible to define the most critical path as the path P_i that has the largest probability to be critical, the maximal circuit delay in (13) must be determined by all paths instead of the most critical path only.

Second, but more importantly, the critical path concept is not so helpful for statistical timing optimization. In the nominal case, the gates and interconnects along the critical (or noncritical) path are repeatedly selected for up (or down) sizing. This strategy becomes ineffective under process variations. One important reason is that many paths may have similar probabilities to be critical and all of them must be considered for timing optimization. Even in the nominal case, many paths in a timing graph can be equally critical, which is the so-called "slack wall" [23]. This multiple-critical-path problem is more pronounced in statistical timing analysis, because more paths can have overlapped delay distributions due to large-scale process variations. In addition to this multiple-critical-path problem, we will demonstrate in Section IV-B that selecting the gates and interconnects along the most critical (or least critical) path for up (or down) sizing is not the best choice under a statistical modeling assumption.

IV. Concept of Statistical Timing Sensitivity

In this section, we mathematically define various statistical timing sensitivities and theoretically prove the equivalence between sensitivity and probability.

A. Path Sensitivity

In nominal timing analysis, the critical path is of great interest because it uniquely determines the maximal circuit delay. If the delay of the critical path is increased (or decreased) by a small perturbation $\delta \to 0$, the maximal circuit delay is correspondingly increased (or decreased) by δ. Therefore, given the maximal circuit delay D in (13), the relation between D and the individual path delay D_{Pi} can be mathematically represented as the path sensitivity

$$S_{Pi}^{\rm Path} = \frac{\partial D}{\partial D_{Pi}} = \begin{cases} 1, & (\text{If } P_i \text{ is critical}) \\ 0, & (\text{Otherwise}). \end{cases} \tag{14}$$

From the sensitivity point of view, a critical path is important because it has nonzero sensitivity and all other noncritical

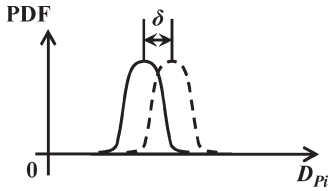


Fig. 4. Path sensitivity in (16) is defined by a small perturbation $\delta \to 0$ on $E(D_{Pi})$ while keeping all high-order central moments $\{E\{[D_{Pi} - E(D_{Pi})]^n\}; n = 2, 3, \ldots\}$ unchanged.

paths have zero sensitivity. The maximal circuit delay can be changed if and only if the critical path delay is changed. This is the underlying reason why the critical path is important for timing optimization. It is the sensitivity, instead of the critical path itself, that provides an important criterion to guide timing optimization. A path is more (or less) important if it has a larger (or smaller) path sensitivity.

In statistical timing analysis, all path delays are random variables and, therefore, can be characterized by their moments [28]. The relation between the maximal circuit delay D and the individual path delay D_{Pi} can be represented by a multidimensional multioutput function that maps the moments of D_{Pi} to the moments of D

$$\begin{bmatrix} E(D_{Pi}) \\ E\left\{[D_{Pi} - E(D_{Pi})]^2\right\} \\ E\left\{[D_{Pi} - E(D_{Pi})]^3\right\} \\ \vdots \end{bmatrix} \to \begin{bmatrix} E(D) \\ E\left\{[D - E(D)]^2\right\} \\ E\left\{[D - E(D)]^3\right\} \\ \vdots \end{bmatrix}. \tag{15}$$

In general, it is possible to define the sensitivity between any mth-order moment of D and nth-order moment of D_{Pi}, where $m, n \in \{1, 2, \ldots\}$. In this paper, we define the path sensitivity by the first-order moments as (16), shown at the bottom of the page.

There are two important clarifications that must be made for the path sensitivity in (16). First, the function in (15) depends on multiple variables: $E(D_{Pi})$ and $\{E\{[D_{Pi} - E(D_{Pi})]^n\}; n = 2, 3, \ldots\}$. When we change $E(D_{Pi})$ by $\delta \to 0$ to calculate the partial derivative in (16), we should keep all other input variables, i.e., all high-order central moments $\{E\{[D_{Pi} - E(D_{Pi})]^n\}; n = 2, 3, \ldots\}$, unchanged. In other words, such a perturbation only shifts the probability distribution by a small amount δ, whereas the shape of the distribution (determined by all high-order central moments) is not changed, as shown in Fig. 4. Second, the perturbation of δ in (16) is defined mathematically. It only changes D_{Pi} and does not impact $\{D_{Pj}; j \neq i\}$. This is different from a perturbation that is physically applied to an arc delay or a process parameter. Such a physical perturbation can concurrently change the delays of multiple paths. These two clarifications are also applicable to other statistical timing sensitivities defined in this section.

The path sensitivity in (16) has several important properties that are summarized by the following theorems.

Theorem 1: The path sensitivity in (16) satisfies

$$\sum_i S_{Pi}^{\text{Path}} = 1. \tag{17}$$

Theorem 2: Given the maximal circuit delay $D = \text{MAX}(D_{P1}, D_{P2}, \ldots)$ where D_{Pi} is the delay of the ith path, if the probability $P[D_{Pi} = \text{MAX}(D_{Pj}; j \neq i)]$ is equal to 0, then the path sensitivity in (14) is equal to the probability that the path P_i is critical, i.e.,

$$S_{Pi}^{\text{Path}} = P(D_{Pi} \geq D_{P1} \quad \& \quad D_{Pi} \geq D_{P2} \quad \& \quad \cdots). \tag{18}$$

The detailed proofs of Theorems 1 and 2 can be found in the Appendix. Theorem 2 relies on the assumption $P[D_{Pi} = \text{MAX}(D_{Pj}; j \neq i)] = 0$. This assumption is valid if any two paths in the circuit are not exactly identical. This conclusion can be summarized by the following Theorem 3 that is formally proven in the Appendix.

Theorem 3: Let D_{Pi} be the delay of the ith path. The probability $P[D_{Pi} = \text{MAX}(D_{Pj}; j \neq i)] = 0$ for any $\{i = 1, 2, \ldots\}$ if the probability $P(D_{Pi} = D_{Pj}) = 0$ for any $i \neq j$.

If D_{Pi} and D_{Pj} are two continuous random variables and they are not fully correlated, the probability $P(D_{Pi} = D_{Pj})$ is equal to 0. In most practical applications, path delays are impacted by both interdie and intradie variations. Even if two path delays have the same mean and variance, they often depend on different intradie variations due to the location difference and, therefore, are not fully correlated.

B. Arc Sensitivity

In nominal timing optimization, the gates and interconnects along the critical path are important, because the maximal circuit delay is sensitive to these gate/interconnect delays. Following this observation, the importance of a given gate or interconnect can be assessed by the following arc sensitivity:

$$S_{Ai}^{\text{Arc}} = \frac{\partial D}{\partial D_{Ai}} = \sum_k \frac{\partial D}{\partial D_{Pk}} \cdot \frac{\partial D_{Pk}}{\partial D_{Ai}} = \sum_k S_{Pk}^{\text{Path}} \cdot \frac{\partial D_{Pk}}{\partial D_{Ai}}$$
$$= \begin{cases} 1, & (A_i \text{ is on the critical path}) \\ 0, & (\text{Otherwise}) \end{cases} \tag{19}$$

where D is the maximal circuit delay defined in (13), D_{Ai} denotes the gate/interconnect delay associated with the ith arc, and D_{Pk} represents the delay of the kth path. In (19), the path sensitivity S_{Pk}^{Path} is nonzero (i.e., equal to 1) if and only if the kth path P_k is critical. In addition, the derivative $\partial D_{Pk}/\partial D_{Ai}$ is nonzero (i.e., equal to 1) if and only if the ith arc A_i sits on the kth path P_k, because the path delay D_{Pk} is equal to the sum of all arc delays D_{Ai}'s that belong to P_k. These observations

$$S_{Pi}^{\text{Path}} = \frac{\partial E(D)}{\partial E(D_{Pi})} = \lim_{\delta \to 0} \frac{E\left[\text{MAX}(D_{p1}, \ldots, D_{Pi} + \delta, \ldots)\right] - E\left[\text{MAX}(D_{p1}, \ldots, D_{Pi}, \ldots)\right]}{\delta} \tag{16}$$

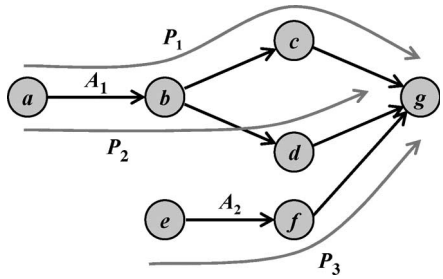


Fig. 5. Simple timing graph to illustrate the application of the proposed arc sensitivity.

yield the conclusion that the arc sensitivity S_{Ai}^{Arc} is nonzero if and only if A_i is on the critical path. The arc sensitivity explains the reason why the gates and interconnects along the critical path are important for timing optimization. A gate/interconnect is more (or less) important if it has a larger (or smaller) arc sensitivity.

The aforementioned sensitivity concept can be extended to statistical timing analysis. In the statistical case, we define the arc sensitivity using the first-order moments

$$S_{Ai}^{\text{Arc}} = \frac{\partial E(D)}{\partial E(D_{Ai})}. \tag{20}$$

The arc sensitivity in (20) has the following property.

Theorem 4: Let D_{Pi} be the delay of the ith path. If the probability $P[D_{Pi} = \text{MAX}(D_{Pj}; j \neq i)] = 0$ for any $\{i = 1, 2, \ldots\}$, then the arc sensitivity in (20) is equal to the following:

$$S_{Ai}^{\text{Arc}} = \sum_{Ai \in Pk} S_{Pk}^{\text{Path}}. \tag{21}$$

The detailed proof of Theorem 4 can be found in the Appendix. Remember that S_{Pk}^{Path} is equal to the probability that the kth path P_k is critical (see Theorem 2). Therefore, Theorem 4 implies the important fact that the arc sensitivity defined in (20) is exactly equal to the probability that an arc sits on the critical path.

The arc sensitivity in (20) provides an effective criterion to select the most important gates and interconnects for up-/downsizing. Roughly speaking, for statistical timing optimization, the gates and interconnects with large arc sensitivities are critical to the maximal circuit delay and, in general, should be upsized to reduce delay, whereas the others with small arc sensitivities should be downsized to save area and/or power. Next, using the concept of arc sensitivity, we explain the reason why repeatedly selecting the gates and interconnects along the most critical (or least critical) path for up (or down) sizing can be ineffective in the statistical case.

Consider a simple timing graph including three paths, as shown in Fig. 5. Assume that the path sensitivity $S_{P1}^{\text{Path}} = S_{P2}^{\text{Path}} = 0.3$ and $S_{P3}^{\text{Path}} = 0.4$. Therefore, P_3 is the most critical path because it has the largest path sensitivity and is most likely to have the maximal delay. Using the traditional concept of critical path, the arc A_2 should be selected for upsizing to reduce delay. However, according to Theorem 4, it is easy to verify that $S_{A1}^{\text{Arc}} = S_{P1}^{\text{Path}} + S_{P2}^{\text{Path}} = 0.6$ and $S_{A2}^{\text{Arc}} = S_{P3}^{\text{Path}} = 0.4$. The arc A_1 has a more significant impact on the maximal circuit delay and should be selected for upsizing, although it does not sit on the most critical path. In this example, using the traditional concept of critical path selects the wrong arc, because it does not consider the nonzero path sensitivities of other less critical paths. These nonzero sensitivities make it possible that changing an arc delay can change the maximal circuit delay through multiple paths. In Fig. 5, the arc A_1 can change the maximal circuit delay through two paths P_1 and P_2, whereas the arc A_2 can change the maximal circuit delay only through one path P_3. Therefore, the arc A_1 eventually becomes more critical than A_2, although neither P_1 nor P_2 is the most critical path.

C. Node Sensitivity

The nominal and statistical node sensitivities can be, respectively, defined as

$$\begin{aligned} S_{Vi}^{\text{Node}} &= \frac{\partial D}{\partial \text{AT}_{Vi}} \\ &= \sum_k \frac{\partial D}{\partial D_{Pk}} \cdot \frac{\partial D_{Pk}}{\partial \text{AT}_{Vi}} \\ &= \sum_k S_{Pk}^{\text{Path}} \cdot \frac{\partial D_{Pk}}{\partial \text{AT}_{Vi}} \\ &= \begin{cases} 1, & (V_i \text{ is on the critical path}) \\ 0, & (\text{Otherwise}) \end{cases} \end{aligned} \tag{22}$$

$$S_{Vi}^{\text{Node}} = \frac{\partial E(D)}{\partial E(\text{AT}_{Vi})} \tag{23}$$

where D is the maximal circuit delay given in (13), AT_{Vi} denotes the arrival time at the ith node, and D_{Pk} represents the delay of the kth path.

The node sensitivities in (22) and (23) are similar to the arc sensitivities defined in (19) and (20). Following the same reasoning of Theorem 4, we can prove that the statistical node sensitivity in (23) is exactly equal to the probability that a node sits on the critical path. This conclusion can be formally stated as the following Theorem 5. The detailed proof of Theorem 5 can be found in the Appendix.

Theorem 5: Let D_{Pi} be the delay of the ith path. If the probability $P[D_{Pi} = \text{MAX}(D_{Pj}; j \neq i)] = 0$ for any $\{i = 1, 2, \ldots\}$, then the node sensitivity in (23) is equal to the following:

$$S_{Vi}^{\text{Node}} = \sum_{Vi \in Pk} S_{Pk}^{\text{Path}}. \tag{24}$$

D. Yield Sensitivity

We define the yield sensitivities for paths, arcs, and nodes

$$\text{SY}_{Pi}^{\text{Path}} = \frac{\partial \text{Yield}}{\partial E(D_{Pi})} \tag{25}$$

$$\text{SY}_{Ai}^{\text{Arc}} = \frac{\partial \text{Yield}}{\partial E(D_{Ai})} \tag{26}$$

$$\text{SY}_{Vi}^{\text{Node}} = \frac{\partial \text{Yield}}{\partial E(\text{AT}_{Vi})}. \tag{27}$$

The yield sensitivities in (25)–(27) quantitatively model how the parametric timing yield changes if $E(D_{Pi})$, $E(D_{Ai})$, or $E(\mathrm{AT}_{Vi})$ is changed. Note that, due to the nonlinearity of the MAX($\bullet$) operation, a small perturbation in $E(D_{Pi})$, $E(D_{Ai})$, or $E(\mathrm{AT}_{Vi})$ not only changes the mean value of the maximal circuit delay D but also changes its variance. Whereas such a variance change is ignored by the sensitivities defined in Sections IV-A–C, it can be captured by the yield sensitivities in (25)–(27).

E. High-Order Sensitivity

The aforementioned sensitivity concept can be extended to high order. One important application of high-order sensitivity is the quadratic MAX($\bullet$) approximation proposed in [24].

For statistical timing analysis, the nonlinear MAX($\bullet$) operator can be approximated as a linear function in (6), where the linear coefficients α and β are determined by the tightness probabilities in (8) and (9). Given the equivalence between probability and sensitivity proven by Theorem 2, the tightness probabilities in (8) and (9) are equal to the first-order sensitivities

$$\alpha = \mathrm{PROB}(x \geq y) = \frac{\partial E\left[\mathrm{MAX}(x,y)\right]}{\partial E(x)} \tag{28}$$

$$\beta = \mathrm{PROB}(y \geq x) = \frac{\partial E\left[\mathrm{MAX}(x,y)\right]}{\partial E(y)}. \tag{29}$$

Although the MAX($\bullet$) operator is not analytical (i.e., does not have continuous derivatives), it can be statistically approximated as the form of (6), (28) and (29) that is similar to the traditional Taylor expansion. Therefore, such a linear approximation is referred to as the first-order statistical Taylor expansion in [24].

The aforementioned statistical Taylor expansion can be extended to the second order to achieve higher approximation accuracy. Consider the simple example MAX$(0, z)$ where z is a zero-mean random variable. The second-order statistical expansion can be expressed as follows:

$$\mathrm{MAX}(0,z) = 0.5 \cdot \frac{\partial^2 E\left[\mathrm{MAX}(0,z)\right]}{\partial\left[E(z)\right]^2} \cdot z^2 + \frac{\partial E\left[\mathrm{MAX}(0,z)\right]}{\partial\left[E(z)\right]} \cdot z + C_{\mathrm{MAX}(0,z)}. \tag{30}$$

The details of the quadratic MAX($\bullet$) approximation is beyond the scope of this paper and can be found in [24].

F. Summary

The proposed sensitivity framework has three unique properties.

1) Distribution-independent. Our discussions do not rely on any specific probability distribution to model the gate/interconnect delays and the arrival times.
2) Correlation-aware. Our proposed sensitivity framework is not restricted to any assumption of statistical independence, and it can handle correlated arrival times.
3) Computation-efficient. The proposed statistical timing sensitivities can be efficiently computed by a single breadth-first graph traversal, as will be demonstrated in Section V.

V. Computation of Statistical Timing Sensitivity

In this section, we first develop the sensitivity equations for two atomic operations: SUM($\bullet$) and MAX($\bullet$). Then, we show how to propagate the sensitivities throughout a timing graph by using a single breadth-first graph traversal. Finally, we discuss the incremental analysis algorithm to quickly update the statistical timing and sensitivity information after changes to a circuit are made.

We assume that all gate/interconnect delays and arrival times are approximated as normal distributions. Such an assumption facilitates an efficient sensitivity computation, even though our proposed sensitivity framework is distribution-independent.

A. Atomic Operations

Because multivariable operations can be broken down into multiple two-variable cases, the remainder of this section focuses on the sensitivity computation for the SUM($\bullet$) and MAX($\bullet$) of two random variables, i.e., $z = x + y$ and $z = \mathrm{MAX}(x, y)$ where x and y are approximated as the linear models in (3) and (4) and z is similarly approximated as a linear function

$$z = B_z^{\mathrm{T}} \cdot \Delta\varepsilon + C_z = \sum_{i=1}^{N} B_{zi} \cdot \Delta\varepsilon_i + C_z. \tag{31}$$

Given the operation $z = x + y$ or $z = \mathrm{MAX}(x, y)$, we define the sensitivity matrix $Q_{z \leftarrow x}$ as follows:

$$Q_{z\leftarrow x} = \begin{bmatrix} \partial C_z/\partial C_x & \partial C_z/\partial B_{x1} & \cdots & \partial C_z/\partial B_{xN} \\ \partial B_{z1}/\partial C_x & \partial B_{z1}/\partial B_{x1} & \cdots & \partial B_{z1}/\partial B_{xN} \\ \vdots & \vdots & \vdots & \vdots \\ \partial B_{zN}/\partial C_x & \partial B_{zN}/\partial B_{x1} & \cdots & \partial B_{zN}/\partial B_{xN} \end{bmatrix}. \tag{32}$$

The sensitivity matrix $Q_{z \leftarrow y}$ can be similarly defined.

The sensitivity matrix in (32) provides the quantitative information on how much the coefficients C_z or $\{B_{zi}; i = 1, 2, \ldots, N\}$ will be changed if there is a small perturbation on C_x or $\{B_{xi}; i = 1, 2, \ldots, N\}$. Next, we derive the mathematical formulas of the sensitivity matrices for both SUM($\bullet$) and MAX($\bullet$) operations.

For the SUM($\bullet$) operation $z = x + y$, it is easy to verify that

$$C_z = C_x + C_y \tag{33}$$

$$B_{zi} = B_{xi} + B_{yi} \qquad (i = 1, 2, \ldots, N). \tag{34}$$

Therefore, the sensitivity matrix $Q_{z \leftarrow x}$ is an identity matrix.

For the MAX($\bullet$) operation $z = \text{MAX}(x, y)$, it can be proven that

$$\frac{\partial C_z}{\partial C_x} = \Phi(\beta) \tag{35}$$

$$\frac{\partial C_z}{\partial B_{xi}} = \frac{\partial B_{zi}}{\partial C_x} = \frac{\varphi(\beta) \cdot (B_{xi} - B_{yi})}{\alpha} \quad (i = 1, \ldots, N) \tag{36}$$

$$\frac{\partial B_{zi}}{\partial B_{xi}} = \Phi(\beta) - \frac{\beta \cdot \varphi(\beta) \cdot (B_{xi} - B_{yi})^2}{\alpha^2} \quad (i = 1, \ldots, N) \tag{37}$$

$$\frac{\partial B_{zi}}{\partial B_{xj}} = -\frac{\beta \cdot \varphi(\beta) \cdot (B_{xi} - B_{yi}) \cdot (B_{xj} - B_{yj})}{\alpha^2} \quad \begin{pmatrix} i = 1, \ldots, N \\ j = 1, \ldots, N \\ i \neq j \end{pmatrix} \tag{38}$$

where $\varphi(\bullet)$ and $\Phi(\bullet)$ are the probability density function (PDF) and the cumulative distribution function (CDF) of the standard normal distribution $N(0, 1)$, respectively, and the coefficients α and β are defined by the following:

$$\alpha = \sqrt{\sum_{i=1}^{N} (B_{xi} - B_{yi})^2} \tag{39}$$

$$\beta = \frac{C_x - C_y}{\alpha}. \tag{40}$$

Equations (35)–(40) can be derived by directly following the mathematical equations in [26]. The sensitivity matrix $Q_{z \leftarrow y}$ can be similarly calculated because both SUM($\bullet$) and MAX($\bullet$) are symmetric.

Finally, it is worth mentioning that the sensitivity matrix defined by (35)–(40) is an approximation for the MAX($\bullet$) operation, because a simple linear function is used in (31) to approximate the nonlinear operation $z = \text{MAX}(x, y)$. It can further be shown that, when a multivariable MAX($\bullet$) is broken down into multiple two-variable operations, the approximation error depends on the ordering of these two-variable operations [25]. More details on this ordering issue are beyond the scope of this paper and will be considered in our future research.

B. Sensitivity Propagation

Once the atomic operations are available, they can be applied to propagate the sensitivity matrices throughout a timing graph. Next, we use the simple timing graph in Fig. 1 as an example to illustrate the key idea of sensitivity propagation.

1) Start from the MAX($\bullet$) operation at the sink node, i.e., $D = \text{MAX}[\text{AT}(f) + D(f, \text{sink}), \text{AT}(g) + D(g, \text{sink})]$ where D denotes the arrival time at the sink node (i.e., the maximal circuit delay), AT(i) represents the arrival time at node i, and $D(i, j)$ stands for the delay of the arc $\langle i, j \rangle$. Compute the sensitivity matrices $Q_{D \leftarrow [\text{AT}(f) + D(f, \text{sink})]}$ and $Q_{D \leftarrow [\text{AT}(g) + D(g, \text{sink})]}$ using (35)–(38).
2) Propagate $Q_{D \leftarrow [\text{AT}(f) + D(f, \text{sink})]}$ to the node f through the arc $\langle f, \text{sink} \rangle$. Based on the chain rule of derivatives

$$Q_{D \leftarrow \text{AT}(f)} = Q_{D \leftarrow [\text{AT}(f) + D(f, \text{sink})]} \cdot Q_{[\text{AT}(f) + D(f, \text{sink})] \leftarrow \text{AT}(f)}$$

and

$$Q_{D \leftarrow D(f, \text{sink})} = Q_{D \leftarrow [\text{AT}(f) + D(f, \text{sink})]} \cdot Q_{[\text{AT}(f) + D(f, \text{sink})] \leftarrow D(f, \text{sink})}.$$

$Q_{[\text{AT}(f) + D(f, \text{sink})] \leftarrow \text{AT}(f)}$ and $Q_{[\text{AT}(f) + D(f, \text{sink})] \leftarrow D(f, \text{sink})}$ are two identity matrices due to the SUM($\bullet$) operation.
3) Similarly, propagate $Q_{D \leftarrow [\text{AT}(g) + D(g, \text{sink})]}$ to the node g through the arc $\langle g, \text{sink} \rangle$. Determine $Q_{D \leftarrow \text{AT}(g)}$ and $Q_{D \leftarrow D(g, \text{sink})}$.
4) Propagate $Q_{D \leftarrow \text{AT}(f)}$ and $Q_{D \leftarrow \text{AT}(g)}$ to the node e, yielding $Q_{D \leftarrow D(e,f)} = Q_{D \leftarrow \text{AT}(f)}$, $Q_{D \leftarrow D(e,g)} = Q_{D \leftarrow \text{AT}(g)}$, and $Q_{D \leftarrow \text{AT}(e)} = Q_{D \leftarrow \text{AT}(f)} + Q_{D \leftarrow \text{AT}(g)}$. Note that the outdegree of the node e is equal to two. Therefore, the sensitivity matrices $Q_{D \leftarrow \text{AT}(f)}$ and $Q_{D \leftarrow \text{AT}(g)}$ should be added together at the node e to compute $Q_{D \leftarrow \text{AT}(e)}$, based on the chain rule of derivatives. Its physical meaning is that a small perturbation on AT(e) can change the maximal circuit delay D through two different paths $\{e \rightarrow f \rightarrow \text{sink}\}$ and $\{e \rightarrow g \rightarrow \text{sink}\}$.
5) Continue propagating the sensitivity matrices until the source node is reached.

After the sensitivity propagation is complete, the sensitivity matrix $Q_{D \leftarrow D(i,j)}$ (or $Q_{D \leftarrow \text{AT}(Vi)}$) between the maximal circuit delay D and any arc delay $D(i, j)$ (or node arrival time AT(V_i)) is determined. The statistical timing sensitivities can be easily computed by a quick postprocessing. For example, the arc sensitivity defined in (20) and the node sensitivity defined in (23) are the (1, 1)th element of $Q_{D \leftarrow D(i,j)}$ and $Q_{D \leftarrow \text{AT}(Vi)}$, respectively

$$S^{\text{Arc}}_{\langle i,j \rangle} = [1 \quad 0 \quad \cdots] \cdot Q_{D \leftarrow D(i,j)} \cdot [1 \quad 0 \quad \cdots]^{\text{T}} \tag{41}$$

$$S^{\text{Node}}_{Vi} = [1 \quad 0 \quad \cdots] \cdot Q_{D \leftarrow \text{AT}(Vi)} \cdot [1 \quad 0 \quad \cdots]^{\text{T}}. \tag{42}$$

Calculating the yield sensitivities in (26) and (27) is more comprehensive because the parametric timing yield is determined by not only the mean value of the maximal circuit delay D but also its variance.

After the statistical timing analysis is complete, D is approximated as the following linear model that is similar to (3) and (4) and (31):

$$D = B_D^{\text{T}} \cdot \Delta\varepsilon + C_D = \sum_{i=1}^{N} B_{Di} \cdot \Delta\varepsilon_i + C_D. \tag{43}$$

Because D is the linear combination of multiple normal distributions, it is also normal and its mean and standard deviations are, respectively, determined by the following [30]:

$$\mu_D = C_D \tag{44}$$

$$\sigma_D = \sqrt{\sum_{i=1}^{N} B_{Di}^2}. \tag{45}$$

Therefore, the CDF of D is equal to the following:

$$\text{cdf}_D(t) = \Phi\left(\frac{t - \mu_D}{\sigma_D}\right) \tag{46}$$

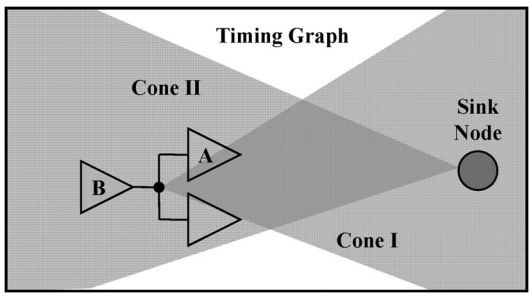


Fig. 6. Incremental statistical timing and sensitivity analysis.

where $\Phi(\bullet)$ stands for the CDF of the standard normal distribution $N(0,1)$.

Assume that the timing constraint is specified by the following:

$$D \leq D_{\text{Spec}} \tag{47}$$

and therefore, the parametric timing yield is equal to the following:

$$\text{Yield} = P(D \leq D_{\text{Spec}}) = \text{cdf}_D(D_{\text{Spec}}) = \Phi\left(\frac{D_{\text{Spec}} - \mu_D}{\sigma_D}\right). \tag{48}$$

We further assume that x denotes the arc delay or the arrival time of interest and that it is approximated as the linear model in (3). Hence, the yield sensitivity can be calculated as follows:

$$\frac{\partial \text{Yield}}{\partial E(x)} = \frac{\partial}{\partial E(x)}\Phi\left(\frac{D_{\text{Spec}} - \mu_D}{\sigma_D}\right). \tag{49}$$

Based on the chain rule of derivatives, we have

$$\frac{\partial \text{Yield}}{\partial E(x)} = \varphi\left(\frac{D_{\text{Spec}} - \mu_D}{\sigma_D}\right) \times \left[\frac{\mu_D - D_{\text{Spec}}}{\sigma_D^3} \cdot \sum_{i=1}^{N} B_{Di} \cdot \frac{\partial B_{Di}}{\partial C_x} - \frac{1}{\sigma_D} \cdot \frac{\partial C_D}{\partial C_x}\right] \tag{50}$$

where $\varphi(\bullet)$ represents the PDF of the standard normal distribution $N(0,1)$ and the derivatives $\{\partial B_{Di}/\partial C_x; i = 1, 2, \ldots, N\}$ and $\partial C_D/\partial C_x$ are the elements of the sensitivity matrix $Q_{D \leftarrow x}$ that is extracted from the sensitivity propagation.

C. Incremental Analysis

The complete statistical timing and sensitivity analysis consists of one forward arrival time propagation from the source node to the sink node and one backward sensitivity propagation from the sink node to the source node. It would be quite expensive, if not impossible, to run such a complete analysis for multiple times within an optimization loop. Therefore, an incremental analysis technique is required to quickly update the statistical timing and sensitivity information after local changes to a circuit are made.

Once a logic cell is modified for timing optimization, the arrival time and the timing sensitivity of a number of nodes are changed. Taking Fig. 6 as an example, if we size logic cell A, the input capacitance, delay, and output slew of cell A are all changed. Due to the input-capacitance change of cell A, the delay and output slew of its fan-in cell (i.e., cell B in Fig. 6) are also changed. Therefore, the arrival time of the fan-out cone of cell B (i.e., cone I in Fig. 6) must be updated, and the timing sensitivity of the fan-in cone of all affected nodes (i.e., cone II in Fig. 6) must also be updated.

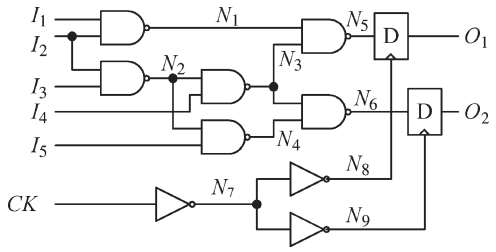


Fig. 7. Circuit schematic of a simple digital circuit.

TABLE I
ARC SENSITIVITY VALUES FOR THE SIMPLE DIGITAL CIRCUIT (SHOWN ARE THE ARCS WITH NONZERO SENSITIVITIES ONLY)

Arc	Proposed Algorithm	Monte Carlo
$\langle I_3, N_2\rangle$	100%	100%
$\langle N_2, N_4\rangle$	0.1%	0.1%
$\langle N_3, N_6\rangle$	29.1%	27.5%
$\langle CK, N_7\rangle$	100%	100%
$\langle N_7, N_9\rangle$	29.2%	27.6%
$\langle N_2, N_3\rangle$	99.9%	99.9%
$\langle N_3, N_5\rangle$	70.8%	72.4%
$\langle N_4, N_6\rangle$	0.1%	0.1%
$\langle N_7, N_8\rangle$	70.8%	72.4%

VI. NUMERICAL EXAMPLES

We demonstrate the efficacy of the proposed statistical timing sensitivity analysis using several circuit examples. All circuits are implemented in either 0.13-μm or 90-nm commercial CMOS technologies. Both interdie and intradie variations on V_{TH}, T_{OX}, W, and L are considered. The probability distribution and the correlation information of these variations are specified in the process design kit from the foundry. All numerical simulations are run on a 2.6-GHz computer with 1-GB memory.

A. Simple Example

Shown in Fig. 7 is a simple digital circuit that consists of nine gates and two D-flip-flops. Such a simple example allows us to intuitively illustrate several key concepts of the proposed sensitivity analysis.

Table I shows the arc sensitivity values computed by the proposed sensitivity analysis and a Monte Carlo simulation with 10^4 samples. The Monte Carlo simulation repeatedly draws random samples and counts the probability that an arc sits on the critical path. Note that the largest arc sensitivity error in Table I is only 1.6%. Such a high accuracy demonstrates that the normal distribution assumption applied to our sensitivity analysis does not incur significant error in this example.

As shown in Table I, $\langle I_3, N_2\rangle$ is the arc that has the largest sensitivity value. This is because $\langle I_3, N_2\rangle$ sits on the three longest paths: $\{I_3 \rightarrow N_2 \rightarrow N_3 \rightarrow N_5\}$, $\{I_3 \rightarrow N_2 \rightarrow N_3 \rightarrow N_6\}$, and $\{I_3 \rightarrow N_2 \rightarrow N_4 \rightarrow N_6\}$. Therefore, a small

TABLE II
STATISTICAL SENSITIVITY ANALYSIS ERROR FOR ISCAS'85 BENCHMARK CIRCUITS

CKT	Min	Avg	Max
c432	0.0%	0.1%	1.6%
c499	0.0%	0.1%	2.4%
c880	0.0%	0.9%	1.3%
c1355	0.4%	0.9%	2.5%
c1908	0.0%	0.4%	3.4%
c2670	0.0%	0.3%	2.6%
c3540	0.0%	0.3%	2.4%
c5315	0.8%	1.8%	2.8%
c6288	0.0%	0.6%	1.9%
c7552	0.7%	1.1%	3.5%

TABLE III
STATISTICAL TIMING AND SENSITIVITY ANALYSIS COST FOR ISCAS'85 BENCHMARK CIRCUITS

CKT	# of RVs	Proposed Algorithm (Sec.)		Monte Carlo (Sec.)
		Timing	Sensitivity	
c432	0.8K	0.01	0.01	128
c499	0.9K	0.02	0.02	154
c880	1.7K	0.03	0.02	281
c1355	2.7K	0.05	0.03	359
c1908	3.8K	0.07	0.06	504
c2670	5.2K	0.09	0.05	771
c3540	7.1K	0.11	0.06	974
c5315	10.6K	0.17	0.11	1381
c6288	12.5K	0.25	0.11	1454
c7552	15.1K	0.26	0.14	1758

perturbation on the delay of $\langle I_3, N_2\rangle$ can significantly change the maximal circuit delay through these three paths. Note that, although such a multiple-path effect cannot be easily identified by a nominal timing analysis, it is successfully captured by the proposed statistical sensitivity analysis.

In addition, it is worth mentioning that the arc $\langle I_2, N_2\rangle$ in Fig. 7 has zero sensitivity, because the NAND gate is asymmetric and the arc delay $D(I_3, N_2)$ is larger than $D(I_2, N_2)$. Even with process variations, $D(I_3, N_2)$ still dominates, because $D(I_2, N_2)$ and $D(I_3, N_2)$ are from the same gate and they are strongly correlated.

B. ISCAS'85 Benchmark Circuits

1) Accuracy and Speed: We conducted statistical timing and sensitivity analysis for the ISCAS'85 benchmark circuits. Table II shows the minimal, average, and maximal sensitivity errors of all timing arcs. These errors are compared against a Monte Carlo simulation with 10^4 samples. Note that the maximal sensitivity error in Table II is less than 3.5% for all circuits. In addition, the proposed sensitivity analysis achieves about 4000 times speedup over the Monte Carlo simulation, as shown in Table III. To fully understand the computational complexity, Table III also lists the number of independent random variables (after PCA) to model both interdie and intradie variations. It is important to note that the proposed statistical sensitivity analysis is slightly cheaper than the statistical timing analysis in this example. The reason is that the proposed sensitivity analysis only involves simple matrix operations, whereas the statistical timing analysis spends substantial computational time on delay calculation (e.g., computing the effective capacitance C_{eff} for interconnects via a number of numerical iterations [27]).

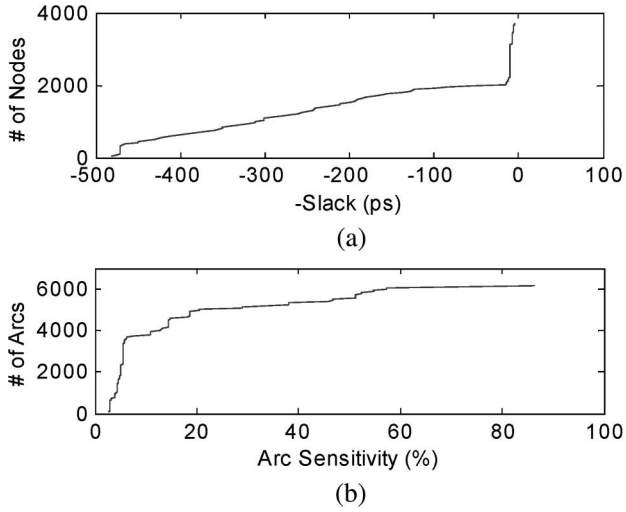


Fig. 8. (a) Cumulative plot of the nominal slacks for ISCAS'85 C7552. (b) Cumulative plot of the statistical sensitivities for ISCAS'85 C7552.

2) Slack and Sensitivity Wall: One important problem in nominal timing optimization is the steep slack wall discussed in [23]. After the nominal timing optimization is complete, many paths have similar delays and are equally critical. We nominally optimized the circuit C7552 and plotted the optimized slacks in Fig. 8(a). (Note that Fig. 8(a) is plotted for "-Slack".) The steep slack wall in Fig. 8(a) implies that a great number of nodes have close-to-zero slacks and, therefore, are equally important in nominal timing optimization.

Next, we ran a statistical sensitivity analysis for the same circuit and plotted the arc sensitivities in Fig. 8(b). Note that the sensitivity wall in Fig. 8(b) is flat. In other words, after process variations are considered, only a small number of arcs dominate the overall timing performance. Although these arcs cannot be identified by nominal timing analysis, they are captured by the proposed statistical sensitivity analysis.

3) Statistical Timing Optimization: We further incorporated the proposed sensitivity analysis into an optimization engine for statistical gate sizing. Because timing optimization is not the major focus of this paper, we only implemented a simple select-and-conquer approach. Namely, we select a small number of the most and least critical cells based on yield sensitivities. The most critical cells are upsized to reduce delay, and the least critical cells are downsized to reduce area and/or power.

For testing and comparison, we applied both corner-based optimization and statistical optimization to all ISCAS'85 benchmark circuits. In both optimizations, the objective is to minimize the total gate area with a given timing constraint. Table IV shows the total gate area after the optimizations are complete. In this example, the statistical timing optimization achieves up to 27.8% area reduction compared with the corner-based method.

C. Industrial Design Examples

1) Statistical Sensitivity Analysis: As a final example, we tested the proposed sensitivity analysis on three large industrial

TABLE IV
NORMALIZED GATE AREA AFTER TIMING OPTIMIZATION FOR ISCAS'85 BENCHMARK CIRCUITS

CKT	Corner	Statistical (Yield = 99%)	Difference (%)
c432	414.75	365.25	11.93
c499	652.00	571.00	12.42
c880	511.00	469.50	8.12
c1355	768.00	675.00	12.11
c1908	747.75	591.50	20.90
c2670	1328.75	1012.00	23.84
c3540	1784.75	1381.50	22.59
c5315	2949.75	2130.50	27.77
c6288	3556.50	3543.50	0.37
c7552	3808.25	2905.75	23.70

TABLE V
STATISTICAL TIMING AND SENSITIVITY ANALYSIS COST FOR LARGE INDUSTRIAL DESIGN EXAMPLES

Design	# of Cells	# of Pins	# of RVs	Time (Sec.) Computational Timing	Time (Sec.) Sensitivity
A	16K	62K	32K	2.4	1.9
B	60K	220K	120K	7.2	5.17
C	330K	1.3M	660K	92.6	75.6

examples. Table V shows the circuit size (i.e., the number of cells, the number of pins, and the number of independent random variables to model both interdie and intradie variations) and the computational cost for these examples. The Monte Carlo simulation is too expensive for these large-size examples and, therefore, is not computationally feasible. As shown in Table V, the computational cost of the proposed sensitivity analysis linearly scales as the circuit size increases (up to 1.3M pins).

2) Statistical Timing Optimization: We further ran a statistical timing optimization for design A that contains 16K cells. The proposed yield sensitivity is utilized as a criterion to select the most critical cells for upsizing and the least critical cells for downsizing. The statistical timing optimization is formulated to minimize the total gate area with a given timing constraint. For the initial design, the normalized gate area is 803 758. The gate area is reduced to 712 221 (11.39% difference) by the statistical timing optimization, whereas the parametric timing yield is guaranteed to be 99%.

Fig. 9 shows the histogram of the mean value of all node slacks before and after the statistical timing optimization is applied. It is apparent that our timing optimization pushes the slack values toward zero to reduce area. However, these slack changes all happen at noncritical nodes, and therefore, no parametric timing yield is surrendered. It is also interesting to note that a number of slack values in Fig. 9(c) are increased after optimization. We believe that it is caused by the load dependence of the delay. Namely, when a cell is downsized to save area, its input capacitance is reduced, which can speed up the driving cell and reduce the total delay.

VII. CONCLUSION

In this paper, we define the statistical timing sensitivities for paths, arcs, and nodes. Our theoretical analysis proves a direct link between probability and sensitivity. An efficient algorithm is developed for fast sensitivity computation. The proposed sensitivity analysis has a linear complexity in circuit size and offers an incremental analysis capability. Our numerical examples demonstrate that the proposed sensitivity analysis yields accurate results and achieves 4000 times speedup over the Monte Carlo simulation with 10^4 samples.

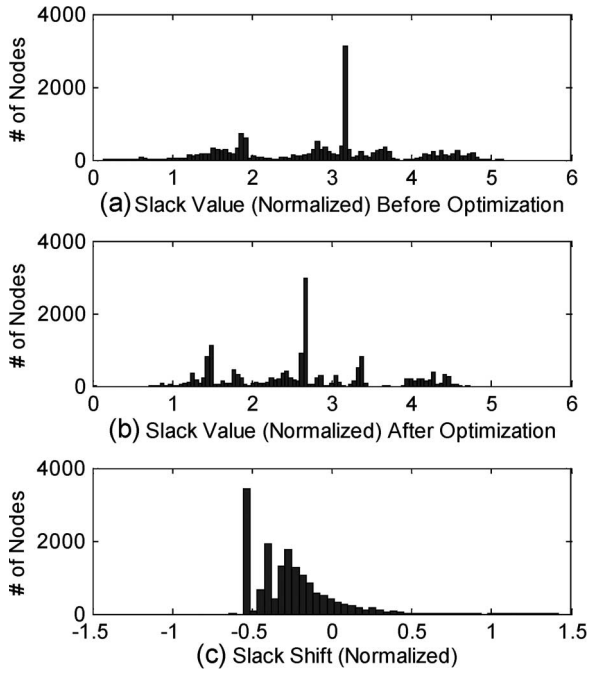


Fig. 9. Node slacks for industrial design A. (a) Histogram of the slack mean value of all nodes before statistical timing optimization. (b) Histogram of the slack mean value of all nodes after statistical timing optimization. (c) Histogram of the slack shift of all nodes.

The proposed sensitivity framework is further incorporated into an optimization engine for statistical gate sizing. Our optimization examples demonstrate that the proposed timing sensitivity can be used to guide statistical gate sizing. Even if a simple sizing algorithm is utilized, the proposed sensitivity-based optimization yields promising results.

APPENDIX

Proof of Theorem 1

Given a small perturbation δ on the mean values of all paths, the mean value of the maximal circuit delay is equal to the following:

$$E\left[\text{MAX}(D_{P1}+\delta, D_{P2}+\delta, \ldots)\right] = E\left[\text{MAX}(D_{P1}, D_{P2}, \ldots)\right] + \delta. \quad (51)$$

According to the path sensitivity definition in (16), the mean value of the maximal circuit delay can also be represented as follows:

$$E\left[\text{MAX}(D_{P1}+\delta, D_{P2}+\delta, \ldots)\right] = E\left[\text{MAX}(D_{P1}, D_{P2}, \ldots)\right] + \sum_i \delta \cdot S_{Pi}^{\text{Path}} + O(\delta^2) \quad (52)$$

where $O(\delta^2)$ is a high-order (order ≥ 2) polynomial of δ. Comparing (51) and (52) yields

$$\delta = \delta \cdot \sum_i S_{Pi}^{\text{Path}} + O(\delta^2). \tag{53}$$

Equation (53) is valid for any sufficiently small δ. Therefore, the first-order coefficient of δ at the left-hand side must equal the first-order coefficient of δ at the right-hand side, yielding

$$1 = \sum_i S_{Pi}^{\text{Path}}. \tag{54}$$

Equation (54) proves Theorem 1.

Studying (53), we would notice another interesting property that the high-order polynomial $O(\delta^2)$ is equal to 0. In other words, there is no high-order term in the Taylor expansion (52). This observation is consistent with the fact that the function $E[\text{MAX}(D_{P1}+\delta, D_{P2}+\delta, \ldots)]$ is actually linear in δ, as shown in (51).

Proof of Theorem 2

Let $A_{Pi} = \text{MAX}(D_{Pj}; j \neq i)$ and we have

$$S_{Pi}^{\text{Path}} = \frac{\partial E\left[\text{MAX}(D_{Pi}, A_{Pi})\right]}{\partial E(D_{Pi})} \tag{55}$$

$$\begin{aligned} &P(D_{Pi} \geq D_{P1} \quad \& \quad D_{Pi} \geq D_{P2} \quad \& \quad \cdots) \\ &\quad = P(D_{Pi} \geq A_{Pi}). \end{aligned} \tag{56}$$

The operation $\text{MAX}(D_{Pi}, A_{Pi})$ can be rewritten as follows:

$$\text{MAX}(D_{Pi}, A_{Pi}) = \text{MAX}(D_{Pi} - A_{Pi}, 0) + A_{Pi}. \tag{57}$$

Substituting (57) into (55) yields

$$S_{Pi}^{\text{Path}} = \frac{\partial E\left[\text{MAX}(D_{Pi} - A_{Pi}, 0)\right]}{\partial E(D_{Pi})} + \frac{\partial E(A_{Pi})}{\partial E(D_{Pi})}. \tag{58}$$

The second term in (58) is independent of $E(D_{Pi})$, and therefore, its derivative to $E(D_{Pi})$ equals zero

$$\begin{aligned} &\frac{\partial E(A_{Pi})}{\partial E(D_{Pi})} \\ &\quad = \lim_{\delta \to 0} \frac{E\left[\text{MAX}(D_{Pj}; j \neq i)\right] - E\left[\text{MAX}(D_{Pj}; j \neq i)\right]}{E(D_{Pi}+\delta) - E(D_{Pi})} \\ &\quad = \lim_{\delta \to 0} \frac{0}{\delta} = 0. \end{aligned} \tag{59}$$

Substituting (59) into (58) yields

$$S_{Pi}^{\text{Path}} = \frac{\partial E\left[\text{MAX}(D_{Pi} - A_{Pi}, 0)\right]}{\partial E(D_{Pi})}. \tag{60}$$

Given a small perturbation $\delta \to 0$ on the mean value of D_{Pi}, (60) yields (61), shown at the bottom of the page. Assume that $\text{pdf}(D_{Pi}, A_{Pi})$ is the joint PDF for D_{Pi} and A_{Pi}, yielding

$$\begin{aligned} S_{Pi}^{\text{Path}} = \iint \lim_{\delta \to 0} \frac{1}{\delta} \cdot &[\text{MAX}(D_{Pi} - A_{Pi} + \delta, 0) \\ &- \text{MAX}(D_{Pi} - A_{Pi}, 0)] \\ &\times \text{pdf}(D_{Pi}, A_{Pi}) \cdot dD_{Pi} \cdot dA_{Pi} \end{aligned} \tag{62}$$

where

$$\begin{aligned} &\lim_{\delta \to 0} \frac{1}{\delta} [\text{MAX}(D_{Pi} - A_{Pi} + \delta, 0) - \text{MAX}(D_{Pi} - A_{Pi}, 0)] \\ &\quad = \begin{cases} 1, & (D_{Pi} > A_{Pi}) \\ 1, & (D_{Pi} = A_{Pi} \quad \& \quad \delta > 0) \\ 0, & (D_{Pi} = A_{Pi} \quad \& \quad \delta < 0) \\ 0, & (D_{Pi} < A_{Pi}). \end{cases} \end{aligned} \tag{63}$$

Therefore, given the assumption that the probability $P(D_{Pi} = A_{Pi})$ is zero, the following integration is equal to zero:

$$\begin{aligned} &\left| \iint_{D_{Pi}=A_{Pi}} \lim_{\delta \to 0} \frac{1}{\delta} [\text{MAX}(D_{Pi} - A_{Pi} + \delta, 0) \right. \\ &\qquad \left. - \text{MAX}(D_{Pi} - A_{Pi}, 0)] \right. \\ &\qquad \left. \times \text{pdf}(D_{Pi}, A_{Pi}) \cdot dD_{Pi} \cdot dA_{Pi} \right| \\ &\quad \leq \iint_{D_{Pi}=A_{Pi}} \text{pdf}(D_{Pi}, A_{Pi}) \cdot dD_{Pi} \cdot dA_{Pi} \\ &\quad = P(D_{Pi} = A_{Pi}) = 0. \end{aligned} \tag{64}$$

Substituting (63) and (64) into (61) yields

$$\begin{aligned} S_{Pi}^{\text{Path}} &= \iint_{D_{Pi}>A_{Pi}} \text{pdf}(D_{Pi}, A_{Pi}) \cdot dD_{Pi} \cdot dA_{Pi} \\ &= P(D_{Pi} > A_{Pi}) = P(D_{Pi} \geq A_{Pi}). \end{aligned} \tag{65}$$

In (65), $P(D_{Pi} \geq A_{Pi}) = P(D_{Pi} > A_{Pi})$ because $P(D_{Pi} = A_{Pi}) = 0$. Substituting (65) into (56) proves the result in (18).

$$\begin{aligned} S_{Pi}^{\text{Path}} &= \lim_{\delta \to 0} \frac{E\left[\text{MAX}(D_{Pi} - A_{Pi} + \delta, 0)\right] - E\left[\text{MAX}(D_{Pi} - A_{Pi}, 0)\right]}{E\left(D_{Pi} + \delta\right) - E(D_{Pi})} \\ &= \lim_{\delta \to 0} \frac{E\left[\text{MAX}(D_{Pi} - A_{Pi} + \delta, 0)\right] - E\left[\text{MAX}(D_{Pi} - A_{Pi}, 0)\right]}{\delta} \end{aligned} \tag{61}$$

$$S_{Ai}^{\text{Arc}} = \frac{\partial \left\{ \int [\text{MAX}(D_{P1}, D_{P2}, \ldots) \cdot \text{pdf}(D_{P1}, D_{P2}, \ldots)] \cdot dD_{P1} \cdot dD_{P2} \cdots \right\}}{\partial E(D_{Ai})}$$
$$= \int \left[\frac{\partial \text{MAX}(D_{P1}, D_{P2}, \ldots)}{\partial E(D_{Ai})} \cdot \text{pdf}(D_{P1}, D_{P2}, \ldots) \right] \cdot dD_{P1} \cdot dD_{P2} \cdots \quad (67)$$

$$S_{Ai}^{\text{Arc}} = \int \left[\sum_{Ai \in Pk} \frac{\partial \text{MAX}(D_{P1}, D_{P2}, \ldots)}{\partial E(D_{Pk})} \cdot \text{pdf}(D_{P1}, D_{P2}, \ldots) \cdot dD_{P1} \cdot dD_{P2} \cdots \right]$$
$$= \sum_{Ai \in Pk} \frac{\partial \int [\text{MAX}(D_{P1}, D_{P2}, \ldots) \cdot \text{pdf}(D_{P1}, D_{P2}, \ldots)] \cdot dD_{P1} \cdot dD_{P2} \cdots}{\partial E(D_{Pk})}$$
$$= \sum_{Ai \in Pk} \frac{\partial E\left[\text{MAX}(D_{P1}, D_{P2}, \ldots)\right]}{\partial E(D_{Pk})} \quad (69)$$

Proof of Theorem 3

Based on probability theorem [30], we have

$$P\left[D_{Pi} = \text{MAX}(D_{Pj}; j \neq i)\right]$$
$$= \sum_{j \neq i} P\left[D_{Pi} = D_{Pj} \;\&\; D_{Pj} \geq \text{MAX}(D_{Pk}; k \neq i, k \neq j)\right]$$
$$\leq \sum_{j \neq i} P(D_{Pi} = D_{Pj})$$
$$= 0. \quad (66)$$

Equation (66) proves Theorem 3.

Proof of Theorem 4

Assume that $\text{pdf}(D_{P1}, D_{P2}, \ldots)$ is the joint PDF of all path delays, yielding (67), shown at the top of the page. Theoretically, the $\text{MAX}(\bullet)$ function is not differentiable at the locations where $D_{Pi} = \text{MAX}(D_{Pj}; j \neq i)$. However, as shown in (64), the integration in (67) is equal to zero at these singular points, given the assumption that $P[D_{Pi} = \text{MAX}(D_{Pj}; j \neq i)] = 0$. Therefore, these singular points have no impact on the final value of S_{Ai}^{Arc} and can be completely ignored

$$S_{Ai}^{\text{Arc}} = \int \left[\sum_{k} \frac{\partial \text{MAX}(D_{P1}, D_{P2}, \ldots)}{\partial E(D_{Pk})} \cdot \frac{\partial E(D_{Pk})}{\partial E(D_{Ai})} \right.$$
$$\left. \times \text{pdf}(D_{P1}, D_{P2}, \ldots) \cdot dD_{P1} \cdot dD_{P2} \cdots \right]. \quad (68)$$

In (68), the derivative $\partial D_{Pk} / \partial E(D_{Ai})$ is nonzero (equal to 1) if and only if the ith arc A_i sits on the kth path P_k. Therefore, we have (69), shown at the top of the page. Substituting (13) and (16) into (69) yields the result in (21).

Proof of Theorem 5

Theorems 4 and 5 are similar. Because we already gave the detailed proof for Theorem 4, we only show the major steps to prove Theorem 5 in this section.

Assume that $\text{pdf}(D_{P1}, D_{P2}, \ldots)$ is the joint PDF of all path delays, yielding

$$S_{Vi}^{\text{Node}} = \int \left[\frac{\partial \text{MAX}(D_{P1}, D_{P2}, \ldots)}{\partial E(\text{AT}_{Vi})} \cdot \text{pdf}(D_{P1}, D_{P2}, \ldots) \right]$$
$$\times dD_{P1} \cdot dD_{P2} \cdots$$
$$= \int \left[\sum_{k} \frac{\partial \text{MAX}(D_{P1}, D_{P2}, \ldots)}{\partial E(D_{Pk})} \cdot \frac{\partial E(D_{Pk})}{\partial E(\text{AT}_{Vi})} \right.$$
$$\left. \times \text{pdf}(D_{P1}, D_{P2}, \ldots) \cdot dD_{P1} \cdot dD_{P2} \cdots \right]$$
$$= \sum_{Vi \in Pk} \frac{\partial E\left[\text{MAX}(D_{P1}, D_{P2}, \ldots)\right]}{\partial E(D_{Pk})}. \quad (70)$$

Substituting (13) and (16) into (70) yields the result in (24).

REFERENCES

[1] X. Li, J. Le, M. Celik, and L. Pileggi, "Defining statistical sensitivity for timing optimization of logic circuits with large-scale process and environmental variations," in *Proc. IEEE Int. Conf. Comput.-Aided Des.*, 2005, pp. 844–851.

[2] S. Nassif, "Modeling and analysis of manufacturing variations," in *Proc. IEEE Custom Integr. Circuit Conf.*, 2001, pp. 223–228.

[3] M. Orshansky and K. Keutzer, "A general probabilistic framework for worst case timing analysis," in *Proc. IEEE Des. Autom. Conf.*, 2002, pp. 556–561.

[4] J. Liou, A. Krstic, L. Wang, and K. Cheng, "False-path-aware statistical timing analysis and efficient path selection for delay testing and timing validation," in *Proc. IEEE Des. Autom. Conf.*, 2002, pp. 566–569.

[5] J. Jess, K. Kalafala, S. Naidu, R. Otten, and C. Visweswariah, "Statistical timing for parametric yield prediction of digital integrated circuits," in *Proc. IEEE Des. Autom. Conf.*, 2003, pp. 932–937.

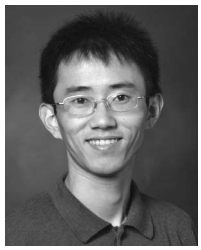
[6] M. Orshansky and A. Bandyopadhyay, "Fast statistical timing analysis handling arbitrary delay correlations," in *Proc. IEEE Des. Autom. Conf.*, 2004, pp. 337–342.

[7] F. Najm and N. Menezes, "Statistical timing analysis based on a timing yield model," in *Proc. IEEE Des. Autom. Conf.*, 2004, pp. 460–465.
[8] C. Amin, N. Menezes, K. Killpack, F. Dartu, U. Choudhury, N. Hakim, and Y. Ismail, "Statistical static timing analysis: How simple can we get?" in *Proc. IEEE Des. Autom. Conf.*, 2005, pp. 652–657.
[9] K. Heloue and F. Najm, "Statistical timing analysis with two-sided constraints," in *Proc. IEEE Int. Conf. Comput.-Aided Des.*, 2005, pp. 828–835.
[10] J. Liou, K. Chen, S. Kundu, and A. Krstic, "Fast statistical timing analysis by probabilistic event propagation," in *Proc. IEEE Des. Autom. Conf.*, 2001, pp. 661–666.
[11] A. Agarwal, V. Zolotov, and D. Blaauw, "Statistical timing analysis using bounds and selective enumeration," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 22, no. 9, pp. 1243–1260, Sep. 2003.
[12] A. Devgan and C. Kashyap, "Block-based static timing analysis with uncertainty," in *Proc. IEEE Int. Conf. Comput.-Aided Des.*, 2003, pp. 607–614.
[13] S. Tsukiyama, M. Tanaka, and M. Fukui, "A statistical static timing analysis considering correlations between delays," in *Proc. IEEE Asia South Pacific Des. Autom. Conf.*, 2001, pp. 353–358.
[14] H. Chang and S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single PERT-like traversal," in *Proc. IEEE Int. Conf. Comput.-Aided Des.*, 2003, pp. 621–625.
[15] A. Agarwal, D. Blaauw, and V. Zolotov, "Statistical timing analysis for intra-die process variations with spatial correlations," in *Proc. IEEE Int. Conf. Comput.-Aided Des.*, 2003, pp. 900–907.
[16] C. Visweswariah, K. Ravindran, K. Kalafala, S. Walker, and S. Narayan, "First-order incremental block-based statistical timing analysis," in *Proc. IEEE Des. Autom. Conf.*, 2004, pp. 331–336.
[17] J. Le, X. Li, and L. Pileggi, "STAC: Statistical timing analysis with correlation," in *Proc. IEEE Des. Autom. Conf.*, 2004, pp. 343–348.
[18] H. Chang, V. Zolotov, S. Narayan, and C. Visweswariah, "Parameterized block-based statistical timing analysis with non-Gaussian parameters, nonlinear delay functions," in *Proc. IEEE Des. Autom. Conf.*, 2005, pp. 71–76.
[19] L. Zhang, W. Chen, Y. Hu, J. Gubner, and C. Chen, "Correlation-preserved non-Gaussian statistical timing analysis with quadratic timing model," in *Proc. IEEE Des. Autom. Conf.*, 2005, pp. 83–88.
[20] Y. Zhan, A. Strojwas, X. Li, L. Pileggi, D. Newmark, and M. Sharma, "Correlation aware statistical timing analysis with non-Gaussian delay distributions," in *Proc. IEEE Des. Autom. Conf.*, 2005, pp. 77–82.
[21] J. Xiong, V. Zolotov, N. Venkateswaran, and C. Visweswariah, "Criticality computation in parameterized statistical timing," in *Proc. IEEE Des. Autom. Conf.*, 2005, pp. 63–68.
[22] V. Zolotov, J. Xiong, and C. Visweswariah, "Computation of yield gradients from statistical timing analysis," in *Proc. IEEE Int. Workshop Timing Issues Specification Synthesis Digital Syst.*, 2006, pp. 119–124.
[23] X. Bai, C. Visweswariah, P. Strenski, and D. Hathaway, "Uncertainty-aware circuit optimization," in *Proc. IEEE Des. Autom. Conf.*, 2002, pp. 58–63.
[24] X. Li and L. Pileggi, "Efficient parametric yield extraction for multiple correlated non-Normal performance distributions of analog/RF circuits," in *Proc. IEEE Des. Autom. Conf.*, 2007, pp. 928–933.
[25] D. Sinha, H. Zhou, and N. Shenoy, "Advances in computation of the maximum of a set of Gaussian random variables," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 26, no. 8, pp. 1522–1533, Aug. 2007.
[26] C. Clark, "The greatest of a finite set of random variables," *Oper. Res.*, vol. 9, no. 2, pp. 145–162, Mar./Apr. 1961.
[27] J. Qian, S. Pullela, and L. Pillage, "Modeling the effective capacitance for the RC interconnect of CMOS gates," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 13, no. 12, pp. 1526–1535, Dec. 1994.
[28] N. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*. Edinburgh, U.K.: Oliver & Boyd, 1965.
[29] G. Seber, *Multivariate Observations*. Hoboken, NJ: Wiley, 1984.
[30] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 2001.
[31] S. Sapatnekar, *Timing*. New York: Springer-Verlag, 2004.

Xin Li (S'01–M'06) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 2005.

He is currently a Research Scientist with the Department of Electrical and Computer Engineering, Carnegie Mellon University. His research interests include modeling, simulation, and synthesis for analog/RF and digital systems.

Dr. Li is the recipient of the IEEE/ACM William J. McCalla International Conference on Computer-Aided Design Best Paper Award in 2004.



Jiayong Le (S'03–M'06) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 2006.

He is currently a Member of the Technical Staff at Extreme DA, Inc., Santa Clara, CA. His current research interests include statistical timing analysis and optimization of digital systems.

Dr. Le is the recipient of the IEEE/ACM William J. McCalla International Conference on Computer-Aided Design Best Paper Award in 2004.

Mustafa Celik (S'89–M'90) received the B.S. degree from Middle East Technical University, Ankara, Turkey, in 1988, and the M.S. and Ph.D. degrees from Bilkent University, Ankara, Turkey, in 1991 and 1994, respectively, all in electrical engineering.

He is currently with Extreme DA, Inc., Santa Clara, CA. His research interests include interconnect analysis and circuit simulation.

Dr. Celik is the recipient of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN Best Paper Award in 1999.

Lawrence T. Pileggi (S'85–M'89–SM'94–F'01) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 1989.

He is the Tanoto Professor of the Department of Electrical and Computer Engineering and the Director of the Center for Silicon System Implementation, Carnegie Mellon University. From 1984 to 1986, he was with Westinghouse Research and Development, where, in 1986, he was recognized with the corporation's highest engineering achievement award. He is the Coauthor of *Electronic Circuit and System Simulation Methods* (McGraw-Hill, 1995) and *IC Interconnect Analysis* (Kluwer, 2002). He has published over 200 refereed conference and journal papers and is the holder of 14 U.S. patents. His research interests include various aspects of digital and analog design and electronic design automation.

Prof. Pileggi served as the Technical Program Chairman of the 2001 International Conference on Computer-Aided Design (ICCAD) and as the Conference Chairman of the 2002 ICCAD. He is the recipient of the TCAD Best Paper Awards in 1991 and 1999 and the best paper awards from the Design Automation Conference in 2003 and ICCAD in 2004. He is also the recipient of the Presidential Young Investigator Award from the National Science Foundation in 1991. In 1991 and again in 1999, he received the SRC Technical Excellence Award, and in 1994, he received the University of Texas Parent's Association Centennial Teaching Fellowship for excellence in undergraduate instruction. In 1995 and 2005, he received the Faculty Partnership Awards from IBM.