Parameterized Block-Based Statistical Timing Analysis with Non-Gaussian Parameters, Nonlinear Delay Functions

Hongliang Chang
Dept. of CSE
University of Minnesota
hchang@cs.umn.edu

Vladimir Zolotov
IBM T.J.W. Research Center
Yorktown Heights, NY
zolotov@us.ibm.com

Sambasivan Narayan
IBM Systems & Technology
Essex Junction, VT
narayans@us.ibm.com

Chandu Visweswariah
IBM T.J.W. Research Center
Yorktown Heights, NY
chandu@us.ibm.com

ABSTRACT

Variability of process parameters makes prediction of digital circuit timing characteristics an important and challenging problem in modern chip design. Recently, statistical static timing analysis (statistical STA) has been proposed as a solution. Unfortunately, the existing approaches either do not consider explicit gate delay dependence on process parameters [3] - [6] or restrict analysis to linear Gaussian parameters only [1, [2]. Here we extend the capabilities of parameterized block-based statistical STA [1] to handle nonlinear function of delays and non-Gaussian parameters, while retaining maximum efficiency of processing linear Gaussian parameters. Our novel technique improves accuracy in predicting circuit timing characteristics and retains such benefits of parameterized block-based statistical STA as an incremental mode of operation, computation of criticality probabilities and sensitivities to process parameter variations. We implemented our technique in an industrial statistical timing analysis tool. Our experiments with large digital blocks showed both efficiency and accuracy of the proposed technique.

Categories and Subject Descriptors

B.8.2 **[Performance and Reliability]**: Performance analysis and design aids

General Terms

Algorithms, performance, design.

1. INTRODUCTION

Electrical characteristics of transistors and interconnects are not the same for different chips and even for the same chip at different locations and time moments [1-7]. Variation of electrical characteristics is due to variation of process parameters, changing of environmental conditions and even chips aging (lithography, CMP, temperature, HCI, NBTI, etc). Variation of electrical characteristics results in variation of timing characteristics of gates and wires. The traditional way to treat variations is to consider process corners at which the gates and wires have the worst combinations of delays. Then chips are designed so that they can properly function at all these process corners assuming that as a result they will function at any other combination of gate and wire delays.

Unfortunately, as the number of sources of variations increases, the corner-based method becomes computational expensive. Moreover, with decreasing size of transistors and interconnect width, the variation of electrical characteristics is getting proportionally higher. The process corner approach, which used to work well, now results in overly conservative and suboptimal designs, wherein much effort is expended in closing timing at extremely low-probability process corners. An alternative approach is to consider the actual statistics of process parameters and use them to compute statistical characteristics of the designed circuit to achieve less pessimistic results in efficient run-times. This approach is known as statistical static timing analysis (statistical STA). There are many varieties of this technique [1-7]. One of the most useful approaches for circuit analysis and optimization is parameterized statistical STA [1, 2]. This technique considers gates or wires delays D as functions of process parameters X_i's:

$$D = D(X_1, X_2, \cdots) \quad \textbf{EQ 1}$$

Using this representation, parameterized statistical STA computes circuit timing characteristics A (arrival and required arrival times, delay, timing slack) as functions of the same process parameters:

$$A = A(X_1, X_2, \cdots) \quad \textbf{EQ 2}$$

Knowing explicit dependencies of circuit timing characteristics on process parameters has two advantages. First, combining this information with the statistical characteristics of process parameters, we can compute probability distributions of circuit delay and predict manufacturing yield. Second, this information is useful for circuit optimization, improving robustness of the design, and manufacturing line tailoring. Non-parameterized statistical STA [3-7] is less attractive because it does not compute relations between circuit characteristics and process parameters.

Statistical STA can be either path-based or block-based. A path-based statistical STA requires enumeration of signal propagation paths and integration over parameter space. A more efficient technique is parameterized block-based statistical STA. It computes signal arrival times as functions of process parameters similarly to propagating arrival times by a deterministic STA. Parameterized statistical STA in [1, 2] assumes that all the parameters have independent normal Gaussian probability distributions and affect gate delays linearly. The independence can be achieved by principal component analysis. According to this assumption, gate delays are represented in first-order canonical form:

$$A = a_0 + \sum_{i=1}^{n} a_i \cdot \Delta X_i + a_{n+1} \Delta R_a \quad \textbf{EQ 3}$$

where:

- a_0 is the mean or nominal delay;
- $\Delta X_i = X_i - \hat{X}_i$ is the variation of parameter X_i, centralized by subtracting its mean value $\hat{X}_i$;
- a_i is the sensitivity of gate delay to process parameter X_i;
- ΔR_a is the variation of an uncorrelated parameter R_a;
- a_{n+1} is the delay sensitivity to uncorrelated variations.

DAC 2005, June 13-17, 2005, Anaheim, California, USA.

The first two terms of the canonical form represent the part of delay predicted by linear approximation. The last term $a_{n+1}\Delta R_a$ represents uncorrelated variation. Figure 1 shows graphically a canonical form for one process parameter. The solid line is a linear approximation of the delay. The dashed lines show the 3σ region of the uncorrelated variation. In the case of multiple process parameters, the canonical form is represented by a hyper-plane defining delay as a linear function of process parameters and two parallel hyper-planes bounding the 3σ region of uncertainty of uncorrelated variation. Using the canonical form of gate delays, parameterized statistical STA approximates signal arrival times in the same first-order canonical form, i.e., theoretically predicts the results of a linear regression.

The assumption about the linear Gaussian nature of process parameters is very convenient for statistical STA. It allows the use of analytical formulas for computing canonical forms of arrival times which make statistical timing analysis efficient. Unfortunately, some process parameters have significantly non-Gaussian probability distributions. For example, via resistance is known to have an asymmetric probability distribution. The linear approximation is justified by the assumption that variations are small. However, with reduction of critical feature size, process variation is getting larger and linear approximation is not accurate enough. For instance, delay dependence on transistor channel length (L_{eff}) is essentially nonlinear, and assuming linear dependency can result in substantially inaccurate results [9].

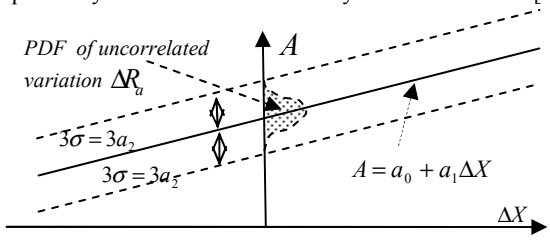

Figure 1. Graphical representation of a canonical form

$$A = a_0 + a_1\Delta X + a_2\Delta R_a$$

Here we focus on block-based parameterized statistical STA. We present a novel technique for handling process parameters that have non-Gaussian distributions and/or affect gate delay nonlinearly (For conciseness, in the remaining of the paper, we will use the term "nonlinear parameter" to refer to parameters with nonlinear effect on delays). The obvious way to handle such inconveniences is to apply a numerical technique. However, it results in losing the desired computational efficiency. We present a combined approach, which processes linear Gaussian parameters analytically and uses a numerical technique only for nonlinear and non-Gaussian parameters. It is efficient for the cases where linear Gaussian approximation is accurate enough for most of the parameters and only a few of them demonstrate complex nonlinear and/or non-Gaussian behaviors. Our experiments show that the latter is the most common practical case. We generalize the first-order canonical form to include non-Gaussian distributions and nonlinear functions. The proposed technique is an extension of the original parameterized block-based statistical STA [1] [2] and fully compatible with it. We implemented our technique to extend the statistical STA tool [1] to be able to process mixture of a linear and nonlinear, Gaussian and non-Gaussian parameters. Our experiments on industrial circuits show that the proposed technique does not reduce efficiency of the original statistical STA tool for handling linear Gaussian parameters and can handle additionally up to 7-8 nonlinear and/or non-Gaussian parameters.

The rest of the paper is organized as follows. Section 2 gives an overview of parameterized block-based statistical STA. In Section 3, we present our technique of handling non-Gaussian and nonlinear parameters. We limit our discussion to computing the latest arrival time. However, using obvious symmetry between minimum and maximum functions, it is easy to extend the proposed technique to computing the earliest arrival time as well as required arrival time and timing slack. Section 4 describes our implementation of the proposed technique in the industrial statistical timing analysis tool EinsStat [1] and presents experimental results to show the accuracy and performance of the proposed technique. Finally, conclusions are drawn in Section 5.

2. PARAMETERIZED STATISTICAL STA

A parameterized block-based statistical STA models a digital circuit with timing graph $G=(N,E,n_s,n_f)$, where: N is a set of nodes, E is a set of edges, n_s is a source node, n_f is a sink node. Each edge is weighted by its delay. The circuit timing characteristics are computed by propagating signals from the source node to the sink. Parameterized block-based statistical STA represents the delays and signal arrival times in first-order canonical form. It uses two basic operations: propagation of arrival time through a timing edge and computation of the latest arrival time at a node. The first operation is basically to compute $C=A+B$, where A and B are in canonical forms. The resulting sum C is also in canonical form and the parameters of C are computed as follows:

$$\begin{aligned} c_0 &= a_0 + b_0 \\ c_i &= a_i + b_i \text{ for } 1 \le i \le n \\ c_{n+1} &= \sqrt{a_{n+1}^2 + b_{n+1}^2} \end{aligned} \quad \textbf{EQ 4}$$

The second operation requires computation of $C=max(A,B)$, where A and B are in canonical form. The maximum is a nonlinear function. Therefore, the maximum of two canonical forms cannot be expressed exactly by a canonical form. The work in [1] proposes the following algorithm for computing statistical approximation C_{appr} of the maximum of two arrival times A and B:

Algorithm 1:

1. Compute variances and covariance of A and B:

$$\sigma_A^2 = \sum_{i=1}^{n+1} a_i^2, \sigma_B^2 = \sum_{i=1}^{n+1} b_i^2, r = \sum_{i=1}^{n} a_i b_i \quad \textbf{EQ 5}$$

2. Compute tightness probability $T_A=P(A>B)$, i.e., the probability that arrival time A is larger than B:

$$T_A = \Phi\left(\frac{a_0 - b_0}{\theta}\right), \Phi(y) = \int_{-\infty}^{y} \phi(x)dx \quad \textbf{EQ 6}$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right), \theta = \sqrt{\sigma_A^2 + \sigma_B^2 - 2r}$$

3. Compute mean c_0, second moment m_{2C} and variance σ_c^2 of $C=max(A,B)$ using the results from Clark's work [8]:

$$\begin{aligned} c_0 &= a_0 T_A + b_0(1 - T_A) + \theta\phi\left(\frac{a_0 - b_0}{\theta}\right) \\ m_{2C} &= (\sigma_A^2 + a_0^2)T_A + (\sigma_B^2 + b_0^2)(1 - T_A) + \\ &\quad (a_0 + b_0)\theta\phi\left(\frac{a_0 - b_0}{\theta}\right) \\ \sigma_C^2 &= m_{2C} - c_0^2 \end{aligned} \quad \textbf{EQ 7}$$

using T_A, θ and $\phi(x)$, computed in step 2.

4. Compute sensitivity coefficients c_i of C_{appr} for $1 \le i \le n$

$$c_i = T_A a_i + (1 - T_A) b_i \quad \textbf{EQ 8}$$

5. Compute sensitivity coefficient c_{n+1} of canonical form C_{appr} to make the variance of C_{appr} equal to the variance of $C=max(A,B)$.

Algorithm 1 uses only analytic expressions and is therefore very efficient. Also, it approximates the maximum of two canonical forms by minimizing the expected value of squared error (proof is not included here). The approximation here has statistical sense, meaning that minimizing the high probability error is more important than the lower probability one. This approximation matches the mean and variance of the exact maximum. The use of tightness probabilities T_A and $T_B=(1-T_A)$ as weighting coefficients is justified by the reasoning that the larger the tightness probability T_A, the more likely that $max(A,B)$ equals A.

Figure 2 shows the approximation of the maximum of two canonical forms A and B depending on one parameter ΔX. Canonical forms A and B are shown with thick dashed lines. The exact maximum is a piecewise linear function $C=max(A,B)$ consisting of two pieces shown with bold solid lines. The approximate canonical form C_{appr} is a line with a slope more than the slope of line A and less than the slope of line B. The line C_{appr} is closer to line A than to line B, because the probability of $A>B$ is larger than the probability of $A<B$. This is obvious from comparing the interval where $A>B$ to the interval of 6σ parameter variation. In the case of multiple process parameters, we have a similar picture but with hyper-planes instead of lines. In this work, we will extend parameterized statistical STA to non-Gaussian and nonlinear parameters using the same ideas of statistical approximation and tightness probabilities that were described in this section for the linear Gaussian case.

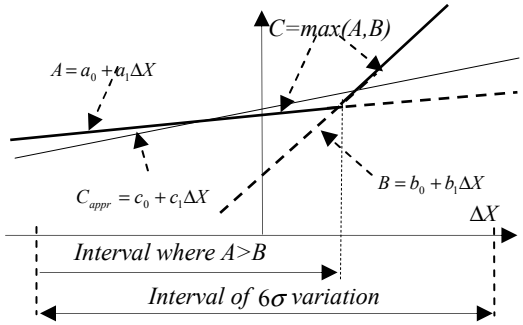

Figure 2. Linear approximation of maximum of two canonical forms *A* and *B*

3. HANDLING NON-GAUSSIAN AND NONLINEAR PARAMETERS

It is known that by applying variable transformation, we can convert non-Gaussian distributions into Gaussian ones and nonlinear dependence into linear ones. However, in the general case, if we have nonlinear function $y=f(x)$ depending on the random variable x with non-Gaussian probability distribution, there is no transformation $z=g(x)$ such that z has a Gaussian distribution and y is a linear function of z. We can convert a non-Gaussian distribution into a Gaussian one, but as a result the function of interest will be nonlinear. On the other hand, we can transform a nonlinear function into a linear one, but as a result the random variable will be non-Gaussian. Therefore, it may appear attractive to extend statistical STA to handling either only non-Gaussian or nonlinear parameters. Then we can apply variable transformation to reduce the general case of having both non-Gaussian and nonlinear parameters into the special one with only either non-Gaussian or nonlinear parameters. However, for statistical STA it is not a convenient solution. A parameter has the same probability distribution for all the delays, but different delays may depend on the same parameter differently, which means different nonlinear functions. On the other hand, we show below that using our approach it is not difficult to handle both nonlinear and non-Gaussian parameters simultaneously and even nonlinear parameters with non-Gaussian probability distributions.

In order to extend parameterized statistical STA to non-Gaussian and nonlinear parameters, we generalize the first-order canonical form (**EQ 3**) to non-Gaussian and nonlinear parameters. Then we construct a statistical approximation for the maximum of two generalized canonical forms by applying the same ideas as in the linear Gaussian case: we build a linear approximation using tightness probabilities as weighting factors, and compute the exact mean and variance values of the maximum of two generalized canonical forms.

3.1. Canonical Form with Nonlinear and Non-Gaussian Parameters

We generalize the first-order canonical form **EQ 3** as follows:

$$A = a_0 + \sum_{i=i}^{n_{LG}} a_{LG,i} \cdot \Delta X_{LG,i} + f_A(\Delta X_N) + a_{n+1} \cdot \Delta R_a \quad \textbf{EQ 9}$$

where:

- a_0 is the mean value of A;
- $\Delta X_{LG,i}$ are linear Gaussian parameters;
- $a_{LG,i}$ is the sensitivity to parameter $\Delta X_{LG,i}$;
- $\Delta X_N = (\Delta X_{N,1}, \Delta X_{N,2}, \ldots)$ is a vector of nonlinear and/or non-Gaussian parameters;
- n_{LG} is the number of linear Gaussian parameters;
- f_A is a function describing the dependence on nonlinear and non-Gaussian parameters. It can be a cross-term of parameters. It should have *0* mean value;
- ΔR_a is a normalized Gaussian parameter responsible for uncorrelated variation.
- a_{n+1} is the sensitivity to the uncorrelated random variable.

The generalization of the first-order canonical form differs from the original one only by term $f_A(\Delta X_N)$ that describes dependencies of A on nonlinear and non-Gaussian parameters. We do not impose any specific restrictions on the type of function f_A. For numerical computations, this function can be specified by a table. We even do not require mutual independence of nonlinear and non-Gaussian parameters. We only require that they are independent with respect to all the linear Gaussian parameters. This is a rather valid assumption: correlated random variables tend to have similar distributions, and if a linear parameter is correlated with a nonlinear one, independence can be achieved by orthogonal transformation techniques such as principal component analysis. Of course, specific restrictions on nonlinear function f_A or on the probability density function of nonlinear parameters can simplify numerical computations. For example, if

$$f_A(\Delta X_N) = \sum_{i=1}^{n_N} f_{A,i}(\Delta X_{N,i}) \quad \textbf{EQ 10}$$

then instead of a multidimensional table for function f_A we can use several one-dimensional tables to define functions $f_{A,i}$, which is

more efficient. Similar simplification is possible if the nonlinear parameters are independent.

Propagation of arrival time in generalized canonical form through a timing edge with delay in the same form is similar to the pure linear Gaussian case. The only difference is the summation of nonlinear functions of the arrival time and the delay, which can be performed numerically by summing tables describing the nonlinear functions of the arrival time and the delay.

3.2. Computing the Maximum Function

For approximation of the maximum of two canonical forms, we use the same concept of tightness probability and computational approach as for the linear Gaussian case, so that the correlation of delays or arrival times can be easily preserved. The parameters of the canonical form C_{appr} approximating the maximum of two generalized canonical forms A and B are computed by the following formulas:

$$\begin{aligned} c_0 &= E[\max(A,B)] \\ c_i &= T_A a_i + (1-T_A) b_i \quad \text{for } 1 < i < n_{LG} \\ f_C(\Delta X_N) &= T_A f_A(\Delta X_N) + (1-T_A) f_B(\Delta X_N) \end{aligned} \qquad \textbf{EQ 11}$$

where $T_A=P(A>B)$ is tightness probability, i.e., the probability that arrival time A is larger than B.

The sensitivity coefficient c_{n+1} to uncorrelated variation is computed so as to make the standard deviation of the approximation C_{appr} equal to the standard deviation of the exact maximum $C=max(A,B)$. Similar to the linear Gaussian case, this approximation of the maximum of two generalized canonical forms is linear: the coefficients c_i and function f_C are computed as linear combinations of coefficients a_i and b_i, and functions f_A and f_B, correspondingly. Figure 3 graphically shows the meaning of a linear approximation for generalized canonical forms that depend only on one nonlinear parameter. Canonical forms A and B are shown with thick dashed curves. The exact maximum $C=max(A,B)$ is shown with a bold solid curve. The approximation of the maximum C_{appr} is shown with a solid thin curve. The curve of the approximate maximum is closer to curve A, because, as we can see in Figure 3, $max(A,B)$ is more often equal to A than to B.

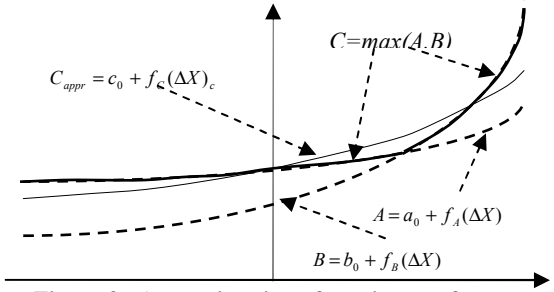

Figure 3. Approximation of maximum of two generalized canonical forms

This approximation of the maximum of two generalized canonical forms requires the computation of the tightness probability T_A, the mean and the second moment of $max(A,B)$, since the standard deviation is expressed through the mean and the second moment as follows:

$$\sigma_C^{\,2} = E[\max(A,B)^2] - (E[\max(A,B)])^2 \quad \textbf{EQ 12}$$

There are two equivalent ways of deriving the required formulas. One uses transformation of the integrals defining tightness probability, mean and the second moment. The other is based on conditional probability and conditional moments. We use here the second approach as it is more concise and intuitive.

Considering nonlinear and non-Gaussian parameter variations fixed, we rewrite the expression for the generalized canonical form by combining the term $f_A(\Delta X_N)$ and the mean value a_0:

$$A = (a_0 + f_A(\Delta X_N)) + \sum_{i=i}^{n_L} a_{LG,i} \cdot \Delta X_{LG,i} + a_{n+1} \cdot \Delta R_a \quad \textbf{EQ 13}$$

This expression can be considered as a canonical form A_{Cond} with a mean value $a_0 + f_A(\Delta X_N)$ and linear Gaussian parameters. All the linear sensitivities in this form are the same as in the original generalized canonical form (**EQ 9**). Now we consider two generalized canonical forms A and B represented in the form of **EQ 13**. The conditional tightness probability, conditional mean and second moments of $max(A,B)$ at the condition that all the nonlinear and non-Gaussian parameters have fixed values ΔX_N are functions of the nonlinear and non-Gaussian parameters ΔX_N:

$$\begin{aligned} T_{A,Cond}(\Delta X_N) &= P(A > B \mid \Delta X_N) \\ c_{0,Cond}(\Delta X_N) &= E[\max(A,B) \mid \Delta X_N] \\ m_{2,Cond}(\Delta X_N) &= E[(\max(A,B))^2 \mid \Delta X_N] \end{aligned} \qquad \textbf{EQ 14}$$

The linear Gaussian parameters are independent of the nonlinear and non-Gaussian ones. Therefore, the joint conditional probability density function of the linear Gaussian parameters at the condition of frozen values of nonlinear and non-Gaussian parameters is simply a joint probability density function of the linear Gaussian parameters.

$$p(\Delta X_{LG} | \Delta X_N) = p(\Delta X_{LG}) \quad \textbf{EQ 15}$$

Thus, we can use Algorithm 1 for computing the conditional tightness probability, mean and second moments for the maximum of two generalized canonical forms at the condition that all nonlinear and non-Gaussian parameters are frozen. In Algorithm 1, we substitute $a_0 + f_A(\Delta X_N)$ and $b_0 + f_B(\Delta X_N)$ for a_0 and b_0. Since Algorithm 1 uses only analytical formulas, the required values can all be computed efficiently. The unconditional tightness probability, mean, and second moment of $max(A,B)$ can be computed by integrating the conditional tightness probability, mean and second moment over the space of nonlinear and non-Gaussian parameters with their joint probability density function:

$$\begin{aligned} T_A &= \int_{-\infty}^{\infty} T_{A,Cond}(\Delta X_N) p(\Delta X_N) d\Delta X_N \\ c_0 &= \int_{-\infty}^{\infty} c_{0,Cond}(\Delta X_N) p(\Delta X_N) d\Delta X_N \\ m_{2,C} &= \int_{-\infty}^{\infty} m_{2,C,Cond}(\Delta X_N) p(\Delta X_N) d\Delta X_N \end{aligned} \qquad \textbf{EQ 16}$$

Having the tightness probability, mean and the second moment, we compute the standard deviation of $max(A,B)$ by **EQ 12** and then all the parameters of the approximate generalized canonical form C by formulas **EQ 11**.

The integration in EQ **16** can be implemented numerically by any technique. In the simplest case, it is performed by integrating numerically in the orthogonal discretized regions of the nonlinear and non-Gaussian parameters ΔX_N. For each grid cell, we compute the conditional tightness probability, mean and second moment by formulas **EQ 14**. Then the integrals of **EQ 16** can be computed approximately as sums over all the cells of the discretization grid. For example the numerical formula for tightness probability is as follows:

$$T_A = \sum_{k=1}^{m} T_{A,Cond,k}\, p_k(\Delta X_N) V_k \quad \textbf{EQ 17}$$

where:

- $T_{A,Cond,k}$ is the conditional tightness probability that A>B at the condition that the nonlinear and non-Gaussian parameters have values inside the *k-th* grid cell;
- $p_k(\Delta X_N)$ is the value of the joint probability density function of the nonlinear and non-Gaussian parameters in *k-th* grid cell;
- V_k is volume of the *k-th* grid cell.

For numerical integration by discretizing the integration region, the computational complexity is exponential with respect to the number of nonlinear and non-Gaussian parameters. Our experiments show that for reasonable accuracy it is enough to have as little as 5-7 discrete points of each variable. This approach is applicable for cases with up to 7-8 nonlinear and non-Gaussian variables. For higher dimensions the integrals can be computed by a Monte-Carlo integration technique.

4. IMPLEMENTATION AND RESULTS

The proposed methodology was implemented on the top of EinsStat [1], an industrial statistical timing analysis tool. In our implementation, a global source of variation can have a non-Gaussian distribution and delay dependence on a parameter can be a nonlinear function. They are both specified by tables using proper discretization. The integrals of **EQ16** are computed by numerical integration.

We tested our implementation on computing *max(A,B)* of two first-order canonical forms *A* and *B* with non-Gaussian parameters[1]:

$$A = 10 + 0.5 \cdot \Delta X_1 + \Delta X_2 + 0.5 \cdot \Delta R_a$$
$$B = 10 + \Delta X_1 + 0.5 \cdot \Delta X_2 + 0.5 \cdot \Delta R_b$$

where ΔX_1 and ΔX_2 are random variables with lognormal probability distributions, ΔR_a and ΔR_b are Gaussian random variables for the uncorrelated randomness. Figure 4(a) shows the probability density function (PDF) of *max(A,B)* computed by the proposed technique, by the original parameterized statistical STA technique of [1], where non-Gaussian distributions are approximated with Gaussian ones having the same mean and standard deviation, and by Monte-Carlo simulation. The PDF computed by the proposed technique matches the Monte-Carlo results much closer than the PDF computed by the original technique. The proposed technique and Monte-Carlo simulation both predict asymmetric PDFs with similar trends especially at the tails of PDFs. The PDF computed by the original technique has a symmetric shape and substantially under-estimates the worst-case value.

We tested our technique on *max(A,B)* with nonlinear (cubic) functions of Gaussian parameters:

$$A = 10 + (\Delta X_1)^3/18 + (\Delta X_2)^3/9 + \Delta R_a/2$$
$$B = 10 + (\Delta X_1)^3/9 + (\Delta X_2)^3/18 + \Delta R_b/2$$

Figure 4(b) compares the PDFs computed by the original technique, by the proposed technique and by Monte-Carlo simulation. The original technique uses linear approximation of nonlinear functions that passes through the same -3sigma and +3sigma points. The proposed technique predicts the same result as Monte-Carlo, while the original technique significantly over-estimates the standard deviation.

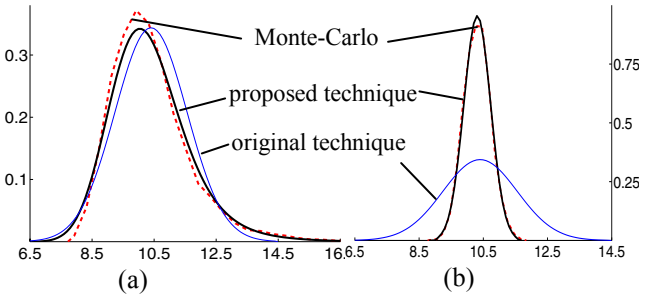

Figure 4. PDFs for maximum of two random

To choose the number of discretizition points providing a good tradeoff between accuracy and run-time, we ran tests on a small industrial design *A* (3,042 gates and 17,579 timing arcs). Table 1 shows the CPU-time of our technique for different numbers of non-Gaussian parameters, for 5 and 10 discretization points. The run time was measured on a single processor IBM Risc System 6000 model 43P-681. It is observed that processing 3 non-Gaussian parameters with 10 discretized points is about 40 times longer than handling all the parameters as Gaussian ones. However for 5 discretization points the run-time is only about 3 times longer. Meanwhile, the PDF plots for design A are provided in Figure 5 for when 5, 10 and 20 discretized points are used. We observe that as the difference between PDF curves for 10 and 20 points is almost undistinguishable, the 5 points one also gives accurate enough result. For nonlinear functions, we saw similar dependence of run-time on the number of discretization points. Therefore, for other experiments, we used only 5 discretized points.

Table 1. Run-time w.r.t. #non-Gaussian and #points

# Non-Gaussians		3	2	1	0
CPU - time (s)	10 points	69.17	7.53	2.14	1.38
	5 points	3.82	1.54	1.40	1.38

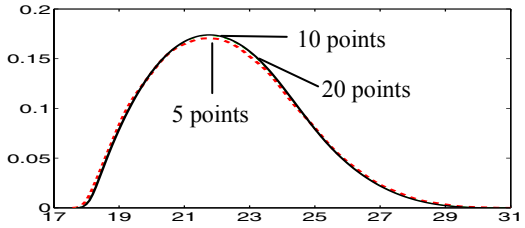

Figure 5. Comparison with different #discretized points

We performed statistical timing analysis of the same design *A* with 3 linear lognormally distributed global sources of parameters variations and a linear Gaussian uncorrelated variation. The average values of delay sensitivities to global parameters and uncorrelated variation were set to 2% and 6% of the corresponding nominal delay values respectively. Figure 6 shows the probability density functions of the latest arrival time computed by three different techniques. The proposed technique gives a close match to the Monte-Carlo result. In contrast, the PDF computed by the original technique [1] deviates substantially from the Monte-Carlo result. Firstly, the PDF computed by Monte-Carlo simulation is not Gaussian, but closer to lognormal because all three global sources of variation have lognormal distributions. However, since the technique of [1] approximates all delays with a Gaussian

[1] Note that the asserted constant sensitivities is only a simplified mode in EinsStat for testing and experiments. EinsStat also has a mode using finite differencing to compute internally all sensitivities accurately handling all interdependences on gate input slew, load capacitance, operating conditions.

distribution, it is hard to obtain a reasonably good estimate. Secondly, Monte-Carlo predicts the 0.1% and 99.9% confidence points of path delays as 19.4 ns and 32.0 ns. The proposed algorithm estimates similar values of 19.6 ns and 31.5 ns, while original technique computes 17.8 ns and 27.0 ns.

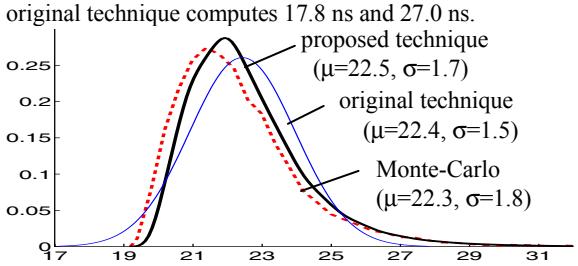

Figure 6. Arrival time PDFs of design *A* with lognormally distributed parameters

In the second set of experiments, the 3 global sources of variation had Gaussian distributions but the delays of circuit gates and wires were cubic functions of these parameters. The values of delay sensitivities to global parameters and uncorrelated variation were set to 2% and 6% of the corresponding nominal delay values respectively. Figure 7 shows PDFs and CDFs of the circuit delay computed by three different techniques. The proposed technique computes the same mean value as Monte-Carlo, while the original technique overestimates it. The original technique computes the 99.9% confidence point as 22.7 ns vs. Monte-Carlo's 22.9 ns, while the original technique over-estimates it as 23.7 ns.

Thus, we can conclude that when parameter variations have non-Gaussian distributions, or gate and wire delay depends on parameters nonlinearly, the proposed technique is essential to correctly predict circuit delay distribution and manufacturing yield.

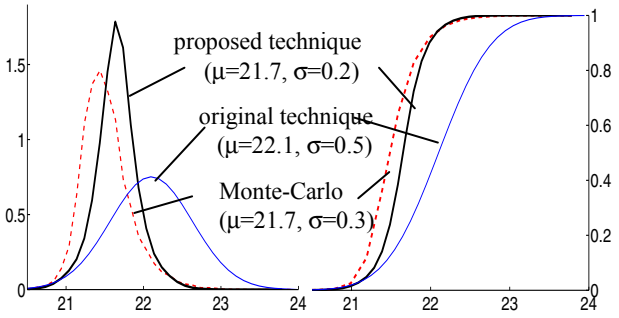

Figure 7. Arrival time PDFs and CDFs of design *A* with nonlinear parameters

Table 2. Run-time results on industrial designs

Ckt	Gates	Timing arcs	CPU-time (s) vs. #Non-Gaussians			
			3	2	1	0
A	3,042	17,579	3.8	1.5	1.4	1.4
B	11,937	57,151	12.3	5.53	4.3	3.07
C	53,317	392,097	79.1	35.8	27.3	18.7
D	70,216	363,537	93.3	41.3	30.5	19.7
E	1,085,034	5,799,545	2,083.1	982.0	788.5	703.6

Table 2 shows run time of statistical timing analysis for five industrial designs for different numbers of non-Gaussian parameters. The size of the designs varies from 3,042 up to 1,085,034 gates. We see that, as the number of non-Gaussian parameters increases to 3, the run-time is only about 3 to 5 times longer compared to the case without any non-Gaussian parameters. For the largest design *E*, the run-time is only about 35 minutes. In contrast, for the smallest design A, the run-time of Monte-Carlo simulation is about 5 hours. However, due to the large size of designs, Monte-Carlo simulations are too unrealistic to finish and thus the run-times are not provided in the table. Statistical timing analysis with nonlinear parameters has approximately the same run time.

5. CONCLUSIONS

In this paper, we presented a novel and efficient technique for handling arbitrary non-Gaussian and nonlinear parameter variations in parameterized block-based statistical timing analysis. Our approach is based on an extension of the first-order canonical form for representing delay and arrival time variations. Therefore this technique is fully compatible with the original parameterized statistical STA and preserves its computational efficiency in processing linear Gaussian parameters. The experimental results showed that the probability distributions of circuit delays computed by the new technique are closer to the results of Monte-Carlo simulations than the original parameterized statistical STA approximating non-Gaussian distributions with Gaussian ones and nonlinear functions with linear ones, especially at the 99.9 confidence level. Our experiments also showed that in many cases non-Gaussian distributions of parameter variations can be approximated with Gaussian ones with reasonable accuracy. Only significantly asymmetric distributions required handling as non-Gaussian. This conclusion is very important in practice because it justifies approximating most parameter distributions by Gaussian ones.

The limitation of the algorithm is that its run-time is exponential to the number of non-Gaussian and/or nonlinear parameters. However, in practice, as the number of nonlinear and/or non-Gaussian is not big, the algorithm is very efficient and provides a general framework for statistical STA handling non-Gaussian parameters and nonlinear functions of delays. The method can be used to validate the linear Gaussian approximation of parameters, and then selectively model crucial parameters as nonlinear or non-Gaussian ones. The method is also important for sign-off timing analysis.

6. REFERENCES

[1] C. Visweswariah, K. Ravindran, K. Kalafala, S. G. Walker and S. Narayan, "First-Order Incremental Block-Based Statistical Timing Analysis", DAC 2004, pp. 331-336.

[2] H. Chang and S. Sapatnekar, "Statistical timing analysis considering spatial correlation in a PERT-like traversal," ICCAD 2003, pp. 621-625.

[3] J. J. Liou, K. T. Cheng, S. Kundu and A. Krstic, "Fast statistical timing analysis by probabilistic event propagation," DAC 2001, pp. 661-666.

[4] M. Orshansky and K. Keutzer, "A general probabilistic framework for worst case timing analysis," DAC 2002, pp. 556-561.

[5] A. Devgan and C. Kashyap, "Block-based static timing analysis with uncertainty," ICCAD 2003, pp. 607-614.

[6] A. Agarwal, V. Zolotov and D. Blaauw, "Statistical timing analysis using bounds and selective enumeration," IEEE Transactions on CAD of Integrated Circuits and Systems, vol. 22, no. 9, Sept 2003, pp.1243-1260.

[7] S. Bhardwaj, S.B.K. Vrudhula and D. Blaauw, "TAU: Timing Analysis Under Uncertainty", ICCAD 2003, pp. 615-620

[8] C.E. Clark, "The Greatest of a Finite Set of Random Variables", Operations Research, vol. 9, 1961, pp. 85-91.

[9] X. Li, J. Le, P. Gopalakrishnan and L. Pileggi, "Asymptotic probability extraction for non-Normal distributions of circuit performance", ICCAD 2004, pp. 1-9.