Asymptotic Probability Extraction for Nonnormal Performance Distributions

Xin Li, *Member, IEEE*, Jiayong Le, *Student Member, IEEE*, Padmini Gopalakrishnan, *Student Member, IEEE*, and Lawrence T. Pileggi, *Fellow, IEEE*

Abstract—While process variations are becoming more significant with each new IC technology generation, they are often modeled via linear regression models so that the resulting performance variations can be captured via normal distributions. Nonlinear response surface models (e.g., quadratic polynomials) can be utilized to capture larger scale process variations; however, such models result in nonnormal distributions for circuit performance. These performance distributions are difficult to capture efficiently since the distribution model is unknown. In this paper, an asymptotic-probability-extraction (APEX) method for estimating the unknown random distribution when using a nonlinear response surface modeling is proposed. The APEX begins by efficiently computing the high-order moments of the unknown distribution and then applies moment matching to approximate the characteristic function of the random distribution by an efficient rational function. It is proven that such a moment-matching approach is asymptotically convergent when applied to quadratic response surface models. In addition, a number of novel algorithms and methods, including binomial moment evaluation, PDF/CDF shifting, nonlinear companding and reverse evaluation, are proposed to improve the computation efficiency and/or approximation accuracy. Several circuit examples from both digital and analog applications demonstrate that APEX can provide better accuracy than a Monte Carlo simulation with 10^4 samples and achieve up to $10\times$ more efficiency. The error, incurred by the popular normal modeling assumption for several circuit examples designed in standard IC technologies, is also shown.

Index Terms—Circuit performance, probability, process variation, response surface modeling.

I. INTRODUCTION

AS IC TECHNOLOGIES are scaled to the deep submicrometer region, process variations are becoming critical and significantly impact the overall performance of a circuit. Table I shows some typical process parameters and their 3σ variations as technologies are scaled from 250 to 70 nm. These large-scale variations introduce uncertainties in circuit behavior, thereby making IC design increasingly difficult. Low product yield or unnecessary overdesign cannot be avoided if

TABLE I
ESTIMATED TECHNOLOGY PARAMETERS AND 3σ VARIATIONS [1]

Year	L_{EFF} (nm)	T_{OX} (nm)	V_{TH} (mV)	W (μm)	H (μm)	ρ (mΩ/□)
1997	250±80	5.0±0.40	500±50	0.80±0.20	1.2±0.30	45±10
1999	180±60	4.5±0.36	450±45	0.65±0.17	1.0±0.30	50±12
2002	130±45	4.0±0.39	400±40	0.50±0.14	0.9±0.27	55±15
2005	100±40	3.5±0.42	350±40	0.40±0.12	0.8±0.27	60±19
2006	70±33	3.0±0.48	300±40	0.30±0.10	0.7±0.25	75±25

process variations are not accurately modeled and analyzed within the IC design flow.

During the past decade, various statistical analysis techniques [1]–[7] have been proposed and utilized in many applications such as statistical timing analysis, mismatch analysis, yield optimization, etc. The objective of these techniques is to model the probability distribution of the circuit performance under random process variations. Nassif [1] applies a linear regression to approximate a given circuit performance f (e.g., delay, gain, etc.) as a function of the process variations (e.g., $\Delta V_{\rm TH}$, $\Delta T_{\rm OX}$, etc.) and assumes that all random variations are normally distributed. As such, the performance f is also a normal distribution, since the linear combination of normally distributed random variables is still a normal distribution [8].

The linear regression model is efficient and accurate when process variations are sufficiently small. However, the large-scale variations in deep submicrometer technologies, which reach almost ±50% in Table I, suggest the need for higher order regression models in order to guarantee high approximation accuracy [4]–[7]. Using a higher order response surface model, however, brings about new challenges due to the nonlinear mapping between the process variations and the circuit performance f. The distribution of f is no longer normal, unlike the case of the linear model. The authors in [3]–[5] utilize the Monte Carlo simulation to evaluate the probability distribution of f, but this is computationally expensive. Note that reducing the computational cost for this probability extraction is crucial, especially when the extraction procedure is an inner loop within an optimization flow.

In this paper, we propose a novel asymptotic-probability-extraction (APEX) approach for estimating the unknown random distribution using the nonlinear response surface modeling. Given a circuit performance f (e.g., the delay of a digital circuit path or the gain of an analog amplifier), the response surface modeling approximates f as a polynomial function of the process variations (e.g., $\Delta V_{\rm TH}$, $\Delta T_{\rm OX}$, etc.). Since the process variations are modeled as random variables,
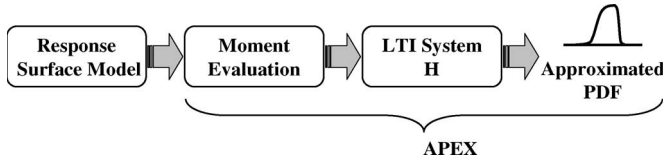
Fig. 1. Overall flow of APEX.

the circuit performance f, which is a function of these random variables, is also a random variable. The APEX applies a moment matching to approximate the characteristic function of f (i.e., the Fourier transform of the probability density function (PDF) [8]) by a rational function H. We conceptually consider H to be of the form of the transfer function of a linear time-invariant (LTI) system, and the PDF and the cumulative distribution function (CDF) of f are approximated by the impulse response and the step response of the LTI system H, respectively. Fig. 1 shows the overall flow of the APEX. In this paper, we assume that the response surface model is already available and it is provided to the APEX for estimating the probability distribution of the circuit performance.

We prove that the moment-matching approach utilized in APEX is asymptotically convergent when applied to quadratic response surface models. In other words, given a quadratic response surface model f that is a quadratic function of normally distributed random variables, the PDF and the CDF of f can be uniquely determined by its moments $\{m_k, k = 1, 2, \ldots, K\}$ when K approaches infinity, i.e., $K \to +\infty$. The PDF and the CDF extracted by the APEX can be used to characterize and/or optimize the statistical performance of analog and digital circuits under process variations.

APEX extends existing moment-matching methods via four important new contributions which significantly reduce the computational cost and improve the approximation accuracy for this particular application. First, a key operation required by the APEX is the computation of the high-order moments, which is extremely expensive when using traditional techniques such as the direct moment evaluation. In APEX, we propose a binomial evaluation scheme to recursively compute the high-order moments for a given quadratic response surface model. The binominal-moment-evaluation scheme is derived from a statistical independence theory and achieves a speedup of more than $10^6 \times$ over the direct moment-evaluation technique in our tested examples.

Second, the APEX approximates the unknown PDF by the impulse response of an LTI system. Directly applying such an approximation to any circuit performance with negative value is infeasible, since it results in an LTI system that is noncausal. To overcome this difficulty, the APEX applies a modified Chebyshev inequality for PDF/CDF shifting.

Third, in several practical applications, we observe that a direct moment matching is not efficient when the unknown PDF has a large negative skewness. In such cases, we propose to apply a nonlinear-companding scheme to automatically compress and/or expand the PDF such that the transformed PDF almost has a zero skewness and can be approximated accurately.

Finally, the best case performance (e.g., the 1% point on CDF) and the worst case performance (e.g., the 99% point

on CDF) are two important metrics to be evaluated in many practical applications. Direct moment matching cannot capture the 1% point value accurately since the moment-matching approximation is most accurate for low-frequency components (corresponding to the final value of a CDF) and least accurate for high-frequency components (corresponding to the initial value of a CDF). To address this problem, a reverse-evaluation technique is proposed in this paper to produce an accurate estimation of the 1% point.

The remainder of this paper is organized as follows. In Section II, we review the background on principal component analysis (PCA), response surface modeling, and classical moment problem. We propose our APEX approach in Section III and discuss several implementation issues, including binomial moment evaluation, PDF/CDF shifting, nonlinear companding, and reverse evaluation in Section IV. We extend the APEX algorithm to handle nonnormal process variations in Section V and discuss the applications of APEX in Section VI. The efficacy of APEX is demonstrated by several circuit examples in Section VII, followed by our conclusions in Section VIII.

II. BACKGROUND

A. PCA

The PCA [9] is a statistical method that finds a set of independent factors to represent a set of correlated random variables. Given N process parameters $X = [x_1, x_2, \ldots, x_N]^{\mathrm{T}}$, the process variations $\Delta X = X - X_0$, where X_0 contains the mean values of X, are often approximated as zero-mean normal distributions, and the correlations of ΔX can be represented by a symmetric positive semidefinite correlation matrix R. PCA decomposes R as

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

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

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

These new random variables ΔY are called the principal components or factors. It is easy to verify that ΔY are independent and satisfy the standard normal distribution $N(0, 1)$ (i.e., zero mean and unit standard deviation).

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 random variables, i.e., a small subset of principal components, to approximate the original N-dimensional space. More details on PCA can be found in [9].

B. Response Surface Modeling

Given a circuit topology, the circuit performance (e.g., gain, delay, etc.) is a function of the design parameters

(e.g., bias current, transistor sizes, etc.) and the process parameters (e.g., V_{TH}, T_{OX}, etc.). The design parameters are optimized and fixed during the design phase; however, the process parameters must be modeled as random variables to account for any random variations. After the PCA shown in (1) and (2), the process variations can be represented by N independent random variables ΔY that satisfy the standard normal distribution $N(0, 1)$. Therefore, given a set of fixed design parameters, the circuit performance f can be approximated by a linear regression model of ΔY [1]

$$f(\Delta Y) = B^{\mathrm{T}} \cdot \Delta Y + C \qquad (3)$$

where $B \in R^N$ stands for the linear coefficients and $C \in R$ is the constant term. The linear regression model in (3) is accurate when process variations are small. However, such a linear model is becoming increasingly inaccurate as process variations become relatively large in nanoscale technologies. In addition, the linear regression model in (3) always yields the conclusion that the worst case performance appears at one of the process corners. This conclusion might not be valid for many circuit performances, especially those analog circuit performances with strong nonlinearities, e.g., the offset voltage of an analog amplifier. For these reasons, applying quadratic response surface models might be required to provide a sufficient modeling accuracy [3]–[5] as

$$f(\Delta Y) = \Delta Y^{\mathrm{T}} \cdot A \cdot \Delta Y + B^{\mathrm{T}} \cdot \Delta Y + C \qquad (4)$$

where $A \in R^{N \times N}$ denotes the quadratic coefficients, $B \in R^N$ represents the linear coefficients, and $C \in R$ is the constant term. Without loss of generality, we assume that A is symmetric in this paper, since any asymmetric quadratic form $X^{\mathrm{T}}AX$ can be easily converted to an equivalent symmetric form $0.5 \cdot X^{\mathrm{T}}(A + A^{\mathrm{T}})X$ [10].

The model coefficients in (3) and (4) can be determined by solving a set of overdetermined linear equations over a number of sampling points [11]–[13]. These sampling points are typically generated from SPICE simulations or measurement results. Since generating each sampling point can be quite expensive, many response surface modeling algorithms [11]–[13] attempt to minimize the required number of sampling points while maintaining good modeling quality. For this purpose, a great number of algorithms for design of experiments (DOE) have been proposed [14], [15]. Many of these DOE algorithms (e.g., fractional factorial design, orthogonal array, etc.) generate deterministic sampling points. These sampling points do not represent the actual probability distributions of process variations and, therefore, cannot be used for estimating performance distributions. Even if a DOE algorithm (e.g., Latin hypercube sampling [15]) generates random sampling points, the number of these sampling points is typically too small (e.g., 10–100) to accurately estimate PDF/CDF functions. For this reason, after the response surface model is created, special techniques are required to extract the probability distribution.

C. Classical Moment Problem

The classical moment problem was first proposed and studied by T. Stieltjes in 1894. The Stieltjes moment problem is defined as follows [16].

Definition 1 (Stieltjes Moment Problem): Given a sequence of numbers $\{m_k, k = 1, 2, \ldots\}$, find a nondecreasing function $F(f)$, where $f \in [0, +\infty)$, such that

$$m_k = \int\limits_0^{+\infty} f^k dF(f). \qquad (5)$$

Note that if the symbol f in (5) represents a random variable, the numbers $\{m_k, k = 1, 2, \ldots\}$ are its moments and the function $F(f)$ is its CDF. In the Stieltjes moment problem, the random variable f is restricted to be nonnegative, i.e., its CDF $F(f)$ is defined only for $f \geq 0$. By varying the interval in which $F(f)$ is valid, two further variations of the classical moment problem can be defined [16].

Definition 2 (Hamburger Moment Problem): Given a sequence of numbers $\{m_k, k = 1, 2, \ldots\}$, find a nondecreasing function $F(f)$, where $f \in (-\infty, +\infty)$, such that

$$m_k = \int\limits_{-\infty}^{+\infty} f^k dF(f). \qquad (6)$$

Definition 3 (Hausdorff Moment Problem): Given a sequence of numbers $\{m_k, k = 1, 2, \ldots\}$, find a nondecreasing function $F(f)$, where $f \in [0, 1]$, such that

$$m_k = \int\limits_0^1 f^k dF(f). \qquad (7)$$

As shown in (6) and (7), the Hamburger moment problem and the Hausdorff moment problem are defined in the interval $(-\infty, +\infty)$ and [0,1], respectively. In this paper, the process variations are modeled as normal distributions, which are unbounded and distributed over $(-\infty, +\infty)$. Therefore, the circuit performance approximated by the quadratic model (4) is also unbounded. The probability extraction problem that we aim to solve is the Hamburger moment problem in (6).

The classical moment problem has been widely studied by mathematicians for over 100 years, focusing on the theoretical aspects of the problem, e.g., the existence and uniqueness of the solution. Details of these theoretical results can be found in [16] or other recent publications, e.g., in [17]. However, the practical applications of this moment problem, especially the computation efficiency of solving the problem, have not been sufficiently explored. In this paper, we develop the APEX algorithm which aims to solve the moment problem efficiently, i.e., to improve the approximation accuracy and reduce the computational cost for practical applications.

III. APEX

Given the quadratic response surface model in (4), the objective of APEX is to estimate the PDF $\mathrm{pdf}(f)$ and the CDF

cdf(f) for the performance f.[1] Instead of running expensive Monte Carlo simulations, the APEX tries to find an Mth order LTI system H whose impulse response $h(t)$ and step response $s(t)$ are the optimal approximations for pdf(f) and cdf(f), respectively.[2] The optimal approximation is determined by matching the first $2M$ moments between $h(t)$ and pdf(f) for an Mth order approximation. In this section, we first describe the mathematical formulation of the APEX algorithm. Then, we link the APEX to probability theory and explain why it is efficient in approximating PDF/CDF functions. Finally, we prove that the moment-matching method utilized in APEX is asymptotically convergent when applied to quadratic response surface models.

A. Mathematical Formulation

We define the time moments [18] for a given circuit performance f whose PDF is pdf(f) as follows:

$$s_k = \frac{(-1)^k}{k!} \int_{-\infty}^{+\infty} f^k \cdot \text{pdf}(f) \cdot df. \tag{8}$$

In (8), the definition of time moments is identical to the traditional definition of moments in probability theory except for the scaling factor $(-1)^k/k!$.

Similarly, the time moments can be defined for an LTI system H [18]. Given an Mth order LTI system whose transfer function and impulse response are

$$H(s) = \sum_{i=1}^{M} \frac{a_i}{s - b_i} \quad \text{and} \quad h(t) = \begin{cases} \sum_{i=1}^{M} a_i e^{b_i t}, & (\text{if } t \geq 0) \\ 0, & (\text{if } t < 0). \end{cases} \tag{9}$$

The time moments of H are defined as [18]

$$s_k = \frac{(-1)^k}{k!} \int_{-\infty}^{+\infty} t^k \cdot h(t) \cdot dt = -\sum_{i=1}^{M} \frac{a_i}{b_i^{k+1}}. \tag{10}$$

In (9), the poles $\{b_i, i = 1, 2, \ldots, M\}$ and residues $\{a_i, i = 1, 2, \ldots, M\}$ are the $2M$ unknowns that need to be determined. Matching the first $2M$ moments in (8) and (10) yields the following $2M$ nonlinear equations:

$$-\left(\frac{a_1}{b_1} + \frac{a_2}{b_2} + \cdots + \frac{a_M}{b_M} \right) = s_0$$

$$-\left(\frac{a_1}{b_1^2} + \frac{a_2}{b_2^2} + \cdots + \frac{a_M}{b_M^2} \right) = s_1$$

$$\vdots \qquad \vdots$$

$$-\left(\frac{a_1}{b_1^{2M}} + \frac{a_2}{b_2^{2M}} + \cdots + \frac{a_M}{b_M^{2M}} \right) = s_{2M-1}. \tag{11}$$

---

[1]In this paper, ΔX, ΔY, and ΔZ represent the random variables for modeling process variations, and f represents the circuit performance of interest. The probability density function and the cumulative distribution function are functions of f. Therefore, they are denoted as pdf(f) and cdf(f), respectively.

[2]The variable t in $h(t)$ and $s(t)$ corresponds to the variable f in pdf(f) and cdf(f).

The nonlinear equations in (11) can be solved using the algorithm proposed in [18], which first solves the poles $\{b_i\}$ and then the residues $\{a_i\}$. In what follows, we briefly describe this two-step algorithm for solving (11).

1) Solving Poles: In order to solve the poles $\{b_i\}$ in (11), Pillage and Rohrer [18] first formulate the following linear equations:

$$-\begin{bmatrix} s_0 & s_1 & \cdots & s_{M-1} \\ s_1 & s_2 & \cdots & s_M \\ \vdots & \vdots & \vdots & \vdots \\ s_{M-1} & s_M & \cdots & s_{2M-2} \end{bmatrix} \cdot \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{M-1} \end{bmatrix} = \begin{bmatrix} s_M \\ s_{M+1} \\ \vdots \\ s_{2M-1} \end{bmatrix}. \tag{12}$$

After solving (12) for $\{c_i, i = 0, 1, \ldots, M - 1\}$, the poles $\{b_i\}$ in (11) are equal to the reciprocals of the roots of the following characteristic polynomial:

$$c_0 + c_1 b^{-1} + c_2 b^{-1} + \cdots + c_{M-1} b^{-M+1} + b^{-M} = 0. \tag{13}$$

The detailed proof of (12) and (13) can be found in [18].

2) Solving Residues: After the poles $\{b_i\}$ are known, substitute $\{b_i\}$ into (11) and the residues $\{a_i\}$ can be solved by using the first M moments

$$-\begin{bmatrix} b_1^{-1} & b_2^{-1} & \cdots & b_M^{-1} \\ b_1^{-2} & b_2^{-2} & \cdots & b_M^{-2} \\ \vdots & \vdots & \vdots & \vdots \\ b_1^{-M} & b_2^{-M} & \cdots & b_M^{-M} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_{M-1} \end{bmatrix}. \tag{14}$$

The aforementioned algorithm assumes that the poles $\{b_i\}$ are distinct. Otherwise, if repeated poles exist, the unknown poles and residues must be solved using a more comprehensive algorithm described in [18]. Once the poles $\{b_i\}$ and residues $\{a_i\}$ are determined, the PDF pdf(f) is optimally approximated by $h(t)$ in (9) and the CDF cdf(f) is optimally approximated by the step response:

$$s(t) = \int_0^t h(\tau) d\tau = \begin{cases} \sum_{i=1}^{M} \frac{a_i}{b_i} \cdot (e^{b_i t} - 1), & (\text{if } t \geq 0) \\ 0, & (\text{if } t < 0). \end{cases} \tag{15}$$

It should be noted that many implementation issues must be considered to make our proposed approach, APEX, feasible and efficient. For example, the impulse response of a causal LTI system is only nonzero for $t \geq 0$, but a PDF in practical applications can be nonzero for $f \leq 0$. In Section IV, we will propose several schemes to address these problems.

The aforementioned moment-matching method was previously applied to IC interconnect order reduction [18], [19], and it is related to the Padé approximation in linear control theory [20]. In the following section, we will explain why such a moment-matching approach is efficient in approximating PDF/CDF functions.
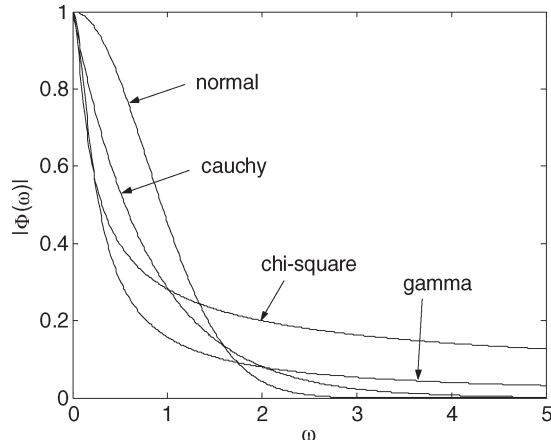
Fig. 2. Characteristic functions of several typical distributions.

B. Connection to Probability Theory

In probability theory, given a random variable f whose PDF is $\text{pdf}(f)$, the characteristic function is defined as the Fourier transform of $\text{pdf}(f)$ [8]

$$\Phi(\omega) = \int_{-\infty}^{+\infty} \text{pdf}(f) \cdot e^{j\omega f} \cdot df = \int_{-\infty}^{+\infty} \text{pdf}(f) \cdot \sum_{k=0}^{+\infty} \frac{(j\omega f)^k}{k!} \cdot df. \tag{16}$$

Substituting (8) into (16) yields

$$\Phi(\omega) = \sum_{k=0}^{+\infty} s_k \cdot (-j\omega)^k. \tag{17}$$

Equation (17) implies an important fact that the time moments defined in (8) are related to the Taylor expansion of the characteristic function at the expansion point $\omega = 0$. Matching the first $2M$ moments in (11) is equivalent to matching the first $2M$ Taylor expansion coefficients between the original characteristic function $\Phi(\omega)$ and the approximated rational function $H(s)$.

To explain why the moment-matching approach is efficient, we first need to show two important properties of the characteristic function that are described in [8].

Property 1: A characteristic function has the maximal magnitude at $\omega = 0$, i.e., $|\Phi(\omega)| \leq \Phi(0) = 1$.

Property 2: A characteristic function $\Phi(\omega) \to 0$ when $\omega \to \infty$.

Fig. 2 shows the characteristic functions for several typical random distributions. The above two properties imply an interesting fact: Namely, given a random variable f, the magnitude of its characteristic function decays as ω increases. Therefore, the optimally approximated $H(s)$ in (9) is a low-pass system. It is well known that a Taylor expansion is accurate around the expansion point. Since a low-pass system is mainly determined by its behavior in the low-frequency range (around $\omega = 0$), it can be accurately approximated by matching the first several Taylor coefficients at $\omega = 0$, i.e., the moments. In addition, the rational function form utilized in APEX is an efficient form to approximate the transfer function $H(s)$ of a low-pass system. These conclusions have been verified in other applications

(e.g., IC interconnect order reduction [18], [19]), and they provide the theoretical background to explain why the APEX works well for PDF/CDF approximation, as will be demonstrated by the numerical examples in Section VII.

C. Proof of Convergence

In the previous section, we have intuitively explained why the moment-matching approach is efficient in approximating PDF/CDF functions. However, there is a theoretical question which might be raised: Given a random variable, can the PDF/CDF functions always be uniquely determined by its moments? In general, the answer is no. It has been observed in mathematics that some probability distributions cannot be uniquely determined by their moments. One example described in [16] is the following PDF:

$$\text{pdf}(f) = \begin{cases} \frac{e^{-0.5 \cdot [\ln(f)]^2}}{\sqrt{2\pi}f} \cdot \{1 + a \cdot \sin[2\pi \cdot \ln(f)]\}, & (\text{if } f > 0) \\ 0, & (\text{if } f \leq 0) \end{cases} \tag{18}$$

where $a \in [-1, 1]$. It can be verified that all moments of the PDF $\text{pdf}(f)$ in (18) are independent of a, although varying a changes $\text{pdf}(f)$ significantly [16]. It, in turn, implies that the PDF in (18) cannot be uniquely determined by its moments.

However, there are special cases for which the moment problem is guaranteed to converge, i.e., the PDF/CDF functions are uniquely determined by the moments. The following Carleman theorem states one of those special cases and gives a sufficient condition for the convergence of the moment problem.

Theorem 1 (Carleman [16]): A probability distribution on the interval $(-\infty, +\infty)$, i.e., the Hamburger moment problem [see (6)], can be uniquely determined by its moments $\{m_k, k = 1, 2, \ldots\}$ if

$$\sum_{k=1}^{+\infty} (m_{2k})^{\frac{-1}{2k}} = \infty. \tag{19}$$

Based on the Carleman theorem, we can prove that the moment-matching approach utilized in APEX is asymptotically convergent when applied to quadratic response surface models. Namely, given a quadratic response surface model f that is a quadratic function of the normally distributed random variables, the probability distribution of f can be uniquely determined by its moments $\{m_k, k = 1, 2, \ldots, K\}$ when K approaches infinity, i.e., $K \to +\infty$. The asymptotic convergence of APEX can be formally stated by the following theorem. The detailed proof of Theorem 2 is given in Appendix.

Theorem 2: Given the quadratic response surface model f in (4) where the random variables ΔY are independent and satisfy the standard normal distribution $N(0, 1)$, the probability distribution of f can be uniquely determined by its moments $\{m_k, k = 1, 2, \ldots\}$.

IV. IMPLEMENTATIONS OF APEX

Our proposed APEX approach is made practically feasible by applying several novel algorithms, including: 1) a binomial scheme for high-order moment computation; 2) a modified

Chebyshev inequality for PDF/CDF shifting; 3) a nonlinear-companding method to improve approximation accuracy; and 4) a reverse-evaluation technique for best case/worst case analysis. In this section, we describe the mathematical formulation of each of these algorithms.

A. Binomial Moment Evaluation

A key operation required in APEX is the computation of the high-order time moments defined in (8) for a given random variable f. Such a time moment evaluation is equivalent to computing the expected values of $\{f^k, k = 0, 1, \ldots, 2M - 1\}$. Given the quadratic response surface model in (4), f^k can be represented as a high-order polynomial of ΔY

$$f^k(\Delta Y) = \sum_i c_i \cdot \Delta y_1^{\alpha_{1i}} \cdot \Delta y_2^{\alpha_{2i}} \cdot \cdots \cdot \Delta y_N^{\alpha_{Ni}} \quad (20)$$

where Δy_i is the ith element in the vector ΔY, c_i is the coefficient of the ith product term, and α_{ij} is the positive integer exponent. Since the random variables ΔY are independent after PCA analysis, we have

$$E(f^k) = \sum_i c_i \cdot E\left(\Delta y_1^{\alpha_{1i}}\right) \cdot E\left(\Delta y_2^{\alpha_{2i}}\right) \cdot \cdots \cdot E\left(\Delta y_N^{\alpha_{Ni}}\right) \quad (21)$$

where $E(\bullet)$ stands for the expected value operator. In addition, remember that the random variables ΔY satisfy the standard normal distribution $N(0, 1)$, which yields [8]

$$E(\Delta y^k) = \begin{cases} 1, & \text{(if } k = 0) \\ 0, & \text{(if } k = 1, 3, 5, \ldots) \\ 1 \cdot 3 \cdots (k-1), & \text{(if } k = 2, 4, 6, \ldots). \end{cases} \quad (22)$$

Substituting (22) into (21), the expected value of f^k can be determined.

The previous computation scheme is called direct moment evaluation in this paper. The key disadvantage of such a moment evaluation is that, as k increases, the total number of the product terms in (21) will exponentially increase, thereby quickly making the computation infeasible. To overcome this difficulty, we propose a novel binomial-moment-evaluation scheme that consists of two steps: quadratic model diagonalization and moment evaluation. The binomial moment evaluation recursively computes the high-order moments without explicitly constructing the high-order polynomial f^k in (20).

1) Quadratic Model Diagonalization: The first step of binomial moment evaluation is to remove the cross product terms in the quadratic response surface model (4), thereby yielding a much simpler, but equivalent quadratic model. According to a matrix theory [10], any symmetric matrix $A \in R^{N \times N}$ can be diagonalized as

$$A = U \cdot \Lambda \cdot U^{\mathrm{T}} \quad (23)$$

where $\Lambda = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_N)$ contains the eigenvalues of A and $U = [U_1, U_2, \ldots, U_N]$ is an orthogonal matrix (i.e.,

$U^{\mathrm{T}}U = I$) containing the eigenvectors. Define the new random variables ΔZ as follows:

$$\Delta Z = U^{\mathrm{T}} \cdot \Delta Y. \quad (24)$$

Substituting (24) into (4) yields

$$f(\Delta Z) = \Delta Z^{\mathrm{T}} \cdot \Lambda \cdot \Delta Z + Q^{\mathrm{T}} \cdot \Delta Z + C$$
$$= \sum_{i=1}^{N} \left(\sigma_i \cdot \Delta z_i^2 + q_i \cdot \Delta z_i\right) + C \quad (25)$$

where Δz_i is the ith element in the vector ΔZ and $Q = [q_1, q_2, \ldots, q_N]^{\mathrm{T}} = U^{\mathrm{T}} B$. Equation (25) implies that there is no cross product term in the quadratic model after the diagonalization. In addition, the following theorem guarantees that the random variables ΔZ defined in (24) are still independent and satisfy the standard normal distribution $N(0, 1)$.

Theorem 3: Given a set of independent random variables ΔY satisfying the standard normal distribution $N(0, 1)$ and an orthogonal matrix U, the random variables ΔZ defined in (24) are independent, and satisfy the standard normal distribution $N(0, 1)$.

Proof: Since the random variables ΔZ are the linear combinations of the zero-mean normal distributions ΔY, ΔZ are also normally distributed and have zero mean. The correlation matrix for ΔZ is given by

$$E(\Delta Z \cdot \Delta Z^{\mathrm{T}}) = E(U^{\mathrm{T}} \cdot \Delta Y \cdot \Delta Y^{\mathrm{T}} \cdot U)$$
$$= U^{\mathrm{T}} \cdot E(\Delta Y \cdot \Delta Y^{\mathrm{T}}) \cdot U. \quad (26)$$

Remember that ΔY are a set of independent random variables satisfying the standard normal distribution $N(0, 1)$, i.e., $E(\Delta Y \cdot \Delta Y^{\mathrm{T}}) = I$, and the matrix U is orthogonal, i.e., $U^{\mathrm{T}}U = I$. Thus, we have

$$E(\Delta Z \cdot \Delta Z^{\mathrm{T}}) = U^{\mathrm{T}} \cdot E(\Delta Y \cdot \Delta Y^{\mathrm{T}}) \cdot U = U^{\mathrm{T}} \cdot I \cdot U = I. \quad (27)$$

Equation (27) implies that the random variables in ΔZ are uncorrelated and their variances are all equal to one. In addition, uncorrelated random variables with normal distributions are also independent [8]. ∎

2) Moment Evaluation: We now demonstrate the use of the simplified quadratic model (25) for fast moment evaluation. Based on (25), we define a set of new random variables

$$g_i = \sigma_i \cdot \Delta z_i^2 + q_i \cdot \Delta z_i$$
$$h_l = \sum_{i=1}^{l} g_i + C = \sum_{i=1}^{l} \left(\sigma_i \cdot \Delta z_i^2 + q_i \cdot \Delta z_i\right) + C. \quad (28)$$

Comparing (28) with (25), it is easy to verify that when $l = N$, $h_N = f$. Instead of computing the high-order moments of f directly, the proposed binomial moment evaluation successively computes the moments of h_l, as shown in the following algorithm.

Algorithm 1 (Binomial Moment Evaluation)

1) Start from $h_0 = C$ and compute $E(h_0^k) = C^k$ for each $\{k = 0, 1, \ldots, 2M - 1\}$. Set $l = 1$.
2) For each $\{k = 0, 1, \ldots, 2M - 1\}$, compute

$$E\left(g_l^k\right) = E\left[\left(\sigma_l \cdot \Delta z_l^2 + q_l \cdot \Delta z_l\right)^k\right]$$
$$= \sum_{i=0}^{k} \binom{k}{i} \cdot \sigma_l^i q_l^{k-i} \cdot E\left(\Delta z_l^{k+i}\right) \quad (29)$$
$$E\left(h_l^k\right) = E\left[(h_{l-1} + g_l)^k\right]$$
$$= \sum_{i=0}^{k} \binom{k}{i} \cdot E\left(h_{l-1}^i\right) \cdot E\left(g_l^{k-i}\right). \quad (30)$$

3) If $l = N$, then go to Step 4). Otherwise, $l = l + 1$ and return Step 2).
4) For each $\{k = 0, 1, \ldots, 2M - 1\}$, we have $E(f^k) = E(h_N^k)$.

Step 2) in Algorithm 1 is the key operation required by the binomial-moment-evaluation algorithm. In Step 2), both (29) and (30) utilize the binomial theorem to get the binomial series. Therefore, we refer to this algorithm as the binomial moment evaluation in this paper.

In (29), the expected values $E(\Delta z_l^{k+i})$ can be easily evaluated using the closed-form expression (22), since Δz_l satisfies the standard normal distribution $N(0, 1)$. Equation (30) utilizes the property that h_{l-1} and g_l are independent, because h_{l-1} is a function of $\{\Delta z_i, i = 1, 2, \ldots, l - 1\}$, g_l is a function of Δz_l and all $\{\Delta z_i, i = 1, 2, \ldots, N\}$ are mutually independent (see Theorem 3). Therefore, $E(h_{l-1}^i \cdot g_l^{k-i}) = E(h_{l-1}^i) \cdot E(g_l^{k-i})$, where the values of $E(h_{l-1}^i)$ and $E(g_l^{k-i})$ have already been computed in previous steps.

The main advantage of the binomial-moment-evaluation algorithm is that, unlike the direct moment evaluation in (21), it does not explicitly construct the high-order polynomial f^k. Therefore, unlike the direct moment evaluation where the total number of the product terms will exponentially increase, both $E(g_l^k)$ in (29) and $E(h_l^k)$ in (30) contain at most $2M$ product terms. Since $\{k = 0, 1, \ldots, 2M - 1\}$ and $\{l = 0, 1, \ldots, N\}$ for an Mth order APEX approximation with N independent random variables, the total number of $E(g_l^k)$ and $E(h_l^k)$ that need to be computed is $O(MN)$. In addition, the matrix diagonalization in (23) is only required once and has a complexity of $O(N^3)$. Therefore, the computational complexity of the proposed algorithm is $O(M^2N) + O(N^3)$. In most circuit-level applications, N is small (around 5–100) after PCA analysis, and selecting $M = 7-10$ provides sufficient accuracy for moment matching. With these typical values of M and N, the proposed binomial moment evaluation is extremely fast, as we will demonstrate by the numerical examples in Section VII.

It should be noted that as long as the circuit performance f is approximated by the quadratic model in (4) and the process variations are normally distributed, the proposed binomial moment evaluation yields the exact high-order moment values (except for numerical errors). There is no further assumption or approximation made by the algorithm.
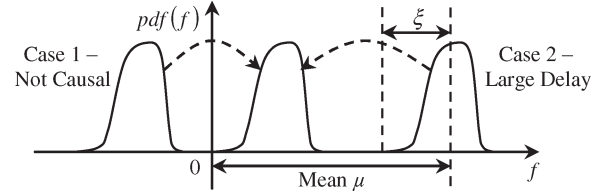

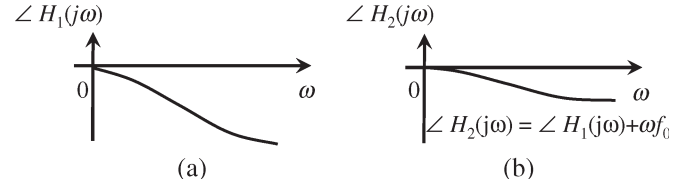Fig. 3. PDF/CDF shifting is required in two cases.

Fig. 4. Phase plots of the Fourier transforms of pdf(f) and pdf$(f + f_0)$. (a) The phase plot of pdf(f) presents a large phase shift. (b) The phase plot of pdf$(f + f_0)$ shows a phase reduction of ωf_0.

In summary, the binomial moment evaluation utilizes a statistical independence theory to efficiently compute the high-order moments that are required by the moment matching of APEX. Compared with the direct moment evaluation in (20)–(22) whose computational complexity is $O(N^M)$, the proposed binomial-moment-evaluation algorithm reduces the complexity to $O(M^2N) + O(N^3)$.

B. PDF/CDF Shifting

APEX approximates the unknown PDF pdf(f) as the impulse response $h(t)$ of an LTI system. The impulse response of a causal system is only nonzero for $t \geq 0$, but a PDF in practical applications can be nonzero for $f \leq 0$. In such cases, we need to right shift the unknown pdf(f) by f_0 and use the impulse response $h(t)$ to approximate the shifted function pdf$(f - f_0)$, as shown in Fig. 3 (case 1).

In addition, even if the unknown PDF pdf(f) is zero for all $f \leq 0$, it can be far away from the origin $f = 0$, as shown in Fig. 3 (case 2). As such, the corresponding impulse response $h(t)$ presents a large delay in the time domain, which corresponds to a large phase shift in the frequency domain, as shown in Fig. 4(a). Since each pole or zero in an LTI system contributes to a phase shift of 90°, the large phase shift must be approximated using a great number of poles or zeros, resulting in a high-order transfer function. In such cases, we need to left shift the unknown pdf(f) by f_0 and, therefore, reduce the phase by ωf_0, as shown in Fig. 4(b). The shifted function pdf$(f + f_0)$ can be accurately approximated by the impulse response $h(t)$ of a low-order LTI system, since it has a much smaller phase shift in the frequency domain.

The previous analysis implies that it is crucial to determine the correct value of f_0 for PDF/CDF shifting. Over-shifting the unknown PDF to either left or right side can increase the approximation error. In this paper, process variations are modeled as normal distributions, which are unbounded and distributed over $(-\infty, +\infty)$. Therefore, any circuit performance f approximated by the quadratic model in (4) is also unbounded. It is difficult to completely shift pdf(f) to the positive axis.

However, since f is a random variable, pdf(f) can be left shifted by f_0[3] such that the probability $P(f - f_0 \leq 0)$ is sufficiently small. As shown in Fig. 3, the PDF/CDF shifting problem can be stated as follows: Find the value ξ and left shift pdf(f) by $f_0 = \mu - \xi$, where μ is the mean value of f, such that the probability $P(f - f_0 \leq 0)$ is not greater than a given error tolerance ε. In addition, we want to select the value ξ to be as small as possible, i.e., find the smallest ξ satisfying $P(f - f_0 \leq 0) \leq \varepsilon$. A small ξ results in a small time-domain delay in $h(t)$ and, therefore, high approximation accuracy for pdf(f). To estimate ξ, we need the following theorem.

Theorem 4: Given a random variable f, for any $\xi \geq 0$ and $k = 2, 4, 6, \ldots$

$$P\left(|f - \mu| \geq \xi\right) \leq \frac{E\left[(f - \mu)^k\right]}{\xi^k} \qquad (31)$$

where μ is the mean value of f.

Proof: For any $k = 2, 4, 6, \ldots$, we have

$$P\left(|f - \mu| \geq \xi\right) = \int_{|f-\mu| \geq \xi} \mathrm{pdf}(f) \cdot df$$

$$\leq \int_{|f-\mu| \geq \xi} \frac{(f - \mu)^k}{\xi^k} \cdot \mathrm{pdf}(f) \cdot df$$

$$\leq \int_{-\infty}^{+\infty} \frac{(f - \mu)^k}{\xi^k} \cdot \mathrm{pdf}(f) \cdot df$$

$$= \frac{E\left[(f - \mu)^k\right]}{\xi^k}. \qquad (32)$$

Note that the proof is not restricted to any specific probability distribution. ∎

Based on (31), if the PDF pdf(f) is left shifted by $f_0 = \mu - \xi$, we have

$$P(f - f_0 \leq 0) = P(f - \mu + \xi \leq 0) = P(-f + \mu \geq \xi)$$

$$\leq P\left(|f - \mu| \geq \xi\right) \leq \frac{E\left[(f - \mu)^k\right]}{\xi^k} \qquad (33)$$

where $k = 2, 4, 6, \ldots$. Therefore, one sufficient condition for $P(f - f_0 \leq 0) \leq \varepsilon$ is

$$\frac{E\left[(f - \mu)^k\right]}{\xi^k} \leq \varepsilon \qquad (k = 2, 4, 6, \ldots) \qquad (34)$$

which is equivalent to

$$\xi \geq \left\{ \frac{E\left[(f - \mu)^k\right]}{\varepsilon} \right\}^{\frac{1}{k}} \qquad (k = 2, 4, 6, \ldots). \qquad (35)$$

Equation (35) estimates ξ using high-order central moments. In an Mth order approximation, after the high-order moments $\{E(f^k), k = 0, 1, \ldots, 2M - 1\}$ are computed by the binomial

[3] f_0 is negative in case of right shifting.

moment evaluation in Algorithm 1, the central moments can be easily calculated using the binomial theorem

$$E\left[(f - \mu)^k\right] = \sum_{i=0}^{k} \binom{k}{i} \cdot E(f^i) \cdot (-\mu)^{k-i} \qquad (36)$$

where $\mu = E(f)$. Then, using (35), an estimated ξ is computed for each $k = 2, 4, \ldots, 2M - 2$, which is denoted as ξ_k. The minimal value of all these ξ_k values is utilized as the final ξ for PDF/CDF shifting, since we aim to find the smallest ξ to achieve high approximation accuracy.

It is worth mentioning that when $k = 2$, (31) is the well-known Chebyshev inequality [8]. We have modified the second-order Chebyshev inequality to higher orders and, therefore, refer to (31) as the modified Chebyshev inequality. In practical applications, we find that high-order moments provide a much tighter (i.e., smaller) estimation of ξ, as will be demonstrated by the numerical examples in Section VII.

In summary, the proposed modified Chebyshev inequality (31) provides an effective way to estimate the boundary for PDF/CDF shifting. As such, the major part of the unknown PDF/CDF can be moved to the positive axis, which is then accurately approximated by the impulse/step response of a causal LTI system. For a specific circuit performance, the PDF/CDF boundary can also be determined by other available knowledge, in addition to using the modified Chebyshev inequality. For example, the bandwidth of an amplifier is always positive, resulting in a lower bound of zero for the bandwidth performance.

C. Nonlinear Companding

Given a random variable f, the deviation of the PDF pdf(f) from a normal distribution can be measured using the skewness

$$\mathrm{Skewness}(f) = \frac{E\left[(f - \mu)^3\right]}{\left\{E\left[(f - \mu)^2\right]\right\}^{1.5}} \qquad (37)$$

where μ is the mean value of f. Note that the skewness is zero if the random variable f is a normal distribution. Although the moment matching is a general methodology, we find that its approximation accuracy depends on the skewness of the unknown probability distribution, as illustrated by the following simple example:

$$f(\Delta x) = \theta \cdot \Delta x^2 + (1 - |\theta|) \cdot \Delta x \qquad (38)$$

where Δx is a random variable satisfying the standard normal distribution $N(0, 1)$ and $\theta \in [-1, 1]$ is a parameter controlling the nonlinearity of the quadratic function $f(\Delta x)$. The skewness of f can be changed by varying the parameter θ, as shown in Fig. 5. For most practical applications, the performance variation should be dominated by the linear term in the response surface model. Given this assumption, the value of θ in (38) should be in the range of $[-1/3, 1/3]$. If $\theta = \pm 1/3$, both the quadratic term and the linear term in (38) are around ± 3 at the $\pm 3\sigma$ point of Δx, i.e., $\Delta x = \pm 3$. Otherwise, if $|\theta| \geq 1/3$, the quadratic term dominates the linear term in (38), which
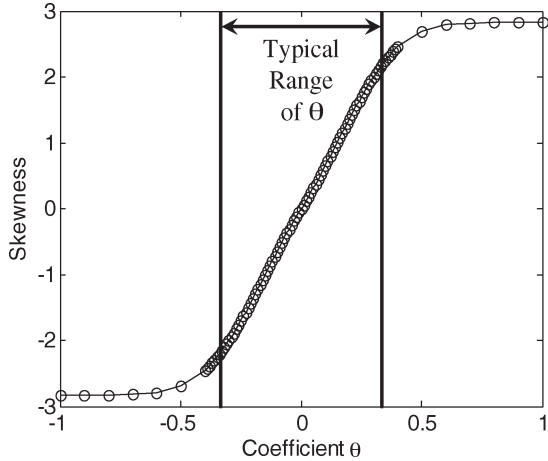
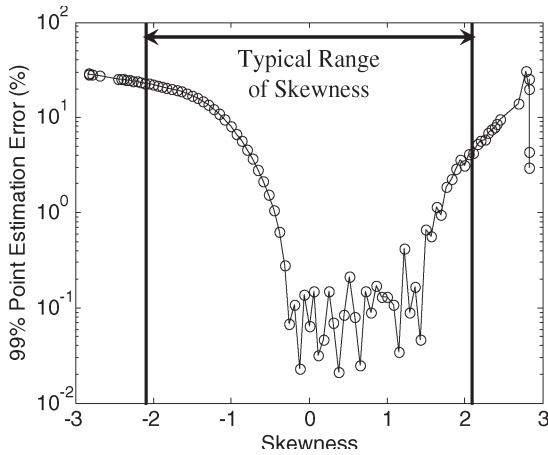Fig. 5. Skewness depends on the coefficient θ in (38).

Fig. 6. Moment-matching error depends on the skewness of the unknown probability distribution.

rarely happens in practice. It, in turn, yields the typical ranges $\theta \in [-1/3, 1/3]$ and Skewness $\in [-2.1, 2.1]$ shown in Fig. 5.

Fig. 6 shows the relation between the moment-matching error and the skewness of the function f in (38). The f value at the 99% point of the CDF is estimated by the moment matching. This moment-matching result is compared with the Monte Carlo simulation result with 10^6 samples. The relative difference between these two results is used as a measure of the estimation error in Fig. 6. Two observations can be made from the data in Fig. 6. First, the moment matching is most accurate when the skewness is close to zero, i.e., the probability distribution is close to symmetric. As the nonlinearity increases and the unknown probability distribution becomes asymmetric, the moment-matching error increases. Second, the moment matching error is not a symmetric function of the skewness. As shown in Fig. 6, within the typical range Skewness $\in [-2.1, 2.1]$, the maximal error for the negative skewness (more than 21% when the skewness is equal to -2.1) is much larger than its counterpart for the positive skewness (less than 5% when the skewness is equal to 2.1). Therefore, it is more critical to reduce the moment-matching error when it is applied to the probability distributions with large negative skewness.

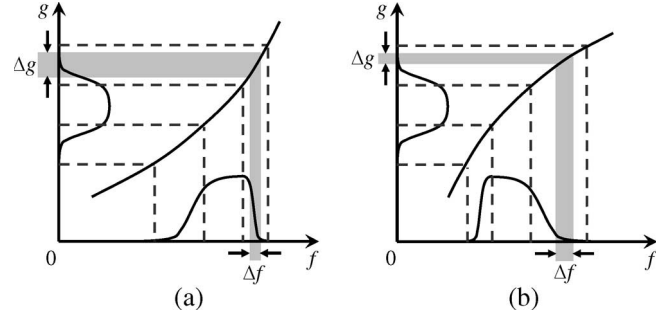In this section, we propose a nonlinear-companding scheme to address this skewness problem and further improve the

Fig. 7. Illustration of nonlinear companding. (a) The probability distribution of f has a negative skewness. (b) The probability distribution of f has a positive skewness.

approximation accuracy of APEX. Our main idea here is to automatically compress and/or expand the unknown probability distribution, such that the transformed probability distribution almost has a zero skewness and can be approximated accurately.

1) Algorithm of Nonlinear Companding: Given a PDF $\mathrm{pdf}_1(f)$, we define a nonlinear mapping $g = \mathrm{map}(f)$. After this nonlinear mapping, the transformed PDF becomes $\mathrm{pdf}_2(g)$. If the original PDF $\mathrm{pdf}_1(f)$ has a nonzero skewness, we can select a specific mapping function $g = \mathrm{map}(f)$ to compress or expand $\mathrm{pdf}_1(f)$ such that the transformed PDF $\mathrm{pdf}_2(g)$ almost has a zero skewness, as shown in Fig. 7. Compared with directly applying the moment matching to estimate $\mathrm{pdf}_1(f)$, applying the moment matching to approximate the transformed PDF $\mathrm{pdf}_2(g)$ can significantly change the approximation error due to the following two reasons.

First, as shown in Fig. 6, the moment matching is most accurate when the unknown PDF has a zero skewness. Therefore, applying the moment matching to $\mathrm{pdf}_2(g)$, which almost has a zero skewness, can achieve better approximation accuracy than directly estimating $\mathrm{pdf}_1(f)$. Second, if the moment matching is applied to estimate the transformed performance value g at a specific point (e.g., the 1% point or the 99% point) on CDF, the estimation error Δg can be represented as

$$\Delta g = g_0 - g \qquad (39)$$

where g_0 and g are the exact and estimated values of the transformed performance, respectively. Substituting $g = \mathrm{map}(f)$ into (39) yields

$$\Delta g = \mathrm{map}(f_0) - \mathrm{map}(f) = \mathrm{map}'(f_0) \cdot \Delta f \qquad (40)$$

where f_0 and f are the exact and estimated values of the original performance respectively, $\mathrm{map}'(f_0)$ denotes the derivative value of the function $g = \mathrm{map}(f)$ at $f = f_0$, and $\Delta f = f_0 - f$ is the estimation error of the original performance. Equation (40) reveals an important fact that, due to nonlinear companding, the estimation error is scaled by the derivative $\mathrm{map}'(f_0)$ when the performance value g is transformed back to the original value f. Therefore, depending on the value of $\mathrm{map}'(f_0)$, Δf can be either reduced [if $\mathrm{map}'(f_0)$ is large] or increased [if $\mathrm{map}'(f_0)$ is small].

The total error change due to nonlinear companding is a combination of the aforementioned two effects. For example, if the probability distribution of f has a negative skewness,

transforming f to g yields an accurate approximation $\mathrm{pdf}_2(g)$, i.e., a small estimation error Δg, due to the close-to-zero skewness of $\mathrm{pdf}_2(g)$. After the performance value g is transformed back to the original performance value f, the estimation error Δf at the 99% point on CDF is further reduced, since the nonlinear function $g = \mathrm{map}(f)$ has a large derivative around the 99% point, as shown in Fig. 7(a).

On the other hand, if the probability distribution of f has a positive skewness, $\mathrm{pdf}_2(g)$ can be once again approximated accurately. However, the nonlinear function $g = \mathrm{map}(f)$ has a small derivative around the 99% point on CDF. Therefore, when the performance value g is transformed back to the original performance value f, the estimation error Δf at the 99% point is increased, as shown in Fig. 7(b). In this case, the total estimation error can be either reduced or increased. In other words, applying the nonlinear companding does not always improve the estimation accuracy of the 99% point on CDF, if the unknown probability distribution has a positive skewness.

In APEX, we apply the nonlinear companding to those probability distributions with large negative skewness, especially when the skewness is smaller than -0.5, in order to reduce the estimation error of the 99% point on CDF. In such cases, the direct moment-matching error is substantially large (shown in Fig. 6), and applying nonlinear companding is guaranteed to improve the accuracy. For those probability distributions with positive skewness, the moment matching can be applied directly to estimate the 99% point on CDF. In our numerical experiments, we do not find any cases where the positive skewness of a circuit performance is so large that the direct moment matching yields a substantially large error.

In summary, the nonlinear-companding technique can be selectively applied to reduce the estimation error of the 99% point on CDF. For estimating the 1% point on CDF, we will apply another reverse-evaluation technique to improve the accuracy, which will be discussed in detail in Section IV-D.

2) Constructing Companding Function: Given a PDF $\mathrm{pdf}_1(f)$, a specific companding function $g = \mathrm{map}(f)$ should be selected such that the transformed PDF $\mathrm{pdf}_2(g)$ almost has a zero skewness. In this section, we describe the methodology for constructing such a companding function automatically. Since nonlinear companding is only applied to negative skewness, we will limit our discussion to those probability distributions with negative skewness.

In addition to the skewness requirement, i.e., the transformed PDF $\mathrm{pdf}_2(g)$ should have a close-to-zero skewness, a good companding function should further satisfy the following two requirements.

1) Monotonicity. A companding function should be monotonically increasing or decreasing. In other words, the companding function $g = \mathrm{map}(f)$ should establish a one-to-one mapping between f and g. Otherwise, given a performance value g, it is not guaranteed to transform g back to f and find the original performance value.
2) Efficiency. In order to make the nonlinear companding efficient, the companding function should be easy to use. This efficiency requirement has a twofold meaning. First, given any PDF $\mathrm{pdf}_1(f)$, the companding function $g =$
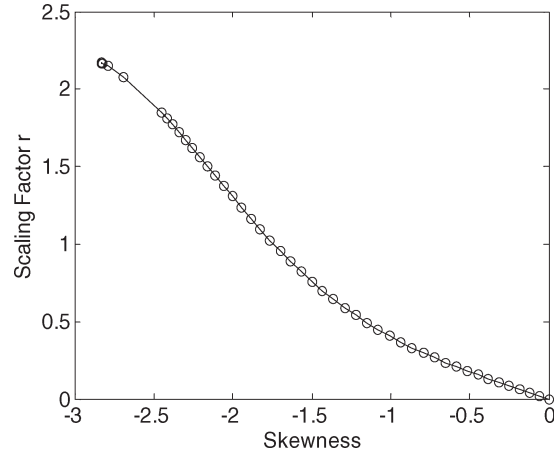

Fig. 8. Relation between the skewness and the scaling factor r for nonlinear companding.

$\mathrm{map}(f)$ should be constructed automatically and quickly. Second, after the PDF $\mathrm{pdf}_1(f)$ is transformed to $\mathrm{pdf}_2(g)$, the high-order moments of $\mathrm{pdf}_2(g)$ should be easy to compute such that the transformed PDF $\mathrm{pdf}_2(g)$ can be efficiently approximated using the moment matching.

Although there are many ways to build the companding function, we will use the following exponential function due to its simplicity:

$$g = \exp(\alpha \cdot f + \beta) \qquad (41)$$

where α and β are two parameters controlling the nonlinearity. When the PDF $\mathrm{pdf}_1(f)$ has a negative skewness, α in (41) should be positive, as shown in Fig. 7(a).

Given a random variable f, we first compute

$$\mu = E(f)$$
$$\sigma = \sqrt{E\left[(f - \mu)^2\right]} \qquad (42)$$

to normalize f, such that the mean and variance of $(f - \mu)/\sigma$ are equal to 0 and 1, respectively. At this stage, the normalization is done using the first- and second-order moments of f. Then, in the next step, the third-order moment, i.e., the skewness, of the normalized random variable $(f - \mu)/\sigma$ is used as a criterion to select a proper scaling factor r and construct the companding function in (41) where

$$\alpha = r/\sigma$$
$$\beta = -r\mu/\sigma. \qquad (43)$$

In order to determine the relation between the skewness of the normalized random variable $(f - \mu)/\sigma$ and the scaling factor r in (43), we repeatedly change the parameter θ in (38) and generate a number of random distributions $\mathrm{pdf}_1(f)$ with negative skewness. Then, using the companding function defined in (41)–(43), an r value is optimized for each of these distributions to force the transformed distribution $\mathrm{pdf}_2(g)$ to have a zero skewness. Based on these experiments, a one-dimensional table is generated, as shown in Fig. 8.

The aforementioned algorithm for constructing the companding function is extremely fast, since it only involves the moment computation up to the third order and a one-dimensional table lookup. However, it should be noted that the curve in Fig. 8 is extracted by using the simple quadratic function f in (38), which only includes one random variable Δx. When such a lookup table is applied to the general multidimensional quadratic function f in (4), the skewness of the transformed PDF $\mathrm{pdf}_2(g)$ might not be exactly zero, since the PDF is not uniquely determined by the first three moments. In such cases, determining the companding function to achieve an exactly zero skewness requires the information of high-order moments, which is computationally expensive. Therefore, in APEX, we simply construct the companding function using the first three moments, in order to make the computation task tractable. It should be noted that, even if the transformed PDF $\mathrm{pdf}_2(g)$ does not have an exactly zero skewness, the proposed companding function defined in (41)–(43) with the positive coefficient α always pushes the negative skewness toward zero, thereby always improving the approximation accuracy.

3) Moment Evaluation in Companding: After the companding function in (41) is determined, the high-order moments of the random variable g must be computed efficiently such that the transformed PDF $\mathrm{pdf}_2(g)$ can be approximated using the moment matching. In this section, we describe the algorithm to compute the moments for the transformed performance g in (41). Since any quadratic response surface model in (4) can be converted to the equivalent model in (25), we will assume that the quadratic function f has the form of (25) in our moment computation.

Substituting the quadratic model in (25) into the companding function in (41) yields

$$
g(\Delta Z) = \exp\left[\sum_{i=1}^{N}\left(\alpha\cdot\sigma_i\cdot\Delta z_i^2 + \alpha\cdot q_i\cdot\Delta z_i\right) + \alpha\cdot C + \beta\right].
$$
(44)

Therefore, the kth order moment of g can be computed as

$$
\begin{aligned}
E(g^k) &= E\left[\exp\left(\begin{array}{c}k\cdot\alpha\cdot C\\+\,k\cdot\beta\end{array}\right)\cdot\prod_{i=1}^{N}\exp\left(\begin{array}{c}k\cdot\alpha\cdot\sigma_i\cdot\Delta z_i^2\\+\,k\cdot\alpha\cdot q_i\cdot\Delta z_i\end{array}\right)\right]\\
&= \exp\left(\begin{array}{c}k\cdot\alpha\cdot C\\+\,k\cdot\beta\end{array}\right)\cdot\prod_{i=1}^{N}E\left[\exp\left(\begin{array}{c}k\cdot\alpha\cdot\sigma_i\cdot\Delta z_i^2\\+\,k\cdot\alpha\cdot q_i\cdot\Delta z_i\end{array}\right)\right].
\end{aligned}
$$
(45)

Equation (45) utilizes the property that all $\{\exp(k\cdot\alpha\cdot\sigma_i\cdot\Delta z_i^2 + k\cdot\alpha\cdot q_i\cdot\Delta z_i), i = 1, 2, \ldots, N\}$ are mutually independent, because each $\exp(k\cdot\alpha\cdot\sigma_i\cdot\Delta z_i^2 + k\cdot\alpha\cdot q_i\cdot\Delta z_i)$ is a function of Δz_i and all $\{\Delta z_i, i = 1, 2, \ldots, N\}$ are mutually independent (see Theorem 3). Therefore, the expected value of the cross product is equal to the cross product of all expected values. Since all $\{\Delta z_i, i = 1, 2, \ldots, N\}$ satisfy the standard

normal distribution $N(0, 1)$, the expected value of each $\exp(k\cdot\alpha\cdot\sigma_i\cdot\Delta z_i^2 + k\cdot\alpha\cdot q_i\cdot\Delta z_i)$ is equal to

$$
\begin{aligned}
&E\left[\exp\left(k\cdot\alpha\cdot\sigma_i\cdot\Delta z_i^2 + k\cdot\alpha\cdot q_i\cdot\Delta z_i\right)\right]\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\infty}\exp\left(\begin{array}{c}k\cdot\alpha\cdot\sigma_i\cdot\Delta z_i^2\\+\,k\cdot\alpha\cdot q_i\cdot\Delta z_i\end{array}\right)\\
&\qquad\qquad\qquad \cdot\exp\left(-0.5\cdot\Delta z_i^2\right)\cdot d\Delta z_i\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\infty}\exp\Big[(k\cdot\alpha\cdot\sigma_i - 0.5)\\
&\qquad\qquad\qquad \cdot\Delta z_i^2 + k\cdot\alpha\cdot q_i\cdot\Delta z_i\Big]\cdot d\Delta z_i\\
&= \frac{1}{\sqrt{2\pi}}\cdot\exp\left[\frac{-k^2\alpha^2 q_i^2}{4\cdot(k\cdot\alpha\cdot\sigma_i - 0.5)}\right]\\
&\quad \cdot\int_{-\infty}^{-\infty}\exp\left\{\left(\begin{array}{c}k\cdot\alpha\cdot\sigma_i\\-0.5\end{array}\right)\right.\\
&\qquad\qquad \left.\cdot\left[\Delta z_i + \frac{k\cdot\alpha\cdot q_i}{2\cdot(k\cdot\alpha\cdot\sigma_i - 0.5)}\right]^2\right\}\cdot d\Delta z_i\\
&= \frac{1}{\sqrt{2\cdot(0.5 - k\cdot\alpha\cdot\sigma_i)}}\cdot\exp\left[\frac{-k^2\alpha^2 q_i^2}{4\cdot(k\cdot\alpha\cdot\sigma_i - 0.5)}\right]\\
&\qquad\qquad\qquad\qquad\qquad \text{(if}\quad k\cdot\alpha\cdot\sigma_i \le 0.5).
\end{aligned}
$$
(46)

Substituting (46) into (45), the kth order moment of the transformed performance g is obtained.

It should be noted that the expected value in (46) exists if and only if the condition $k\cdot\alpha\cdot\sigma_i \le 0.5$ is satisfied. Otherwise, the integration in (46) does not converge. In (46), the maximal value of k is determined by the APEX approximation order M. The moments up to $2M - 1$th order are required in an Mth order approximation. The value of σ_i in (46) is the coefficient of the quadratic response model in (25), and it is fixed during model fitting. Finally, the value of α in (46) is the coefficient of the companding function in (41), and it is determined when constructing the companding function in (41)–(43). For an Mth order approximation with a given quadratic response model, the values of k and σ_i are fixed. Therefore, in order to satisfy the condition $k\cdot\alpha\cdot\sigma_i \le 0.5$, the value of α in the companding function (41) must be restricted by

$$
\alpha \le \frac{0.5}{(2M - 1)\cdot\sigma_i} \quad (i = 1, 2, \ldots, N \text{ where } \sigma_i > 0).
$$
(47)

Equation (47) poses the new constraints for the companding function $g = \mathrm{map}(f)$. If the value of α is too large when using the equations in (41)–(43) and the lookup table in Fig. 8 to construct the companding function, α must be reduced such that the constraints in (47) are satisfied. In these cases, the skewness of the transformed PDF $\mathrm{pdf}_2(g)$ might not be exactly zero; however, the proposed companding function defined in (41)–(43) with the positive coefficient α always pushes the negative skewness toward zero, thereby always improving the approximation accuracy.
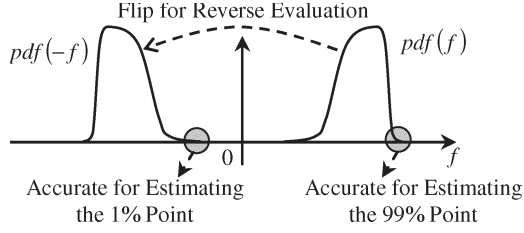
Fig. 9.   Illustration of the reverse evaluation.

For an Mth order approximation with N independent random variables $\{\Delta z_i, i = 1, 2, \ldots, N\}$, computing the first $2M - 1$ moments using (45) and (46) has a complexity of $O(MN)$. In addition, the matrix diagonalization in (23) is only required once and has a complexity of $O(N^3)$. Therefore, the computational complexity of the moment evaluation in (45) and (46) is $O(MN) + O(N^3)$, which is even lower than the binomial moment evaluation proposed in Section IV-A.

In summary, the aforementioned nonlinear-companding scheme is applied to improve the APEX approximation accuracy, if the skewness of the unknown probability distribution is smaller than -0.5. In many practical applications, we find that the nonlinear companding is not required since the skewness of the circuit performance is not so far away from zero. The nonlinear-companding scheme is applied only to a few strongly nonlinear circuit performances, e.g., the offset voltage of an analog amplifier, as will be demonstrated by the numerical examples in Section VII.

D. Reverse Evaluation

In many practical applications, such as robust circuit optimization [4], [21], the best case performance (e.g., the 1% point on CDF) and the worst case performance (e.g., the 99% point on CDF) are two important metrics to be evaluated. As discussed in Section III-B, APEX matches the first $2M$ Taylor expansion coefficients between the original characteristic function $\Phi(\omega)$ and the approximated rational function $H(s)$. Recall that the Taylor expansion is most accurate around the expansion point $\omega = 0$. According to the final value theorem of Laplace transform, accurately approximating $\Phi(\omega)$ at $\omega = 0$ provides an accurate pdf(f) at $f \to \infty$. It, in turn, implies that the proposed approach can accurately estimate the 99% point of the random distribution, as shown in Fig. 9.

The previous analysis motivates us to apply a reverse-evaluation scheme for accurately estimating the 1% point. As shown in Fig. 9, the reverse-evaluation algorithm flips the original pdf(f) to pdf$(-f)$. The 1% point of the original pdf(f) now becomes the 99% point of the flipped pdf$(-f)$ which can be accurately evaluated by APEX.

After the reverse evaluation is applied, however, the skewness of the unknown probability distribution will be changed. For example, if the original pdf(f) has a negative skewness, the flipped pdf$(-f)$ will have a positive skewness, as shown in Fig. 9. Depending on the skewness of the unknown probability distribution, Table II summarizes three cases where the nonlinear companding and/or the reverse evaluation should be selectively applied.

E. Summary

Algorithm 2 (Asymptotic Probability Extraction)
 1. Start from the quadratic response surface model in (4) and a given approximation order M.
 2. Diagonalize the quadratic model based on (23)–(25).
 3. Compute the first three moments $\{E(f^k), k = 1, 2, 3\}$ using the binomial moment evaluation in Algorithm 1.
 4. Compute the central moments $\{E[(f - \mu)^k], k = 1, 2, 3$ using (36), where μ is the mean value of f.
 5. Compute the skewness of f and based on Table II, determine if the nonlinear companding is required or not. If the nonlinear companding is not required:
 6. Compute the high-order moments $\{E(f^k), k = 1, 2, \ldots, 2M - 1\}$ using the binomial moment evaluation in Algorithm 1.
 7. Compute the central moments $\{E[(f - \mu)^k], k = 1, 2, \ldots, 2M-1\}$using (36), where μ is the mean value of f.
 8. Determine the value of ξ using (35) and $f_0 = \mu - \xi$, where μ is the mean value of f.
 9. Compute the time moments $\{s_k = (-1)^k/k! \cdot E[(f-f_0)^k], k = 1, 2, \ldots, 2M - 1\}$, where $E[(f - f_0)^k]$ is similarly evaluated by replacing μ by f_0 in (36).
10. Substitute $\{s_k, k = 1, 2, \ldots, 2M - 1\}$ into (11) and solve the problem unknowns $\{a_i, i = 1, 2, \ldots, M\}$ and $\{b_i, i = 1, 2, \ldots, M\}$.
11. The shifted pdf$(f + f_0)$ is approximated by the impulse response in (9) and the shifted cdf$(f + f_0)$ is approximated by the step response in (15).
12. Shift pdf$(f + f_0)$ and cdf$(f + f_0)$ back to pdf(f) and cdf(f).
    Else if the nonlinear companding is required:
13. Determine the companding function $g = \text{map}(f)$ using the equations in (41)–(43), the lookup table in Fig. 8 and the constraints in (47).
14. Compute the high-order moments $\{E(g^k), k = 1, 2, \ldots, 2M - 1\}$ using (45) and (46).
15. Compute the central moments $\{E[(g - \mu)^k], k = 1, 2, \ldots, 2M - 1\}$ using (36), where μ is the mean value of g.
16. Determine the value of ξ using (35) and $g_0 = \mu - \xi$, where μ is the mean value of g.
17. Compute the time moments $\{s_k = (-1)^k/k! \cdot E[(g-g_0)^k], k = 1, 2, \ldots, 2M - 1\}$, where $E[(g - g_0)^k]$ is similarly evaluated by replacing μ by g_0 in (36).
18. Substitute $\{s_k, k = 1, 2, \ldots, 2M - 1\}$ into (11) and solve the problem unknowns $\{a_i, i = 1, 2, \ldots, M\}$ and $\{b_i, i = 1, 2, \ldots, M\}$.
19. The shifted pdf$_2(g + g_0)$ is approximated by the impulse response in (9) and the shifted cdf$_2(g + g_0)$ is approximated by the step response in (15).
20. Shift pdf$_2(g + g_0)$ and cdf$_2(g + g_0)$ back to pdf$_2(g)$ and cdf$_2(g)$.
21. Transform pdf$_2(g)$ and cdf$_2(g)$ back to pdf$_1(f)$ and cdf$_1(f)$ using the relation:

$$\begin{aligned} \text{pdf}_1(f) &= \text{pdf}_2\left[\text{map}(f)\right] \cdot \text{map}'(f) \\ \text{cdf}_1(f) &= \text{cdf}_2\left[\text{map}(f)\right]. \end{aligned} \qquad (48)$$

    End If

TABLE II
SELECTIVELY APPLYING THE NONLINEAR COMPANDING AND/OR THE REVERSE EVALUATION IN APEX

	1% Point on CDF		99% Point on CDF	
	Flipped *pdf*(–*f*)	Apply Companding?	Original *pdf*(*f*)	Apply Companding?
Case 1	Skewness > 0.5	No	Skewness < –0.5	Yes
Case 2	–0.5 ≤ Skewness ≤ 0.5	No	–0.5 ≤ Skewness ≤ 0.5	No
Case 3	Skewness < –0.5	Yes	Skewness > 0.5	No

Algorithm 2 summarizes the overall implementation of APEX without the reverse evaluation. If the reverse evaluation is required to improve the estimation accuracy for the 1% point on CDF, we need to replace f by $-f$ in Step 3) and run all the subsequent steps for computing $\mathrm{pdf}(-f)$ and $\mathrm{cdf}(-f)$.

Algorithm 2 is based on a given approximation order M. Pillage and Rohrer [18] and Celik *et al.* [19] proposed several methods for iteratively determining M based on the approximation error. The approximation order should be increased if the error is large. These methods can also be applied to APEX. In addition, it is worth mentioning that using an approximation order greater than ten can result in serious numerical problems [18], [19]. In most practical applications, we find that selecting M in the range of 7–10 can achieve the best accuracy.

V. HANDLING NONNORMAL PROCESS VARIATIONS

The moment-matching method used in APEX is general and is not limited to normal process variations. However, the binominal moment evaluation proposed in Section IV-A relies on the normal distribution assumption for process parameters. If the binominal moment evaluation fails to work due to nonnormal process variations, the moment values cannot be easily computed and, therefore, the APEX algorithm is no longer computationally efficient. In this section, we propose a nonlinear transform method to overcome this problem: i.e., if the nonnormal random variables are mutually independent, they can be converted to normal distributions by a nonlinear mapping. However, the nonlinear transform approach is not valid for correlated nonnormal distributions. Handling the correlated nonnormal distributions is extremely difficult and even Monte Carlo simulation becomes impractical in such cases.

A. Independent Nonnormal Process Variations

Given a set of random variables $\Delta X = [\Delta x_1, \Delta x_2, \ldots, \Delta x_N]^{\mathrm{T}}$, we assume that all these random variables $\{\Delta x_i, i = 1, 2, \ldots, N\}$ are mutually independent.[4] A set of one-dimensional functions $\{\Delta y_i = g_i(\Delta x_i), i = 1, 2, \ldots, N\}$ can be constructed to convert ΔX to $\Delta Y = [\Delta y_1, \Delta y_2, \ldots, \Delta y_N]^{\mathrm{T}}$ such that $\{\Delta y_i, i = 1, 2, \ldots, N\}$ satisfy normal distributions [8]. The detailed algorithm for finding the transform function $g_i(\bullet)$ is described in [8]. Since the random variables

---

[4]Correlated normal distributions can be decomposed into independent ones by principal component analysis.
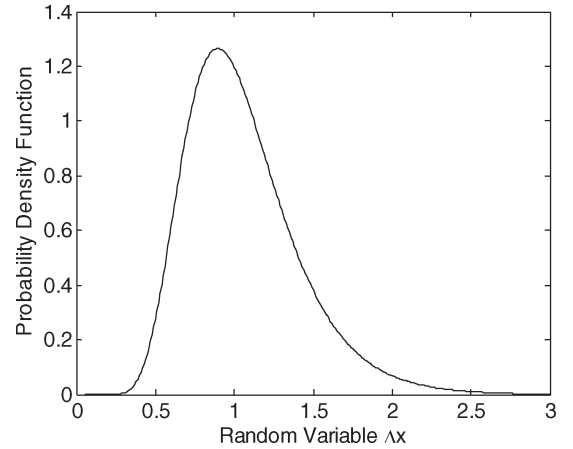

Fig. 10. Probability density function of the log-normal distribution $\Delta x \sim \exp[N(0, 1/3)]$, i.e., the logarithm of Δx is a normal distribution whose mean is 0 and standard deviation is 1/3.

ΔX are mutually independent, ΔY are also mutually independent and their joint PDF is given by

$$\mathrm{pdf}(\Delta y_1, \Delta y_2, \ldots, \Delta y_N) = \mathrm{pdf}(\Delta y_1) \cdot \mathrm{pdf}(\Delta y_2)$$
$$\cdots \cdots \cdot \mathrm{pdf}(\Delta y_N). \quad (49)$$

It is easy to verify that the random variables ΔY constitute a multivariate normal distribution [8]. Therefore, if the response surface model is constructed as a function of the transformed random variables ΔY, i.e., $f(\Delta Y)$ as shown in (4), the APEX algorithm can be directly applied to estimate the probability distribution of the performance f.

The additional nonlinear transforms $\{\Delta y_i = g_i(\Delta x_i), i = 1, 2, \ldots, N\}$, however, generally increase the nonlinearity of the response surface model $f(\Delta Y)$, thereby making it more difficult to approximate $f(\Delta Y)$ as a quadratic function. This nonlinearity problem can be illustrated by the following simple example.

Given a log-normal distribution $\Delta x \sim \exp[N(0, 1/3)]$ as shown in Fig. 10, the logarithm of x is a normal distribution $\Delta y = \log(\Delta x) \sim N(0, 1/3)$. The mean and standard deviation of Δy are 0 and 1/3, respectively. For simplicity, we assume that the performance f is a simple linear function of Δx, i.e., $f = \Delta x$. Substituting $\Delta y = \log(\Delta x)$ into $f = \Delta x$ yields $f = \exp(\Delta y)$ which is no longer a linear or quadratic function. Approximating $f = \exp(\Delta y)$ as a quadratic response surface model introduces additional errors, as shown in Fig. 11. If the process variations $\{\Delta x_i, i = 1, 2, \ldots, N\}$ are more
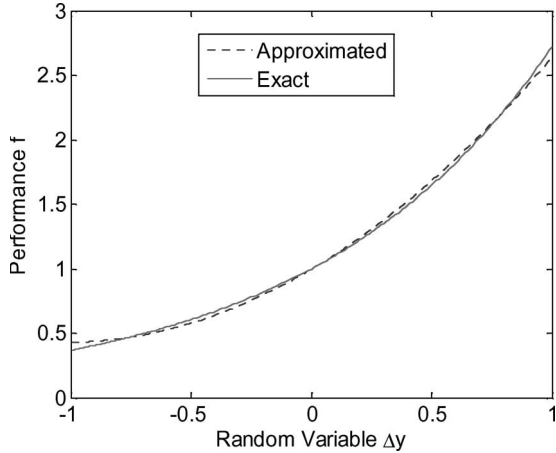
Fig. 11. Function $f = \exp(\Delta y)$ is approximated as a quadratic model within the $\pm 3\sigma$ range of Δy.

different from normal distributions, the transforms $\{g_i(\bullet), i = 1, 2, \ldots, N\}$ become more nonlinear and, therefore, larger response surface modeling error is expected for approximating $f(\Delta Y)$.

In summary, we propose to convert the independent nonnormal process variations ΔX to the normal distributions ΔY by a nonlinear transform. As such, the APEX algorithm can be directly applied to the quadratic response surface model $f(\Delta Y)$. The proposed nonlinear transform approach, however, cannot be applied to correlated nonnormal variations, as will be discussed in the next section.

B. Correlated Nonnormal Process Variations

At first glance, the aforementioned nonlinear transform approach seems applicable to the correlated nonnormal variations. The one-dimensional nonlinear functions $\{\Delta y_i = g_i(\Delta x_i), i = 1, 2, \ldots, N\}$ can be constructed to convert each nonnormal random variable Δx_i to a normal variable Δy_i. However, after the nonlinear transform is performed, the random variables $\{\Delta y_i, i = 1, 2, \ldots, N\}$ are not mutually independent and their joint PDF is not equal to the product of the marginal PDFs, i.e.

$$\text{pdf}(\Delta y_1, \Delta y_2, \ldots, \Delta y_N)$$
$$\neq \text{pdf}(\Delta y_1) \cdot \text{pdf}(\Delta y_2) \cdot \cdots \cdot \text{pdf}(\Delta y_N). \quad (50)$$

In this case, the random variables $\{\Delta y_i, i = 1, 2, \ldots, N\}$ are not guaranteed to constitute a multivariate normal distribution. In other words, even if the random variables $\{\Delta y_i, i = 1, 2, \ldots, N\}$ are marginally normal, they might not be jointly normal [8]. This property can be understood from the following example described in [8].

Consider two random variables Δy_1 and Δy_2, and their joint PDF

$$\text{pdf}(\Delta y_1, \Delta y_2) = \text{pdf}_1(\Delta y_1) \cdot \text{pdf}_2(\Delta y_2)$$
$$\cdot \left\{ \begin{array}{c} 1 + \rho \cdot [2 \cdot \text{cdf}_1(\Delta y_1) - 1] \\ \cdot [2 \cdot \text{cdf}_2(\Delta y_2) - 1] \end{array} \right\} \quad (51)$$

where $|\rho| < 1$, and $\text{pdf}_1(\Delta y_1)$ and $\text{pdf}_2(\Delta y_2)$ are two PDFs with respective CDFs $\text{cdf}_1(\Delta y_1)$ and $\text{cdf}_2(\Delta y_2)$. It is easy to verify that [8]

$$\text{pdf}(\Delta y_1, \Delta y_2) \geq 0$$
$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \text{pdf}(\Delta y_1, \Delta y_2) \cdot d\Delta y_1 \cdot d\Delta y_2 = 1. \quad (52)$$

Equation (52) shows that the function in (51) is a valid joint PDF. In addition, directly integrating the joint PDF yields [8]

$$\int_{-\infty}^{+\infty} \text{pdf}(\Delta y_1, \Delta y_2) \cdot d\Delta y_2 = \text{pdf}_1(\Delta y_1)$$
$$\int_{-\infty}^{+\infty} \text{pdf}(\Delta y_1, \Delta y_2) \cdot d\Delta y_1 = \text{pdf}_2(\Delta y_2) \quad (53)$$

implying that $\text{pdf}_1(\Delta y_1)$ and $\text{pdf}_2(\Delta y_2)$ in (51) are the marginal PDFs of Δy_1 and Δy_2, respectively. In particular, let $\text{pdf}_1(\Delta y_1)$ and $\text{pdf}_2(\Delta y_2)$ be normal distributions. In this case, both Δy_1 and Δy_2 are marginally normal; however, their joint PDF in (51) is not a multivariate normal distribution.

In addition, correlated nonnormal random variables cannot be decomposed into independent ones using the PCA [9]. They must be characterized by their joint PDF, thereby making them extremely difficult to handle in statistical analysis. Even Monte Carlo simulation becomes impractical, if not impossible, in such cases, since it is difficult to draw random samples from a general multidimensional joint PDF [22]. In addition, using the correlated nonnormal distributions also increases the difficulty of process characterization, because extracting the multidimensional joint PDF from the process testing data is not trivial. Most IC foundries do not use correlated nonnormal distributions in their process characterizations. Therefore, handling correlated nonnormal process variations is beyond the scope of this paper.

VI. APPLICATIONS OF APEX

APEX can be applied to many statistical analysis and optimization problems (e.g., [3]–[7] and [21]) where estimating the probability distribution of the circuit performance is required. For example, we have applied the APEX to a robust analog design flow in [21]. The robust analog design method proposed in [21] starts from building a quadratic response surface model $f(\text{DV}, \Delta Y)$, where f is the circuit performance of interest, $\text{DV} = [d\nu_1, d\nu_2, \ldots]^{\text{T}}$ contains the design variables that can be controlled by the designer (e.g., bias current, transistor sizes, etc.), and $\Delta Y = [\Delta y_1, \Delta y_2, \ldots, \Delta y_N]^{\text{T}}$ models the process variations. The quadratic model $f(\text{DV}, \Delta Y)$ is typically created by running a number of SPICE simulations. Next, based on the model $f(\text{DV}, \Delta Y)$, a robust optimization is performed to search the optimal design variable values that yield the best circuit performance. During this optimization step, the probability distribution of the performance f is repeatedly estimated with

different design variable values. Since the probability extraction is repeated for many times, its computational cost can be quite expensive, if using Monte Carlo simulation in the loop. In such cases, significant speedup can be achieved if APEX is utilized for probability extraction, as demonstrated in [21].

VII. NUMERICAL EXAMPLES

In this section, we demonstrate the efficacy of the APEX using several circuit examples. All of our experiments are run in MATLAB on a 1-GHz SUN SPARC server. For testing and comparison, both linear and quadratic response surface models are created using the standard model fitting algorithm [12], and these models are utilized for generating performance distributions. The accuracy and cost of the probability extraction are compared against Monte Carlo simulation. A Monte Carlo simulation typically consists of three steps, including: 1) generating a set of random samples; 2) evaluating the performance values at all sampling points; and 3) estimating the probability distribution of the performance. In order to make a fair comparison, the following two issues are considered when implementing the Monte Carlo simulation.

First, instead of running transistor-level simulations, our Monte Carlo simulation evaluates the performance values using the quadratic response surface model in (4). As such, the APEX error can be accurately tested by excluding the response surface modeling error.

Second, various advanced techniques exist to improve the accuracy of Monte Carlo simulation. For example, instead of directly drawing sampling points from a random number generator, they can be created by a controlled sampling scheme, e.g., the Latin hypercube sampling [14]. In addition, a kernel smoothing method [23] can be further utilized to accurately extract the performance distributions (i.e., PDF/CDF functions). Given an accuracy requirement, these advanced techniques can reduce the total number of the required sampling points; however, they involve an additional computational effort for preprocessing and/or postprocessing. Applying these advanced techniques is efficient in those applications where evaluating the circuit performance is expensive, e.g., by running a transistor-level simulator. Otherwise, if the circuit performance can be easily evaluated, e.g., by using a response surface model, the additional computational cost required by these advanced techniques can become a dominant portion of the total cost. For this reason, a simple random number generator is used to generate the sampling points, and a simple bin-based histogram is used to estimate the performance distributions in this paper.

A. Single-Variable Quadratic Model

We first consider a simple example of the single-variable quadratic model in (38). Such a simple example allows us to make a full comparison between the APEX and other traditional probability extraction techniques. By varying the parameter θ in (38), we change the nonlinearity of the function and generate a number of unknown probability distributions with different skewness. As discussed at the beginning of Section IV-C, we assume that the linear term dominates the quadratic term in
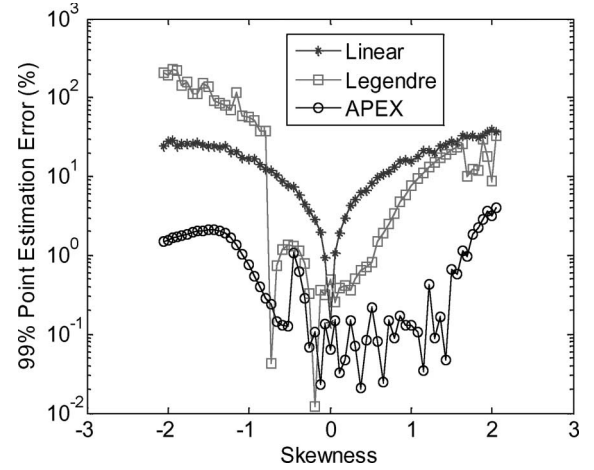

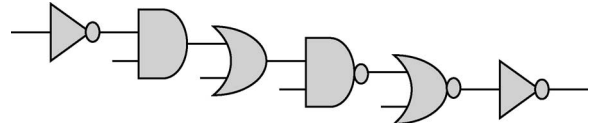Fig. 12. Estimation error of various probability extraction techniques.

Fig. 13. Circuit schematic of the longest path in the ISCAS'89 S27 benchmark circuit.

(38), which is the typical case in most practical applications. Therefore, the value of θ is in the range of $[-1/3, 1/3]$ and the skewness of f is in the range of $[-2.1, 2.1]$.

Three different approaches, i.e., linear regression, Legendre approximation, and APEX, are respectively applied to estimate the unknown distributions. The linear regression approach approximates the quadratic function f in (38) by a best fitted linear model with least squares error, resulting in a normal probability distribution for f. The Legendre approximation is often utilized in traditional mathematics [23]. It expands the unknown PDF by the Legendre polynomials and determines the expansion coefficients based on moment matching.

For each probability extraction technique, the f value at the 99% point of the CDF is estimated and compared with the Monte Carlo simulation result with 10^6 samples. Their relative difference is used as a measure of the estimation error for accuracy comparison, as shown in Fig. 12. The estimation error at the 1% point on CDF has a similar pattern, since the reverse-evaluation method proposed in Section IV-D flips the unknown PDF pdf(f) and the 1% point of the original pdf(f) is the 99% point of the flipped pdf($-f$). As shown in Fig. 12, the proposed APEX approach is the most accurate one in this example. In addition, compared with the error curve in Fig. 6, the nonlinear-companding technique proposed in Section IV-C significantly reduces the estimation error (up to $20\times$ in this example), when the skewness of the unknown distribution is less than -0.5.

B. ISCAS'89 S27

1) Response Surface Modeling: We create a physical implementation for the ISCAS'89 S27 benchmark circuit using the ST CMOS 0.13-μm process. Given a set of fixed gate sizes, the longest path delay in the benchmark circuit (shown in Fig. 13)

TABLE III
COMPUTATIONAL COST FOR MOMENT EVALUATION

Model Order	Direct # of Product Terms	Direct Time (Sec.)	Binomial Time (Sec.)
1	28	1.00×10^{-2}	0.01
3	924	3.02×10^{0}	0.01
5	8008	2.33×10^{2}	0.01
6	18564	1.57×10^{3}	0.01
7	38760	8.43×10^{3}	0.01
8	74613	3.73×10^{4}	0.02
10	—	—	0.02
15	—	—	0.04
20	—	—	0.07

Fig. 14. Estimated ξ values using high-order moments.

Fig. 15. Cumulative distribution function of the performance delay.

is a function of the process variations (e.g., ΔV_{TH}, ΔT_{OX}, ΔL, etc.). Since the circuit only consists of six gates which can be put close to each other in the layout, the interdie variations will dominate the intradie variations and the gate delay will dominate the (local) interconnect delay in this example. Therefore, for simplicity, we only consider the interdie variations for the CMOS transistors in this example. The probability distributions and the correlation information of the interdie transistor variations are obtained from the ST design kit. After PCA analysis, six principal random factors are identified to represent these process variations. We should note, however, that nothing precludes us from including more detailed intradie and/or interconnect variation models in APEX as well.

We approximate the longest path delay as a function of the process variations by a linear regression model and a quadratic response surface (i.e., second-order polynomial) model, respectively. The fitting error is 4.48% for the linear model and 1.10% for the quadratic model ($4\times$ difference). The skewness of the approximated quadratic delay model is equal to 0.28. Therefore, based on Table II, the nonlinear-companding technique proposed in Section IV-C is not required in this example.

It is worth noting that while the linear modeling error in this example is not very large, as IC technologies are scaled to finer feature size, the process variations will become relatively larger, thereby making the nonlinear terms in the quadratic model even more important.

2) Moment Evaluation: Table III compares the computation time required for the direct moment evaluation and our proposed binomial moment evaluation. During the direct moment evaluation, the number of product terms increases exponentially, thereby making the computation task quickly infeasible. The binomial moment evaluation, however, is extremely fast and achieves more than $10^6\times$ speedup over the direct moment evaluation in this example. In addition, we verify that the moment values obtained from both approaches are identical except for numerical errors.

3) PDF/CDF Shifting: As discussed in Section IV-B, PDF/CDF shifting is necessary to make the proposed APEX approach feasible and efficient. A key operation for PDF/CDF shifting is to determine the ξ value based on (35) (also see Fig. 3). We select an error tolerance $\varepsilon = 10^{-3}$ in (35). Fig. 14 shows the estimated ξ values using various high-order moments. From Fig. 14, we find that the high-order moments ($k > 2$) provide a much tighter (i.e., smaller) estimation of ξ.
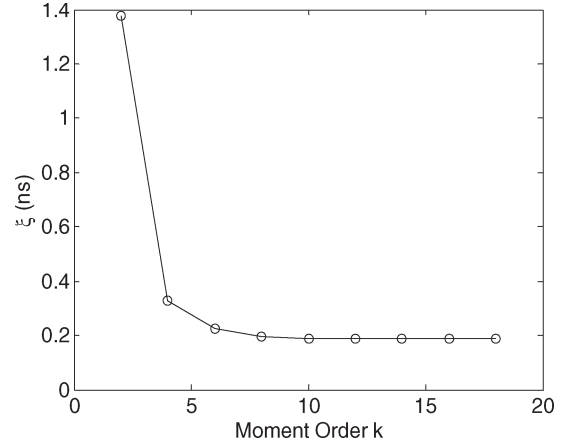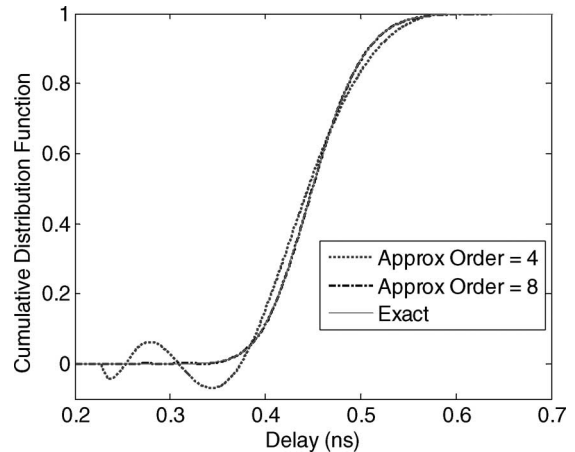
However, after the moment order $k > 10$, further increases in k do not have a significant impact on reducing ξ.

4) PDF/CDF Evaluation: Fig. 15 shows the CDFs for two different approximation orders. In Fig. 15, the "exact" CDF is evaluated by a Monte Carlo simulation with 10^6 samples. Note that the CDF obtained from the high-order approximation (Order = 4) is not accurate and contains numerical oscillations. However, once the approximation order is increased to eight, these oscillations are eliminated and the approximated CDF asymptotically approaches the exact CDF. Similar behavior has been noted in moment matching of the LTI models of interconnect circuits [18], [19].

5) Comparison of Accuracy and Speed: Table IV compares the estimation accuracy and speed for four different probability extraction approaches: linear regression, Legendre approximation, Monte Carlo simulation with 10^4 samples, and APEX. The delay values at several specific points of the CDF are estimated by these probability extraction techniques. The 1% point and the 99% point, for example, denote the best case delay and the worst case delay, respectively. After the CDF is obtained, the best case delay, the worst case delay, and any other specific points on CDF can be easily found using a binary search algorithm. These delay values are compared with the Monte Carlo simulation results with 10^6 samples. Their relative

TABLE IV
ESTIMATION ERROR (COMPARED AGAINST MONTE CARLO WITH 10^6 SAMPLES) AND COMPUTATIONAL COST

	Linear	Legendre	MC (10^4 Samples)	APEX
1% Point	1.43%	0.87%	0.13%	0.04%*
10% Point	4.63%	0.02%	0.22%	0.01%
25% Point	5.76%	0.12%	0.04%	0.03%
50% Point	6.24%	0.04%	0.13%	0.02%
75% Point	5.77%	0.03%	0.26%	0.02%
90% Point	4.53%	0.15%	0.38%	0.03%
99% Point	0.18%	0.77%	0.86%	0.09%
Cost (Sec.)	0.04	0.16	1.56	0.18

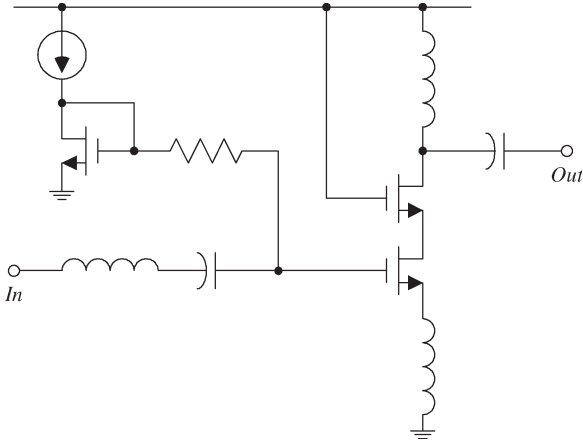* This 1% point error is computed by using reverse evaluation.

Fig. 16. Circuit schematic of a low-noise amplifier (LNA).

difference is used as a measure of the estimation error for accuracy comparison, as shown in Table IV. The computational cost in Table IV is the total computation time for estimating the unknown PDF and CDF.

Note from Table IV that the linear regression approach has the largest error. APEX achieves a speedup of $8.7\times$ over the Monte Carlo simulation with 10^4 samples while still providing better accuracy. In this example, applying a reverse evaluation on pdf$(-f)$ reduces the 1% point estimation error by $4\times$, from 0.20% to 0.04%. This observation demonstrates the efficacy of the reverse-evaluation method proposed in Section IV-D.

C. Low-Noise Amplifier (LNA)

1) Response Surface Modeling: We consider a LNA designed in the IBM BiCMOS 0.25-μm process, as shown in Fig. 16. In this example, the variations on both MOS transistors and passive components (resistor, capacitor, and inductor) are considered. The probability distributions and the correlation information of these variations are provided in the IBM design kit. After PCA analysis, eight principal factors are identified to represent the process variations.

The performance of the LNA is characterized by eight different specifications. Given a fixed circuit design, each circuit performance is a function of the process variations. We approximate these unknown functions by linear regression models and quadratic response surface (i.e., second-order polynomial) models, respectively. Table V shows the response surface modeling error and the model skewness for all these eight perfor-

TABLE V
RESPONSE SURFACE MODELING ERROR AND MODEL SKEWNESS OF THE LNA

Performance	Linear	Quadratic	
	Error	Error	Skewness
F0	1.76%	0.14%	0.23
S11	6.40%	1.32%	0.35
S12	3.44%	0.61%	−0.01
S21	2.94%	0.34%	−0.18
S22	5.56%	3.47%	0.29
NF	2.38%	0.23%	0.13
IIP3	4.49%	0.91%	−0.05
Power	3.79%	0.70%	0.29

mances. In this example, the quadratic modeling error is $7.5\times$ smaller than the linear modeling error on average. In addition, the skewness of all performances is in the range of $[-0.5, 0.5]$. Therefore, based on Table II, the nonlinear-companding technique proposed in Section IV-C is not required in this example.

2) Comparison of Accuracy and Speed: Tables VI and VII compare the estimation accuracy for five different statistical analysis approaches: corner simulation, linear regression, Legendre approximation, Monte Carlo simulation with 10^4 samples, and APEX. The lower bound and upper bound of each LNA performance are estimated by these statistical analysis techniques. These performance values are compared with the Monte Carlo simulation results with 10^6 samples. Their relative difference is used as a measure of the estimation error for accuracy comparison.

The corner simulation approach computes the best case and worst case performances by enumerating all process corners, i.e., combining the extreme values of all process parameters. The corner simulation approach is simple, but it can result in extremely large errors, as shown in Tables VI and VII. The linear regression approach provides more accurate results than the corner simulation, but the errors are expected to increase as IC technologies continue to scale. APEX achieves better accuracy than the Monte Carlo simulation with 10^4 samples and is $10\times$ faster, as shown in Table VIII.

D. Operational Amplifier

1) Response Surface Modeling: Fig. 17 shows a two-stage folded-cascode operational amplifier designed in the IBM BiCMOS 0.25-μm process. In this example, 49 principal factors are extracted by the PCA to represent the process variations, including both the interdie variations and the device mismatches. Modeling the mismatches is extremely important for this operational amplifier, since they significantly impact many circuit performances, e.g., input offset voltage. The probability distributions and the correlation information of all random variations are obtained from the IBM design kit.

The performance of the Op Amp is characterized by eight different specifications. Given a fixed circuit design, each circuit performance is a function of the process variations. We approximate these unknown functions by linear regression models and quadratic response surface (i.e., second-order polynomial) models, respectively. Table IX shows the response surface modeling error and the model skewness for all these

TABLE VI
ESTIMATION ERROR FOR THE LOWER BOUND OF THE LNA PERFORMANCE (1% POINT ON CDF)

Performance	Corner	Linear	Legendre	MC (10^4 Samples)	APEX
F0	15.80%	1.11%	0.20%	0.78%	0.06%
S11	45.44%	5.78%	2.94%	0.15%	0.09%
S12	38.87%	3.88%	0.39%	0.38%	0.14%
S21	60.25%	2.91%	0.37%	0.23%	0.17%
S22	23.08%	1.01%	1.11%	0.05%	0.07%
NF	51.91%	3.70%	0.34%	0.01%	0.06%
IIP3	54.64%	5.02%	0.29%	0.30%	0.33%
Power	16.56%	0.01%	0.92%	0.07%	0.09%

TABLE VII
ESTIMATION ERROR FOR THE UPPER BOUND OF THE LNA PERFORMANCE (99% POINT ON CDF)

Performance	Corner	Linear	Legendre	MC (10^4 Samples)	APEX
F0	20.12%	1.10%	0.55%	0.35%	0.05%
S11	51.53%	1.40%	3.28%	0.80%	0.08%
S12	44.64%	1.16%	0.27%	2.41%	0.28%
S21	51.63%	4.69%	0.01%	1.22%	0.19%
S22	36.01%	5.61%	0.84%	0.80%	0.19%
NF	72.82%	3.52%	0.37%	1.36%	0.12%
IIP3	59.66%	5.93%	0.43%	0.04%	0.26%
Power	42.53%	1.24%	0.93%	0.59%	0.02%

TABLE VIII
COMPUTATIONAL COST FOR THE PROBABILITY EXTRACTION OF THE LNA (SEC.)

Performance	Linear	Legendre	MC (10^4 Samples)	APEX
F0	0.04	0.14	1.68	0.16
S11	0.03	0.14	1.58	0.14
S12	0.04	0.13	1.58	0.17
S21	0.04	0.14	1.57	0.19
S22	0.04	0.13	1.60	0.19
NF	0.03	0.14	1.58	0.15
IIP3	0.03	0.14	1.57	0.14
Power	0.04	0.14	1.55	0.15
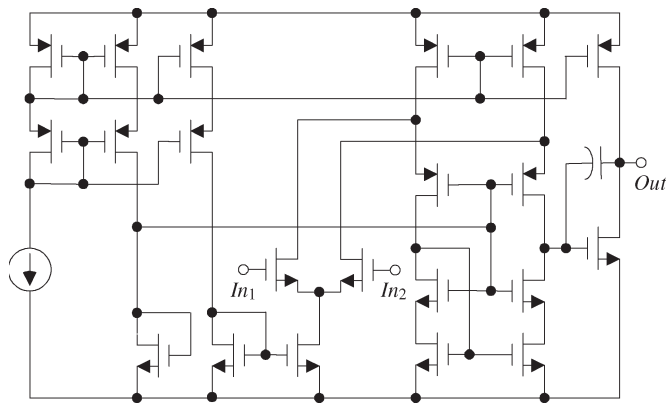
Fig. 17. Circuit schematic of an operational amplifier.

TABLE IX
RESPONSE SURFACE MODELING ERROR AND MODEL SKEWNESS OF THE OP AMP

Performance	Linear	Quadratic	
	Error	Error	Skewness
Gain	3.92%	1.57%	−1.30
Offset	21.80%	7.49%	2.80
UGF	1.14%	0.45%	0.25
Gain Margin	0.96%	0.52%	−0.16
Phase Margin	1.11%	0.41%	−0.26
Slew Rate (+)	0.82%	0.66%	0.06
Slew Rate (−)	1.27%	0.44%	0.25
Power	1.00%	0.64%	0.04

eight performances. In this example, the skewness of two performances, i.e., gain and offset, is not in the range of [−0.5, 0.5]. Therefore, based on Table II, the nonlinear-companding technique proposed in Section IV-C should be applied to these two performances to improve the approximation accuracy.

2) Comparison of Accuracy and Speed: Tables X and XI compare the estimation accuracy for four different statistical analysis approaches: linear regression, Legendre approximation, Monte Carlo simulation with 10^4 samples, and APEX. The lower bound and upper bound of each Op-Amp performance are estimated by these statistical analysis techniques. These perfor-

mance values are compared with the Monte Carlo simulation results with 10^6 samples. Their relative difference is used as a measure of the estimation error for accuracy comparison.

As shown in Tables X and XI, the linear regression and the Legendre approximation yield large estimation errors, especially when the circuit performance has a large skewness, e.g., gain and offset. Because of their strong nonlinearities, the probability distributions of gain and offset are extremely difficult to approximate. Most traditional approaches in Tables X and XI fail to estimate the upper bounds and the lower bounds of these two performances accurately. By using the proposed nonlinear-companding technique, APEX achieves a significant error reduction. Without the nonlinear companding, the APEX

TABLE X
ESTIMATION ERROR FOR THE LOWER BOUND OF THE OP-AMP PERFORMANCE (1% POINT ON CDF)

Performance	Linear	Legendre	MC (10^4 Samples)	APEX
Gain	22.71%	22.02%	2.39%	1.45%
Offset	11.50%	222.66%	1.21%	0.58%
UGF	3.78%	0.39%	0.34%	0.03%
Gain Margin	2.72%	0.37%	0.36%	0.08%
Phase Margin	4.41%	0.40%	1.36%	0.13%
Slew Rate (+)	0.81%	0.35%	1.07%	0.11%
Slew Rate (−)	3.83%	0.24%	0.60%	0.13%
Power	0.69%	0.35%	1.12%	0.11%

TABLE XI
ESTIMATION ERROR FOR THE UPPER BOUND OF THE OP-AMP PERFORMANCE (99% POINT ON CDF)

Performance	Linear	Legendre	MC (10^4 Samples)	APEX
Gain	10.39%	81.72%	0.03%	0.32%
Offset	74.68%	158.53%	5.82%	3.20%
UGF	4.30%	0.33%	0.76%	0.18%
Gain Margin	2.46%	0.20%	1.02%	0.04%
Phase Margin	3.79%	0.52%	0.06%	0.02%
Slew Rate (+)	0.97%	0.34%	1.45%	0.07%
Slew Rate (−)	4.31%	0.27%	0.58%	0.24%
Power	0.65%	0.41%	0.46%	0.00%

TABLE XII
COMPUTATIONAL COST FOR THE PROBABILITY EXTRACTION OF THE OP AMP (SEC.)

Performance	Linear	Legendre	MC (10^4 Samples)	APEX
Gain	0.04	0.40	3.28	0.42
Offset	0.02	0.38	3.20	0.44
UGF	0.02	0.41	3.13	0.42
Gain Margin	0.03	0.40	3.13	0.43
Phase Margin	0.03	0.40	3.10	0.41
Slew Rate (+)	0.02	0.36	3.07	0.41
Slew Rate (−)	0.04	0.40	3.15	0.42
Power	0.02	0.37	3.14	0.44

error for the 99% point of the performance gain is 5.37% (16× larger) and the APEX error for the 1% point of the performance offset is 23.47% (40× larger). For most circuit performances in this example, APEX achieves better accuracy than the Monte Carlo simulation with 10^4 samples and is 7.2× faster, as shown in Table XII.

VIII. CONCLUSION

As IC technologies reach nanoscale, process variations are becoming relatively large and nonlinear (e.g., quadratic) response surface models might be required to accurately characterize the large-scale variations. In this paper, we propose an APEX method for estimating the nonnormal random distribution resulting from the nonlinear response surface modeling. Four novel algorithms, i.e., binomial moment evaluation, CDF/PDF shifting, nonlinear companding, and reverse evaluation, are proposed to reduce the computational cost and improve the estimation accuracy. As is demonstrated by the numerical examples, applying the APEX results in better accuracy than a Monte Carlo simulation with 10^4 samples and achieves up to 10× more efficiency. APEX can be incorporated into a parametric yield optimization loop or a statistical timing analysis environment for efficient probability extraction and worst case analysis. For example, the efficacy of applying APEX to robust analog design is further discussed in [21].

APPENDIX
PROOF OF THEOREM 2

Proof: The model diagonalization procedure proposed in Section IV-A demonstrates that any quadratic response surface model in (4) can be converted to an equivalent model in (25) without any cross-product terms. Therefore, from now on, we will assume that there is no cross-product term in the quadratic function. The quadratic response surface model in (25) can be rewritten as

$$f(\Delta Z) = \sum_{i=1}^{N} \left(\sigma_i \cdot \Delta z_i^2 + q_i \cdot \Delta z_i \right) + C$$
$$= \sum_{i=1}^{N} \left[\sigma_i \cdot \left(\Delta z_i + \frac{q_i}{2\sigma_i} \right)^2 \right] + \left(-\sum_{i=1}^{N} \frac{q_i^2}{4\sigma_i} + C \right).$$
(54)

Therefore, we have

$$|f(\Delta Z)| \leq \sum_{i=1}^{N} \left[|\sigma_i| \cdot \left(|\Delta z_i| + \left| \frac{q_i}{2\sigma_i} \right| \right)^2 \right] + \left| -\sum_{i=1}^{N} \frac{q_i^2}{4\sigma_i} + C \right|.$$
(55)

Let

$$\alpha = \max_i |\sigma_i| \quad \beta_i = \frac{q_i}{2\sigma_i} \quad \gamma = \sqrt{\frac{1}{\alpha N} \cdot \left| -\sum_{i=1}^{N} \frac{q_i^2}{4\sigma_i} + C \right|} \quad (56)$$

and substituting (56) into (55) yields

$$|f(\Delta Z)| \le |\alpha| \cdot \left[\sum_{i=1}^{N} (|\Delta z_i| + |\beta_i|)^2 + N\gamma^2 \right]$$

$$\le |\alpha| \cdot \sum_{i=1}^{N} (|\Delta z_i| + |\beta_i| + |\gamma|)^2$$

$$\le |\alpha| \cdot \left[\sum_{i=1}^{N} (|\Delta z_i| + |\beta_i| + |\gamma|) \right]^2. \quad (57)$$

The high-order moments of f satisfy

$$m_{2k} = E(f^{2k}) \le E\left[|\alpha|^{2k} \cdot \left(\sum_{i=1}^{N} |\Delta z_i| + \sum_{i=1}^{N} |\beta_i| + \sum_{i=1}^{N} |\gamma| \right)^{4k} \right]$$

$$= |\alpha|^{2k} \cdot \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(\sum_{i=1}^{N} |\Delta z_i| + \sum_{i=1}^{N} |\beta_i| + \sum_{i=1}^{N} |\gamma| \right)^{4k}$$

$$\cdot \, \text{pdf}(\Delta z_1, \ldots, \Delta z_N) \cdot d\Delta z_1 \cdots d\Delta z_N. \quad (58)$$

Since the random variables ΔZ are independent, and satisfy the standard normal distribution $N(0,1)$ (see Theorem 3), the integration in (58) is over a symmetric function. Using this symmetric property yields

$$m_{2k} \le 2^N \cdot |\alpha|^{2k} \cdot \int_{0}^{+\infty} \cdots \int_{0}^{+\infty} \left(\sum_{i=1}^{N} \Delta z_i + \sum_{i=1}^{N} |\beta_i| + \sum_{i=1}^{N} |\gamma| \right)^{4k}$$

$$\cdot \, \text{pdf}(\Delta z_1, \ldots, \Delta z_N) \cdot d\Delta z_1 \cdots d\Delta z_N. \quad (59)$$

Compared with (58), the absolute operator is removed for $\{\Delta z_i, i = 1, 2, \ldots, N\}$ in (59), because the function in (59) is only integrated over positive $\{\Delta z_i, i = 1, 2, \ldots, N\}$. The function

$$\left(\sum_{i=1}^{N} \Delta z_i + \sum_{i=1}^{N} |\beta_i| + \sum_{i=1}^{N} |\gamma| \right)^{4k} \cdot \text{pdf}(\Delta z_1, \ldots, \Delta z_N) \quad (60)$$

is nonnegative. Therefore, expanding the integration interval in (59) yields

$$m_{2k} \le 2^N \cdot |\alpha|^{2k} \cdot \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(\sum_{i=1}^{N} \Delta z_i + \sum_{i=1}^{N} |\beta_i| + \sum_{i=1}^{N} |\gamma| \right)^{4k}$$

$$\cdot \, \text{pdf}(\Delta z_1, \ldots, \Delta z_N) \cdot d\Delta z_1 \cdots d\Delta z_N$$

$$= 2^N \cdot |\alpha|^{2k} \cdot E\left[\left(\sum_{i=1}^{N} \Delta z_i + \sum_{i=1}^{N} |\beta_i| + \sum_{i=1}^{N} |\gamma| \right)^{4k} \right]$$

$$= 2^N \cdot |\alpha|^{2k} \cdot E\left[\left(\sqrt{N} \cdot \Delta z + \sqrt{N} \cdot \beta \right)^{4k} \right] \quad (61)$$

where

$$\Delta z = \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \Delta z_i \quad \text{and} \quad \beta = \frac{1}{\sqrt{N}} \cdot \left(\sum_{i=1}^{N} |\beta_i| + \sum_{i=1}^{N} |\gamma| \right). \quad (62)$$

It is easy to verify that the random variable Δz in (62) also satisfies the standard normal distribution $N(0,1)$, since it is the linear combination of N independent random variables $\{\Delta z_i, i = 1, 2, \ldots, N\}$ with the standard normal distribution $N(0,1)$. Therefore, (61) can be rewritten as

$$m_{2k} \le 2^N \cdot |\alpha|^{2k} \cdot N^{2k} \cdot E\left[(\Delta z + \beta)^{4k} \right]$$

$$= 2^N \cdot |\alpha|^{2k} \cdot N^{2k} \cdot \left[\begin{array}{l} \int_{-\infty}^{-\beta} (\Delta z + \beta)^{4k} \cdot \text{pdf}(\Delta z) \cdot d\Delta z \\ + \int_{-\beta}^{+\beta} (\Delta z + \beta)^{4k} \cdot \text{pdf}(\Delta z) \cdot d\Delta z \\ + \int_{+\beta}^{+\infty} (\Delta z + \beta)^{4k} \cdot \text{pdf}(\Delta z) \cdot d\Delta z \end{array} \right] \quad (63)$$

where β is a positive number defined in (62) and $\text{pdf}(\Delta z)$ is the PDF of the random variable Δz. Note that the three integrations in (63) are over different intervals. It is easy to verify the following inequalities:

$$\begin{cases} (\Delta z + \beta)^{4k} \le (2\Delta z)^{4k} & (\text{if } |\Delta z| \ge \beta) \\ (\Delta z + \beta)^{4k} \le (2\beta)^{4k} & (\text{if } |\Delta z| \le \beta). \end{cases} \quad (64)$$

Since the PDF $\text{pdf}(\Delta z)$ is nonnegative, (64) yields

$$\begin{cases} \int_{-\infty}^{-\beta} (\Delta z + \beta)^{4k} \cdot \text{pdf}(\Delta z) \cdot d\Delta z \\ \quad \le \int_{-\infty}^{-\beta} (2\Delta z)^{4k} \cdot \text{pdf}(\Delta z) \cdot d\Delta z \\ \int_{-\beta}^{+\beta} (\Delta z + \beta)^{4k} \cdot \text{pdf}(\Delta z) \cdot d\Delta z \\ \quad \le \int_{-\beta}^{+\beta} (2\beta)^{4k} \cdot \text{pdf}(\Delta z) \cdot d\Delta z \\ \int_{+\beta}^{+\infty} (\Delta z + \beta)^{4k} \cdot \text{pdf}(\Delta z) \cdot d\Delta z \\ \quad \le \int_{+\beta}^{+\infty} (2\Delta z)^{4k} \cdot \text{pdf}(\Delta z) \cdot d\Delta z. \end{cases} \quad (65)$$

Substituting (65) into (63) yields

$$m_{2k} \leq 2^N \cdot |\alpha|^{2k} \cdot N^{2k} \cdot \begin{bmatrix} \int_{-\infty}^{-\beta} (2\Delta z)^{4k} \cdot \mathrm{pdf}(\Delta z) \cdot d\Delta z \\ + \int_{-\beta}^{+\beta} (2\beta)^{4k} \cdot \mathrm{pdf}(\Delta z) \cdot d\Delta z \\ + \int_{+\beta}^{+\infty} (2\Delta z)^{4k} \cdot \mathrm{pdf}(\Delta z) \cdot d\Delta z \end{bmatrix}.$$

(66)

$$m_{2k} \leq 2^N \cdot |\alpha|^{2k} \cdot N^{2k} \cdot \begin{bmatrix} \int_{-\infty}^{+\infty} (2\Delta z)^{4k} \cdot \mathrm{pdf}(\Delta z) \cdot d\Delta z \\ + \int_{-\infty}^{+\infty} (2\beta)^{4k} \cdot \mathrm{pdf}(\Delta z) \cdot d\Delta z \end{bmatrix}$$

$$= 2^{N+4k} \cdot |\alpha|^{2k} \cdot N^{2k} \cdot E\left[(\Delta z)^{4k}\right]$$
$$+ 2^{N+4k} \cdot |\alpha|^{2k} \cdot N^{2k} \cdot \beta^{4k}.$$

(67)

Recall that the random variable Δz has the standard normal distribution $N(0,1)$ and their high-order moments can be computed using the closed-form expression (22). Substituting (22) into (67) yields

$$m_{2k} \leq 2^{N+4k} \cdot |\alpha|^{2k} \cdot N^{2k} \cdot [1 \cdot 3 \cdot \cdots \cdot (4k-1)]$$
$$+ 2^{N+4k} \cdot |\alpha|^{2k} \cdot N^{2k} \cdot \beta^{4k}$$
$$\leq 2^{N+4k} \cdot |\alpha|^{2k} \cdot N^{2k} \cdot (4k)^{2k}$$
$$+ 2^{N+4k} \cdot |\alpha|^{2k} \cdot N^{2k} \cdot \beta^{4k}.$$

(68)

Therefore, we have

$$\sum_{k=1}^{+\infty} (m_{2k})^{\frac{-1}{2k}} \geq \frac{1}{4 \cdot |\alpha| \cdot N} \cdot \sum_{k=1}^{+\infty} \left[2^N \cdot (4k)^{2k} + 2^N \cdot \beta^{4k}\right]^{\frac{-1}{2k}}$$

$$= \frac{1}{4 \cdot |\alpha| \cdot N} \cdot \sum_{k=1}^{K} \left[2^N \cdot (4k)^{2k} + 2^N \cdot \beta^{4k}\right]^{\frac{-1}{2k}}$$

$$+ \frac{1}{4 \cdot |\alpha| \cdot N} \cdot \sum_{k=K}^{+\infty} \left[2^N \cdot (4k)^{2k} + 2^N \cdot \beta^{4k}\right]^{\frac{-1}{2k}}.$$

(69)

The value of K in (69) can be selected to be sufficiently large such that

$$4k \geq \beta^2 \quad (\forall k \geq K).$$

(70)

Combining (69) and (70) yields

$$\sum_{k=1}^{+\infty} (m_{2k})^{\frac{-1}{2k}} \geq \frac{1}{4 \cdot |\alpha| \cdot N} \cdot \sum_{k=1}^{K} \left[2^N \cdot (4k)^{2k} + 2^N \cdot \beta^{4k}\right]^{\frac{-1}{2k}}$$

$$+ \frac{1}{4 \cdot |\alpha| \cdot N} \cdot \sum_{k=K}^{+\infty} \left[2^{N+1} \cdot (4k)^{2k}\right]^{\frac{-1}{2k}}$$

$$\geq \frac{1}{4 \cdot |\alpha| \cdot N} \cdot \sum_{k=1}^{K} \left[2^N \cdot (4k)^{2k} + 2^N \cdot \beta^{4k}\right]^{\frac{-1}{2k}}$$

$$+ \frac{1}{4 \cdot |\alpha| \cdot N} \cdot \sum_{k=K}^{+\infty} \left(\frac{1}{\sqrt[2k]{2^{N+1}}} \cdot \frac{1}{4k}\right)$$

$$\geq \frac{1}{4 \cdot |\alpha| \cdot N} \cdot \sum_{k=1}^{K} \left[2^N \cdot (4k)^{2k} + 2^N \cdot \beta^{4k}\right]^{\frac{-1}{2k}}$$

$$+ \frac{1}{16 \cdot |\alpha| \cdot N \cdot 2^{N+1}} \cdot \sum_{k=K}^{+\infty} \frac{1}{k}$$

$$= \infty.$$

(71)

Equation (71) demonstrates that the condition in (19) is satisfied. Therefore, based on Theorem 1, the probability distribution of f can be uniquely determined by its moments $\{m_k, k = 1, 2, \ldots\}$, which proves the asymptotic convergence of APEX. ∎

REFERENCES

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

[2] C. Michael and M. Ismail, "Statistical modeling of device mismatch for analog MOS integrated circuits," *IEEE J. Solid-State Circuits*, vol. 27, no. 2, pp. 154–166, Feb. 1992.

[3] Z. Wang and S. Director, "An efficient yield optimization method using a two step linear approximation of circuit performance," in *Proc. IEEE Eur. Des. and Test Conf.*, 1994, pp. 567–571.

[4] A. Dharchoudhury and S. Kang, "Worse-case analysis and optimization of VLSI circuit performance," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 14, no. 4, pp. 481–492, Apr. 1995.

[5] E. Felt, S. Zanella, C. Guardiani, and A. Sangiovanni-Vincentelli, "Hierarchical statistical characterization of mixed-signal circuits using behavioral modeling," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, 1996, pp. 374–380.

[6] J. Swidzinski, M. Styblinski, and G. Xu, "Statistical behavioral modeling of integrated circuits," in *Proc. IEEE Int. Symp. Circuits and Syst.*, 1998, vol. 6, pp. 98–101.

[7] A. Graupner, W. Schwarz, and R. Schüffny, "Statistical analysis of analog structures through variance calculation," *IEEE Trans. Circuits Syst. I, Fundam Theory Appl.*, vol. 49, no. 8, pp. 1071–1078, Aug. 2002.

[8] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 2001.

[9] G. Seber, *Multivariate Observations*. Hoboken, NJ: Wiley, 1984.

[10] G. Golub and C. Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[11] G. Box and N. Draper, *Empirical Model-Building and Response Surfaces*. Hoboken, NJ: Wiley, 1987.

[12] R. Myers and D. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York: Wiley-Interscience, 2002.

[13] X. Li, J. Le, L. Pileggi, and A. Strojwas, "Projection-based performance modeling for inter/intra-die variations," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, 2005, pp. 721–727.

[14] M. Mckay, R. Beckman, and W. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, May 1979.

[15] D. Montgomery, *Design and Analysis of Experiments*. Hoboken, NJ: Wiley, 2004.

[16] N. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*. Edinburgh, U.K.: Oliver and Boyd, 1965.

[17] B. Simon, "The classical moment problem as a self-adjoint finite difference operator," *Adv. Math.*, vol. 137, no. 1, pp. 82–203, Jul. 1998.

[18] L. Pillage and R. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 9, no. 4, pp. 352–366, Apr. 1990.

[19] M. Celik, L. Pileggi, and A. Odabasioglu, *IC Interconnect Analysis*. Norwell, MA: Kluwer, 2002.

[20] M. Bosley and F. Lees, "A survey of simple transfer-function derivations from high-order state-variable models," *Automatica*, vol. 8, no. 6, pp. 765–775, Nov. 1972.

[21] X. Li, P. Gopalakrishnan, Y. Xu, and L. Pileggi, "Robust analog/RF circuit design with projection-based posynomial modeling," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, 2004, pp. 855–862.

[22] C. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 2005.

[23] B. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman and Hall, 1986.

Xin Li (S'01–M'06) received the B.S. and M.S. degrees in electronics engineering from Fudan University, Shanghai, China, in 1998 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 2005.

He was a summer Intern with Extreme DA, Palo Alto, CA, in 2004. Since the summer of 2005, he has been a Systems Scientist with the Department of Electrical and Computer Engineering, Carnegie Mellon University. His current research interests include modeling, simulation, and synthesis for analog/RF and digital systems.

Dr. Li received the IEEE/ACM William J. McCalla International Conference on Computer Aided Design Best Paper Award in 2004.
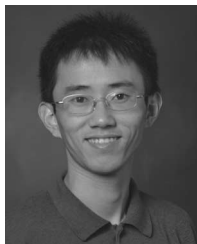
Jiayong Le (S'03) received the B.S. and M.S. degrees in electronics engineering from Shanghai Jiaotong University, Shanghai, China, in 1999 and 2001, respectively. He is currently working toward the Ph.D. degree in electrical and computer engineering at Carnegie Mellon University, Pittsburgh, PA.

In 2001, he was a Student Intern with Intel Shanghai Research Lab, Shanghai. Since the spring of 2004, he has been a member of the technical staff with Extreme DA, Palo Alto, CA. His current research interests include statistical timing analysis and optimization of digital systems.

Mr. Le received the IEEE/ACM William J. McCalla ICCAD Best Paper Award in 2004.

Padmini Gopalakrishnan (S'98) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1997, and the M.S. degree in electrical and computer engineering from the University of Texas, Austin, in 1999. She is currently working toward the Ph.D. degree in electrical and computer engineering at Carnegie Mellon University, Pittsburgh, PA.

She spent three years with the R&D Team of Monterey Design Systems, Sunnyvale, CA, from 1999 to 2002. Her research interests span various aspects of VLSI design methodologies and CAD algorithms, including applications of graph theory and optimization theory.

Ms. Gopalakrishnan received the IEEE/ACM William J. McCalla International Conference on Computer Aided Design Best Paper Award in 2004.

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 electrical and computer engineering and the Director of the Center for Silicon System Implementation at Carnegie Mellon University. From 1984 through 1986, he was with Westinghouse Research and Development, where he was recognized with the corporation's highest engineering achievement award in 1986. From 1989 through 1995, he was a Faculty Member with the University of Texas, Austin. In January of 1996, he joined the faculty at Carnegie Mellon University. He has consulted for several electronic design automation (EDA) and semiconductor companies. He is a 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 holds 14 U.S. patents. His research interests include various aspects of digital and analog design and EDA.

Dr. 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 received the Best CAD Transactions Paper Awards in 1991 and 1999. He received a Best Paper Award from the Design Automation Conference in 2003, and a Best Paper Award from the 2004 ICCAD. He received a Presidential Young Investigator Award from the National Science Foundation in 1991. In 1991 also, and again in 1999, he received the Semiconductor Research Corporation (SRC) Technical Excellence Award. In 1993, he received an Invention Award from the SRC and, subsequently, a U.S. patent for the RICE simulation tool. 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.