

Asymptotic Probability Extraction for Non-Normal Distributions of Circuit Performance

Xin Li, Jiayong Le, Padmini Gopalakrishnan and Lawrence T. Pileggi
Dept. of ECE, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
{xinli, jiayongl, pgopalak, pileggi}@ece.cmu.edu

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 (e.g. quadratic) response surface models can be utilized to capture larger scale process variations; however, such models result in non-Normal distributions for circuit performance which are difficult to capture since the distribution model is unknown. In this paper we propose an asymptotic probability extraction method, APEX, for estimating the unknown random distribution when using nonlinear response surface modeling. APEX first uses a novel binomial moment evaluation to efficiently compute the high order moments of the unknown distribution, and then applies moment matching to approximate the characteristic function of the random circuit performance by an efficient rational function. A simple statistical timing example and an analog circuit example demonstrate that APEX can provide better accuracy than Monte Carlo simulation with 10^4 samples and achieve orders of magnitude more efficiency. We also show the error incurred by the popular Normal modeling assumption using standard IC technologies.

1. Introduction

As IC technologies are scaled to the deep sub-micron region, process variations are becoming critical and significantly impact the overall performance of a circuit. Table 1 shows some typical process parameters and their 3σ variations as technologies are scaled from 0.25 μm to 70 nm. These large-scale variations introduce uncertainties in circuit behavior, thereby making IC design increasingly difficult. Low product yield or unnecessary over-design cannot be avoided if these process variations are not accurately modeled and analyzed within the IC design flow.

Table 1. Technology parameters and 3σ variations [1]

Leff (nm)	Tox (nm)	Vth (mV)	W (μm)	H (μm)
250 \pm 80	5.0 \pm 0.40	500 \pm 50	0.80 \pm 0.20	1.2 \pm 0.3
180 \pm 60	4.5 \pm 0.36	450 \pm 45	0.65 \pm 0.17	1.0 \pm 0.3
130 \pm 45	4.0 \pm 0.39	400 \pm 40	0.50 \pm 0.14	0.9 \pm 0.27
100 \pm 40	3.5 \pm 0.42	350 \pm 40	0.40 \pm 0.12	0.8 \pm 0.27
70 \pm 33	3.0 \pm 0.48	300 \pm 40	0.30 \pm 0.10	0.7 \pm 0.25

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. The author in [1] applies linear regression to approximate a given circuit performance f (e.g.

delay, gain, etc.) as a function of the process variations (e.g. V_{th} , T_{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 still has a Normal distribution [8].

The linear regression model is efficient and accurate when the process variations are sufficiently small. However, the large-scale variations in the deep sub-micron technologies, which can reach $\pm 35\%$ as shown in Table 1, suggest applying high order regression models in order to guarantee high approximation accuracy [4]-[7]. Using a high 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, as is the case for the linear regression model. The authors in [3]-[5] utilize Monte Carlo simulation to evaluate the probability distribution of f , which is computationally expensive. Note that the computation cost for this probability extraction is crucial, especially when the extraction procedure is an inner loop within the optimization flow.

In this paper, we propose a novel Asymptotic Probability EXtraction approach, APEX, for estimating the unknown PDF/CDF functions using nonlinear response surface modeling. Given a circuit performance f (e.g. a digital circuit path delay or the performance parameter of an analog/RF circuit), the response surface modeling approximates f as a polynomial function of the process parameters (e.g. V_{th} , T_{ox} , etc.). Since the process parameters are modeled as random variables, the circuit performance f is a function of these random variables, which is also a random variable. APEX applies moment matching to approximate the characteristic function of f (i.e. the Fourier transform of the probability density function [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 probability distribution function (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. The resulting probability distribution function can then 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 three important new contributions which significantly reduce the computation cost and improve the approximation accuracy for this particular application. Firstly, a key operation required by APEX is to compute the high order moments, which is extremely expensive when using traditional techniques. In APEX, we propose a binomial evaluation scheme to recursively compute the high order moments for a given quadratic response surface model. The binomial moment evaluation is derived from statistical independence theory and principal component analysis (PCA) methods. It can achieve more than $10^6\times$ speedup compared with direct moment evaluation.

Secondly, APEX approximates the unknown probability distribution function by the impulse response of an LTI system. Directly applying such an approximation to any circuit performance with negative values is infeasible, since it results in

an LTI system that is non-causal. To overcome this difficulty, APEX applies a generalized Chebyshev inequality for PDF/CDF shifting.

Lastly, 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. 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 values of CDF), and least accurate for high frequency components (corresponding to the initial values of 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 the paper is organized as follows. In Section 2 we review the background on response surface modeling. Then we propose our APEX approach in Section 3. We discuss several implementation issues, including the high order moment evaluation, PDF/CDF shifting and reverse PDF/CDF evaluation, in Section 4. The efficacy of APEX is demonstrated by several circuit examples in Section 5. Finally, we conclude in Section 6.

2. Background

Given a circuit topology, the circuit performance (e.g. gain, delay) is a function of the design parameters (e.g. bias current, transistor sizes) and the process parameters (e.g. V_{TH} , T_{OX}). The design parameters are optimized and fixed during the design process; however, the process parameters must be modeled as random variables to account for any uncertain variations. Given a set of fixed design parameters, the circuit performance f can be approximated by a linear regression model [1]:

$$f(X) = f(\bar{X}) + \hat{B}^T \cdot \Delta X \quad (1)$$

where $X = [x_1, x_2, \dots, x_N]^T$ denotes the process parameters, \bar{X} is the mean value of X , $\Delta X = X - \bar{X}$ represents the process variations, $\hat{B} \in R^N$ stands for the linear model coefficients and N is the total number of these random variations.

The process variations in (1), i.e. ΔX , are often approximated by zero-mean Normal distributions^{*}. In addition, correlated process variations can be expressed in terms of independent factors using principal component analysis (PCA) [9]. Given a set of normally distributed random variables ΔX and their symmetric, positive semi-definite correlation matrix R , PCA decomposes R as:

$$R = V \cdot \Sigma \cdot V^T \quad (2)$$

where $\Sigma = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ contains the eigenvalues of R , and $V = [V_1, V_2, \dots, V_N]$ contains the corresponding eigenvectors that are orthonormal, i.e. $V^T V = I$ (I is the identity matrix). Based on Σ and V , PCA defines a set of new random variables:

$$\Delta Y = \Sigma^{-0.5} \cdot V^T \cdot \Delta X \quad (3)$$

It is easy to verify that the random variables ΔY are independent and satisfy the Normal distribution $N(0,1)$ (i.e. zero mean and unit standard deviation).

The factors extracted from PCA can be interpreted as coordinate rotations of the space defined by the original random variables. In addition, if the magnitude of the eigenvalues λ_i decreases quickly, it is possible to use a small number of principal components to approximate the original N -dimensional space.

^{*} If a process parameter ζ satisfies a log-Normal distribution, it can also be transformed to a Normal distribution by taking the logarithmic operator, i.e. $\ln(\zeta)$ is normally distributed.

More details on PCA can be found in [9].

Substituting (3) into (1) yields:

$$f(\Delta Y) = C + B^T \cdot \Delta Y \quad (4)$$

where $C = f(\bar{X})$ and $B = \Sigma^{0.5} \cdot V^T \cdot \hat{B}$. The linear regression model in (4) is accurate when the process variations are small. However, the large-scale variations that are expected for nanoscale technologies suggest that applying quadratic response surface models might be required to provide sufficient accuracy:

$$f(\Delta Y) = C + B^T \cdot \Delta Y + \Delta Y^T \cdot A \cdot \Delta Y \quad (5)$$

In (5), $C \in R$ is the constant term, $B \in R^N$ represents the linear coefficients and $A \in R^{N \times N}$ denotes the quadratic coefficients. Without loss of generality, we assume that A is symmetric in this paper, since any asymmetric quadratic form can be easily converted to an equivalent symmetric form [10].

3. Asymptotic Probability Extraction

Given the quadratic response surface model in (5), the objective of probability extraction is to estimate the unknown probability density function $pdf(f)$ and cumulative distribution function $cdf(f)$ for performance f . Instead of running expensive Monte Carlo simulations, APEX tries to find an M-th order LTI system H whose impulse response $h(t)$ and step response $s(t)$ are the optimal approximations for the $pdf(f)$ and $cdf(f)$ respectively. Here, the variable t in $h(t)$ and $s(t)$ corresponds to the variable f in $pdf(f)$ and $cdf(f)$. The optimal approximation is determined by matching the first 2M moments between $h(t)$ and $pdf(f)$ for an M-th order approximation. We first describe the mathematical formulation of APEX in Section 3.1. Then, in Section 3.2 we will link APEX to traditional probability theory and explain why it can be used to efficiently approximate PDF/CDF functions.

3.1 Mathematical Formulation

We define the *time moments* [11] for a given circuit performance f whose probability density function is $pdf(f)$ as follows:

$$m_k = \frac{(-1)^k}{k!} \int_{-\infty}^{+\infty} f^k \cdot pdf(f) \cdot df \quad (6)$$

In (6), 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, time moments can be defined for an LTI system H [11]. Given an M-th order LTI system whose transfer function and impulse response are:

$$H(s) = \sum_{i=1}^M \frac{a_i}{s - b_i} \quad \text{and} \quad h(t) = \begin{cases} \sum_{i=1}^M a_i e^{b_i t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \quad (7)$$

The time moments of H are expressed as [11]:

$$m_k = \frac{(-1)^k}{k!} \int_{-\infty}^{+\infty} t^k \cdot h(t) \cdot dt = - \sum_{i=1}^M \frac{a_i}{b_i^{k+1}} \quad (8)$$

In (7), the poles $\{b_i, i=1, 2, \dots, M\}$ and residues $\{a_i, i=1, 2, \dots, M\}$ are the 2M unknowns that need to be determined. Matching the first 2M moments in (6) and (8) yields the following 2M nonlinear equations:

$$\begin{aligned}
-\left(\frac{a_1}{b_1} + \frac{a_2}{b_2} + \dots + \frac{a_M}{b_M}\right) &= m_0 \\
-\left(\frac{a_1}{b_1^2} + \frac{a_2}{b_2^2} + \dots + \frac{a_M}{b_M^2}\right) &= m_1 \\
&\vdots \\
-\left(\frac{a_1}{b_1^{2M}} + \frac{a_2}{b_2^{2M}} + \dots + \frac{a_M}{b_M^{2M}}\right) &= m_{2M-1}
\end{aligned} \tag{9}$$

The nonlinear equations in (9) can be solved using the algorithm proposed in [11]. Once the poles b_i and residues a_i have been determined, the probability density function $pdf(f)$ is optimally approximated by $h(t)$ in (7), and the cumulative distribution function $cdf(f)$ is optimally approximated by the step response:

$$s(t) = \int_0^t h(\tau) d\tau = \begin{cases} \sum_{i=1}^M \frac{a_i}{b_i} \cdot (e^{b_i t} - 1) & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{10}$$

It is worth noting 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 4, we will propose several schemes to address these problems.

The aforementioned moment-matching method was previously applied for IC interconnect order reduction [11], [12] and is related to the Padé approximation in linear control theory [13]. In the following section, we will explain why this moment-matching approach is an efficient way to approximate PDF/CDF functions.

3.2 Connection to Probability Theory

In probability theory, given a random variable f whose probability density function is $pdf(f)$, the characteristic function is defined as the Fourier transform of $pdf(f)$ [8].

$$\Phi(\omega) = \int_{-\infty}^{+\infty} pdf(f) \cdot e^{j\omega f} \cdot df = \int_{-\infty}^{+\infty} pdf(f) \cdot \sum_{k=0}^{+\infty} \frac{(j\omega f)^k}{k!} \cdot df \tag{11}$$

Substituting (6) into (11) yields:

$$\Phi(\omega) = \sum_{k=0}^{+\infty} m_k \cdot (-j\omega)^k \tag{12}$$

This implies an important fact: the time moments defined in (6) are related to the Taylor expansion of the characteristic function at the expansion point $\omega = 0$. Matching the first $2M$ moments in (9) 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 [8]:

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

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

Fig. 1 shows the characteristic functions for several typical random distributions. The above two properties imply an interesting fact: namely, given a random variable f , its

characteristic function decays as ω increases. Therefore, the optimally approximated $H(s)$ in (7) 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 band (around $\omega = 0$), it can be accurately approximated by matching the first several Taylor coefficients at $\omega = 0$, i.e. the moments. This conclusion has been verified in other applications (e.g. IC interconnect order reduction [11], [12]) and it provides the theoretical background to explain why moment-matching works well for the PDF/CDF evaluations that we will demonstrate in Section 5.

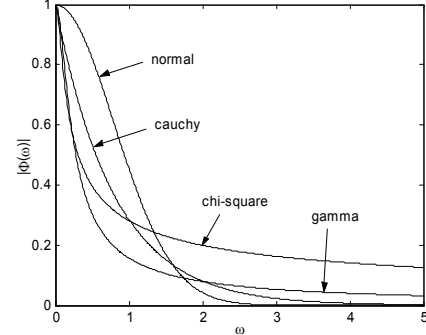


Fig. 1. Characteristic function for typical distributions.

4. Implementation 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 generalized Chebyshev inequality for PDF/CDF shifting and 3) a reverse evaluation technique for best-case/worst-case analysis. In this section, we describe the mathematical formulation of each of these algorithms.

4.1 Binomial Moment Evaluation

A key operation required in APEX is the computation of the high order time moments defined in (6) for a given random variable f . Such a moment evaluation is equivalent to computing the expectation of $\{f^k, k=0,1,\dots,2M-1\}$. Given the quadratic response surface model in (5), f^k is a high order polynomial in ΔY :

$$f^k(\Delta Y) = \sum_i c_i \cdot \Delta y_1^{\alpha_{i1}} \cdot \Delta y_2^{\alpha_{i2}} \cdot \dots \cdot \Delta y_N^{\alpha_{iN}} \tag{13}$$

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

$$E(f^k) = \sum_i c_i \cdot E(\Delta y_1^{\alpha_{i1}}) \cdot E(\Delta y_2^{\alpha_{i2}}) \cdot \dots \cdot E(\Delta y_N^{\alpha_{iN}}) \tag{14}$$

where $E(\bullet)$ stands for the expectation. In addition, remember that each random variable Δy_i has a Normal distribution $N(0,1)$, which yields [8]:

$$E(\Delta y_i^k) = \begin{cases} 1 & (k=0) \\ 0 & (k=1,3,5,\dots) \\ 1 \cdot 3 \cdot \dots \cdot (k-1) & (k=2,4,6,\dots) \end{cases} \tag{15}$$

Substituting (15) into (14), the expectation of f^k can be determined.

The above computation scheme is called *direct moment*

evaluation in this paper. The key disadvantage of the direct moment evaluation is that, as k increases, the total number of the product terms in (14) will increase exponentially, thereby quickly making their 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 scheme recursively computes the high order moments, instead of explicitly constructing the high order polynomials f^k in (14).

A. Quadratic Model Diagonalization

The first step in binomial moment evaluation is to remove the cross product terms in the quadratic response surface model (5), thereby yielding a much simpler, but equivalent, quadratic model.

According to matrix theory [10], any symmetric matrix $A \in R^{N \times N}$ can be diagonalized as:

$$A = U \cdot \Lambda \cdot U^T \quad (16)$$

where $\Lambda = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N)$ contains the eigenvalues of A and $U = [U_1, U_2, \dots, U_N]$ is an orthogonal matrix (i.e. $U^T U = I$) containing the eigenvectors. Define new random variables ΔZ as follows:

$$\Delta Z = U^T \cdot \Delta Y \quad (17)$$

Substituting (17) into (5) yields:

$$\begin{aligned} f(\Delta Z) &= C + Q^T \cdot \Delta Z + \Delta Z^T \cdot \Lambda \cdot \Delta Z \\ &= C + \sum_{i=1}^N (q_i \cdot \Delta z_i + \sigma_i \cdot \Delta z_i^2) \end{aligned} \quad (18)$$

where Δz_i is the i -th element in the vector ΔZ and $Q = [q_1, q_2, \dots, q_N]^T = U^T \cdot B$. Equation (18) implies that there is no cross product term in the quadratic model after diagonalization. In addition, the following theorem guarantees that the random variables ΔZ defined in (17) are still independent and satisfy the Normal distribution $N(0,1)$.

Theorem 1: Given a set of independent random variables ΔY with the Normal distribution $N(0,1)$ and an orthogonal matrix U , the random variables ΔZ defined in (17) are independent and satisfy the Normal distribution $N(0,1)$.

Proof: Since the random variables ΔZ are linear combinations of normally distributed random variables ΔY , they are normally distributed. The correlation matrix for ΔZ is given by:

$$E(\Delta Z \cdot \Delta Z^T) = E(U^T \cdot \Delta Y \cdot \Delta Y^T \cdot U) = U^T \cdot E(\Delta Y \cdot \Delta Y^T) \cdot U \quad (19)$$

Remember that ΔY is a set of independent random variables with a Normal distribution $N(0,1)$, i.e. $E(\Delta Y \cdot \Delta Y^T) = I$, and matrix U is orthogonal, i.e. $U^T U = I$. Thus, we have:

$$E(\Delta Z \cdot \Delta Z^T) = U^T \cdot E(\Delta Y \cdot \Delta Y^T) \cdot U = U^T \cdot I \cdot U = I \quad (20)$$

Equation (20) implies that the random variables in ΔZ are uncorrelated. In addition, uncorrelated random variables with Normal distributions are also independent [8]. ■

B. Moment Evaluation

We now demonstrate the use of the simplified quadratic model (18) for fast moment evaluation. Based on (18), we define a set of new random variables:

$$\begin{aligned} g_i &= q_i \cdot \Delta z_i + \sigma_i \cdot \Delta z_i^2 \\ h_l &= C + \sum_{i=1}^l g_i = C + \sum_{i=1}^l (q_i \cdot \Delta z_i + \sigma_i \cdot \Delta z_i^2) \end{aligned} \quad (21)$$

Comparing (21) with (18), it is obvious that when $l = N$, $h_N = f$. Instead of computing the high order moments of f directly, the proposed binomial moment evaluation scheme successively computes the moments of h_l , as shown in Fig. 2.

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

$$E(g_i^k) = E\left[(q_i \cdot \Delta z_i + \sigma_i \cdot \Delta z_i^2)^k\right] = \sum_{i=0}^k \binom{k}{i} \cdot q_i^i \sigma_i^{k-i} \cdot E(\Delta z_i^{2k-i}) \quad (22)$$

$$E(h_l^k) = E\left[(h_{l-1} + g_l)^k\right] = \sum_{i=0}^k \binom{k}{i} \cdot E(h_{l-1}^i) \cdot E(g_l^{k-i}) \quad (23)$$
3. If $l = N$, then go to Step 4. Otherwise, $l = l + 1$ and return Step 2.
4. For each $k = 0, 1, \dots, 2M - 1$, we have $E(f^k) = E(h_N^k)$.

Fig. 2. Binomial moment evaluation algorithm.

Step 2 in Fig. 2 is the key operation required by the binomial moment evaluation algorithm. In Step 2, both (22) and (23) utilize the binomial theorem to get the binomial series. Therefore, we refer to this algorithm as *binomial moment evaluation* in this paper.

In (22), the expectation $E(\Delta z_i^{2k-i})$ can be easily evaluated using the closed-form expression (15), since Δz_i is normally distributed $N(0,1)$. Equation (23) utilizes the property that h_{l-1} and g_l are independent, because h_{l-1} is a function of $\{\Delta z_i, i = 1, 2, \dots, l-1\}$, g_l is a function of Δz_l and all Δz_i are mutually independent. 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 is that, unlike the direct moment evaluation in (14), it does not explicitly construct the high order polynomial f^k . Therefore, unlike direct moment evaluation, where the total number of the product terms will exponentially increase, both $E(g_i^k)$ in (22) and $E(h_l^k)$ in (23) contain at most $2M$ product terms. Since $k = 0, 1, \dots, 2M - 1$ and $l = 0, 1, \dots, N$ for an M -th order APEX approximation with N independent random variations, the total number of $E(g_i^k)$ and $E(h_l^k)$ that need to be computed is $O(MN)$. In addition, the matrix diagonalization in (16) only needs to be computed once and has a complexity of $O(N^3)$. Therefore, the computational complexity of the proposed algorithm is $O(M^2 N) + O(N^3)$. In most circuit applications, N is small (around 5~100) after PCA analysis, and selecting $M = 7 \sim 10$ provides sufficient accuracy for moment matching. With these typical values for M and N , the proposed binomial moment evaluation is extremely fast, as we will demonstrate with numerical examples in Section 5.

It should be noted that as long as the circuit performance f is represented by the quadratic model in (5) and the process variations are normally distributed, binomial moment evaluation provides the exact high order moment values (except for numerical errors). There is no further assumption or

approximation made by the algorithm.

In summary, binomial moment evaluation utilizes statistical independence theory and principal component analysis (PCA) to efficiently compute high order moment values, which are required in moment matching for probability extraction.

4.2 PDF/CDF Shifting

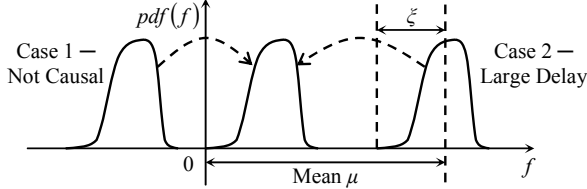


Fig. 3. Illustration for PDF/CDF shifting.

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 PDF $pdf(f - f_0)$, 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, as shown in Fig. 3 (case 2). As such, the corresponding impulse response $h(t)$ presents a large delay in time domain, which cannot be accurately captured by a low-order approximation. In such cases, we need to left-shift the unknown $pdf(f)$ by f_0 and use the impulse response $h(t)$ to approximate the shifted PDF $pdf(f + f_0)$.

The above 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 represented by the quadratic model in (5) is also unbounded. It is impossible 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 (f_0 is negative in case of right-shifting) 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 2: Given a random variable f , for any $\xi > 0$ and $k = 2, 4, 6, \dots$,

$$P(|f - \mu| \geq \xi) \leq \frac{E[(f - \mu)^k]}{\xi^k} \quad (24)$$

where μ is the mean value of f .

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

$$\begin{aligned} P(|f - \mu| \geq \xi) &= \int_{|f - \mu| \geq \xi} pdf(f) \cdot df \leq \int_{|f - \mu| \geq \xi} \frac{(f - \mu)^k}{\xi^k} \cdot pdf(f) \cdot df \\ &\leq \int_{-\infty}^{\infty} \frac{(f - \mu)^k}{\xi^k} \cdot pdf(f) \cdot df = \frac{E[(f - \mu)^k]}{\xi^k} \end{aligned} \quad (25)$$

Note that the above proof is not restricted to any special probability distribution. ■

Based on (24), if the unknown PDF $pdf(f)$ is left-shifted by $f_0 = \mu - \xi$, we have:

$$\begin{aligned} P(f - f_0 \leq 0) &= P(f - \mu + \xi \leq 0) = P(-f + \mu \geq \xi) \\ &\leq P(|f - \mu| \geq \xi) \leq \frac{E[(f - \mu)^k]}{\xi^k} \end{aligned} \quad (26)$$

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

$$\frac{E[(f - \mu)^k]}{\xi^k} \leq \varepsilon \quad (k = 2, 4, 6, \dots) \quad (27)$$

which is equivalent to:

$$\xi \geq \left\{ \frac{E[(f - \mu)^k]}{\varepsilon} \right\}^{\frac{1}{k}} \quad (k = 2, 4, 6, \dots) \quad (28)$$

Equation (28) estimates ξ using high order central moments. In an M -th order approximation, after the high order expectations $E(f^k)$ ($k = 0, 1, \dots, 2M - 1$) are computed by the binomial moment evaluation algorithm in Fig. 2, the central moments can be easily calculated using the binomial theorem:

$$E[(f - \mu)^k] = \sum_{i=0}^k \binom{k}{i} \cdot E(f^i) \cdot (-\mu)^{k-i} \quad (29)$$

where $\mu = E(f)$. Then, using (28), an estimated ξ is computed for each $k = 2, 4, \dots, 2M - 2$, 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 for $pdf(f)$.

It is worth mentioning that when $k = 2$, equation (24) is the well-known Chebyshev inequality [8]. We have generalized the 2nd order Chebyshev inequality to higher orders and, therefore refer to (24) as the *generalized Chebyshev inequality*. In practical applications we find that high order moments provide a much tighter (i.e. smaller) estimation of ξ , as is demonstrated by the numerical examples in Section 5.

In summary, the proposed generalized Chebyshev inequality (24) 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.

4.3 Reverse PDF/CDF Evaluation

In many practical applications, such as robust circuit optimization [4], [14], 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 3.2, APEX matches the first $2M$ Taylor expansion coefficients between the original characteristic function $\Phi(\omega)$ and the approximated rational function. Remember that the

Taylor expansion is most accurate around the expansion point $\omega = 0$. According to the final value theorem of the Laplace transform, accurately approximating $\Phi(\omega)$ at $\omega = 0$ provides an accurate $pdf(f)$ at $f \rightarrow \infty$. This, in turn, implies that the proposed approach can accurately estimate the 99% point of the random distribution, as shown in Fig. 4.

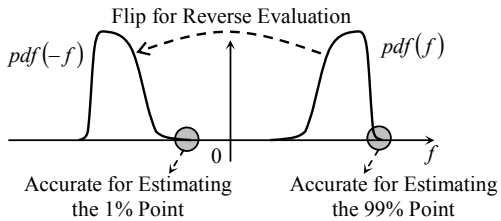


Fig. 4. Illustration for reverse PDF/CDF evaluation.

The above analysis motivates us to apply a reverse evaluation scheme for accurately estimating the 1% point. As shown in Fig. 4, 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.

4.4 Summary

1. Start from the quadratic response surface model in (5) and a given approximation order M .
2. Diagonalize the quadratic model based on (16)~(18) and compute the high order expectations $E(f^k)$ using the binomial moment evaluation algorithm in Fig. 2.
3. Compute the central moments $E[(f - \mu)^k]$ based on (29).
4. Determine the value of ξ using (28) and $f_0 = \mu - \xi$, where μ is the mean value of f .
5. Compute time moments $m_k = (-1)^k / k! \cdot E[(f - f_0)^k]$, where $E[(f - f_0)^k]$ is similarly evaluated by replacing μ by f_0 in (29).
6. Substitute m_k into (9) and solve the problem unknowns a_i and b_i .
7. The shifted $pdf(f + f_0)$ is approximated by the impulse response in (7) and the shifted $cdf(f + f_0)$ is approximated by the step response in (10).
8. Shift $pdf(f + f_0)$ and $cdf(f + f_0)$ back to $pdf(f)$ and $cdf(f)$.

Fig. 5. Overall implementation of APEX.

Fig. 5 summarizes the overall implementation of APEX except for reverse evaluation. If reverse evaluation is required to improve estimation accuracy for the 1% point, we need to compute the high order expectations $E[(-f)^k]$ in Step 2 and repeat Step 3~8 for computing $pdf(-f)$ and $cdf(-f)$. However, using reverse evaluation doesn't require explicitly computing the high order expectations again. Note that:

$$E[(-f)^k] = (-1)^k \cdot E(f^k) \quad (30)$$

where the high order expectations $E(f^k)$ have already been calculated in previous computations.

The algorithm in Fig. 5 is based on a given approximation

order M . The authors in [11] and [12] 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 here for APEX. In addition, it is worth mentioning that using an approximation order greater than 10 can result in serious numerical problems [11], [12]. In most practical applications, we find that selecting M in the range of 7~10 can achieve the best accuracy.

5. Numerical Examples

In this section we demonstrate the efficacy of APEX using several circuit examples. All experiments are run on a SUN Sparc - 1GHz server.

5.1 ISCAS'89 S27

A. Response Surface Modeling

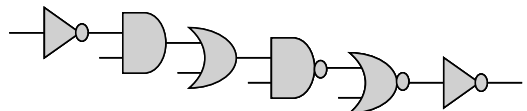


Fig. 6. Longest path in ISCAS'89 S27.

We create a physical implementation for the ISCAS'89 S27 benchmark circuit using the ST CMOS 0.13 μm process. This benchmark circuit is simple, but it enables us to make a full comparison of APEX with various PDF/CDF estimation methods. Given a set of fixed gate sizes, the longest path delay in the benchmark circuit (shown in Fig. 6) is a function of the process variations (e.g. V_{th} , T_{ox} , ΔL , etc.). Since the circuit only consists of six gates which can be put close to each other in the layout, inter-die variation will dominate intra-die variation, and gate delays will dominate (local) interconnect delays in this example. Therefore, for simplicity, we only consider inter-die variations for CMOS transistors in this example. The probability distributions and the correlation information of the inter-die transistor variations are obtained from the ST design kit. After PCA analysis, 6 principal random factors are identified to represent these process variations. We should note, however, that nothing precludes us from including more detailed intra-die and/or interconnect variation models in APEX as well.

We approximate the longest path delay as a function of 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 (4x difference).

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

B. Moment Evaluation

Table 2 compares the computation time for direct moment evaluation and our proposed binomial moment evaluation. In direct moment evaluation, the number of the total product terms increases exponentially, thereby making the computation task quickly infeasible. Binomial moment evaluation, however, is extremely fast and achieves more than 10^6 x speedup over direct moment evaluation. In addition, we verify that the moment values obtained from both approaches are identical except for numerical errors.

Table 2. Computation time for moment evaluation

Moment Order	Direct		Binomial
	# of Terms	Time (Sec.)	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

C. PDF/CDF Shifting

As discussed in Section 4.2, PDF/CDF shifting is necessary to make the proposed APEX approach feasible and efficient. A key operation for PDF/CDF shifting is determining the ζ value based on (28) (also see Fig. 3). We select an error tolerance $\varepsilon = 10^{-3}$ in (28). Fig. 7 shows the estimated ζ value using various high order moments. From Fig. 7, we find that the high order moments ($k > 2$) provide a much tighter (i.e. smaller) estimation of ζ . However, after the moment order $k > 10$, further increases in k do not have a significant impact on reducing ζ .

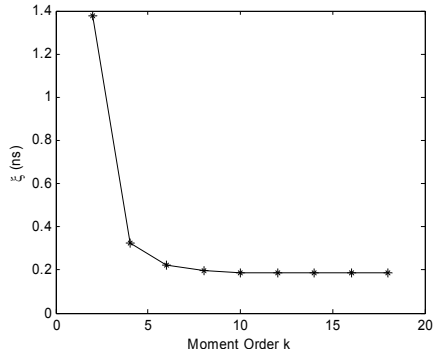


Fig. 7. Estimated ζ value using high order moments.

D. PDF/CDF Evaluation

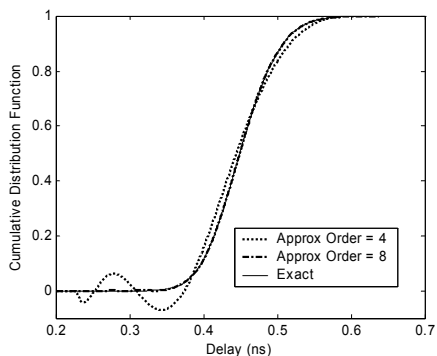


Fig. 8. Cumulative distribution function for delay.

Fig. 8 shows the cumulative distribution function using various approximation orders. In Fig. 8, the “exact” cumulative distribution function is evaluated by Monte Carlo simulation with 10^6 samples. Note that, the CDF obtained from the low order approximation (Order = 4) is not accurate and contains numerical oscillations. However, once the approximation order is increased to 8, these oscillations are eliminated and the approximated CDF asymptotically approaches the exact CDF. Similar behavior has

been noted in moment matching of LTI models of interconnect circuits [11], [12].

E. Comparison of Accuracy and Speed

Table 3 compares the accuracy and speed for three different probability extraction approaches: linear regression, Monte Carlo analysis with 10^4 samples, and the proposed APEX approach. Several specific points on the cumulative distribution function are utilized for comparing the accuracy. The 1% point and the 99% point, for example, denote the best-case delay and the worst-case delay respectively. After the cumulative distribution function is explicitly obtained in the closed-form expression (10), the best-case delay, worst-case delay and any other specific points on CDF can be easily found using a binary search algorithm. The error values in Table 3 are calculated against the “exact” CDF obtained by Monte Carlo simulation with 10^6 samples.

Note from Table 3 that the linear regression approach has the largest error. APEX achieves more than 200x speedup over the Monte Carlo analysis with 10^4 samples, while still providing better accuracy. In this example, reverse evaluation on $pdf(-f)$ reduces the 1% point estimation error by 4x, from 0.20% to 0.04%. This observation demonstrates the efficacy of the reverse evaluation method proposed in Section 4.3.

Table 3. Estimation error (compared to Monte Carlo with 10^6 samples) and computation cost

	Linear	MC (10^4 Runs)	APEX
1% Point	1.43%	0.34%	0.04%*
10% Point	4.63%	0.64%	0.01%
25% Point	5.76%	0.47%	0.03%
50% Point	6.24%	0.32%	0.02%
75% Point	5.77%	0.25%	0.02%
90% Point	4.53%	0.66%	0.03%
99% Point	0.18%	0.78%	0.09%
Cost (Sec.)	0.04	43.44	0.18

5.2 Low Noise Amplifier

A. Response Surface Modeling

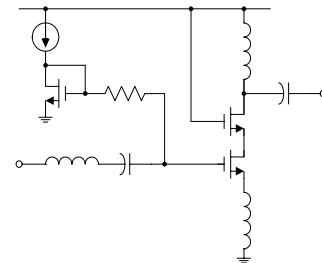


Fig. 9. Circuit schematic for LNA.

As a second example we consider a low noise amplifier designed in the IBM BiCMOS 0.25 μm process, as shown in Fig. 9. In this example, the variations on both MOS transistors and passive components (capacitance and inductance) are considered. The probability distributions and the correlation information of these variations are provided in the IBM design kit. After PCA analysis, 8 principal factors are identified to represent the process variations.

The performance of the LNA is characterized by 8 different specifications. Given a set of determined circuit sizes, each circuit performance is a function of the process variations. We approximate these unknown functions by linear regression models

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

and quadratic response (i.e. second order polynomial) models respectively. Table 4 shows the modeling error for all these 8 performances. In this example, the quadratic modeling error is 7.5x smaller than the linear modeling error on average.

Table 4. Regression modeling error for LNA

Performance	Linear	Quadratic
F0	1.76%	0.14%
S11	6.40%	1.32%
S12	3.44%	0.61%
S21	2.94%	0.34%
S22	5.56%	3.47%
NF	2.38%	0.23%
IIP3	4.49%	0.91%
Power	3.79%	0.70%

B. Comparison of Accuracy and Speed

Table 5. Estimation error for lower bound (1% point)

Performance	Corner	Linear	MC (10^4 Runs)	APEX
F0	15.80%	1.11%	0.24%	0.06%
S11	45.44%	5.78%	0.47%	0.09%
S12	38.87%	3.88%	0.59%	0.14%
S21	60.25%	2.91%	0.08%	0.17%
S22	23.08%	1.01%	0.29%	0.07%
NF	51.91%	3.70%	0.10%	0.06%
IIP3	54.64%	5.02%	0.22%	0.33%
Power	16.56%	0.01%	0.47%	0.09%

Table 6. Estimation error for upper bound (99% point)

Performance	Corner	Linear	MC (10^4 Runs)	APEX
F0	20.12%	1.10%	1.31%	0.05%
S11	51.53%	1.40%	1.74%	0.08%
S12	44.64%	1.16%	0.25%	0.28%
S21	51.63%	4.69%	0.02%	0.19%
S22	36.01%	5.61%	0.29%	0.19%
NF	72.82%	3.52%	1.28%	0.12%
IIP3	59.66%	5.93%	0.01%	0.26%
Power	42.53%	1.24%	1.53%	0.02%

Table 7. Computation cost for statistical analysis (Sec.)

Performance	Linear	MC (10^4 Runs)	APEX
F0	0.04	41.73	0.16
S11	0.03	41.86	0.14
S12	0.04	41.75	0.17
S21	0.04	41.84	0.19
S22	0.04	41.78	0.19
NF	0.03	41.70	0.15
IIP3	0.03	41.80	0.14
Power	0.04	41.91	0.15

Table 5 and Table 6 compare the estimation accuracy for four different statistical analysis approaches: corner simulation, linear regression, Monte Carlo analysis with 10^4 samples and the proposed APEX approach. These error values are calculated against the “exact” CDF obtained by Monte Carlo simulation with 10^6 samples.

The corner simulation approach computes the best-case and worst-case performance 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 Table 5 and Table 6. Linear

regression provides more accurate results than corner simulation, but the errors are expected to increase as IC technologies continue to scale. APEX achieves better accuracy than the Monte Carlo analysis with 10^4 samples, and is more than 200x faster, as shown in Table 7.

6. Conclusion

As IC technologies reach nanoscale, process variations are becoming relatively large and nonlinear (quadratic) response surface models might be required to accurately characterize the large-scale variations. In this paper we propose an asymptotic probability extraction (APEX) method for estimating the non-Normal random distribution resulting from the nonlinear response surface modeling. Three novel algorithms, i.e. binomial moment evaluation, CDF/PDF shifting and reverse PDF/CDF evaluation, are proposed to reduce the computation cost and improve the estimation accuracy. As is demonstrated by the numerical examples, applying APEX results in better accuracy than the Monte Carlo analysis with 10^4 samples, and achieves more than 200x speedup. APEX can be incorporated into a yield optimization loop or a timing analysis environment, for efficient probability extraction and worst-case analyses. For example, the efficacy of applying APEX to robust analog design is further discussed in [14].

7. Acknowledgements

This work was funded in part by the MARCO Focus Center for Circuit & System Solutions (C2S2, www.c2s2.org) under contract 2003-CT-888.

8. Reference

- [1] S. Nassif, “Modeling and analysis of manufacturing variations,” *IEEE CICC*, pp. 223-228, 2001.
- [2] C. Michael and M. Ismail, “Statistical modeling of device mismatch for analog MOS integrated circuits,” *IEEE JSSC*, 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,” *IEEE EDAC*, pp. 567-571, 1994.
- [4] A. Dharchoudhury and S. Kang, “Worse-case analysis and optimization of VLSI circuit performance,” *IEEE Trans. CAD*, 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,” *IEEE ICCAD*, pp. 374-380, 1996.
- [6] J. Swidzinski, M. Styblinski and G. Xu, “Statistical behavioral modeling of integrated circuits,” *IEEE ISCAS*, Vol. 6, pp. 98-101, 1998.
- [7] A. Graupner, W. Schwarz and R. Schüffny, “Statistical analysis of analog structures through variance calculation,” *IEEE Trans. CAS-I*, Vol. 49, No. 8, pp. 1071-1078, Aug. 2002.
- [8] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 2001.
- [9] G. Seber, *Multivariate Observations*, Wiley Series, 1984.
- [10] G. Golub and C. Loan, *Matrix Computations*, The Johns Hopkins Univ. Press, 1996.
- [11] L. Pillage and R. Rohrer, “Asymptotic waveform evaluation for timing analysis,” *IEEE Trans. CAD*, Vol. 9, No. 4, pp. 352-366, Apr. 1990.
- [12] M. Celik, L. Pileggi and A. Odabasioglu, *IC Interconnect Analysis*, Kluwer Academic Publishers, 2002.
- [13] M. Bosley and F. Lees, “A survey of simple transfer-function derivations from high-order state-variable models,” *Automatica*, Vol. 8, pp. 765-775, 1972.
- [14] X. Li, P. Gopalakrishnan, Y. Xu and L. Pileggi, “Robust analog/RF circuit design with projection-based posynomial modeling,” *IEEE ICCAD*, 2004.