

Fast Stochastic Analysis of Electromigration in Power Distribution Networks

Palkesh Jain, *Member, IEEE*, Vivek Mishra, *Student Member, IEEE*, and Sachin S. Sapatnekar, *Fellow, IEEE*

Abstract—A fast and stochastic analysis methodology for electromigration (EM) assessment of power distribution networks is presented in this work. We examine the impact of variability on EM time-to-failure (TTF), considering altered current densities due to global/local process variations as well as the fundamental factors that cause the conventional EM TTF distribution. Through novel variations-aware current density model based on Hermite polynomial chaos, we demonstrate significant margins in EM lifetime when compared with the traditional worst-case approach. On the other hand, we show that the traditional approach is altogether incompetent in handling transistor-level local variations leading to significantly optimistic lifetime estimates for lower metal level interconnects of PDN. Subsequently, we attempt to bridge the conventional, component-level EM verification method to the system level failures, inspired by the extreme order statistics. We make use of asymptotic order models to determine the TTF for the k^{th} component failure due to EM, and demonstrate application of this approach in developing IR drop aware system-level failure criteria.

Keywords: Electromigration, extreme value theory, worst-case corner, lognormal distribution

I. CIRCUIT-LEVEL ELECTROMIGRATION VERIFICATION

Electromigration (EM) in copper interconnects is caused by the current-driven movement of metal atoms and remains the foremost challenge to interconnect reliability. The flow of a contemporary industrial EM verification cycle is represented in Fig. 1, which outlines a comparison of specified EM limits on the current density in an interconnect (J_{th}) against the calculated actual current density (J) in the circuit. The procedure is based upon the characterization of failure data on serially interconnected test structures [1], where failure is defined by the first break in any element of the serial structure. The time to failure (TTF) of the structure primarily depends on the current density and stress temperature, empirically related through Black’s equation [2], [3], and characterization is performed under accelerated aging conditions of high voltage and temperature, in a regime where the failure fraction (FF) is high ($\sim 0.1 - 0.5$). To capture the stochastic nature of EM, the TTF is obtained over several test structures as a function of the current density, and then modeled as a lognormal distribution for the failure of a single wire.

To apply this characterized distribution to compute the EM TTF distribution in a manufactured product, it is important to account for the differences between the product use

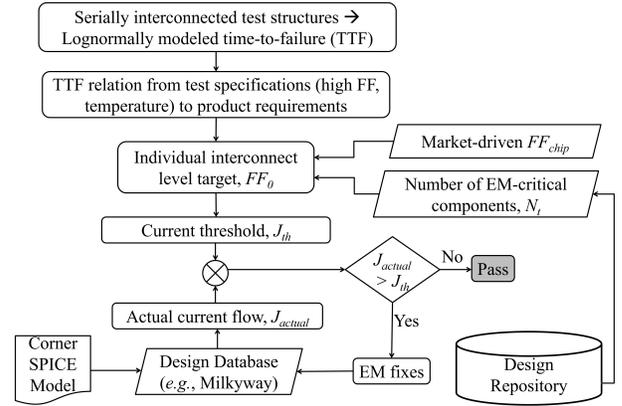


Figure 1: A schematic of the traditional EM verification flow.

case and the characterization scenario described above. First, the product operates at a nominal temperature and voltage that is lower than the accelerated stress conditions during characterization. The TTF data based on the lognormal distribution is therefore scaled to these use case values [4]. Second, while characterization addresses the failure of one wire, a chip typically consists of millions of interconnects, and the prevalent EDA approach requires the acceptable failure fraction, FF_{chip} , at the chip level to be translated to the component-level failure fraction, FF_0 , on individual wires. Since $FF_0 \ll 1$, the chip- and component-level FF s can be related as [3]:

$$FF_{chip} = 1 - (1 - FF_0)^{N_t} \implies FF_0 \approx FF_{chip}/N_t \quad (1)$$

Here, N_t is the number of EM-critical resistors [5]. In a typical industry flow [6], [7], the value of N_t is estimated at point of design definition, based on design experience and historical data. Using the computed value of FF_0 , the lognormal PDF is used to determine the current density threshold, J_{th} , for each wire in the design.

Traditionally, amongst various interconnects, the on-chip power delivery network (PDN) has been the primary EM concern. Industrial EDA tools verify these structures in a design by computing the actual current, J , in each wire at a given process corner, for a known clock frequency, application, and parasitics. Immortal wires are first filtered out using the Blech criterion [8]. For each of the remaining wires, J is compared against J_{th} . In case the threshold is violated, EM fixes are invoked through PDN optimization procedures such as wire widening and current flow reduction.

Palkesh Jain is with Qualcomm Technologies Inc., India, Bangalore. E-mail: palkesh@qti.qualcomm.com. Vivek Mishra and Sachin S. Sapatnekar are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA

Manuscript received xxxx-xx-xxx.

II. LIMITATIONS OF EXISTING EM METHODOLOGIES

As an increasing number of PDN wires becomes susceptible to EM in scaled technologies, several significant limitations in the methodology of Fig. 1 become more acute, resulting in the incorrect identification of EM-critical wires. The objective of this paper is to identify these factors, as outlined in the remainder of this section, and to propose a new EM verification methodology that addresses these issues.

A. Statistical variations in J

The PDN carries both leakage and switching currents, both of which are susceptible to statistical process variations. Variations in switching current are moderate and are captured by a Gaussian. **With shrinking technology nodes, the instantaneous transistor drive current and thereby the timing impact must be modeled in non-Gaussian manner [9]. However, since we are primarily interested in the average-switching current (dependent on the capacitance variations), Gaussian assumptions serve well [10].** Leakage currents, on the other hand, have a much wider non-Gaussian spread owing to the exponential dependency of leakage on threshold voltages [11]. In traditional designs, the switching current often limits the EM TTF, but as leakage currents become more significant, their impact can become dominant in some scenarios, particularly due to their large statistical spread. Finding the worst-case (WC) corner for current evaluation is difficult: using the timing WC corner can lead to unwarranted pessimism (up to $2\times$, as we will demonstrate in Table I). Prior work has largely neglected statistical variations, barring a few studies that assume Gaussian variations [12]–[14] and may not appropriately model leakage.

To demonstrate the impact of leakage current in on-chip power grids, we consider the example in Fig. 2, which represents an industrial octacore chip. **For the ease of illustration, we extract the wire currents in the upper metal layer of the chip, and assume equal wire widths, so that the current density in the wires can be approximated by the current alone.** The chip is shown to operate under four different workloads, in each of which a different numbers of cores is in active mode (shown using a solid outline), or in an idle or power-gated state (shown using a dotted outline). For each workload, we report the ratio of the total leakage current to the total switching current at the nominal process corner. Following common design practice, all cores share the PDN to contain the cost and complexity [15]. This sharing results in a mix of leakage and switching currents in upper metal layers, which are illustrated through current contours overlaid on the octacore layout. For example, under workloads b) and c), the active quad cluster sees identical activity, but the cores in the other quad cluster are either idle or power gated, altering the leakage:switching current ratios.

To examine the effect of process variations, we consider the impact of global process variations on these workloads through Monte Carlo (MC) simulations. These variations are further aggravated for local within-die variations. **For a 28nm block implemented in an industrial setup, we furnish the normalized results from 1000 Monte Carlo simulations in Fig. 3 for the**

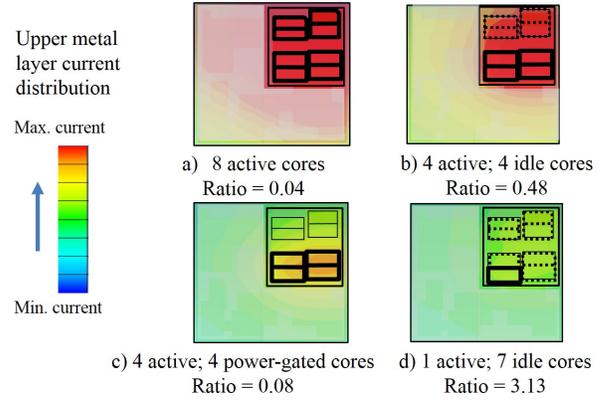


Figure 2: An octacore SoC, with the eight CPUs shown on the upper right, under various workloads. Depending on whether the CPUs are in active, idle, or power-gated mode, the ratio of total active power to total leakage power may vary, and the nominal current in the power grid (shown by the contours) may show different distributions.

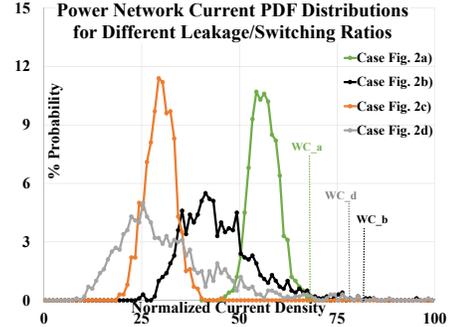


Figure 3: Current density PDFs in a power network for various cases mapping to Fig. 2.

four workloads in Fig. 2. It can be seen that although workload a) has the highest nominal current density, workloads b) and d) have a much larger variance due to their leakage-dominant nature. At a 99.7% yield point, the current density of 66.3 units for case a) is overshadowed by the value of 83.4 units for case b), implying that workload b) is the case that limits EM lifetime. On the other hand, at the $\sim 95\%$ yield point, the largest current density corresponds to the switching-dominated workload a).

The above data indicate that the worst estimation due to statistical variations depends not just on the workload, but also on the yield requirement. The traditional WC model is neither workload-dependent nor yield-configurable. Traditional WC SPICE models are often targeted to correlate with either 3σ transistor drive-current for the timing corner, or transistor leakage for the leakage corner, but not both, as required for EM verification. An EM-specific analysis is warranted because process variations alter both the total interconnect currents and the underlying failure kinetics of EM [16], neither of which is captured by the timing or leakage WC corner. Moreover, incorporating workload dependency in a WC model, i.e., deriving a unique 3σ -matching artificial process point from the

current spread for every workload, is logistically impractical. Additionally, while we assumed a multicore system with shared PDN in the above example, the traditional WC model is incapable even within a single core system in accounting for current-density variations induced due to local, random variations on transistors.

B. Outline of the proposed methodology

In the first step in Fig. 1, the characterization of the lognormal relies on test data collection at the foundry based on chain-like serially-interconnected test structures [1]. However, the actual circuit topologies that are evaluated are much complex. To simplify the analysis of the PDN, the traditional flow uses the weakest link approximation (WLA), wherein the entire PDN is deemed to have failed when the first component fails [17]. By breaking down system level reliability to a component-level problem, the WLA has enabled EDA methodologies for PDN verification to scale up to millions of wires. Unfortunately, this simplification does not grasp that system-level failure occurs beyond the point of first component failure, e.g., on-chip interconnects such as the PDN can satisfy IR drop constraints even after multiple EM failures due to its inherent redundancy [18], [19]. A Monte Carlo based approach has been proposed to solve this problem [19], modeling a cascade of EM events leading the system to successively degraded states, but is computationally prohibitive for large systems. Therefore, there is a strong need to bridge the gap between verification scalability and system-level feedback.

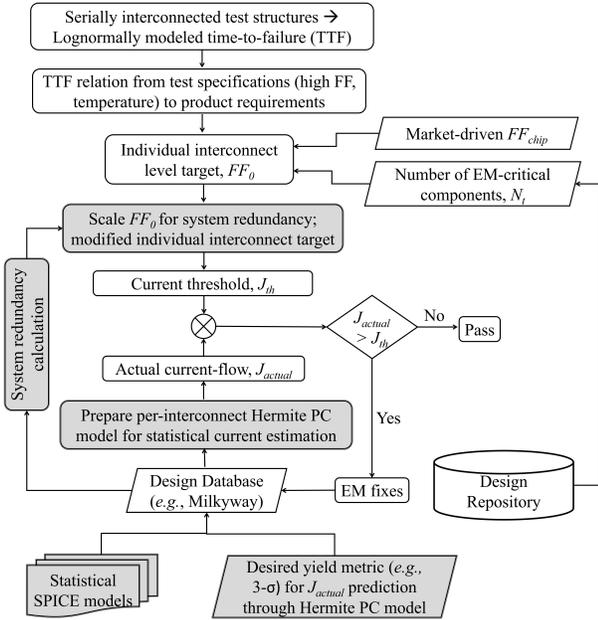


Figure 4: The proposed EM verification flow, where the highlighted regions indicate modifications to the traditional flow (Fig. 1).

In this work, we present a new EM verification procedure that modifies the traditional approach by addressing these limitations. We augment the traditional methodology in Fig. 1 through additional steps represented by the highlighted

blocks in Fig. 4. Our technical contributions as follows are summarized as follows:

- We incorporate system redundancy by applying the theory of order statistics and address system failure criteria using an asymptotic failure model to determine the TTF for the k^{th} component failure due to EM. Our approach is cognizant of the typical current-day EDA framework and arrives at a modified component level FF target incorporating a known extent of system redundancy. For a given system, the extent of redundancy is computed one-time as average number of sustainable failures (k) through Monte Carlo means during the early design phase. As we rely on Monte Carlo simulation only for one-time, it makes our method faster against performing such simulations for every iteration till the PDN optimization is achieved. Thus, our approach efficiently bridges the gap between component and system reliability for a given system.
- To incorporate the aggravated impact of non-Gaussian statistical process variations under arbitrary workloads, we statistically derive a value for the worst-case TTF. Our model is based on multivariate Hermite polynomial chaos.

The remainder of the paper is organized as follows. We recapitulate the mathematical basis for the current-day deterministic EM methodology in Section III. Section IV details the multivariate Hermite polynomial based method for incorporating global variations. Next, Section V describes a formulation for incorporating system redundancies based on extreme order statistics. Finally, Section VI shows a list of experimental results and their analysis in an industrial context and Section VII concludes the paper.

III. MODELING EM AND WIRE CURRENTS

A. TTF modeling

We now recapitulate the traditional, deterministic, EM modeling approach. EM failures are accelerated by two operational parameters for a circuit: the current density, J , and the temperature, T . Black's equation [2] specifies the mean TTF, t_{50} , as:

$$t_{50} = \frac{A}{J^m} \quad (2)$$

where m is the current exponent, and a typical value is $m = 1$ [3]. Based on [20],

$$A = \frac{L_c kT}{eZ \rho D_{eff}} e^{\frac{E_a}{kT}} \quad (3)$$

Here, E_a is the activation energy, k is Boltzmann's constant, eZ is the effective electron charge, ρ is the resistivity, L_c is the critical void length that causes a failure, and D_{eff} is the effective diffusivity, a constant.

The 50% fail fraction of (2) is too high for real applications. Since t_{use} is a lognormal function of random variable z_{use} [1], we represent a realistic fail fraction FF at time t_{use} as:

$$FF(z_{use}) = \Phi(z_{use}) \text{ where } z_{use} = \frac{\ln t_{use}/t_{50}}{\sigma} \quad (4)$$

where Φ is the standard Gaussian CDF. From (2) and (4), with $m = 1$,

$$t_{use} = \frac{A}{J} e^{\sigma z_{use}} \quad (5)$$

Finally, we will relate the value of z and t under two different sets of stresses, a and b , with currents J_a and J_b and temperatures T_a and T_b . Typically, condition a corresponds to the reference condition provided by the foundry, and condition b is the use condition at which EM reliability is evaluated. Substituting (2) for both cases into the corresponding expression for z in (4), we obtain:

$$z_b = z_a + \frac{1}{\sigma} \left[\ln \left(\frac{t_b J_b}{t_a J_a} \right) - \frac{E_a}{k} \left(\frac{1}{T_b} - \frac{1}{T_a} \right) \right] \quad (6)$$

This expression relates the lognormal random variable, z , related to the failure fraction, with the time to failure, t , under two stress conditions a and b . As we will see, in Section V, we formulate the shift in z required to correctly model system redundancy, and then map that solution back to an effective TTF using this equation.

B. Evaluating the PDN

A critical ingredient of TTF modeling is based on determining the current through each wire. Consider a PDN with N nodes, and let the conductance matrix of the power grid be represented by \mathbf{G} . Given a vector of excitations, \mathbf{I} , representing the current consumption at the N nodes, the behavior of the PDN is described by:

$$\mathbf{G}\mathbf{V} = \mathbf{I} \quad (7)$$

where \mathbf{V} is the vector of node voltages in the PDN.

Let I_k denote the k^{th} element of \mathbf{I} . The gate currents that contribute to I_k comprise the switching ($I_{S,k}$) and leakage ($I_{L,k}$) components, which are given by:

$$I_{S,k} = \sum_i \alpha_i C_i V_i f_i \quad (8)$$

$$I_{L,k} = \sum_i I_{0,i} e^{\beta_i V_t} \quad (9)$$

Here, α_i is the switching activity factor, V_i is the the supply voltage V_k , f_i is the operating frequency, and C_i is the equivalent switching capacitance for gate i . $I_{0,i}$ is the base leakage, and β_i is the sensitivity to the threshold voltage V_t for gate i . **In this work, we account for subthreshold leakage only, which is the dominant leakage mechanism [21] over the gate leakage or gate-induced drain leakage.** In both equations, the summations are taken over all gates whose currents contribute to node k . We write the switching current equation in terms of its nominal current, $I_{S,k}^{nom}$ without variations and its components due to global and local variations, $I_{S,k}^g$ and $I_{S,k}^l$, respectively, as

$$I_{S,k} = I_{S,k}^{nom} + I_{S,k}^g + I_{S,k}^l \quad (10)$$

Similarly,

$$I_{L,k} = I_{L,k}^{nom} + I_{L,k}^g + I_{L,k}^l \quad (11)$$

The switching and leakage currents are subject to process variations. The key process parameters that undergo global/local statistical variations are the capacitances and the threshold voltages, which are modeled by Gaussian distributions. These impact of the global and local variations on each component of current is described next.

Switching current: In the expression for $I_{S,k}$ in (8), the process-dependent term is C_i , which follows a normal distribution under global variations. This implies that its global component, $I_{S,k}^g$, is a weighted sum of normal distributions, which is also normally distributed. We represent $I_{S,k}^g \sim \mathcal{N}(\mu_{S,k}^g, \sigma_{S,k}^g{}^2)$.

Since the switching current is affected linearly by statistical local on-die variations, as seen in (8), the weighted sum of zero-mean Gaussian-distributed switching current terms used to obtain the local component of switching current, $I_{S,k}^l$, has zero mean. Further, since the summation that computes $I_{S,k}^l$ corresponds to the addition of a large number of gate switching currents, the variance of this sum virtually vanishes due to cancellation effects. Therefore, this variance is negligible as compared to the nominal switching current and can be neglected, i.e., $I_{S,k}^l \approx 0$.

Leakage current: Owing to global normal variations in threshold voltage, $I_{L,k}$ becomes a weighted sum of lognormal distributions. **The sum of lognormal random distributions is not known to possess a closed form. However, using moment matching based approaches like the widely-used Wilkinson's method [22] or inverse gamma distribution method [23], we can reasonably approximate the characteristics of the distribution.** In this work, we follow the Wilkinson's method, using which, the global leakage component, $I_{L,k}^g$, of the current at node k of the PDN is given by (9), and can be written as

$$I_{L,k}^g \approx I_{0,k}^g e^{\beta_k V_t^g} \quad (12)$$

where the terms $I_{0,k}^g$ and β_k are obtained from Wilkinson's formula. Note that in case dual V_t is used, there will be two terms in the above expression, one corresponding to each value of V_t , and the subsequent analysis is very similar.

For the leakage current component associated with local variations, denoted by $I_{L,k}^l$, since the variations are independent, it is easy to show that the variance of the sum of a large number of such terms is negligible, and that the effect of summation of a large set of independent lognormals is to shift the mean. Its local leakage component, $I_{L,k}^l$ is given by

$$I_{L,k}^l = \sum_i E[I_{0,i} e^{\beta_i V_t^l}] = I_{0,k}^l \quad (13)$$

When the number of leakage components to be added is smaller (e.g., on lower metal layers of the power grid), the evaluation of the sum of these lognormals proceeds using Wilkinson's method, in a manner similar to the global variation case except that the lognormals are all uncorrelated.

C. Modeling the distributions of wire current densities

To compute the current density, J_M , for any metal segment M between the nodes (i, j) of the grid one can solve (7) as

$$J_M = (V_i - V_j)/(R_M w_M) \quad (14)$$

where R_M is the branch resistance and w_M is the segment width. Technically, the current density is J_M/t_M , where t_M is the segment thickness, but we use this simpler definition since t_M is constant for a metal layer and can be incorporated into a maximum current density limit. Using (7), the voltage at the node i is given by:

$$V_i = \sum_{k=1}^N q_{ik} I_k \quad (15)$$

where q_{ik} is the $(i, k)^{\text{th}}$ entry of \mathbf{G}^{-1} . This leads to:

$$J_M = \sum_{k=1}^N q_{ij,k} \frac{I_k}{R_M w_M} \quad (16)$$

where $q_{ij,k} = q_{ik} - q_{jk}$. Note that the computation of \mathbf{G}^{-1} is impractical, but we use this notion for ease of exposition. As we will soon see, this computation is not necessary.

In order to assess the impact of these variations on the segment current density J_M , we must first separately compute the leakage and switching current density components ($J_{S,M}$ and $J_{L,M}$, respectively), and subsequently determine the impact their statistics on the distribution of J_M . Noting the individual compositions of leakage and switching currents for I_k , we can represent J_M as follows:

$$J_M = \underbrace{\sum_{k=1}^N q_{ij,k} \frac{I_{S,k}}{R_M w_M}}_{J_{S,M}} + \underbrace{\sum_{k=1}^N q_{ij,k} \frac{I_{L,k}}{R_M w_M}}_{J_{L,M}} \quad (17)$$

Next, we show how $J_{S,M}$ and $J_{L,M}$ can be evaluated without the expensive step of explicitly inverting \mathbf{G} . To compute $J_{S,M}$, we begin by solving (7) under switching current excitations only, i.e., for the case where $I_k = I_{S,k}^g$, since $I_{S,k}^l = 0$. Through LU factorization, the system (7) can be rewritten as $(\mathbf{L}\mathbf{U})\mathbf{V} = \mathbf{I}$, where $\mathbf{G} = \mathbf{L}\mathbf{U}$, and the solution can be obtained through:

$$\mathbf{L}\mathbf{y} = \mathbf{I} \quad (18)$$

$$\mathbf{U}\mathbf{V} = \mathbf{y} \quad (19)$$

Forward substitution in (18) obtains each $y_k, 1 \leq k \leq N$, as

$$y_k = \left(I_k - \sum_{l=1}^{k-1} L_{kl} y_l \right) / L_{kk} \quad (20)$$

Since $I_{S,k}^g$ is a Gaussian, the evaluation of each y_k involves the summation of some constants and/or Gaussians, and therefore it is easy to represent each y_k as a Gaussian.

The backward substitution step in (19) evaluates voltages as

$$V_k = \left(y_k - \sum_{l=k+1}^N U_{kl} V_l \right) / U_{kk} \quad (21)$$

As before, each step involves a summation of Gaussians and/or constants, and therefore each voltage can be obtained as a Gaussian distribution. Using (14), the switching component of each branch current density is thus expressed as a Gaussian distribution.

A similar approach can be used to find $J_{L,M}$ in each wire. The excitation is now set either to $I_k = I_{L,k}^l$, a constant (on the upper metal layers), or to the lognormal sum (on the lower metal layers), and the first component of $J_{L,M}$ is computed for each wire as a constant. Next, setting $I_k = I_{L,k}^g$, a lognormal, we perform forward substitution, using Wilkinson's method to approximate the sums of lognormals as a lognormal at each step. The same approach is used in backward substitution, and this leads to expressing the second component of $J_{L,M}$ as a lognormal.

Finally, adding up $J_{S,M}$ and $J_{L,M}$, each branch current distribution is expressed as a sum of a Gaussian, a lognormal, and a constant.

IV. MODELING WIRE CURRENT VARIATION

To assess the impact of variations in a single wire, they must be incorporated into (5). We consider the electrical and physical parameters that affect the EM lifetime, namely, (a) the switching and leakage current variations, as discussed in the previous section, driven by shifts in the threshold voltage, V_t , and interconnect parasitics, and (b) EM kinetics. These are modeled as set of uncorrelated random variables (RVs). **In case of correlated RVs, we can use techniques such as principal component analysis (PCA) to arrive at a different coordinate system, wherein all translated RVs are uncorrelated.**

In some contexts where the impact of perturbations is relatively small, such as statistical timing analysis, it is common to use a first-order Taylor series expansion to capture the performance impact of variations, but for EM lifetime estimation, the existence of exponential terms implies that first-order expansions are inadequate and a higher order Taylor series expansion is mandated. An alternative approach to incorporate higher order expansions and non-Gaussian variations is to treat the varying electrical and physical parameters as a continuous stochastic process. This process can be represented as an infinite series of orthogonal polynomial chaos (PC) in a Hilbert space of random variables, truncated later by finite-dimensional projections while minimizing the error. This results in a response expression as a multidimensional polynomial in the random variables that represent the variations [24]. Indeed, using these polynomials, much higher order expansions to capture nonlinear terms are efficiently possible when compared to perturbation techniques. Orthogonal PC based methods have been previously applied for analyzing IC performance in [25], [26], but have not yet been employed for reliability analysis.

In this work, we employ the Hermite PC scheme and Galerkin procedure [24] to convert the stochastic reliability problem to a set of deterministic problems, later solved through standard matrix manipulations. This leads to the mean and variance estimation of interconnect EM lifetime, and correspondingly, a worst-case estimate (e.g., at the 3σ value).

A. Hermite PC based model

The basic principle of Hermite PC based approach is to use a series of orthogonal polynomials (of orthonormal Gaussian random variables) to facilitate stochastic analysis. Let $\xi = [\xi_1, \xi_2, \dots, \xi_n]$ denote a vector of n independent unit Gaussian random variables, modeling the variations in V_t , interconnect, and EM kinetics. As a result, t_{use} is also a random variable that is a function of ξ , and the impact of the random variable ξ on (5) can be explicitly shown as:

$$t_{use}(\xi) = \frac{A(\xi)}{J(\xi)} e^{\sigma(\xi)z_{use}} \quad (22)$$

Based on the principle of orthogonal polynomials, t_{use} can be approximated by a truncated Hermite PC expansion:

$$\hat{t}_{use}(\xi) = \sum_{k=0}^P t_k H_k^n(\xi) \quad (23)$$

where $H_k^n(\xi)$ is an n -dimensional Hermite polynomial with deterministic coefficients t_k . The number of terms, P , is related to n [24]. For a single variable, ξ_1 , we can represent:

$$H_0^1(\xi_1) = 1; H_1^1(\xi_1) = \xi_1, H_2^1(\xi_1) = \xi_1^2 - 1; \quad (24)$$

The weighting function for Hermite polynomials is the Gaussian probability density function and they are orthogonal with respect to this weighting function [27] in the Hilbert space:

$$\langle H_i(\xi), H_j(\xi) \rangle = \langle H_i^2(\xi) \rangle \delta_{ij} \quad (25)$$

where δ_{ij} is the Kronecker delta and $\langle *, * \rangle$ denotes the inner product. The coefficients t_k can be evaluated by the projection operation onto the Hermite PC basis. From (23), the mean $\mu_{t_{use}}$ and variance $\sigma_{t_{use}}^2$ of $\hat{t}_{use}(\xi)$ are:

$$\mu_{t_{use}} = t_0; \quad \sigma_{t_{use}}^2 = \sum_{k=1}^P t_k^2 E[H_k^2] \quad (26)$$

We define the error, $\Delta(\xi)$, as the difference between the exact t_{use} and its value from (22), i.e., $\left(\hat{t}_{use}(\xi) - \frac{A'}{J(\xi)} \right)$, where $A' = Ae^{\sigma(\xi)z_{use}}$. This implies that

$$J(\xi)\Delta(\xi) = (J(\xi)\hat{t}_{use}(\xi) - A') \quad (27)$$

To obtain t_k , we use Galerkin's method, which states that for the best approximation of $t_{use}(\xi)$, the error is orthogonal to the polynomials [24], i.e.,

$$\langle J(\xi)\Delta(\xi), H_k(\xi) \rangle = 0, k = 0, 1, \dots, P \quad (28)$$

This approach transforms the stochastic analysis to the deterministic task of computing the Hermite PC coefficients.

Next, we consider how to represent the process-dependent parameters – the leakage current $J_L(\xi)$, the switching current $J_S(\xi)$, and the term $A(\xi)$ that captures the EM kinetics – as functions of the underlying orthonormal Gaussians.

We represent the lognormal **leakage current** in a wire, J_L , as a function of the normally-distributed V_t . The global variations in V_t have mean and variance as (μ_L, σ_L^2) , and in turn, are modeled using the unit normal random variable

$\xi_L \sim \mathcal{N}(0, 1)$. Therefore:

$$J_L(\xi) = J_0 e^{\beta(\mu_L + \xi_L \sigma_L)} \quad (29)$$

For the lognormal relationship of leakage current with threshold voltage, we use a second-order Hermite polynomial:

$$J_L(\xi) = \sum_{k=0}^2 J_{Lk} H_k^n(\xi) \quad (30)$$

$$= J_{L0} \left(1 + \beta \sigma_L \xi_L + \frac{1}{2} \beta^2 \sigma_L^2 (\xi_L^2 - 1) \right) \quad (31)$$

Here, $J_{L0} = J_0 e^{\beta(\mu_L + \sigma_L^2/2)}$, $J_{L1} = J_{L0} \beta \sigma_L$, and $J_{L2} = J_{L0} \beta^2 \sigma_L^2 / 2$, where (μ_L, σ_L) come from the V_t distribution.

The **switching current** in a wire, J_S , is altered linearly with the normally-distributed global capacitance variations, with mean and variance as (μ_S, σ_S^2) , modeled as $\mu_s + \sigma_s \xi_s$ where $\xi_s \sim \mathcal{N}(0, 1)$.

$$J_S(\xi_S) = \sum_{k=0}^1 J_{Sk} H_k^n(\xi) = J_{S0} + J_{S1} \xi_S \quad (32)$$

Here, $J_{S0} = \mu_S$ and $J_{S1} = \sigma_S$.

The impact of **EM kinetics** is felt in the form of global variations in the term $A' = Ae^{\sigma z_{use}}$, caused by the process-dependent elements of the prefactor A in (3), and the variance, σ of the lognormal [16]. We model these variations in EM kinetics as lognormal, wherein the underlying normal distribution has mean and variance as (μ_K, σ_K^2) and is modeled through the unit random variable $\xi_K \sim \mathcal{N}(0, 1)$. Thus, if $A'_0 = Ae^{z_{use}(\mu_K + \sigma_K^2)}$, for a specified z_{use} , we can represent this nonlinear dependence using the Hermite polynomial:

$$A'(\xi_K) = \sum_{k=0}^2 A_k H_k^n(\xi) \quad (33)$$

$$= A'_0 \left(1 + z_{use} \sigma_K \xi_K + \frac{1}{2} z_{use}^2 \sigma_K^2 (\xi_K^2 - 1) \right) \quad (34)$$

where A_0, A_1 , and A_2 can be deduced from the above.

B. Hermite PC: Coefficient estimation

Next, we assess the eventual influence of ξ_L, ξ_S , and ξ_K on the EM lifetime through the Hermite PC representation for t_{use} , represented in the second order form as:

$$\hat{t}_{use}(\xi) = t_0 + t_1 \xi_L + t_2 \xi_S + t_3 \xi_K + t_4 (\xi_L^2 - 1) + t_5 (\xi_S^2 - 1) + t_6 (\xi_K^2 - 1) + t_7 \xi_L \xi_S + t_8 \xi_S \xi_K + t_9 \xi_K \xi_L \quad (35)$$

To compute the Hermite PC coefficients, we apply (28):

$$\begin{aligned} \langle J(\xi)\Delta(\xi), 1 \rangle &= 0; & \langle J(\xi)\Delta(\xi), \xi_L \rangle &= 0 \\ \langle J(\xi)\Delta(\xi), \xi_S \rangle &= 0; & \langle J(\xi)\Delta(\xi), \xi_K \rangle &= 0 \\ \langle J(\xi)\Delta(\xi), \xi_L^2 - 1 \rangle &= 0; & \langle J(\xi)\Delta(\xi), \xi_S^2 - 1 \rangle &= 0 \\ \langle J(\xi)\Delta(\xi), \xi_K^2 - 1 \rangle &= 0; & \langle J(\xi)\Delta(\xi), \xi_L \xi_S \rangle &= 0 \\ \langle J(\xi)\Delta(\xi), \xi_S \xi_K \rangle &= 0; & \langle J(\xi)\Delta(\xi), \xi_K \xi_L \rangle &= 0 \end{aligned}$$

Next, we substitute $J(\xi) = J_L(\xi_L) + J_S(\xi_S)$, and use (27) and (35). Comparing the coefficients of like terms on both sides

of each of the 10 inner products above, we obtain:

$$\tilde{J} \tilde{t}_{use} - \tilde{A} = 0 \quad (36)$$

where $\tilde{A} = [A'_0, 0, 0, A'_1, 0, 0, A'_2, 0, 0, 0]^T$
 $\tilde{t}_{use} = [t_0, t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9]^T$

Here, \tilde{J} is the 10×10 matrix:

$$\begin{pmatrix} J_{00} & J_{L1} & J_{S1} & 0 & 2J_{L2} & 0 & 0 & 0 & 0 & 0 \\ J_{L1} & J_{00} + 2J_{L2} & 0 & 0 & 2J_{L1} & 0 & J_{S1} & 0 & 0 & 0 \\ J_{S1} & 0 & J_{00} & 0 & 0 & 2J_{S1} & 0 & J_{L1} & 0 & 0 \\ 0 & 0 & 0 & J_{00} & 0 & 0 & 0 & J_{S1} & 0 & J_{L1} \\ 2J_{L2} & 2J_{L1} & 0 & 0 & 2J_{00} + 8J_{L2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & J_{S1} & 0 & 0 & 2J_{00} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & J_{00} & 0 & 0 & 0 \\ 0 & J_{S1} & 0 & J_{L1} & 0 & 0 & 0 & J_{00} + 2J_{L2} & 0 & 0 \\ 0 & 0 & 0 & J_{S1} & 0 & 0 & 0 & 0 & J_{00} & 0 \\ 0 & 0 & 0 & J_{L1} & 0 & 0 & 0 & 0 & 0 & J_{00} \end{pmatrix}$$

where ($J_{00} = J_{L0} + J_{S0}$). Thus, depending on coefficients of the leakage distributions (J_{L0}, J_{L1} and J_{L2}), switching current distribution (J_{S0}, J_{S1}) and the EM kinetics distributions (A_0, A_1 , and A_2), the matrix can be solved to find the \tilde{t}_{use} , which solely defines the statistics of EM lifetime.

The TTF formulations in (26) can be used to determine a realistic worst-case estimate of the EM lifetime due to global variations. **For example in the traditional context, we can compute the 3σ lifetime as $\mu t_{use} - 3\sigma t_{use}$ value. On the other hand, the statistical lifetime can also be easily derived for an arbitrary yield requirement.**

C. Relevance to alternative EM checking paradigms

We would like to reiterate that the formulations developed so far have been based on void-growth dominated Black's equation (2): the *de facto* model applied across industry. However, it is also extendible to alternative EM checking paradigms such as the modified nucleation based Black's approaches [28], [29] or flux divergence based approaches [30], [31].

For example, incorporating modified Black's equation requires alteration to the starting representation as: $t_{use}(\xi) = \left(\frac{A(\xi)}{J(\xi)} + \frac{B(\xi)}{J^2(\xi)}\right) e^{\sigma(\xi)z_{use}}$, where the void-nucleation related current term gets added. Consequently, the Hermite PC coefficients must be reworked in (37). On the other hand, flux divergence based approach like via-node vector method is directly applicable on (22) and (37), with a single change that the current density J , changes from an individual wire density to the effective current density at the via-node [30].

V. EM UNDER CIRCUIT REDUNDANCY

The lognormal model is empirically driven and is based on the weakest link failure data from a set of test structures, i.e., it assumes that the first wire failure results in system failure. However, the applicability of the weakest link model has been often questioned [3], [16], since real multimillion interconnect systems have significant redundancy and survivability even after the first component failure. Some studies on power grid and signal interconnects have proposed TTF models where failure is declared when a critical system parameter (e.g., skew or IR drop) exceeds a specification [18], [19]. Notably, such exceedances occur after multiple individual components fail, highlighting the need to incorporate redundancy into EM TTF estimations.

One way to address the problem is to perform MC simulations, which can model a cascade of EM events that lead the system to successively degraded states, but this is computationally prohibitive. Alternatively, if we could develop a framework to accurately predict the time to the k^{th} component failure in a system, it could be utilized for high-level assessments of the reliability benefit due to system redundancies, or even in deriving the EM guidelines. Interestingly, this problem statement fits into the *Order Statistics* branch of EVT [32], [33], which is widely used in financial risk management. In this work, we explore its usage to predict the time to k^{th} component failure and now recapitulate some of the required mathematical concepts.

Order Statistics Distribution: If we draw n samples from a population described by a random variable z , i.e., $Z = [Z_1, Z_2, \dots, Z_n]$, and rearrange them in increasing order as $[Z_{1,n} \leq Z_{2,n} \leq \dots \leq Z_{n-1,n} \leq Z_{n,n}]$, then the k^{th} term, $Z_{k,n}$, of this sequence is the k^{th} order statistic. For finite n and k , we can represent the k^{th} order statistic as:

$$F_{k,n} = P(Z_{k,n} \leq z) = \sum_{i=k}^n \binom{n}{i} F(z)^i \{1 - F(z)\}^{n-i} \quad (37)$$

where $F(z)$ is the CDF of z . Intuitively, this is the probability of *at least* k (out of n) draws of Z_i to be less than or equal to z . Two common cases of order statistics are:

- **first minima**, or the first order ($F_{1,n} = 1 - (1 - F(z))^n$), where one draw out of n is $\leq z$.
- **first maxima**, or the last order ($F_{n,n} = F(z)^n$), where all draws out of n must be $\leq z$.

The k^{th} order statistic, $F_{k,n}$, corresponds to the k^{th} minima. In reliability terminology, $F(z)$ corresponds to the interconnect failure CDF. Consequently, the *minima* maps to the series system, where the system failure occurs at *first* component failure, whereas *maxima* maps to the parallel, where the system fails at the *last* component failure. The first *minima* forms the basis of the weakest link based EM checking practices in industry [5], whereas *maxima* is not of interest to our problem.

To assess the applicability of order statistics on TTF estimation, we generate and analyze the TTF of an ensemble of hundred independent interconnects through Monte Carlo means. We proceed in following manner:

- We first assign 1000 random FF values (generated through normal distribution) per interconnect, which are used to estimate the lognormal TTF for every interconnect.
- These hundred TTF distributions are subsequently ordered.
- The first, second, third and fourth time to failures across all hundred distributions are collected to obtain the distribution of first four failures of the ensemble.
- We plot these TTF from various failure orders on a Gumbel-form scale, $\ln(-\ln(1 - CDF))$ versus $\ln(t)$.

For above case, the outcome is illustrated through Fig. 5, where the x -axis indicates the normalized logarithmic TTF and the y -axis is the Gumbel-form failure probability, $\ln(-\ln(1 - CDF))$. Indeed, when we examine the plot of first failure (also the worst, represented in grey color) across all the hundred interconnects, we notice that it follows a linear trend with a high correlation coefficient. The second, third and

fourth TTF distributions also exhibit similar behaviour. This observation is indeed a signature of ordered behaviour and a strong motivation for its applicability on TTF estimation, as also anticipated through [16].

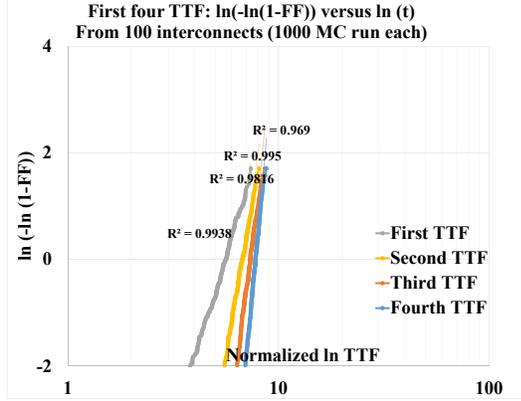


Figure 5: Time to k^{th} failure on a Gumbel plot demonstrating applicability of order statistics.

Asymptotic Order Statistics: While (37) represented the k^{th} order statistics for finite n , MC simulation is far too expensive. We will resort to asymptotic EVT as $n \rightarrow \infty$ and $k \ll n$, both of which are relevant in dealing with power grid system of multi-million interconnects. This also has the benefit of providing closed-form analytical solutions.

For any given CDF, $F(z)$, an asymptotic maxima distribution, $Z_{n,n}$, is said to exist iff

$$F(a_n z + b_n)^n \rightarrow G(z) \quad (38)$$

where a_n and b_n are fitting constants and $G(z)$ is a function. For such cases, $F(z)$ is also said to be in the domain of maximal attraction of $G(z)$. Now, a normal CDF, $F(z)$, satisfies the convergence criteria, and the limiting minima function is given by the Gumbel form [33]:

$$\hat{G}(z) = 1 - \exp(-e^z) \quad (39)$$

The above relation is for the asymptotic minima, but we are interested in the asymptotic k^{th} minima ($\hat{G}^{(k)}(z)$), for which we reuse the results from Gumbel [32] as:

$$\hat{G}^{(k)}(z) = 1 - [\exp(-ke^{z_k})] \sum_{j=0}^{k-1} k^j \frac{e^{z_k j}}{j!} \quad (40)$$

where $z_k = (z - b_k)/a_k$, with a_k and b_k being the empirically derived fitting constants. The reader is referred to [32] for a detailed derivation. It can be verified easily that for $k = 1$, the ordered distribution evaluates to that of the asymptotic minima, $\hat{G}(z)$.

In our methodology, we obtain initial guidance of the extent of system redundancy as an input from the designer or through system-level statistical Monte Carlo simulations. These simulations provide us the system failure CDF as well as an approximation of the number of failures the system can sustain before the system requirement breaks, which becomes our guidance for the value of k . Notice that even though the formulation (40) assumes that all the

underlying CDFs, $F(z)$, are identical, its applicability to TTF estimation of PDN interconnects is still motivated by the fact that the PDN weaknesses are mostly found in clusters containing interconnects of similar nature. Thus, with the help of appropriately derived fitting constants a_k and b_k to approximate the failure CDF obtained through a set of Monte Carlo simulations, we can fit the results to a Gumbel distribution to arrive at order statistics based CDF formulation.

Application for TTF Estimation:

If the system can tolerate k failures, then the k^{th} minima, $\hat{G}^{(k)}(z)$, represents the failure CDF for the system, and this can be used for reliability prediction under system redundancy as follows:

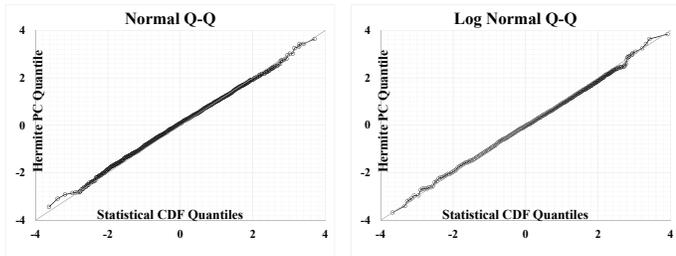
- 1) Given a customer requirement, FF_{chip} on the acceptable fail fraction for the chip, (1) can be applied to translate it to a component-level fail fraction, FF_0 .
- 2) Through a one-time assessment of the system level reliability using Monte Carlo simulations, we obtain the system failure CDF, which is then approximated through the Gumbel distribution $\hat{G}^{(k)}(z)$, wherein, the parameters k , a_k and b_k are empirically estimated.
- 3) Based on the value of k , the number of wire failures that can be tolerated before system failure, the corresponding Gumbel distribution, $\hat{G}^{(k)}(z)$, is used to map FF_0 to $z_{use} = [\hat{G}^{(k)}]^{-1}(FF_0)$.
- 4) Using (4), the foundry failure specification, specified as $(z_{ref}, J_{ref}, t_{ref}, T_{ref})$ is translated to z_{use} , as computed above, t_{use} , the lifetime specification on the chip, and T_{use} , the operating temperature specification for reliability evaluation of the chip. This results in a current limit, J_{th} , on the wire.
- 5) For each wire, the actual current, J_{actual} , through the wire is compared against J_{th} to verify whether it passes EM verification or not.

Alternatively, if the lifetime associated with a current, J_{actual} , is to be computed, we may use (4) to translate the reference values, along with z_{use} and T_{use} , to obtain the lifetime, t_{use} .

VI. RESULTS

We now present the results from the methods developed so far. For consistency, all of our results and implementation are based on the standard IBM power grid benchmark circuits [34]. We take the technology constants from an industry environment, and we normalize the data for confidentiality. We proceed as follows:

- Firstly, we validate the Hermite PC based framework to model the statistical variability in Section VI-A. We benchmark our results against statistical SPICE simulations (involving transistor global/local variability). The outcome of this statistical analysis is the current density for every resistor at any given yield point, for example 3σ which can be then used for verification against the EM thresholds.
- Subsequently, we take the current densities of individual resistors and incorporate system redundancy through order statistics based approach in VI-B. We benchmark our results against system-level statistical reliability simulations [18].



(a) Gaussian current-density distribution (b) Non-Gaussian current density distribution

Figure 6: Q-Q representation of the CDFs derived through statistical SPICE simulations and Hermite PC based approach for Gaussian and non-Gaussian cases.

A. Statistical Variability Estimation

1) *Power Network: Experimental Setup:* The IBM PG grid benchmarks are representative of different classes of industrial designs and vary over a reasonable range of complexity. The original intent of these PG grids was to benchmark improvements in the capability of simulation tools for PG grid simulation. Therefore, they only contain grid parasitics and voltage/current sources, instead of transistor level subcircuits of standard cells. The grids represent various blocks on the chip, and each block contains number of current sources as a representative of the switching instances of standard-cells. These current sources are DC and correspond to the cumulative current requirement of the standard-cell (including switching and leakage contributions).

In contrast with this original intent, our primary purpose is to study the impact of statistical variations (in the form of transistor V_t variations and switching capacitances) on EM lifetime, and therefore, we must separate the current-flow per standard cell in switching and leakage components. This is achieved by assigning Gaussian and lognormal distributions to the switching and leakage components, respectively, of the current sources. The nominal value of the cumulative current requirement of the standard-cell matches the original IBMPG grid value, and the ratio between nominal values of leakage and switching contributions is globally controlled. We utilize the HSPICE Monte Carlo variability simulation setup, in which the parameter governing leakage variation of the standard-cell could be treated as a globally (and identically) distributed or locally (and independently) distributed parameter. In this way, through statistical SPICE simulations, we can arrive at the current distribution spread in every resistor of the power grid, considering global and/or local variations. Subsequently, we derive (a) Hermite PC model and (b) WC based estimates to perform comparison against the statistical results.

2) Evaluation of the Hermite PC approach:

Through statistical SPICE simulations on the IBMPG2 grid and with global variations enabled, we directly obtain the current density CDFs for all the resistors.

We validate the results of our Hermite PC model under global variations against these simulations. We follow the procedure outlined earlier in Section III, and compute the

nominal values of leakage and switching currents per resistor. For the resistors in the power grid, we extract the $q_{ij,k}$ coefficients in (16). As outlined earlier, these are used to derive the variance of leakage and switching current density per wire using Wilkinson’s method, noting that these variations are identical and fully correlated. Subsequently, using (37), we set up the Hermite PC model and evaluate it to produce the CDFs of the current density.

These CDFs, obtained from statistical SPICE simulations and Hermite PC based approach are then plotted on a Q-Q plot against each other in Fig. 6, for two cases, representing the switching-dominated case and a case with significant leakage contributions. As we can see, for both the cases, the Hermite PC based CDF is in close agreement with the CDF derived through statistical SPICE simulations.

Next, to evaluate the Hermite PC based model under local variations, we first set the leakage variability parameter in HSPICE to be locally and independently distributed. Thus, through statistical SPICE simulations, we can directly obtain the current density CDFs for various resistors in the power grid under local variations. For illustration purposes, we pick resistors corresponding to lower, middle and upper metal layers of the PG grid, and their CDFs are plotted using black solid lines in Fig. 7 where the data for every resistor is normalized with respect to its median value.

As discussed earlier in Section IV, the current density in a given resistor is a weighted summation of the individual gate currents ((16)), wherein the position of the resistor in the power grid affects the weightages. Unlike global variations, which are fully correlated, the local variations are independent and uncorrelated. Thus, a resistor in lower metal layer, which sees the current-flow primarily from adjacent standard-cells, experiences a relatively high spread in its distribution, as shown through Fig. 7a. On the other hand, as we move to upper metal layers, the number of influencing standard-cells for a given resistor keeps increasing, resulting in a much tighter distribution shown through Figs. 7b and 7c (Note that the x-axis range grows progressively smaller in these figures). The CDFs obtained from the Hermite PC model are shown in Fig. 7 through green solid lines. For all the three cases, the Hermite PC based CDF correlates very well with the CDF derived through statistical SPICE simulations.

The influence of wires in various metal layers is further illustrated when we graphically plot the $q_{ij,k}$ coefficients in (16) for these resistors in Fig. 8. We normalize and rank order the coefficients for the three resistors and plot the top 100 values. As we can see, for resistor R176902, which is on lower metal layer, there are only few cells that influence its value, as signified by fewer coefficients with high values. On the other hand for resistor R208433 on a middle metal layer, there are more cells with high coefficients. Using these coefficients in Wilkinson’s method, and noting that these variations are now uncorrelated, we can compute the leakage and switching variances per resistor, eventually leading to the Hermite PC based model using (37). Thus, the impact of local variations is more prominent on lower metal layers and Hermite PC based model rightly comprehends this behaviour.

Next, we estimate the current density from a timing-based

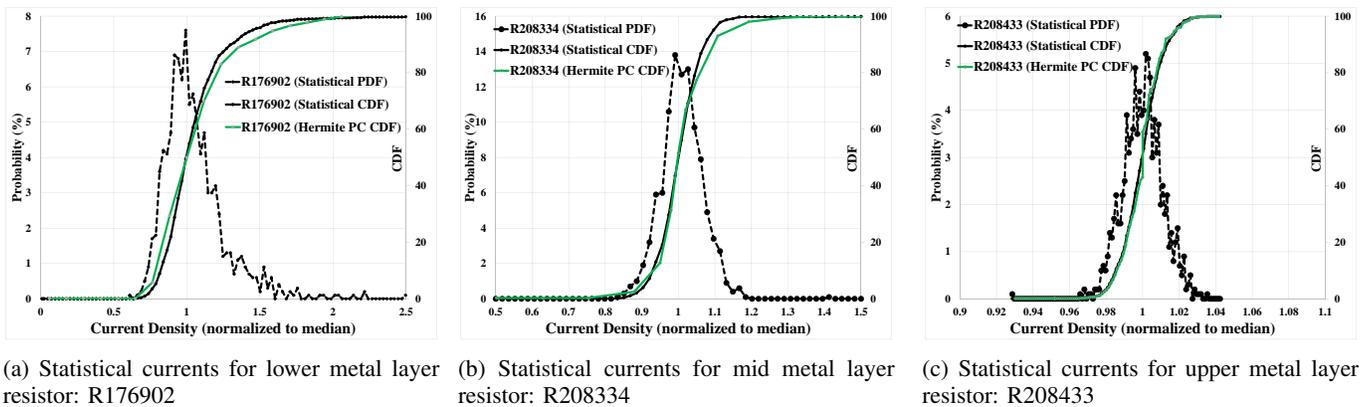


Figure 7: Current density PDFs and CDFs derived through statistical SPICE simulations and Hermite PC based approach for three resistor cases, corresponding to lower, mid and upper metal layers, incorporating local variations. Distributions becomes narrower as the resistors move to upper metal layers.

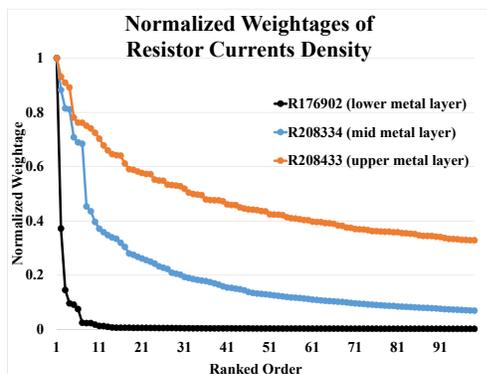


Figure 8: Normalized and ranked-order sensitivity of different resistors to individual current sources. x -axis indicates the identifier for one amongst several current-sources in the design.

WC approach, which assumes a strong transistor (optimized to 3σ transistor current) and worst-case capacitance. We set the individual standard-cell current sources to their corresponding timing-based WC values and perform the current measurements for the entire power grid for a given ratio of leakage and switching current per cell. For the same scenario, we also compute the lifetime from the Hermite PC model and the results are as tabulated in the Table I, where, we can see that Hermite PC based lifetime estimates are over $1.3\times$ better than the WC ones for 3σ yield point. **Indeed, our method can predict the lifetime for different yield targets (for example, 95% or 99% as against only 3σ), and the benefit with respect to WC estimation significantly increases for these cases, as seen from the second and third rows of Table I. Lastly, if we consider both global and local variations in our 3σ estimates, we see that the lifetime of lower metal resistor decreases as compared to the WC, owing to the fact that WC completely lacks accounting of local variations.**

In summary, our Hermite PC model agrees very well with statistical SPICE estimates for comprehending global as well as local variations. For the cases studied, our model

Design Entity		R176902 (lower metal layer)	R208334 (mid metal layer)	R208433 (upper metal layer)
Distribution Spread (global/local, 3σ /Mean)		2.38	1.15	1.03
t_{use} Improvement	global (95%/WC)	1.57	1.51	1.39
	global (99%/WC)	1.39	1.38	1.24
	global (3σ /WC)	1.37	1.32	1.21
	global + local (3σ /WC)	0.72	1.16	1.12

Table I: Comparison of our analytical EM lifetime prediction (at different yield points) against a timing-based WC approach.

reduces pessimism in EM lifetime estimates as compared to WC approach for global variations, whereas the timing-based WC approach is altogether incompetent in modeling local variations. Additionally, our model can be directly applied to project the lifetime for a given failure fraction requirement, thus enabling the yield tradeoffs which provides an opportunity to designers to cope with outlier violations.

3) *Runtime Comparison*: Earlier in Sec. III, we outlined the derivation of statistical lifetime coefficients using LU decomposition and subsequent calculations. From an industrial context, we make use of the HSPICE framework to compute these coefficients through the adjoint sensitivity means, since the percentage of high current-carrying resistors in the entire power grid is low. With the help of these coefficients, the wire statistics are then be estimated. For various IBMPG benchmark circuits, the runtime for our approach as against the full Monte Carlo approach is as listed in the Table II. Notice that the MC simulations take a significantly longer time due to incorporation of local variations. Overall, significant gain in runtime with an acceptable accuracy is demonstrated using Hermite PC based approach.

B. Application of Order Statistics

We now present results based on the methodology developed in Section V on the IBM power grid benchmark circuits [34], wherein our intention is to mimic the system failure through order statistics based prediction. We first discuss the methodology for progressive interconnect degradation in the

Grid name	CPU time per sim	1000 MC sim (global + local)	Our approach runtime
IBMPG1	0.10m	10.25m	0.35m
IBMPG2	0.15m	15.05m	0.47m
IBMPG3	3.93m	6.55h	11.79m
IBMPG4	5.37m	8.95h	16.14m
IBMPG5	0.55m	54.70m	1.66m
IBMPG6	1.99m	3.32h	6.00m
IBMPGNEW1	3.99m	6.65h	11.98m
IBMPGNEW2	9.00m	15.00h	27.03m

Table II: CPU runtime comparison of our analytical EM lifetime prediction versus Monte Carlo simulations.

PG grid resulting in increased IR drop. We define system failure when the IR drop increases by 50mV at any node in the grid from its $t = 0$ value.

The PG benchmark circuits are solved to obtain the currents and IR drop at every node; the currents are scaled to create a 100mV IR drop at $t = 0$. We obtain the system failure time through MC simulation, similar to the method in [18]. Each MC simulation involves:

- computing $t = 0$ currents and assigning normally distributed random FF to resistors
- using the current-flow, operating temperature and FF per resistor to derive the TTF for every resistor ((6))
- rank-ordering of resistors to open-circuit the resistor with least TTF
- recomputing currents and IR drop, and
- iterating till system criterion failure criterion is met.

Next, we perform MC simulations to obtain the statistical distribution, with different numbers of iterations per simulation warranted by different PG benchmark circuits to breach the system failure criterion. Recall that in our approach, MC simulations serve to provide the initial guidance of redundancy in the system, or an estimation of the average number of acceptable interconnect-failures. Since mean-estimation converges rapidly, 100 samples are sufficient for ($> 95\%$ confidence) when the underlying distribution is normal [35]. Thus, 100 MC sims gives a good estimate of the average number of sustainable failures.

Specifically looking at IBMPG1, from the MC runs, the TTF for the first, third and fifth (out of $\sim 15K$) failing resistors from every MC simulation can be obtained. Next, using the $t = 0$ currents data and the transformation from z to t , we obtain the ordered statistics model from (40). The TTF distributions for both the MC-based and the ordered statistics model are plotted in Fig. 9, where the x -axis is the normalized time. The region of interest corresponds to small FF values (< 0.25), and in this region, order-based failure estimations are in good agreement with MC estimations.

Additionally, from the MC simulations, we also extract the system TTF, i.e., the time at which, due to progressive EM degradation, any node in the PG grid suffers a drop that is 50mV higher than its $t = 0$ value. This data is also plotted in Fig. 9. It is clear that this system-level criterion is violated *much* later than the first component failure. Moreover, the system failure CDF can be approximated by the k^{th} component failure CDF. Here, the third failure CDF could be used to approximate the TTF to a 50mV drop. Notice

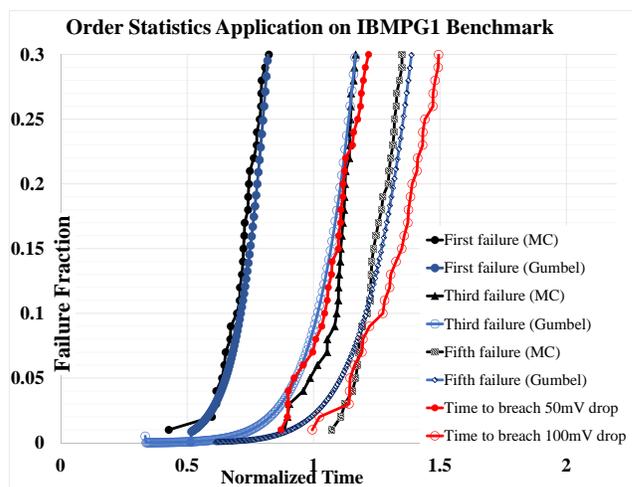


Figure 9: Application of Order Statistics Based EM Prediction on PG benchmark, IBMPG1.

that the order statistics based model (40) is guided one-time through the MC means to derive the average number of components which must fail before the system failure. This is an important result, since even though MC based system TTF estimation provides an accurate picture of failures, performing full MC analysis on a production PG grid is computationally prohibitive for every iteration cycle till the PDN optimization is achieved.

Next, we look at larger PG grid benchmark circuits for applicability of this principle. For one such grid, IBMPGNEW1, the voltage drop maps at $t = 0$ is shown at left in Fig. 10, which shows the inherent drop of the circuit before any wire failures. As the stress builds and we take the system through progressive degraded states, at the point of system failure, there will be at least one node in the PG grid whose voltage drop is 50mV higher as compared to its $t = 0$ value. The voltage drop map at that time instant is shown at right in Fig. 10. Notice that prior to this point, the circuit is functional and is able to tolerate several EM failures.

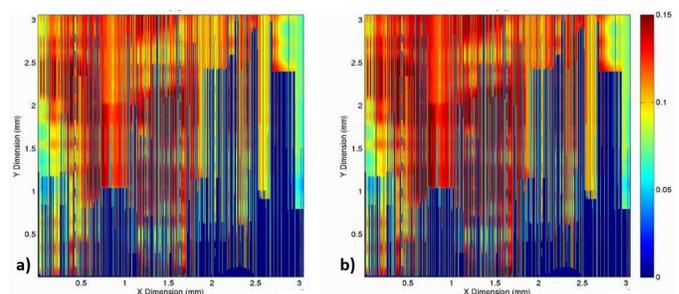


Figure 10: Voltage drop maps of the power grid, IBMPGNEW1 (left) at $t = 0$, showing the inherent IR drop of the circuit with no wire failures (right) after the circuit undergoes 20 EM events, after which there is at least one node whose voltage drop is 50mV higher as compared to its $t = 0$ value. The IR drop scale is described at right.

The results for the other power grid benchmarks are

Grid name	Total voltage sources	Total interconnects	Avg. failures to system fail (rounded)	CPU runtime		Normalized mean TTF (hrs)		
				System MC	Proposed Model	WLA	System MC	Proposed model
IBMPG1	100	14,750	3	5.13m	0.15m	1.000	1.626	1.507
IBMPG2	120	101,718	8	7.53m	0.23m	0.571	0.999	1.097
IBMPG3	494	677,388	10	3.28h	5.90m	0.593	1.089	1.120
IBMPG4	312	780,699	8	4.48h	8.06m	0.696	1.220	1.352
IBMPG5	100	495,756	7	27.35m	0.82m	0.568	0.970	1.073
IBMPG6	132	797,711	9	1.66h	2.98m	0.612	1.186	1.160
IBMPGNEW1	494	698,595	18	3.33h	5.99m	0.754	1.566	1.577
IBMPGNEW2	494	1,157,849	16	7.50h	13.50m	0.807	1.590	1.866

Table III: Comparison of our analytical EM lifetime prediction against Monte Carlo and WC approach, performed on different power grid benchmarks. The failure criterion is 50mV higher voltage drop on any node as compared to its $t = 0$ value.

Grid name	Number of EM violations	
	Default thresholds	Order model
IBMPG2	372	85
IBMPGNEW2	116	41

Table IV: Application of order model for threshold based verification of circuits.

presented in Table III. As noted before, we perform 100 MC simulations to obtain the statistical distribution, with different numbers of iterations per simulation warranted by different PG benchmark circuits to breach the system failure criterion. Here, we list the average number of failures required per PG grid to violate the system criterion (rounded off to nearest integer). We also report the normalized mean TTF (hrs) from WLA, system MC and proposed order based method. Recall that as the order method requires only one-time estimation of system redundancy through MC means, it is computationally superior as compared to performing costly MC sims for every iteration of design-closure till the PDN optimization is achieved. As we can see from the Table III, the lifetime estimates from the proposed method agrees well with the more expensive MC-based computation and reduce significant pessimism as compared to the WLA method. It must be mentioned, however, that due to topological differences, the number of interconnects required to create a system failure are different for various grids (column 4 in Table III), and that is an input to the order statistics based TTF generation procedure.

Lastly, we share the results of EM thresholds based verification from the proposed order based method, using the guidance of average number of failures required per grid. The default thresholds are targetted for a tighter reliability specification so as to expose the EM violations. To derive the thresholds from order approach, we follow the steps enumerated earlier in Sec. V, and the results are as shown in Table IV. While Table III establishes the accuracy of order based method against the system level MC simulations, we now see the translation of lifetime benefit in form of reduced number of violations. Indeed, for IBMPG2 circuit, a designer must *only* fix 85 (of 372) violations and still, safe operation of the circuit is guaranteed as per the given system criterion. Thus, using the order statistics model, we demonstrate that the system time-to-failure can be approximated analytically as well as its application for threshold based EM verification in a conventional context.

VII. CONCLUSION

A fast and stochastic methodology for electromigration analysis of power distribution networks has been presented in this work. The impact of statistical global/local process variation on EM TTF assessment has been examined. Through novel variations-aware current density models, we demonstrate significant margins in EM lifetime when compared with the traditional worst-case approach. On the other hand, we show that the traditional approach is altogether incompetent in handling transistor-level local variations leading to significantly optimistic lifetime estimates for lower metal level interconnects. Additionally, we show that the traditional component-level model is inadequate in predicting the system failure since system failure often occurs after multiple components fail. Through an extreme order statistics based approach, we have demonstrated that system failures, in form of IR drop exceedances, can be approximated reasonably by an asymptotic k^{th} component failure model. As our method requires only one-time estimation of system redundancy through Monte Carlo means, it is computationally superior as compared to performing costly MC simulations for every iteration of design-closure till the PDN optimization is achieved. Thus, our approach efficiently bridges the gap between component and system reliability for a given system.

REFERENCES

- [1] K.-D. Lee, "Electromigration critical length effect and early failures in Cu/oxide and Cu/low k interconnects," Ph.D. dissertation, University of Texas at Austin, Austin, Texas, USA, 2003.
- [2] J. R. Black, "Electromigration failure modes in aluminum metallization for semiconductor devices," *Proceedings of the IEEE*, vol. 57, no. 9, pp. 1587–1594, 1969.
- [3] J. R. Lloyd and J. Kitchin, "The electromigration failure distribution: The fine-line case," *Journal of Applied Physics*, vol. 69, no. 4, pp. 2117–2127, 1991.
- [4] P. Jain, J. Cortadella, and S. S. Sapatnekar, "A fast and retargetable framework for logic-ip-internal electromigration assessment comprehending advanced waveform effects," *IEEE Transactions on VLSI Systems*, vol. 24, no. 6, pp. 2345–2358, Jun. 2016.
- [5] B. Li, P. McLaughlin, J. Bickford, P. Habitz, D. Netrabile, and T. Sullivan, "Statistical evaluation of electromigration reliability at chip level," *IEEE Transactions on Device and Materials Reliability*, vol. 11, no. 1, pp. 86–91, 2011.
- [6] J.-G. Ahn, M. F. Lu, N. Navale, D. Graves, P.-C. Yeh, J. Chang, and S. Y. Pai, "Product-level reliability estimator with budget-based reliability management in 20nm technology," in *Proceedings of the IEEE International Reliability Physics Symposium*, 2015, pp. 6B–2.1–6.B–2.5.
- [7] B. Li, C. Christiansen, D. Badami, and C.-C. Yang, "Electromigration challenges for advanced on-chip Cu interconnects," *Microelectronics Reliability*, vol. 54, no. 4, pp. 712–724, 2014.
- [8] I. A. Blech, "Electromigration in thin aluminum films on titanium nitride," *Journal of Applied Physics*, vol. 47, no. 4, pp. 1203–1208, 1976.

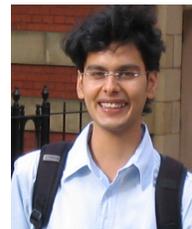
- [9] S. Balasubramanian, N. Pimparkar, M. Kushare, V. Mahajan, J. Bansal, T. Shimizu, V. Joshi, K. Qian, A. Dasgupta, K. Chandrasekaran, and C. Weintraub, "Near-threshold circuit variability in 14nm finfets for ultra-low power applications," in *Proceedings of the International Symposium on Quality Electronic Design*, 2016.
- [10] Y. Yang and N. K. Jha, "Finprin: Finfet logic circuit analysis and optimization under pvt variations," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 12, pp. 2462–2475, 2014.
- [11] Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices*. Cambridge, UK: Cambridge University Press, 2013.
- [12] Z. Guan, M. Marek-Sadowska, and S. Nassif, "Statistical analysis of process variation induced SRAM electromigration degradation," in *Proceedings of the International Symposium on Quality Electronic Design*, 2014, pp. 700–707.
- [13] Y. Ban, C. Choi, H. Shin, Y. Kang, and W. H. Paik, "Analysis and optimization of process-induced electromigration on signal interconnects in 16nm FinFET SoC (system-on-chip)," in *SPIE Advanced Lithography*, 2014, pp. 90 530P–1–90 530P–11.
- [14] J. Lienig, "Electromigration and its impact on physical design in future technologies," in *Proceedings of the ACM International Symposium on Physical Design*, 2013, pp. 33–40.
- [15] "High-level Considerations for Power Management of a big.LITTLE System," <http://www.infocenter.arm.com/help/topic/com.arm.doc.dai0424a/DAI0424A.pdf>, 2011.
- [16] A. E. Hubbard and J. R. Lloyd, "Study of the variation in electromigration performance from lot to lot variations and the implications for design rule generation," in *IEEE International Integrated Reliability Workshop Final Report*, 2014, pp. 123–126.
- [17] D. F. Frost and K. F. Poole, "Reliant: a reliability analysis tool for vlsi interconnect," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 2, pp. 458–462, 1989.
- [18] P. Jain, S. S. Sapatnekar, and J. Cortadella, "Stochastic and topologically aware electromigration analysis for clock skew," in *Proceedings of the IEEE International Reliability Physics Symposium*, 2015, pp. 3D–4.1–3D–4.6.
- [19] V. Mishra and S. S. Sapatnekar, "The impact of electromigration in copper interconnects on power grid integrity," in *Proceedings of the 50th Annual Design Automation Conference*, 2013, article 88.
- [20] M. Hauschildt, "Statistical analysis of electromigration lifetimes and void evolution in Cu interconnects," Ph.D. dissertation, University of Texas at Austin, Austin, TX, USA, 2005.
- [21] M. Agostinelli, M. Alioto, D. Esseni, and L. Selmi, "Leakage–delay tradeoff in finfet logic circuits: A comparative analysis with bulk technology," *IEEE Transactions on VLSI Systems*, vol. 18, no. 2, p. 232, 2010.
- [22] P. Cardieri and T. S. Rappaport, "Statistics of the sum of lognormal variables in wireless communications," in *Proceedings of the Vehicular Technology Conference Proceedings*, vol. 3, 2000, pp. 1823–1827.
- [23] A. K. Acar, E. and S. Nassif, "Characterization of total chip leakage using inverse (reciprocal) gamma distribution," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 2006, pp. 4–pp.
- [24] R. G. Ghanem and P. D. Spanos, *Stochastic finite elements: a spectral approach*. North Chelmsford, MA: Courier Corporation, 2003.
- [25] S. Vruthula, J. M. Wang, and P. Ghanta, "Hermite polynomial based interconnect analysis in the presence of process variations," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 25, no. 10, pp. 2001–2011, 2006.
- [26] N. Mi, J. Fan, S. X.-D. Tan, Y. Cai, and X. Hong, "Statistical analysis of on-chip power delivery networks considering lognormal leakage current variations with spatial correlation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 7, pp. 2064–2075, 2008.
- [27] E. Zeraouia, *Models and Applications of Chaos Theory in Modern Sciences*. Boca Raton, FL: CRC Press, 2011.
- [28] R. L. De Orio, H. Ceric, and S. Selberherr, "Physically based models of electromigration: From Black's equation to modern TCAD models," *Microelectronics Reliability*, vol. 50, no. 6, pp. 775–789, 2010.
- [29] J. R. Lloyd, "Black's law revisited–Nucleation and growth in electromigration failure," *Microelectronics Reliability*, vol. 47, no. 9, pp. 1468–1472, 2007.
- [30] Y.-J. Park, P. Jain, and S. Krishnan, "New electromigration validation: Via node vector method," in *Proceedings of the IEEE International Reliability Physics Symposium*, May 2010, pp. 698–704.
- [31] Z. Guan, M. Marek-Sadowska, S. Nassif, and B. Li, "Atomic flux divergence based current conversion scheme for signal line

electromigration reliability assessment," in *Proceedings of the IEEE International Interconnect Technology Conference/Advanced Metallization Conference*, 2014, pp. 245–248.

- [32] E. J. Gumbel, *Statistics of extremes*. North Chelmsford, MA: Courier Corporation, 2012.
- [33] H. A. David and H. N. Nagaraja, *Order statistics*. Hoboken, NJ: Wiley Online Library, 1970.
- [34] S. R. Nassif, "Power grid analysis benchmarks," in *Proceedings of the Asia and South Pacific Design Automation Conference*, 2008, pp. 376–381.
- [35] A. DasGupta, *Probability for statistics and machine learning: fundamentals and advanced topics*. New York, NY: Springer Science & Business Media, 2011.



Palkesh Jain (M'04) graduated from the Indian Institute of Technology Bombay, in 2004 with Bachelors and Masters in Electrical Engineering. Subsequently, he joined the ASIC group at Texas Instruments India, where he defined and developed, several of the GHz enabling reliability methodologies. In 2014, he joined the Central Engineering team at Qualcomm India, where he is presently involved with system level power and thermal management methodologies and initiatives. He holds 15 US patents (granted/ pending) and is also a part time doctoral student at the Universitat Politècnica de Catalunya.



Vivek Mishra (S'12) received the B. Tech and M. Tech degrees from the Indian Institute of Technology, Bombay in 2008. He worked as an ASIC design engineer at Taiwan Semiconductor Manufacturing Company in Hsinchu, Taiwan, from 2008 to 2011. He completed his Ph.D. degree in 2016 from the Department of Electrical and Computer Engineering at the University of Minnesota, where he was the recipient of the 2015-16 Doctoral Dissertation Fellowship. He is presently working as a Research and Development Engineer at Synopsys Inc. His research interests include circuit modeling and development of efficient algorithms to analyze the impact of interconnects on circuit performance.



Sachin S. Sapatnekar (S'86, M'93, F'03) received the B. Tech. degree from the Indian Institute of Technology, Bombay, the M.S. degree from Syracuse University, and the Ph.D. degree from the University of Illinois. He taught at Iowa State University from 1992 to 1997 and has been at the University of Minnesota since 1997, where he holds the Distinguished McKnight University Professorship and the Robert and Marjorie Henle Chair. His current research interests lie in developing efficient techniques for computer-aided design of integrated circuits, and are primarily centered around physical design, timing and simulation issues, and optimization algorithms. He has served on the editorial boards of several IEEE journals, including as Editor-in-Chief of the IEEE Transactions on CAD. He was a Technical Program Co-chair in 2006 and 2007 for the Design Automation Conference (DAC), and the General Chair in 2010. He has also been the Technical Program Chair and General Chair for the International Symposium on Physical Design (ISPD) and the Tau workshop, and Program Chair for the International Conference on VLSI Design in India. He has received six conference Best Paper awards, a Best Poster Award, an ICCAD 10-year Retrospective Most Influential Paper Award, the SRC Technical Excellence award and the SIA University Researcher Award. He is a Fellow of the IEEE.