Stochastic Variational Analysis of Large Power Grids Considering Intra-die Correlations

Praveen Ghanta
EE Dept.,
Arizona State University,
Tempe, AZ
praveen.ghanta@asu.edu

Sarma Vrudhula
CSE Dept.
Arizona State University,
Tempe, AZ
sarma.vrudhula@asu.edu

Sarvesh Bhardwaj
EE Dept.,
Arizona State University,
Tempe, AZ
sarvesh.bhardwaj@asu.edu

Rajendran Panda
Freescale Semiconductor Inc.,
Austin, TX
Rajendran.Panda@asu.edu

ABSTRACT

For statistical timing and power analysis that are very important problems in the sub-100nm technologies, stochastic analysis of power grids that characterizes the voltage fluctuations due to process variations is inevitable. In this paper, we propose an efficient algorithm for the variational analysis of large power grids in the presence of a significant number of Gaussian intra-die process variables that are correlated. We consider variations in the power grid's electrical parameters as spatial stochastic processes and express them as linear expansions in an orthonormal series of random variables using the Karhunen-Loève (KLE) method. The voltage response is then represented as an orthonormal polynomial series and the coefficients are obtained optimally using the Galerkin method. We propose a novel method to separate the stochastic analysis for the random variables that effect only the inputs (e.g, drain currents) and for those that effect the system parameters as well (e.g., conductance, capacitance). We show that this parallelism can result in significant speed-ups in addition to the speed-ups inherent to Galerkin-based methods. Our analysis has been applied to several industrial power grids and the results show speed-ups of up to two orders of magnitude over Monte Carlo simulations for comparable accuracy.

Categories and Subject Descriptors: B.7.2 [Integrated Circuits]: Design Aids—Simulation, Verification

General Terms: Algorithms, Performance, Verification

Keywords: Power Grids, Process Variations, Correlations, Stochastic Analysis, Polynomial Chaos, Orthonormal Polynomials

1. INTRODUCTION

In state-of-the-art VLSI designs, large power distribution networks that span across the entire chip area and across several layers are widely used. With technology scaling and the resulting increase in the density of on chip transistors, there is a continuous reduction in supply voltages (about $1 - 1.2$V for sub-100nm) while the drain currents drawn from the power grids are increasing. This can result in severe voltage drops in power grids and can have a detrimental impact on the signal integrity and reliability. The sheer size and the complexity of the power grids presents significant challenges in it's analysis and optimization, which have been well addressed by an established body of literature [3–5, 7, 9, 12].

The most important challenge posed by sub-100nm IC technologies is the increasing uncertainty in the performance of CMOS ICs due to manufacturing process variations [11, 15]. Given the high performance sensitivity of today's ICs to even minor voltage fluctuations, it is important to characterize the impact of process variations on power grids in addition to it's analysis under nominal process conditions [1, 10]. As in the case of statistical timing analysis, process corner based analysis for power grids may no longer be accurate or even computationally feasible given the increasing number of process variables due to intra-die and inter-die variations.

Two recent works have considered the impact of process variations on power grids. Ferzli et al. [10] propose a method to obtain the mean and bounds on the variance of voltage drops considering intra-die process variations in system inputs (drain leakage currents), which they assume are mutually independent. By combining statistical importance sampling with random walks, Li [1] proposed a method to efficiently perform localized sensitivity analysis of a few power grid nodes to process variations. While Li [1] considers variations in both the system parameters (conductance of power grid lines) and inputs (drain currents), at certain stages of a chip design cycle like the timing signoff a complete variational analysis of all the power grid nodes is desirable. Also in the presence of intra-die variations, locating a few critical nodes of interest may be a hard task. Both the previous works, perform only a static analysis of the power grids.

The outline of our paper is as follows. In Section 2, we discuss the core contributions of our work. Section 3 contains the problem definition. In Section 4, we discuss our approach and the solution to our problem in detail. In Section 5 we discuss the computational cost. The experimental results are shown in Section 6 and conclusions are presented in Section 7.

2. OUR CONTRIBUTIONS

In this paper, we propose a method to efficiently compute the stochastic transient voltage response of power grids in the presence of correlated intra-die process variations in both the system param-

DAC 2006, July 24–28, 2006, San Francisco, California, USA.

eters viz., Conductance and Capacitance and the system inputs viz., the drain currents. Due to the manufacturing variations in the interconnect width W and thickness T; and device width (W_d) and threshold voltage (V_{th}), the electrical parameters G, C are modeled as continuous parameter (spatial) stochastic processes. From the existing literature, we select an analytical correlation function in the spatial coordinates (x,y) for these parameters. We then propose a novel and an exact method to express the electrical parameters as a linear orthogonal series in the spatial coordinates and an orthonormal set of random variables.

We then represent the voltage response as an orthonormal polynomial series in the random variables. The expansion is truncated to the required order of accuracy and the expansion coefficients are obtained using the Galerkin method which ensures mean-square optimality. We also propose a novel way to explicitly separate the stochastic analysis for the random variables that effect only the system inputs (e.g, drain currents) and those that effect only the system parameters (e.g, Conductance, Capacitance). We show that this parallelism results in significant additional speed-ups and provides an attractive alternative to the computationally expensive Monte Carlo simulations[1].

3. PROBLEM DEFINITION

The power grid is modeled as an RC network. The metal V_{DD} and ground interconnects and vias are modeled as RC circuits. The external power supplies are modeled as ideal voltage sources between the V_{DD} and the ground lines in series with the contact resistances of the supply pins. The CMOS on chip functional blocks that span across the chip are modeled as transient current sinks between the V_{DD} nodes and ground nodes, in parallel with the gate capacitances of the non-switching devices in the functional blocks. The transient profiles are typically obtained through simulation of the CMOS blocks for a large sequence of input patterns at full supply rails.

The time domain response of the power grid can be described by the MNA equation:

$$G V(t) + C \frac{dV(t)}{dt} = U(t) \tag{1}$$

where G, C are the conductance and capacitance matrices and $V(t)$ is the voltage response. $U(t) = i(t) + G_1 V_{DD}$ is the known excitation; vector $i(t)$ represents the functional block drain currents; G_1 is a diagonal matrix that consists of series resistances of the power grid V_{DD} pin contacts.

The dependence of the grid's electrical parameters on the physical dimensions of the CMOS devices and interconnects can be described as shown in Table 1. W, T are the width and thickness of power grid interconnects, C_{GS} and C_{GD}) are device capacitances of the non-switching and decoupling CMOS devices, W_d, L_{eff} are the device width and channel length, and V_{th} is the device threshold voltage. The current $i(t)$ consists of the CMOS on state currents i_{DS} and the offstate leakage currents i_{leak}. In sub-100nm technologies, i_{leak} can be as large as i_{DS} [10]. i_{leak} is primarily an exponential function of the threshold voltage V_{th} and i_{DS} is dependent on the W_d, C_{ox} and L_{eff} of the devices. The process variables we consider in this paper are shown in column 3 of Table 1. Process variations lead to significant variations in each of the physical parameters (Table 1) and the electrical parameters G, C and $U(t)$. Both the physical and electrical variations in general have *strong* intra-die, inter-die and random components. Inter-die variations generally

[1]An extended version of the paper considering 2^{nd} order expansions in all the process variables is available at http://veda.eas.asu.edu/people/praveen/pubs/DAC06-extended.pdf

Table 1: Relationship between electrical and physical params. [8]

Electric params.	Physical relation	Params. considered
G	$\propto \frac{WT}{\rho L}$	W, T
C	$\propto P_A\,(C_{GS}+C_{GD})$ $\propto P_A W_d L_{eff} C_{ox}$	 W_d
i_{DS}	$\propto \frac{W_d}{L_{eff}}$	W_d
i_{leak}	$\propto e^{-V_{th}/S_0}$	V_{th}

lead to a shift in the mean values of the parameters. Intra-die variations can arise due to a number of environment, layout and process factors [11] and are thus modeled as spatial stochastic processes that exhibit high frequency trends across a chip. In this paper, we consider the impact of correlated Gaussian intra-die variations. We assume that the process variables in different layers are independent of each other.

In the presence of intra-die variations W, T, W_d and V_{th} are also a function of the location on a chip. For a single metal layer, they thus become a function of (x,y) coordinates, $W(x,y)$, $T(x,y)$, $W_d(x,y)$ and $V_{th}(x,y)$, where $D = \{(x,y) \in \Re^2 | |x| \leq 0.5a, |y| \leq 0.5a\}$ denotes the chip area and 'a' is the length of the square chip. Thus, W, T, and W_d are spatial stochastic processes that vary across the chip area and over the fabrication sample space.

In general, if Ω denotes the fabrication sample space, for every $\omega \in \Omega$, $Z(x,y,\omega) : D \times \omega \rightarrow R$ denotes a random process. Let $Z_S(x,y,\omega) = (W(x,y,\omega), T(x,y,\omega), W_d(x,y,\omega))$ denote the vector of random processes that impact the system parameters and the inputs. And let $Z_I(x,y,\omega) = (V_{th}(x,y,\omega))$ denote a vector of random processes that impact only the system inputs. In the presence of intra-die process variations, the variational power grid analysis problem can be described in the form of MNA equations as:

$$\begin{aligned} & G(Z_S(x,y,\omega))\, V(t, Z_S(x,y,\omega), Z_I(x,y,\omega)) \\ + \;& C(Z_S(x,y,\omega))\, \frac{dV}{dt}(t, Z_S(x,y,\omega), Z_I(x,y,\omega)) \\ = \;& U(t, Z_S(x,y,\omega), Z_I(x,y,\omega)) \end{aligned} \tag{2}$$

Equation (2) is a stochastic differential equation dependent on a vector of random variables $Z_S(x,y,\omega)$, $Z_I(x,y,\omega)$ and time t. Further, the excitation vector $U(t, Z_S(x,y,\omega), Z_I(x,y,\omega))$ has both deterministic and random components. For each manufacturing outcome ω, $V(t, Z_S(x,y,\omega), Z_I(x,y,\omega))$ denotes the fixed response of the system for that particular manufacturing outcome.

Based on the available literature [10, 14, 15], the models for G, C and $U(t)$ in the parameters considered (W, T, W_d and V_{th}) can generally be approximated as,

$$\begin{aligned} & W(x,y,\omega) = \bar{W} + \Delta W(x,y,\omega), \quad T(x,y,\omega) = \bar{T} + \Delta T(x,y,\omega) \\ & W_d(x,y,\omega) = \bar{W}_d + \Delta W_d(x,y,\omega), \quad V_{th}(x,y,\omega) = \bar{V}_{th} + \Delta W_d(x,y,\omega) \end{aligned} \tag{3}$$

$$\begin{aligned} G(Z_S(x,y,\omega)) &= \bar{G}_a + G_b\, \Delta W(x,y,\omega) + G_c \Delta T(x,y,\omega) \\ C(Z_S(x,y,\omega)) &= \bar{C}_a + C_b \Delta W_d(x,y,\omega) \\ U(t, Z_S(x,y,\omega), Z_I(s,y,\omega)) &= \bar{U}_a(t) + U_b(t) \Delta W_d(x,y,\omega) \\ &+ U_c\, exp(-\Delta V_{th}(x,y,\omega)/S_0) \\ G_a = f(\bar{W}, \bar{T}), \quad C_a &= f(\bar{W}_d), \quad U_a(t) = f(\bar{W}_d, \bar{V}_{th}) \end{aligned} \tag{4}$$

To solve Equation (2), we first need to express each element of the spatial stochastic process vectors $Z_S(x,y,\omega)$, and $Z_I(x,y,\omega)$, as a function of an uncorrelated set of random variables. But, for that we need a description of the correlation (covariance) function relating two locations $C((x_1,y1),(x_2,y_2))$ on a chip. Once such a func-

tion is given, we can use the Karhunen-Loève expansion to express the stochastic processes as a series of uncorrelated variables.

4. OUR APPROACH

In this section, we first present a brief overview of the theory behind our approach. The only assumption we need in our analysis is that the response $V(t, Z_S(x,y,\omega), Z_I(x,y,\omega))$ is a second order process i.e. $\forall t, V(t, Z_S(x,y,\omega), Z_I(x,y,\omega))$ has a finite variance. This is a robust assumption for all practical power grids where the parameters are bounded. Our approach first represents the spatial stochastic processes $Z_S(x,y,\omega)$ and $Z_I(x,y,\omega)$ as a series of uncorrelated random variables and eigenfunctions in x and y given a correlation function $C((x_1,y_1),(x_2,y_2))$ using a method called the Karhunen-Loève expansion [2]. The second step involves expanding the voltage response as an orthogonal polynomial series in these uncorrelated random variables. Let's define a radial coordinate $r_i = \sqrt{x_i^2 + y_i^2}$, which is the distance from the reference origin (0,0) on a chip. The Karhunen-Loève expansion can be stated as follows.

4.1 Karhunen-Loève expansion (KLE)

Consider the random process $\Delta W(r,\omega) : D \times \omega \rightarrow R$, indexed on the domain $D = \{\sqrt{x^2+y^2} | |x| \leq 0.5a, |y| \leq 0.5a\}$, 'a' being the side length of the square chip. We assume the process has a zero mean and a finite variance, $E[\Delta W(r,\omega)]^2$, that is bounded $\forall\ r \in D$. The process can then be expressed as a mean-square convergent series as [2]

$$\Delta W(r,\omega) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\, \xi_{W_i}(\omega)\, f_{W_i}(r) \tag{5}$$

in which λ_i and $f_i(r)$ are the eigenvalues and eigenfunctions of the covariance function $C(r_1,r_2)$. Mean square convergence implies convergence in the norm $\| \cdot \|$, where $||f|| = < f, f > = E(f.f)$ (Expectance over f^2). The explicit dependence of ξ_{W_i} on ω will not be shown further and $\xi_{W_i}(\omega)$ will always be further referred to as ξ_{W_i}. $C(r_1,r_2)$ is by definition bounded, symmetric and positive definite. The eigenvalues and eigenfunctions of the above equation are the solutions of the Fredholm integral equation of the second kind [13] are given by

$$\int_D C(r_1,r_2)\, f_{W_i}(r_1) dr_1 = \lambda_i\, f_{W_i}(r_2) \tag{6}$$

The set of variables ξ_{W_i} are zero mean, orthonormal random variables, which implies that $E(\xi_{W_i}) = 0$ and $E(\xi_{W_i}\, \xi_{W_j}) = \delta_{ij}$, where E is the expectance. *We assume in this paper the spatial processes $\Delta W(r,\omega)$ are Gaussian, and for such a process, the random variables $\xi_{W_i}(\omega)$ are also Gaussian.* Karhunen-Loève expansion in Equation (5) is bi-orthonormal, i.e., in addition to ξ_i being orthonormal, the deterministic basis functions $f_{W_i}(r)$ are also orthonormal, which implies that $\int_D f_{W_i}(r)\, f_{W_i}(r) dr = \delta_{ij}$.

Intra-die variations are generally analyzed by dividing a chip in to several grids and then considering a separate set of variables for each grid. The unique advantage of the Karhunen-Loève expansion (see Equation (5)) is that a *very fine* grid can be considered to the extent of considering every resistor and capacitor to be in a different grid and yet limit the number of variables to a small number. A discretized version of this algorithm will be the principal component analysis (PCA), where the number of grids that can be considered are limited by the computational complexity of the algorithm and the significant numerical error due to discretization.

A experimentally verified [18] characteristic of correlation functions is that the correlation decreases to a minimum as the distance increases between two locations on a chip. An exponential [2] function can be efficiently used to model such a correlation, which is what we consider in this paper. We consider the covariance function as $C(r1,r2) = e^{-c|r_1 - r_2|}$. We fit the coefficient c based on data provided in [18]. *We consider that each resistor and capacitor in our grid is in a different location and hence has a different KL expansion.* The processes $\Delta W(x,y,\omega), \Delta T(x,y,\omega), \Delta W_d(x,y,\omega)$ and $\Delta V_{th}(x,y,\omega)$ can be expressed using the expansion in Equation (5), where

$$\lambda_{W_i} = \left\{ \frac{2c}{K_{W_i}^2 + c^2} \quad \text{for odd i} \right\}, \quad \left\{ \frac{2c}{{K_{W_i}^*}^2 + c^2} \quad \text{for even i} \right\} \tag{7}$$

and K_{W_i} and $K_{W_i}^*$ are respectively the solutions of the transcendental equations $c - Ktan(K\sqrt{2}a) = 0$ and $c + Ktan(K\sqrt{2}a) = 0$. Also,

$$\begin{aligned} f_{W_i}(r) &= cos(K_{W_i} r) / \sqrt{\sqrt{2}a + \frac{sin(2K_{W_i}\sqrt{2}a)}{2K_{W_i}}} \quad \text{for odd i} \\ &= sin(K_{W_i}^* r) / \sqrt{\sqrt{2}a - \frac{sin(2K_{W_i}^*\sqrt{2}a)}{2K_{W_i}^*}} \quad \text{for even i} \end{aligned} \tag{8}$$

The Karhunen-Loève expansion in $\Delta W(x,y,\omega)$ is truncated to a required order N based on a user defined threshold for the ratio $\lambda_N / \sum_{k=1}^{N} \lambda_k$. For all the processes, similarly we have after truncation,

$$\begin{aligned} \Delta W(x,y,\omega) &= \sum_{i=1}^{N_1} \sqrt{\lambda_{W_i}}\, \xi_{W_i}\, f_{W_i}(r) \\ \Delta T(x,y,\omega) &= \sum_{i=1}^{N_2} \sqrt{\lambda_{T_i}}\, \xi_{T_i}\, f_{T_i}(r) \\ \Delta W_d(x,y,\omega) &= \sum_{i=1}^{N_3} \sqrt{\lambda_{Wd_i}}\, \xi_{Wd_i}\, f_{Wd_i}(r) \\ \Delta V_{th}(x,y,\omega) &= \sum_{i=1}^{N_4} \sqrt{\lambda_{Vth_i}}\, \xi_{Vth_i}\, f_{Vth_i}(r) \end{aligned} \tag{9}$$

In this paper, we truncate the variables $\Delta W(x,y,\omega)$, $\Delta T(x,y,\omega)$, $\Delta W_d(x,y,\omega)$, $\Delta V_{th}(x,y,\omega)$ to order 5 based on tradeoffs between accuracy and computational complexity. Thus in our paper, we have $N_1 = N_2 = N_3 = N_4 = 5$. Figure 5 shows how the correlation of the center of a chip falls off as we move away, for an actual exponential covariance function and the covariance function using a 5 term Karhunen-Loève truncation. Truncating the above expansions and plugging these Equations (5) in Equations (3), we have G, C, and U(t) expressed as a function of the random variables as,

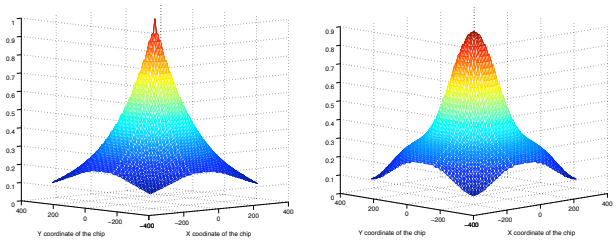
$$\begin{aligned} G(Z_S(x,y,\omega)) &= \bar{G}_a + \sum_{i=1}^{N_1} G_{b_i}\, \xi_{W_i} + \sum_{j=1}^{N_2} G_{c_i} \xi_{T_i} \\ C(Z_S(x,y,\omega)) &= \bar{C}_a + \sum_{i=1}^{N_3} C_{b_i} \xi_{W_d} \\ U(t, Z_S(x,y,\omega), Z_I(s,y,\omega)) &= \bar{U}_a(t) + \sum_{i=1}^{N_3} U_{b_i}(t)\, \xi_{Wd_i} \\ &+ U_c\, exp(-\sum_{j=1}^{N_4} \xi_{Vth_j} / S_0) \end{aligned} \tag{10}$$

[2]For other covariance functions, the eigenfunctions f_{W_i} are modeled using a deterministic orthonormal polynomial basis, e.g., Trigonometric polynomials, Legendre polynomials, etc., and Galerkin method is used to solve Equation (6) numerically. More details in the extended version of this paper mentioned in footnote 1

$U(t,Z_S(x,y,\omega),Z_I(s,y,\omega))$ can be simplified considering a second order Hermite series [3] in $\vec{\xi}_{Vth}$ as

$$U(t,Z_S(x,y,\omega),Z_I(s,y,\omega)) = \bar{U}_a(t) + \sum_{i=1}^{N_3} U_{b_i}(t)\,\xi_{Wd_i} + \sum_{j=1}^{N_4} U_{d_j}\xi_{Vth_j} + \sum_{k=1}^{N_4} U_{e_k}(\xi^2_{Vth_k} - 1) \quad (11)$$

$$\text{Let } \vec{\xi} = (\vec{\xi}_W, \vec{\xi}_T, \vec{\xi}_{W_d}, \vec{\xi}_{V_{th}})^T$$
$$\text{where } \vec{\xi}_W = (\vec{\xi}_{W_1},\cdots,\vec{\xi}_{W_{N_1}})^T, \quad \vec{\xi}_T = (\vec{\xi}_{T_1},\cdots,\vec{\xi}_{T_{N_2}})^T$$
$$\vec{\xi}_{W_d} = (\vec{\xi}_{Wd_1},\cdots,\vec{\xi}_{Wd_{N_3}})^T, \quad \vec{\xi}_{Vth} = (\vec{\xi}_{Vth_1},\cdots,\vec{\xi}_{Vth_{N_4}})^T \quad (12)$$



(a) (b)

Figure 1: Center of Chip Correlation with other points on the chip (a) Exponential Co-variance function (b) Co-variance function with 5 term Karhunen-Loève approximation

4.2 Orthogonal polynomial expansion

We now represent the voltage response as a mean-square convergent series of orthonormal polynomials in the Gaussian random variables $\vec{\xi}$ from KLE. The theory behind the existence of such an expansion is now called the polynomial chaos theory [13]. In this paper, we focus more on it's application and refer the interested readers to [13] for a detailed discussion. The voltage response $V(t,\vec{\xi})$ can be represented as a orthonormal polynomial series in Hermite Polynomials of $\vec{\xi}$ as,

$$V(t,\vec{\xi}) = \sum_{k=0}^{\infty} a_k(t)\,H_k(\vec{\xi}) \quad (13)$$

An order p Hermite polynomial is defined as,

$$H_p(\{\xi_{a_1},\xi_{a_2},\cdots,\xi_{a_p}\}) = (-1)^p e^{\frac{1}{2}\vec{\xi}^T\vec{\xi}} \frac{\partial^p}{\partial\xi_{a_1}\partial\xi_{a_2}\cdots\partial\xi_{a_p}} e^{-\frac{1}{2}\vec{\xi}^T\vec{\xi}} \quad (14)$$

where $\{\xi_{a_i}\}$ are any set of p variables chosen from the vector $\vec{\xi}$ allowing repetitions (see Equation (11)).

Equation (13) is a general representation of the random process $V(t,\vec{\xi})$ in terms of the orthonormal basis functions $H_k(\vec{\xi})$ of the random variables. The different optimal polynomial families for other distributions are identified from the Askey scheme of polynomials [13].

4.3 Galerkin Method

We have the MNA equation for power grids as $G(\vec{\xi})\,V(t,\vec{\xi}) + C(\vec{\xi})\,\frac{dV}{dt}(t,\vec{\xi}) = U(t,\vec{\xi})$. Let r be the number of random variables or the length of vector $\vec{\xi}$. For computational purposes, we need to truncate the expansion in $V(t,\vec{\xi})$ to an order p and determine the coefficients $a_k(t)$.

$$\widehat{V}(t,\vec{\xi}) = \sum_{k=0}^{NT(p,r)} a_k(t)H_k(\vec{\xi}) \quad (15)$$

where the number of terms NT is a function of p, r

$$NT(p,r) = \sum_{j=0}^{p} \binom{r-1+j}{j} \quad (16)$$

The error then between the actual response and the approximated response is

$$\Delta_p(t,\vec{\xi}) = U(t,\vec{\xi}) - G(\vec{\xi})V(t,\vec{\xi}) - C(\vec{\xi})\frac{d\widehat{V}}{dt}(t,\vec{\xi}) \quad (17)$$

The Galerkin method states that the best approximation to the response $V(t,\vec{\xi})$ is the one that is orthogonal to the error term $\Delta_p(t,\vec{\xi})$. The orthogonality is w.r.t the underlying norm $\|\,.\,\|$, where $||f|| =< f,f >= E(f.f)$. Appying this method, the deterministic coefficients $a_k(t)$ can be obtained by solving the following system of equations.

$$\Big\langle \Delta_p(t,\vec{\xi}), H_k(t,\vec{\xi}) \Big\rangle = E[\Delta_p(t,\vec{\xi}), H_k(t,\vec{\xi})] = 0,$$
$$\text{for } k = 0,..,(NT(r,p)-1) \quad (18)$$

Note that the expectation in Equation (18) is over the basis of the random variables $\vec{\xi}$. Hence the resulting system of equations from Equation (18) is a deterministic system in the unknown coefficients $\{a_k\}$ and is completely independent of the random variables $\vec{\xi}$. We thus simplified our analysis from trying to solve a stochastic system to solving a deterministic system. The deterministic system is $NT(p,r)$ times larger than the original stochastic system.

4.4 Separation of Analysis

We know that in the random variable vector $\vec{\xi}$, $\vec{\xi}_{Vth}$ only impacts the inputs and vector $\vec{\xi}_{W_d}$ impacts both the system parameters and inputs. A brute force approach would be to just consider Hermite polynomials in all the vectors, $H_k(\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d},\vec{\xi}_{Vth})$ as is done in Equation (15). After applying the Galerkin method, this would result in a large deterministic system as the number of random variables $r = (N_1+N_2+N_3+N_4)$ (see Equation (16)). Especially so, since in this paper we use an expansion of order 1 (p = 1)in variables of vector $\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d}$ and of order 2 (p = 2) in variables of vector $\vec{\xi}_{Vth}$. To avoid this huge cost, we now separate the analysis for these two variable vectors based on the orthogonality of the polynomial chaos expansion. From Equations (11) and (12), $U(t,Z_S(x,y,\omega),Z_I(x,y,\omega))$ can be re-expressed as,

$$U(t,\vec{\xi}_{W_d},\vec{\xi}_{Vth}) = U_0(t) + \sum_{i=1}^{NT(1,N_3)} U_i(t)H_i(\vec{\xi}_{W_d}) + \sum_{k=1}^{NT(2,N_4)} U_k(t)H_k(\vec{\xi}_{Vth})$$

Based on the results of Polynomial chaos theory [13], we can re-express $V(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d},\vec{\xi}_{Vth})$ as

$$V(t,\vec{\xi}_{W_d},\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{Vth}) = \sum_{k=0}^{NT(2,N_4)} V_k(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d},)H_k(\vec{\xi}_{Vth}),$$
$$V_k(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d}) = \sum_{l=0}^{NT(1,N_1+N_2+N_3)} b_l(t)H_l(\vec{\xi}_{W_d},\vec{\xi}_W,\vec{\xi}_T) \quad (19)$$

[3] Stochastic Response Surface Methods [17] employing SPICE runs are used to obtain an accurate fit

We thus have from the MNA equations and Equation (19),

$$\begin{aligned} & G(\vec{\xi}_W,\vec{\xi}_T)\left(\sum_{k=0}^{NT(2,N_4)} V_k(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d})H_k(\xi_i)\right) \\ + \quad & C(\xi_{W_d})\left(\sum_{k=0}^{NT(2,N_4)} \frac{dV_k(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d})}{dt}H_k(\xi_W,\vec{\xi}_T,\vec{\xi}_{W_d})\right) \\ & = U_0(t)+\sum_{i=1}^{NT(1,N_3)} U_i(t)H_i(\vec{\xi}_{W_d})+\sum_{k=1}^{NT(2,N_4)} U_k(t)H_k(\vec{\xi}_{Vth}) \end{aligned} \tag{20}$$

From the orthogonality of the basis functions $H_k(\xi_i)$, it follows that the analysis of Equation (20) can be broken up in to $NT(2,N_4)$ individual systems given below, *which can all be solved in parallel*.

$$\begin{aligned} G(\vec{\xi}_W,\vec{\xi}_T)V_0(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d}) \quad + \quad & C(\xi_{W_d})\frac{dV_0(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d})}{dt} \\ = \quad & U_0(t)+\sum_{i=1}^{NT(1,N_3)} U_i(t)H_i(\vec{\xi}_{W_d}) \\ G(\vec{\xi}_W,\vec{\xi}_T)V_k(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d}) \quad + \quad & C(\xi_{W_d})\frac{dV_k(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d})}{dt} \\ = \quad & U_k(t) \\ \text{for } k = 1,\cdots,NT(2,N_4) & \end{aligned} \tag{21}$$

Now each of the above individual systems can be analyzed using the Galerkin method described in this section. Not only have we decoupled the analysis of both $\{\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d}\}$ and $\vec{\xi}_{Vth}$, we can now choose expansions of different orders in them.

4.5 Solving separate systems

As mentioned in Section IV (A), we consider a 1^{st} order expansion for $V_k(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d})$ in $\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d}$, $\forall k$. Let's consider the first equation for $V_0(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d})$ in the system of Equations (21) as an example to demonstrate the solution process for each of the independent equations in Equation (21). We have from Equation (19), a first order expansion in $\vec{\xi}_{W_d},\vec{\xi}_W,\vec{\xi}_T$

$$\begin{aligned} V_0(t,\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d}) \quad = \quad & \sum_{l=0}^{NT(1,N_1+N_2+N_3)} b_l(t)H_l(\vec{\xi}_W,\vec{\xi}_T,\vec{\xi}_{W_d}) \\ = \quad & b_0(t)+\sum_{i=1}^{N_1} b_i(t)\xi_{W_i}+\sum_{j=N_1+1}^{N_2} b_j(t)\xi_{T_j} \\ + \quad & \sum_{k=N_2+1}^{N_3} b_k(t)\xi_{Wd_k} \end{aligned} \tag{22}$$

Using the equations for $G(\vec{\xi}_W,\vec{\xi}_T)$ and $C(\vec{\xi}_{W_d})$ from Equation (10) and (11) and following the Galerkin procedure [13] described in Section 4.3, we can obtain the deterministic system (from the stochastic system) in the Laplace domain as,

$$\begin{aligned} (G_a+sC_a)\,b_0(s) \quad = \quad & U_0(s)-\sum_{i=1}^{N_1} G_{W_i}b_i(s)-\sum_{j=1}^{N_2} G_{T_j}b_{j+N_1}(s) \\ - \quad & \sum_{k=1}^{N_3} s\,C_{Wd_k}b_{k+N_1+N_2}(s) \\ (G_a+sC_a)\,b_i(s) \quad = \quad & -G_{W_i}b_0(s) \quad \forall\, i\,=\,1,\cdots,N_1 \\ (G_a+sC_a)\,b_{i+N_1}(s) \quad = \quad & -G_{T_i}b_0(s) \quad \forall\, i\,=\,1,\cdots,N_2 \\ (G_a+sC_a)\,b_{i+N_1+N_2}(s) \quad = \quad & U_i(s)-sC_{Wd_i}b_0(s) \quad \forall\, i\,=\,1,\cdots,N_3 \end{aligned} \tag{23}$$

The above system of equations can be solved efficiently in time domain by the following iterative steps: I) Discretize the system using Backward Euler method with a fixed time step, II) At time $t=0$, initialize $b_0(t)=b_i(t)=0$, III) Using the old b_0 and b_i values, repeat through the loop of equations in (23) for a few steps. In practice, we have observed that repeating for three steps is sufficient for accuracy up to 2^{nd} decimal in mean and variance in our analysis.

The other individual systems can be solved similarly, using the Galerkin method and the iterative procedure. Once the individual system of equations in (21) are solved, the voltage response in all the variables can be stitched together using the equations in (19).

5. COMPUTATIONAL COMPLEXITY

The primary cost associated with our method is solving the iterative set of equations in Equation (23). On average three recursive iterations are found to be sufficient for 2^{nd} decimal accuracy in mean and variance in solving the equations in (23). For each step of the transient analysis, the cost of the iterative method includes the cost of LU factorization of the original system matrix (G_a+sC_a), the number of variables times the cost of each LU solve, and the cost of having to solve all the independent systems in Equation (21). Let's say we have N_1 variables in W, N_2 in T, N_3 in W_d and N_4 in V_{th} after Karhunen-Loève truncation. And let the cost of 1 LU solve be $C_1(l)$ (forward + backward substitution) and one LU factorization be $C_2(l)$ and let the number of transient steps be K. We consider an order 1 expansion in W,T,W_d, and order 2 expansion in V_{th}. Hence, the cost of our method is approximately $(1.5K(N_1+N_2+N_3)C_1(l)(N_4^2+3N_4+1)+C_2(l))$ when the software is run serially on a single processor. When the software is executed on a parallel machine or a cluster with M nodes, the cost of the method due to the simultaneous solution of the independent systems becomes $(\frac{1.5K(N_1+N_2+N_3)C_1(l)(N_4^2+3N_4+1)}{M}+C_2(l))$. If the input variables (V_{th}) are assumed to be mutually independent and only a static (DC) analysis is performed as in [1, 10], then the cost is $(3(N_1+N_2+N_3)C_1(l)N_4+C_2(l))$, when executed serially .

6. EXPERIMENTAL RESULTS

We have extensively verified the accuracy of our approach for the *transient analysis* of several power grids. As mentioned earlier, the *intra-die Gaussian process variations* we considered are 3σ variations of 30 % in W, T, W_d, and 25 % in V_{th}. We assumed an exponential correlation function and used the Karhunen-Loève expansion to truncate the number of variables in W, T and W_d to 5 each (with order 1 expansions), and in V_{th} to 5 (with an order 2 expansion). We thus have 20 variables for all our experiments. We consider a linear model for the variations in G, C and modeled *every* resistor and capacitor in the circuit to be at a different location. The drain current profiles used for the transient analysis of the power grid were such that the peak drop in the voltage at any grid node was less than 20 % of the V_{DD}. Backward Euler method with a fixed time step was used for transient simulation and 10 transient steps were considered. *For a larger number of transient time steps, the key to using our method is to exploit the nice parallelism available in our approach.*

The results obtained from our analysis on 6 industrial power grids are shown in Table 2. All the results were run on a 1-4 processors of HP/Compaq Alpha cluster with 1.15 GHz Alpha EV7 processors. The matrix package UFsparse [16] (including UMFPACK) was used for LU factorization and subsequent solves and FORTRAN BLAS routines were used for speed-ups. For the first 5 grids, we ran 5000 Monte Carlo simulations in all the process variables to compare the accuracy of our approach. For the grid with

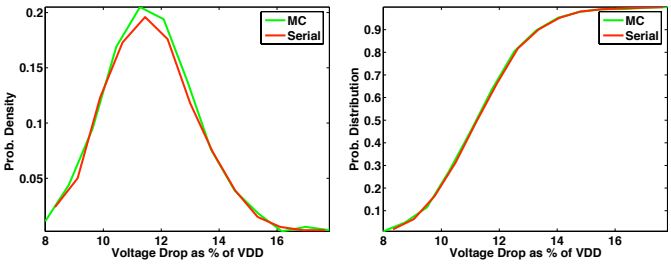
Table 2: Results from our approach (serial) and 5000 Monte Carlo (MC) simulations

Size (# nodes)	Ave. % Error in μ	Max. % Error in μ	Ave. % Error in σ	Max. % Error in σ	CPU time MC	CPU time serial	Transient Speed-up (serial solve/MC)	Static Speed-up (serial solve/MC)
25K	0.025	0.070	2.234	3.341	3.0 hrs	13.3 min	13	75
35K	0.043	0.573	1.896	12.262	2.3 hrs	10.7 min	13	81
50K	0.108	0.176	4.618	6.283	7.7 hrs	31.8 min	15	73
91K	0.041	0.379	5.651	12.569	7.3 hrs	31.3 min	14	79
350K	0.253	0.717	1.302	11.714	35.3 hrs	37.5 min	57	191
1.3M	MC infeasible, projected MC time from few runs - 50.0 hrs					2.0 hrs	25	122

1.3 Million nodes, 5000 Monte Carlo (MC) samples are too expensive and hence we only obtain the MC time from a few sample runs. We obtain the average, maximum errors in Mean and Variance of our approach vs. MC simulations across all the transient simulation steps, along with the speed-ups obtained over MC for transient analysis. We also show the speed-up of our approach in case only a static (DC) variational analysis of the power grids is performed along with the assumption that the input variables (variations in V_{th}) are mutually independent, as addressed in the previous works [1, 10].

The time consumed for our approach when executed serially vs. the time taken for Monte Carlo is also shown in the Table 2. The speed-ups obtained for the transient and static analysis were in the range of 10X-200X over MC simulations, with speed-ups (typically) increasing with the size of the power grid. A parallel execution of our transient and static analysis algorithms on 4 processors produced a *further speed-up* of about 4X on average (40X-800X in all), there by showing the utility of separating the analysis for variables that only impact the system inputs. Also, on average over all grids, the difference between the mean voltage drop with process variations was close to the the voltage drop with no variations. However, on average across all grids the $\pm\sigma$ variation in the voltage drops at the power grid nodes varied from $\pm 20-30\%$ of the voltage drops with no variations .

To further verify the accuracy of our analysis, we plotted the voltage drop distributions of several nodes of the grids at random and observed a good match in all the cases. As an example, the probability density and distribution of voltage response for a random grid node of the 91K power grid is shown in Figures 2(a) and (b).



(a) (b)

Figure 2: Voltage distribution at a random node

7. CONCLUSIONS

We proposed an efficient algorithm for the stochastic variational analysis of large power grids in the presence of intra-die process correlations. We consider variations in the power grid's electrical parameters as spatial stochastic processes and model the voltage response as an orthonormal polynomial series in these random variables. We propose an efficient means to separate the stochastic analysis for the random variables that effect only the inputs and for those that effect the system parameters, which results in significant speed-ups. We applied our analysis exhaustively to several power grids and our results show speed-ups of up to two orders of magnitude over Monte Carlo simulations, while showing a very good match in accuracy.

8. ACKNOWLEDGMENTS

This work was carried out under the auspices of the National Science Foundation's Grants #CCR-0205227 and #EEC-9523338. It was conducted at the National Science Foundation's State/Industry/University Cooperative Research Centers' (NSF-S/IUCRC) Center for Low Power Electronics (CLPE) which is supported by the NSF, the State of Arizona, and an industrial consortium. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

9. REFERENCES

[1] P. Li. Variational analysis of large power grids by exploring statistical sampling sharing and spatial locality., *In ACM/IEEE ICCAD, Nov. 2005, pp. 644-650.*

[2] M. Loeve. Probability theory., *4th Edition. NY. Springer Verlag*

[3] A. Dharchoudhury et al. Design and analysis of power distribution networks in powerPC microprocessors., *In ACM/IEEE DAC, June 1998, pp. 738-743.*

[4] D. Stark, and M. Horowitz. Techniques for calculating currents and voltages in VLSI power supply networks., *IEEE TCAD, vol. CAD-9, Feb. 1979, pp. 126-132.*

[5] M. Zhao et al. Hierarchical analysis of power distribution networks. *In ACM/IEEE DAC, June 2000, pp. 481-486.*

[6] H. Qian, S. R. Nassif, S. S. Sapatnekar. Random walks in a supply network., *In ACM/IEEE DAC, June 2003, pp. 93-98.*

[7] H. Kriplani, F. Najm, and I. Hajj. Pattern independent minimum current estimation in power and ground buses of CMOS VLSI Circuits., *In IEEE TCAD, vol. 14, Aug.1995, pp. 998-1012.*

[8] R. Chaudhry et al. Design and analysis of power distribution networks with accurate RLC models., *In IEEE Int'l Conf. on VLSI Design, Jan. 2000, pp. 151-155*

[9] S. Pant, D. Blaauw, V. Zolotov, S. Sundareswaran, and R. Panda. A stochastic approach to power grid analysis., *In ACM/IEEE DAC, June 2004, pp. 171-176.*

[10] I. A. Ferzli, and F. N. Najm. Statistical verification of power grids considering process-induced leakage current variations., *In IEEE/ACM ICCAD, Nov. 2003, pp. 770-777.*

[11] D. Boning, and S. R. Nassif. Models of process variations in device and interconnect in design of high performance microprocessor circuits. *A. Chandrakasan, W.J. Bowhill and F. Fox ed. John Wiley and Sons, NY, 2001.*

[12] S. R. Nassif, and J. N. Kozhaya. Fast power grid simulation. *In ACM/IEEE DAC, June 2000, pages 156-161.*

[13] P. D.Spanos, and R. G. Ghanem. Stochastic finite elements: a spectral approach. *Dover Publications, 2003.*

[14] R. R. Rao et al. Parametric yield estimation considering leakage variability. In *In ACM/IEEE DAC, June 2004, pp. 442-447.*

[15] V. Mehrotra et al. A methodology for modeling the effects of systematic within-die interconnect and device variation on circuit performance, *In ACM/IEEE DAC, June 2000, pp. 172-175.*

[16] T. Davis. `http://www.cise.ufl.edu/research/sparse/UFsparse`. *Sparse Matrix Package, Univ. of Florida.*

[17] P. Ghanta, S. Bhardwaj, and S. Vrudhula. `http://www.tauworkshop.com/05_Slides/TAU_ghanta.pdf`. *Presentation Slides, ACM/IEEE TAU Workshop, 2005.*

[18] Friedberg et al. Modeling within-die spatial correlation effects for process-design co-optimization. *In IEEE ISQED, March 2005, pp. 516-521.*