Performance-Oriented Statistical Parameter Reduction of Parameterized Systems via Reduced Rank Regression[*]

Zhuo Feng
Department of ECE
Texas A&M University
College Station, TX 77843
fengzhuo@ece.tamu.edu

Peng Li
Department of ECE
Texas A&M University
College Station, TX 77843
pli@neo.tamu.edu

ABSTRACT

Process variations in modern VLSI technologies are growing in both magnitude and dimensionality. To assess performance variability, complex simulation and performance models parameterized in a high-dimensional process variation space are desired. However, the high parameter dimensionality, imposed by a large number of variation sources encountered in modern technologies, can introduce significant complexion in circuit analysis and may even render performance variability analysis completely intractable. We address the challenge brought by high-dimensional process variations via a new *performance-oriented* parameter dimension reduction technique. The basic premise behind our approach is that the dimensionality of performance variability is determined not only by the statistical characteristics of the underlying process variables, but also by the *structural information* imposed by a given design. Using the powerful reduced rank regression (RRR) and its extension as a vehicle for variability modeling, we are able to systematically identify statistically significant reduced parameter sets and compute not only *reduced-parameter* but also *reduced-parameter-order* models that are far more efficient than what was possible before. For a variety of interconnect modeling problems, it is shown that the proposed parameter reduction technique can provide more than one order of magnitude reduction in parameter dimensionality. Such parameter reduction immediately leads to reduced simulation cost in sampling-based performance analysis, and more importantly, highly efficient parameterized interconnect reduced order models. As a general parameter dimension reduction methodology, it is anticipated that the proposed technique is broadly applicable to a variety of statistical circuit modeling problems, thereby offering a useful framework for controlling the complexity of statistical circuit analysis.

1. INTRODUCTION

As IC technologies enter the nanometer regime, capturing various process variations and assessing their impacts on circuit performance become increasingly critical and difficult [1]. While the growing magnitude of process variations pushes for more complex parametric models that may go beyond those based on the first-order sensitivities, the increasing sources of process variability impose a formidable high-dimensional parameter space in which a given design must be verified and optimized.

While the notorious issue of *curse of dimensionality*, coined by Bellman [2], emerges in many fields of science and engineering, its manifestation in variation-aware circuit design is particularly problematic. For instance, a full consideration of inter-/intra-die wire width, thickness and dielectric thickness variations in multi-layer interconnect structures can easily introduce several tens of geometrical variation parameters. Modeling interconnect variations and performing timing verification in such high-dimensional parameter space involve obvious challenges. *Curse of dimensionality* impacts a wide range of CAD problems since the feasibility as well as the efficiency of many CAD algorithms critically depend on the dimension of the parameter space. For example, the cost and complexity of many empirical macromodeling techniques (e.g. RSM based performance modeling) grow *exponentially* in the number of parameters[3, 4]. The same issue appears in a large body of more formal parameterized interconnect reduced order modeling algorithms and variational analysis techniques developed for capturing interconnect variability [5, 6, 7, 8, 9, 10, 11]. For many of these techniques, the inclusion of a large set of variational parameters can make the circuit modeling and analysis extremely costly, and under many cases, may even render those tasks impractical. Furthermore, we notice that the efficiency of many statistical timing analysis techniques also depend on the dimension of underlying parametric variations as well as the way in which these variations are processed [12, 13, 14].

In the CAD community, the standard practice employs PCA (principle component analysis) and its variants for parameter reduction [15, 16]. Although widely adopted, these techniques are limited since parameter reduction is achieved by only considering the statistics of the controlling parameters while neglecting the important correspondence between these parameters and circuit performances under modeling. Parameter screening is often applied under the context of response surface modeling [3], however, the technique is empirical in nature as it prunes parameters one at a time based on sensitivity-like measures.

Given the fact that systematic CAD specific parameter reduction methodologies are lacking, in this paper we propose a new *performance-oriented* parameter reduction approach. Unlike the standard principle component analysis (PCA), our approach is *performance-oriented* in the sense that not only the statistical properties of underlying process parameters but also the correspondence between these parame-

---

[*] This work was support in part by SRC under contract 2005-TJ-1298 and C2S2, the MARCO Focus Center for Circuit & System Solution, under MARCO contract 2003-CT-888.

ICCAD'06 November 5-9, 2006, San Jose, CA

ters and circuit performances of interest are *simultaneously* exploited. To build our parameter reduction methodology on a rigorous statistical foundation, we adopt the powerful reduced-rank regression (RRR) [17] and extend it to for practical circuit modeling purpose. This new methodology allows us to perform systematic parameter dimension reduction while exploiting the valuable *structural information* imposed by a given design, making it possible to achieve design specific parameter reduction in a way that is much powerful than what was possible before. We show that the proposed technique can lead to more than one order of magnitude parameter reduction for a variety of interconnect circuit examples. Our performance-oriented parameter reduction technique reduces the number of statistical samples required to derive accurate performance statistics. Furthermore, the achieved parameter dimension reduction dramatically simplifies parameterized interconnect model order reduction and leads to highly accurate and compact interconnect simulation models. We demonstrate the proposed techniques by extracting highly efficient *reduced-parameter* as well as *reduced-parameter-order* interconnect models. In the latter, the parameter dimension and the circuit size are reduced simultaneously by applying parameter reduction and model order reduction. As a general parameter reduction methodology, it is well expected that the proposed framework can be broadly applied to a variety of other circuit modeling problems hence providing a new way to reduce the cost of statistical circuit analysis.

2. PRINCIPLE COMPONENT ANALYSIS

In this section, we review the conventional principle component analysis (PCA) and its limitations. Then, we point out the need for more powerful parameter reduction techniques specifically for variational circuit analysis.

The objective of PCA is to identify data patterns and describe the data in a more descriptive way. Using PCA, one tries to achieve data reduction by performing variable transformations and computing a few linear combinations of the original variables to capture the most of statistical variance of the data. PCA is usually conducted in four steps: **a)** collect data and get the covariance matrix of the data; **b)** compute the eigenvectors and eigenvalues of the covariance matrix; **c)** select the eigenvectors which correspond to the first few largest eigenvalues to be the principle components; **d)** represent the original data set using the selected principal components.

Useful parameter dimension reduction can be achieved if the eigenvalues of the covariance matrix drop off quickly. However, one critical observation is that such parameter reduction is achieved by merely considering the statistical characteristics (correlations) of a given data set without any account for other statistical data that may depend on the data set under analysis. Under the context of statistical circuit analysis, PCA is commonly used to perform data reduction in the process parameter space. Once the process variables are compressed based upon their variances and correlations, statistical circuit analysis is carried out based on the reduced process variables in an independent subsequent step.

It is worth noting that the main objective and challenge of most statistical circuit analysis tasks is to analyze the system *performance variability*, which is a function of underlying process variations. Therefore, performing parameter reduction while considering only the statistical property of process variations using a standard technique such as PCA can be rather limited, and under certain cases, it may even lead to misleading parameter reduction. To see this issue more clearly, let us consider an RC circuit with a single voltage source input and no grounded resistors, representing the widely used on-chip RC interconnect model used for timing analysis. We assume that the RC circuit is perturbed by the manufacturing fluctuations in the forms of wire width, thickness and dielectric layer thickness variations. It is well known that the DC voltage response of such circuit can be trivially determined by the input voltage excitation regardless of any RC element variations. However, if one blindly applies PCA to reduce the dimension of RC variations for the purpose of modeling the DC circuit performance, one will fail to identify the trivial fact that the dimensionality of the variability of the DC performance is essentially zero. Next, we consider the more useful issue of modeling the timing performance variations of the RC circuit. Suppose that variance of wire width W is greater than that of dielectric layer thickness H. Then a relevant question to ask is: which variation is more critical (statistically) in terms of the delay variability? Without taking any circuit information into account, PCA may just pick W since it has a larger variance. However, in terms of delay the W variation may not necessarily be a more dominant factor, since the increase in W leads to an increase in wire capacitance but also a decrease in wire resistance so that the delay may not be influenced much.

3. PARAMETER DIMENSION REDUCTION VIA REDUCED RANK REGRESSION

To achieve more powerful parameter dimension reduction, it is clear that a framework that can take into account the meaningful structural information of a given design is desired. To facilitate a new statistical parameter reduction approach rigorously, we adopt reduced rank regression (RRR) as a suitable modeling tool and extend it for practical circuit modeling needs.

3.1 Linear reduced rank regression

Regression analysis has been widely used in statistical data analysis. We consider the general multivariate linear model

$$Y = CX + \varepsilon, \tag{1}$$

where Y is an $m \times N$ matrix containing N-samples of m dependent variable vectors, X is an $n \times N$ matrix containing N-samples of n predictor variables, C is an $m \times n$ regression coefficient matrix and ε is the zero-mean random errors of the regression. As a standard approach, C can be found by using the least square regression. The least squares criterion is to minimize the trace (sum of the diagonal elements) of the covariance matrix, $\Sigma_{\varepsilon\varepsilon}$ of ε, such that an optimal solution for C can be obtained as

$$C = YX^T(XX^T)^{-1}. \tag{2}$$

It is easy to show that the minimization of the trace of $\Sigma_{\varepsilon\varepsilon}$ also implies the minimization of the standard deviation error for each dependent variable Y.

Notice that the above linear regression model does not lend itself to parameter reduction. The standard regression model does not exploit any statistical redundancy and correlation between Y in the model. In practical problems, however, it is very likely that significant model redundancy may exist, which manifests in the possibility of constructing a rank-reduced regression matrix $\tilde{C}$.

Suppose that we have a predictor variable vector $X \in \mathbb{R}^n$ and a dependent variable vector $Y \in \mathbb{R}^m$, with each having a zero mean. We denote the covariance matrix of X as $Cov(X) = \Sigma_{xx}$, and the covariance matrix between X and Y as $Cov(Y, X) = \Sigma_{yx} = \Sigma_{xy}^T$. The following theoretical result can be shown [17]:

Theorem 1. *For any positive-definite matrix* Ω*, an* $m \times r$ *matrix* A_r *and* $r \times n$ *matrix* B_r *can be found to minimize the trace*

$$tr\{E[\Omega^{1/2}(Y - A_rB_rX)(Y - A_rB_rX)^T\Omega^{1/2}]\}, \tag{3}$$

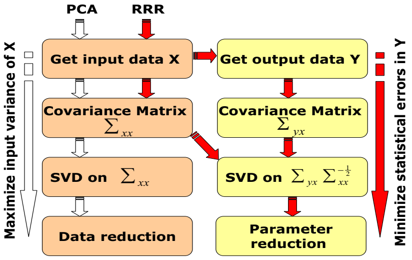

Figure 1: Comparison between PCA and RRR

where

$$A_r = \Omega^{-1/2}U, B_r = U^T\Omega^{-1/2}\Sigma_{yx}\Sigma_{xx}^{-1}, \tag{4}$$

and $U = [U_1, ..., U_r]$ *contains* r *normalized eigenvectors corresponding to the* r *largest eigenvalues of the matrix*

$$D = \Omega^{1/2}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Omega^{1/2}. \tag{5}$$

It is straightforward to show that (4) can be found equivalently by computing the SVD of the matrix

$$E = \Omega^{1/2}\Sigma_{yx}\Sigma_{xx}^{-\frac{1}{2}}. \tag{6}$$

The complete proof of the theorem can be found in [17].

It is critical to note that, a successful application of RRR also implies the possibility of parameter reduction. In other words, by the previously described rigorous procedure, the inherent redundancy in the predictor variables can be filtered out statistically. To see this point, we first notice that we have computed a rank-r regression model that minimizes the statistical errors in Y in the sense of (3)

$$Y = A_rB_rX + \tilde{\varepsilon}, \tag{7}$$

where $\tilde{\varepsilon}$ represents the model error. We can construct a new set of variable $Z \in \mathbb{R}^r$ $(r < n)$ as

$$Z = B_rX, \tag{8}$$

leading to an optimal regression model

$$Y \approx A_rZ. \tag{9}$$

Under our context of circuit modeling, it is important to notice that a reduced rank model such as (9) is computed *not* to simplify a given more complex model (e.g. (1)), instead, it is used as a means to reveal the redundancy in the predictor variables (e.g. process variations) to fulfill the purpose of parameter reduction. In our circuit modeling task, Y does not have to be the circuit performances of interest, more generally it can be chosen to be some other easily computed circuit responses that are closely related to the performances, as described in the following sections of the paper. Furthermore, it can be noted easily that the standard PCA can only be applied to reduce data redundancy in either X or Y, but not the both simultaneously. We show the differences between PCA and RRR in Fig. 1.

3.2 Nonlinear reduced rank regression

For many realistic circuit problems, we have noticed that the linear regression models in (1)(7) are not completely adequate to capture the noticeable nonlinear relationship between process variables and circuit performances, especially when the range of the process variations is relatively large. To seek a more robust parameter reduction under these cases, we adopt the same notion of reduced rank regression as described in the previous subsection but cast it under a more general quadratic model. Consider the following quadratic regression model

$$Y = f(X) \approx \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} X \\ X \otimes X \end{bmatrix}, \tag{10}$$

where the quadratic terms of X are expressed using the tensor product: $X \otimes X = [x_1^2, x_1x_2, \cdots, x_1x_n, \cdots, x_n^2]^T$, C_1 and C_2 are the first order and second order coefficient matrices, respectively. To apply the reduced rank approximation, ideally one would like to find some regression matrices $\tilde{A}_{r_1} \in \mathbb{R}^{m\times r}$, $\tilde{A}_{r_2} \in \mathbb{R}^{m\times r^2}$, and $\tilde{B}_r \in \mathbb{R}^{r\times n}$ such that the error of the following reduced-rank regression model can be minimized in a statistical sense

$$Y \approx \begin{bmatrix} \tilde{A}_{r_1} & \tilde{A}_{r_2} \end{bmatrix} \begin{bmatrix} \tilde{B}_rX \\ \left(\tilde{B}_rX\right) \otimes \left(\tilde{B}_rX\right) \end{bmatrix}. \tag{11}$$

However, it turns out that an optimal model in the form of (11) is difficult to solve. Instead, we include the quadratic terms $X \otimes X$ in the linear RRR model as additional predictor variables by defining a new predictor vector

$$\tilde{X} = \begin{bmatrix} X \\ X \otimes X \end{bmatrix}. \tag{12}$$

We compute the new covariance matrices $Cov(Y, \tilde{X}) = \Sigma_{Y,\tilde{X}}$ and $Cov(\tilde{X}) = \Sigma_{\tilde{X},\tilde{X}}$ and follow the linear RRR procedure to get a reduced-rank model

$$Y \approx A_r \begin{bmatrix} B_{r_1} & B_{r_2} \end{bmatrix} \begin{bmatrix} X \\ X \otimes X \end{bmatrix}, \tag{13}$$

where $A_{r_1} \in \mathbb{R}^{m\times r}$, $B_{r_1} \in \mathbb{R}^{r\times n}$ and $B_{r_2} \in \mathbb{R}^{r\times n^2}$. The above model is optimal in a sense similar to (3) (the regression model is cast in a quadratic form here). Compared with the model in (11), here we have

$$\begin{aligned} A_rB_{r1} &\approx \tilde{A}_{r1}\tilde{B}_r, \\ A_rB_{r2} &\approx \tilde{A}_{r2}(\tilde{B}_r \otimes \tilde{B}_r). \end{aligned} \tag{14}$$

The reduced parameter set $Z \in \mathbb{R}^r$ is expressed in a quadratic form of X

$$Z = B_{r1}X + B_{r2}(X \otimes X). \tag{15}$$

4. STATISTICAL CIRCUIT MODEL GENERATION WITH PARAMETER REDUCTION

In this section, we apply the nonlinear RRR based parameter reduction to practical circuit applications. We focus on statistical interconnect modeling problems and develop specific techniques to develop compact parameterized simulation models.

4.1 Capturing interconnect parametric variations

We use the standard modified nodal analysis (MNA) equations to describe an interconnect network

$$\begin{cases} (G + sC)x = Bu \\ y = L^Tx \end{cases}, \tag{16}$$

where $u \in R^{n\times 1}$ and $y \in R^{m\times 1}$ represent the inputs and outputs, $x \in R^{N\times 1}$ represents the system unknowns, $G, C \in R^{N\times N}$ are the conductance and capacitance matrices, $B \in R^{N\times n}$ and $L \in R^{N\times m}$ are the input and output matrices, respectively.

In order to possibly capture process variations, without loss of generality, we consider the RC circuit as shown in Fig. 2 as an example. The circuit has one nonlinear driver providing the input and three output circuit nodes driving three downstream stages. The circuit is divided into several regions spatially and the local geometrical variations are introduced on a per region basis to capture possible spatial

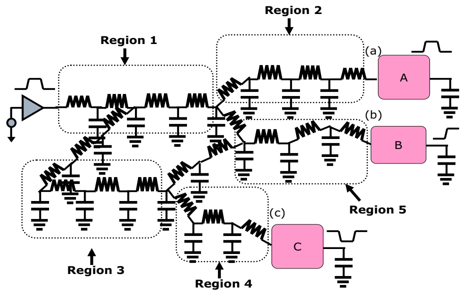

Figure 2: An RC circuit with parametric variations.

process variations. Variations considered in this paper only include various geometrical parameters such as wire width and thickness, dielectric layer thickness, though other types of local or global parameters can be treated in a similar way. Generally, we consider a set of n_p local and global geometrical variation variables: $\vec{p} = [p_1, p_2, \cdots, p_{n_p}]^T$. Without loss of generality, we capture their influences in (16) by expanding conductance and capacitance matrices into quadratic forms in $\vec{p}$ as

$$G = G_0 + \sum_i G_i p_i + \sum_{ij} G_{ij} p_i p_j, \qquad (17)$$

$$C = C_0 + \sum_i C_i p_i + \sum_{ij} C_{ij} p_i p_j. \qquad (18)$$

In practice, we may only consider variations in resistances and capacitances and neglect inductance variations, which have been observed to be small.

4.2 RRR-based interconnect parameter reduction

A full account of global and local variations in a large multi-layer interconnect network can lead to the consideration of a large set of geometrical variables, i.e., n_p is large. However, if we are only interested in analyzing the circuit performances at the output nodes, the effective parameter dimension of a given network may not be very large since the specific circuit structure can hide certain parametric variations and may even introduce canceling effects between multiple variations. To seek the true parameter dimension in a statistically rigorous fashion, we exploit the proposed nonlinear RRR based parameter reduction.

To apply nonlinear RRR, one would very naturally choose the underlying process variations, i.e., $\vec{p}$, as the predictor variables $(X/\tilde{X})$. In the proposed approach, RRR is only employed as a tool to perform parameter reduction but is not used for performance modeling. Therefore, the dependent variables (Y) may not have to be chosen as certain performance measures such as circuit delays. In practice, this flexibility is particularly useful because in many cases a compact simulation model is often needed but not a performance model. For interconnect models, we use transfer function moments as the dependent variables based on their strong correlation with timing performance. One important benefit of such choice is that transfer function moments are also easy to compute. We have developed computationally efficient procedures to generate closed-form expressions for transfer function moments and their dependency on the underlying geometrical variations. As such, statistical measures required by RRR, e.g. $\Sigma_{\tilde{x}\tilde{x}}$ and $\Sigma_{\tilde{x}y}$, can be efficiently obtained in closed-form without resorting Monte-Carlo sampling, leading to high efficiency of the proposed parameter reduction.

Without loss of generality,a transfer function moment at a particular output of interest can be written as

$$m_k = m_{k0} + \sum_{i=1}^{n_p} \alpha_{k,i} p_i + \sum_{i=1}^{n_p} \beta_{k,i,i} p_i^2 + \sum_{i=1}^{n_p} \sum_{j=1}^{i-1} \beta_{k,i,j} p_i p_j, \quad (19)$$

where $k = 1, \cdots, n_s$ and n_s is the number of moments to be observed. For example, if we want to capture the first three moments for five output nodes, then n_s will be equal to 15. In the above equation, m_{k0} is the nominal case moment, $\alpha_{k,i}$ and $\beta_{k,i,j}$ are the first and second order coefficients capturing the dependency of m_k on $\vec{p}$. For many interconnect networks, we have observed that considering the first few (three) moments using the second order formulas is usually sufficient for parameter reduction purpose under the typical ranges of interconnect variations (30% 3σ variations [1]).

In the light of (12), we use $\tilde{X}$ to denote the zero-mean linear and quadratic terms associated with the geometrical variations $\vec{p}$, which are assumed to have zero mean. We partition $\tilde{X}$ into $\tilde{X} = \left[\begin{array}{ccc} X_f^T & X_c^T & X_s^T \end{array} \right]^T$, where $X_f = \left[p_1, p_2, \cdots, p_{n_p}\right]^T$ consists of the 1st order terms, $X_s = \left[p_1^2 - \sigma_{p_1}^2, \cdots, p_{n_p}^2 - \sigma_{p_{n_p}}^2\right]^T$ consists of the pure square terms, $X_c = \left[p_1 p_2, \cdots, p_{n_p - 1} p_{n_p}\right]^T$ consists of the 2nd order cross terms, and σ_{p_i} is the standard deviation of each p_i. Notice that, X_f, X_c and X_f all have zero mean.

Expressing all the moments as quadratic functions of $\vec{p}$ gives

$$Y = \left[\begin{array}{c} \Delta m_1 \\ \vdots \\ \Delta m_{n_s} \end{array} \right] = \left[\begin{array}{ccc} S_f & S_c & S_s \end{array} \right] \left[\begin{array}{c} X_f \\ X_c \\ X_s \end{array} \right] = S\tilde{X}, \quad (20)$$

where Y contains all the n_s moments but subtracted by their mean values, S_f, S_c and S_s are coefficients for the first order terms, the pure square terms and the cross terms, which can be computed efficiently.

Given the joint probability function (jpdf) of the process variables $\vec{p}$, the covariance matrices required by the nonlinear RRR algorithm described in Section 3.2 can be computed using (20). In the following, we consider a widely assumed special case in which the process variables $\vec{p}$ are jointly Gaussian with zero mean. In this case, covariance matrices can be obtained in simple closed-form expressions. We further assume that $\vec{p}$ are independent with standard deviations σ_{p_i}'s since otherwise the standard PCA analysis can be always applied to obtain a set of independent Gaussian variables. We compute the covariance matrix $Cov(\tilde{X})$ in a partitioned form as

$$\Sigma_{\tilde{X}\tilde{X}} = E\left\{\tilde{X}\tilde{X}^T\right\} = \left[\begin{array}{ccc} \sum_f & & \\ & \sum_c & \\ & & \sum_s \end{array} \right], \quad (21)$$

where the diagonal elements of the above matrix are categorized into three groups. The covariances of the first order terms are given as: $\left(\Sigma_f\right)_{ii} = \sigma_{p_i}^2$ while those of the pure square terms and the second order cross terms are given as $\left(\Sigma_s\right)_{ii} = 2\sigma_{p_i}^4$ and $\left(\Sigma_c\right)_{ij} = \sigma_{p_i}^2 \sigma_{p_j}^2$, respectively. The covariance matrix between Y and $\tilde{X}$ is given as

$$\Sigma_{Y\tilde{X}} = E\left\{S\tilde{X}\tilde{X}^T\right\} = S\Sigma_{\tilde{X}\tilde{X}}. \quad (22)$$

We set the positive-definite matrix Ω (4) to be the identity matrix, and compute the SVD (Singular Value Decomposition) of E matrix (6) to obtain the matrix U as

$$U\Sigma V^T = \Sigma_{Y\tilde{X}} \Sigma_{\tilde{X}\tilde{X}}^{-\frac{1}{2}} = S\Sigma_{\tilde{X}\tilde{X}}^{\frac{1}{2}}, \quad (23)$$

where matrix U contains the first few singular vectors which have the largest singular values. Finally, the reduced set of parameters Z can be expressed using (15), from which the statistical distributions of Z can be also computed.

4.3 Reduced-parameter interconnect models and parameterized model order reduction

To be benefited by parameter reduction in simulation, we need to cast our circuit model such as (16) in the reduced parameter set Z. Hence, the dependency of the system matrices on the new parameters should be computed

$$G = G_0 + \sum_i G_{z_i} z_i + \sum_{ij} G_{z_{ij}} z_i z_j, \tag{24}$$

$$C = C_0 + \sum_i C_{z_i} z_i + \sum_{ij} C_{z_{ij}} z_i z_j. \tag{25}$$

Applying the chain rule gives the first order sensitivities with respect to the new parameters as

$$G_{z_k} = \frac{\partial G}{\partial z_k} = \sum_i \frac{\partial G}{\partial p_i}\frac{\partial p_i}{\partial z_k};\; C_{z_k} = \frac{\partial C}{\partial z_k} = \sum_i \frac{\partial C}{\partial p_i}\frac{\partial p_i}{\partial z_k}. \tag{26}$$

To fully compute the above expressions, we still have to find $\frac{\partial p_i}{\partial z_k}$ first. To simplify the computation, a good approximation is to retain only the dominant linear terms (matrix B_{r_1}) in (15) to solve p_i's (or x_i's) using z_i's in the form

$$p_i = \sum_{j=1}^{n_z} t_{ij} z_j, i = 1, ..., n_p. \tag{27}$$

Notice that the linear portion of (15) represents a set of under-determined linear equations since the number of z_i's (n_z) is less than the number of p_i's (n_p) due to the parameter reduction. However, we shall recall the a successful application of parameter reduction also implies that statistically not all the original parameters (p_i's) are important but only a few combinations of them (z_i's) are. This observation allows us to use the standard solution methods for under-determined systems such as pseud-inverse to express p_i's in terms z_i's. Consequently, we are able to compute $\frac{\partial p_i}{\partial z_k}$ so that the sensitivity matrices in (26) can be handled. The second order dependencies of the system matrices on Z can be obtained by substituting (27) into (17) and (18) and collecting the coefficients matrices which correspond to the second order terms $z_i z_j$.

Upon obtaining the new simulation models in the reduced parameter set z, the immediate benefit of parameter reduction is to conduct Monte Carlo simulation by sampling in the new parameter space, which is much more efficient. We have applied variance reduction techniques such as Latin Hypercube Sampling (LHS) to reduce the number of random samples needed to estimate performance statistics by working in the reduced parameter space. Due to the application of our RRR based parameter reduction, LHS becomes an effective variance-reduction tool in the low-dimensional parameter space.

Equally important, the reduction of parameter dimension is also a key to enable parameterized model order reduction techniques to compute compact simulation models while considering the impact of process variations [9, 10, 11]. It is important to notice that the efficiency and the cost of these algorithms critically depend on the parameter dimension. By performing parameter dimension reduction, we are able to compute highly efficient reduced order models while capturing a large set of (original) process variables. This leads to compact reduced-parameter-order models.

5. NUMERICAL RESULTS

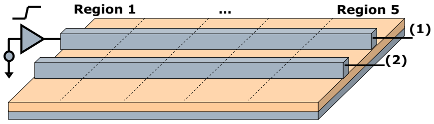

Figure 3: Two coupled lines.

We demonstrate the application of the proposed techniques on several interconnect circuit examples. We apply the RRR based algorithms to significantly reduce the parameter dimension and compute the compact parameterized reduced order models in the reduced parameter space. We assume that the random interconnect geometrical variations are independent and Gaussian, although our methodology can be applied to other types of statistical variations. The accuracy of our reduced-parameter models as well as the reduced-parameter-order models are verified by examining $50\% V_{dd}$ delays and frequency domain responses.

5.1 Two coupled lines

First, we consider two coupled long RC lines as shown in Fig. 3. The wire width W and thickness T of each line are both 1 μm, and the dielectric layer thickness H is 0.5 μm. The spacing S between two lines is $0.8\mu m$. We divide the two lines into five regions and include 20 resistors and 60 capacitors (20 of them are coupling capacitors) in each region. To realistically relate the RC parameters with the geometrical parameters, which are subject to process variation, capacitance values are calculated using the closed-form formulas based on the geometrical values [18] while the unit length resistance is calculated using the cross section area and the conductor resistivity. To model process variations, in this example for each region we consider four geometrical variations in wire width (W), wire thickness (T), wire spacing (S), and dielectric layer thickness (H). Therefore, there are a total of 20 variation variables. Since the second order sensitivities of capacitance and resistance with respect to the geometry are quite small compared with the first order terms, the R and C values can be safely expressed in the first order sensitivity in these geometrical parameters. The 3 σ geometrical variation ranges are from 15% to 30%. We apply the nonlinear RRR-based parameter reduction algorithm in Section 3.2 to generated three parameter-reduced models with one, two and three parameters, respectively. Therefore, the maximum parameter reduction achieved is 20x for this example.

5.1.1 Delay distributions

We compare the original model and two reduced-parameter models by examining the delay at terminal (1) when a ramp input is applied, as shown in Fig. 3. For the 3-parameter model, we demonstrate the reduction on the number of random samples required to collect the delay distribution when the 3 σ variations of all parameters are set to be 30%. First, we perform Latin Hypercube Sampling (LHS) [19] sampling in the full 20-parameter model to get the delay distribution. LHS is used as a variance reduction technique to improve the sampling efficiency. It is observed that a minimum of $4,000$ samples are required in order to get a stable delay distribution. If sampling in the 3-parameter model using LHS, it is observed that 800 samples are enough to provide an accurate estimation of the same distribution. We compare the PDFs and CDFs of the two models in Fig. 4.

In Fig. 5, we verify the accuracy of two reduced-parameter models on a per sample basis. We generate $1,000$ statistical samples in the full parameter space and compute the reference delay for each circuit sample using circuit simulation. Then, we map these $1,000$ samples in the original 20-dimensional parameter space to the new reduced (1/3 di-

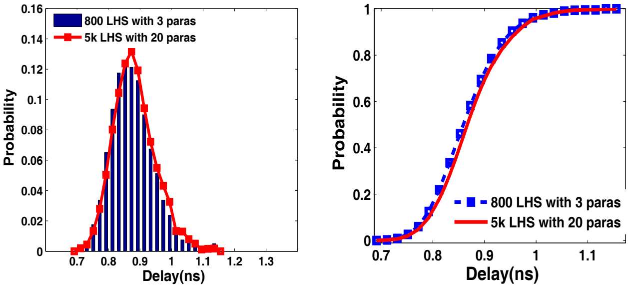

Figure 4: Comparison of the full and the reduced-parameter models on the delay PDF and CDF.

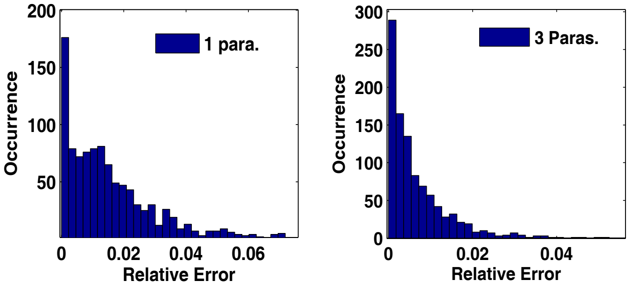

Figure 5: Relative delay errors of the reduced-parameter models for 1,000 random samples.

mensional) parameter space using (15). For each sample we obtain the corresponding delay based on reduced-parameter models and compare it against the reference value. As can be seen, both reduced parameter models are rather accurate while keeping three parameters in the model can improve the accuracy further.

More experiments are conducted on the three reduced-parameter models in one, two and three parameters, respectively in Table 1. Four different combinations of geometrical variations are considered. For each case, we use $10,000$ Monte Carlo samples in the full 20-parameter model to get stable estimation of the delay distribution and compute the mean and the standard deviation (std) as reference values. Then, we verify the accuracy (relative error in mean/std.) of the three reduced-parameter models by generating $4,000$ and 800 LHS samples, respectively. As can be seen, 800 LHS samples of the reduced models can provide quite accurate estimations on the mean and standard deviation values while the three-parameter model is offering an excellent accuracy.

5.1.2 Formation of the reduced parameter space

As shown in Section 3.2, the original set of parameters in X can be reduced into new variables in Z using a transformation as in (15). From another angle, (15) reveals the importance of each old parameter with respect to the performance of the circuit in a statistical sense, which can be clearly understood by examining the weighing coefficients. For example, the (i,j) entry of matrix B_{r1} describes the linear contribution of the j^{th} original parameter x_j to the i^{th} new parameter z_i. To show a more clear picture of each x_j's statistical importance, we plot the linear weighing coefficients for the first three new parameters (z_1, z_2 and z_3) in Fig. 6. We designate the variation sources from each of the five regions using the corresponding the region number. It is evident that the wire width and thickness variations, especially those in the first few regions, contribute most to the new parameters. This result can be well explained by circuit intuition. However, our approach provides a statistical approach to reveal the importance of the variation sources quantitatively.

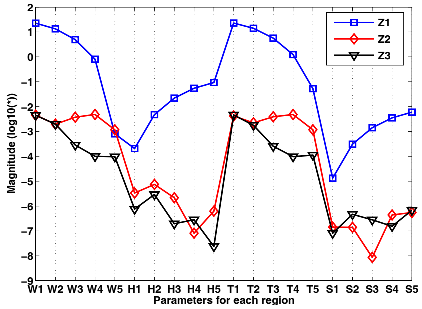

Figure 6: Weighing coefficients for the two coupled lines.

Table 2: Average delay errors of the linear and nonlinear reduced-parameter models for the RC circuit.

Var. of Paras.			1 para.	2 paras.		3 paras.	
σ_W	σ_H	σ_T	Lin.	Lin.	Non.	Lin.	Non.
15%	10%	10%	1.7%	0.9%	0.7%	0.3%	0.2%
10%	15%	10%	2.0%	1.2%	1.2%	0.4%	0.2%
10%	10%	15%	1.5%	1.0%	0.9%	0.8%	0.7%

5.2 An RC circuit

We consider an RC circuits with 776 circuit unknowns and $1,276$ RC elements. The nominal wire width and thickness are $W = 0.35$ μm and $T = 0.65$ μm, and the nominal dielectric layer thickness is $H = 0.65$ μm. The circuit is divided into five regions and three local variations associated with the above geometrical parameters are introduced for each region, leading to 15 variation sources in total. 19 widely separated nodes are selected as the output nodes and are considered in the reduced-rank regression. We compute three reduced-parameter models in one, two and three parameters. For a given parameter size, we consider two reduced-parameter models, one computed by the linear RRR-based reduction (Section 3.1) and the other by the nonlinear RRR-based reduction (Section 3.2).

In this example, we demonstrate the accuracy of these reduced-parameter modelings by performing statistical sampling. We generate 500 Monte Carlo samples based on the full 15-parameter model and compute the circuit delay at one particular node when a ramp input is applied. We transform the above 500 samples from the original parameter space into the reduced parameter space and compare the delay differences between the original and the reduced-parameter models. In Table 2, three parameter variation combinations with different standard deviations are considered. For various reduced-parameter models, we compute the average delay error of the 500 samples by resulting the delays computed by the full model as reference. We only give the results of the linear RRR-based model for the one parameter case because the nonlinear model leads to a almost identical accuracy. As observed, the accuracy of the reduced model can be improved by including a larger number of parameters. Unlike the previous example, for this particular circuit, both the linear and nonlinear RRR models are fairly accurate. The accuracy can be somewhat improved by adopting nonlinear models.

5.3 An RC mesh

We consider an RC mesh which is divided into nine blocks

Table 1: Comparison of delays between the original model and the new model

Variations of Parameters				10K M.C (20 Paras.)		4K LHS Rel. Err.	800 LHS Rel. Err.		
σ_W	σ_H	σ_T	σ_S	Mean	Std.	20 paras.	1 para.	2 paras.	3 paras.
5%	10%	10%	10%	885.9 ps	52.4 ps	0.00%/1.11%	0.94%/3.96%	0.54%/2.78%	0.01%/0.16%
10%	5%	10%	10%	893.2 ps	68.7 ps	0.00%/0.78%	1.50%/5.44%	0.74%/1.68%	0.00%/1.86%
10%	10%	5%	10%	885.3 ps	51.3 ps	0.00%/2.07%	0.88%/3.32%	0.49%/0.95%	0.21%/0.29%
10%	10%	10%	5%	892.4 ps	69.0 ps	0.09%/0.56%	1.40%/5.67%	0.64%/2.05%	0.73%/2.55%

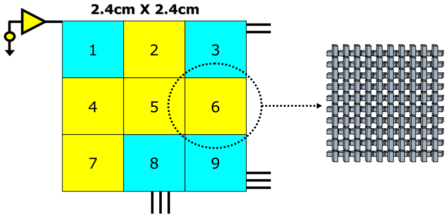

Figure 7: An RC mesh with nine blocks.

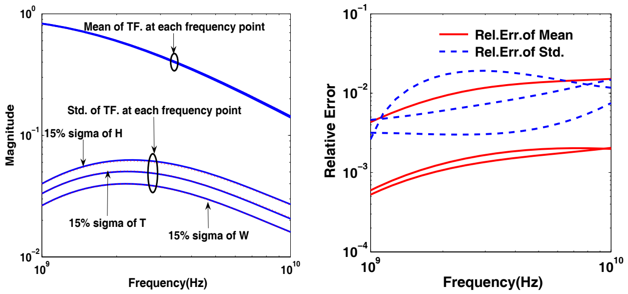

Figure 8: The frequency response of the two-parameter model for the RC mesh.

as shown in Fig. 7. In each block there is an 8-by-8 sub-mesh containing about 120 resistors and 56 capacitors. The whole design has 576 circuit unknowns and 1,609 circuit elements. The nominal wire dimensions are set in a way identical to the previous circuit example. We neglect the effect of coupling capacitance since the line spacing is relatively large. We consider the variations in wire width and thickness (W/T) and dielectric layer thickness (H) for each of the nine regions, giving a total of 27 local geometrical variations. For a single circuit input, we select the first three transfer function moments of eight circuits nodes located in different blocks (as shown in Fig. 7) to be the dependent variables (observations) in the nonlinear reduced-rank regression. We compute a two-parameter reduced model and verify the model accuracy by observing the frequency response of one selected observation nodes. We conduct Monte-Carlo simulations using the two-parameter model under three different settings where in each setting we choose the standard deviation of one variation to be 15% while keeping the standard deviations of the remaining parameters as 10%. The mean values and the standard deviations of the frequency response obtained from the reduced-parameter model are compared against with the true values (obtained from the full model) in Fig. 8. In the left plot, the results of the full and reduced-parameter models are indistinguishable. The right plot shows the relative errors in the three different cases. As observed, the two-parameter model is very accurate while reducing the parameter dimension from 27 to only two.

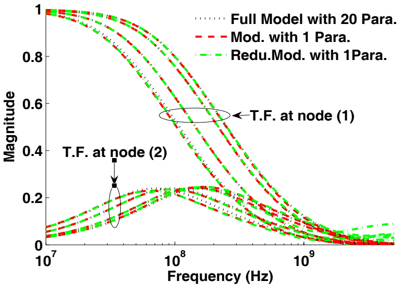

Figure 9: Comparison between the full, reduced-parameter, and reduced-parameter-order models.

5.4 Combining parameter dimension reduction with model order reduction

In the previous subsections, we have demonstrated the accuracy of the reduced-parameter interconnect models as well as the improved efficiency brought by these models in sampling-based circuit analysis. To tackle the statistical analysis complexity brought by the high parameter dimension and the large design size simultaneously, we combine parameter reduction and model order reduction techniques to compute compact *reduced-parameter-order* models. It should be noticed that the cost of most parameterized interconnect model order reduction algorithms grow *exponentially* in the number of the parameters, thus a significant reduction in the parameter space will lead to highly efficient parameterized models as shown by the following circuit examples.

5.4.1 Two coupled RC lines

For the two coupled RC line circuit modeled using 204 circuit unknowns in Fig. 3, we first apply the nonlinear RRR based algorithm to reduce the parameter dimension from 20 to one and then use the parameterized model order reduction algorithm in [10] to compute a passive one-parameter 12th-order reduced model. Six transfer function moments of nodes (1) and (2) are selected as the dependent variables in the RRR procedure. Since the model order reduction algorithm performs moment-matching with respect to the process variable, a direct inclusion of 20 parameters will lead to an explosion in model size. This difficulty is completely avoided by performing a reduction in the parameter space first.

We compare the frequency responses of the full model and the one-parameter 12th-order model on circuit samples generated by perturbing all the 20 geometrical parameters by ±10% and ±20%, respectively. In Fig. 9, four samples of the frequency responses at nodes (1) and (2) are obtained based on three models: 20-parameter full-order model, one-parameter full-order model and one-parameter 12th-order reduced model, are plotted. We also plot the transfer functions of three circuit nodes located in different regions (as shown in Fig. 10). Not surprisingly, the accuracy of the reduced models becomes worse at the node (region 2) that

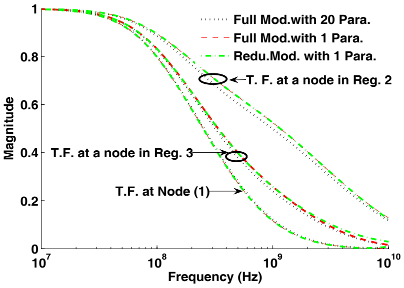

Figure 10: Accuracy of the reduced models in different regions.

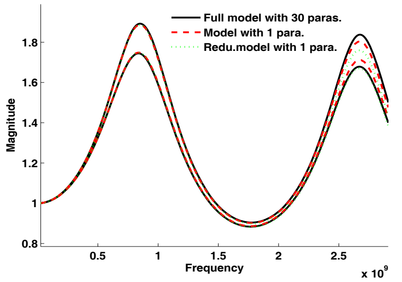

Figure 11: Frequency responses of various models for the RLC line.

is far away from the observation nodes (nodes (1) and (2) in region 5) used in the RRR procedure.

5.4.2 An RLC line

We apply the same reduction procedure to an RLC line. The line is 4 mm long and it contains 120 resistors, inductors and capacitors. We divide the line into ten regions and each region has three geometrical variations with the nominal values as: wire width $W = 1.2\ \mu m$, wire thickness $T = 1\mu m$, and dielectric layer thickness $H = 1\ \mu m$.

Again we apply the nonlinear RRR algorithm to reduce the number of the variation parameters from 30 to one , resulting a 30x reduction. Then a one-parameter reduced order model is computed, which has a size of 16. We introduce ±25% variations on all 30 geometrical parameters to generate a set of circuit samples.

In Fig. 11, two circuit samples are selected and the full model, the one- parameter full model and the one-parameter 16th-order reduced model are compared in terms of the frequency response at the output. These models show indistinguishable curves for the lower frequency band. The error produced by the reduced models in the high frequency region is well expected since the first three transfer moments expanded at DC are used as observations in the RRR procedure. Including other moments into the regression model will improve the accuracy. The linear weighing coefficients of the first three new parameters are plotted in Fig. 12. Similar trends (Fig. 6) have been observed.

6. CONCLUSIONS AND FUTURE WORK

A performance-oriented statistical parameter reduction algorithm is proposed based on reduced-rank regression. This novel approach enables us to analyze interconnect variations by reducing the cost of sampling-based simulation and generating very compact parameterized interconnect models with only a few *compressed* parameters. We are currently extending the proposed parameter reduction technique from interconnect modeling to more challenging nonlinear analog circuit problems while achieving encouraging initial results. It is expected that the proposed parameter reduction technique and its extensions will facilitate effective circuit/system-level variation modeling for a variety of circuit applications.

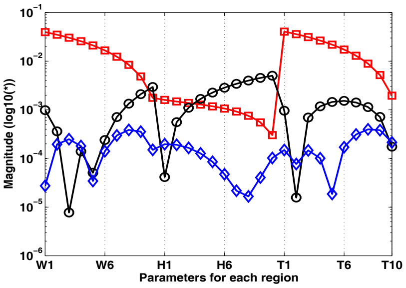

Figure 12: Weighing coefficients of the RLC line.

7. REFERENCES

[1] S. R. Nassif. Modeling and analysis of manufacturing variations. In *IEEE CICC*, pages 223–228, 2001.

[2] R. E. Bellman. *Adaptive Control Processes*. Princeton University Press, Princeton, NJ, 1961.

[3] K. K. Low and Stephen W. Director. An efficient methodology for building macromodels of ic fabrication processes. *IEEE Trans. on Computer-Aided Design*, 8(12):1299–1313, December 1989.

[4] C. Chao and L. Milor. Performance modeling using additive regression splines. *IEEE Trans. on Semiconductor Manufacturing*, 8(3):239–251, August 1995.

[5] Y. Liu, L. T .Pileggi, and A. Strojwas. Model order-reduction of rc(l) interconnect including variational analysis. In *IEEE/ACM DAC*, pages 201–206, 1999.

[6] P. Heydari and M. Pedram. Model reduction of variable-geometry interconnects using variational spectrally-weighted balanced truncation. In *IEEE/ACM ICCAD*, pages 586–591, 2001.

[7] J.Phillips. Variational interconnect analysis via pmtbr. In *IEEE/ACM ICCAD*, pages 872–879, 2004.

[8] J. Wang, P. Ghanta, and S. Vrudhula. Stochastic analysis of interconnect performance in the presence of process variations. In *IEEE/ACM ICCAD*, pages 880–886, 2004.

[9] L. Daniel, O. Siong, L. Chay, K. Lee, and J. White. A multiparameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models. *IEEE Trans. Computer-Aided Design*, 23(5):678–693, May 2004.

[10] P. Li, F. Liu, X. Li, L. Pileggi, and S. Nassif. Modeling interconnect variability using efficient parametric model order reduction. In *IEEE/ACM DATE*, pages 958–963, 2005.

[11] X. Li, P. Li, and L. Pileggi. Parameterized interconnect order reduction with explicit-and-implicit multi-parameter moment matching for inter/intra-die variations. In *IEEE/ACM ICCAD*, pages 806–812, 2005.

[12] H. Chang and S. Sapatnekar. Statistical timing analysis considering spatial correlations using a single pert-like traversal. In *IEEE/ACM ICCAD*, pages 621–625, November 2003.

[13] C. Visweswariah, K. Ravindran, K. Kalafala, S. Walker, and S. Narayan. First-order incremental block-based statistical timing analysis. In *IEEE/ACM DAC*, pages 331–336, June 2004.

[14] Y. Zhan, A. Strojwas, X. Li, and L. Pileggi. Correlation-aware statistical timing analysis with non-guassian delay distributions. In *IEEE/ACM DAC*, pages 77–82, June 2005.

[15] D. Morrison. *Multivariate Statistical Methods*. New York: McGraw-Hill, 1976.

[16] Z. Li, X. Lu, and W. Shi. Process variation dimension reduction based on svd. In *IEEE ISCAS*, pages 672–675, 2003.

[17] G. Reinsel and R. Velu. *Multivariate Reduced-Rank Regression, Theory and Applications*. Springer-Verlag, 1998.

[18] D. Ma S. Wong, T. Lee and C. Chao. An empirical three-dimensional crossover capacitance model for multilevel interconnect vlsi circuits. *IEEE Trans. Semiconductor Manufacturing,*, 13(2):219–227, May 2000.

[19] M. Stein. Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151, May 1987.