Parameter Reduction for Variability Analysis by Slice Inverse Regression (SIR) Method

Alexander Mitev Michael Marefat Dongsheng Ma Janet Wang

e-mail: {mitev,marefat,ma,wml}@ece.arizona.edu
Electrical Engineering Department
University of Arizona
Tucson, AZ - 85716
Tel: 520-621-2434
Fax: 520-621-8076

Abstract— With semiconductor fabrication technologies scaled below 100 nm, the design-manufacturing interface becomes more and more complicated. The resultant process variability causes a number of issues in the new generation IC design. One of the biggest challenges is the enormous number of process variation related parameters. These parameters represent numerous local and global variations, and pose a heavy burden in today's chip verification and design. This paper proposes a new way of reducing the statistical variations (which include both process parameters and design variables) according to their impacts on the overall circuit performance. The new approach creates an effective reduction subspace (ERS) and provides a transformation matrix by using the mean and variance of the response surface. With the generated transformation matrix, the proposed method maps the original statistical variations to a smaller set of variables with which we process variability analysis. Thus, the computational cost due to the number of variations is greatly reduced. Experimental results show that by using new method we can achieve 20% to 50% parameter reduction with only less than 8% error on average.

I. Introduction

The increasing variations of today's nanometer process technology cause circuit performance such as delay and power to deviate from their designed values. On the other hand, performance information can be used for yield improvement in manufacturing process. However, the dimensionality of design parameters depending on the number of variation sources has been ever increasing with the development of process technologies.

In previous technology generations, variability was dominated by front-end elements such as active transistors and devices. Device parameters that are susceptible to variations typically include, but are not limited to, effective gate length, mobility, gate-oxide thickness and threshold voltage. In recent technologies, back-end, or interconnect variability has become equally significant. Interconnect parameters such as metal width, metal thickness and interlayer dielectric thickness can vary significantly from their nominal values. In the era of subwavelength manufacturing, line-width variation stemming from reticle and proximity effects during lithography has become more significant with each new technology generation. In addition, interconnect thickness variation of copper interconnect due to polishing significantly affects interconnect resistance and capacitance. Thickness variation depends on local density, metal line width and spacing to the next metal line. To summarize, the number of variation sources has greatly increased as we move to the nanometer regime. To make things worse, same variation source different locations may follow different distribution. The resultant number of random variables representing variations can be of thousands. Even with clustering reduction mechanism, the random variables are still of hundreds. This fact is known as "the curse of dimensionality".

One possibility to reduce dimensionality is to apply principal component analysis (PCA) on variation sources or input random variables (denoted as x vector). PCA keeps only the first few principal components for modeling the relationship with performance or output variable (denoted as y). Because PCA based reduction is totally independent of the output variable y, it may not be effective. For example, if we have $y_1 = f_1(x)$ and $y_2 = f_2(x)$ two different functions, as long as the input variables x have the same distributions, PCA will provide the same reduction results. In reality, to address the dimension reduction issue in today's VLSI design, one must not treat x separately from performance y. This is the reason why in this paper we introduce a new reduction mechanism. In this paper we propose a novel dimension reduction method based on Sliced Inverse Regression (SIR) approach by Duan and Li [1] [2]. The new method aims to find a low dimensional subspace that carries essential information of the relationship

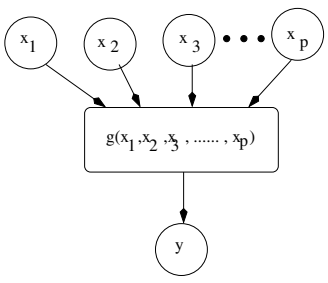

Fig. 1. General regression model with g as regression function

1-4244-0630-7/07/$20.00 ©2007 IEEE.

between the variation sources and the responses. It is easier to see how this model is related to dimension reduction by comparing Figure 1 to Figure 2[11]. Here we consider y as a univariate performance variable (e.g. delay and power). The dimension of x is denoted by p. Figure 1 shows y as a function of x vector:

$$y = g(x_1, x_2, \cdots, x_p) \tag{1}$$

As one can see in Figure 1, there are p input nodes. The function g is represented by the black box in the middle, feeding into the output node y at the bottom. Figure 2 has an intermediate layer of nodes which linearly combine the input nodes using weights indicated along the line segments. It is clear that the relationship between x and y is determined only through $\beta_1^t, \beta_2^t, \cdots, \beta_K^t$. The black box represents the function m in (2).

$$y = m(\beta_1^t x, \beta_2^t x, \cdots, \beta_K^t x) \tag{2}$$

Our goal is to find the K p-dimensional subspace $span\{\beta_1^t, \beta_2^t, \cdots, \beta_K^t\}$. Taking advantage of the generated subspace, the proposed method obtains implicitly new response surface models that operate on a reduced dimension variable set. Different from existing dimension reduction approaches such as clustering and PCA, the new SIR based method not only considers the behavior of dependant variables but also the response surface impact on dimension reduction.

Note that the dimension reduction approach we discuss here is also different from Projection-Based Extraction (PROBE) for quadratic response surface modeling [3]. While PROBE greatly reduces number of sampling points required to achieve an accurate quadratic response surface model, the proposed approach targets parameter reduction in the response surface model. In addition, the new method is applicable to all existing digital and analog designs where dimension becomes a problem. In the current paper we apply the new method to block-based SSTA. All gate/interconnect delays and signal arrival times are represented in quadratic form over a base set of variational process parameters. From experiments, we verify that nonlinear delays at all levels can be accurately approximated by quadratic models with much less number of variables.

The organization of the rest of the paper is as follows: In Section 2, we focus on explaining the basics for dimensionality reduction. Section 3 discusses the reduction framework in block-based SSTA and the new SIR based reduction algorithm. Experimental results and algorithm complexity are discussed in Section 4. Finally, Section 5 concludes the current paper.

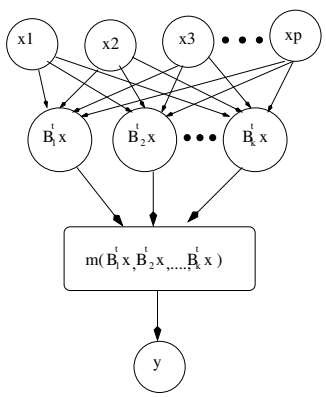

Fig. 2. Reduced regression model with m as new after reduction function

II. Effective Dimension Reduction

We start this section by reviewing PCA method, then introduce effective dimension concept and finally provide the basics of the slice inverse approach.

A. Principal component analysis (PCA)

No doubt that PCA is perhaps the most popular procedure of dimension reduction. It in fact is a special case for the Sliced Inverse Regression (SIR) approach. This is why in this subsection, we first take a brief look at PCA. PCA projects the high dimensional space to a lower dimensional space with the hope that the essential structure in the original space can be kept as much as possible. The projected space is chosen so that the distributions can spread out as much as possible.

Assume x is the p-dimensional variable of interest. The first principal component is a linear combination x denoted as $b_1^t x$ which has the largest variance among all b with unitary length. Equation (3) provides the description of such procedure.

$$max_{||b||=1} b^t \Sigma_x b \tag{3}$$

Here Σ_x denotes the covariance matrix of x. After finding the first direction b_1, we repeat the same procedure by restricting to those that yield projections uncorrelated with $b_1^t x$: To illustrate, we get the the second principal direction b_2 by using (4).

$$cov(b^t x, b_1^t x) = b^t \Sigma_x b_1 = 0 \tag{4}$$

Continuing this process in (4), we can get all other directions $b_3, b_4, \cdots, b_p$. It can be shown that

$$\Sigma_x b_i = \lambda_i b_i \tag{5}$$

Here the variance of $b_i^t x$ is λ_i.

The above equations provide the original concept of PCA. However they may give readers the impression that PCA is a complicated procedure. In fact, we need only to conduct the eigenvalue decomposition on the covariance matrix of x. Eigenvalues of PCA often decrease rapidly. This is because most of the variables spread out wide along the first few directions. Therefore, it is possible to have dimension reduction while still keep the original structure of the input variable space. Unfortunately, the reduction is not a guarantee. To achieve correct PCA reduction, one often needs to rescale each input variable appropriately before applying PCA. Standard deviation is generally selected as rescaling factor.

PCA provides principal dimensions for input variables x. Principal dimensions, however as mentioned in the introduction part of this paper, can not reflect the influence of x on output variable y. As long as the input variables x have the same distribution, even if we have two different functions, that is, the relationship between x and y is not the same, the same x input variables would always reduce to the same linear combinations. Therefore, principal dimensions are not

the effective ones if we consider the output variable y. Finding effective dimensions is the key point of this paper and will be further discussed in the next section.

B. Effective Dimension Reduction

Effective dimension reduction (EDR) concept is the center of the SIR based reduction scheme.

Definition 1: Under (2), the space $\mathcal{B}$ generated by $\beta_1^t, \beta_2^t, \cdots, \beta_K^t$ is called the EDR space. Any non-zero vector in the EDR space is called an EDR direction.

By changing m suitably, one can reparameterize (2) by any set of K linearly independent EDR directions. Thus it is the EDR space B that can be identified; the individual vectors β_1^t, β_2^t, $\cdots$,β_K^t themselves are not identifiable unless further structural conditions on g are imposed. Our primary goal of this paper is to find the EDR space or a subspace of it.

Though (2) should be interpreted as an approximation to reality, the fundamental difference between this and other statistical models is that (2) takes the weakest form to reflect our hope that a low dimensional projection of a high dimensional regressor variable contains most of the information that can be gathered from the original space. (2) does not impose any structures on how the projected variable effects the output variable. In addition, we may vary K to reflect the degree of the anticipated dimension reduction. To illustrate, at $K = p$, (2) becomes a redundant assumption. We want to emphasize that the estimation of the projection directions can be a more important statistical issue than the estimation of the function m itself. In fact, the structure of m is impossible to identify unless we have other evidence. For example, let m be the delay model after reduction in SSTA analysis, we can alway assume that m is a quadratic polynomial. In reality, one can obtain two different versions of m to represent the same joint distribution of y and x. Thus what we can estimate at most are statistical quantities such as the conditional mean or quantities of y given x. On the other hand, at the beginning stage of analysis when one does not have a fixed objective in mind, the need for estimating such quantities is not as pressing as that for finding ways to simplify the design space. Our formulation of estimating the EDR directions is one way to address such a need in analysis. After finding a good EDR space, we can project data to this smaller space. Then we are in a better position to identify what should be pursued further: model building by response surface estimation after reduction.

(2) is equivalent to: the conditional distribution of y given x depends on x only through the K dimensional variable ($\beta_1^t x$, $\beta_2^t x$, $\cdots$, $\beta_K^t x$), or, to put it slightly differently, conditional on $\beta_1^t x$, $\beta_2^t x$, $\cdots$,$\beta_K^t x$, y and x are independent. The reduced variable, $\beta_1^t x$, $\beta_2^t x$, $\cdots$,$\beta_K^t x$ is as informative as the original x in predicting y.

The key in the notion of EDR space is to find the one with the smallest dimension. Now one question arise: is this space unique? Cook [8] explored the answer to this question. It turns out that under certain regularity conditions, the EDR space with the smallest dimension is unique. We shall assume this is the case from now on.

C. Basics of Sliced Inverse Regression (SIR) based Dimension Reduction

In this subsection, we introduce a method for finding EDR directions – sliced inverse regression (SIR). First we present a theory for justifying SIR.

As pointed out by Li in [11], the natural way to think about response surface estimation is going from x to Y. Of special importance is the first moment $E(Y|x)$ or the second moment $var(Y|x)$. Unlike general Response Surface approaches, SIR does not follow the above one-way traffic of going from x to Y. Instead, SIR reverses the role of x and Y. That is, SIR treats Y as if it were the independent variable and treat x as if it were the dependent variable. This fundamental difference can be further illustrated as "given $x = x_0$, what value will Y take?" One straightforward conventional answer would be examining the data points close to x_0 and taking their average Y values. This is the basic idea of SIR methodology. To gain global insight on Y, SIR based reduction scheme studies how the associated x region varies as Y changes.

One advantage is immediate. The general response surface $E(Y|x)$ is p-dimensional, which is very difficult to estimate directly. When p is large, even with some "smart" sampling techniques, the resulting response surface techniques may perform poorly because of lack of sufficient data points in some relevant local region. However, for inverse regression, the conditional expectation $E(x|Y)$ can be taken one coordinate at a time $E(x_i|Y)$, for $i = 1,\cdots,p$. The estimation of $E(x_i|Y)$ should be easy to handle because this is just a one-dimensional problem. This is the reason why we can, from a different angle, take care of the curse of dimensionality problem.

The most important question remaining is how to relate inverse regression to forward regression. To fill up the gap, we shall derive Theorem 1, which is the foundation of the SIR theory. Generally speaking, inverse regression factorizes the joint density of x and y into the condition density $h(x|y)$ and the marginal density $k(y)$. While only $E(x|y)$ is considered in this section, other quantities can be utilized as well. For example, we shall also discuss how to use conditional covariance $cov(x|y)$ for extending the basic SIR algorithm.

Theorem 1: Assume the condition provided in (2) is true. In addition, for any b in $\mathcal{R}^p$, the conditional expectation $E(b^t x|\beta_1^t x, \beta_2^t x, \cdots, \beta_K^t x)$ is linear in $\beta_1^t x$,$\beta_2^t x$,$\cdots$, $\beta_K^t x$; that is, for some constants, c_0, c_1, $\cdots$, c_K,

$$E(b^t x|\beta_1^t x, \beta_2^t x, \cdots, \beta_K^t x) = c_0 + c_1\beta_1^t x + \cdots + c_K\beta_K^t x \quad (6)$$

then, the difference of $E(x|y) - E(x)$ is contained in the linear subspace spanned by $\Sigma_x\beta_k$, $k = 1, 2, \cdots, K$, where Σ_x denotes the covariance matrix of x.

The condition (6) is also referred as Linear Design Condition (LDC). It has been proved that the sufficient condition for LDC is elliptical symmetric.

Definition 2: A random vector $X \in \mathcal{R}^n$ with probability density function f is said to to be elliptical symmetric if f can be written as

$$f(x) = g(||A(x - \mu)||) \quad (7)$$

where A is a positive definite symmetric matrix and vector $\mu \in \mathcal{R}^n$, and g is a function $g : [0, \infty) \to [0, \infty)$.

The elliptical symmetric property holds for the majority of statistical problems. We will show in the later sections that elliptical symmetry is also applicable to process variation problems.

Theorem 1 relates the inverse regression to regular response surface methods. To illustrate, we present a simple example with the standardized variable $z = {\Sigma_x}^{-\frac{1}{2}}x - E(x)$ where $\Sigma_x^{-\frac{1}{2}}$ is the covariance matrix of x. the inverse response (IR) curve $m_1(y) = E(z|Y = y)$ lies in $span(\eta_1, \eta_2, \cdots, \eta_k)$, where $\eta_i = \Sigma_x^{\frac{1}{2}}\beta_i$. This means that the conditional expectation $m_1(y) = E(z|Y = y)$ is moving in $span(\eta_1, \eta_2, \cdots, \eta_k)$ depending on y. With b orthogonal to $span(\eta_1, \eta_2, \cdots, \eta_k)$, it follows that

$$b^t m_1(y) = 0 \tag{8}$$

and further that

$$m_1(y)m_1^t(y)b = Cov\{m_1(y)\}b = 0 \tag{9}$$

As a consequence $Cov\{E(z|y)\}$ is degenerated in each direction orthogonal to all EDR-directions η_i of Z. This suggests the following steps. First, estimate $Cov\{m_1(y)\}$ and then calculate the orthogonal directions of this matrix (for example, with eigenvalue/eigenvector decomposition). In general, the estimated covariance matrix will have full rank because of random variability, estimation errors and numerical imprecision. Therefore, we investigate the eigenvalues of the estimate and ignore eigenvectors having small eigenvalues. These eigenvectors $\hat{\eta_i}$ are estimates for the EDR-direction η_i of z.

We can easily rescale them to estimate $\hat{\beta}_i$ for the EDR-directions of x by multiplying by $\hat{\Sigma}^{-\frac{1}{2}}$, but then they are not necessarily orthogonal.

III. Sliced Inverse Regression Algorithm for Parameter Reduction

As an effective reduction technique, SIR can be applied to any performance driven design methodology. In the current paper, we employ the proposed method to statistical static timing analysis as an example.

A. Static Statistical Timing Analysis (SSTA) with Sliced Inverse Regression based Parameter Reduction

Let's now start with the sum and max operations for SSTA. We assume delay models as quadratic functions with correlated random variables. For all quadratic functions, we can represent them in the following format:

$$y = x^t Ax + Bx + C \tag{10}$$

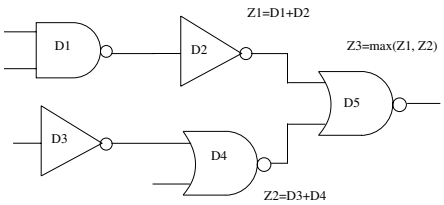

Fig. 3. Block-wise Static Statistical Timing Analysis (SSTA) Basic Operations

in which $x = x_1, x_2, \cdots, {x_n}^t$ is the process parameter vector, A is a coefficient matrix, B is a $1 \times n$ vector and C is a constant term. Assume we have two random variables, $y_1 = \tilde{x}^t A_1 \tilde{x} + B_1\tilde{x} + C_1$, and $y_2 = \hat{x}^t A_2 \hat{x} + B_2\hat{x} + C_2$. Here we differentiate $\tilde{x}$ from $\hat{x}$, as in reality gates and interconnects may be far apart from each other. y_1 and y_2 may be delays that depend on two different sets of process parameters. This is different from most existing references in which $\tilde{x}$ is regarded as the same set as $\hat{x}$. The sum operation gives:

$$y = sum(y_1, y_2) = x^t Ax + Bx + C$$
$$A = A_1 + A_2, \quad B = B_1 + B_2, \quad C = C_1 + C_2 \tag{11}$$

where $x = \tilde{x}U\hat{x}$. We abuse the notation "+" in A and B computation as the sum operation adds up the coefficients of the same random variables. Likewise, we can present the max operation as

$$y = max(y_1, y_2) = x^t Ax + Bx + C \tag{12}$$

The max operation may follow ideas presented in [4]-[7]. Here we only focus on the sum and max operation results y which is also a quadratic function with $x = \bar{x}U\hat{x}$. If the sizes of the vectors $\bar{x}$ and $\hat{x}$ are $\bar{n}$ and $\hat{n}$ respectively, then the size of x n_x will follow $n_x \geq \bar{n}$ and $n_x \geq \hat{n}$. When the number of parameters exceed certain user defined threshold (as shown in Fig.4), SIR based reduction procedure will be employed.

B. Sliced Inverse Regression based Parameter Reduction Algorithm

Suppose the PDF function of x is $f(x)$ and $x \in [a_x, b_x]$. The PDF function $f(x)$ is positive so as to a_x and b_x for most process variation cases. Even if x is negative, by shifting and linear transformation, we can always get the new x with the desired property as in Definition 2, that is, elliptical symmetric property. Therefore, Theorem 1 can be easily satisfied.

The algorithm of SIR consists of the following steps:

Step 1: Standardize x. Let $\bar{x}$ denote the mean of random variable vector x, $\hat{\Sigma}_x$ covariance of x, and z be the standardized vector, we have

$$z = \hat{\Sigma}_x^{-\frac{1}{2}}(x - \bar{x}) \tag{13}$$

with $\bar{x} = \int_{a_x}^{b_x} xf(x)dx$ and $\hat{\Sigma}_x = \int_a^b (x - \bar{x})(x - \bar{x})^t f(x)dx$. The standardized vector z is $\in [b_z, a_z]$. $b_z = \hat{\Sigma}_x^{-\frac{1}{2}}(b_x - \bar{x})$

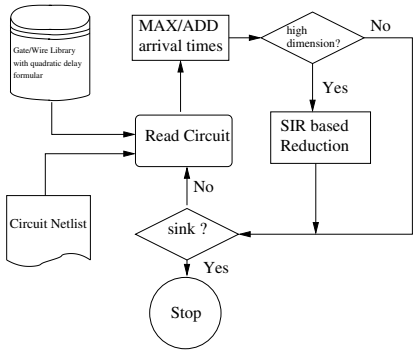

Fig. 4. Static Statistical Timing Analysis (SSTA) flow with Sliced Inverse Regression based reduction

and $a_z = \hat{\Sigma}_x^{-\frac{1}{2}}(a_x - \bar{x})$. As $x = \hat{\Sigma}_x^{\frac{1}{2}} z + \bar{x}$, we can represent PDF function $f(x)$ as $f(\hat{\Sigma}_x^{\frac{1}{2}} z + \bar{x})$ or $f_z(z)$.

Step 2: Sort the data by y. With (11)-(12) and known x data, we can generate data for y denoted as y_i, where $i = 1, \cdots, N$. We can also use the fact that $y = g(x)$ is a continuous function. When $y = g(x)$ is linear, 3rd order or any odd order function, g^{-1} exists. Therefore, $x = g^{-1}(y)$. Sorting y is straightforward as g^{-1} is monotonic. If $y = g(x)$ is quadratic (as assumed in the current paper), then we first divide y into two monotonic regions by finding the optimal points $\partial g(x)/\partial x = 0$. Let x_{opt} be the optimal point of x, $y = g_1(x)$ be the function for $x \in [a_x, x_{opt}]$ and $y = g_2(x)$ be the function for $x \in (x_{opt}, b_x]$. Both g_1 and g_2 have inverse functions g_1^{-1} and g_2^{-1}. We may apply the following Step 3 on both g_1^{-1} and g_2^{-1} defined in region $[g_1(a_x), g_1(x_{opt})]$ and $(g_2(x_{opt}), g_2(b_x)]$.

Step 3: Divide the range of y into S nonoverlapping intervals (slices) H_h,$h = 1, 2, \cdots, S$. The number of slices S is a user-specified parameter. For example, we find between 10 to 20 slices to be reasonable for a sample of size 300. There are theoretical results indicating that SIR outputs do not change much for a wide range of S[1].

Step 4: Compute the mean of z over all slices. This is a crude estimate for the inverse regression curve $m_1(y) = E(z|Y = y)$:

$$\bar{z}_h = \int_{a_h}^{b_h} z f_z(z) dz \tag{14}$$

where f_z is the PDF for z, and a_h and b_h are the starting and ending point of z for each slice. Note that SIR uses Y values only to create slices. Once slices are formed, they can be discarded.

Step 5: Calculate the estimate for the covariance of $m_1(y) = E(z|Y = y)$ denoted as $Cov(m_1(y)) = \hat{\Sigma}_\eta$:

$$\hat{\Sigma}_\eta = (b_z - a_z)^{-1} \sum_{h=1}^{S} (b_h - a_h)(\bar{z}_h - \bar{z})(\bar{z}_h - \bar{z})^t \tag{15}$$

Step 6: Find the SIR directions by conducting the eigenvalue decomposition of $\hat{\Sigma}_\eta$ with respect to $\hat{\Sigma}_x$:

$$\hat{\Sigma}_\eta \hat{\eta}_i = \hat{\lambda}_i \hat{\Sigma}_x \hat{\eta}_i \tag{16}$$

The i-th eigenvector $\hat{\beta}_i$ is called the i-th SIR direction with ordering: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. The first few SIR directions can be used for dimension reduction. They serve as the coefficients linking the input nodes to the intermediate nodes in Figure 7.

Step 7: Project z along the SIR directions; that is, use each SIR direction to form a linear combination of z. We shall call $\hat{\beta}_1^t z$ the first SIR variate, the $\hat{\beta}_2^t z$ second SIR variate, and so on. Now the estimates for the EDR-directions are given by

$$\hat{\beta}_i = \hat{\Sigma}_x^{-\frac{1}{2}} \hat{\eta}_i \tag{17}$$

Step 8: Evaluate the effectiveness of an estimated EDR direction. An obvious criterion is to evaluate the squared Euclidean distance between the estimated EDR direction $\hat{\beta}_i$ and the true e.d.r. space $\mathcal{B}$.

$$R^2(\hat{\beta}_i) = max_{\beta \in \mathcal{B}} \frac{(\hat{\beta}_i^t \Sigma_x \beta)^2}{(\hat{\beta}_i^t \Sigma_x \hat{\beta}_i \dot{\beta}^t \Sigma_x \beta} \tag{18}$$

the squared multiple correlation coefficient between the projected variable $\hat{\beta}_i^t x$ and the ideally-reduced variables $\beta_1^t x, \beta_2^t x, \cdots, \beta_K^t x$. For a collection of K estimated directions $\hat{\beta}_1$, $\hat{\beta}_2$, $\cdots$, $\hat{\beta}_K$ which generate a linear subspace $\mathcal{B}$. As $\mathcal{B}$ will not be known beforehand, we replace β with $\hat{\beta}$ vectors to obtain the relative squared correlation. After SIR parameter reduction, we reconstruct function $m()$ in $y = m(\hat{\beta}_1^t x, \hat{\beta}_2^t x, \cdots, \hat{\beta}_K^t x)$ by minimum residue based curve fitting. Then $m()$ function joins further sum and max operations.

We can summarize SIR based reduction as follows: (1) partitioning the cases into H groups according to the y values; (2) finding the H slice means of z ; (3) applying a eigendecomposition on covariance of z. It is important to remember that our use of eigendecomposition differs from PCA. We use y to form slices while PCA does not use any information from y at all.

C. Special Cases

If the covariance of $x = (x_1, x_2, \cdots, x_p)$ is an identity matrix I, then all random variables x_i, $i = 1, \cdots, p$, have the same variance (=1) and are uncorrelated with each other. Then on the right side of the equality in (17), the matrix $\hat{\Sigma}_x$ can be removed. Thus Step 6 is merely eigendecomposition applied to the slice means of z just like PCA in (5). However, it is important to remember that our use of eigendecomposition differs from PCA. We use y to form slices while PCA does not use any information from y at all.

The proposed SIR based reduction method is able to provide EDR directions in all our experimental cases (Section IV). However, theoretically, SIR based method may have trouble in finding EDR directions when $E(z|y) = 0$. We overcome this difficulty by considering the conditional covariance $Cov(z|y)$ instead of the inverse response curve. Therefore, in the aforementioned SIR reduction algorithm, we can replace $E(z|y)$ with $Cov(z|y)$. This kind of technique is referred as SIR II in some references [9][10].

IV. Numerical examples

In this section we demonstrate the results of the proposed parameter reduction method by a number of examples.

First, lets take a look at a simple example as shown in Figure 3. Assume we use quadratic delay model. At D_1, the output delay depends on x_1, x_2, x_3 that is $z_1 = y_1(x_1, x_2, x_3)$. Likewise, at D_4, $z_2 = y_4(x_4, x_5, x_6)$ and D_3, $z_3 = y_5(x_7, x_8, x_9)$. All the input variables x_1 to x_9 are correlated. The delay model at the output of $D5$ is $y = max(sum(z_1, z_3), sum(z_2, z_3))$,essentially a quadratic expression with respect entire variable space $(x_1, ..., x_9)$. The new reduction algorithm reduces the original 9 variable design space to 3, a reduction of 60%. However if PCA is applied, the reduced space will consider the 6 variables, leading to only 34% reduction. In addition, the PCA reduction in this case does not work well. Since the PCA reduction is totally independent of performance, not only PCA produces less accurate results (from PDF distribution), but also provide less reduction. The reconstruction of new function $m(\hat{\beta}_1^t x, \hat{\beta}_2^t x, \cdots, \hat{\beta}_K^t x)$ is carried through least square

TABLE I
RESULTS OF SIR ON ISCAS 85 BENCHMARKS

Circuit	Number of Gates	PCA delay mean error (%)	New delay mean error (%)	PCA delay variance error (%)	New delay variance error (%)	PCA reduction (%)	New reduction (%)
C17	6	0.5	0.4	6.5	4.3	25	42
C432	160	2.3	1.9	9.2	7.1	37	61
C499	202	5.4	2.2	11.9	8.5	23	39
C880	383	3.0	3.0	6.0	6.0	44	44
C1355	546	3.2	3.0	5.5	6.0	30	38
C1908	880	1.1	1.0	6.7	7.2	10	41
C2670	1193	3.1	2.9	5.0	4.5	20	30
C3540	1669	5.4	4.0	8.0	8.0	25	35
C5315	2307	5.0	5.0	8.3	8.3	35	35
C6288	2416	1.6	2.6	6.6	7.0	23	36
C7552	3512	2.7	2.2	4.0	7.4	20	53

approximation. Figure 5,6 demonstrates the distributions of the delay distribution at the outputs of D4 and D5. The solid line (red) represents the result from our proposed approach. The dash line (blue) is the result from Monte Carlo. The doted-dash line (green) comes from PCA.

We also test our proposed method on ISCASS'85 benchmark circuits for 0.13 μm. Assume we fix the yield at 98%. The simulations are performed in the following way: for particular output, we assigned different process variation distributions for all gates through the related paths. Further we assume quadratic timing model with tree process variations $(Weff, Leff, Tox)$ for each gate model. The model coefficients are determined by response surface method. All simulations were run on Pentium 2GHz. The results show clearly the advantage of using the proposed reduction scheme over PCA: new method can achieve 20% to 50% parameter reduction with only less than 8% error on average.

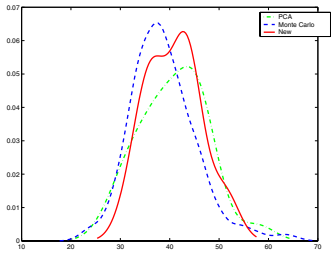

Fig. 5. Delay PDF distribution comparison at the output of D_4

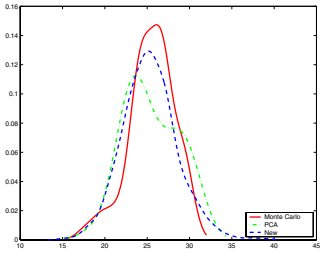

Fig. 6. Delay PDF distribution comparison at the output of D_5

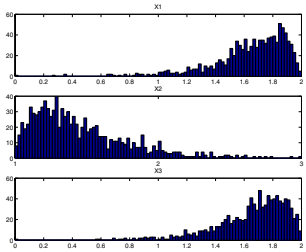

Fig. 7. New variables and their PDF distribution after new reduction approach

V. CONCLUSION

This paper proposes a new way of reducing the statistical variations. The new approach creates an effective reduction subspace (ERS) and provides a transformation matrix by using the mean and variance of the response surface. With the generated transformation matrix, the proposed method maps the original statistical variations to a smaller set of variables with which we process variability analysis. Thus, the computational cost due to the number of variations is greatly reduced.

REFERENCES

[1] K.C. Li, "Sliced inverse regression for dimension reduction", with discussions, J. Amer. Statist. Assoc. 86. 316-342, 1991
[2] N. Duan and K.C. Li, "Slicing Regression: a Link-Free Regression Method", Ann. Stat. 19, 505-530. 1991
[3] X. Li, J. Le, L. Pileggi, and A. Strojwas, "Projection-Based Performance Modeling for Inter/Intra-Die Variations", International Conference of Computer-aided Design (ICCAD) 2005
[4] H. Chang, V. Zolotov, S. Narayan, and C. Visweswariah, "Parameterized Block-Based Statisitical Timing Analysis with Non-Gaussian Parameters, Nonlinear Delay Functions", Design Automation Conference (DAC) 2005
[5] Y. Zhang, A. Strojwas, X. Li, and L. Pileggi, "Correlation-Aware Statistical Timing Analysis with Non-Gaussian Delay Distributions", Design Automation Conference (DAC) 2005
[6] L. Zhang, W. Chen, Y. Hu, J. Gubner, and C. Chen, "Correlation-Preserved Non-Gaussian Statistical Timing Analysis with Quadratic Timing Model", Design Automation Conference (DAC) 2005
[7] V. Khandelwal, and A. Srivastava, "A General Framework for Accurate Statistical Timing Analysis Considering Correlations", Design Automation Conference (DAC) 2005
[8] R.D. Cook, "On the Interpretation of Regression Plots", J. Amer. Stat. Assoc., 89, 177-189, 1994
[9] R.D. Cook, and S. Weisberg, "Discussion of "Sliced inverse regression", J. Amer. Stat. Assoc., 86, 328-332, 1991.
[10] K.C. Li, "Uncertainty analysis for mathematical models with SIR", in Probability and Statistics, 138-162, 1992.
[11] K.C. Li, "Dimension reduction and data visulaization", Lecture Notes, http://www.stat.ucla.edu/ kcli/