

Statistical Regression for Efficient High-Dimensional Modeling of Analog and Mixed-Signal Performance Variations

Xin Li

Department of ECE, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
xinli@ece.cmu.edu

Hongzhou Liu

Cadence Design Systems
246 Alpha Drive, Pittsburgh, PA 15238
hliu@cadence.com

ABSTRACT

The continuous technology scaling brings about high-dimensional performance variations that cannot be easily captured by the traditional response surface modeling. In this paper we propose a new statistical regression (STAR) technique that applies a novel strategy to address this high dimensionality issue. Unlike most traditional response surface modeling techniques that solve model coefficients from over-determined linear equations, STAR determines all unknown coefficients by moment matching. As such, a large number of (e.g., $10^3\sim 10^5$) model coefficients can be extracted from a small number of (e.g., $10^2\sim 10^3$) sampling points without over-fitting. In addition, a novel recursive estimator is proposed to accurately and efficiently predict the moment values. The proposed recursive estimator is facilitated by exploiting the interaction between different moment estimators and formulating the moment estimation problem into a special form that can be iteratively solved. Several circuit examples designed in commercial CMOS processes demonstrate that STAR achieves more than 20x runtime speedup compared with the traditional response surface modeling.

Categories and Subject Descriptors

B.7.2 [Integrated Circuits]: Design Aids – Verification

General Terms

Algorithms

Keywords

Response Surface Modeling, Process Variations, Circuits

1. INTRODUCTION

As IC technologies are scaled to nanoscale regime, it becomes increasingly difficult to control the variations in manufacturing process [1]. Process variations can be classified into two broad categories: inter-die variations and intra-die variations. Inter-die variations model the common/average variations across the die, while intra-die variations model the individual, but spatially correlated, local variations (e.g., random device mismatches) within the same die. Both inter-die and intra-die variations introduce substantial uncertainty in circuit behavior and significantly impact parametric yield. Hence, accurately modeling and analyzing process variations to ensure manufacturability has been identified as one of the most critical problems for today's IC design.

To address this variation issue, response surface modeling

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DAC 2008, June 8–13, 2008, Anaheim, California, USA

Copyright 2008 ACM 978-1-60558-115-6/08/0006...5.00

(RSM) has been widely applied to various statistical circuit analysis problems [2]-[8], [12]. The objective of response surface modeling is to approximate the circuit performance (e.g., delay, gain) as an analytical (either linear or nonlinear) function of process parameters (e.g., V_{TH} , T_{OX}). Once response surface models are fitted from transistor-level simulation data, they can be used to predict performance distributions and, more importantly, identify the critical device-level variation sources that cause circuit-level performance variations. While response surface modeling was extensively studied in the past, two recent trends in advanced IC technologies suggest a need to revisit this area.

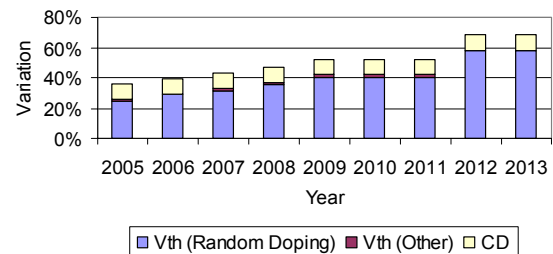


Fig 1. Predicted relative process variations ($3\sigma/\text{mean}$) [1].

Firstly, among all sources of variations, random mismatches are expected to become dominant within the next few technology generations [1], as shown in Fig 1. Hence, any two transistors on the same die can have significantly different electrical behaviors (e.g., mobility, V_{TH} , etc.). To accurately model this effect, a large number of random variables must be utilized, rendering a *high-dimensional* variation space. Secondly, the aggressive scaling of VLSI technologies makes it possible to integrate more and more devices onto a single chip, further increasing the problem size of statistical performance modeling. As will be demonstrated by the numerical examples in Section 5, more than 20,000 independent random variables must be used to model the inter-die and intra-die variations of a simplified SRAM critical path in a commercial 65nm CMOS process. In this case, even if a linear model template is used, we must solve more than 20,000 unknown model coefficients!

Recently, a number of advanced response surface modeling techniques [5]-[8] have been proposed to address this high dimensionality issue. The key idea of these methods is to find an optimal projection subspace to map the high-dimensional modeling problem to a low-dimensional problem that is easy to solve. Such a projection strategy substantially reduces the number of unknown model coefficients and, therefore, only a small number of simulation samples are required to determine all model coefficients. The aforementioned projection techniques have been successfully applied to many practical applications with 10~100 variables. However, they seem still ill-equipped to address the modeling challenges of today's mixed-signal IC designs that contain $10^3\sim 10^5$ variables. How do we make response surface modeling feasible for such a *large* problem size?

Unlike the projection-based techniques, we attempt to approach the high dimensionality issue from a completely different avenue. We aim to develop an efficient numerical algorithm that allows us to fit a large number of model coefficients from a small set of sampling points. In general, all existing modeling techniques solve model coefficients from over-determined linear equations. As such, the number of sampling points must be equal to or greater than the number of model coefficients. In contrast, we ask the following fundamental question: *can we uniquely determine more model coefficients than sampling points?* Certainly, this is not feasible if the coefficients are solved from deterministic equations. However, we will show in this paper that such an ambitious task is possible if we determine the model coefficients *statistically*.

In this paper, we propose a novel statistical regression (STAR) algorithm for high-dimensional linear/nonlinear response surface modeling. STAR determines all unknown model coefficients from a set of random samples by *moment matching*. The proposed moment matching approach is derived from statistics theory [9], [11], [13]. It is especially tuned for circuit modeling applications such that a large number of (e.g., $10^3\sim 10^5$) unknown model coefficients can be accurately and efficiently extracted from a small set of (e.g., $10^2\sim 10^3$) sampling points without over-fitting. We theoretically prove that such a moment-matching strategy is optimal where the weighted fitting error is minimized.

An important contribution of this paper is to develop a new *recursive estimator* to accurately predict the moment values. The proposed recursive estimator is facilitated by exploiting the interaction between different moment estimators and formulating the moment estimation problem into a special form that can be iteratively solved. It is provably unbiased. Most importantly, compared with the direct estimator that is traditionally used in statistics, our recursive estimator offers substantial (more than 20x for our examples in Section 5) error reduction and, therefore, makes the proposed moment matching approach of practical interest and utility.

The remainder of this paper is organized as follows. In Section 2, we review the background on principal component analysis and response surface modeling. Then, we propose our STAR algorithm in Section 3. Several implementation issues, including the recursive estimator, are discussed in detail in Section 4. The efficacy of STAR is demonstrated by several numerical examples in Section 5, followed by the conclusions in Section 6.

2. BACKGROUND

2.1 Principal Component Analysis

Given N process parameters $X = [x_1 \ x_2 \ \dots \ x_N]^T$, the process variations $\Delta X = X - X_0$, where X_0 contains the mean values of X , are often approximated as zero-mean, correlated Normal distributions. Principal component analysis (PCA) [10] is a statistical method that finds a set of independent factors to represent the correlated Normal distributions. Assume that the correlations of ΔX are represented by a symmetric, positive semi-definite covariance matrix R . PCA decomposes R as:

$$R = U \cdot \Sigma \cdot U^T \quad (1)$$

where $\Sigma = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ contains the eigenvalues of R , and $U = [U_1 \ U_2 \ \dots \ U_N]$ contains the corresponding eigenvectors that are orthonormal, i.e., $U^T U = I$. (I is the identity matrix.) Based on Σ and U , PCA defines a set of new random variables $\Delta Y = [\Delta y_1 \ \Delta y_2 \ \dots \ \Delta y_N]^T$:

$$\Delta Y = \Sigma^{-0.5} \cdot U^T \cdot \Delta X. \quad (2)$$

These new random variables in ΔY are called the principal components or factors. It is easy to verify that all principal components in ΔY are independent and standard Normal (i.e., zero mean and unit standard deviation).

The essence of PCA can be interpreted as a coordinate rotation of the space defined by the original random variables. In addition, if the magnitude of the eigenvalues $\{\lambda_n; n = 1, 2, \dots, N\}$ decreases dramatically, it is possible to use a small number of random variables, i.e., a small subset of principal components, to approximate the original N -dimensional space. More details on PCA can be found in [10].

2.2 Response Surface Modeling

Given a circuit design, the circuit performance f (e.g., delay, gain) is a function of the process variations ΔY defined in (2). If the process variations are sufficiently small, the performance function $f(\Delta Y)$ can be approximated as a linear response surface model [2]-[8], [12]:

$$f(\Delta Y) \approx \sum_{i=1}^N \alpha_i \cdot \Delta y_i + \alpha_0 \quad (3)$$

where $\{\alpha_n; n = 0, 1, \dots, N\}$ are the model coefficients. If the performance function $f(\Delta Y)$ is strongly nonlinear due to large-scale variations, nonlinear (e.g., quadratic) response surface models may be used to improve modeling accuracy [2], [5], [12].

Traditionally, the unknown model coefficients in (3) are determined by solving the over-determined linear equations at a number of sampling points [12]:

$$\sum_{i=1}^N \alpha_i \cdot \Delta y_i^{(m)} + \alpha_0 = f^{(m)} \quad (m = 1, 2, \dots, M) \quad (4)$$

where $\Delta y_i^{(m)}$ and $f^{(m)}$ are the values of Δy_i and $f(\Delta Y)$ at the m -th sampling point respectively, and M is the total number of sampling points. To solve the over-determined linear equations in (4), the number of samples must be equal to or greater than the number of coefficients. Taking the linear response surface model in (3) as an example, the total number of model coefficients is $N+1$ and, therefore, at least $N+1$ simulation samples are required to determine these coefficients. It, in turn, makes response surface modeling extremely expensive, if N is large (e.g., around $10^3\sim 10^5$). For this reason, the traditional response surface modeling is limited to small-size or medium-size applications (around 10~100 variables). In this paper, we propose a novel response surface modeling scheme that aims to solve a large number of model coefficients from a small set of sampling points with low computational cost.

3. STATISTICAL REGRESSION

Most traditional response surface modeling techniques rely on two major assumptions: (1) all model coefficients are solved from over-determined linear equations; and (2) all sampling points are deterministically located and the locations are optimized by design of experiments (DOE) [16]. In contrast, our proposed statistical regression (STAR) applies a completely different strategy: it randomly distributes all sampling points in the variation space and statistically extracts the model coefficients by moment matching. In this section, we describe the mathematical formulation of STAR and prove that the proposed statistical regression is optimal where the weighted fitting error is minimized.

3.1 Mathematical Formulation

Given the linear response surface model in (3) where all random variables in ΔY are independent and standard Normal, it is easy to verify the following relations:

$$E(f) = E\left(\sum_{i=1}^N \alpha_i \cdot \Delta y_i + \alpha_0\right) = \alpha_0 \quad (5)$$

$$E(f \cdot \Delta y_n) = E\left[\left(\sum_{i=1}^N \alpha_i \cdot \Delta y_i + \alpha_0\right) \cdot \Delta y_n\right] = E(\alpha_n \cdot \Delta y_n^2) = \alpha_n \quad (6)$$

$(n=1,2,\dots,N)$

where $E(\bullet)$ represents the expected value. Eqn. (5)-(6) imply an important fact that the model coefficients $\{\alpha_n; n=0,1,\dots,N\}$ can be uniquely determined by the moments $E(f)$ and $\{E(f \cdot \Delta y_n); n=1,2,\dots,N\}$.

The aforementioned linear statistical regression can be extended to nonlinear cases using orthogonal polynomials [14]. Define a set of N -dimensional polynomial functions $\{g_k(\Delta Y); k=1,2,\dots,K\}$ that are normalized and orthogonal:

$$E(g_k) = 0 \quad (k=1,2,\dots,K) \quad (7)$$

$$E(g_i \cdot g_j) = \begin{cases} 1 & (i=j) \\ 0 & (i \neq j) \end{cases} \quad (i,j=1,2,\dots,K). \quad (8)$$

In our case, since all random variables in ΔY are independent and standard Normal, the functions $\{g_k(\Delta Y); k=1,2,\dots,K\}$ can be constructed by the expansion of Hermite series [14] and, therefore, are referred to as the Hermite polynomials. For example, the one-dimensional Hermite polynomials can be expressed as:

$$g_1(\Delta y) = \Delta y \quad g_2(\Delta y) = \frac{1}{\sqrt{2}} \cdot (\Delta y^2 - 1) \quad \dots \quad (9)$$

The key idea of nonlinear statistical regression is to approximate the performance function $f(\Delta Y)$ as the linear combination of all polynomial basis functions:

$$f(\Delta Y) \approx \sum_{i=1}^K \beta_i \cdot g_i(\Delta Y) + \beta_0 \quad (10)$$

where $\{\beta_k; k=0,1,\dots,K\}$ are the model coefficients. Since the polynomials $\{g_k(\Delta Y); k=1,2,\dots,K\}$ are normalized and orthogonal as shown in (7)-(8), we can solve all coefficients $\{\beta_k; k=0,1,\dots,K\}$ by moment matching:

$$E(f) = E\left[\sum_{i=1}^K \beta_i \cdot g_i + \beta_0\right] = \beta_0 \quad (11)$$

$$E(f \cdot g_k) = E\left[\left(\sum_{i=1}^K \beta_i \cdot g_i + \beta_0\right) \cdot g_k\right] = E[\beta_k \cdot g_k^2] = \beta_k \quad (12)$$

$(k=1,2,\dots,K)$

Intuitively, STAR calculates the correlation between f and Δy_n (or g_k). If there is a strong positive (or negative) correlation between f and Δy_n (or g_k), the corresponding model coefficient is large and positive (or negative). Because STAR fundamentally changes the modeling algorithm, many traditional limitations for response surface modeling are not applicable to STAR. For example, it is no longer required that the number of sampling points is equal to or greater than the number of model coefficients. In contrast, STAR can estimate a large number of (e.g., $10^3 \sim 10^5$) model coefficients from a small set of (e.g., $10^2 \sim 10^3$) sampling points without over-fitting. This is one of the most attractive features of STAR, especially when a high-dimensional modeling

problem is considered.

The proposed STAR algorithm is related to the Pearson correlation method that was developed by the statistical learning community for feature selection [9], [13]. However, such a moment matching idea was not successfully applied to response surface modeling in the past, mainly because it is not trivial to accurately estimate the moment values in (5)-(6) and (11)-(12) to determine the model coefficients. In Section 4, we will develop a novel recursive estimator to address this moment estimation problem so that STAR is made of practical interest and utility.

To simplify the notation, we will use linear statistical regression as an example to further discuss the detailed properties and implementations of STAR in the following sections. It should be noted, however, that many results reported in this paper can also be extended to nonlinear statistical regression.

3.2 Proof of Optimality

While STAR determines all model coefficients by moment matching, the following theorem proves that it implicitly utilizes the same optimal criterion as the traditional response surface modeling. Namely, the weighted modeling error is implicitly minimized by moment matching.

Theorem 1: Given the linear performance model $f(\Delta Y)$ in (3), the model coefficients determined by moment matching in (5)-(6) minimize the weighted fitting error:

$$\begin{aligned} & \text{Error}(\alpha_0, \alpha_1, \dots, \alpha_N) \\ &= \int_{-\infty}^{+\infty} \left[f(\Delta Y) - \sum_{i=1}^N \alpha_i \cdot \Delta y_i - \alpha_0 \right]^2 \cdot \text{pdf}(\Delta Y) \cdot d(\Delta Y) \quad (13) \\ &= E\left[\left(f - \sum_{i=1}^N \alpha_i \cdot \Delta y_i - \alpha_0\right)^2\right] \end{aligned}$$

where $\text{pdf}(\Delta Y)$ is the joint probability density function of ΔY .

Proof: The error function in (13) is quadratic and positive-semi-definite. Therefore, the function $\text{Error}(\alpha_0, \alpha_1, \dots, \alpha_N)$ is convex. The optimal coefficients for minimal $\text{Error}(\alpha_0, \alpha_1, \dots, \alpha_N)$ are determined by the conditions [15]:

$$\frac{\partial \text{Error}(\alpha_0, \alpha_1, \dots, \alpha_N)}{\partial \alpha_n} = 0 \quad (n=0,1,\dots,N). \quad (14)$$

Substituting (13) into (14) yields:

$$\begin{aligned} \frac{\partial \text{Error}}{\partial \alpha_0} &= 2 \cdot E\left[\left(\sum_{i=1}^N \alpha_i \cdot \Delta y_i + \alpha_0 - f\right)\right] \\ &= 2 \cdot \alpha_0 - 2 \cdot E(f) = 0 \end{aligned} \quad (15)$$

$$\begin{aligned} \frac{\partial \text{Error}}{\partial \alpha_n} &= 2 \cdot E\left[\left(\sum_{i=1}^N \alpha_i \cdot \Delta y_i + \alpha_0 - f\right) \cdot \Delta y_n\right] \\ &= 2 \cdot \alpha_n - 2 \cdot E(f \cdot \Delta y_n) = 0 \\ & \quad (n=1,2,\dots,N) \end{aligned} \quad (16)$$

Note that Eqn. (15)-(16) are exactly identical to Eqn. (5)-(6) that determine the model coefficients by moment matching. ■

Theorem 1 creates a theoretical link between the proposed statistical regression and the traditional response surface modeling. It demonstrates an important property that if the moment values are accurately estimated, the model coefficients extracted by STAR are as optimal as those solved from over-determined linear equations.

4. IMPLEMENTATION OF STAR

An important operation required by STAR is to estimate the moment values. In this section, we first show the traditional moment estimator used in statistics and study its estimation accuracy. Next, we develop a novel recursive estimator that can significantly improve accuracy and, therefore, make STAR of practical interest and utility.

4.1 Direct Estimator

Unlike the traditional response surface modeling that deterministically locates all sampling points by design of experiments (DOE) [16], STAR randomly draws sampling points to statistically estimate moment values. Given M random sampling points:

$$\begin{aligned} \Delta Y^{(m)} &= [\Delta y_1^{(m)} \quad \Delta y_2^{(m)} \quad \dots \quad \Delta y_N^{(m)}]^T \\ f^{(m)} &= f[\Delta Y^{(m)}] \\ (m &= 1, 2, \dots, M) \end{aligned} \quad (17)$$

$E(f)$ and $\{E(f \cdot \Delta y_n); n = 1, 2, \dots, N\}$ can be estimated by [11]:

$$E(f) \approx \frac{1}{M} \cdot \sum_{m=1}^M f^{(m)} = \tilde{a}_0. \quad (18)$$

$$E(f \cdot \Delta y_n) = \frac{1}{M} \cdot \sum_{m=1}^M f^{(m)} \cdot \Delta y_n^{(m)} = \tilde{a}_n \quad (n = 1, 2, \dots, N) \quad (19)$$

where $\{\tilde{a}_n; n = 0, 1, \dots, N\}$ denote the estimated values for the model coefficients $\{\alpha_n; n = 0, 1, \dots, N\}$. Eqn. (18)-(19) have been widely used in statistics [11] and they are referred to as the *direct estimators* in this paper.

If the performance function $f(\Delta Y)$ is approximated by the linear model in (3), it can be shown that the standard deviations of the estimators $\{\alpha_n; n = 0, 1, \dots, N\}$ are equal to [11]:

$$\sigma(\tilde{a}_0) = \sqrt{\frac{1}{M} \cdot \sum_{i=1}^N \alpha_i^2} \quad (20)$$

$$\sigma(\tilde{a}_n) = \sqrt{\frac{1}{M} \cdot \left(\sum_{i=0}^N \alpha_i^2 + \alpha_n^2 \right)} \quad (n = 1, 2, \dots, N). \quad (21)$$

Eqn. (20)-(21) define the theoretical measure of the estimation accuracy. For example, if $\{\sigma(\tilde{a}_n); n = 0, 1, \dots, N\}$ are large, the estimators $\{\tilde{a}_n; n = 0, 1, \dots, N\}$ have substantial variations, meaning that the estimation error is large.

As shown in (20)-(21), the estimation accuracy depends on the total number of sampling points M . If M increases, $\{\sigma(\tilde{a}_n); n = 0, 1, \dots, N\}$ decrease and the estimation accuracy is improved. Intuitively, more sampling points provide more information about the statistical distribution and, hence, offer more accurate moment estimation.

In addition to the parameter M , the model coefficients $\{\alpha_n; n = 0, 1, \dots, N\}$ also have a significant impact on the estimation accuracy. When the estimator $\{\tilde{a}_n; n = 1, 2, \dots, N\}$ is calculated, all coefficients $\{\alpha_n; n = 0, 1, \dots, N\}$ add up to $\{\sigma(\tilde{a}_n); n = 1, 2, \dots, N\}$ and contribute as estimation noise. In practice, the error incurred by the direct estimators is typically too large to accurately predict the model coefficients. As will be demonstrated by the numerical examples in Section 5, 40%~50% approximation error is observed if the direct estimators are used. It, in turn, motivates us to propose a novel recursive estimator that can predict the moment values more accurately.

4.2 Recursive Estimator

The direct estimators in (18)-(19) predict all moments separately and no connection is exploited between the estimators. Namely, $\{\tilde{a}_n; n = 0, 1, \dots, N\}$ in (18)-(19) are directly represented as functions of the sampling points $\{\Delta Y^{(m)}, f^{(m)}; m = 1, 2, \dots, M\}$. For our proposed statistical regression, however, the moments $E(f)$ and $\{E(f \cdot \Delta y_n); n = 1, 2, \dots, N\}$ are related to the same performance function $f(\Delta Y)$. Therefore, instead of considering the estimators $\{\tilde{a}_n; n = 0, 1, \dots, N\}$ as independent operations, it is possible to identify the interaction between these estimators and utilize it to improve accuracy.

Our proposed recursive estimator is motivated by studying the standard deviation equations in (20)-(21). For the direct estimators, all coefficients $\{\alpha_n; n = 0, 1, \dots, N\}$ add up to the estimation noise. This is one of the key reasons why the direct estimators yield large error. If we estimate the coefficient α_n and subtract the component $\alpha_n \cdot \Delta y_n$ from the performance function $f(\Delta Y)$ before calculating \tilde{a}_{n+1} , then $\alpha_n \cdot \Delta y_n$ will not impact the estimator \tilde{a}_{n+1} any more. As a result, α_n will no longer contribute to $\sigma(\tilde{a}_{n+1})$, thereby decreasing the standard deviation $\sigma(\tilde{a}_{n+1})$ and improving the estimation accuracy. This observation motivates us to iteratively calculate the coefficient α_n and then use the result \tilde{a}_n to improve the estimation accuracy of α_{n+1} in the next iteration step. In addition, we should first extract the most dominant coefficient α_d , i.e., $|\alpha_d| \geq |\alpha_i|$ ($i \neq d$), such that the error reduction is maximized by removing $\alpha_d \cdot \Delta y_d$.

Algorithm 1: Recursive moment estimation

1. Start from a set of sampling points $\{\Delta Y^{(m)}, f^{(m)}; m = 1, 2, \dots, M\}$. Let the set $S = \{0, 1, \dots, N\}$ and the iteration index $p = 1$.
2. Apply the direct estimators in (18)-(19) to calculate the coefficients \tilde{a}_i ($i \in S$).
3. Find the most dominant coefficient \tilde{a}_d , where $|\tilde{a}_d| \geq |\tilde{a}_i|$ ($i \in S$).
4. If $|\alpha_d| \leq \varepsilon$ where ε is a pre-defined threshold, go to Step 9.
5. Update the sampling points:

$$f^{(m)} = \begin{cases} f^{(m)} - \tilde{a}_d & (\text{if } d = 0) \\ f^{(m)} - \tilde{a}_d \cdot \Delta y_d^{(m)} & (\text{if } d \geq 1) \end{cases} \quad (22)$$

$$(m = 1, 2, \dots, M)$$

6. Scale the coefficient \tilde{a}_d to make the estimator unbiased:

$$\tilde{a}_d = \left(\frac{M}{M-1} \right)^{p-1} \cdot \tilde{a}_d. \quad (23)$$

where p is the iteration index and M is the total number of sampling points.

7. Remove the index d from the set S , i.e., $S = S - \{d\}$.
8. $p = p + 1$ and return Step 2.
9. Set $\tilde{a}_i = 0$ ($i \in S$) and the approximated performance model is:

$$f(\Delta Y) \approx \sum_{i=1}^N \tilde{\alpha}_i \cdot \Delta y_i + \tilde{\alpha}_0. \quad (24)$$

Algorithm 1 shows the simplified flow of the proposed recursive moment estimation where the model coefficients are repeatedly extracted. In each iteration, the dominant coefficient \tilde{a}_d is identified and it is used to update the performance function $f(\Delta Y)$ so that the estimation accuracy for the next step can be improved. The recursive iteration stops, if all dominant coefficients are extracted and the non-dominant coefficients are smaller than the pre-defined threshold ε . From this point of view, the proposed recursive moment estimation implicitly provides a screening capability. Namely, it identifies a subset of the critical

random variables that significantly impact the performance of interest.

The scaling factor in (23) makes the estimator *unbiased*, i.e., $E(\tilde{\alpha}_d) = \alpha_d$. This property can be formally stated by the following theorem. The detailed proof of Theorem 2 is not included in this paper due to the limited number of available pages.

Theorem 2: Given the linear performance model $f(\Delta Y)$ in (3), the recursive estimators $\{\tilde{\alpha}_n; n = 0, 1, \dots, N\}$ defined in Algorithm 1 are unbiased, i.e.:

$$E(\tilde{\alpha}_n) = \alpha_n \quad (n = 0, 1, \dots, N). \quad (25)$$

The recursive estimators defined in Algorithm 1 are mutually coupled, since the estimator $\tilde{\alpha}_i$ extracted at the p -th step is used to update the performance function $f(\Delta Y)$ and, therefore, it directly impacts the estimator $\tilde{\alpha}_j$ extracted at the $(p+1)$ -th step. This coupling nature makes the proposed recursive estimators completely different from the direct estimators in (18)-(19).

It is worth mentioning that the performance function $f(\Delta Y)$ is updated in (22) using the estimated coefficient $\tilde{\alpha}_d$. Therefore, the component α_d ($d = 0$) or $\alpha_d \Delta y_d$ ($d \geq 1$) may not be exactly removed from $f(\Delta Y)$, depending on the accuracy of the estimator $\tilde{\alpha}_d$. However, such a performance update still significantly improves the estimation accuracy for most practical applications, as will be demonstrated by the numerical examples in Section 5.

5. NUMERICAL EXAMPLES

In this section we demonstrate the efficacy of STAR using several examples. All numerical experiments are performed on a 2.8GHz Linux server.

5.1 Simple Mathematical Example

We first consider a simple example of the linear model:

$$f(\Delta Y) = 4 \cdot \Delta y_1 + 16 \cdot \Delta y_2 + 64 \cdot \Delta y_3 + 256. \quad (26)$$

To make a full comparison between the direct estimators and the proposed recursive estimators, we repeatedly estimate the model coefficients by using different random sampling sets with different sizes. In addition, given a fixed number of sampling points, we independently draw the random samples and estimate the coefficients for 100 times so that the standard deviations of the estimators can be checked. If the standard deviations are large, it implies that the estimators are not sufficiently accurate.

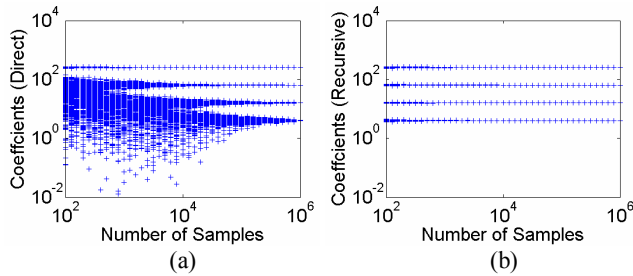


Fig 2. Estimated model coefficients by STAR. (a) Direct estimators. (b) Recursive estimators.

Fig 2 shows the simulation results for both estimators. Note that, in order to provide an accurate estimation for the smallest coefficient, the direct estimators require about 10^6 sampling points. The proposed recursive estimators, however, accurately capture all coefficients, even if the number of sampling points is as small as 100. In this example, given the same error tolerance, the proposed recursive estimators reduce the required number of

sampling points by 10^4 x over the traditional direct estimators.

5.2 Low Noise Amplifier

Fig 3 shows the simplified circuit schematic of a low noise amplifier (LNA) designed in a commercial $0.25\mu\text{m}$ CMOS process. In this example, we attempt to model the variations of the S parameters. The inter-die and intra-die variations of both active/passive devices and layout parasitics are considered. After PCA based on foundry data, 6,418 independent random variables are extracted to model these variations.

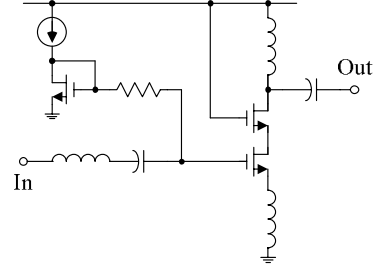


Fig 3. Simplified circuit schematic of a low noise amplifier.

Table 1 compares the linear response surface modeling accuracy for four different techniques. The traditional approach creates a set of sampling points using a simplex-based DOE [16] and then solves all model coefficients from linear equations. STAR (Direct) and STAR (Recursive) apply the proposed statistical regression using the direct estimators and the recursive estimators, respectively. STAR (Screening) first fits a linear model using STAR (Recursive) and identifies a small subset of the critical variables. Next, the model coefficients of these critical variables are solved from a set of over-determined linear equations. The modeling errors in Table 1 are measured from 1000 random samples that are independently collected for testing and comparison purpose.

Table 1. Linear response surface modeling error for LNA

	Traditional	STAR (Direct)	STAR (Recursive)	STAR (Screening)
S11	0.87%	42.37%	1.84%	1.06%
S12	0.64%	25.55%	0.62%	0.45%
S21	0.60%	26.36%	1.17%	0.62%
S22	1.04%	49.69%	2.11%	1.38%

Table 2. Linear response surface modeling cost for LNA

	Traditional	STAR (Direct)	STAR (Recursive)	STAR (Screening)
# of Samples	6419	500	500	500
Spectre (Sec.)	102639	7995	7995	7995
Fitting (Sec.)	6.36	0.04	4.66	4.78
Total (Sec.)	102646	7995	8000	8000

As shown in Table 1, compared with STAR (Direct), STAR (Recursive) reduces the modeling errors by more than 20x. STAR (Screening) further improves the accuracy and makes the modeling error as small as that of the traditional approach. In this example, the modeling cost is dominated by the Spectre simulation time, as shown in Table 2. While the traditional response surface modeling takes more than one day to finish, STAR reduces the computational time to 2.2 hours (12x faster) by using a much smaller set of sampling points.

5.3 Simplified SRAM Read Path

Shown in Fig 4 is the simplified circuit schematic of an SRAM read path designed in a commercial 65nm CMOS process. The simplified read path contains memory array, replica path for self-timing and sense amplifier. In this example, the performance of interest is the delay from the word line (WL) to the sense amplifier output (Out). Both inter-die and intra-die variations are considered. After PCA based on foundry data, 20,276 independent random variables are extracted to model these variations.

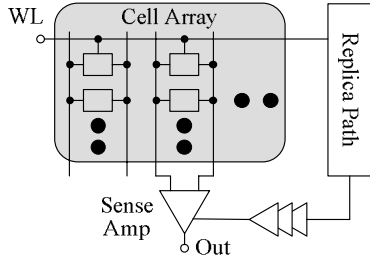


Fig 4. Simplified circuit schematic of an SRAM read path.

The response surface model of the delay variation is extracted by three different techniques: the traditional approach using a simplex-based DOE [16], STAR (Direct) using the direct estimators and STAR (Recursive) using the recursive estimators. Fig 5 shows the magnitude of the estimated linear model coefficients. Note that the recursive estimators result in much smaller estimation noise than the direct estimators. In this example, STAR (Recursive) achieves more than 23x error reduction over STAR (Direct), as shown in Table 3. Compared with the traditional approach, STAR reduces the computational time from 6.4 days to 7.6 hours, yielding more than 20x runtime speedup.

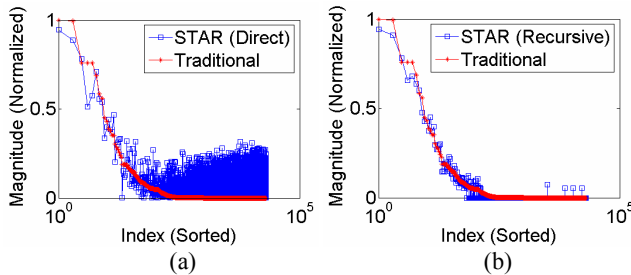


Fig 5. Magnitude of the estimated linear model coefficients. (a) Direct estimators. (b) Recursive estimators.

Table 3. Linear response surface modeling error and cost for SRAM read path

	Traditional	STAR (Direct)	STAR (Recursive)
Modeling Error	0.84%	44.38%	1.89%
# of Samples	20277	1000	1000
Spectre (Sec.)	552751	27260	27260
Fitting (Sec.)	23.57	0.08	7.65
Total (Sec.)	552774	27260	27268

6. CONCLUSIONS

In this paper we propose a novel statistical regression (STAR) approach to efficiently extract high-dimensional linear/nonlinear response surface models. STAR utilizes a new recursive estimator to accurately determine model coefficients by moment matching.

The proposed recursive estimator is facilitated by exploiting the interaction between different moment estimators. It offers more than 20x error reduction over the direct estimator and, therefore, makes STAR of practical interest and utility. Our numerical examples demonstrate that compared with the traditional response surface modeling, STAR achieves more than 20x runtime speedup without substantially increasing modeling error. STAR can be further incorporated into a statistical analysis/optimization environment for accurate and efficient parametric yield analysis/optimization.

7. ACKNOWLEDGEMENTS

This work has been supported by the Semiconductor Research Corporation and the National Science Foundation.

8. REFERENCES

- [1] Semiconductor Industry Associate, *International Technology Roadmap for Semiconductors*, 2005.
- [2] Dharchoudhury and S. Kang, "Worse-case analysis and optimization of VLSI circuit performance," *IEEE Trans. CAD*, vol. 14, no. 4, pp. 481-492, Apr. 1995.
- [3] F. Schenkel, M. Pronath, S. Zizala, R. Schwencker, H. Graeb and K. Antreich, "Mismatch analysis and direct yield optimization by spec-wise linearization and feasibility-guided search," *IEEE DAC*, pp. 858-863, 2001.
- [4] X. Li, J. Le, P. Gopalakrishnan and L. Pileggi, "Asymptotic probability extraction for nonnormal performance distributions," *IEEE TCAD*, vol. 26, no. 1, pp. 16-37, Jan. 2007.
- [5] X. Li, J. Le, L. Pileggi and A. Strojwas, "Projection-based performance modeling for inter/intra-die variations," *IEEE ICCAD*, pp. 721-727, 2005.
- [6] Z. Feng and P. Li, "Performance-oriented statistical parameter reduction of parameterized systems via reduced rank regression," *IEEE ICCAD*, pp. 868-875, 2006.
- [7] A. Singhee and R. Rutenbar, "Beyond low-order statistical response surfaces: latent variable regression for efficient, highly nonlinear fitting," *IEEE DAC*, pp. 256-261, 2007.
- [8] A. Mitev, M. Marefat, D. Ma and J. Wang, "Principle Hessian direction based parameter reduction for interconnect networks with process variation," *IEEE ICCAD*, pp. 632-637, 2007.
- [9] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.
- [10] G. Seber, *Multivariate Observations*, Wiley Series, 1984.
- [11] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 2001.
- [12] R. Myers and D. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley-Interscience, 2002.
- [13] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer, 2003.
- [14] G. Sansone, *Orthogonal Functions*, Dover Publications, 2004.
- [15] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [16] D. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, 2005.