Toward Efficient Large-Scale Performance Modeling of Integrated Circuits via Multi-Mode/Multi-Corner Sparse Regression

Wangyang Zhang
Mentor Graphics Corporation
1001 Ridder Park Drive
San Jose, CA 95131
wangyan1@ece.cmu.edu

Tsung-Hao Chen and Ming-Yuan Ting
Mentor Graphics Corporation
1001 Ridder Park Drive
San Jose, CA 95131
{howard_chen, ming_ting}@mentor.com

Xin Li
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
xinli@ece.cmu.edu

ABSTRACT

In this paper, we propose a novel multi-mode/multi-corner sparse regression (MSR) algorithm to build large-scale performance models of integrated circuits at multiple working modes and environmental corners. Our goal is to efficiently extract multiple performance models to cover different modes/corners with a small number of simulation samples. To this end, an efficient Bayesian inference with shared prior distribution (i.e., model template) is developed to explore the strong performance correlation among different modes/corners in order to achieve high modeling accuracy with low computational cost. Several industrial circuit examples demonstrate that the proposed MSR achieves up to 185× speedup over least-squares regression [14] and 6.7× speedup over least-angle regression [7] without surrendering any accuracy.

Categories and Subject Descriptors

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

General Terms

Algorithms

Keywords

Process Variations, Performance Modeling

1. INTRODUCTION

As IC technology scales to nanoscale region, process variation has become a critical issue that must be carefully addressed within today's IC design flow [1]. Modeling, analyzing and optimizing process variation is now a critical task in order to guarantee high parametric yield for silicon chips. For this reason, various response surface modeling (RSM) techniques have been developed and used as an efficient method to analyze circuit variability [2]-[7]. The objective of RSM is to approximate the circuit performance (e.g., delay, gain, etc.) as an analytical (either linear or nonlinear) function of device parameters (e.g., V_{TH}, T_{OX}, etc.). Once these performance models are created, they can be used for many circuit analysis and optimization applications [8]-[9], e.g., parametric yield prediction, robust circuit design, etc.

While RSM was extensively studied in the past, several recent trends of advanced IC technology pose a number of new challenges in this area.

- **Strong nonlinearity**: As process variation becomes relatively large, nonlinear (e.g., quadratic) models are required to accurately capture performance variability [2]-[7].

DAC'10, June 13-18, 2010, Anaheim, California, USA

- **High dimensionality**: As random device mismatch becomes dominant at sub-65nm technology node, a large number of random variables must be used to model device-level variation, rendering a high-dimensional variation space [2]-[7].
- **Multi-mode/multi-corner operation**: Today's integrated circuits often operate at multiple working modes (e.g., high performance vs. low power) and multiple environmental corners (e.g., high temperature vs. low temperature) [10], [15]. To accurately capture performance variability, different modes/corners must be characterized by different performance models. It, in turn, results in an enormously large modeling problem, as a huge number of performance models must be created to cover all modes/corners.

The aforementioned issues make RSM increasingly difficult. To accurately model and analyze performance variability of analog and mixed-signal integrated circuits, we have to build hundreds of performance models and each model contains thousands of unknown model coefficients. While a number of new RSM techniques have been recently developed [2]-[7], they remain ill-equipped to address the tremendous challenges that we face today. For instance, the least-angle regression (LAR) algorithm in [7] can fit a high-dimensional performance model with about 10^3 simulation samples. If more than 10^2 performance models are required to cover all working modes and environmental corners (as will be demonstrated by several industrial circuit examples in Section 4), over 10^5 sampling points must be generated by SPICE simulation to build all these models. This requires a huge amount of simulation time and, hence, suggests a need to re-think our fundamental strategy for RSM and develop a completely new modeling algorithm to accommodate such a *large* problem size.

In this paper, we propose a novel *multi-mode/multi-corner sparse regression* (MSR) method to address the aforementioned modeling challenges. The key idea is to reduce the number of required simulation samples and, hence, modeling cost by exploring the following two unique properties observed for nanoscale integrated circuits.

- **Sparse model coefficients**: While a large number of basis functions must be used to span the high-dimensional, strongly-nonlinear variation space, not all of them play an important role for a given performance of interest. In other words, only a small subset of model coefficients corresponding to the important basis functions are non-zero, rendering to a unique *sparse* structure [7].
- **Correlated performance variability**: The performance models associated with different modes/corners are not independent. Instead, since these models capture the performance variability of the same circuit [10], [15], they are strongly *correlated*.

However, we do not know the location of the non-zero model

coefficients or the correlation value of the performance variability in advance. The fundamental question is how to develop an efficient algorithm that can accurately "extract" this information from a limited number of simulation samples.

The main contribution of this paper is to borrow a new Bayesian inference recently developed by statistics community [11] to derive an efficient numerical algorithm for MSR. This Bayesian framework aims to build a shared sparse model template for all modes/corners so that their strong correlation can be effectively taken into account to improve modeling accuracy and/or reduce modeling cost. As will be demonstrated by several industrial circuit examples in Section 4, the proposed MSR achieves up to 185× speedup over least-squares regression (LSR) [14] and 6.7× speedup over least-angle regression (LAR) [7] without surrendering any accuracy.

The rest of this paper is organized as follows. In Section 2, the background on principal component analysis, response surface modeling, and sparse regression is reviewed. The proposed multi-mode/multi-corner sparse regression (MSR) will then be described in Section 3. The efficacy of MSR will be demonstrated by several industrial circuit examples in Section 4, followed by our conclusion in Section 5.

2. BACKGROUND

2.1. Principal Component Analysis

Given N process parameters $X = [x_1\ x_2\ ...\ x_N]^T$, the process variation $\Delta X = X - X_0$, where X_0 denotes the mean value of X, is often modeled by multiple zero-mean, correlated Normal distributions [2]-[7]. Principal component analysis (PCA) [12] is a statistical method that finds a set of independent factors to represent the correlated Normal distributions. Assume that the correlation of ΔX is represented by a symmetric, positive semi-definite covariance matrix R. PCA decomposes R as [12]:

$$R = U \cdot \Sigma \cdot U^T \tag{1}$$

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

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

The new random variables in ΔY are called the principal components. It is easy to verify that all principal components in ΔY are independent and standard Normal (i.e., zero mean and unit variance). More details on PCA can be found in [12].

2.2. Response Surface Modeling

Given a circuit design, the circuit performance f (e.g., delay, gain, etc.) is a function of the process variation ΔY defined in (2). RSM approximates the performance function $f(\Delta Y)$ as the linear combination of M basis functions [2]-[7]:

$$f(\Delta Y) \approx \sum_{m=1}^{M} \alpha_m \cdot g_m(\Delta Y) \tag{3}$$

where $\{\alpha_m;\ m = 1,2,...,M\}$ are the model coefficients, and $\{g_m(\Delta Y);\ m = 1,2,...,M\}$ are the basis functions (e.g., linear, quadratic, etc.). The unknown model coefficients in (3) can be determined by solving the following linear equation at K sampling points:

$$G \cdot \alpha = F \tag{4}$$

where

$$\alpha = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_M \end{bmatrix}^T \tag{5}$$

$$F = \begin{bmatrix} f^{(1)} & f^{(2)} & \cdots & f^{(K)} \end{bmatrix}^T \tag{6}$$

$$G = \begin{bmatrix} g_1(\Delta Y^{(1)}) & g_2(\Delta Y^{(1)}) & \cdots & g_M(\Delta Y^{(1)}) \\ g_1(\Delta Y^{(2)}) & g_2(\Delta Y^{(2)}) & \cdots & g_M(\Delta Y^{(2)}) \\ \vdots & \vdots & \vdots & \vdots \\ g_1(\Delta Y^{(K)}) & g_2(\Delta Y^{(K)}) & & g_M(\Delta Y^{(K)}) \end{bmatrix}. \tag{7}$$

In (5)-(7), $\Delta Y^{(k)}$ and $f^{(k)}$ are the values of ΔY and $f(\Delta Y)$ at the k-th sampling point respectively.

The traditional least-squares regression (LSR) [14] attempts to solve the least-squares solution for (4). Hence, the number of samples (K) must be equal to or greater than the number of coefficients (M). Such an LSR becomes intractable, if M is large (e.g., 10^4~10^6). For this reason, a number of sparse regression techniques [6]-[7] were recently proposed to address this computational cost issue.

2.3. Sparse Regression

While a large number of basis functions must be used to span the high-dimensional and strongly nonlinear variation space, only a few of them are required to approximate a specific performance function of interest. Sparse regression is motivated by this observation. It assumes that the unknown vector α in (4) contains a large number of zeros and, hence, is sparse. To find such a sparse solution, the following L_1-norm regularization problem:

$$\begin{aligned} \underset{\alpha}{\text{minimize}} \quad & \left\| G \cdot \alpha - F \right\|_2^2 \\ \text{subject to} \quad & \left\| \alpha \right\|_1 \le \lambda \end{aligned} \tag{8}$$

is solved by least-angle regression (LAR) [7]. In (8), $\|\bullet\|_1$ and $\|\bullet\|_2$ denote the L_1-norm and L_2-norm of a vector, respectively. The value of λ should be optimally determined by the cross-validation scheme described in [7].

It has been demonstrated that LAR can fit a high-dimensional performance model with about 10^3 simulation samples. It, however, remains ill-equipped to address the multi-mode/multi-corner modeling problem that we study in this paper. For instance, if more than 10^2 performance models are required to cover all modes and corners, over 10^5 sampling points must be generated by SPICE simulation to build all these models, resulting in unaffordable computational cost. We will develop a new sparse regression algorithm in this paper that is particularly tuned for multi-mode/multi-corner applications.

3. MULTI-MODE/MULTI-CORNER PERFORMANCE MODELING

Our proposed multi-mode/multi-corner sparse regression (MSR) is facilitated by a novel Bayesian inference that is derived from advanced statistical theories [11], [13]. In this section, we describe its mathematical formulation and highlight the novelties.

3.1. Problem Definition

The objective of MSR is to generate different performance models for different working modes and environmental corners. Today's integrated circuits often operate at multiple modes/corners, e.g., high performance vs. low power, high V_{DD} vs. low V_{DD}, high temperature vs. low temperature, etc. It is difficult to accurately capture the performance variability at all these modes/corners by a unified model. Instead, multiple models must be generated to cover all modes/corners:

$$f_{(l)}(\Delta Y) \approx \sum_{m=1}^{M} \alpha_{(l),m} \cdot g_m(\Delta Y) \quad (l = 1,2,\cdots,L) \tag{9}$$

where L represents the total number of modes/corners. Similar to

(4)-(7), simulation samples can be generated for $\{f_{(l)}(\Delta Y);\ l = 1,2,...,L\}$ and a number of linear equations can be created to solve the model coefficients $\{\alpha_{(l),m};\ l = 1,2,...,L;\ m = 1,2,...,M\}$:

$$G_{(l)} \cdot \alpha_{(l)} = F_{(l)} \quad (l = 1,2,\cdots,L) \tag{10}$$

where

$$\alpha_{(l)} = \begin{bmatrix} \alpha_{(l),1} & \alpha_{(l),2} & \cdots & \alpha_{(l),M} \end{bmatrix}^T \tag{11}$$

$$F_{(l)} = \begin{bmatrix} f_{(l)}^{(1)} & f_{(l)}^{(2)} & \cdots & f_{(l)}^{(K)} \end{bmatrix}^T \tag{12}$$

$$G_{(l)} = \begin{bmatrix} g_1\left(\Delta Y_{(l)}^{(1)}\right) & g_2\left(\Delta Y_{(l)}^{(1)}\right) & \cdots & g_M\left(\Delta Y_{(l)}^{(1)}\right) \\ g_1\left(\Delta Y_{(l)}^{(2)}\right) & g_2\left(\Delta Y_{(l)}^{(2)}\right) & \cdots & g_M\left(\Delta Y_{(l)}^{(2)}\right) \\ \vdots & \vdots & \vdots & \vdots \\ g_1\left(\Delta Y_{(l)}^{(K)}\right) & g_2\left(\Delta Y_{(l)}^{(K)}\right) & & g_M\left(\Delta Y_{(l)}^{(K)}\right) \end{bmatrix}. \tag{13}$$

In (11)-(13), $\Delta Y_{(l)}^{(k)}$ and $f_{(l)}^{(k)}$ are the values of ΔY and $f_{(l)}(\Delta Y)$ at the k-th sampling point of the l-th performance function.

One straightforward way to solve $\{\alpha_{(l),m};\ l = 1,2,...,L;\ m = 1,2,...,M\}$ is to consider each $f_{(l)}(\Delta Y)$ independently and solve all linear equations in (10) one by one. This simple approach, however, is not optimal, since it completely ignores the correlation among the performance functions $\{f_{(l)}(\Delta Y);\ l = 1,2,...,L\}$. Such correlation exists, because these performance functions capture the performance variability of the same circuit. It, in turn, motivates us to develop a new multi-mode/multi-corner sparse regression (MSR) algorithm that explores the extra correlation information to improve modeling accuracy and/or reduce modeling cost. This goal is achieved by adapting a novel Bayesian inference recently developed by statistics community [11] to build a shared sparse model template for all modes/corners. In what follows, we first present the Bayesian framework for sparse regression, and then show how to adapt and apply such a Bayesian inference to multi-mode/multi-corner performance modeling problems.

3.2. Bayesian Inference

Given a set of simulation samples $\{G_{(l)},F_{(l)};\ l = 1,2,...,L\}$, most traditional regression methods [2]-[7] solve the linear equations in (10) and find the deterministic values of all model coefficients $\{\alpha_{(l)};\ l = 1,2,...,L\}$. Bayesian inference, however, takes a completely different strategy. It considers all unknown model coefficients as random variables and applies Bayes' theorem to find the probability density function $\{pdf(\alpha_{(l)} \mid G_{(l)},F_{(l)});\ l = 1,2,...,L\}$. The distribution $pdf(\alpha_{(l)} \mid G_{(l)}, F_{(l)})$ contains important information about $\alpha_{(l)}$ that is "learned" from the simulation samples $\{G_{(l)},F_{(l)};\ l = 1,2,...,L\}$. It tells us: (1) the value of $\alpha_{(l)}$ that is most likely to occur (i.e., determined by the mode of the PDF), and (2) the uncertainty of $\alpha_{(l)}$ that corresponds to performance modeling error (i.e., determined by the covariance matrix of the PDF).

A. Prior Definition

To formulate a Bayesian inference, we first need to define a so-called *prior distribution* for $\{\alpha_{(l)};\ l = 1,2,...,L\}$. Intuitively, the prior distribution represents our prior knowledge about $\{\alpha_{(l)};\ l = 1,2,...,L\}$ without seeing any simulation samples. In general, if we do not have any prior information, a uniform distribution over $(-\infty, +\infty)$ can be used to model the prior distribution $\{pdf(\alpha_{(l)});\ l = 1,2,...,L\}$, implying that each coefficient $\alpha_{(l),m}$ can possibly take any value with equal probability. However, in our application of sparse regression, we know that the coefficient vector $\alpha_{(l)}$ in (10) is sparse. Hence, we should define an appropriate prior distribution to carry this unique information of sparsity.

To this end, we borrow the idea of hierarchical Bayesian inference described in [11]. We first model the regression error as a zero-mean Normal distribution. Eq. (9) is re-written as:

$$f_{(l)}(\Delta Y) = \sum_{m=1}^{M} \alpha_{(l),m} \cdot g_m(\Delta Y) + \varepsilon_{(l)} \quad (l = 1,2,\cdots,L) \tag{14}$$

where $\varepsilon_{(l)}$ represents the regression error and its PDF is:

$$pdf\left(\varepsilon_{(l)} \mid \theta_{(l),0}\right) = \sqrt{\frac{\theta_{(l),0}}{2\pi}} \cdot \exp\left(-\frac{\theta_{(l),0} \cdot \varepsilon_{(l)}^2}{2}\right) \quad (l = 1,2,\cdots,L). \tag{15}$$

In (15), $\theta_{(l),0}$ is the *precision* of the Normal distribution. It is equal to the inverse of the variance. In our performance modeling application, the precision $\theta_{(l),0}$ represents the accuracy of the l-th performance model. The value of $\theta_{(l),0}$ can be estimated by a maximum likelihood method, as will be discussed in detail in Section 3.2.C.

Next, we define the prior distribution of each $\alpha_{(l),m}$ as a parameterized, zero-mean Normal distribution:

$$pdf\left(\alpha_{(l),m} \mid \theta_{(l),m}\right) = \sqrt{\frac{\theta_{(l),m}}{2\pi}} \cdot \exp\left(-\frac{\theta_{(l),m} \cdot \alpha_{(l),m}^2}{2}\right) \tag{16}$$

$$(l = 1,2,\cdots,L; \quad m = 1,2,\cdots,M)$$

where $\theta_{(l),m}$ is the precision of the Normal distribution. The key idea of the hierarchical Bayesian inference in [11] is to assign appropriate values to the parameters $\{\theta_{(l),m};\ m = 1,2,...,M\}$ so that the coefficient vector $\alpha_{(l)}$ has a sparse solution. Intuitively, if the precision $\theta_{(l),m}$ is large (i.e., variance is small), the corresponding model coefficient $\alpha_{(l),m}$ is likely to be zero. Otherwise, if the precision $\theta_{(l),m}$ is small (i.e., variance is large), the corresponding $\alpha_{(l),m}$ can be far away from zero. However, we only know that the coefficient vector $\alpha_{(l)}$ is sparse, but we do not know the exact location of zeros. For this reason, it is not trivial to determine the "optimal" values of $\{\theta_{(l),m};\ m = 1,2,...,M\}$. In Section 3.2.C, we will present a maximum likelihood method to address this issue.

To complete the definition of the prior distribution, we assume that the random variables $\varepsilon_{(l)}$ and $\{\alpha_{(l),m};\ m = 1,2,...,M\}$ are mutually independent. Hence, the joint PDF of $\{\varepsilon_{(l)},\alpha_{(l)}\}$ is:

$$pdf\left(\varepsilon_{(l)},\alpha_{(l)} \mid \theta_{(l),0,\cdots,M}\right) = pdf\left(\varepsilon_{(l)} \mid \theta_{(l),0}\right) \cdot \prod_{m=1}^{M} pdf\left(\alpha_{(l),m} \mid \theta_{(l),m}\right). \tag{17}$$

$$(l = 1,2,\cdots,L)$$

The independence assumption we made simply implies that we do not know the correlation between $\varepsilon_{(l)}$ and $\{\alpha_{(l),m};\ m = 1,2,...,M\}$ in advance. Such correlation information will be considered, once the simulation samples $\{G_{(l)},F_{(l)};\ l = 1,2,...,L\}$ are available to calculate the posterior distribution.

B. Performance Correlation

The prior distributions defined in (15)-(17) promote a sparse solution of $\{\alpha_{(l)};\ l = 1,2,...,L\}$, if the values of $\{\theta_{(l),m};\ l = 1,2,...,L,\ m = 1,2,...,M\}$ are properly selected. For our proposed multi-mode/multi-corner sparse regression (MSR), we expect that all performance models corresponding to different working modes and environmental corners are similar and, hence, strongly correlated. Such correlation information should be explicitly explored to improve modeling accuracy and/or reduce modeling cost. In this sub-section, we further present a systematic methodology to take into account the strong correlation among all performance models.

A close look at the prior distributions (15)-(17) motivates several additional assumptions that we can make to explore the "similarity" among $\{f_{(l)}(\Delta Y);\ l = 1,2,...,L\}$. First, we assume that

the values of $\{\theta_{(l),0};\ l = 1,2,...,L\}$ in (15) are identical:

$$\theta_{(1),0} = \theta_{(2),0} = \cdots = \theta_{(L),0} = \theta_0 . \quad (18)$$

Eq. (18) implies that the error of all performance models follows the same Normal distribution with identical variance. (Note that even with the same prior distribution $\{pdf(\varepsilon_{(l)} \mid \theta_0);\ l = 1,2,...,L\}$, the regression error $\varepsilon_{(l)}$ can still vary for different performance models.) This assumption is typically valid, especially if a proper scaling is applied to all performance functions (e.g., dividing each performance function by its nominal value).

Next, we further assume that the values of $\{\theta_{(l),m};\ l = 1,2,...,L\}$ in (16) are identical:

$$\theta_{(1),m} = \theta_{(2),m} = \cdots = \theta_{(L),m} = \theta_m \quad (m = 1,2,\cdots,M). \quad (19)$$

Remember that the values of $\{\theta_{(l),m};\ l = 1,2,...,L,\ m = 1,2,...,M\}$ indicate the importance of the corresponding basis function $g_m(\Delta Y)$ for the l-th performance function $f_{(l)}(\Delta Y)$. Hence, Eq. (19) simply implies that a particular basis function $g_m(\Delta Y)$ may either simultaneously impact all performance functions or affect no performance function at all. In other words, all performance functions share the same model template, because they represent the behavior of the same circuit at different modes/corners. It is important to note that $\{\theta_{(l),m};\ l = 1,2,...,L,\ m = 1,2,...,M\}$ are not the unknown model coefficients $\{\alpha_{(l),m};\ l = 1,2,...,L,\ m = 1,2,...,M\}$ and, hence, Eq. (19) does not necessarily lead to identical model coefficients. The solution of $\{\alpha_{(l),m};\ l = 1,2,...,L,\ m = 1,2,...,M\}$ will be determined by calculating the posterior distribution of the proposed Bayesian inference, as will be discussed in detail in Section 3.2.C.

The equality constraints defined by (18)-(19) offer a systematic approach to explore the correlation among all performance functions. In other words, the same prior distribution $\{pdf(\varepsilon_{(l)},\alpha_{(l)} \mid \theta_{0,...,M});\ l = 1,2,...,L\}$ is now shared by all modes/corners, since they are defined by using the same parameters $\{\theta_m;\ m = 0,1,...,M\}$. This unique modeling structure for MSR enables us to achieve superior modeling accuracy and/or cheap modeling cost over the traditional sparse regression algorithms (e.g., least-angle regression [7]), as will be demonstrated by several industrial circuit examples in Section 4.

In addition, Eq. (18)-(19) also reduce the complexity of the proposed prior distribution. Namely, they reduce the number of independent parameters that define the prior distribution. In what follows, we will describe an efficient algorithm to determine the values of $\{\theta_m;\ m = 0,1,...,M\}$ from a limited number of simulation samples. Once $\{\theta_m;\ m = 0,1,...,M\}$ are found, the prior distribution in (15)-(19) is uniquely determined.

C. Posterior Calculation

A critical component of the aforementioned Bayesian inference is to determine the values of $\{\theta_m;\ m = 0,1,...,M\}$ so that the prior distribution of $\{\alpha_{(l)};\ l = 1,2,...,L\}$ is accurate. In this sub-section, we will first present an efficient maximum likelihood estimation (MLE) method to solve $\{\theta_m;\ m = 0,1,...,M\}$. Next, we will further show a maximum posterior method (MAP) to determine the values of the unknown model coefficients $\{\alpha_{(l)};\ l = 1,2,...,L\}$.

Given a set of simulation samples $\{G_{(l)},\ F_{(l)};\ l = 1,2,...,L\}$ collected for all performance functions, the key idea of MLE is to determine $\{\theta_m;\ m = 0,1,...,M\}$ for the shared prior distribution so that the sampling points $\{G_{(l)},\ F_{(l)};\ l = 1,2,...,L\}$ are most likely to occur. Namely, we aim to maximize the following conditional probability:

$$\underset{\theta_0,\cdots,\theta_M}{\text{maximize}} \quad pdf\left(F_{(1,\cdots,L)} \mid \theta_{0,\cdots,M}\right) \quad (20)$$

If all simulation samples $\{G_{(l)},\ F_{(l)};\ l = 1,2,...,L\}$ are created independently, $pdf(F_{(1,...,L)} \mid \theta_{0,...,M})$ can be represented as:

$$pdf\left(F_{(1,\cdots,L)} \mid \theta_{0,\cdots,M}\right) = \prod_{l=1}^{L} pdf\left(F_{(l)} \mid \theta_{0,\cdots,M}\right). \quad (21)$$

Taking the logarithm of (21) results in the *log-likelihood*:

$$\log\left[pdf\left(F_{(1,\cdots,L)} \mid \theta_{0,\cdots,M}\right)\right] = \sum_{l=1}^{L} \log\left[pdf\left(F_{(l)} \mid \theta_{0,\cdots,M}\right)\right]. \quad (22)$$

Since $\log(\bullet)$ monotonically increases, maximizing the PDF in (20) is equivalent to maximizing the log-likelihood in (22), yielding:

$$\underset{\theta_0,\cdots,\theta_M}{\text{maximize}} \quad \sum_{l=1}^{L} \log\left[pdf\left(F_{(l)} \mid \theta_{0,\cdots,M}\right)\right]. \quad (23)$$

To solve (23), we need to find an efficient way to calculate the log-likelihood function. Towards this goal, we apply the Bayes' theorem [13]:

$$pdf\left(\alpha_{(l)} \mid F_{(l)},\theta_{0,\cdots,M}\right) = \frac{pdf\left(\alpha_{(l)} \mid \theta_{0,\cdots,M}\right)\cdot pdf\left(F_{(l)} \mid \alpha_{(l)},\theta_{0,\cdots,M}\right)}{pdf\left(F_{(l)} \mid \theta_{0,\cdots,M}\right)}. \quad (24)$$

By simply rearranging (24), we have:

$$pdf\left(F_{(l)} \mid \theta_{0,\cdots,M}\right) = \frac{pdf\left(\alpha_{(l)} \mid \theta_{0,\cdots,M}\right)\cdot pdf\left(F_{(l)} \mid \alpha_{(l)},\theta_{0,\cdots,M}\right)}{pdf\left(\alpha_{(l)} \mid F_{(l)},\theta_{0,\cdots,M}\right)}. \quad (25)$$

Eq. (25) implies that we can easily calculate $pdf(F_{(l)} \mid \theta_{0,...,M})$ if we know $pdf(\alpha_{(l)} \mid \theta_{0,...,M})$, $pdf(F_{(l)} \mid \alpha_{(l)},\theta_{0,...,M})$ and $pdf(\alpha_{(l)} \mid F_{(l)},\theta_{0,...,M})$.

First, we note that $pdf(\alpha_{(l)} \mid \theta_{0,...,M})$ is the prior distribution for $\alpha_{(l)}$, as defined in (16)-(17) and (19):

$$pdf\left(\alpha_{(l)} \mid \theta_{0,\cdots,M}\right) = \prod_{m=1}^{M} \sqrt{\frac{\theta_m}{2\pi}} \cdot \exp\left(-\frac{\theta_m \cdot \alpha_{(l),m}^2}{2}\right). \quad (26)$$

Second, $pdf(F_{(l)} \mid \alpha_{(l)},\theta_{0,...,M})$ is referred to as the *likelihood function*. Based on the performance model in (14)-(15), we know that once $\alpha_{(l)}$ and $\{\theta_m;\ m = 0,1,...,M\}$ are fixed, $pdf(F_{(l)} \mid \alpha_{(l)},\theta_{0,...,M})$ is a Normal distribution determined by the modeling error [11], [13]:

$$pdf\left(F_{(l)} \mid \alpha_{(l)},\theta_{0,\cdots,M}\right) = \left(\sqrt{\frac{\theta_0}{2\pi}}\right)^K \cdot \exp\left(-\frac{\theta_0 \cdot \left\|G_{(l)}\alpha_{(l)} - F_{(l)}\right\|_2^2}{2}\right) \quad (27)$$

where K is the number of simulation samples for the l-th performance function. Finally, $pdf(\alpha_{(l)} \mid F_{(l)},\theta_{0,...,M})$ is the *posterior distribution* for $\alpha_{(l)}$. It models the uncertainty of $\alpha_{(l)}$ after we observe the simulation samples $\{G_{(l)},\ F_{(l)}\}$. Given the prior distribution in (26) and the likelihood function in (27), the posterior distribution $pdf(\alpha_{(l)} \mid F_{(l)},\theta_{0,...,M})$ is Normal and its covariance $\Sigma_{(l)}$ and mean $\mu_{(l)}$ are respectively equal to [11], [13]:

$$\Sigma_{(l)} = \left(\theta_0 \cdot G_{(l)}^T \cdot G_{(l)} + A\right)^{-1} \quad (28)$$

$$\mu_{(l)} = \theta_0 \cdot \Sigma_{(l)} \cdot G_{(l)}^T \cdot F_{(l)} \quad (29)$$

where

$$A = \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_M \end{bmatrix} \quad (30)$$

is a diagonal matrix.

Combining (25)-(29), it can be shown that the optimization in (23) is equivalent to:

$$\underset{\theta_0,\cdots,\theta_M}{\text{minimize}} \quad \sum_{l=1}^{L} \begin{bmatrix} \log\left[\det\left(\theta_0^{-1} I + G_{(l)} A^{-1} G_{(l)}^T\right)\right] + \\ F_{(l)}^T \cdot \left(\theta_0^{-1} I + G_{(l)} A^{-1} G_{(l)}^T\right)^{-1} \cdot F_{(l)} \end{bmatrix} \quad (31)$$

where $\det(\bullet)$ stands for the determinant of a matrix. Eq. (31) presents an analytical form of the cost function that we should

minimize. Given this complicated cost function, the optimization in (31) is not convex. An efficient greedy algorithm has been developed in [11] to solve $\{\theta_m;\ m = 0,1,...,M\}$ from (31). Even though global convergence is not guaranteed, the greedy algorithm in [11] has been demonstrated as a robust numerical solver for many practical problems. Due to limited space, we will not discuss the numerical solver here. More details on the greedy algorithm can be found in [11].

Once the parameters $\{\theta_m;\ m = 0,1,...,M\}$ are determined, the posterior distribution of $\alpha_{(l)}$ is Normal and its mean $\mu_{(l)}$ and covariance $\Sigma_{(l)}$ are specified by (28)-(29). Since a Normal distribution is peaked at the mean value $\mu_{(l)}$, the maximum posterior solution of $\alpha_{(l)}$ (i.e., the value of $\alpha_{(l)}$ that is most likely to occur) is exactly equal to $\mu_{(l)}$:

$$\alpha_{(l)} = \mu_{(l)} = \theta_0 \cdot \Sigma_{(l)} \cdot G_{(l)}^T \cdot F_{(l)}. \quad (32)$$

Eq. (32) solves the unknown model coefficients $\{\alpha_{(l)};\ l = 1,2,...,L\}$ for our proposed multi-mode/multi-corner performance modeling problem.

3.3. Summary

As shown in Figure 1, the key idea of the proposed multi-mode/multi-corner spare regression (MSR) is to apply the same prior distribution (i.e., model template) to all performance functions. Algorithm 1 summarizes the major steps of the MSR method. Starting from a number of random sampling points $\{G_{(l)}, F_{(l)};\ l = 1,2,...,L\}$, Algorithm 1 first solves the parameters $\{\theta_m;\ m = 0,1,...,M\}$ of the shared prior distribution by using the greedy algorithm described in [11]. Next, the maximum posterior solution of the unknown model coefficients $\{\alpha_{(l)};\ l = 1,2,...,L\}$ is determined by Bayesian inference. Due to the shared prior distribution applied by MSR, the number of required simulation samples for each performance model can be significantly reduced, as will be demonstrated by industrial circuit examples in Section 4.

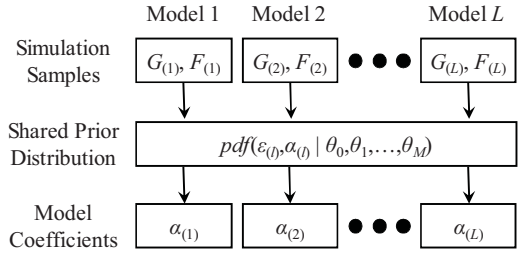

Figure 1. The same prior distribution (i.e., model template) is shared by all performance models for multi-mode/multi-corner sparse regression (MSR).

Algorithm 1: Multi-Mode/Multi-Corner Sparse Regression

1. Randomly generate the sampling points $\{G_{(l)},\ F_{(l)};\ l = 1,2,...,L\}$ for L performance functions where K sampling points are associated with each performance function, as shown in (12) and (13).
2. Define the prior distribution for the model coefficients $\{\alpha_{(l)};\ l = 1,2,...,L\}$ and the regression errors $\{\varepsilon_{(l)};\ l = 1,2,...,L\}$, as shown in (15)-(19).
3. Solve the optimization in (31) to determine the parameters $\{\theta_m;\ m = 0,1,...,M\}$ by using the greedy algorithm described in [11].
4. Find the posterior distribution (which is Normal) for the model coefficients $\{\alpha_{(l)};\ l = 1,2,...,L\}$ using (28)-(29).
5. Determine the maximum posterior solution of the model coefficients $\{\alpha_{(l)};\ l = 1,2,...,L\}$ using (32).

4. NUMERICAL EXAMPLES

In this section, we demonstrate the efficacy of MSR using several industrial circuit examples. All circuits are designed in a commercial CMOS process. For each circuit example, there are multiple working modes and environmental corners. One performance model is extracted for each of these modes/corners. Towards this goal, two independent random sampling sets, called training set and testing set respectively, are generated. The training set is used for coefficient fitting (i.e., Algorithm 1), while the testing set is used for model validation. All numerical experiments are performed on a 3GHz Linux server.

4.1. High-Speed Adder

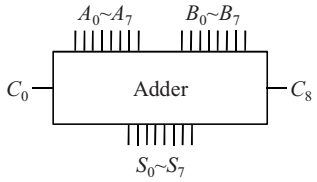

Figure 2. Block diagram of an 8-bit adder

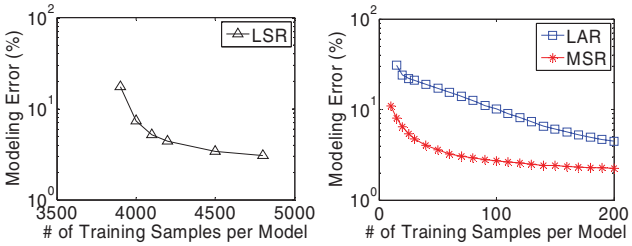

Figure 3. Linear performance modeling error

Table 1. Linear performance modeling cost

	LSR	LAR	MSR
# Samples per Model	4200	190	30
Simulation (Sec.)	4.94×10^6	2.23×10^5	3.52×10^4
Fitting (Sec.)	1.24×10^4	1.04×10^3	3.1×10^1
Total (Sec.)	4.95×10^6	2.24×10^5	3.52×10^4

Figure 2 shows the block diagram of an 8-bit adder. In this example, we aim to model the path delay from the input A_0 to the output S_7. Both global variation and local mismatch are taken into account. After PCA based on foundry data, 3876 independent random variables are extracted to model the variations. In addition, we consider the circuit to operate at 3 different V_{DD} corners with 7 different input slopes at A_0 and 8 different output load capacitors at S_7, resulting in 168 different modes/corners. Such a multi-mode/multi-corner modeling problem occurs in several practical applications, e.g., statistical timing library characterization [15].

Figure 3 shows the linear modeling error for three different techniques: least-squares regression (LSR) [14], least-angle regression (LAR) [7], and the proposed multi-mode/multi-corner sparse regression (MSR). Given the same number of training samples, MSR yields substantially better accuracy than LSR and LAR. This is because MSR explores the strong correlation between different modes/corners by applying the same prior distribution (i.e., model template), as shown in Algorithm 1.

On the other hand, for a given accuracy requirement, MSR requires much less training samples than LSR and LAR. Studying Figure 3, one would notice that to achieve 5% modeling error,

MSR only requires 30 samples per model (168 models in total), while LSR and LAR require 4200 and 190 samples per model, respectively. Table 1 further shows the computational cost for the aforementioned three techniques. Note that MSR is computationally cheaper than LSR and LAR in both simulation cost (i.e., to generate simulation samples) and fitting cost (i.e., to solve model coefficients). In this example, MSR achieves 140× speedup over LSR and 6.3× speedup over LAR.

4.2. Simplified SRAM Read Path

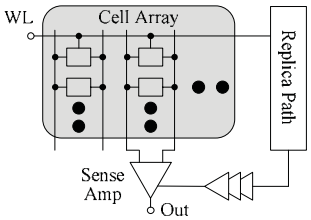

Figure 4. Simplified circuit schematic of an SRAM read path

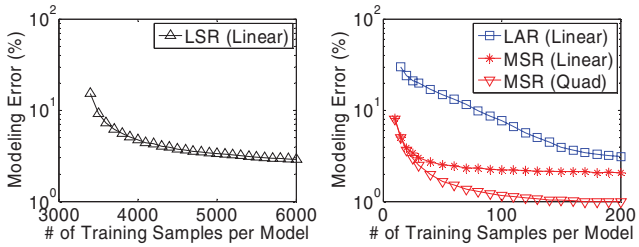

Figure 5. Linear and quadratic performance modeling error

Table 2. Linear performance modeling cost

	LSR	LAR	MSR
# Samples per Model	5600	200	30
Simulation (Sec.)	2.35×10^{6}	8.40×10^{4}	1.26×10^{4}
Fitting (Sec.)	1.52×10^{4}	8.36×10^{2}	2.50×10^{1}
Total (Sec.)	2.36×10^{6}	8.48×10^{4}	1.26×10^{4}

Shown in Figure 4 is the simplified circuit schematic of an SRAM read path. After PCA based on foundry data, 3345 independent random variables are extracted to model both global variation and local mismatch. Similar to the previous example, we consider the circuit to operate at 3 different V_{DD} corners with 7 different word line slopes and 8 different output load capacitors, resulting in 168 different modes/corners.

In this example, we first build linear performance models for read path delay using three different techniques: LSR [14], LAR [7], and MSR. Figure 5 shows the modeling error as a function of the number of training samples per model (168 models in total). To achieve 3% modeling error, MSR only requires 30 samples per model, while LSR and LAR require 5600 and 200 samples per model, respectively. Table 2 further shows the computational cost for these three modeling techniques. In this example, MSR achieves 185× speedup over LSR and 6.7× speedup over LAR, as shown in Table 2.

To further improve accuracy, we select a subset of important process parameters corresponding to non-zero linear model coefficients. Next, we create quadratic performance models for these critical process parameters using MSR. Figure 5 also shows the quadratic modeling error for MSR. Compared to simple linear modeling, the proposed quadratic modeling scheme further reduces modeling error from 2% to 1% (2× reduction), if a sufficient number of simulation samples (more than 100 samples per model) are available.

5. CONCLUSIONS

In this paper, we propose a novel multi-mode/multi-corner spare regression (MSR) method to efficiently generate large-scale performance models of integrated circuits at multiple working modes and environmental corners. An efficient Bayesian inference with shared prior distribution (i.e., model template) is applied to explore the strong performance correlation among different modes/corners to improve modeling accuracy and/or reduce modeling cost. Several industrial circuit examples demonstrate that MSR achieves up to 185× speedup over least-squares regression (LSR) [14] and 6.7× speedup over least-angle regression (LAR) [7] without surrendering any accuracy. The proposed MSR algorithm can be applied to a number of practical applications for both analog and digital circuits such as statistical timing library characterization and parametric yield estimation.

6. REFERENCES

[1] Semiconductor Industry Associate, *International Technology Roadmap for Semiconductors*, 2007.

[2] X. Li, J. Le, L. Pileggi and A. Strojwas, "Projection-based performance modeling for inter/intra-die variations," *IEEE ICCAD*, pp. 721-727, 2005.

[3] Z. Feng and P. Li, "Performance-oriented statistical parameter reduction of parameterized systems via reduced rank regression," *IEEE ICCAD*, pp. 868-875, 2006.

[4] 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.

[5] 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.

[6] X. Li and H. Liu, "Statistical regression for efficient high-dimensional modeling of analog and mixed-signal performance variations," *IEEE DAC*, pp. 38-43, 2008.

[7] X. Li, "Finding deterministic solution from underdetermined equation: large-scale performance modeling by least angle regression," *IEEE DAC*, pp. 364-369, 2009.

[8] 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.

[9] T. McConaghy and G. Gielen, "Template-free symbolic performance modeling of analog circuits via canonical-form functions and genetic programming," *IEEE Trans. CAD*, vol. 28, no. 8, pp. 1162-1175, Aug. 2009.

[10] B. Floyd and D. Ozis, "Low-noise amplifier comparison at 2GHz in 0.25-µm and 0.18-µm RF-CMOS and SiGe BiCMOS," *IEEE RFIC*, pp. 185-188, 2004.

[11] S. Ji, D. Dunson and L. Carin, "Multi-task compressive sensing," *IEEE Trans. Signal Processing*, vol. 57, no. 1, pp. 92-106, Jan. 2009.

[12] G. Seber, *Multivariate Observations*, Wiley Series, 1984.

[13] G. Casella and R. Berger, *Statistical Inference*, Duxbury Press, 2001.

[14] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer, 2003.

[15] S. Sapatnekar, *Timing*, Springer, 2004.