

Efficient Spatial Variation Modeling of Nanoscale Integrated Circuits Via Hidden Markov Tree

Changhai Liao, Jun Tao, *Member, IEEE*, Xuan Zeng, *Member, IEEE*, Yangfeng Su, Dian Zhou, *Senior Member, IEEE*, and Xin Li, *Senior Member, IEEE*

Abstract—In this paper, we propose a novel spatial variation modeling method based on hidden Markov tree (HMT) for nanoscale integrated circuits, which could efficiently improve the accuracy of full-wafer/chip spatial variations recovery at extremely low measurement cost. Applying this method, HMT is introduced to set up a statistical model for coefficients after exploring the underlying correlated representation of the spatial variation in the frequency domain. Accordingly, two key inherent properties of the modeling coefficients, i.e., correlations and sparse presentations in the frequency domain, can be captured exactly and the modeling accuracy can be improved evidently. Then, maximum-*a-posteriori* estimation is applied to formulate the original problem as a convex optimization that could be solved efficiently and robustly. Numerical results based on industrial data demonstrate that the proposed method can achieve superior accuracy over other existing approaches including orthogonal matching pursuit, l_1 -norm regularization, and reweighted l_1 -norm regularization.

Index Terms—Compressive sensing, hidden Markov tree (HMT), process variation, virtual probe (VP).

I. INTRODUCTION

AS THE integrated circuits (ICs) scale to nanoscale feature size, process variation has become the major roadblock for circuit design, especially at advanced technology nodes [1]. The increasing fluctuations posed by IC manufacturing process lead to significant performance

variations and substantial yield loss [2]. Therefore, yield improvement has been considered as one of the most urgent tasks for today's IC design. Toward this goal, various techniques have been developed, e.g., statistical timing analysis [3]–[6] and post-silicon tuning [7]–[9], to minimize performance variability and enhance parametric yield. The efficiency of all these techniques relies heavily upon the accuracy of the underlying statistical model (e.g., probability distribution and correlation) that characterizes the process variation of interest.

Spatial variation modeling and characterization, however, is not a trivial task. Silicon wafers and chips must be carefully characterized with multiple test structures (e.g., ring oscillators (ROs)) deployed in wafer scribe lines and/or within product chips [10]. The traditional variation modeling methods often require a large number of test structures to monitor the spatial variation of interest. For instance, today's advanced microprocessor chip typically contains hundreds of on-chip ROs [11] to monitor process variations. Physically measuring all these test structures through a limited number of I/O pins is extremely time-consuming [12]. In addition, wafer probe test may even physically damage the wafer/chip due to mechanical stress [12]. The significant overhead in silicon area, testing time, and chip reliability leads to continuously growing modeling cost, as more and more test structures must be used to capture the increasingly complicated spatial variation pattern due to technology scaling [13].

Conventionally, the number of required test structures for spatial variation modeling is determined by the well-known Nyquist–Shannon sampling theory. However, a large amount of redundant data could be generated since the Nyquist–Shannon sampling theory assumes that all the frequency components below the maximum frequency may exist. Recently, several statistical methods, such as virtual probe (VP) [14]–[22] and Gaussian process [23]–[26], have been developed in the literature to capture the spatial variation pattern with low-measurement cost. Taking VP as an example, it exploits the fact that spatial variation often has a sparse presentation in the frequency domain based on discrete cosine transform (DCT). Namely, a large number of the DCT coefficients are close to zero. By taking advantage of such a sparse structure, VP is able to minimize the required measurement data based on compressive sensing [27]–[30].

In this paper, we aim to further improve the accuracy of VP by revisiting its fundamental assumption on frequency-domain sparsity. To explore the underlying sparse structure,

Manuscript received December 25, 2014; revised April 22, 2015, July 7, 2015, and August 25, 2015; accepted September 2, 2015. Date of publication September 24, 2015; date of current version May 18, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61376041, Grant 61125401, Grant 61376040, Grant 91330201, Grant 61574046, and Grant 61574044, in part by the National Basic Research Program of China under Grant 2011CB309701, in part by the Recruitment Program of Global Experts (the Thousand Talents Plan), in part by the National Science Foundation under Research Project 1115556 and Research Project CCF-1316363, and in part by the Laboratory of Mathematics for Nonlinear Science, Fudan University. This paper was recommended by Associate Editor H. Qian. (*Corresponding authors: Jun Tao and Xuan Zeng.*)

C. Liao, J. Tao, and X. Zeng are with the ASIC and System State Key Laboratory, Department of Microelectronics, Fudan University, Shanghai 200433, China (e-mail: taojun@fudan.edu.cn; xzeng@fudan.edu.cn).

Y. Su is with the School of Mathematical Sciences, Fudan University, Shanghai 200433, China (e-mail: yfsu@fudan.edu.cn).

D. Zhou is with the ASIC and System State Key Laboratory, Department of Microelectronics, Fudan University, Shanghai 200433, China, and also with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: zhou@utdallas.edu).

X. Li is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: xinli@andrew.cmu.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCAD.2015.2481868

VP statistically models the DCT coefficients as a zero-mean Laplace distribution where the DCT coefficients at different frequencies are considered to be mutually independent. In this way, the correlation among different DCT coefficients is completely ignored. However, as will be demonstrated by the industrial data in Section III, the low-frequency and high-frequency DCT coefficients are strongly correlated in the sense that if the low-frequency coefficients are close to zero, the corresponding high-frequency coefficients are likely to be zero as well.

To appropriately capture the correlation information and, hence, improve the accuracy of VP, we propose an efficient statistical model based on hidden Markov tree (HMT). Specifically, we model each DCT coefficient as a mixture distribution with a hidden state variable. Markovian dependency is used to describe the correlation between these hidden state variables based on a probabilistic tree (i.e., the HMT) [32]. Next, maximum-*a-posteriori* (MAP) estimation is used to optimally solve the unknown DCT coefficients and reconstruct the spatial variation pattern.

The idea of HMT has been developed for signal processing [32]. In this paper, we further adopt this powerful technique for spatial variation modeling of ICs. As will be demonstrated by the industrial examples in Section V, the proposed HMT method achieves up to 40% error reduction over other conventional approaches including orthogonal matching pursuit (OMP) [35], l_1 -norm regularization [19], [36], and reweighted l_1 -norm regularization [37].

The remainder of this paper is organized as follows. In Section II, we briefly review the background of spatial variation modeling by using VP. In Section III, we develop the proposed modeling algorithm based on HMT and then discuss several implementation issues in Section IV. The efficacy of HMT is demonstrated by several industrial examples in Section V. Finally, we conclude in Section VI.

II. BACKGROUND

A. Virtual Probe

Without loss of generality, in order to intuitively characterize the spatial variation, an interested performance (e.g., the frequency of an RO) can be expressed as a 2-D function $g(x, y)$, where x and y represent the coordinates of spatial location on a wafer [3], [19]. To capture the information in spatial frequency domain, $g(x, y)$ can be mapped to frequency domain by several kinds of transforms, such as Fourier transform, DCT, and wavelet transform. In this paper, similar as VP [19], we take DCT transform as an example. After discretization, the coordinates x and y can be denoted as integers $x \in \{1, 2, \dots, P\}$ and $y \in \{1, 2, \dots, Q\}$ and DCT transform [46] for $g(x, y)$ can be represented as

$$G(u, v) = \sum_{x=1}^P \sum_{y=1}^Q \alpha_u \cdot \beta_v \cdot g(x, y) \cdot \cos \frac{\pi \cdot (2x-1)(u-1)}{2 \cdot P} \times \cos \frac{\pi \cdot (2y-1)(v-1)}{2 \cdot Q} \quad (1)$$

where

$$\alpha_u = \begin{cases} \sqrt{1/P} & (u=1) \\ \sqrt{2/P} & (2 \leq u \leq P) \end{cases} \quad (2)$$

$$\beta_v = \begin{cases} \sqrt{1/Q} & (v=1) \\ \sqrt{2/Q} & (2 \leq v \leq Q). \end{cases} \quad (3)$$

In (1), $\{G(u, v); u = 1, 2, \dots, P, v = 1, 2, \dots, Q\}$ represents a set of DCT coefficients. The sampling value $\{g(x, y); x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$ can also be expressed as the linear combination of $G(u, v)$ by inverse DCT (IDCT) [46]

$$g(x, y) = \sum_{u=1}^P \sum_{v=1}^Q \alpha_u \cdot \beta_v \cdot G(u, v) \cdot \cos \frac{\pi \cdot (2x-1) \cdot (u-1)}{2 \cdot P} \times \cos \frac{\pi \cdot (2y-1) \cdot (v-1)}{2 \cdot Q}. \quad (4)$$

Based on (1)–(4), it is easy to verify that once all the sampling values $\{g(x, y); x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$ are known, the DCT coefficients $\{G(u, v); u = 1, 2, \dots, P, v = 1, 2, \dots, Q\}$ can be uniquely determined, and vice versa. However, in order to reduce testing cost, we always would like to recover $\{g(x, y); x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$ accurately from an extremely limited number of samples at locations $\{(x_m, y_m); m = 1, 2, \dots, M\}$, where $M \ll PQ$. So the recovery can be formulated as the following linear equation [19]:

$$\mathbf{A} \cdot \boldsymbol{\eta} = \mathbf{B} \quad (5)$$

where

$$\mathbf{A} = \begin{bmatrix} A_{1,1,1} & A_{1,1,2} & \cdots & A_{1,P,Q} \\ A_{2,1,1} & A_{2,1,2} & \cdots & A_{2,P,Q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{M,1,1} & A_{M,1,2} & \cdots & A_{M,P,Q} \end{bmatrix} \quad (6)$$

$$A_{m,u,v} = \alpha_u \cdot \beta_v \cdot \cos \frac{\pi \cdot (2x_m-1) \cdot (u-1)}{2 \cdot P} \times \cos \frac{\pi \cdot (2y_m-1) \cdot (v-1)}{2 \cdot Q} \quad (7)$$

$$\boldsymbol{\eta} = [G(1, 1) \quad G(1, 2) \quad \cdots \quad G(P, Q)]^T \quad (8)$$

$$\mathbf{B} = [g(x_1, y_1) \quad g(x_2, y_2) \quad \cdots \quad g(x_M, y_M)]^T. \quad (9)$$

If all the unknown coefficients $\boldsymbol{\eta}$ can be calculated from (5) based on the sampling data at locations $\{(x_m, y_m); m = 1, 2, \dots, M\}$, all the function values $\{g(x, y); x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$ can be obtained by (4). However, solving (5) is not trivial since $M \ll PQ$, i.e., the equations in (5) are profoundly underdetermined and cannot be uniquely solved by a simple matrix inverse [19]. As derived in [19], MAP estimation [46] is preferred to be used here to solve (5) based on the key idea that the optimal solution $\boldsymbol{\eta}$ should maximize the posterior distribution, namely

$$\max_{\boldsymbol{\eta}} \text{pdf}(\boldsymbol{\eta}|\mathbf{B}). \quad (10)$$

According to Bayes' theorem, we have

$$\text{pdf}(\boldsymbol{\eta}|\mathbf{B}) \propto \text{pdf}(\boldsymbol{\eta}) \cdot \text{pdf}(\mathbf{B}|\boldsymbol{\eta}). \quad (11)$$

Then the optimization problem (5) can be reformulated as

$$\max_{\boldsymbol{\eta}} \text{pdf}(\boldsymbol{\eta}) \cdot \text{pdf}(\mathbf{B}|\boldsymbol{\eta}). \quad (12)$$

Applying VP algorithm [19], (12) could be formulated as an l_1 -norm regularization problem by deducing $\text{pdf}(\boldsymbol{\eta})$ and $\text{pdf}(\mathbf{B}|\boldsymbol{\eta})$ as follows.

- 1) Based on the prior knowledge of spatial variation sparseness in frequency domain, each DCT coefficient is modeled as an independent random variable with zero-mean Laplace distribution. Then, we have

$$\text{pdf}(\boldsymbol{\eta}) = \left(\frac{1}{2\lambda}\right)^{PQ} \cdot \exp\left(-\frac{\|\boldsymbol{\eta}\|_1}{\lambda}\right) \quad (13)$$

where $\|\bullet\|_1$ denotes l_1 -norm, i.e., the summation of the absolute value of all the elements in a vector.

- 2) The likelihood function $\text{pdf}(\mathbf{B}|\boldsymbol{\eta})$ is characterized by a Dirac delta function, namely, $\text{pdf}(\mathbf{B}|\boldsymbol{\eta})$ is nonzero if and only if $\mathbf{A} \cdot \boldsymbol{\eta}$ equals \mathbf{B}

$$\text{pdf}(\mathbf{B}|\boldsymbol{\eta}) = \begin{cases} \infty & (\mathbf{A} \cdot \boldsymbol{\eta} = \mathbf{B}) \\ 0 & (\mathbf{A} \cdot \boldsymbol{\eta} \neq \mathbf{B}) \end{cases} \quad (14)$$

where

$$\int_{\mathbf{A} \cdot \boldsymbol{\eta} = \mathbf{B}} \text{pdf}(\mathbf{B}|\boldsymbol{\eta}) \cdot d\mathbf{B} = 1. \quad (15)$$

Then, according to (13)–(15), the optimization (12) can be rewritten as

$$\begin{aligned} \min \quad & \|\boldsymbol{\eta}\|_1 \\ \text{s.t.} \quad & \mathbf{A} \cdot \boldsymbol{\eta} = \mathbf{B}. \end{aligned} \quad (16)$$

Equation (16) is referred to as l_1 -norm regularization problem in compressive sensing field and can be solved by linear programming in [45]. Once the DCT coefficients $\boldsymbol{\eta}$ are calculated, the spatial variation model can be constructed by IDCT as in (4).

B. Embedded Zero-Tree

Note that the traditional VP method models each DCT coefficient as mutually independent. In contrast, the correlations among different DCT coefficients have been exploited in the field of image processing (e.g., image browsing, digital watermarking [34], and image compression [38]–[41]) to achieve superior recovery accuracy. For example, in image compression [41], by exploiting the self-similarity of the hierarchically decomposed sub-bands in the frequency domain, an embedded zero-tree (EZT) can be constructed to convey the significance information of the DCT coefficients. Next, the prior knowledge of insignificant coefficients could be efficiently encoded as part of the EZT. As a result, the error of compressed images could be largely reduced for the same compression ratio. In the next section, we will study the correlation model of DCT coefficients and the EZT representation in detail.

III. SPATIAL VARIATION MODELING VIA HIDDEN MARKOV TREE

In this section, we propose a novel spatial variation modeling method based on HMT to further improve the accuracy and efficiency of VP. This novel algorithm takes full advantage of the two kinds of prior knowledge for spatial variations in frequency domain, i.e., correlations and sparse structure of its DCT transform, while the previous one is completely ignored by VP. Here, we will study the DCT coefficient correlation by examining some industrial data at first. Then HMT is introduced for complete DCT coefficients modeling to capture these two properties. Based on the HMT training result and random Gaussian distribution assumption of the modeling residue, MAP estimation can be applied to formulate (5) as a convex optimization. In addition, an iteration scheme is proposed at last to refine the constructed HMT model and obtain a convergent result.

A. Probabilistic Model for Prior Knowledge

As previously mentioned, VP [19] has demonstrated the rationality of the coefficient sparseness structure and also taken advantage of it to reduce the modeling cost. Here, we will focus on studying another property, i.e., the strong correlations among DCT coefficients with different frequencies. Since spatial patterns of process variations are often smooth as is demonstrated in [19] and [44], the magnitude of DCT coefficients used for modeling will decay along with the frequency increasing, which is also referred as “decaying spectrum” in many other fields [40]–[43]. Suppose that the DCT coefficients correlation really exists and define that a coefficient is insignificant with respect to a given threshold T if its magnitude is smaller than T . Then along the similar frequency orientation, e.g., u -dimension in Fig. 1, it is reasonable to assume that if the low-frequency coefficient is insignificant, the higher-frequency coefficients will also be insignificant with a large probability.

In fact, we can decompose the frequency domain of DCT hierarchically [38]–[41] with many sub-bands sharing the similar frequency orientation as what follows. Initially, assume that a 2-D frequency domain can be divided at each dimension by using “low” and “high” frequency filters vertically and horizontally as shown in Fig. 1(a). The low frequency at this scale means $0 \leq |\omega| < \pi/2$ (where ω denotes the angular frequency), and the high frequency means $\pi/2 \leq |\omega| < \pi$. Then, totally four sub-bands (labeled LL_1 , HL_1 , LH_1 , and HH_1) arise, and HH_1 represents the sub-band of the finest scale (i.e., of the highest frequencies). To obtain the coefficients of lower frequency which are regarded to be more significant, the sub-band LL_1 could be further decomposed at a coarser scale as in Fig. 1(b), where the low frequency at this scale means $0 \leq |\omega| < \pi/4$, and the high frequency means $\pi/4 \leq |\omega| < \pi/2$. Such decomposition process stops until the required scale is reached [40]. Denote the angular frequency at u -dimension as ω_u , and the frequency at v -dimension as ω_v . It could be found clearly that, in Fig. 1(b), HH_2 and HH_1 share a similar frequency orientation since they are both selected out by high frequency filters at each dimension. For HH_2 , we have $\pi/4 \leq \omega_u < \pi/2$ and $\pi/4 \leq \omega_v < \pi/2$. For HH_1 ,

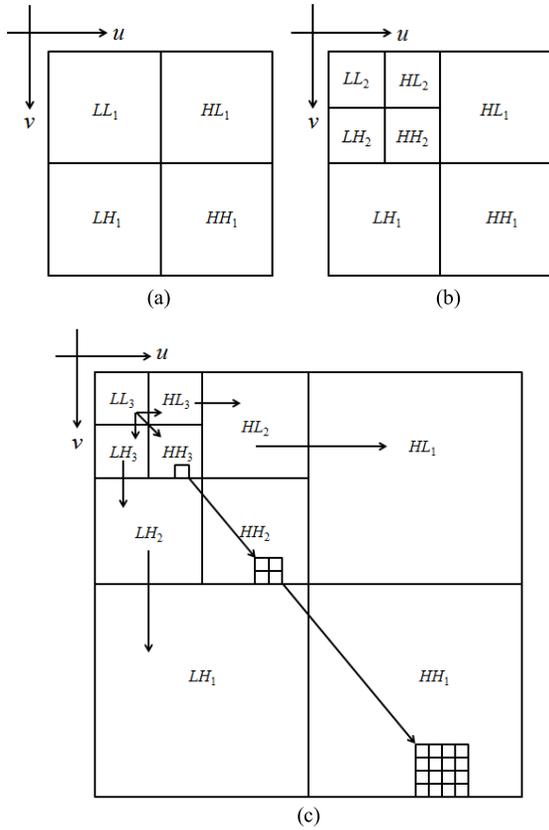


Fig. 1. (a)–(c) Hierarchical decomposition of the DCT in frequency domain, where 2-D of the frequency domain are denoted as u -dimension and v -dimension, respectively. Note that (c) shows the construction of EZT and the arrow points from parent to children. The lowest frequency sub-band is at the top left, and the highest frequency sub-band is at the bottom right.

we have $\pi/2 \leq \omega_u < \pi$ and $\pi/2 \leq \omega_v < \pi$. So the frequencies of HH_1 are around twice as much as those of HH_2 in each dimension. On the other hand, HL_2 and HL_1 also share a similar frequency orientation that is different from the one shared by HH_2 and HH_1 . This is because HL_2 and HL_1 are both selected out by using high filters at u -dimension but low filters at v -dimension. For HL_2 , we have $\pi/4 \leq \omega_u < \pi/2$ and $0 \leq \omega_v < \pi/4$, and for HL_1 , we have $\pi/2 \leq \omega_u < \pi$ and $0 \leq \omega_v < \pi/2$. If the spatial variation components at the frequency sub-band HL_2 are found to be sufficiently small, we expect that the DCT coefficients of HL_1 are also small.

Then, a tree could be constructed by connecting the sub-bands of similar frequency orientations at two successive scales as shown in Fig. 1(c). For the corresponding coefficients of these connected sub-bands, the one at the coarse scale (i.e., of low frequency) is called parent, and the others at the next finer scale (i.e., of higher frequency) of the similar frequency orientation are called children. Note that all the parents have four children except the LL sub-band at the coarsest scale that have three (i.e., the other three sub-bands at the coarsest scale) [38], [39]. Consequently, considering the similar frequency orientations shared by the connected sub-bands and the “decaying spectrum” of the coefficients [38], [39], [41], it is reasonable to assume that if the parent coefficient

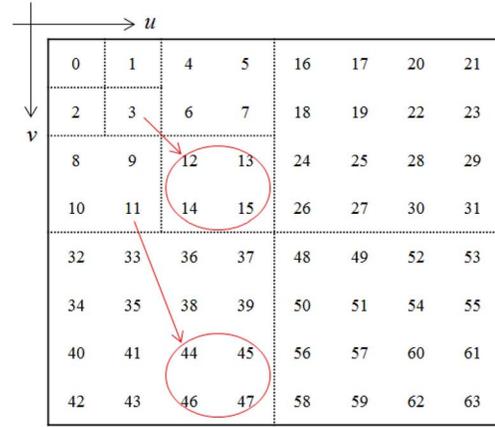


Fig. 2. 8×8 DCT coefficients are labeled to 0–63. For the top left component, $u = 1$ and $v = 1$, while for the bottom right component, $u = 8$ and $v = 8$. G_3 is the parent coefficient of $G_{\{12,13,14,15\}}$, and G_{11} is the parent coefficient of $G_{\{44,45,46,47\}}$.

is insignificant, the children coefficients will also be insignificant with a large probability. Based on this assumption, this tree structure might result in a few subtrees whose coefficients are all near to zero. Consequently, it is referred as EZT [38]–[41]. Further details about EZT could be founded in [38]–[41].

Take DCT with 8×8 coefficients as an example, which can be treated as an depth-3 tree as shown in Fig. 2, where coefficients $G(u, v)$ ($u = 1, 2, \dots, 8, v = 1, 2, \dots, 8$) is labeled as G_i ($0 \leq i \leq 63$). The parent of coefficient G_i is $G_{\lfloor i/4 \rfloor}$ for $1 \leq i \leq 63$, while the set of four children associated with coefficient G_j is $G_{\{4j, 4j+1, 4j+2, 4j+3\}}$ for $1 \leq j \leq 15$. Coefficient G_0 is the root of the DCT coefficient tree, which has only three children: $G_{\{1,2,3\}}$ [38].

Of course, if we consider different types of dependences among the coefficients, other kinds of trees could also be set up, but EZT seems to be the most popular one [38]. It has been previously used in image processing, including image browsing, progressive transmission [32], [33] and digital watermarking [34], and achieved superior performance.

In order to verify the rationality of the assumption of coefficients correlation with EZT for spatial variation modeling, let us conduct an experiment based on an industrial IC design here. At first, we measure the power of a set of RO from 16 wafers (each wafer contains around 110 chips). Then for each wafer, the DCT coefficients of the power spatial variation can be calculated and the corresponding EZT can be set up. To analyze correlations among these coefficients, we can calculate the conditional probability $P_{i|p(i)} = P(|G_i| < T | |G_{p(i)}| < T)$, where $G_{p(i)}$ is the parent of G_i , and $P(\bullet|\bullet)$ actually denotes the probability that the coefficient is insignificant given that its parent is insignificant. Then, it is found that when we set T to 0.2, which is much smaller than the maximum magnitude of the coefficients, the conditional probability $P_{i|p(i)}$ ranges from 0.78 to 0.9. In addition, the correlation coefficient between $|G_{p(i)}|$ and $|G_i|$ in this case is around 0.38. It has been observed by the image processing community that the correlation coefficients with EZT are around 0.35 for a large number of image applications [40]. Consequently, there indeed

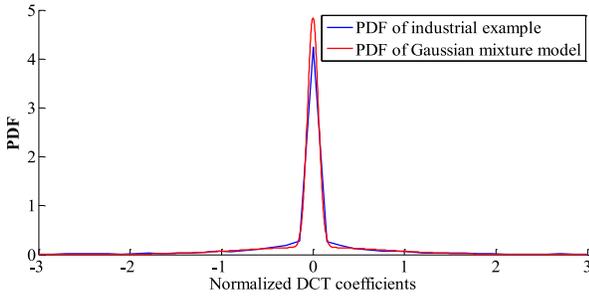


Fig. 3. Two-state, zero-mean Gaussian mixture model can closely fit the PDF of DCT coefficients obtained from the industrial data.

exist strong correlations among DCT coefficients of our spatial variation modeling, and EZT could capture this kind of correlations exactly.

Because the prior knowledge of coefficients characterization could further constrains the underdetermined equations (5), accurate and efficient characterization of coefficients correlations can provide tractable and unique pattern to the solution of (5). Based on EZT, all the dependencies among DCT coefficients can be completely characterized by using the joint probability density function. However, it could be very difficult or even impossible to estimate and use such kind of complete joint probability density because the number of coefficient combination grows exponentially in terms of the number of coefficients. Conversely, if we treat these coefficients independently as in VP [14], [19], all the intercoefficient correlations revealed above are totally disregarded. Therefore, it is required to strike a balance between the above two extremes, which means, the key and only the key correlations should be captured. Toward this goal, we introduce a probabilistic tree, i.e., HMT based on EZT, for coefficients modeling in this paper. The following two features contribute to the efficiency and flexibility of the proposed modeling scheme.

- 1) Considering the sparse structure, each coefficient is modeled as a mixture Gaussian density with a hidden state variable.
- 2) To characterize the key correlations between the coefficients, Markovian dependencies are introduced between the hidden state variables. Based on the EZT of DCT coefficients, these Markovian dependencies can be described by HMT.

Next, we will develop the mixture density model for an individual DCT coefficient in detail, and then extend it to HMT model for the whole DCT transform.

1) *Probabilistic Model for Marginal Distribution*: In order to characterize an individual DCT coefficient, let us restudy the previous empirical experiment where the interesting circuit performance is the power of ROs from 16 wafers. Fig. 3 shows the distribution of all the calculated DCT coefficients for the power spatial variation. It can be obviously found that for this industrial design, probability distribution function (PDF) of the coefficients implies a sparse structure, namely, only a few coefficients have large value, while most coefficients are near zero. The similar structure has been observed in [3] and [15], including VP for spatial variation characterization [14], [19].

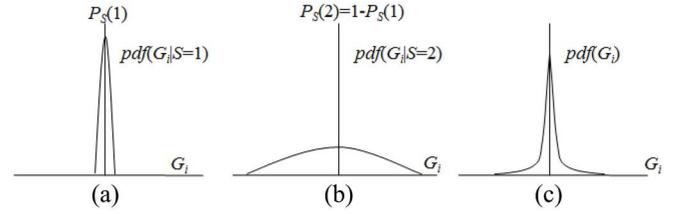


Fig. 4. Two-state, zero-mean Gaussian mixture model for an individual DCT coefficient. S denotes the hidden state variable. $P_S(1)$ denotes the corresponding probability massive function (PMF) of state variable when $S = 1$ and the coefficient has the low-variance Gaussian PDF. $P_S(2)$ denotes the corresponding PMF of state variable when $S = 2$ and the coefficient has the high-variance Gaussian PDF. (a) Low State. (b) High State. (c) Gaussian Mixture State.

To capture this sparse property and fit the distribution histogram perfectly, we can set up a simple model for each coefficient with a mixture of two states as shown in Fig. 4.

- 1) Low state associated with a Gaussian distribution that has zero-mean and small variance σ_L . This low state corresponds to the coefficients with very small values. Note that the small variance means the coefficient value in this state is near to zero with a large probability.
- 2) High state associated with a Gaussian distribution that has zero-mean and large variance σ_H . This high state corresponds to the coefficients with large values. The large variance means the coefficient in this state is very likely to hold large value far from zero.

As shown in Fig. 4, with this two-state mixture model, each coefficient belongs either to a low state or to a high state. Therefore, these coefficients could be completely modeled by the variances (i.e., σ_L and σ_H) of the Gaussian distribution corresponding to each state and the PMF of the state variable S . Let $P_S(1)$ denote the PMF of state variable when $S = 1$ and $P_S(2)$ denote the PMF when $S = 2$. Then with this two-state mixture model, $P_S(2) = 1 - P_S(1)$ and the PDF of individual DCT coefficient can be expressed as

$$\text{pdf}(G_i) = P_S(1) \cdot \text{pdf}(G_i|S = 1) + P_S(2) \cdot \text{pdf}(G_i|S = 2). \quad (17)$$

In our application, the value of DCT coefficient could be observed while the value of S is not, so we say that the value of state variable is hidden.

To prove validity, we use this simple two-state mixture model to fit the PDF of DCT coefficients obtained from the previous industrial data for the power of processors. As shown in Fig. 3, PDF of the modeling result is almost the same as that of the industrial data, which demonstrates its efficiency. It is obvious that the modeling accuracy could be improved if we use $N > 2$ states and allow nonzero mean value. However, since the two-state mixture model is simple, effective, and easy to use, we will focus on it in our applications. But the proposed algorithms can also be applied to handle general mixture models. Actually, such kind of Gaussian mixture model has been applied in many other fields and proved to be extremely useful [32], [33].

In summary, the general N -state Gaussian mixture model consists of the following parameters.

- 1) A discrete random state variable S with the value s , $s \in \{1, 2, \dots, N\}$, and the corresponding PMF, $P_S(1), P_S(2), \dots, P_S(N)$.
- 2) The Gaussian conditional PDF for each state $\text{pdf}(G_i|S = s)$, $i = 0, 1, \dots, PQ - 1$. Then the PDF of an individual DCT coefficient G_i can be expressed as

$$\text{pdf}(G_i) = \sum_{s=1}^N P_S(s) \cdot \text{pdf}(G_i|S = s). \quad (18)$$

2) *Probabilistic Model for Joint Distribution:* Intuitively, (18) can be used to capture the sparse structure of spatial variation modeling efficiently, but the correlations between DCT coefficients as we have discussed before are totally discarded. Ideally, we would like to find a model that not only matches the PDF of individual coefficients but also captures coefficient correlations. Toward this goal, we can extend the previous two-state mixture model to Gaussian mixtures with interdependent state variables, i.e., model each coefficient as a Gaussian mixture, and allow probabilistic dependencies between the hidden state variables so as to explore the correlations between DCT coefficients.

Then the remaining problem is how to model the dependencies between the interdependent state variables appropriately. It is obvious that if we take into account all possible dependencies to establish the complete joint PDF, the complexity will become unacceptable. Fortunately, as we have discussed before, the major correlations between the coefficients could be characterized by EZT. Therefore, we could construct HMT based on EZT to characterize the dependencies among state variables and model the whole DCT transform.

HMT is one kind of general probabilistic graph that is useful to model the local dependencies between random variables [32]. In this paper, the proposed HMT shares the same topology as that of EZT defined in Fig. 2, but each node corresponds to the hidden state variable instead of DCT coefficient. Dependencies between pairs of states are represented by links connecting the corresponding nodes. Actually, HMT specifies the dependency between the parent and children state variables as the first order Markov process, i.e., the children state depends only on its parent state. For example, as in Fig. 5, all the state variables S_4 – S_7 are children of S_1 , so they all depend on S_1 and only on S_1 . However, it should be noted that dependencies are not simply limited to parent–children interactions. S_4 – S_7 may be highly correlated due to their relationship with S_1 .

State variable dependencies can be characterized by state transition probabilities. According to our numerical results given in Section V, a large state transition probability often appears when one node and its parent are both in low state, while a small probability appears when the parent is in low state but the children is in high state.

Let S_i denote the hidden state variable corresponding to DCT coefficient G_i , $i = 0, 1, \dots, PQ - 1$. The parameters

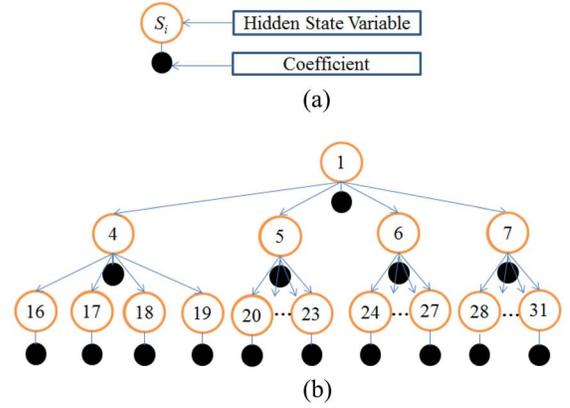


Fig. 5. For an individual coefficient G_i , it is modeled as Gaussian mixture with a hidden state variable S_i . In (a), the smaller black circle represents the corresponding coefficient, and the bigger circle represents S_i . Meanwhile, to represent the intercoefficient dependencies, we link the hidden states and establish an HMT model. (b) Truncation of HMT.

for HMT with N -state Gaussian mixture model include the following.

- 1) $P_{S_0}(s)$, the PMF of the root S_0 when it takes value $s \in \{1, 2, \dots, N\}$.
- 2) $\varepsilon_{i,p(i)}^{sr} = P_{S_i|S_{p(i)}}(S_i = s|S_{p(i)} = r)$, the state transition probability that S_i is in state s given $S_{p(i)}$ is in state r , where $S_{p(i)}$ is parent of S_i and $s, r \in \{1, 2, \dots, N\}$.
- 3) $\mu_{i,s}$ and $\sigma_{i,s}^2$, the mean value and variance of the DCT coefficient G_i with Gaussian distribution given S_i is in state s .

All the model parameters can be grouped together as a model parameter vector θ . In this paper, recall that we focus on two-state zero-mean Gaussian mixture model, so $N = 2$ and $\mu_{i,s} = 0$.

It should be noted that some conditional dependencies for the hidden state variables are implicitly contained in the Markovian tree connection. For state variable S_i , given its parent state $S_{p(i)}$, the value of S_i is independent of the entire tree except for S_i 's descendants. Conversely, given its children state $S_{ch(i)}$, the value of S_i are independent of $S_{ch(i)}$'s descendants. For instance, Fig. 5(b) represents a truncation of our HMT based on EZT. For state S_4 , given the value of its parent state S_1 , the value of S_4 is independent of the entire tree except for its descendants. Meanwhile, if given its children state S_{16} , the value of state S_4 is independent of S_{16} 's descendants. Combining these correlations together, we see that given the value of parent state S_1 and the children states S_{16} – S_{19} , the value of S_4 is conditionally independent of the rest of the tree.

3) *Training of HMT via EM Algorithm:* In order to obtain the value of HMT parameters θ , training procedure should be done for the DCT coefficients based on the industrial data. Maximum likelihood (ML) principle [46] is always used for parameter estimation. However, direct ML method [46] is intractable here for the training of θ since the hidden state variables $S = [S_0, S_1, \dots, S_{PQ-1}]$ are invisible. Therefore, we introduce another widely used method, i.e., iterative expectation maximization (EM) algorithm.

Algorithm 1 EM algorithm for HMT Training

1. Initialize the model parameters as θ^0 , set iteration counter $l = 0$;
2. E step: Calculate the probabilities $P(S_i = s|\eta, \theta^l)$ and $P(S_i = s, S_{p(i)} = r|\eta, \theta^l)$;
3. M step: Update the entries of θ^{l+1} to maximize the likelihood function as:

$$P_{S_i}(s) = P(S_i = s|\eta, \theta^l), \quad (19)$$

$$\varepsilon_{i,p(i)}^{s,r} = \frac{P(S_i = s, S_{p(i)} = r|\eta, \theta^l)}{P(S_{p(i)} = r|\eta, \theta^l)}, \quad (20)$$

$$\mu_{i,s} = \frac{G_i P(S_i = s|\eta, \theta^l)}{P_{S_i}(s)}, \quad (21)$$

$$\sigma_{i,s}^2 = \frac{(G_i - \mu_{i,s})^2 P(S_i = s|\eta, \theta^l)}{P_{S_i}(s)}. \quad (22)$$

4. Set $l = l+1$. If θ converges, then stop; else, go to step 2.

The goal of EM algorithm given in Algorithm 1 is to maximize a likelihood function $\ln f(\eta|\theta)$ which measures how well the model parameter θ describes the coefficients η . Toward this goal, it decouples the complex likelihood function maximization into iterations between two simple steps, i.e., the *E* and *M* steps. At $l+1$ th iteration, in *E* step, it calculates the two kinds of PMF for hidden state variables: 1) $P(S_i = s|\eta, \theta^l)$ and 2) $P(S_i = s, S_{p(i)} = r|\eta, \theta^l)$, $i = 0, 1, \dots, PQ-1$ and $s, r \in \{1, 2, \dots, N\}$ (θ^l represents the HMT parameters obtained in l th iteration) by using upward-downward algorithm [32], [46]. In *M* step, it optimizes the model parameters θ to achieve maximum value of the likelihood function. After the initialization of θ , the EM algorithm iterates *E* and *M* steps until convergence. Further details about the algorithm are referred in [31] and [32].

Additionally, since the number of available data for training could be very limited, “tying” strategy [32] can be applied here to obtain more robust parameter estimations. It is rational to assume that the DCT coefficients at the same level of EZT have similar properties since they correspond to similar components in frequency domain. Then, with the tying strategy, we can allow these coefficients to share the same model parameters, i.e., $\varepsilon_{i,p(i)}^{s,r}$, $\mu_{i,s}$, and $\sigma_{i,s}^2$ (recall that in this paper, $\mu_{i,s}$ is set to 0). The superior performance of this tying strategy has been proved in [32].

Once HMT training result is obtained, the PDF of each DCT coefficient G_i can be expressed as

$$\text{pdf}(G_i) = \sum_{s=1}^N P_{S_i}(s) \cdot \text{pdf}(G_i|S_i = s) \quad (23)$$

where $P_{S_i}(s)$ can be obtained with θ , the training result of HMT parameter vector. Then MAP estimation can be used to optimize the unknown DCT coefficients $\eta = [G_0, G_1, \dots, G_{PQ-1}]$ by solving (12) and reconstruct the spatial variation pattern as illustrated in the following section.

B. Maximum-a-Posteriori Estimation

According to Bayesian inference and (11), to maximize the $\text{pdf}(\eta|\mathbf{B})$, we should deduce $\text{pdf}(\eta)$ and $\text{pdf}(\mathbf{B}|\eta)$ first.

Based on (23), the PDF of η can be presented as

$$\begin{aligned} \text{pdf}(\eta) &= \prod_{i=0}^{PQ-1} \text{pdf}(G_i) \\ &= \prod_{i=0}^{PQ-1} \left(\sum_{s=1}^N P_{S_i}(s) \cdot \text{pdf}(G_i|S_i = s) \right). \end{aligned} \quad (24)$$

Note that when HMT parameters are obtained by EM training, the probability $P_{S_i}(s)$ [means $P_{S_i}(S_i = s)$, $s = 1, 2, \dots, N$] can be calculated as a series of constants, and $\text{pdf}(G_i)$ with the unknown coefficient G_i is actually a sum of Gaussian distributions $\text{pdf}(G_i|S_i = s)$ with the weight $P_{S_i}(s)$. Then $\text{pdf}(\eta)$ becomes a nonconvex function, so does the objective function $\text{pdf}(\eta) \cdot \text{pdf}(\mathbf{B}|\eta)$ in (12). But it is always complicated or even impossible to find out the global optimum of a nonconvex function numerically. Fortunately, as shown in the numerical results given in Section V, for the state variable S_i in HMT, the probability of a primary option $S_i = s_{ip}$ could be much larger than that of others, i.e., the coefficient G_i will be located in the primary state s_{ip} with a much larger probability compared with other states. Therefore, the PDF of G_i can be simplified as only one Gaussian distribution $\text{pdf}(G_i|S_i = s_{ip})$ with a weight $P_{S_i}(s_{ip})$. This means that $\text{pdf}(\eta)$ can be approximated by the product of a series of Gaussian distributions corresponding to the most probable sequence of hidden states s_{ip} , $i = 0, 1, \dots, PQ-1$ for each state

$$\text{pdf}(\eta) = \prod_{i=0}^{PQ-1} \left(P_{S_i}(s_{ip}) \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_{ip}} \cdot \exp\left(-\frac{G_i^2}{2 \cdot \sigma_{ip}^2}\right) \right) \quad (25)$$

where σ_{ip} corresponds to the variance of the Gaussian distribution of the primary state s_{ip} and is referred as primary variance at the rest of this paper. Then $\text{pdf}(\eta)$ becomes a convex function.

Intuitively, we can get the most probable sequence of hidden states by enumerating all the possibilities, and choose the one with the ML. However, the computational complexity could be up to $O(N^{PQ})$, which is really expensive. Luckily, Viterbi algorithm [46] is a standard method for HMT that can be used here to find out this sequence. It should be noted that the complexity of Viterbi algorithm is linear [i.e., $O(PQ)$] for its fusion of dynamic programming.

The prior PDF in (25) has a twofold meaning.

- 1) Most of σ_{ip} , $i = 0, 1, \dots, PQ-1$ are very small as shown in the numerical results in Section V, which indicates sparse structure of DCT transform, i.e., most of the DCT coefficients are close to zero with a large probability.
- 2) The probability $P_{S_i}(s_{ip})$ of the state variable S_i is strongly affected by its parent with the consideration of the state transition probability $\varepsilon_{i,p(i)}^{s,r}$ during the calculation. The dependencies among the state variables represent the correlations of DCT coefficients as we have discussed before.

On the other side, in order to calculate $\text{pdf}(\mathbf{B}|\boldsymbol{\eta})$, we can commonly assume that the residue between the measurement data and the recovered results $\mathbf{t} = \mathbf{B} - \mathbf{A} \cdot \boldsymbol{\eta}$ can be modeled as a sequence of white Gaussian noise with the variance σ_t^2 . Then we have

$$\text{pdf}(\mathbf{B}|\boldsymbol{\eta}) = \frac{1}{\sqrt{2\pi} \cdot \sigma_t} \cdot \exp\left(-\frac{1}{2\sigma_t^2} \sum_{m=1}^M (B_m - A_{m,:} \cdot \boldsymbol{\eta})^2\right). \quad (26)$$

Recall that M represents the number of obtained samples and $M \ll PQ$. The value of σ_t can be determined by a statistical technique, i.e., cross-validation, in [46].

The unknowns in (25) and (26) are the coefficients $\boldsymbol{\eta}$, while $P_{Si}(s_{ip})$, σ_{ip} , and σ_t are constants previously calculated. Then substitute (25) and (26) into (12), the optimization problem in (10) can be transformed to

$$\min_{\boldsymbol{\eta}} \sum_{i=0}^{PQ-1} \frac{G_i^2}{\sigma_{ip}^2} + \sum_{m=1}^M \left(\frac{B_m - A_{m,:} \cdot \boldsymbol{\eta}}{\sigma_t} \right)^2. \quad (27)$$

Note that the optimization of (27) tries to minimize the sum of two parts.

- 1) The first part can be regarded as a weighted l_2 -norm for DCT coefficients. The weight $1/\sigma_{ip}^2$ is obtained from HMT model and indicates the coefficient correlation information. The minimization of this part can be regarded to guarantee the sparsity of DCT coefficients while considering their intercorrelations. But the HMT model used here is obtained from the input, i.e., the initialized DCT coefficients that could be coarse and noisy, which means the estimated σ_{ip} could be inexact.
- 2) The second part refers to the normalized mean squared error (MSE) for spatial variation modeling. Obviously the goal of minimization for this part is to improve the modeling accuracy. Note that the measurement samples \mathbf{B} are introduced in this part to revise the DCT coefficients $\boldsymbol{\eta} = [G_0, G_1, \dots, G_{PQ-1}]^T$.

It is obvious that the objective function of (27) is convex. Therefore, the traditional convex optimization methods, such as interior point algorithm [45], can be applied here to obtain the optimal coefficients $\boldsymbol{\eta}$ robustly and efficiently.

C. Iteration Flow

As discussed in the above section, MAP estimation takes advantage of HMT training result to optimize the unknown coefficients $\boldsymbol{\eta}$. But the accuracy of HMT model partly depends on the input initial coefficients, which might be coarse and noisy. In order to refine the HMT model and σ_{ip} in (27), and figure out a convergent and accurate result, we propose an iteration scheme based on HMT and Bayesian inference. During each iteration, EM algorithm is applied first to construct the HMT model based on the coefficients obtained from the previous iteration. Then Viterbi algorithm is used to figure out the most possible sequence of HMT hidden states and MAP estimation formulates the original spatial variation characterization as a convex optimization problem (27). By solving (27), the refined coefficients for the next iteration can be calculated.

This iteration terminates until $\|\boldsymbol{\eta}^{l+1} - \boldsymbol{\eta}^l\|_2 < \delta$, where $\|\bullet\|_2$ denotes the l_2 -norm, $\boldsymbol{\eta}^l$ denotes the coefficients obtained at l th iteration, and δ is the threshold set as a small value, e.g., 10^{-4} .

IV. IMPLEMENTATION DETAILS

In order to make the proposed spatial variation modeling method practically efficient, two important implementation issues should be considered carefully during the above iterations based on HMT and Bayesian inference, i.e., the coefficient initialization for the iteration scheme and the refinement of Bayesian inference (27). In this section, we will describe the two issues in detail.

A. Initialization of Bayesian Inference

According to the HMT modeling process and MAP estimation (27) based on Bayesian inference, the efficiency of the proposed iteration scheme depends on both the initialized DCT coefficient and the measurement samples \mathbf{B} . It is obvious that the more accurate the initialized coefficients are, the faster the convergence rate of the iteration scheme is. Actually, we can apply the traditional compressive sensing methods, such as OMP algorithm [35] or l_1 -norm regularization [36], to initialize the coefficients by taking full advantage of the sparse property. Then HMT modeling and Bayesian inference can be used to capture the coefficient correlations and refine the coarse initialization. In this paper, we take OMP as an example for the initialization.

The goal of OMP is to figure out the sparse solution for the optimization problem

$$\begin{aligned} \min_{\boldsymbol{\eta}^0} \quad & \|\mathbf{A}\boldsymbol{\eta}^0 - \mathbf{B}\|_2 \\ \text{s.t.} \quad & \|\boldsymbol{\eta}^0\|_0 \leq p \end{aligned} \quad (28)$$

where \mathbf{A} and \mathbf{B} are given in (6) and (9), respectively. \mathbf{A} can be represented as $\mathbf{A} = [A_{:,0}, A_{:,1}, \dots, A_{:,PQ-1}]$, where $A_{:,i}$, $i = 0, 1, \dots, PQ - 1$, is the column vector of $\mathbf{A} \cdot \boldsymbol{\eta}^0$ is the initialized coefficients required to be solved for the iteration scheme, $\|\bullet\|_0$ denotes the l_0 -norm, i.e., the number of nonzero components. p is an arbitrary integer meaning the upper bound of the number of nonzero elements in $\boldsymbol{\eta}^0$ calculated by (28). According to the sparse structure of DCT coefficients in our application, we have $p \ll M$. Then the rest $M-p$ elements in $\boldsymbol{\eta}^0$ are set zero.

The key idea of OMP [35] is to pick up the most useful column of \mathbf{A} in a greedy fashion, which makes the chosen column and the present residue related to the greatest extent. Then its contribution is subtracted from \mathbf{B} . The above procedure should be repeated until the number of iterations is up to p . Subsequently, the optimal result obtained at p th iteration is regarded as the initial coefficients $\boldsymbol{\eta}^0$ for the iteration scheme proposed in Section III-C. Algorithm 2 shows the detailed flow of OMP in our application.

Algorithm 2 assumes that p is given as the input. In practice, p is unknown and often determined automatically by cross-validation technique [46]. The key idea here is to run Algorithm 2 repeatedly for different values of p and calculate

Algorithm 2 OMP Algorithm

1. Initialize the residual $r_0 = \mathbf{B}$, the index set $\Lambda_0 = \phi$, $\Xi_0 = \{0, 1, \dots, PQ - 1\}$, the chosen columns $\mathbf{A}_0 = \phi$, and the iteration counter $l = 1$;
2. Find the index λ_l by solving the optimization problem:

$$\lambda_l = \arg \max_{\lambda \in \Xi_{l-1}} |\langle r_{l-1}, \mathbf{A}_{:, \lambda} \rangle|. \quad (29)$$

where $\langle \bullet, \bullet \rangle$ denotes the inner product of the two vectors.

3. Augment the index set and the matrix of chosen atoms: $\Lambda_l = \Lambda_{l-1} \cup \{\lambda_l\}$ and $\mathbf{A}_l = [\mathbf{A}_{l-1}, \mathbf{A}_{:, \lambda_l}]$.
4. Solve the least square problem to obtain a new signal estimate:

$$\boldsymbol{\eta}_l^0 = \arg \min_{\boldsymbol{\eta}^0} \|\mathbf{B} - \mathbf{A}_l \boldsymbol{\eta}^0\|_2. \quad (30)$$

5. Calculate the new residual:

$$r_l = \mathbf{B} - \mathbf{A}_l \boldsymbol{\eta}_l^0. \quad (31)$$

6. Set $l = l + 1$, $\Xi_l = \Xi_{l-1} - \{\lambda_l\}$. If $l < p$, return to Step 2.

Algorithm 3 Spatial Variation Modeling Based on HMT

1. Collect M measurement data at locations $\{(x_m, y_m); m = 1, 2, \dots, M\}$.
2. Initialize the DCT coefficients as $\boldsymbol{\eta}^0$ by applying OMP given in Algorithm 2.
3. Set the iteration counter $l = 1$.
4. Set $\Delta^0 = \|\boldsymbol{\eta}^0 - 0\|_2$ and $\delta = 10^{-4}$, where δ is the convergence threshold.
5. **while** ($\Delta^{l-1} > \delta$) **do**
6. Construct the HMT based on DCT coefficients $\boldsymbol{\eta}^{l-1}$ via EM algorithm given in Algorithm 1;
7. Obtain the most possible primary state sequence via Viterbi algorithm and determine the primary variance σ_{ip} in (25) for each coefficient;
8. Calculate the variance of the residue σ_t in (26) via cross validation;
9. Calculate the coefficients $\boldsymbol{\eta}^l$ by solving the convex optimization (32);
10. Update $\Delta^l = \|\boldsymbol{\eta}^l - \boldsymbol{\eta}^{l-1}\|_2$;
11. Set $l = l + 1$;
12. **end while**

the modeling error associated with each p . Once the relationship between modeling error and p is known, the optimal value of p is determined by minimizing the modeling error.

B. Refinement of Bayesian Inference

According to MAP and Bayesian inference, (27) is utilized to optimize the DCT coefficients as discussed in Section III-B. As shown in (27), the sum of two parts should be minimized, i.e., the weighted l_2 -norm of coefficients and the normalized MSE of spatial variation modeling. Note that for the first part, if the estimated σ_{ip} is close to zero, the correlated coefficient G_i would be forced to take the value of zero, which might deviate from accurate estimation especially during the early iterations. In order to improve stability and ensure that a zero-valued σ_{ip} does not strictly prohibit a nonzero estimate of G_i at the next step, we introduce $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \dots, \gamma_{PQ-1}]$ ($\gamma_i \neq 0$) in (27) to refine this cost function as

$$\min_{\boldsymbol{\eta}} \sum_{i=0}^{PQ-1} \frac{G_i^2}{\sigma_{ip}^2 + \gamma_i} + \sum_{m=1}^M \left(\frac{B_m - A_{m, :} \cdot \boldsymbol{\eta}}{\sigma_t} \right)^2. \quad (32)$$

Actually, the similar refinement technique has also been applied in reweighted l_1 -norm regularization method [37]. Meanwhile, (32) is still a convex optimization problem that can be solved efficiently and accurately.

Furthermore, since spatial patterns of process variations are often smooth as is demonstrated in [44], the magnitude of DCT coefficients used for modeling will decay along with the frequency increasing [42]. Hence, in order to improve the modeling accuracy, it is rational to let the value of corresponding γ_i decrease with the increase of coefficient frequency. In this paper, we assume that γ_i for the coefficients at the same level of the HMT have the same value due to the tying strategy mentioned before. For different levels, γ_i can be empirically set as $10^{(0.5-l_i)}$ ($l_i = 1, 2, \dots, L_t$, L_t is the maximum level of HMT).

C. Summary

Algorithm 3 summarizes the major steps of the proposed spatial variation modeling method in this paper. At the beginning, OMP algorithm is applied to initialize the DCT coefficients. Then during each iteration, with the coefficients obtained from previous iteration, HMT model is established, based on which, the original modeling problem (5) is transformed to a convex programming (32) via MAP estimation. Afterwards, DCT coefficients could be updated. Once we figure out the convergent result, the spatial variation model $\{g(x, y); x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$ can be established by IDCT as shown in (4). Applying HMT, the correlations and sparsity of the modeling coefficients can be captured exactly. Therefore, given the same number of sampling points, the modeling accuracy of proposed method can be improved evidently compared with the existing methods, which will be demonstrated by the industrial examples in Section V. Actually, even in the cases with little correlations among the DCT coefficients, the proposed method could still achieve the similar modeling accuracy as the previous VP method [19] (whose key algorithm is l_1 -norm regularization). This is because sparsity of the coefficients is considered carefully in all these methods.

In what follows, we will demonstrate the efficacy of the proposed spatial variation modeling method by some industrial examples.

V. NUMERICAL EXAMPLES

In this section, we validate the proposed algorithm by constructing the spatial variation models of power and frequency based on the measurement data collected for a set of ROs. We do not use the data set in [19] for our experiment, because it is proprietary and inaccessible to us. All the numerical

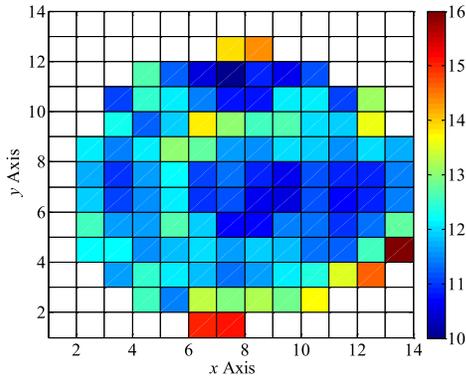


Fig. 6. Measurement power values of 108 industrial chips from the same wafer are mapped to 10–16, which show significant spatial variations.

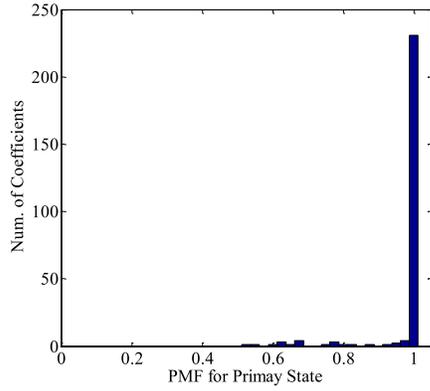


Fig. 7. PMFs of the primary states for all the 256 DCT coefficients. It can be found that the PMFs for more than 92% primary states are larger than 0.9.

experiments are performed on a computer with 3.2 GHz CPU and 8 GB memory.

A. Power Measurement Data

We consider the power data measured from 108 industrial chips on the same wafer as shown in Fig. 6. It is observed that the power value significantly varies from chip to chip due to process variation. In order to capture the power spatial variations at wafer-level, a 2-D function $g(x, y)$ is utilized to model the power, where $x = 1, 2, \dots, 14$ and $y = 1, 2, \dots, 14$. Each coordinate point (x, y) corresponds to a chip. Meanwhile, for simplicity, a two-state zero-mean HMT is applied for DCT coefficient modeling.

Since the two types of prior knowledge, i.e., sparse structure and coefficient correlations, are the key fundamentals of our proposed HMT method, we will give the training result of HMT based on power data to demonstrate the rationality of these assumptions first, then show the whole modeling results for power variations.

1) *Prior Knowledge of DCT Coefficients*: Fig. 7 shows the PMF of primary states and Fig. 8 gives the histogram of primary variances corresponding to all coefficients. From Fig. 7, we can find that the PMFs for more than 92% primary states obtained by Viterbi algorithm are larger than 0.9. This means that for this two-state HMT model, the probability of the primary state is much larger than the other option. Therefore, it is

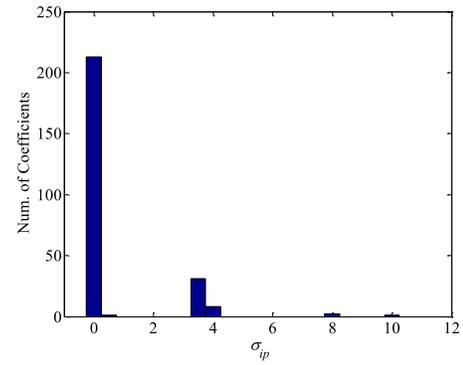


Fig. 8. Histogram of primary variances $\sigma_{ip}, i = 0, \dots, PQ - 1$. σ_{ip} corresponding to more than 80% coefficients are close to zero.

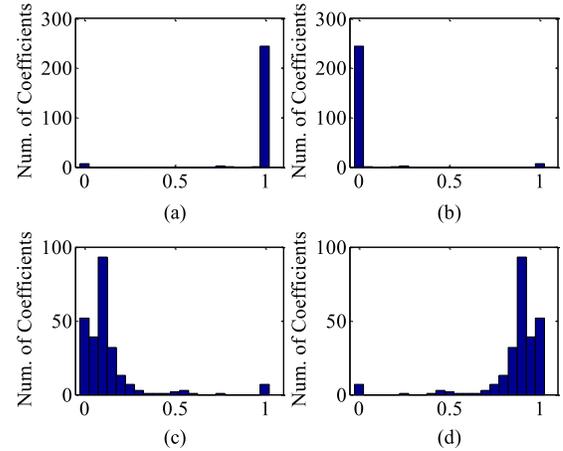


Fig. 9. Conditional probabilities $P(S_i = s, S_{p(i)} = r)$ for 256 DCT coefficients, where $s, r \in \{1, 2\}$, “1” represents low state, and “2” represents high state. (a) $P(S_i = 1, S_{p(i)} = 1)$. (b) $P(S_i = 2, S_{p(i)} = 1)$. (c) $P(S_i = 1, S_{p(i)} = 2)$. (d) $P(S_i = 2, S_{p(i)} = 2)$.

exactly rational to approximate $\text{pdf}(\eta)$ by only the product of a series of Gaussian distributions corresponding to the primary states as given in (25). From Fig. 8, it can be found that the primary variances ($\sigma_{ip}, i = 0, 1, \dots, PQ - 1$) corresponding to more than 80% coefficients are close to zero. This indicates the sparse structure of the coefficients, i.e., most coefficients possess low-variance Gaussian PDF and would be close to zero with very large probability.

Fig. 9 gives the conditional probabilities $P(S_i = s, S_{p(i)} = r), s, r \in \{1, 2\}$ for all the coefficients obtained from the two-state zero-mean HMT training result. From Fig. 9, we can find that $P(S_i = 1, S_{p(i)} = 1)$ is concentrated near 1 as in Fig. 9(a), while $P(S_i = 2, S_{p(i)} = 1)$ is concentrated near 0 as in Fig. 9(b). This implies that for this HMT model, if the parent is in low state, i.e., $S_{p(i)} = 1$, the children are more likely to be in low state (i.e., $S_i = 1$) than in high state (i.e., $S_i = 2$). Similarly, from Fig. 9(c) and (d), we can find that if the parent is in high state, the children are more likely to be in high state than in low state. The second property of DCT transform, i.e., coefficient correlations, is reflected by Fig. 9 exactly.

From the above numerical results, we can find that it is rational to regard the two types of prior knowledge, i.e., sparse

structure and coefficient correlations, as the key fundamentals for our modeling method based on HMT.

2) *Spatial Variation Modeling*: In order to quantitatively evaluate the accuracy of the proposed HMT method for power variation modeling, we repeatedly run Algorithm 3 to predict the wafer-level spatial variations with different numbers (i.e., M) of spatial samples. For testing and comparison purposes, we further consider three other existing methods: 1) l_1 -norm regularization [19]; 2) OMP [35]; and 3) reweighting l_1 -norm regularization [37]. The comparison against l_1 -norm regularization is based on the source code implemented by Carnegie Mellon University. All the heuristic techniques described in [19], e.g., modified-Latin hypercube sampling and cross-validation, have also been applied in our experiment.

The accuracy of prediction metric is defined as the average error

$$E_{\text{AVG}} = \sqrt{\frac{\sum_x \sum_y [g(x, y) - \hat{g}(x, y)]^2}{\sum_x \sum_y [g(x, y)]^2}} \quad (33)$$

where $g(x, y)$ and $\hat{g}(x, y)$ denote the measured value and the estimated value of the power at location (x, y) , respectively.

Fig. 10 shows the prediction error [both the average error and its standard deviation] as a function of the number of samples (i.e., M) for different algorithms. It is clear that the proposed HMT method achieves superior accuracy than the other algorithms. This is because the proposed HMT method takes full advantage of two inherent properties of DCT coefficient, i.e., the sparse structure and coefficient correlations, while the later one is completely ignored in all the traditional methods. Furthermore, OMP method and reweighted l_1 -norm regularization method show relatively larger recovery error in this case. This is because both of these algorithms favor strong sparsity in frequency domain. However, as shown in Fig. 11, the DCT coefficients of dynamic power estimated in one run are only approximately (but not extremely) sparse, i.e., most coefficients are observed with small values but the number of coefficients with large values is still considerable.

Meanwhile, another two important observations could be made based on Fig. 11. First, similar as VP methods [14], [19], substantial high-frequency components are contained in $G(u, v)$, implying that the spatial variation sampling rate could not be largely reduced due to Nyquist–Shannon sampling theorem by using the traditional modeling methods. Second, $G(u, v)$ is correlated, i.e., if the parent coefficient is close to zero, the children are also very likely to be zero.

Particularly, Fig. 12 shows the power data predicted from 40 tested chips (i.e., $M = 40$) by the proposed HMT method. Comparing Figs. 6 and 12, it can be found that the power spatial variation is predicted well. To quantitatively assess the prediction accuracy of each chip, we calculate the following relative error:

$$E_{\text{REL}}(x, y) = \left| \frac{g(x, y) - \hat{g}(x, y)}{g(x, y)} \right|. \quad (34)$$

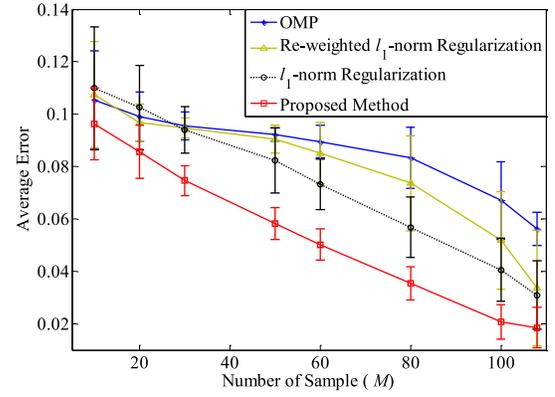


Fig. 10. Average prediction error (both the average error and the standard deviation) of different algorithms is estimated by 100 repeated runs. It suggests that the proposed HMT method could largely reduce the prediction error, especially when the number of samples M ranges from 30 to 80 in this example.

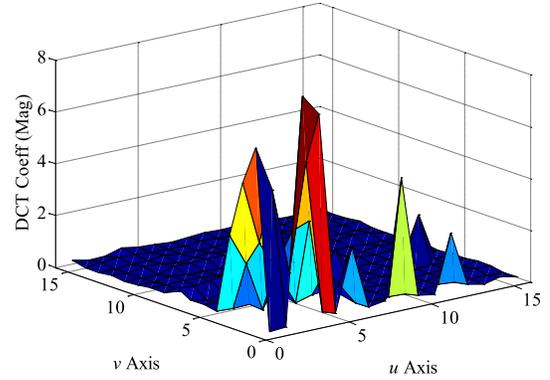


Fig. 11. DCT coefficients (magnitude) of the power show a unique pattern that is approximately sparse and correlated.

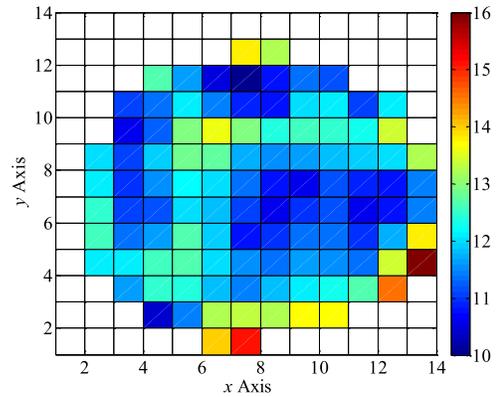


Fig. 12. Power value predicted from 40 tested chips by the proposed HMT method.

This metric quantitatively measures the difference between the measurement data (i.e., Fig. 6) and prediction results (i.e., Fig. 12) for every chip. Fig. 13 shows the histogram of the relative error calculated for all the chips on the same wafer with 40 tested chips. Note that the relative error is less than 10% for most chips in this example.

For the modeling cost of spatial variations, compared with the coefficient computation in (5), the silicon area overhead

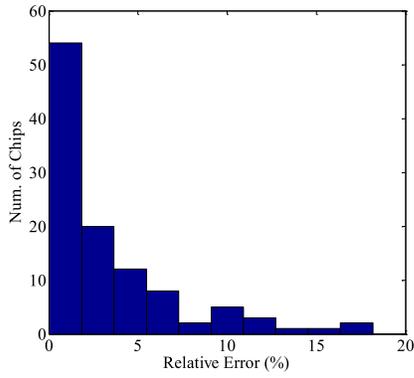


Fig. 13. Histogram of the relative errors \bar{E}_{REL} calculated by (34) for all chips in the same wafer by applying the proposed HMT method.

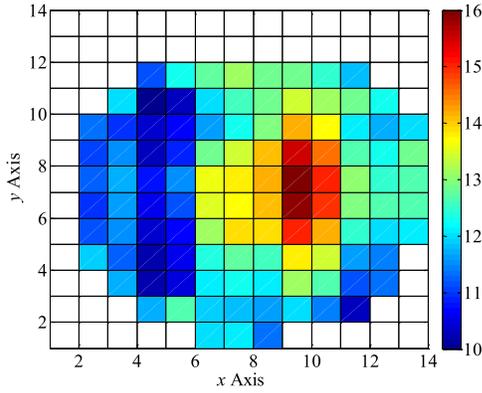


Fig. 14. Measurement frequency values of 109 industrial chips from the same wafer show significant spatial variations.

and the testing time required to generate sampling points is much more expensive. For instance, today's advanced processor chip typically applies hundreds of on-chip ROs to monitor process variations [11] and the reliability test of one chip often consumes days of time. On the contrary, for the calculation of the modeling coefficients, our proposed method takes averagely 100 s, l_1 -norm regularization (i.e., the existing VP method [19]) takes averagely 480 s (because cross-validation is utilized to screen out high frequency components), reweighted l_1 -norm regularization takes more than 2500 s (because an iterative scheme is applied to enhance sparsity further), and OMP only takes about 5 s (because it only uses simple inner-product computations to select important basis function). Therefore, the number of sampling points could be taken as a metric for cost comparison. Note that as shown in Fig. 10, in order to model the power distribution with a given error specification $E_{AVG} = 0.06$ in our case, the number of sampling points required for the proposed method is around 50, while the smallest number for other approaches is around 85. Namely, compared to the existing approaches, the proposed method could achieve up to 75% cost reduction without any surrender of modeling accuracy.

B. Frequency Measurement Data

We consider the frequency of ROs collected from the same wafer for the same circuit design at this section. These RO

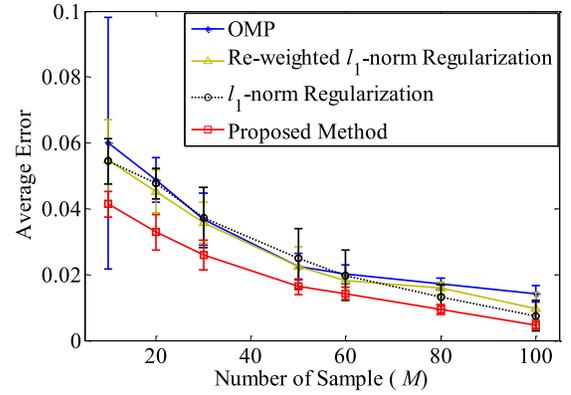


Fig. 15. Average prediction error (both average error and standard deviation) of different algorithms is estimated by 100 repeated runs.

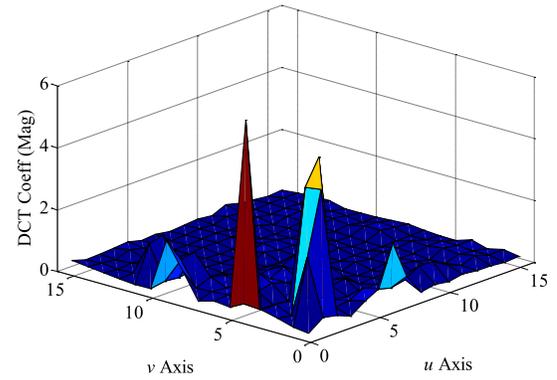


Fig. 16. DCT coefficients (magnitude) of the frequency show a unique pattern that is sparse and correlated.

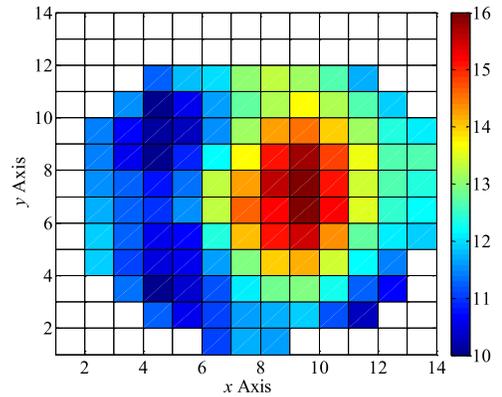


Fig. 17. Frequency value predicted from 40 tested powers by the proposed HMT algorithm.

data are of great significance for process monitoring and control because they are strongly correlated with the final chip performance [11], [14]. Fig. 14 shows the measurement data of frequency as a function of location (x, y) .

We apply different algorithms to predict the spatial variation based on different number of sampling data. Fig. 15 shows the average error calculated by (33). Similar to the previous example, the average errors are calculated from 100 repeated runs. Note that for this example, the proposed HMT method also achieves better accuracy than other three traditional techniques in general.

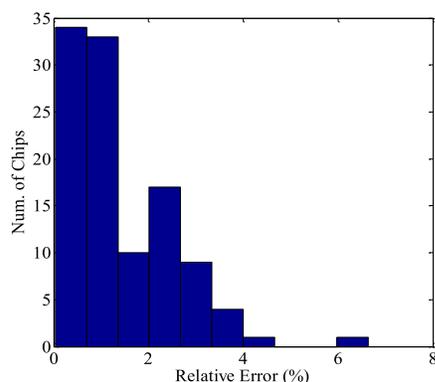


Fig. 18. Histogram of the relative errors of the proposed HMT method calculated by (34) for all chips in the same wafer.

Fig. 16 gives DCT coefficients after applying the proposed method in one run. Similar to the previous example, some high-frequency components and the correlations of DCT coefficients can also be observed. However, compared with the dynamic power modeling, whose coefficients for one run are shown in Fig. 11, the frequency components of this case are obviously further sparser. Therefore, the modeling errors of OMP, reweighted l_1 -norm regularization and l_1 -norm regularization are almost the same as shown in Fig. 15.

Fig. 17 shows the frequency of RO predicted from 40 examples by our proposed HMT method. Fig. 18 further shows the histogram of the relative error calculated for all chips using (34). The relative error is less than 5% for most chips in this example.

VI. CONCLUSION

In this paper, we propose a novel method based on HMT for efficient spatial variation modeling. Applying the proposed method, HMT is introduced to model DCT coefficients accurately by covering both the sparse structure and the correlation properties for DCT transform in frequency domain. MAP estimation is used to reformulate the spatial variation modeling as a typical convex optimization problem. Numerical results demonstrate that the proposed HMT method could achieve up to 40% accuracy improvement at the same low measurement cost compared with the existing approaches including OMP, l_1 -norm regularization, and reweighted l_1 -norm regularization.

Furthermore, though EZT is utilized in this paper to exploit the correlations between DCT coefficients, other correlation models may be applied for spatial variation modeling and analysis. In our future research, we will further study the modeling accuracy and computational cost of different correlation models.

REFERENCES

- [1] Semiconductor Industry Associate. (2013). *International Technology Roadmap for Semiconductors*. [Online]. Available: <http://www.itrs.net/Links/2013ITRS/Home2013.htm>
- [2] S. Borkar *et al.*, "Parameter variations and impact on circuits and microarchitecture," in *Proc. IEEE Design Autom. Conf.*, Anaheim, CA, USA, Jun. 2003, pp. 338–342.

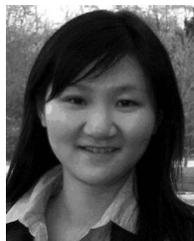
- [3] H. Chang and S. Sapatnekar, "Statistical timing analysis under spatial correlations," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 9, pp. 1467–1482, Sep. 2005.
- [4] C. Visweswariah, K. Ravindran, K. Kalafala, S. Walker, and S. Narayan, "First-order incremental block-based statistical timing analysis," in *Proc. IEEE Design Autom. Conf.*, San Diego, CA, USA, Jul. 2004, pp. 331–336.
- [5] Y. Zhan *et al.*, "Correlation aware statistical timing analysis with non-Gaussian delay distributions," in *Proc. IEEE Design Autom. Conf.*, Anaheim, CA, USA, Jun. 2005, pp. 77–82.
- [6] K. Heloue and F. Najm, "Statistical timing analysis with two-sided constraints," in *Proc. IEEE Int. Conf. Comput. Aided Design*, San Jose, CA, USA, Nov. 2005, pp. 829–836.
- [7] M. Mani, A. Singh, and M. Orshansky, "Joint design-time and post-silicon minimization of parametric yield loss using adjustable robust optimization," in *Proc. IEEE Int. Conf. Comput. Aided Design*, San Jose, CA, USA, Nov. 2006, pp. 19–26.
- [8] S. Kulkarni, D. Sylvester, and D. Blaauw, "A statistical framework for post-silicon tuning through body bias clustering," in *Proc. IEEE Int. Conf. Comput. Aided Design*, San Jose, CA, USA, Nov. 2006, pp. 39–46.
- [9] Q. Liu and S. Sapatnekar, "Synthesizing a representative critical path for post-silicon delay prediction," in *Proc. IEEE Int. Symp. Phys. Design*, San Diego, CA, USA, Mar./Apr. 2009, pp. 183–190.
- [10] P. Girard, "Survey of low-power testing of VLSI circuits," *IEEE Des. Test Comput.*, vol. 19, no. 3, pp. 80–90, May/Jun. 2002.
- [11] M. Bhushan, A. Gattiker, M. Ketchen, and K. Das, "Ring oscillators for CMOS process tuning and variability control," *IEEE Trans. Semicond. Manuf.*, vol. 19, no. 1, pp. 10–18, Feb. 2006.
- [12] W. R. Mann, F. L. Taber, P. W. Seitzer, and J. J. Broz, "The leading edge of production wafer probe test technology," in *Proc. IEEE Int. Test Conf.*, Charlotte, NC, USA, Oct. 2004, pp. 1168–1195.
- [13] M. Bushnell and V. Agrawal, *Essentials of Electronic Testing for Digital, Memory, and Mixed-Signal VLSI Circuits*. Norwell, MA, USA: Kluwer, 2000.
- [14] X. Li, R. Rutenbar, and R. Blanton, "Virtual probe: A statistically optimal framework for minimum-cost silicon characterization of nanoscale integrated circuits," in *Proc. IEEE Int. Conf. Comput. Aided Design*, San Jose, CA, USA, Nov. 2009, pp. 433–440.
- [15] W. Zhang, X. Li, and R. Rutenbar, "Bayesian virtual probe: Minimizing variation characterization cost for nanoscale IC technologies via Bayesian inference," in *Proc. IEEE Design Autom. Conf.*, Anaheim, CA, USA, Jun. 2010, pp. 262–267.
- [16] W. Zhang, X. Li, E. Acar, F. Liu, and R. Rutenbar, "Multi-wafer virtual probe: Minimum-cost variation characterization by exploring wafer-to-wafer correlation," in *Proc. IEEE Int. Conf. Comput. Aided Design*, San Jose, CA, USA, Nov. 2010, pp. 47–54.
- [17] H. Chang, K. Cheng, W. Zhang, X. Li, and K. Butler, "Test cost reduction through performance prediction using virtual probe," in *Proc. IEEE Int. Test Conf.*, Anaheim, CA, USA, Sep. 2011, pp. 1–9.
- [18] W. Zhang, K. Balakrishnan, X. Li, D. Boning, and R. Rutenbar, "Toward efficient spatial variation decomposition via sparse regression," in *Proc. IEEE Int. Conf. Comput. Aided Design*, San Jose, CA, USA, Nov. 2011, pp. 162–169.
- [19] W. Zhang *et al.*, "Virtual probe: A statistical framework for low-cost silicon characterization of nanoscale integrated circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 30, no. 12, pp. 1814–1827, Dec. 2011.
- [20] W. Zhang, X. Li, S. Saxena, A. Strojwas, and R. Rutenbar, "Automatic clustering of wafer spatial signatures," in *Proc. IEEE Design Autom. Conf.*, Austin, TX, USA, May/Jun. 2013, pp. 1–6.
- [21] W. Zhang *et al.*, "Efficient spatial pattern analysis for variation decomposition via robust sparse regression," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 32, no. 7, pp. 1072–1085, Jul. 2013.
- [22] H. Goncalves *et al.*, "A fast spatial variation modeling algorithm for efficient test cost reduction of analog/RF circuits," in *Proc. IEEE Design Autom. Test Eur. Conf.*, Grenoble, France, Mar. 2015, pp. 1042–1047.
- [23] N. Kupp, K. Huang, J. Carulli, and Y. Makris, "Spatial estimation of wafer measurement parameters using Gaussian process models," in *Proc. IEEE Int. Test Conf.*, Anaheim, CA, USA, Nov. 2012, pp. 1–8.
- [24] N. Kupp, K. Huang, J. Carulli, and Y. Makris, "Spatial correlation modeling for probe test cost reduction in RF devices," in *Proc. IEEE Int. Conf. Comput.-Aided Design*, San Jose, CA, USA, Nov. 2012, pp. 23–29.
- [25] K. Huang, N. Kupp, J. Carulli, and Y. Makris, "Handling discontinuous effects in modeling spatial correlation of wafer-level analog/RF tests," in *Proc. IEEE Design Autom. Test Eur. Conf.*, Grenoble, France, Mar. 2013, pp. 553–558.

- [26] K. Huang, N. Kupp, J. Carulli, and Y. Makris, "On combining alternate test with spatial correlation modeling in analog/RF ICs," in *Proc. IEEE Eur. Test Symp.*, Avignon, France, May 2013, pp. 64–69.
- [27] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.
- [28] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [29] E. J. Candes, "Compressive sampling," in *Proc. Int. Congr. Math.*, vol. 3, Madrid, Spain, 2006, pp. 1433–1452.
- [30] J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.
- [31] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [32] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
- [33] H. Choi, J. Romberg, R. Baraniuk, and N. Kingsbury, "Hidden Markov tree modeling of complex wavelet transforms," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Istanbul, Turkey, Jun. 2000, pp. 133–136.
- [34] C. Wu and W. Hsieh, "Digital watermarking using zerotree of DCT," *IEEE Trans. Consum. Electron.*, vol. 46, no. 1, pp. 87–94, Feb. 2000.
- [35] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [36] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [37] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 877–905, Dec. 2008.
- [38] Z. Xiong, O. G. Guleryuz, and M. T. Orchard, "A DCT-based embedded image coder," *IEEE Signal Process. Lett.*, vol. 3, no. 11, pp. 289–290, Nov. 1996.
- [39] Z. Xiong, K. Ramchandran, M. Orchard, and Y. Zhang, "A comparative study of DCT- and wavelet-based image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 5, pp. 692–695, Aug. 1999.
- [40] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.
- [41] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
- [42] E. C. Reed and J. S. Lim, "Efficient coding of DCT coefficients by joint position-dependent encoding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Seattle, WA, USA, May 1998, pp. 2817–2820.
- [43] A. Nowroz, R. Cochran, and S. Reda, "Thermal monitoring of real processors: Techniques for sensor allocation and full characterization," in *Proc. IEEE Design Autom. Conf.*, Anaheim, CA, USA, Jun. 2010, pp. 56–61.
- [44] S. Reda and S. Nassif, "Analyzing the impact of process variations on parametric measurements: Novel models and applications," in *Proc. IEEE Design Autom. Test Eur. Conf.*, Nice, France, Apr. 2009, pp. 375–380.
- [45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [46] C. Bishop, *Pattern Recognition and Machine Learning*. Upper Saddle River, NJ, USA: Prentice-Hall, 2007.



Changhai Liao received the B.S. degree in microelectronics from the School of Microelectronics and Solid-State Electronics, University of Electronic Science and Technology of China, Chengdu, China, in 2013. He is currently pursuing the Ph.D. degree in microelectronics and solid-state electronics with Fudan University, Shanghai, China.

His current research interests include statistical methods for computer-aided design under process variation and design automation for analog/mixed-signal integrated circuits.



Jun Tao (M'07) received the B.S. and Ph.D. degrees in electrical engineering from Fudan University, Shanghai, China, in 2002 and 2007, respectively.

Since 2007, she has been an Assistant Professor with the Department of Microelectronics, Fudan University. She was a Visiting Scholar with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA, in 2012. Her current research interests include modeling, simulation, and design for analog and mixed-signal systems.



Xuan Zeng (M'97) received the B.S. and Ph.D. degrees in electrical engineering from Fudan University, Shanghai, China, in 1991 and 1997, respectively.

She is currently a Full Professor with the Department of Microelectronics, Fudan University. She was a Visiting Professor with the Department of Electrical Engineering, Texas A&M University, College Station, TX, USA, and the Department of Microelectronics, Technische Universiteit Delft, Delft, The Netherlands, in 2002 and 2003, respectively. Her current research interests include design for manufacturability, high-speed interconnect analysis and optimization, and circuit simulation.

Dr. Zeng was a recipient of the Chinese National Science Funds for Distinguished Young Scientists in 2011, the First-Class of Natural Science Prize of Shanghai in 2012, and the Changjiang Distinguished Professor from the Ministry of Education Department of China in 2014.



Yangfeng Su received the B.S. and Ph.D. degrees in computational mathematics from Fudan University, Shanghai, China, in 1986 and 1992, respectively.

He is a Full Professor with the School of Mathematical Sciences, Fudan University. His current research interests include model order reduction, linear and nonlinear eigenvalue problems, and sparse expression and its applications.



Dian Zhou (M'89–SM'07) received the B.S. and M.S. degrees in physics from Fudan University, Shanghai, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 1990.

He is currently a Professor with the University of Texas at Dallas, Richardson, TX, USA.

Prof. Zhou was a recipient of the IEEE Circuits and Systems Society Darlington Award in 1993, the National Science Foundation (NSF) Young Investigator Award in 1994, the Outstanding Overseas Young Investigator Award from the NSF of China in 2000, the Changjiang Honor Professor from the Ministry of Education Department of China in 2002, and the Grand Award HDTV Terrestrial Broadcasting ASIC Chip: Chinese Television Number One from the Electronics Society of China in 2005. He was selected as the Thousand Talents Plan Professor from China in 2010.



Xin Li (S'01–M'06–SM'10) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2005.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Carnegie Mellon University. His current research interests include integrated circuit and signal processing.

Dr. Li was a recipient of the National Science Foundation Faculty Early Career Development Award in 2012, the IEEE Donald O. Pederson Best Paper Award in 2013, the Best Paper Award from Design Automation Conference in 2010, two IEEE/ACM William J. McCalla International Conference on Computer-Aided Design Best Paper Awards in 2004 and 2011, and the Best Paper Award from the International Symposium on Integrated Circuits in 2014.