View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DSpace@MIT

Spatial Variation Decomposition via Sparse Regression

Wangyang Zhang[1], Karthik Balakrishnan[3], Xin Li[1], Duane Boning[2], Emrah Acar[3], Frank Liu[4] and Rob A. Rutenbar[5]
[1]Carnegie Mellon University, Pittsburgh, PA 15213, wangyan1@ece.cmu.edu, xinli@ece.cmu.edu
[2]Massachusetts Institute of Technology, Cambridge, MA 02139, boning@mtl.mit.edu
[3]IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, kbalakr@us.ibm.com, emrah@us.ibm.com
[4]IBM Austin Research Lab, Austin, TX 78758, frankliu@us.ibm.com
[5]University of Illinois at Urbana-Champaign, Urbana, IL 61801, rutenbar@illinois.edu

***Abstract*—In this paper, we briefly discuss the recent development of a novel sparse regression technique that aims to accurately decompose process variation into two different components: (1) spatially correlated variation, and (2) uncorrelated random variation. Such variation decomposition is important to identify systematic variation patterns at wafer and/or chip level for process modeling, control and diagnosis. We demonstrate that the spatially correlated variation can be accurately represented by the linear combination of a small number of "templates". Based upon this observation, an efficient algorithm is developed to accurately separate spatially correlated variation from uncorrelated random variation. Several examples based on silicon measurement data demonstrate that the aforementioned sparse regression technique can capture systematic variation patterns with high accuracy.**

Keywords—process variation; integrated circuit; variation decomposition

I. INTRODUCTION

With the continued scaling of CMOS technology, process variation has become a critical issue for design and manufacture of integrated circuits [1]-[2]. Large-scale performance variability has been observed for integrated circuits fabricated at advanced technology nodes, resulting in significant parametric yield loss. For this reason, accurate process characterization and modeling is required in order to fully understand the variation sources [10].

Towards this goal, identifying and modeling systematic variation is of great importance. Once the systematic variation sources are found, it is possible to optimize the manufacturing process and/or modify the circuit design to improve yield. It has been demonstrated in the literature that systematic variation often presents a unique spatial pattern [2]. Namely, systematic variation is spatially correlated. For example, it has been observed in [3] that the spatial correlation in gate length is partially caused by the systematic variation due to lithography. Motivated by these observations, several prior works [2]-[3] uncover the systematic variation pattern by modeling the spatially correlated variation with a small number of pre-determined "templates" (e.g., linear and quadratic functions). However, the most appropriate templates to model the spatial variation change from process to process. Moreover, if the number of templates is too small, the spatially correlated variation cannot be modeled accurately. On the other hand, too many templates also lead to inaccuracy caused by the over-fitting problem [13].

Motivated by these observations, we developed a novel sparse regression technique to accurately model spatially correlated variation and separate it from uncorrelated random variation [4]. To apply sparse regression, only a *dictionary* of templates is needed, which includes all possible patterns of systematic variation. The optimal templates to model the systematic variation of a given wafer/die will be automatically selected by the sparse regression algorithm. The dictionary can be customized by process engineers based on their experience. When such custom dictionaries are not available, a general dictionary containing Discrete Cosine Transform (DCT) [11] functions can be applied. It has been observed that spatially correlated variation typically carries a unique sparse structure in frequency domain [4]-[7], implying that it can be accurately represented by a small number of dominant DCT coefficients. On the other hand, uncorrelated random variation has a white frequency spectrum and the corresponding DCT coefficients are evenly distributed over all frequencies. Variation decomposition can be accurately performed by exploring this unique sparsity in frequency domain.

The proposed variation decomposition technique has been validated by using two sets of silicon measurement data. The first data set contains within-chip measurements of contact plug resistance and the second data set contains within-wafer measurements of ring oscillator period. In both examples the sparse regression method captures the systematic variation pattern with high accuracy.

The remainder of this paper is organized as follows. In Section II, we first derive the mathematical formulation for the proposed variation decomposition problem and then describe the sparse regression algorithm in Section III. The efficacy of sparse regression is demonstrated by two examples in Section IV. Finally, we conclude in Section V.

II. VARIATION DECOMPOSITION

Let $g(x, y)$ be a two-dimensional function representing the spatial variation of interest, where x and y denote the coordinate of a spatial location within the two-dimensional plane. The spatial variation g can be the device-level threshold voltage variation within a chip, chip-level leakage current variation on a wafer, etc. In this paper, due to the space constraint, we only show the core idea of our algorithm by assuming that the spatial function $g(x, y)$ is measured on one chip or wafer. An extended version of the algorithm to model

This work was supported in part by the C2S2 Focus Center and the Interconnect Focus Center, two of six research centers funded under the Focus Center Research Program (FCRP), a Semiconductor Research Corporation entity. This work was also supported in part by the National Science Foundation under contract CCF–0915912.

multiple chips and/or wafers is described in [4]. Mathematically, we aim to decompose the spatial variation function $g(x, y)$ into two different components:

$$g(x,y) = s(x,y) + r(x,y) \quad (1)$$

where $s(x, y)$ and $r(x, y)$ stand for the spatially correlated variation and the uncorrelated random variation, respectively.

To perform variation decomposition, we model the spatial variation function as follows:

$$g(x,y) = \sum_{j=1}^{M} \eta_j \cdot A_j(x,y) + r(x,y) \quad (2)$$

where the spatially correlated variation $s(x, y)$ is represented by a dictionary of basis functions $\{A_j(x, y); j = 1, 2, \ldots, M\}$ through coefficients $\{\eta_j; j = 1, 2, \ldots, M\}$. Each basis function can be viewed as a particular "template" to model the spatially correlated variation. The dictionary only needs to be designed to include templates for all possible systematic variation patterns. The relevant templates will be automatically selected by a *sparse regression* algorithm, which will be introduced in the next section. Once the spatially correlated variation is accurately determined, the uncorrelated random variation $r(x, y)$ can be represented by the residual.

In practice, the dictionary of templates for sparse regression can be customized by process engineers based on their experience. When such custom dictionaries are not available, a general dictionary containing Discrete Cosine Transform (DCT) [11] functions can be applied. To introduce the DCT functions, we note that the spatial variation g is measured at a finite number of spatial locations. Therefore, without loss of generality, the spatial coordinates x and y can be labeled as integer numbers: $x \in \{1, 2, ..., P\}$ and $y \in \{1, 2, ..., Q\}$, as shown in [5]-[7]. The DCT functions can then be defined as:

$$A_{u,v}(x,y) = \alpha_u \cdot \beta_v \cdot \cos\frac{\pi(2x-1)(u-1)}{2 \cdot P} \cdot \cos\frac{\pi(2y-1)(v-1)}{2 \cdot Q} \quad (u = 1,2,\cdots,P; v = 1,2,\cdots,Q) \quad (3)$$

where

$$\alpha_u = \begin{cases} \sqrt{1/P} & (u = 1) \\ \sqrt{2/P} & (2 \le u \le P) \end{cases} \quad (4)$$

$$\beta_v = \begin{cases} \sqrt{1/Q} & (v = 1) \\ \sqrt{2/Q} & (2 \le v \le Q) \end{cases}. \quad (5)$$

The DCT coefficients $\{\eta(u, v); u = 1, 2, ..., P; v = 1, 2, ..., Q\}$ in (2) represent the frequency-domain components of the spatial variation function $\{g(x, y); x = 1, 2, ..., P; y = 1, 2, ..., Q\}$. An important property of the DCT coefficients is that if the spatial variation exhibits a spatially correlated pattern, a vast majority of the DCT coefficients are close to 0, and therefore the spatial pattern can be accurately represented by a small number of large DCT coefficients. This unique property of sparseness has been observed in many image processing tasks and serves as a key component for the compression algorithm of JPEG [11]. On the other hand, uncorrelated random variation can be characterized as white noise [12] and evenly distributed among all frequencies. Therefore, the corresponding DCT coefficients are relatively small. These properties, in turn, demonstrate that spatially correlated variation can be accurately represented by a small number of dominant DCT coefficients.

III. SPARSE REGRESSION ALGORITHM

A. Orthogonal Matching Pursuit

The objective of sparse regression is to determine a small number of templates to approximate the spatially correlated variation in (2). Mathematically, the sparse regression problem can be formulated as the following optimization:

$$\begin{aligned} \underset{\eta}{\text{minimize}} \quad & \|A \cdot \eta - B\|_2^2 \\ \text{subject to} \quad & \|\eta\|_0 \le \lambda \end{aligned} \quad (6)$$

where the function $g(x, y)$ is measured at N different spatial locations, and $B = [b_1\ b_2\ \ldots\ b_N]^T$ is a vector of these measurements. A is matrix where A_{ij} represents the value of $A_j(x, y)$ at the ith measurement location, and $\eta = [\eta_1\ \eta_2\ \ldots\ \eta_M]^T$ is a vector of unknown coefficients. The symbols $\|\bullet\|_2$ and $\|\bullet\|_0$ stand for the L_2-norm (i.e., the square root of the summation of the squares of all elements) and the L_0-norm (i.e., the number of non-zero elements) of a vector respectively. The optimization in (6) attempts to use a small number of (i.e., λ) dominant templates to approximate the measurement data B with least-squares error.

In general, solving the optimization in (6) is not trivial, since the problem is NP-hard. Orthogonal matching pursuit (OMP) [9] is an efficient greedy algorithm to approximate the solution of (6). In what follows, we briefly review the major steps of the OMP algorithm. More details on OMP can be found in [9].

The key idea of OMP is to iteratively use the inner product to identify a small number of important templates. Towards this goal, we re-write the matrix A by its column vectors:

$$A = [A_1 \quad A_2 \quad \cdots \quad A_M] \quad (7)$$

where each column vector A_j can be conceptually viewed as a basis vector associated with the template $A_j(x, y)$. The inner product $\langle B, A_j \rangle$ measures the "correlation" between the measurement data B and the basis vector A_j. A strong correlation between B and A_j implies that the template $A_j(x, y)$ is an important component to approximate the spatially correlated variation $s(x, y)$.

Based on this idea, OMP applies an iterative process to find a set of important templates, as summarized in Algorithm 1. At each iteration, OMP performs two major operations. First, it selects the basis vector A_s that is most "correlated" to the residual *Res*. Second, the coefficients associated with all selected basis vectors are solved by least-squares fitting. It should be noted that Algorithm 1 relies on a given input parameter λ. In practice, the value of λ is not known in advance. However, it can be accurately estimated by cross-validation, as will be discussed in detail in the next sub-section.

Algorithm 1: Orthogonal Matching Pursuit (OMP)

1. Start from the optimization problem in (6) with a given integer λ specifying the total number of basis vectors.
2. Initialize the residual $Res = B$, the set $\Omega = \{\}$, and the iteration index $p = 1$.

3. Select the new basis vector A_s according to the following criterion:

$$\underset{s}{\text{maximize}} \quad \left|\left\langle Res, A_s \right\rangle\right|. \tag{8}$$

4. Update Ω by $\Omega = \Omega \cup \{s\}$.
5. Solve the least-squares fitting problem:

$$\underset{\eta, i\in\Omega}{\text{minimize}} \quad \left\| \sum_{i\in\Omega} A_i \cdot \eta_i - B \right\|_2^2 . \tag{9}$$

6. Calculate the residual:

$$Res = B - \sum_{i\in\Omega} A_i \cdot \eta_i . \tag{10}$$

7. If $p < \lambda$, $p = p + 1$ and go to Step 3.
8. For any $i \notin \Omega$, set $\eta_i = 0$.

B. *Cross-Validation*

The OMP algorithm (i.e., Algorithm 1) relies on a user defined parameter λ to control the number of basis vectors that should be selected. In practice, λ is not known in advance. The appropriate value of λ must be determined by considering the following two important issues. First, if λ is too small, OMP cannot select a sufficient number of basis vectors to represent the spatially correlated variation, thereby leading to large modeling error. On the other hand, if λ is too large, OMP can incorrectly select too many basis vectors and some of the corresponding coefficients are associated with uncorrelated random variation, instead of spatially correlated systematic variation. It, again, results in large modeling error due to over-fitting. In order to achieve the best accuracy, we must accurately estimate the modeling error for different λ values and then find the optimal λ with minimum error.

We adopt the cross-validation method [13] to estimate the modeling error for our variation decomposition application. An F-fold cross-validation partitions the entire data set into F groups. Modeling error is estimated according to the cost function in (6) from F independent runs. In each run, one of the F groups is used to estimate the modeling error and all other groups are used to calculate the coefficients. Note that the training data for coefficient estimation and the testing data for error estimation are not overlapped. Hence, over-fitting can be easily detected. In addition, different groups should be selected for error estimation in different runs. As such, each run results in an error value ε_f (f = 1, 2, ..., F) that is measured from a unique group of data points. The final modeling error is computed as the average of $\{\varepsilon_f; f = 1, 2, ..., F\}$, i.e., $\varepsilon = (\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_F) / F$.

IV. NUMERICAL EXAMPLES

In this section, we demonstrate the efficacy of the sparse regression algorithm in variation decomposition using several examples. All numerical experiments are performed on a 2.8GHz Linux server.

A. *Measurement Data for Contact Plug Resistance*

We consider the contact plug resistance measurement data from a test chip in a 90 nm CMOS process. The chip contains 36,864 test structures (i.e., contacts) arranged as a 144×256 array, as described in [8].

Figure 1 (a) shows the measured contact plug resistance (normalized) from the test chip. Studying Figure 1 (a), we would notice that there is a unique spatial pattern due to layout dependency. However, the spatial pattern is not clearly visible because of the large-scale uncorrelated random variation found in this example. Figure 1 (b) further shows the DCT coefficients (magnitude) of the measured contact plug resistance. Note that there only exist a small number of dominant DCT coefficients with large magnitude. These DCT coefficients are distributed over a small number of frequencies, representing a unique signature of the layout-dependent systematic variation in frequency domain. All other DCT coefficients are small in magnitude and have a white frequency spectrum (i.e., evenly distributed over all frequencies). They correspond to the uncorrelated random variation that we observe from Figure 1 (a). These observations demonstrate the important fact that the spatially correlated systematic variation can be extracted by identifying the dominant DCT coefficients in frequency domain.

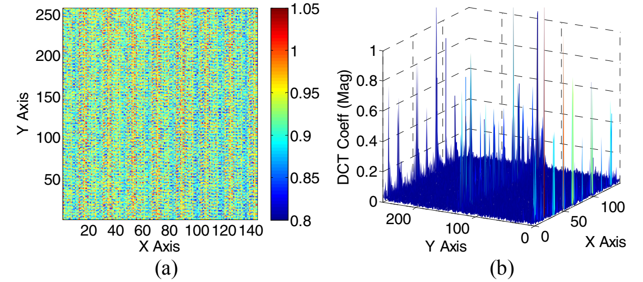

Figure 1. (a) Measured contact plug resistance (normalized) of a 144×256 array. (b) Discrete cosine transform (DCT) coefficients (magnitude) of the measured contact plug resistance.

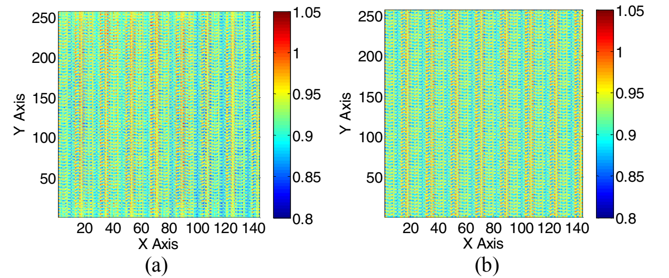

Figure 2. (a) Extracted layout-dependent systematic variation (normalized) of contact plug resistance. (b) Spatial distribution of different contact layout patterns in the test chip.

We apply sparse regression to extract the layout-dependent systematic variation for the test chip. The extracted systematic variation is shown in Figure 2 (a). Comparing Figure 2 (a) with Figure 1 (a), we would notice that the spatial pattern of systematic variation becomes clear, after sparse regression is applied. Such a spatial variation pattern can serve as an important basis for diagnosing the sources of systematic variation. In this example, the systematic variation is mainly caused by different layout patterns regularly distributed over the entire chip. To verify the layout dependency, we plot the spatial distribution of different layout patterns in Figure 2 (b) where there exist 55 layout patterns in total and different layout patterns are shown in different colors. Note that Figure 2 (b) perfectly matches Figure 2 (a). It, in turn, demonstrates that the

aforementioned layout dependency is the dominant source for the extracted systematic variation in Figure 2 (a).

We further estimate the percentage of systematic variation compared to random variation by calculating the variance of spatially correlated and uncorrelated random components from sparse regression. Our analysis indicates that 68.8% of the spatial variation is systematic and the other 31.2% is random. This result shows that the layout-dependent systematic variation is the dominant variation source. On the other hand, the spatial variation pattern in Figure 2 cannot be captured by simple templates such as linear and quadratic functions. If these functions are used, it will lead to the incorrect result that 99.1% of the spatial variation is random.

B. Measurement Data for Ring Oscillator Period

We consider the ring oscillator (RO) period measurement data from one wafer at an advanced technology node. The wafer contains 117 ROs distributed over different spatial locations.

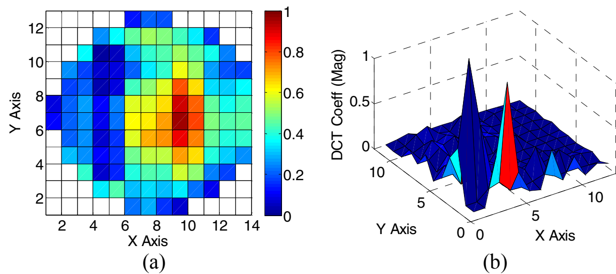

Figure 3. (a) Measured RO period (normalized) of 117 ROs. (b) DCT coefficients (magnitude) of the measured RO period.

Figure 3 shows the measured RO period and the magnitude of DCT coefficients. From Figure 3 (a) it can be intuitively seen that the wafer already has a clear pattern and from Figure 3 (b) it can be seen that the DCT coefficients are relatively sparse. These observations indicate that the spatially correlated variation is strongly dominant for this wafer.

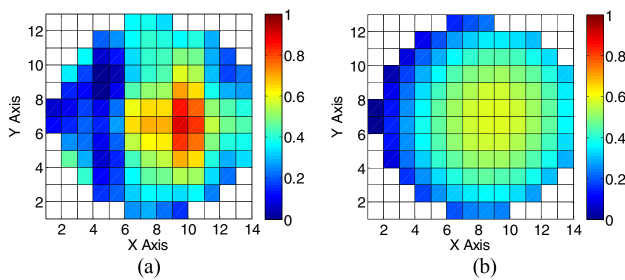

Figure 4. (a) Extracted spatially correlated systematic variation (normalized) of RO period by sparse regression. (b) Extracted spatially correlated systematic variation by using linear and quadratic templates.

The extracted systematic variation by sparse regression is shown in Figure 4 (a). Note that Figure 4 (a) is not significantly different from Figure 3 (a). By inspecting a number of other wafers for the same product, we found that a spatial pattern close to Figure 4 (a) can indeed be observed on all wafers. Based on Figure 4 (a), we estimate that the systematic variation is 89.3% and the random variation is 10.7%. If we extract the spatially correlated variation using linear and quadratic templates, the result is shown in Figure 4 (b), which clearly does not fully capture the spatial pattern. Additional analysis of Figure 4 (b) shows that 48.2% of the spatial variation is systematic and the other 51.8% is random. This result will lead to the incorrect conclusion that random variation is dominant and it may result in ineffective efforts in yield improvement.

V. CONCLUSIONS

In this paper, we discuss the recent development of a novel sparse regression technique that accurately decomposes process variation into spatially correlated variation and uncorrelated random variation. The technique is based upon the fact that the spatially correlated variation can be accurately represented by the linear combination of a small number of "templates". An efficient OMP algorithm is borrowed from the statistics community to accurately find these templates from a large dictionary. Our experimental results on silicon measurement data demonstrate that sparse regression can accurately capture the true systematic pattern. Such a goal cannot be achieved by using a small number of simple templates such as linear and quadratic functions.

REFERENCES

[1] Semiconductor Industry Associate, *International Technology Roadmap for Semiconductors*, 2009.

[2] A. Gattiker, "Unraveling variability for process/product improvement," *IEEE ITC*, pp. 1-9, 2008.

[3] P. Friedberg, Y. Cao, J. Cain, R. Wang, J. Rabaey, and C. Spanos, "Modeling within-field gate length spatial variation for process-design co-optimization," *Proceedings of SPIE*, vol. 5756, pp. 178-188, May. 2005.

[4] W. Zhang, K. Balakrishnan, Xin Li, D. Boning, and R. Rutenbar, "Toward efficient spatial variation decomposition via sparse regression," *IEEE ICCAD*, pp. 162-169, 2011.

[5] X. Li, R. Rutenbar and R. Blanton, "Virtual probe: a statistically optimal framework for minimum-cost silicon characterization of nanoscale integrated circuits," *IEEE ICCAD*, pp. 433-440, 2009.

[6] W. Zhang, X. Li and R. Rutenbar, "Bayesian virtual probe: minimizing variation characterization cost for nanoscale IC technologies via Bayesian inference," *IEEE DAC*, pp. 262-267, 2010.

[7] 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," *IEEE ICCAD*, pp. 47-54, 2010.

[8] K. Balakrishnan and D. Boning, "Measurement and analysis of contact plug resistance variability," *IEEE CICC*, pp. 416-422, 2009.

[9] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[10] M. Orshansky, S. Nassif, and D. Boning, *Design for Manufacturability and Statistical Design: A Constructive Approach*, Springer, 2007.

[11] R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice Hall, 2007.

[12] A. Oppenheim, *Signals and Systems*, Prentice Hall, 1996.

[13] C. Bishop, *Pattern Recognition and Machine Learning*, Prentice Hall, 2007.