

Multi-Wafer Virtual Probe: Minimum-Cost Variation Characterization by Exploring Wafer-to-Wafer Correlation

Wangyang Zhang¹, Xin Li¹, Emrah Acar², Frank Liu³ and Rob Rutenbar⁴

¹Carnegie Mellon University, Pittsburgh, PA 15213

²IBM T.J. Watson Research Center, Yorktown Heights, NY 10598

³IBM Austin Research Lab, Austin, TX 78758

⁴University of Illinois at Urbana-Champaign, Urbana, IL 61801

ABSTRACT

In this paper, we propose a new technique, referred to as *Multi-Wafer Virtual Probe* (MVP) to efficiently model wafer-level spatial variations for nanoscale integrated circuits. Towards this goal, a novel Bayesian inference is derived to extract a shared model template to explore the wafer-to-wafer correlation information within the same lot. In addition, a robust regression algorithm is proposed to automatically detect and remove outliers (i.e., abnormal measurement data with large error) so that they do not bias the modeling results. The proposed MVP method is extensively tested for silicon measurement data collected from 200 wafers at an advanced technology node. Our experimental results demonstrate that MVP offers superior accuracy over other traditional approaches such as VP [7] and EM [8], if a limited number of measurement data are available.

1. INTRODUCTION

Parametric variations such as gate length variations, dopant fluctuations and metal thickness variations widely exist in today's manufacturing process and will continue to increase in the future [1]-[2]. These variations manifest themselves in different scales, including lot-to-lot, wafer-to-wafer, within-wafer and within-reticle/die, due to both systematic and random effects [14]. A lot of IC analysis and optimization techniques (e.g. statistical timing analysis [4], post-silicon tuning [5], etc.) have been proposed in recent years to combat parametric variations. These methods heavily rely on the accuracy of the variation models that capture the corresponding random distributions and their correlations.

Accurately characterizing and modeling process variations, however, can be extremely expensive. Silicon wafers and chips must be extensively tested and characterized using a large number of test structures (e.g., ring oscillators) deployed in wafer scribe lines and/or within product chips [3]. Physically measuring all test structures through a limited number of I/O ports is time-consuming. Moreover, silicon measurement such as wafer probe test may damage the wafer being tested due to mechanical stress, causing additional yield loss [6]. All these technical issues lead to today's prohibitive cost for nanoscale manufacturing technology.

Recently, two different methods, *Virtual Probe* (VP) [7] and *Expectation-Maximization* (EM) [8], have been proposed to reduce the cost of silicon characterization. Both methods aim to physically sample a small subset of test structures, and predict parametric variations at other locations by a numerical algorithm. On one hand, VP takes the measurement data from a single wafer and predicts the spatial variations for that wafer. Since VP does not simultaneously consider the measurement results from multiple wafers, the similarity among different wafers (i.e., wafer-to-wafer correlation) is completely ignored. On the other hand, EM requires to measure a large number of (e.g., more than 100) wafers to extract an accurate statistical model. Hence, it cannot efficiently handle a number of practical cases where measurement

data are limited and/or non-stationary (e.g., due to low-volume manufacturing or equipment aging).

Motivated by these observations, it is desirable to develop a new prediction technique that can efficiently explore the wafer-to-wafer correlation information from a small number of (e.g., 10~20) wafers. In practice, since the variation between wafers within the same lot tends to be much smaller than the within-wafer variation between chips [14], wafers from the same lot present strongly-correlated parametric variations, as will be demonstrated by our industrial examples in Section 5. If such correlation information is appropriately extracted, it is possible to substantially improve prediction accuracy. In other words, since strong correlation exists between different wafers, we can predict the spatial variations of a wafer by borrowing measurement data from the other wafers in the same lot. The open question, however, is how to develop a *new* modeling and prediction algorithm to achieve this goal.

In this paper, we propose a new *Multi-Wafer Virtual Probe* (MVP) technique that is derived from the VP framework. Given multiple wafers from the same lot, MVP takes a few physical samples from different locations of each wafer. By exploring wafer-to-wafer correlation, the spatial variations of all wafers are estimated together through a two-step technique. In the first step, a Bayesian inference is adopted from the statistics community [12] to find a shared model template for spatial variations of the wafers being tested. Next, a detailed modeling scheme is applied to predict spatial variations of each wafer using the extracted template. The proposed two-step approach is the key technique that makes MVP superior over other traditional methods such as VP and EM.

Another important contribution of this paper is to develop a robust regression algorithm to accurately detect and remove measurement outliers (i.e., large measurement errors). Silicon measurement data are usually error-prone. Taking wafer probe test as an example, probe misalignment and/or manufacturing defect can cause a number of measurement results that are greatly deviated from the actual values [6]. To address this issue, MVP uses robust statistics such as interquartile range [15] to distinguish outliers from regular measurement data. An iterative algorithm is developed and integrated with MVP to automatically detect and remove all outliers.

The proposed MVP method has been extensively tested for a large set of silicon measurement data that are collected from 200 wafers at an advanced technology node. As will be demonstrated by the experimental results in Section 5, to accurately extract the wafer-to-wafer correlation information, the number of wafers required by MVP is up to 14.3× less than that of EM [8]. On the other hand, MVP reduces the number of required sampling points by up to 3.9× compared to VP [7], while offering the same prediction accuracy. Note that even though MVP must measure different locations of different wafers and, hence, does not reduce the silicon area of test structures, it minimizes testing/characterization time by measuring fewer test structures

for each wafer.

The remainder of this paper is organized as follows. We briefly review the background of VP in Section 2 and then describe the proposed MVP method in Section 3. In Section 4, we further develop a robust regression method for MVP to automatically detect and remove outliers. The efficacy of MVP is demonstrated by several industrial examples in Section 5. Finally, we conclude in Section 6.

2. BACKGROUND

In this section, we briefly summarize the existing algorithms for spatial variation modeling. Since our proposed MVP technique is derived from the VP framework [7], we mainly focus on the background for VP here. The technical details of EM can be found in [8].

Let $g(x, y)$ be the two-dimensional function of the performance of interest, where x and y represent the coordinate of a location within the two-dimensional plane. The performance g can be the frequency of a ring oscillator, the threshold voltage of a transistor, etc. Without loss of generality, we discretize the two-dimensional function $g(x, y)$ and denote the coordinates x and y as integers $x \in \{1, 2, \dots, P\}$ and $y \in \{1, 2, \dots, Q\}$. Mathematically, the relation between the performance value and its frequency-domain component can be represented by a two-dimensional linear transform such as discrete cosine transform (DCT) [16]:

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

where $\{G(u, v); u = 1, 2, \dots, P, v = 1, 2, \dots, Q\}$ is a set of DCT coefficients and:

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

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

Equivalently, the performance value $\{g(x, y); x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$ can be represented as the linear combination of $\{G(u, v); u = 1, 2, \dots, P, v = 1, 2, \dots, Q\}$ by inverse discrete cosine transform (IDCT):

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

Virtual Probe (VP [7]) aims to measure a small number of (say, M) samples at the locations $\{(x_m, y_m); m = 1, 2, \dots, M\}$ and recover the performance value $g(x, y)$ at other locations $\{(x_m, y_m); m = M+1, M+2, \dots, PQ\}$ where $M \ll PQ$. Towards this goal, we formulate the following linear equation based on the measurement data $\{g(x_m, y_m); m = 1, 2, \dots, M\}$:

$$A \cdot \eta = B \quad (5)$$

where

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

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

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

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

We need to solve $\{G(u, v); u = 1, 2, \dots, P, v = 1, 2, \dots, Q\}$ from (5)-(9). Once the DCT coefficients are known, the unknown performance values $\{g(x_m, y_m); m = M+1, M+2, \dots, PQ\}$ can be easily calculated by IDCT in (4).

Note that the linear equation (5)-(9) is profoundly underdetermined. To solve (5), the authors of [7] further assume that the solution η is sparse. Since the exact locations of the non-zeros in η are unknown, VP formulates the following optimization to find the sparse solution η :

$$\begin{aligned} & \underset{\eta}{\text{minimize}} \quad \|\eta\|_0 \\ & \text{subject to} \quad A \cdot \eta = B \end{aligned} \quad (10)$$

where $\|\bullet\|_0$ stands for the L_0 -norm of a vector, i.e., the number of non-zeros in the vector. The optimization problem in (10) is NP hard and, hence, is extremely difficult to solve [9]-[10]. For this reason, VP further adopts an alternative formulation based on L_1 -norm regularization – a relaxed version of L_0 -norm [7], [9]-[10]:

$$\begin{aligned} & \underset{\eta}{\text{minimize}} \quad \|\eta\|_1 \\ & \text{subject to} \quad A \cdot \eta = B \end{aligned} \quad (11)$$

where $\|\bullet\|_1$ denotes the L_1 -norm of a vector, i.e., the summation of the absolute value of all elements in the vector. The L_1 -norm regularization in (11) can be re-formulated as a linear programming problem and solved efficiently [7].

The aforementioned VP algorithm aims to efficiently model within-wafer spatial variations from a minimum number of sampling points. However, the current implementation of VP completely ignores the wafer-to-wafer correlation that can be observed for multiple wafers within the same lot. Motivated by this observation, we will develop a new Multi-Wafer Virtual Probe (MVP) algorithm in order to further improve modeling accuracy by integrating the wafer-to-wafer correlation information into the VP framework.

3. MULTI-WAFER VIRTUAL PROBE

Our proposed Multi-Wafer Virtual Probe (MVP) consists of two major steps: (1) solve a shared model template for all wafers being tested, and (2) apply a detailed modeling scheme to predict the spatial variations of each wafer. In this section, we describe the mathematical formulation of MVP and highlight its novelties.

3.1 Problem Definition

Given L wafers from the same lot, we denote the measurement data from all wafers as $\{g_{(l)}(x_{(l),m}, y_{(l),m}); l = 1, 2, \dots, L, m = 1, 2, \dots, M\}$, where $g_{(l)}$ and $(x_{(l),m}, y_{(l),m})$ are the performance function and the m -th sampling point of the l -th wafer, respectively. Note that the number of samples can vary from wafer to wafer. For each wafer, a linear equation can be formulated based on (4) to solve the DCT coefficients:

$$A_{(l)} \cdot \eta_{(l)} = B_{(l)} \quad (l = 1, 2, \dots, L) \quad (12)$$

where

$$A_{(l)} = \begin{bmatrix} A_{(l),1,1} & A_{(l),1,2} & \cdots & A_{(l),1,P,Q} \\ A_{(l),2,1,1} & A_{(l),2,1,2} & \cdots & A_{(l),2,P,Q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{(l),M,1,1} & A_{(l),M,1,2} & \cdots & A_{(l),M,P,Q} \end{bmatrix} \quad (13)$$

$$A_{(l),m,u,v} = \alpha_u \cdot \beta_v \cdot \cos \frac{\pi(2x_{(l),m} - 1)(u-1)}{2 \cdot P} \cdot \cos \frac{\pi(2y_{(l),m} - 1)(v-1)}{2 \cdot Q} \quad (14)$$

$$\eta_{(l)} = [G_{(l)}(1,1) \ \cdots \ G_{(l)}(P,Q)]^T \quad (15)$$

$$B_{(l)} = [g_{(l)}(x_{(l),1}, y_{(l),1}) \ \cdots \ g_{(l)}(x_{(l),M_l}, y_{(l),M_l})]^T. \quad (16)$$

In (15), $G_{(l)}(u,v)$ is the DCT coefficient of the l -th wafer at the frequency (u,v) .

In (12), each equation $A_{(l)} \eta_{(l)} = B_{(l)}$ can be individually solved by the L_1 -norm regularization in (11). Such an approach, however, is not optimal, since it ignores the wafer-to-wafer correlation among these L wafers. This, in turn, motivates us to develop a new algorithm in the following sub-sections to further improve the modeling accuracy for spatial variations of multiple wafers.

3.2 L_2 -Norm Regularization

In this sub-section, we first introduce the L_2 -norm regularization method, as it is one of the key mathematical tools that facilitate the proposed MVP algorithm. L_2 -norm regularization [17] is a classic method in statistics to solve the linear equation $A_{(l)} \eta_{(l)} = B_{(l)}$ in (12). It can be formulated as the following optimization:

$$\begin{aligned} & \underset{\eta_{(l)}}{\text{minimize}} \quad \|\eta_{(l)}\|_2^2 \\ & \text{subject to} \quad \|A_{(l)} \cdot \eta_{(l)} - B_{(l)}\|_2 \leq e_{(l)} \end{aligned} \quad (l=1,2,\dots,L) \quad (17)$$

where $\|\bullet\|_2$ denotes the L_2 -norm of a vector, i.e., the square root of sum of squared value of all elements in the vector. Comparing (17) to the L_1 -norm regularization in (11), we would notice two key differences. First, the cost function in (17) is defined as L_2 -norm, instead of L_1 -norm. Second, the inequality constraint in (17) forces the sum of squared error to be less than or equal to a threshold $e_{(l)}$. Here, $e_{(l)}$ is added to consider measurement noise. In the extreme case, if $e_{(l)}$ equals 0 and, hence, there is no measurement noise, the inequality constraint in (17) is equivalent to the equality constraint in (11).

Eq. (17) can be converted to the following unconstrained optimization based on the Lagrange multiplier theory [18]:

$$\underset{\eta_{(l)}}{\text{minimize}} \quad w_{(l),0} \|A_{(l)} \eta_{(l)} - B_{(l)}\|_2^2 + \|\eta_{(l)}\|_2^2 \quad (l=1,2,\dots,L) \quad (18)$$

where $w_{(l),0}$ is the Lagrange multiplier of the inequality constraint in (17). Given a fixed value of $e_{(l)}$, there exists a unique $w_{(l),0}$ that makes (17) and (18) equivalent, i.e., both optimizations share the same solution $\eta_{(l)}$.

While Eq. (18) shows the basic formulation of L_2 -norm regularization, we can further derive a weighted formulation that is slightly different from (18) [11]:

$$\underset{\eta_{(l)}}{\text{minimize}} \quad w_{(l),0} \|A_{(l)} \eta_{(l)} - B_{(l)}\|_2^2 + \|W_{(l)}^{-1} \eta_{(l)}\|_2^2 \quad (l=1,2,\dots,L) \quad (19)$$

where $W_{(l)}$ is a diagonal matrix with $w_{(l)} = [w_{(l),1}, w_{(l),2}, \dots, w_{(l),PQ}]^T$ as its diagonal entries and $w_{(l),i}$ is the weight associated with the DCT coefficient $\eta_{(l),i}$. It can be shown that the optimization in (19) is convex [18] and, hence, can be solved efficiently and robustly.

Compared to (18), Eq. (19) further allows us to specify the ‘‘importance’’ of each DCT coefficient $\eta_{(l),i}$ by assigning an appropriate weight $w_{(l),i}$ to it. Intuitively, if the weight $w_{(l),i}$ of the DCT coefficient $\eta_{(l),i}$ is small, $\eta_{(l),i}$ is likely to be zero, as it is strongly penalized in the cost function of (19). In the extreme case, if $w_{(l),j}$ equals 0, $\eta_{(l),j}$ must be 0 in order to minimize the cost function in (19). For this reason, the L_2 -norm regularization in (19) results in a sparse solution $\eta_{(l)}$, if the weights $\{w_{(l),i}; i =$

$1,2,\dots,PQ\}$ are appropriately designed.

The aforementioned discussion implies an important fact that the weights $\{w_{(l),i}; i = 1,2,\dots,PQ\}$ in (19) serve as a model template to determine whether the DCT coefficient $\eta_{(l),i}$ is likely to be zero or non-zero. Once these weight values are known, solving the DCT coefficients becomes trivial, as shown in (19). In what follows, we will derive a Bayesian algorithm to determine the optimal weight values by exploring the wafer-to-wafer correlation information.

3.3 Wafer-to-Wafer Correlation

An efficient way to explore the correlation among different wafers can be motivated by closely observing the wafer-level measurement data. Figure 1 (a)-(c) show the normalized ring oscillator (RO) period measured for three randomly chosen industrial wafers, where Figure 1 (a)-(b) are within the same lot. Studying Figure 1, it can be seen that wafers across different lots exhibit much larger variation than wafers within the same lot. Moreover, it can be seen that the within-wafer variation between chips is much larger than the variation between wafers within the same lot. The same observation is also stated in [14]. If we further study the wafer variation within the same lot, it can be seen that the variation mainly lies in the mean value, instead of the within-wafer spatial variation pattern. Such a variation in mean, for example, can be caused by the systematic shift of manufacturing equipment that uniformly affects the entire wafer rather than a particular location on the wafer [14]. Therefore, if we remove the wafer-level mean value, the resulting spatial pattern presents strong similarity among different wafers within the same lot.

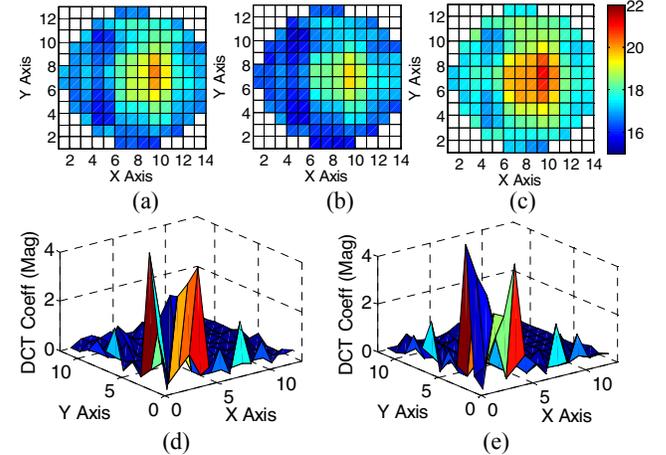


Figure 1. (a)-(c) measured ring oscillator (RO) period (normalized by a randomly selected constant) for three wafers, where (a) and (b) are within the same lot. (d)-(e) DCT coefficients (magnitude) of the measured RO period for (a) and (b) respectively.

The aforementioned discussion implies an important fact that since different wafers have similar spatial variation patterns, the corresponding DCT coefficients are strongly correlated among these wafers. To intuitively illustrate this concept, Figure 1 (d)-(e) show the magnitude of the DCT coefficients for two different wafers within the same lot. Note that these DCT coefficients almost share the same pattern in this example.

Motivated by this observation, it is possible for us to develop a shared ‘‘model template’’ in order to accurately estimate spatial variations of multiple wafers. Remember that the weights $\{w_{(l),i}; i = 1,2,\dots,PQ\}$ in (19) serve as a model template to determine the

corresponding DCT coefficients, as mentioned in Section 3.2. Hence, we propose to use identical weight values among L different wafers:

$$w_{(1),i} = w_{(2),i} = \dots = w_{(L),i} = w_i \quad (i = 1, 2, \dots, PQ). \quad (20)$$

It, in turn, results in a shared model template that efficiently explores the wafer-to-wafer correlation information.

To complete the definition of our model template, we further set up the following constraint for the Lagrange multipliers $\{w_{(l),0}; l = 1, 2, \dots, L\}$ in (19):

$$w_{(1),0} = w_{(2),0} = \dots = w_{(L),0} = w_0. \quad (21)$$

Eq. (21) implies that the measurement noise is an intrinsic property of the testing equipment and, hence, it should be similar over different wafers. Note that the measurement noise of the l -th wafer is expressed as $\|A_{(l)}\eta_{(l)} - B_{(l)}\|_2$ in (19). Even though different wafers share the same Lagrange multiplier w_0 , the exact value of $\|A_{(l)}\eta_{(l)} - B_{(l)}\|_2$ can still be different. In other words, given a testing equipment, the statistics (e.g., variance) of its measurement noise are fixed. However, since measurement noise is random, we can get different noise values from different measurements.

Once the shared model template is defined in (20)-(21), we need to further determine the optimal value of $\{w_i; i = 0, 1, \dots, PQ\}$ based on the measurement data collected from all wafers. In the next sub-section, we will borrow a Bayesian method from the statistics community [12] to address this problem.

3.4 Bayesian Inference

The essence of Bayesian inference is to model unknown parameters using probability. In this sub-section, we borrow the Bayesian inference idea in [12] to solve $\{w_i; i = 0, 1, \dots, PQ\}$ in a probabilistic setting. Towards this goal, we first derive a probabilistic formulation for the L_2 -norm regularization in (19).

Substituting (20)-(21) into (19) yields:

$$\min_{\eta_{(l)}} w_0 \|A_{(l)}\eta_{(l)} - B_{(l)}\|_2^2 + \|W^{-1}\eta_{(l)}\|_2^2 \quad (l = 1, 2, \dots, L) \quad (22)$$

where W is a diagonal matrix with $w_{(l)} = [w_1, w_2, \dots, w_{PQ}]^T$ as its diagonal entries. In (22), $w_0 \|A_{(l)}\eta_{(l)} - B_{(l)}\|_2^2$ is included in the cost function to penalize the measurement noise (or equivalently, the measurement uncertainty). In a probabilistic sense, such uncertainty can be modeled by representing $B_{(l)}$ as the sum of the following two terms:

$$B_{(l)} = A_{(l)}\eta_{(l)} + \varepsilon_{(l)}. \quad (23)$$

In (23), $A_{(l)}\eta_{(l)}$ denotes the actual performance values which depend on the DCT coefficients $\eta_{(l)}$. On the other hand, $\varepsilon_{(l)}$ is a vector of random variables that models the measurement noise of the l -th wafer.

The noise $\varepsilon_{(l)}$ is typically modeled as independent and identically distributed (*i.i.d.*) zero-mean Normal random variables:

$$pdf(\varepsilon_{(l)} | w_0) = \left(\frac{w_0}{2\pi}\right)^{\frac{M_l}{2}} \prod_{i=1}^{M_l} \exp\left(-\frac{w_0 \varepsilon_{(l),i}^2}{2}\right). \quad (l = 1, 2, \dots, L) \quad (24)$$

In (24), the Lagrange multiplier w_0 is used to define the *precision* of the measurement $B_{(l),i}$ (i.e., the inverse of the variance of the measurement noise). Intuitively, if w_0 is large (or equivalently, the variance of noise is small), the L_2 -norm $\|A_{(l)}\eta_{(l)} - B_{(l)}\|_2$ is strongly penalized in (22) so that the residue $A_{(l)}\eta_{(l)} - B_{(l)}$ is small. The equivalence between the L_2 -norm representation $w_0 \|A_{(l)}\eta_{(l)} - B_{(l)}\|_2^2$ in (22) and the probability formulation in (24) can be mathematically proven by applying several statistical theorems [17]. More details on this topic can be found in [17].

Due to the random noise $\varepsilon_{(l)}$, the measured performance $B_{(l)}$ is

a random variable, even if the DCT coefficients $\eta_{(l)}$ are fixed. Given the noise distribution in (24), it is easy to verify that $B_{(l)}$ satisfies the following Multivariate Normal distribution:

$$pdf(B_{(l)} | \eta_{(l)}, w_0) = \left(\frac{w_0}{2\pi}\right)^{\frac{M_l}{2}} \cdot \exp\left(-\frac{1}{2} w_0 \|A_{(l)}\eta_{(l)} - B_{(l)}\|_2^2\right). \quad (25)$$

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

Eq. (25) models the statistical uncertainty of the measured performance $B_{(l)}$ due to the measurement noise $\varepsilon_{(l)}$.

Next, the second term $\|W^{-1}\eta_{(l)}\|_2$ in (22) regularizes the magnitude of the DCT coefficients $\eta_{(l)}$. It is this regularization term that allows us to find a unique solution from the underdetermined linear equation: $A_{(l)}\eta_{(l)} = B_{(l)}$. The L_2 -norm regularization $\|W^{-1}\eta_{(l)}\|_2$ can be interpreted by the following statistical framework. Assume that the DCT coefficients $\eta_{(l)}$ are statistically modeled as independent zero-mean Normal random variables:

$$pdf(\eta_{(l)} | w_{1,\dots,PQ}) = \prod_{i=1}^{PQ} \sqrt{\frac{1}{2\pi w_i^2}} \cdot \exp\left(-\frac{\eta_{(l),i}^2}{2w_i^2}\right) \quad (26)$$

$$= \prod_{i=1}^{PQ} \sqrt{\frac{1}{2\pi w_i^2}} \cdot \exp\left(-\frac{\|W^{-1}\eta_{(l)}\|_2^2}{2}\right) \quad (l = 1, 2, \dots, L)$$

Intuitively, the probability distribution in (26) indicates the ‘‘uncertainty’’ of the unknown DCT coefficients $\eta_{(l)}$. Studying (26), we would have two important observations. First, the probability density function decreases, as the magnitude of $\eta_{(l)}$ increases. It, in turn, implies that small DCT coefficients are more likely to occur than large ones. This observation is consistent with the L_2 -norm regularization in (22) where the term $\|W^{-1}\eta_{(l)}\|_2$ is added to the cost function to penalize large DCT coefficients. Second, the weight w_i is used to define the standard deviation of the DCT coefficients $\{\eta_{(l),i}; l = 1, 2, \dots, L\}$ in (26). If w_i is small, the corresponding DCT coefficients $\{\eta_{(l),i}; l = 1, 2, \dots, L\}$ are likely to be small. In the extreme case, if w_i equals 0, the distribution of $\{\eta_{(l),i}; l = 1, 2, \dots, L\}$ becomes a Dirac delta function centered at $\{\eta_{(l),i} = 0; l = 1, 2, \dots, L\}$, meaning that $\{\eta_{(l),i}; l = 1, 2, \dots, L\}$ must be 0 for all wafers. This observation is also consistent with the formulation in (22) where the weight w_i specifies the ‘‘importance’’ of the DCT coefficients $\{\eta_{(l),i}; l = 1, 2, \dots, L\}$.

Based on the Bayes’ theorem [17], combining (25) and (26) yields:

$$pdf(B_{(l)}, \eta_{(l)} | w_{0,\dots,PQ}) = pdf(B_{(l)} | \eta_{(l)}, w_0) \cdot pdf(\eta_{(l)} | w_{1,\dots,PQ}). \quad (27)$$

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

The objective of deriving the Bayesian inference in (27) is to solve the optimal value of $\{w_i; i = 0, 1, \dots, PQ\}$. Towards this goal, we apply Maximum Likelihood Estimation (MLE) to find the optimal $\{w_i; i = 0, 1, \dots, PQ\}$ that are most likely to occur. Namely, we aim to maximize the following conditional probability:

$$\underset{w_{0,\dots,PQ}}{\text{maximize}} \quad pdf(B_{(1,2,\dots,L)} | w_{0,\dots,PQ}). \quad (28)$$

If the measurement of different wafers is performed independently, Eq. (28) can be represented as:

$$\underset{w_{0,\dots,PQ}}{\text{maximize}} \quad \prod_{l=1}^L pdf(B_{(l)} | w_{0,\dots,PQ}). \quad (29)$$

Since the function $\log(\bullet)$ monotonically increases, Eq. (29) is equivalent to maximizing the following *log-likelihood*:

$$\underset{w_{0,\dots,PQ}}{\text{maximize}} \quad \sum_{l=1}^L \log[pdf(B_{(l)} | w_{0,\dots,PQ})]. \quad (30)$$

Combining (27) and (30), we would notice that

$pdf(B_{(l)}|w_0, \dots, PQ)$ is the marginal probability density function of $pdf(B_{(l)}, \eta_{(l)}|w_0, \dots, PQ)$. Therefore, $pdf(B_{(l)}|w_0, \dots, PQ)$ can be found from (27) by integrating out $\eta_{(l)}$:

$$pdf(B_{(l)} | w_0, \dots, PQ) = \int pdf(B_{(l)}, \eta_{(l)} | w_0, \dots, PQ) \cdot d\eta_{(l)}. \quad (31)$$

$$= \int pdf(B_{(l)} | \eta_{(l)}, w_0) \cdot pdf(\eta_{(l)} | w_1, \dots, PQ) \cdot d\eta_{(l)}$$

Note that $pdf(B_{(l)}|\eta_{(l)}, w_0)$ and $pdf(\eta_{(l)}|w_1, \dots, PQ)$ are defined in (25) and (26) respectively. Combining (25) and (26), we can derive an analytical expression for (31):

$$pdf(B_{(l)} | w_0, \dots, PQ) = (2\pi)^{-\frac{M_l}{2}} \det(C_{(l)})^{-\frac{1}{2}} \exp\left(-\frac{1}{2} B_{(l)}^T C_{(l)}^{-1} B_{(l)}\right) \quad (32)$$

where

$$C_{(l)} = w_0^{-1} I + A_{(l)} W^2 A_{(l)}^T \quad (33)$$

and I represents the identity matrix and $\det(\bullet)$ stands for the determinant of a matrix. Substituting (32) into (30), we have:

$$\underset{w_0, \dots, PQ}{\text{minimize}} \sum_{l=1}^L \left\{ B_{(l)}^T C_{(l)}^{-1} B_{(l)} + \log[\det(C_{(l)})] \right\}. \quad (34)$$

Eq. (34) shows the analytical form of the cost function that we should minimize, in order to determine the optimal value of $\{w_i; i = 0, 1, \dots, PQ\}$. It should be noted that the optimization in (34) is not convex. However, an efficient greedy algorithm has been developed in [12] to solve $\{w_i; i = 0, 1, \dots, PQ\}$ from (34), and it has been demonstrated as an efficient numerical method for many practical problems. More details on the numerical solver can be found in [12]. Once $\{w_i; i = 0, 1, \dots, PQ\}$ are determined, the DCT coefficients $\eta_{(l)}$ of each wafer can be efficiently solved from (22).

3.5 Summary

Algorithm 1 summarizes the major steps of the proposed MVP method. Starting from the measurement data $\{g_{(l)}(x_{(l),m}, y_{(l),m})\}; l = 1, 2, \dots, L, m = 1, 2, \dots, M_l\}$, MVP first adopts a recently developed statistical algorithm [12] to determine the shared weights $\{w_i; i = 0, 1, \dots, PQ\}$ by solving the optimization in (34). The objective of this step is to extract a shared model template by exploring the wafer-to-wafer correlation information. Next, we solve the convex optimization in (22) to determine the DCT coefficients $\{G_{(l)}(u, v)\}; l = 1, 2, \dots, L, u = 1, 2, \dots, P, v = 1, 2, \dots, Q\}$ for all wafers. Finally, the spatial variations $\{g_{(l)}(x, y)\}; l = 1, 2, \dots, L, x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$ are re-constructed by IDCT, as discussed in Section 2.

Algorithm 1: Multi-Wafer Virtual Probe (MVP)

1. Start from the measurement data $\{g_{(l)}(x_{(l),m}, y_{(l),m})\}; l = 1, 2, \dots, L, m = 1, 2, \dots, M_l\}$ collected from L different wafers.
 2. Formulate the linear equations in (12)-(16).
 3. Apply the greedy algorithm in [12] to solve the optimization in (34) where the matrices $\{C_{(l)}; l = 1, 2, \dots, L\}$ are defined in (33). Determine the weights $\{w_i; i = 0, 1, \dots, PQ\}$.
 4. Solve the convex optimization in (22) for each wafer to determine the DCT coefficients for all wafers.
 5. Apply IDCT to estimate the performance values $\{g_{(l)}(x, y)\}; l = 1, 2, \dots, L, x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$.
-

4. OUTLIER REMOVAL

As discussed in the previous section, the proposed MVP method (i.e., Algorithm 1) models the measurement noise as Normal random variables. However, such a modeling assumption is not always valid, because a number of non-ideal effects exist in reality. For example, silicon chips may fail to work due to manufacturing defects, and wafer probe test may produce wrong

measurement results due to probe misalignment [6]. Such non-ideal effects may cause the measurement results to greatly deviate from the actual values. These abnormal observations are referred to as *outliers* [15] by the statistics community. If these outliers are not appropriately removed from the measurement data, they will introduce substantial error when applying the proposed MVP algorithm. For this reason, we propose a *robust regression* method to further improve the accuracy of MVP by automatically detecting and removing all outliers. Our proposed outlier removal consists of two major steps: (1) data processing, and (2) outlier detection. In this section, we first illustrate the detailed procedure for data processing in Section 4.1, and then discuss the outlier detection algorithm in Section 4.2.

4.1 Data Processing

Outliers must be clearly separable from the regular data in order to accurately detect and remove them. In our application of wafer-level spatial variation modeling, the measured performance g at the location (x, y) is a random variable. In other words, if $g(x, y)$ is measured from L different wafers, we get L different performance values. The key idea of outlier detection is to extract the statistics of $g(x, y)$, i.e., probability distribution, lower bound, upper bound, etc. As such, outliers can be identified, since they do not follow the same statistics due to large measurement error.

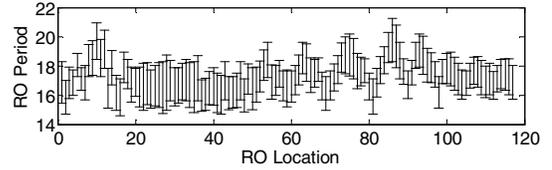


Figure 2. Normalized ring oscillator (RO) period ($\mu \pm 3\sigma$) varies as a function of RO location where silicon measurement data are collected from 24 wafers of the same lot.

To intuitively illustrate this concept, Figure 2 shows the normalized ring oscillator (RO) period as a function of the RO location. In this example, silicon measurement data are collected from 24 wafers of the same lot. For each wafer, RO period is measured at 117 different spatial locations. Next, the mean (i.e., μ) and the standard deviation (i.e., σ) are calculated from the measurement data (without including outliers) for each spatial location. The variation range (i.e., $\mu \pm 3\sigma$) is plotted in Figure 2. If a particular measurement result does not follow the statistics in Figure 2, it is likely to be an outlier and should be removed from the data set.

While the aforementioned idea for outlier detection can be directly applied to MVP, it can be further improved. Ideally, we want to process the measurement data so that the corresponding variation range is minimized. As such, we can accurately detect and remove the outliers that are abnormal. In what follows, we propose a number of comprehensive data processing steps to achieve this goal.

Remember that a significant portion of the wafer-to-wafer variation lies in the systematic shift that uniformly affects the entire wafer, instead of a particular location on the wafer, as shown in Figure 1. In other words, the performance g measured at different wafer locations share a common “global” variation component. If we can extract and remove that global component, we will substantially reduce the variance of $g(x, y)$ and, consequently, improve the accuracy of outlier detection.

Following this idea, we first estimate the median of the spatial

variation of each wafer, i.e., $\{Median(l); l = 1, 2, \dots, L\}$. Here, “median” is chosen instead of “mean”, because “mean” can be highly biased by outliers. Next, we subtract the median $\{Median(l); l = 1, 2, \dots, L\}$ from the measured performance $\{g_{(l)}(x, y); l = 1, 2, \dots, L\}$, resulting in a new data set for outlier detection:

$$f_{(l)}(x, y) = g_{(l)}(x, y) - Median(l) \quad (35)$$

$(l = 1, 2, \dots, L, x = 1, 2, \dots, P, y = 1, 2, \dots, Q)$

To illustrate the impact of the proposed data processing steps, we re-use the measurement data shown in Figure 2. Figure 3 plots the statistics of $\{f_{(l)}(x, y); l = 1, 2, \dots, L\}$, after $\{Median(l); l = 1, 2, \dots, L\}$ is removed for each wafer. Comparing Figure 2 and Figure 3, we would notice that the variation range is greatly reduced (2× smaller). It, in turn, allows us to accurately detect and remove outliers. The outlier detection algorithm will be discussed in detail in the next sub-section.

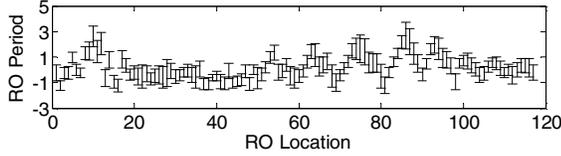


Figure 3. Normalized ring oscillator (RO) period ($\mu \pm 3\sigma$) varies as a function of RO location where silicon measurement data are collected from 24 wafers of the same lot and the median of the spatial variation is removed for each wafer.

4.2 Outlier Detection

Given the data set $\{f_{(l)}(x, y); l = 1, 2, \dots, L\}$ in (35), our outlier detection method aims to estimate the variation range of $\{f_{(l)}(x, y); l = 1, 2, \dots, L\}$ and then remove the outliers that sit outside the estimated range. In this paper, we borrow the *Interquartile Range* (IQR) method [15] from statistics and apply it to automatic outlier detection. *Quartiles* are defined as the three values in ascending order, $[Q_1, Q_2, Q_3]$, which divide the sorted data set into four equal parts. In other words, Q_1, Q_2 and Q_3 correspond to the 25%, 50% and 75% points of the cumulative distribution function (CDF) of the data.

Based on the IQR method, we first compute the following quartiles:

$$\begin{aligned} Q_1(x, y) &= Q_1\{f_{(l)}(x, y); l = 1, 2, \dots, L\} \\ Q_3(x, y) &= Q_3\{f_{(l)}(x, y); l = 1, 2, \dots, L\} \end{aligned} \quad (36)$$

$(x = 1, 2, \dots, P, y = 1, 2, \dots, Q)$

Next, we compute the *IQR* of the data:

$$IQR(x, y) = Q_3(x, y) - Q_1(x, y) \quad (37)$$

to estimate the variability. If the measurement data satisfy a Normal distribution, *IQR* is nearly equal to $4/3 \cdot \sigma$ where σ denotes the standard deviation of the distribution.

Finally, for each location $\{(x, y); x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$, we consider the measured performance $g_{(l)}(x, y)$ as an outlier, if the corresponding $f_{(l)}(x, y)$ is outside the following variation range:

$$R(x, y) = [Q_1(x, y) - 3 \cdot IQR(x, y), Q_3(x, y) + 3 \cdot IQR(x, y)]. \quad (38)$$

The scaling factor 3 in (38) is decided empirically by the statistics community. If the measurement data satisfy a Normal distribution, the IQR method removes the data outside $\pm 4.7\sigma$ range.

To intuitively illustrate the idea of IQR, Figure 4 shows an outlier detection example containing 9 regular data points and 1 outlier. In this example, the variability of the regular data is correctly captured by the quartiles, and the outlier is beyond the

boundary defined by $Q_3 + 3 \cdot IQR$. On the other hand, both the mean (i.e., μ) and the standard deviation (i.e., σ) cannot be accurately estimated due to the influence of the outlier. As a result, the outlier remains inside the interval specified by $\mu \pm 3\sigma$. It, in turn, implies that outlier detection cannot be performed by using simple statistics (e.g., mean and standard deviation), since they are extremely sensitive to the large measurement error posed by outliers.

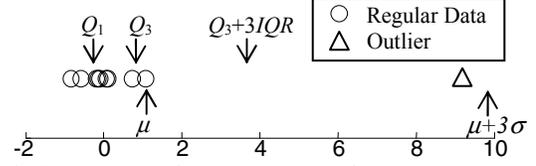


Figure 4. The proposed IQR method successfully detects the outlier that strongly biases the estimation of mean and standard deviation.

4.3 Iterative Solver

The aforementioned IQR method can be directly applied as a preprocessing step of MVP for outlier detection and removal. However, since IQR must estimate the quartiles in (36), its accuracy depends on the number of available data samples. In our application of MVP, we measure an extremely small number of samples per wafer. As will be demonstrated by the numerical examples in Section 5, when MVP is applied to 14 wafers from the same lot, only about 2 data samples are available at each spatial location. In this case, it is not feasible to estimate the quartiles accurately.

Motivated by this observation, we borrow the idea of robust regression from statistics [15] to develop an iterative algorithm. As shown in Algorithm 2, we first apply Algorithm 1 to predict the spatial variation and then use the IQR method to detect and remove outliers. In this case, Algorithm 1 can help to generate a large number of “estimated” data samples so that the quartiles can be accurately estimated. On the other hand, once the outliers are detected and removed, Algorithm 1 should be applied again in order to predict the accurate spatial variation pattern that is not influenced by the outliers. The aforementioned two steps (i.e., Algorithm 1 and IQR) are repeatedly performed until convergence is reached, i.e., no new outliers are detected and removed between two successive iteration steps. For most experimental examples that we tested, Algorithm 2 converges within two iteration steps.

Algorithm 2: Robust Regression for MVP

1. Randomly select M sampling locations for each of the L wafers.
 2. Collect the measurement data $\{g_{(l)}(x_{(l),m}, y_{(l),m}); l = 1, 2, \dots, L, m = 1, 2, \dots, M\}$ from all wafers.
 3. Run Algorithm 1 on all wafers to predict the performance values $\{g_{(l)}(x, y); l = 1, 2, \dots, L, x = 1, 2, \dots, P, y = 1, 2, \dots, Q\}$.
 4. Perform data processing using (35).
 5. Use the IQR method (36)-(38) to detect and remove outliers from the measurement data set.
 6. If any new outliers are detected and removed in step 5, go to 3. Otherwise, stop.
-

5. NUMERICAL EXAMPLES

In this section, we demonstrate the efficacy of MVP using silicon measurement data collected from 200 wafers at an advanced technology node. Each wafer contains 117 ring oscillators (ROs) distributed over different spatial locations. For

testing and comparison, three different techniques are implemented: (1) the VP method [7], (2) the EM method [8], and (3) the proposed MVP method. For all three methods, Latin Hypercube Sampling [13] is used to generate a set of random spatial locations to collect measurement data. All numerical experiments are performed on a 2.8GHz Linux server.

5.1 RO Period Measurement Data

To demonstrate the efficiency of the proposed wafer-to-wafer correlation modeling, we first compare VP and MVP (i.e., Algorithm 1) for the RO period data measured from the same lot (say, lot A). This lot contains 24 wafers with no measurement outliers. Our objective is to predict the period of each RO by measuring a small number of ROs per wafer.

Figure 5 shows the prediction error of RO period as a function of the number of measured ROs. In this paper, we define the prediction error as:

$$Error = \frac{1}{L} \sum_{l=1}^L \sqrt{\frac{\sum_{x=1}^P \sum_{y=1}^Q [g_{(l)}(x,y) - g'_{(l)}(x,y)]^2}{\sum_{x=1}^P \sum_{y=1}^Q [g_{(l)}(x,y)]^2}} \quad (39)$$

where $g_{(l)}(x,y)$ and $g'_{(l)}(x,y)$ denote the exact value and the estimated value of the performance of interest respectively, and L is the total number of the wafers being tested.

Studying Figure 5, we would notice that MVP achieves up to 2.8× error reduction over VP. The error of MVP is around 2.7% when 16 ROs (out of 117 ROs in total) are tested. To achieve the same accuracy, VP has to measure around 50 ROs (3.1× more). In this example, MVP offers superior accuracy over VP, since Algorithm 1 carefully models the wafer-to-wafer correlation information and uses it to predict spatial variation patterns.

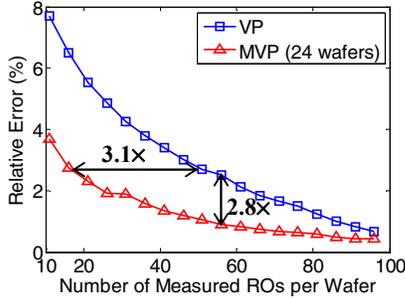


Figure 5. Prediction error of RO period for 24 wafers from the same lot (i.e., lot A) without measurement outliers.

Next, we compare VP and MVP for the RO period data measured from another lot (say, lot B). This lot contains 14 wafers with a number of measurement outliers. Hence, it serves as a good example for us to demonstrate the accuracy of the proposed outlier removal algorithm (i.e., Algorithm 2).

Figure 6 shows the prediction error of RO period as a function of the number of measured ROs. In this example, the prediction error is calculated by (39) where all outliers are excluded. Note that MVP achieves up to 3.5× error reduction over VP. The error of MVP is around 2.5% when 16 ROs (out of 117 ROs in total) are tested. To achieve the same accuracy, VP has to measure around 62 ROs (3.9× more). Compared to Figure 5, MVP offers additional accuracy improvement in this example, as it accurately detects and removes all measurement outliers. On the other hand,

since VP does not remove these outliers, they substantially bias the prediction result.

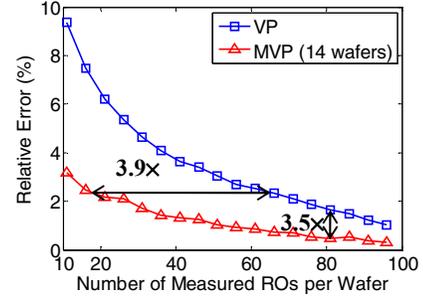


Figure 6. Prediction error of RO period for 14 wafers from the same lot (i.e., lot B) with measurement outliers.

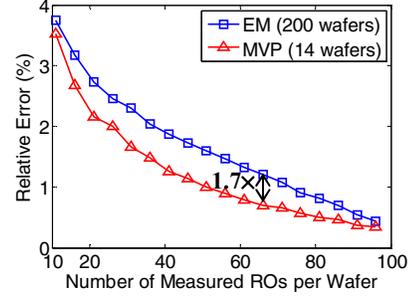


Figure 7. Prediction error of RO period where measurement outliers are automatically detected and removed.

Finally, we further compare the accuracy of MVP and EM for the RO period data with measurement outliers. In this example, MVP is applied to 14 wafers in lot B, similar to the previous case. On the other hand, since EM does not utilize the prior knowledge that variation has spatial pattern, its solution space is much larger than MVP and therefore it requires a large number of wafers to capture spatial correlation [8]. In our measurement setup, since there are 117 ROs on each wafer, 200 wafers from multiple lots are used to extract the spatial correlation model required by EM.

It should be noted that EM can automatically detect measurement outliers at the wafer level [8]. Namely, a wafer will be considered as an outlier and removed from the data set, if it contains abnormal measurement results. In this example, 40 wafers (out of 200 wafers in total) are identified as the outliers by EM. Among these 40 wafers with outliers, 6 wafers are from lot B. In other words, since lot B contains 14 wafers in total, EM identifies 8 regular wafers without outliers in this lot. These 8 wafers will be used to measure and compare the prediction error of MVP and EM in this example.

Figure 7 shows the prediction error of RO period as a function of the number of measured ROs. Note that MVP achieves up to 1.7× accuracy improvement compared to EM. More importantly, MVP requires only 14 wafers to extract the correlation information, while EM uses 200 wafers (14.3× more) to achieve the same goal. For this reason, EM cannot be easily applied to a number of practical applications where the number of available wafers is limited due to low-volume manufacturing and/or equipment aging.

To further understand the difference between MVP and EM, Figure 8 (a) shows one of the measured wafer maps of RO period from lot B. Studying Figure 8 (a), we would notice several non-ideal effects. First, a number of measurement data are missing, as indicated by the white spaces in Figure 8 (a). Second, this

particular wafer contains a measurement outlier at the lower right corner. The measured RO period for this outlier is significantly larger than other regular measurement data. From Figure 8 (b), it can be seen that EM fails to appropriately detect the outlier and accurately capture the wafer map. Due to spatial correlation, this single outlier contaminates the prediction result of the entire wafer. On the other hand, by applying the proposed MVP algorithm, the outlier is appropriately detected and removed in this example, and the predicted wafer map accurately gives the correct spatial variation pattern, as shown in Figure 8 (c).

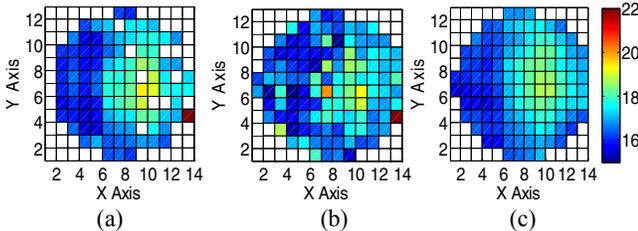


Figure 8. (a) Measured RO period (normalized by a randomly selected constant) of a wafer from lot B. (b)-(c) RO period predicted by EM and MVP by measuring 16 ROs per wafer respectively.

5.2 Leakage Power Measurement Data

In this sub-section, we compare the prediction accuracy of VP, EM and MVP for the leakage power data measured for ROs. Our objective is to predict the leakage power, $\log_{10}(P_{\text{leak}})$ (after logarithmic transform), for each RO by measuring a small number of ROs per wafer. VP and MVP are applied to 24 wafers from lot A. On the other hand, 200 wafers from multiple lots are used to extract the spatial correlation model required by EM.

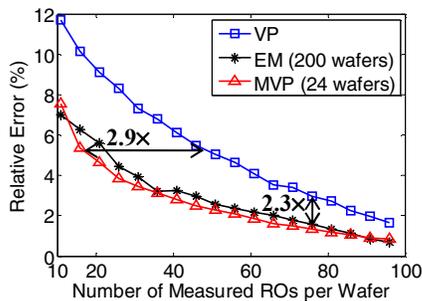


Figure 9. Prediction error of leakage power, i.e., $\log_{10}(P_{\text{leak}})$, with different number of measured ROs.

Figure 9 shows the prediction error of leakage power as a function of the number of measured ROs. Note that MVP achieves up to $2.3\times$ error reduction over VP. The error of MVP is around 5.4% when 16 ROs (out of 117 ROs in total) are tested. To achieve the same accuracy, VP has to measure around 47 ROs ($2.9\times$ more). On the other hand, while EM and MVP results in similar prediction accuracy, EM must use a large number of wafers to extract the correlation information. From this point of view, MVP is preferred over EM in a number of practical applications where the number of available wafers is limited.

6. CONCLUSIONS

In this paper, we propose a novel Multi-Wafer Virtual Probe (MVP) technique to efficiently model wafer-level spatial variations for nanoscale integrated circuits. MVP explores the wafer-to-wafer correlation information by extracting a shared

model template for multiple wafers from the same lot. In addition, a robust regression algorithm is proposed to automatically detect and remove outliers to further improve modeling accuracy. Our experimental results based on 200 industrial wafers demonstrate that MVP is preferred over other traditional approaches such as VP and EM, if a limited number of measurement data are available. The proposed MVP method can be applied to a wide range of integrated circuit applications such as variation characterization, wafer probe test, and speed binning.

7. ACKNOWLEDGEMENTS

The authors acknowledge the support of the C2S2 Focus Center, one of six research centers funded under the Focus Center Research Program (FCRP), a Semiconductor Research Corporation entity. This work is also supported in part by the National Science Foundation under contract CCF-0915912.

8. REFERENCES

- [1] S. Nassif, "Modeling and analysis of manufacturing variations," IEEE CICC, pp. 223-228, 2001.
- [2] Semiconductor Industry Associate, International Technology Roadmap for Semiconductors, 2007.
- [3] M. Bhushan, A. Gattiker, M. Ketchen and K. Das, "Ring oscillators for CMOS process tuning and variability control," IEEE Trans. Semiconductor Manufacturing, vol. 19, no. 1, pp. 10-18, Feb. 2006.
- [4] D. Blaauw, K. Chopra, A. Srivastava and L. Scheffer, "Statistical timing analysis: from basic principles to state of the art," IEEE Trans. CAD, vol. 27, no. 4, pp. 589-607, Apr. 2008.
- [5] M. Mani, A. Singh, and M. Orshansky, "Joint design-time and post-silicon minimization of parametric yield loss using adjustable robust optimization," IEEE ICCAD, pp. 19-26, 2006.
- [6] W. Mann, F. Taber, P. Seitzer and J. Broz, "The leading edge of production wafer probe test technology," IEEE ITC, pp. 1168-1195, 2004.
- [7] 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.
- [8] S. Reda and S. Nassif, "Analyzing the impact of process variations on parametric measurements: novel models and applications," IEEE DATE, pp. 375-380, 2009.
- [9] E. Candes, "Compressive sampling," International Congress of Mathematicians, 2006.
- [10] D. Donoho, "Compressed sensing," IEEE Trans. Information Theory, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.
- [11] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," IEEE Trans. Signal Processing, vol. 45, no. 3, pp. 600-616, Mar. 1997.
- [12] S. Ji, D. Dunson and L. Carin, "Multi-task compressive sensing," IEEE Trans. Signal Processing, vol. 57, no. 1, pp. 92-106, Jan. 2009.
- [13] M. McKay, R. Beckman, and W. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from computer code," Technometrics, vol. 42, no. 1, pp. 55-61, May. 1979.
- [14] M. Orshansky, S. Nassif, and D. Boning, Design for Manufacturability and Statistical Design: A Constructive Approach, Springer, 2007.
- [15] R. Maronna, R. Martin, and V. Yohai, Robust Statistics: Theory and Methods, John Wiley and Sons, 2006.
- [16] R. Gonzalez and R. Woods, Digital Image Processing, Prentice Hall, 2007.
- [17] C. Bishop, Pattern Recognition and Machine Learning, Prentice Hall, 2007.
- [18] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.