Accurate Spatial Estimation and Decomposition Techniques for Variability Characterization

Sherief Reda, *Member, IEEE,* and Sani R. Nassif, *Fellow, IEEE*

***Abstract*—In this paper, we show that the impact of process variations on the parametric measurements of semiconductor circuits can be modeled using multivariate statistical techniques. We show that it is possible to devise data transformation methods to model different kinds of measurements such as timing and leakage using multivariate statistical analysis. We use these models to propose new semiconductor spatial estimation and variability decomposition techniques. We demonstrate a new semiconductor spatial estimation technique based on the expectation-maximization algorithm. Our technique can be used to fill in the expected values of measurements at wafer locations that have been skipped or missed during parametric testing. Furthermore, we use our proposed spatial estimation method together with nested analysis of variance techniques to arrive to an accurate variability decomposition method. We extensively verify our models and results with timing and leakage variability data measurements collected from a large volume of manufactured wafers at 65 nm SOI process. Using this data we explore and quantify the trade-off between the accuracy of estimations and the reductions in the number of required parametric measurements. We demonstrate the superiority of the proposed technique for spatial estimation in comparison to geostatistical Kriging-based estimators and traditional cubic b-spline-based interpolation methods. We also show the impact of wafer sampling techniques on the accuracy of spatial estimation, and we reveal the spatial structure of various variability sources.**

***Index Terms*—Analysis, characterization, modeling, variability.**

I. Introduction

AGGRESSIVE technology scaling has led to large levels of manufacturing process variations due to statistical fluctuations inherent in the manufacturing process. Manifestations of these variations include gate length variations, line edge roughness, dopant fluctuations, and variations in the dimensions of interconnects [2], [22], [27]. These variations impact the key electrical parameters of semiconductor devices and interconnects, and correspondingly determine their metrics like performance, power, yield, and reliability [4], [5], [22], [26]. For example, at the time of writing this paper, the price difference between the fastest and slowest versions of a popular 45 nm quad-core processor is more than $700. Extrapolatary studies predict that the sources and magnitudes of process variabilities will further increase in future technology nodes [3], [8], [14], [20].

Differences in manufacturing outcomes can be attributed to *systematic* and *random variability* sources [2]. Systematic sources impact different die or wafers in a deterministic manner, while random variations are unique to each die or wafer or lot. Random variability can be decomposed into different components that reflect the hierarchy implied by the various steps in the semiconductor manufacturing process [9]. This hierarchy includes the following components: *within-die* variations that arise within the same die; *die-to-die or across wafer* variations that arise among die manufactured on the same wafer; *wafer-to-wafer* variations that appear in wafers within the same lot; and *lot-to-lot* variations that arise among different lots. Because different variations are caused by different physical phenomena, identification and estimation of each variation component is critical for determining effective variability reduction techniques [9], [15].

The increased number of variability sources and the corresponding necessary inline characterization test structures lead to increases in the costs and time dedicated to variability characterization during manufacturing [7]. Thus, there is a real incentive for techniques that can reduce the number of needed measurements without compromising the accuracy of variability characterization. *Spatial estimation* techniques achieve such reduction by substituting large numbers of physical characterization measurements (e.g., full wafer) with a smaller number of measurements and then use this smaller set of measurements to estimate the skipped measurements.

Given the parametric test measurements from the process sensitive test structures, our objectives are: 1) to develop a statistical model that accurately captures variability characterization measurements across the entire process; and 2) to utilize the developed statistical model to devise spatial estimation and variability decomposition techniques that are of benefit to both process engineers and designers. We summarize our contributions as following.

1) We show that multivariate normal (MVN) statistical techniques can be used to model our variability mea-

Manuscript received October 30, 2009; revised April 5, 2010; accepted April 9, 2010. The work of S. Reda was partially supported by the NSF CAREER, under Grant 0952866. An earlier version of this paper appeared in Design, Automation, Test in Europe 2009 [23]. This paper has an expanded coverage and new material, including: (1) a new section for parametric leakage modeling and data transformation; (2) comparisons against geostatistical-based spatial estimation techniques and cubic b-spline interpolation techniques; (3) variability decomposition using nested ANOVA techniques; (4) different scenarios for the applicability of the expectation-maximization algorithm; and (5) new experiments using both timing and leakage variability measurements using a complete set of wafer of measurements.

S. Reda is with the Division of Engineering, Brown University, Providence, RI 02912 USA (e-mail: sherief_reda@brown.edu).

S. R. Nassif is with Austin Research Laboratories, IBM Corporation, Austin, TX 78758 USA (e-mail: nassif@us.ibm.com).

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/TSM.2010.2051752

surements data across the entire process. In case the measurement data does not lend itself to MVN assumptions, we show that statistical data transformation techniques can enable the application of our modeling methods. We also describe a procedure to verify the accuracy of our multivariate model and to detect any outliers in the measurements.

2) We observe that in medium-volume to high-volume semiconductor fabrication there are typically thousands of wafers that are generated using the same process steps. Thus, by exploiting the correlation structure among the measurements at different wafer locations, it is possible to devise a highly accurate framework for spatial estimation. Our proposed approach carries out parametric test measurements at a few sites and then uses the *expectation-maximization algorithm* to estimate the expected values for the measurements at all other sites. The proposed approach reduces drastically the volume of variability characterization measurements needed.
3) We use the proposed spatial estimation technique together with nested analysis of variance (ANOVA) techniques to decompose variability into systematic and random variability sources, and we show how the variability sources can be decomposed into: lot-to-lot, wafer-to-wafer, die-to-die and within-die components. We also provide analysis techniques to uncover the spatial structure in these systematic and random sources.
4) Using thousands of measurements from process sensitive test structures, we validate the proposed techniques and demonstrate their applicability and accuracy for timing and leakage variability modeling and estimation. We elucidate: 1) the trade-off between the accuracy of spatial estimation as a function of the number of physical characterization measurements that are available to the estimation algorithm; and 2) the impact of the sampling plan of the measurements on the accuracy of the estimation results.

The organization of this paper is as follows. In Section II, we provide background information on spatial estimation techniques in the literature and nested analysis of variance (nested ANOVA) methods. In Section III, we develop the proposed multivariate statistical framework and describe the makeup of our set of variability measurement data. In Section IV, we describe our spatial estimation techniques which are based on the expectation-maximization algorithm. In Section V, we describe new methods to decompose the observed variations into systematic and random sources and explore the spatial structure of each of the sources. Finally, we summarize the main conclusions of this paper in Section VI.

II. Background

In this section, we discuss two relevant techniques that we use either as bases for some of our analysis techniques or for comparison purposes. In Subsection II-A, we discuss variograms as a technique to model spatial correlations and Kriging estimators as one potential technique for estimating missing data in spatial fields. In Subsection II-B, we discuss nested analysis of variance techniques which can be used to estimate the contributions of various variability sources.

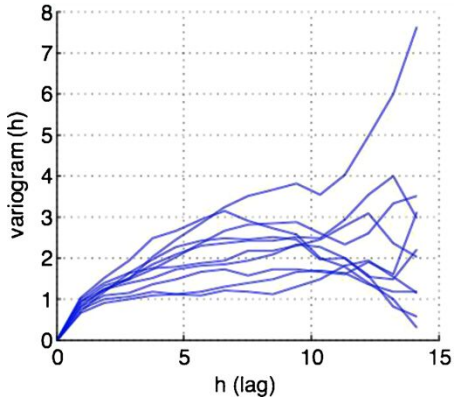

Fig. 1. Variograms of 10 random fields generated for a hypothetical field of 10 mm × 10 mm and with a $\lambda = 0.5$. The x-axis of the plot gives the lag (h) and the y-axis gives the variogram function $\gamma(h)$.

A. Variograms and Kriging-Based Estimation

Geostatistical-based techniques have been recently proposed as a general framework for spatial variability estimation [18]. In this framework the variability measurement data is assumed to form a 2-D spatial Gaussian random field that is *stationary*, i.e., the mean and variance of a random variable do not depend on the variable's location in the field, and that the covariance $C(Y(l_i), Y(l_j))$ between two random variables $Y(l_i)$ and $Y(l_j)$ at locations l_i and l_j, respectively, depends only on the (Euclidean) distance $h = \|l_i - l_j\|$ between the two locations. That is

$$C(Y(l_i), Y(l_j)) = \sigma^2 \varrho(h)$$

where the parameter $\sigma > 0$ is a *scale parameter* (σ^2 is the variance of the field) and the function ϱ is called the *correlation function*. A simple and natural model that allows for correlation between different locations is the exponential model. For this model, the correlation function decays exponentially as a function of the distance h, that is

$$\varrho(h) = e^{-\lambda h} \quad \lambda > 0. \tag{1}$$

Note that as λ increases the correlation decays faster as a function of the distance. In this respect, λ can be interpreted as the strength of correlation. Under this model, the random field has three parameters: the mean level μ, the scale parameter σ, and the strength of correlation λ. In many occasions, this spatial correlation is expressed using the *variogram* function, where the variogram $\gamma(h)$ is defined as

$$2\gamma(h) = E[(Y(l_i) - Y(l_i + h))^2] = 2\sigma^2(1 - \varrho(h)). \tag{2}$$

For example, Fig. 1 gives examples of variograms of ten synthetic random fields generated for a hypothetical spatial field with dimensions 10 mm × 10 mm with a $\lambda = 0.5$ and

$\sigma^2 = 1$. We use the procedure described by Hargreaves *et al.* [13] for generating the synthetic random fields.

Kriging is a linear estimator that estimates the value at a desired unknown location of a random field as a weighted linear combination of the measurements at known locations of the field. That is

$$Y^*(l_u) = \sum_{i=1}^{k} \alpha_i Y(l_i) \tag{3}$$

where $Y(l_i)$ are the measurements at the k known locations, $Y^*(l_u)$ is the estimate at the unknown location l_u, and α_i are the linear combination weights. Kriging estimation finds the optimal weights that minimize the error variance as computed by $E[(Y^*(l_u) - Y(l_u))^2]$. It can be shown [24] that the optimal weights satisfy the following set of linear equations

$$\sum_{j=1}^{k} \alpha_i C(Y(l_i), Y(l_j)) = C(Y(l_u), Y(l_i)) \quad \text{for } i = 1, \ldots, k. \tag{4}$$

The covariances in the linear equations described by (4) can be computed directly through the variogram function of (2). Then, these k linear equations can be solved with standard linear algebra techniques to find the α_i's required for Kriging estimation as described by (3).

While the utilization of Kriging estimators for semiconductor spatial estimation has provided encouraging results [18], the technique cannot leverage measurements from various manufactured wafers toward the estimations required for a particular wafer. Kriging estimation is good for geostatistical studies where there is typically only one set of measurements on a given domain; however, in semiconductor fabrication there are typically hundreds and thousands of wafers that are generated roughly using the same process. The covariance structure between the various measurements on the different wafers can be exploited to provide far accurate measurements as we will demonstrate in Section IV.

B. ANOVA Techniques

The hierarchical decomposition of variability in semiconductor fabrication mirrors the division of variability in many other *batch* manufacturing systems [17], [15]. In these systems, nested ANOVA methods are used to decompose the total observed variance into the components of interest. Consider the simplest case of batch manufacturing in which sample products are manufactured in batches, where each batch contains a number of samples. We would have a *two-factor* ANOVA design: *Factor A* is responsible for batch-by-batch variations, and *Factor B* is responsible for sample-to-sample variations. In this case, the change of *Factor B* is nested within the change of *Factor A*, which would be typically written as $B(A)$. If y_{ij} denotes the measurement taken from sample $j \in \{1, \ldots, s\}$ of batch $i \in \{1, \ldots, b\}$ then y_{ij} can be expressed as $y_{ij} = \mu + \alpha_i + \beta_{j(i)}$. The overall variability as computed by the total sum of squares is equal to

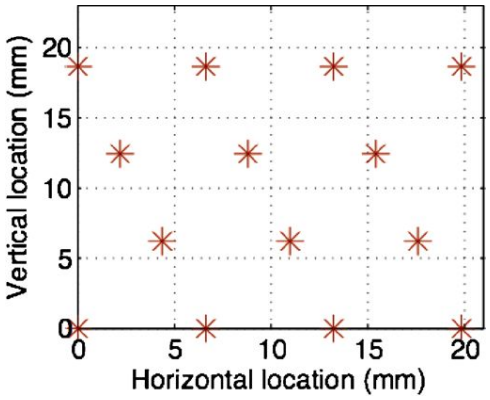

Fig. 2. Locations of PSROs within the layout of the chip.

$$\mathrm{SS}_T = \sum_{i=1}^{b} \sum_{j=1}^{s} (y_{ij} - \mu)^2. \tag{5}$$

The within batch variability expressed as the sum of squares is equal to

$$\mathrm{SS}_{B(A)} = \sum_{i=1}^{b} \sum_{j=1}^{s} (y_{ij} - \bar{y}_{i.})^2 \tag{6}$$

where $\bar{y}_{i.} = \sum_{j=1}^{b} y_{ij}/b$. The across batch variability expressed as the sum of squares is equal to

$$\mathrm{SS}_A = s \times \sum_{i=1}^{b} (\bar{y}_{i.} - \mu)^2. \tag{7}$$

It can be shown that $\mathrm{SS}_T = \mathrm{SS}_A + \mathrm{SS}_{B(A)}$ [17], and hence it is possible to decompose the total variability into two factors or components that reflect the hierarchical manufacturing structure. In an analogous method, it is possible to generalize the given two-factor nested model to nested models with a larger number of factors. We will use this technique in Section V to decompose the total variability of our measurement data into lot-to-lot, wafer-to-wafer, die-to-die, and within-die variations.

III. Proposed Variability Characterization Model

Parametric measurements occur throughout the fabrication process to track its quality control [1], [11], [21]. These measurements probe various process characteristics such as critical dimensions, transistor thresholds, timing and leakage. Our parametric data set comes from a 65 nm SOI semiconductor process, and the obtained measurements include: 1) frequency measurements from process sensitive ring oscillators (PSROs), where the frequencies are normalized by the number of stages in the ring oscillator design; and 2) leakage (IDDQ) measurements.

Each instance of our production chip holds 14 PSROs that are spatially organized along a grid that spans the die's area as shown in Fig. 2. The PSROs are used as a quick

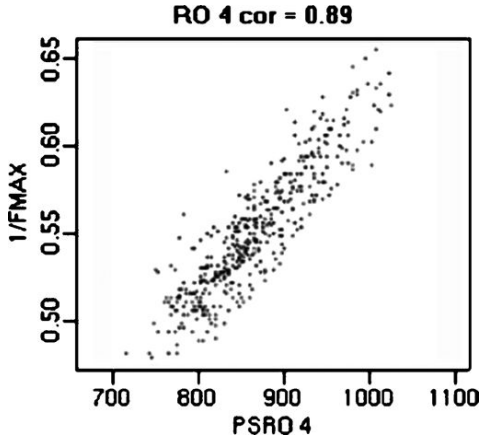

Fig. 3. Correlation between the measurements of PSRO number 4 and the actual speed measurements of the part of the chip that it resides in.

parametric indicator of the overall process quality. Previous results in the literature demonstrate that the speeds of PSROs are strongly correlated with the speeds of their embedded chips [1], [12]. In Fig. 3, we give the measurements of the fourth PSRO together with the actual timing results of the part of the chip that it resides in. Our measurements indicate a very strong correlation with a coefficient of 0.89. The leakage of each chip is evaluated using one lumped measurement. Our data set consists of measurements from 244 wafers that span 19 wafer lots and where each wafer holds 111 die. In each wafer few parametric measurements are missing. These missing measurements arise from a number of reasons including, for example, errors in test probe landing locations, failure in test probes, and manufacturing defects. The locations of the missing measurements are typically random and differ depending on the wafer. Furthermore, a few wafers have received "special" processing steps that are different from the rest of the wafers.

In this section, we describe the proposed multivariate statistical modeling framework that will be used to model our variability measurements data. In Subsection III-A, we describe the main assumptions of our multivariate statistical techniques. To address the problem of variability measurements that might not lend themselves to our statistical assumptions, we propose data transformation techniques in Subsection III-B. In Subsection III-C, we describe techniques to verify the correctness of our statistical assumptions and to detect any outliers in the data set of measurements.

A. Multivariate Normal Modeling

In this paper, we assume that a given data set of variability measurements comes from a MVN distribution with potentially a few outliers. Even if the data set does not lend itself directly to this assumption, we provide in Subsection III-B data transformation techniques that enable us to carry out such assumption safely. In our model, we consider the measurements obtained from a single wafer as an *observation* that is mathematically represented as a random vector $\mathbf{w}$. Each observation vector $\mathbf{w}$ consists of measurements on p *variables* corresponding to the number of parametric test sites on each wafer. The number of variables or sites is equal to $p = d \times r$, where d is the number of die on a wafer and r is the number of measurements per die.[1] In the MVN distribution, the probability density function for a random vector $\mathbf{w}$ is equal to

$$p(\mathbf{w}) = \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}} e^{-\frac{(\mathbf{w}-\mu)'\Sigma^{-1}(\mathbf{w}-\mu)}{2}} \tag{8}$$

where μ is a $p \times 1$ vector that gives the expected value of the random vector $\mathbf{w}$, and Σ is the $p \times p$ covariance matrix. If $\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_n}$ denote the measurements from some n wafers, then the joint density function of all observations is the product of the marginal normal densities

$$\left\{ \begin{array}{c} \text{joint density} \\ \mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_n} \end{array} \right\} = \prod_{j=1}^{n} p(\mathbf{w_j}) = \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{w_j}-\mu)'\Sigma^{-1}(\mathbf{w_j}-\mu).} \tag{9}$$

Equation (9) considered as a function of μ and Σ for the fixed set of observations $\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_n}$ is called the *likelihood* function, and

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{w_j} \tag{10}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n} (\mathbf{w_j} - \hat{\mu})(\mathbf{w_j} - \hat{\mu})' \tag{11}$$

are the maximum likelihood estimators (MLE) of μ and Σ, respectively [16].

B. Data Transformations

In many cases variability characterization data (e.g., leakage) might not appear MVN. In this case, it is necessary to *transform* the data to make it MVN. Transforming the data is just a re-expression of the data in different units [16]. The appropriate transformation could be based on a theoretical basis given the nature of the variability characterization data; for example, leakage has exponential dependency on gate length which may suggest using lognormal transformations. Another approach for transformation is based on using the data itself. This latter approach works when a physical model is hard to describe accurately. One of the popular methods for data transformation is the Box–Cox method [6]. For a univariate distribution, the Box–Cox transformation of a random variable x is given by

$$x^{(t)} = \begin{cases} \frac{x^t - 1}{t} & t \neq 0 \\ \ln x & t = 0. \end{cases} \tag{12}$$

Given a set of observations $x_1, \ldots, x_n$, the appropriate transformation t is the one that maximizes the likelihood of the

[1] If the size of the data set (that is the number of observations) is relatively small compared to $p = d \times r$, then the data can be modeled using r independent MVN distributions each with dimension p.

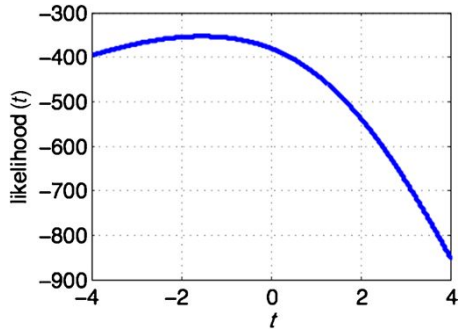

Fig. 4. Plot of likelihood function as given by (13) as a function of t to find the best transformation for the leakage measurements of the center die across all wafers.

following expression:

$$L(t) = -\frac{n}{2}\ln\left(\frac{1}{n}\sum_{j=1}^{n}(x_j^{(t)} - \overline{x^{(t)}})^2\right) + (t-1)\sum_{j=1}^{n}\ln x_j \quad (13)$$

where $\overline{x^{(t)}}$ is the arithmetic average value of transformed observations. The maximum value of the likelihood function of (13) occurs when $dL/dt = 0$, and it can be readily found through numerical techniques. In our particular implementation, we use the Nelder–Mead algorithm which is an unconstrained nonlinear minimization algorithm [19]. For example, Fig. 4 gives the likelihood of the leakage measurements of the center die from all wafers as a function of the Box–Cox transformation value t. From the plot, it is apparent that $t = -1.5235$ maximizes the likelihood function, and thus, it is the best choice to transform the leakage measurements to look normal. For illustration, we plot the delay-leakage measurements (before and after transformation) in Fig. 5.

To verify that the Box–Cox transformation provides a better transformation technique than the standard logarithm transformation for leakage power, we compare the normality of the resultant data from the two transformations. We conduct the Kolmogorov–Smirnov test on both sets of the transformed data under the null hypothesis that the data has a normal distribution. The null hypothesis was rejected at the 5% significance level for the logarithm transformed data but was accepted for the Box–Cox transformed data. In accordance with the Kolmogorov–Smirnov test, we measure the maximum distance between the cumulative distribution function (CDF) of the transformed data and the ideal normal distribution as shown in Fig. 6. The Box–Cox transformed report a distance of 1.3 while the logarithm transformed data report a distance of 3.05, which further confirms that the Box–Cox transformed data is closer to normality than the standard logarithm transformation.

As a wafer observation has many die rather than a single one, it is necessary to devise data transformation techniques that are applicable for multivariate observations. In this case, a power transformation must be selected for each of the variables of the multivariate distribution. If $t_1, t_2, \ldots, t_p$ denote the power transformations of the p variables, then a multivariate observation $\mathbf{w_j} = (x_{j1}, x_{j2}, \ldots, x_{jp})$ is transformed to $\mathbf{w_j}^{(t)} = (x_{j1}^{(t_1)}, x_{j2}^{(t_2)}, \ldots, x_{jp}^{(t_p)})$. Each t_i can be computed by using the maximum likelihood approach on each individual variable as given by (13). While applying the transformation on variables individually does not ensure that the joint distribution is normal, in many cases such approach is practically good enough [16].

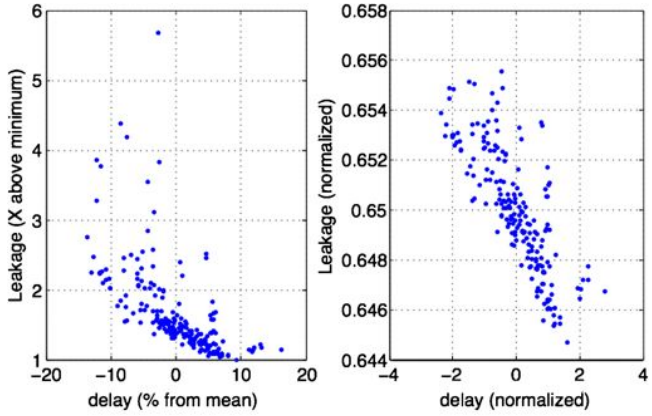

Fig. 5. Application of Box–Cox transformation to transform nonnormal leakage measurements (of the center die across all wafers) to normal data.

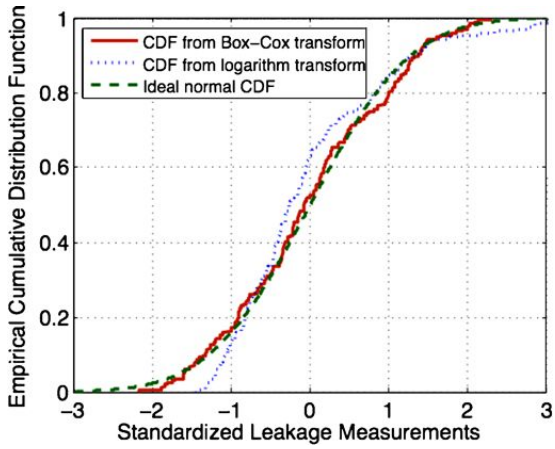

Fig. 6. Empirical CDF plot of leakage power after transformation.

C. Verifying the MVN Assumption

It is necessary to verify the MVN assumption on either the original data or the transformed data set. A formal way to assess the joint normality of a data set is based on calculating the squared *Mahalanobis* distances d_j^2 of the observations, where

$$d_j^2 = (\mathbf{w_j} - \mu)'\Sigma^{-1}(\mathbf{w_j} - \mu) \quad \mathrm{j} = 1, \ldots, n. \quad (14)$$

Essentially (14) summarizes all the measurements of a wafer by a single number. It can be shown [16] that

$$d_j^2 = (\mathbf{w_j} - \mu)'\Sigma^{-1}(\mathbf{w_j} - \mu) = \sum_{i=1}^{p} z_i^2 \quad (15)$$

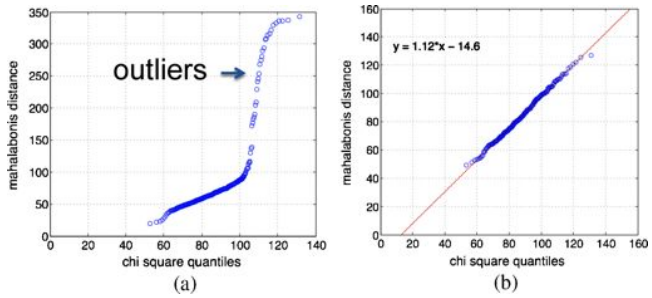

Fig. 7. (a) Chi-square plot for the entire data set. (b) Chi-square plot after removing the few outlier wafers. The x-axis of the plots gives the chi-square quantiles, and the y-axis gives the Mahalanobis distance quantiles.

where $z_1, z_2, \ldots, z_p$ are independent standard normal variables. Since $\sum_{i=1}^{p} z_i^2 = \chi_p^2$, then the *Mahalanobis* distance has a *chi-square distribution* with p-degrees of freedom. We can assess the multivariate normality of a data set of $\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_n}$ by plotting the *chi-square plot* of the Mahalanobis distances of its wafer observations (either directly or after an applicable data transformation). The *chi-square plot* is a generalization of the Q–Q plot for the case of MVN observations. We use the following standard procedure to construct the chi-square plot.

1) Use (14) to compute the squared Mahalanobis distances for each of the n wafers. Then, order the computed distances from smallest to largest $d_{(1)}^2 \leq d_{(2)}^2 \leq \cdots \leq d_{(n)}^2$.
2) On a 2-D plane, plot the pairs $(q_p((j - \frac{1}{2})/n), d_j^2)$, where $q_p((j - 1/2)/n)$ is the $100((j - \frac{1}{2})/n)$ quantile of the standard chi-square distribution with p degrees of freedom.

If the MVN assumption is indeed true, then the chi-square plot should be a straight line that passes through the origin of the plane with a slope equal to 1.

To tune the fabrication process, process engineers occasionally experiment with some wafers to assess any proposed process changes. The measurements of these wafers, which are included with the rest of the measurements, may constitute unusual observations, or *outliers*, within the population of observations. One method to detect outlier wafers is to examine the calculated Mahalanobis distances using (14) for unusually large numbers. In a chi-square plot, the outliers would be the points farthest from the origin.

Fig. 7(a) gives the chi-square plot for our entire data set of PSRO frequency measurements. The x-axis of the plot gives the chi-square quantiles, and the y-axis gives the Mahalanobis distance quantiles. As it is clear from the figure, there are points with extreme values that are apart from the rest of the bulk points. Checking these points against the manufacturing recipes of the different wafers confirmed that the wafers corresponding to these points received special processing steps. After removing these few outlier wafers from our data set, we re-plot the chi-square plot in Fig. 7(b). In contrast to Fig. 7(a), Fig. 7(b) displays quite a linear plot that passes close to the origin with a near unity slope. The linearity of the chi-square plot verifies that our MVN assumption is an accurate way to model the parametric measurements in hand.

IV. Proposed Spatial Variability Estimation Technique

Based on the proposed statistical model, we propose in this section a new method for spatial estimation of semiconductor variability measurements. There are two reasons that motivate spatial estimation of variability measurements. The first reason is that in each wafer, a good number of the parametric measurements are naturally missing because of errors in test probe landing locations, failure in test probes, and manufacturing defects. The locations of the missing measurements are typically random and differ depending on the wafer. The second reason is that process engineers could intentionally skip some sites from measuring due to limitations in the cost and time required for variability measurements. Our method substitutes expensive physical measurements by *soft* computational methods that are capable of accurately estimating the expected values of the skipped measurements.

A. Proposed Method

To compute the MLE distribution parameters from incomplete data, we propose utilizing the expectation maximization (EM) algorithm [10]. The EM algorithm enables parameter estimation in multivariate statistical models with incomplete data. The algorithm is an iterative procedure for estimating the expected values of some unknown quantities, given the values of some correlated, known quantities. EM assumes that the quantities are represented as values in some parameterized probability distribution such as the MVN distribution. The EM algorithm involves two main steps. The two steps are the *Expectation* Step and the *Maximization* Step. A general framework for the EM algorithm is given as following.

1) Initialize the MLEs of distribution parameters ($\hat{\mu}$ and $\hat{\Sigma}$).
2) Repeat until convergence.
   a) *E-Step:* estimate the **E**xpected value of the missing measurements given the current MLEs ($\hat{\mu}$ and $\hat{\Sigma}$) of the distribution.
   b) *M-Step:* given the expected estimates of the missing measurements, re-estimate the distribution parameters ($\hat{\mu}$ and $\hat{\Sigma}$) to **M**aximize the likelihood of the data.

Because the results of the E-Step and M-Step depend on each other, the EM algorithm is iterated a number of times until the convergence of $\hat{\mu}$ and $\hat{\Sigma}$. The key to the success of the EM algorithm lies in the operation of the E-Step. To understand how the estimation is carried out in the E-Step, it is necessary to introduce some notation. For an observation vector $\mathbf{w_j}$ with some missing values, let $\mathbf{w_j^u}$ denote the *unknown* or *missing* measurements, and let $\mathbf{w_j^k}$ denote the *known* measurements. Thus, $\mathbf{w_j}$ can be partitioned as $\mathbf{w_j} = \begin{bmatrix} \mathbf{w_j^u} \\ \mathbf{w_j^k} \end{bmatrix}$, and accordingly $\hat{\mu}$ and $\hat{\Sigma}$ can be partitioned as $\hat{\mu} = \begin{bmatrix} \hat{\mu}^u \\ \hat{\mu}^k \end{bmatrix}$ and $\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{u,u} & \hat{\Sigma}_{u,k} \\ \hat{\Sigma}_{k,u} & \hat{\Sigma}_{k,k} \end{bmatrix}$. Then, the conditional probability of $\mathbf{w_j^u}$ given $\mathbf{w_j^k}$ is normal, where the

$$\text{mean of } \mathbf{w_j^u} = \hat{\mu}^{\mathbf{u}} + \hat{\Sigma}_{u,k}\hat{\Sigma}_{k,k}^{-1}(\mathbf{w_j^k} - \hat{\mu}^{\mathbf{k}}). \tag{16}$$

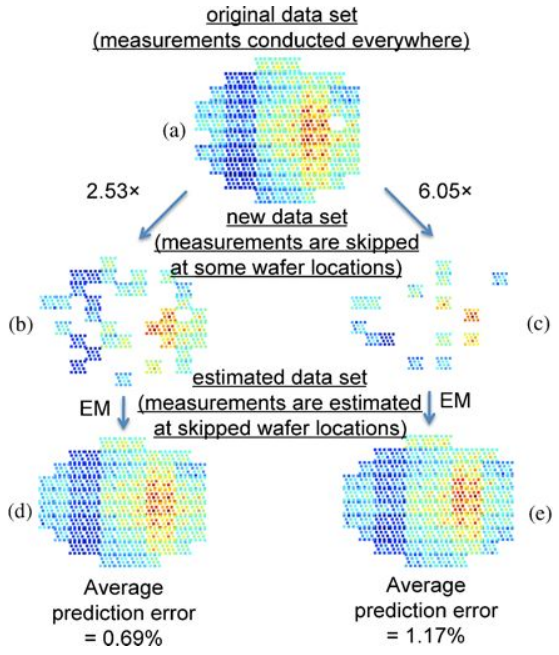

Fig. 8. Estimating skipped PSRO measurements.

Equation (16) is the key method used for the E-Step, and it can be intuitively explained as follows (formal derivations can be found in [10] and [16]). The equation basically says that the expected values for the missing measurements of wafer j are equal to the estimated means at their locations $\hat{\mu}^{\mathbf{u}}$ plus some term that estimates the deviations of the unknown measurements of wafer j from their mean $\hat{\mu}^{\mathbf{u}}$. This term is equal to the deviations of the known measurements of wafer j from their mean $(\mathbf{w}_{\mathbf{j}}^{\mathbf{k}} - \hat{\mu}^{\mathbf{k}})$ multiplied by some weight. This weight is the product of:

1) the covariance, $\hat{\Sigma}_{u,k}$, between the unknown and the known measurements, which reflects the dependencies between the known measurements and the unknown measurements;
2) the inverse of the variance of the known measurements $\hat{\Sigma}_{k,k}^{-1}$, which reduces the contribution of the known measurements toward the estimation if they have large variances, and consequently they should not be quite "trusted."

For example, Fig. 8(a) shows a visual imagery of the PSRO variability measurements if all the sites of a wafer are measured. Fig. 8(b) shows the measured variability after skipping the measurements of 68 die leading to a reduction factor of 2.53× in the number of required measurements. Fig. 8(d) shows the estimations of the skipped measurements with an average estimation error of 0.69%. Fig. 8(c) shows the test results after skipping the measurements of 93 die. Fig. 8(e) gives the expected estimations of the skipped measurements with an average estimation error of 1.17%.

Another interesting aspect of the EM algorithm is that it also computes the covariance of the estimated measurements as follows:

$$\text{covariance of } \mathbf{w}_{\mathbf{j}}^{\mathbf{u}} = \hat{\Sigma}_{u,u} - \hat{\Sigma}_{u,k}\hat{\Sigma}_{k,k}^{-1}\hat{\Sigma}_{k,u}. \tag{17}$$

The diagonal elements of the covariance matrix of (17) give the variances in the estimations of the missing measurements. The ability of the EM algorithm to calculate the variances in its estimations is useful as it gives the *confidence intervals* of the EM algorithm in its estimation of a skipped measurement. Small variances indicate that the EM algorithm is confident in its estimation of the missing measurements, and large variances indicate that the estimated values for the missing measurements might significantly deviate from their true values. The calculated variance of an estimated measurement provides a *confidence interval* where the true measurement of a skipped site would likely fall. These confidence intervals provide "safety nets" for the estimated measurements. For example, if we denote an estimated measurement of a skipped site by m_s and the variance in estimation by $\sigma_{s,s}$ then one can be almost certain that the true value of the measurement at the skipped site will fall between $m_s - 3\sqrt{\sigma_{s,s}}$ and $m_s + 3\sqrt{\sigma_{s,s}}$.

B. Empirical Validation

1) *First Experiment—Random Sampling:* To illustrate the performance of the proposed EM-based spatial estimation algorithm, we conduct a first experiment where we assess the accuracy of the estimations as a function of the percentage of skipped variability measurements. The main steps of our experiment are given as following.

1) We process all original wafers and intentionally delete from each wafer a fixed percentage of its measurements. The locations of these deleted measurements are random both within each wafer and across all wafers.
2) We run the EM algorithm to estimate the expected values of all deleted measurements. The EM algorithm only uses the modified wafers; i.e., it has no information about the original set of measurements.
3) We compare the EM estimations against the measurements from the original wafers.

We report in Table I the main comparison results of the EM algorithm as a function of the percentage of the measurements deleted. We report these comparison results for both the PSRO and leakage measurements. These results include: 1) the *normalized average estimation error* where the estimation error is defined as the absolute difference between an estimation and its true value normalized with respect to the true value; 2) the *normalized standard error* which is the standard deviation of the estimation errors normalized to the standard deviation of the measurements; 3) the *estimation error threshold* where 50% of the all estimations lie below it (i.e., median error); and 4) the estimation error where 95% of the all estimations lie below it. For clarification, the first line in the table indicates that the average estimation error is 0.59% when 10% of the measurement are deleted, and that 50% of the estimation errors are below 0.36%, and that 95% of the estimation errors are below 1.85%. For leakage power measurements, we apply the appropriate Box–Cox transformations before the execution of the EM algorithm, and then apply the inverse transformations after the EM algorithm computes the estimated measurements.

The results in Table I lead to a number of insightful conclusions on the performance of the proposed EM algorithm

TABLE I
STATISTICS OF ESTIMATION ERRORS FOR PSRO AND LEAKAGE SKIPPED MEASUREMENTS USING RANDOM SAMPLING WITHIN AND ACROSS ALL WAFERS

Skipped	PSRO				Leakage			
	Normalized Avg. Error	Normalized Std. Dev.	Error Threshold 50%	Error Threshold 95%	Normalized Avg. Error	Normalized Std. Dev.	Error Threshold 50%	Error Threshold 95%
10%	0.59%	1.65%	0.36%	1.84%	3.91%	0.17%	2.59%	11.69%
20%	0.69%	1.73%	0.43%	2.15%	4.91%	0.97%	3.02%	13.37%
30%	0.71%	1.79%	0.44%	2.17%	4.88%	0.21%	3.15%	14.93%
40%	0.73%	1.85%	0.46%	2.24%	5.24%	0.24%	3.36%	16.10%
50%	0.75%	1.92%	0.49%	2.21%	5.53%	0.26%	3.54%	17.12%
60%	0.77%	1.95%	0.51%	2.27%	6.04%	0.31%	3.85%	18.87%
70%	0.86%	2.00%	0.59%	2.45%	7.04%	0.89%	4.29%	20.70%
80%	1.06%	2.17%	0.78%	2.86%	9.56%	2.11%	5.75%	24.97%
90%	1.98%	3.30%	1.50%	5.45%	16.44%	1.75%	11.22%	44.20%

We report the following metrics: normalized average estimation error (normalized average error), normalized standard error (normalized standard deviation), and the estimation error threshold (error threshold).

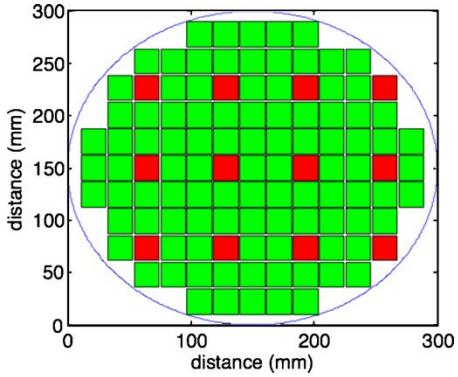

Fig. 9. Sampling plan across all wafers with a pitch of 3. Red locations are sampled. Green locations are skipped.

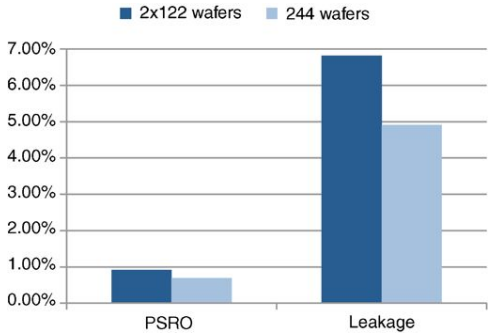

Fig. 10. Error in estimation as function of the available number of wafers.

as a spatial estimator.

1) The tight values in estimation errors demonstrate that the proposed EM algorithm provides very accurate estimation results even when large numbers of the measurements are skipped. Thus, it is possible to used the proposed method to drastically cut down the required parametric measurements.
2) The true values of the vast majority (95%) of the skipped PSRO measurements lie within a tight range of the estimations provided by the EM algorithm which further confirmins the accuracy of the proposed method.
3) While the leakage estimations appear to be less accurate than the estimations of the PSRO measurements in absolute values, the errors in leakage estimation measurements are relatively better as measured by the normalized standard deviation. That is, the spread in leakage estimation errors normalized by the overall spread in leakage measurements is much smaller than the spread of PSRO estimation errors normalized by the overall spread in PSRO measurements. In general, it is necessary to be careful when power data transformations are applied because small errors in estimations might be magnified after the appropriate inverse data transformations are applied.

2) *Second Experiment—Identical Sampling:* In our first experiment the skipped or missing measurements were random across all wafers, and thus, accurate estimation is possible as the locations of the missing measurements of one wafer are likely to have measurements in some other wafers. In some cases (e.g., in process control monitoring), the locations of the sampled measurements and the skipped locations are the same across all wafers. For example, Fig. 9 shows a sampling plan with a *pitch* of 3. In this case, to estimate the electrical characteristics at wafer locations that are not sampled, it is necessary to know the mean (μ) and covariance structure (Σ) in order to utilize (16). The mean and covariance structures could be obtained from either previously fully characterized wafers or using empirical models (e.g., variograms) based on historical fabrication data. In this second experiment, we repeat experiment 1 but using identical wafer sampling plans with different pitches. We report the results in Table II. The results show competitive estimation results with the random sampling results (Table I) of the first experiment. In this experiment, we used the known mean and covariance structures before sampling the measurements.

3) *Third Experiment—Impact of Data Set Size:* As a third experiment, we assess the accuracy of EM-based estimation as a function of the available number of wafers. It is expected that the larger the number of available wafer, the more accurate the estimation will be. This experiment attempts to quantify this intuitive reasoning. For this experiment, we delete 20% of the measurements as outlined in the first

TABLE II
STATISTICS OF ESTIMATION ERRORS FOR PSRO AND LEAKAGE SKIPPED MEASUREMENTS USING IDENTICAL SAMPLING ACROSS ALL WAFERS

Pitch	Skipped	PSRO				Leakage			
		Normalized Avg. Error	Normalized Std. Dev.	Error Threshold 50%	Error Threshold 95%	Normalized Avg. Error	Normalized Std. Dev.	Error Threshold 50%	Error Threshold 95%
2	76.58%	0.56%	1.30%	0.38%	1.55%	4.23%	0.18%	2.63%	13.71%
3	89.19%	0.94%	2.09%	0.59%	3.12%	7.65%	0.39%	4.15%	27.83%
4	94.59%	1.12%	2.07%	0.82%	3.02%	7.91%	0.29%	5.48%	23.32%
5	96.40%	1.38%	2.46%	1.05%	3.68%	9.86%	0.34%	7.06%	28.52%

We report the following metrics: normalized average estimation error (normalized average error), normalized standard error (normalized standard deviation), and the estimation error threshold (error threshold).

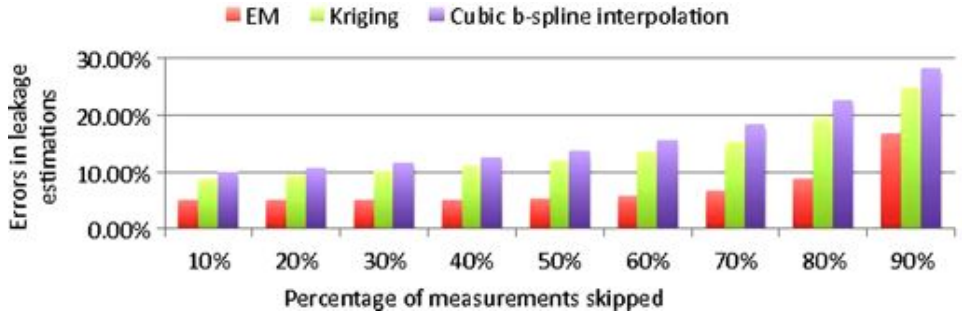

Fig. 11. Performance of EM algorithm vs. Kriging estimation and cubic b-spline interpolation in estimated skipped PSRO parametric test measurements.

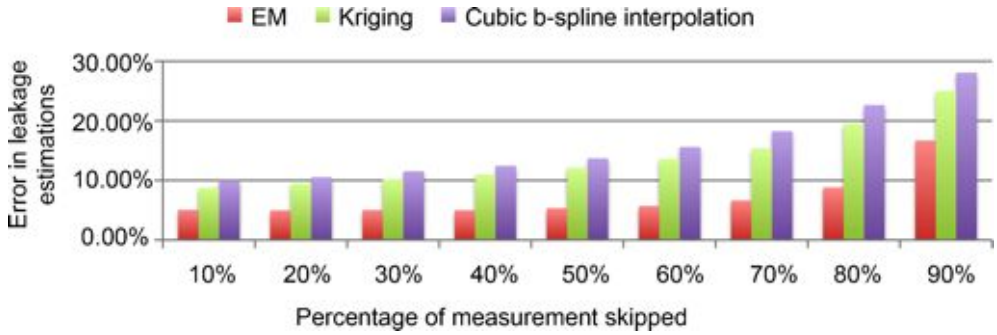

Fig. 12. Performance of EM algorithm vs. Kriging estimation and cubic b-spline interpolation in estimating skipped leakage parametric test measurements.

experiment, and then we execute the EM algorithm on: 1) the entire 244 wafers; and 2) two independent subsets of 122 wafers each. We report the absolute average estimation error as a function of the number of wafers in Fig. 10. The results show that using all wafers toward the estimations reduces the error from an average of 0.92% to 0.69% for the PSRO measurements and from 6.82% to 4.91% for the leakage measurements.

4) *Fourth Experiment—Comparisons:* It is informative to compare the estimation accuracy of the proposed EM-based algorithm against existing variability spatial estimation techniques. For this purpose, we implement both the Kriging algorithm as summarized in Section II-A, and classical cubic b-spline interpolation. Cubic b-spline interpolation estimates the skipped measurements by convolving a 2-D cubic b-spline function with the field measurements. In contrast to the proposed EM method, these methods use only the measurements of a given wafer to interpolate the missing measurements in the same wafer. Such approaches have an advantage of computational runtime but they suffer from low accuracy, especially when the number of missing measurements in a wafer increases. We report in Figs. 11 and 12 the average absolute error of the EM algorithm PSRO and leakage estimations, respectively, vs. the estimations of both the Kriging estimation and cubic b-spline interpolations. The results show a stable, superior performance of the EM algorithm in comparison to existing methods. Kriging estimation and cubic b-spline interpolations deteriorate drastically when only a few of measurements are available.

Our method is more powerful than spatial-based Kriging estimators [18] because it makes use of the entire data set, i.e., measurements from all wafers, to estimate the missing measurements for each wafer. Note that our method does not make any explicit use of the spatial locations on the wafers, so even if the measurements are permuted across all wafers in the same manner (e.g., measurements of two locations across all wafers are swapped), our method will still correctly estimate the missing values. Kriging estimators are good for geospatial studies where there is typically only one set of measurements on a given spatial field; however, in semiconductor fabrication there are typically hundreds and thousands of wafers that are generated roughly using the same process. Our proposed method exploits the variance-covariance structure between the

TABLE III
COMPARISON BETWEEN DIFFERENT ESTIMATION TECHNIQUES

	Expectation-Maximization	Kriging	Spatial Interpolation
Advantages	• extremely accurate estimation • tolerates large volume of missing data • provides uncertainty estimates • works even with lack of spatial structure	• accurate only for few missing measurements • computational runtime • provides uncertainty estimates	• accurate only when few missing measurements are present • computational runtime
Disadvantages	• computational runtime • needs sizable number of wafers to compute accurate results	• accuracy deteriorates with larger volume of missing data • fails when no spatial structure exists	• accuracy deteriorates with larger volume of missing data • fails when no spatial structure exists • no guarantee on estimation certainty • worst estimation accuracy

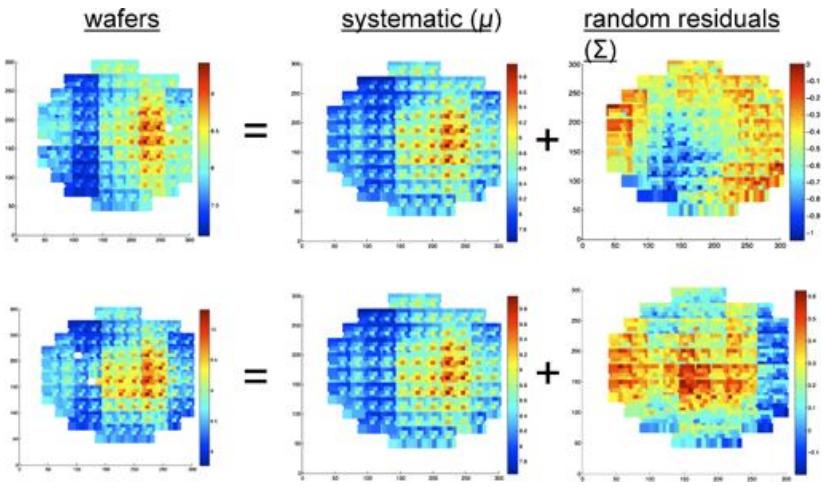

Fig. 13. Decomposition of the process variation into systematic and random residuals.

various measurements on the different wafers to estimate the missing measurements. Table III provides a summary of comparison between the proposed EM-based spatial estimation technique and previous approaches. Furthermore, our model is more accurate and versatile than deterministic approaches [25] that attempt to fit a deterministic mathematical model that is a function of spatial location onto a given data set. In contrast to deterministic approaches, our approach does not require calculating fitting constants that generally change depending on the wafer, and besides estimating the measurements, it provides confidence intervals for the estimations.

V. PROPOSED VARIABILITY DECOMPOSITION TECHNIQUE

Given the completed data set and the MLEs ($\hat{\mu}$ and $\hat{\Sigma}$) of the underlying MVN distribution, the objective of this section is: 1) to carry out further analysis on the spatial structure of both $\hat{\mu}$ and $\hat{\Sigma}$; and 2) to provide hierarchical variability decomposition techniques. The vector $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_p)$ gives the mean value of the parameter under test for every location on the wafers; thus, $\hat{\mu}$ can be regarded as the result of the *systematic* manufacturing sources, and it represents the expected wafer measurements across the entire process. On the other hand, the *residual* of each wafer, $\mathbf{r_j} = \mathbf{w_j} - \hat{\mu} = (x_{j1} - \hat{\mu}_1, x_{j2} - \hat{\mu}_2, \ldots, x_{jp} - \hat{\mu}_p)$, can be considered as the result of the process variations that impact each wafer differently. These residuals form a spatial field with a correlation structure between its variables at the different p locations. This correlation structure is captured by the covariance matrix $\hat{\Sigma}$ of the MVN distribution. Fig. 13 shows the measurements of two wafers broken into two parts: 1) the $\hat{\mu}$ part which is shared between the two wafers (and any other wafer as well); and 2) the random residual part which is unique for every wafer. Thus, a measurement $x_{ji} = \mu_i + r_{ij}$ on wafer j and location (or die) i is the sum of two numbers that are contributed from the systematic and variability sources.

A. Variability Decomposition Using Nested ANOVA

To decompose the variability sources into a hierarchy of lot-to-lot, wafer-to-wafer, die-to-die and within-die variations, we consider a four-factor nested ANOVA model. If r_{lkji} denotes the random residual at site $i \in \{1, \cdots s\}$ of die $j \in \{1, \ldots, p\}$ located on wafer $k \in \{1, \ldots, w\}$ of lot $l \in \{1, \ldots, L\}$ then the nested ANOVA model decomposes variability as

$$r_{lkji} = \tau_l + \beta_{k(l)} + \gamma_{j(kl)} + \epsilon_{i(jkl)} \tag{18}$$

where τ_l is the lot effect, and $\beta_{k(l)}$ is the wafer effect nested under the lth lot effect, $\gamma_{j(kl)}$ are the die j effect nested under the kth die and lth wafer effects, and finally $\epsilon_{i(jkl)}$ is the site ith effect nested under the jth die, the kth wafer, and the lth lot effects. The sum of squares due to lot-lot variability is given by

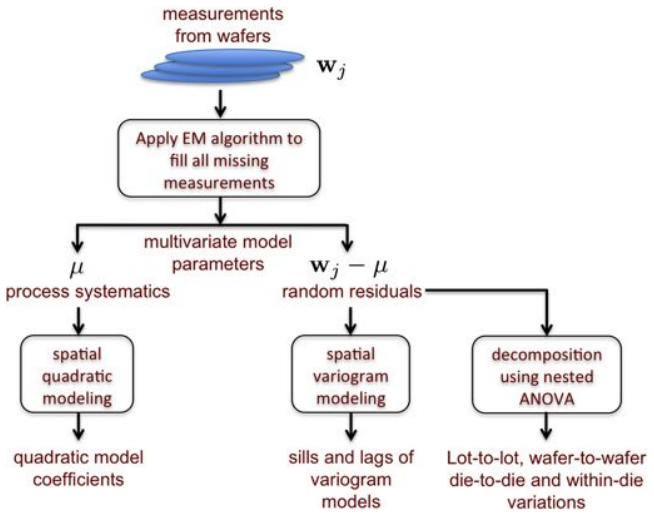

Fig. 14. Proposed flow for variability decomposition.

$$\mathrm{SS}_{lot-to-lot} = w \times p \times s \sum_{l} (\bar{r}_{l...} - \bar{r}_{....})^2 \tag{19}$$

$$= w \times s \times p \sum_{l} (\bar{r}_{l...})^2 \tag{20}$$

with $L-1$ degrees of freedom and where $r_{l...} = \sum_k \sum_j \sum_i r_{lkji}$. Note that (19) has been simplified to (20) because we already decomposed the measurement data into systematic and residual variability sources which leads to residuals with a mean of zero. The sum of squares due to wafer-to-wafer variability is given by

$$\mathrm{SS}_{wafer-to-wafer} = p \times s \sum_{l} \sum_{k} (\bar{r}_{lk..} - \bar{r}_{l...})^2 \tag{21}$$

with $L \times (w-1)$ degrees of freedom and where $r_{lk..} = \sum_j \sum_i r_{lkji}$. The sum of squares of due to die-to-die variability

$$\mathrm{SS}_{die-to-die} = s \sum_{l} \sum_{k} \sum_{j} (\bar{r}_{lkj.} - \bar{r}_{lk..})^2 \tag{22}$$

with $L \times w \times (p-1)$ degrees of freedom and $r_{lkj.} = \sum_i r_{lkji}$. Finally within-die variability is given by

$$\mathrm{SS}_{within-die} = \sum_{l} \sum_{k} \sum_{j} \sum_{i} (r_{lkji} - \bar{r}_{lkj.})^2 \tag{23}$$

with $p \times L \times w \times (s-1)$ degrees of freedom. The total variability $\mathrm{SS}_T = \sum_l \sum_k \sum_j \sum_i (r_{lkji} - \bar{r}_{....})^2 = \sum_l \sum_k \sum_j \sum_i r_{lkji}^2$ with $Lwps - 1$ degrees of freedom. Thus, the contribution of lot-to-lot variability is equal to $\mathrm{SS}_{lot-to-lot}/\mathrm{SS}_T$; the contribution of wafer-to-wafer variability is equal to $\mathrm{SS}_{wafer-to-wafer}/\mathrm{SS}_T$, the contribution of die-to-die variability is equal to $\mathrm{SS}_{die-to-die}/\mathrm{SS}_T$, and finally the within-die contribution is $\mathrm{SS}_{within-die}/\mathrm{SS}_T$. We compute these contributions for our data set after filling all missing values in accordance with the flow of Fig. 14 and give the results in the pie chart given in Fig. 15.

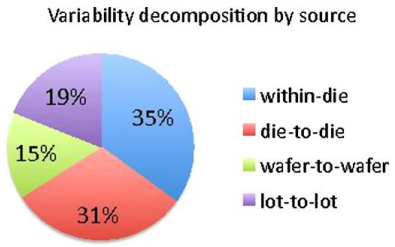

Fig. 15. Budgeting the contribution of process variations from within-die, die-to-die, wafer-to-wafer, and lot-to-lot sources.

B. Spatial Analysis of Systematic Sources

The vector $\hat{\mu}$ gives the mean of the PSRO frequency measurements at each wafer location as illustrated in Fig. 16(a). The spatial dependency in $\hat{\mu}$ can be observed by plotting the values of $\hat{\mu}$ as a function of the radius from the center of the wafer as shown in Fig. 16(b). The plot shows that the average values of the measurements generally decreases as the distance from the center increases. This systematic dependency can be also expressed by a quadratic function where the systematic component u_j at location (x, y) can be captured as $\mu_j = ax^2 + by^2 + cxy + dx + ey + f$, where a, b, c, d, e, and f are constants that can be found using least square fitting [25].

C. Spatial Analysis of Random Sources

To reveal insights into the wafer-level spatial structure of the residuals, we propose using *experimental variograms*. Variograms can reveal any spatial trends in the random variations. Variograms have been previously proposed in the literature to analyze the spatial trends of within-die process variations [13], [18], [28]. In experimental variograms plots, the variance in measurements is plotted as function of the distance, or *lag* h, between them. Since, we have subtracted the mean $\hat{\mu}$ from the measurements, we can directly express the variogram function of the residuals of some wafer j as follows:

$$\gamma_j(h) = \frac{1}{2N_j(h)} \sum_{N_j(h)} (r_j(l) - r_j(l+h))^2 \tag{24}$$

where $r_j(l)$ is the residual of wafer j at location l, $r_j(l+h)$ is the residual of wafer j at location $l+h$, and $N_j(h)$ gives the number of measurements that are at distance h from each other on wafer j. If there is a spatial structure in the data, then we would expect nearby measurements to have similar values, and thus, $\gamma_j(h)$ would be close to zero in this case. As the distance between the measurements increases, the measurements would be more independent, and thus, $\gamma_j(h)$ would increase (potentially leveling up at a particular value). Fig. 17 gives the variograms of wafers from three different lots. The variograms reveal a spatial correlation structure in the random residuals, where the independency in the measurements increases exponentially (or equivalently the dependency between the measurements decays exponentially) as the distance between the measurements increases. The computed experimental variograms show the same trends as the synthetic theoretical variograms plotted earlier in Fig. 1. More interestingly, Fig. 17 shows that variograms of wafers

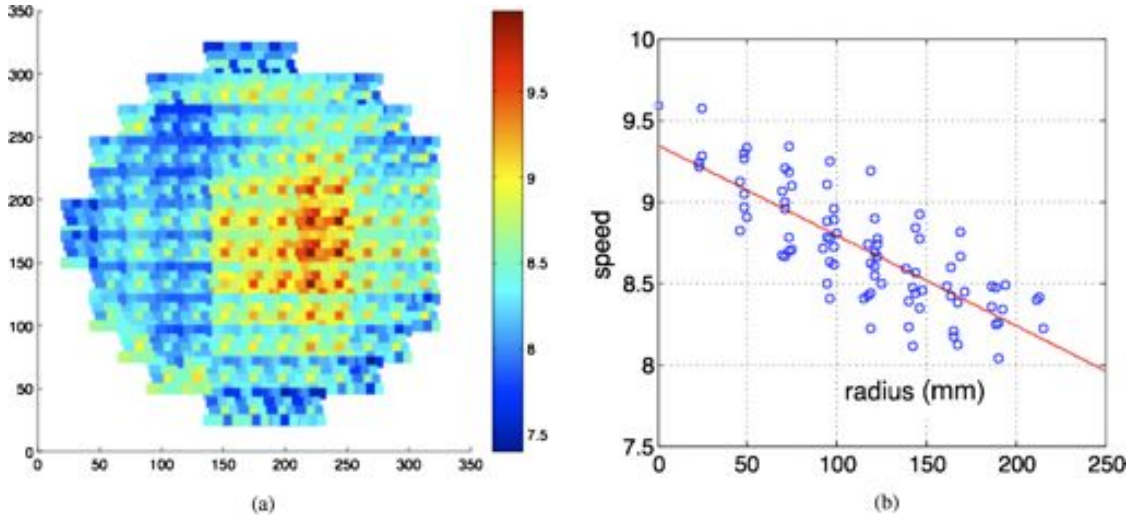

Fig. 16. (a) Mean of the parametric measurements ($\hat{\mu}$) as a function of its location on the wafer. (b) Average value (speed) of the parameter under test as a function of the distance (radius) from the center of the wafer.

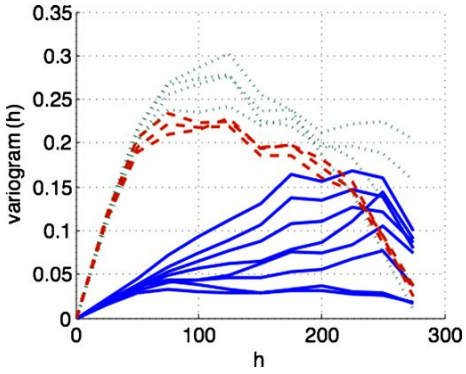

Fig. 17. Variograms for the wafers in the first three lots. The first lot variograms are given by the solid blue lines. The second lot variagrams are given by the red dashed lines. The third lot variograms are given by the green short dotted lines.

that come from the same lot share shape similarities more than variograms of wafers in different lots.

VI. Conclusion and Future Work

In this paper, we proposed a novel modeling technique to characterize process variability parametric measurements based on statistical multivariate techniques. We proposed using the expectation-maximization algorithm to estimate the expected values of the missing test measurements, and to accurately estimate the statistical model parameters. Using chi-square plots, we proposed techniques to verify the accuracy of our model and to detect any outliers. We also carried out further analysis to reveal spatial trends in the systematic and random sources of the variations. We elucidated the trade-off between direct measurements and estimation as a function of the number of available measurements, and contrasted the performance of our proposed method against other spatial estimation techniques such as geospatial Kriging estimation and traditional cubic b-spline interpolation. Finally, we proposed a variability decomposition flow that uses our spatial estimation techniques together with nested ANOVA method, to split the variability into lot-to-lot, wafer-to-wafer, die-to-die, and within-die components. We quantified each of these components and provided analysis techniques to uncover their spatial structure.

Future work includes researching further applications for our techniques. For example, one possible application is part-average testing (PAT) which is used to remove electronic parts with abnormal characteristics from the semiconductors supplied to the automative electronic industry. Most PAT applications assume Gaussian distribution for the electrical measurements. In case the electrical measurements (e.g., leakage) are not Gaussian, one possible future work is to incorporate the Box–Cox transformation into PAT procedures.

Acknowledgment

We thank the anonymous reviewers for their comments which improved many sections of this paper.

References

[1] M. Bhushan, A. Gattiker, M. Ketchen, and K. K. Das, "Ring oscillators for CMOS process tuning and variability control," *IEEE Trans. Semicond. Manuf.*, vol. 19, no. 1, pp. 10–18, Feb. 2006.

[2] D. Boning and S. Nassif, "Models of process variations in device and interconnect," in *Design of High-Performance Microporcessor Circuits*, A. Chandrakasan, W. J. Bowhill, and F. Cox, Eds., 1st ed. Piscataway, NJ: IEEE Press, 2001, pp. 98–115.

[3] S. Borkar, "Designing reliable systems from unreliable components: The challenges of transistor variability and degradation," *IEEE Micro*, vol. 25, no. 6, pp. 10–16, Nov.–Dec. 2005.

[4] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. Design Autom. Conf.*, 2003, pp. 338–342.

[5] K. Bowman, S. Duvall, and J. Meindl, "Impact of die-to-die and within-die parameter fluctations on the maximum clock frequency distribution for gigascale integration," *IEEE J. Solid State Electron.*, vol. 37, no. 2, pp. 183–190, Feb. 2002.

[6] G. Box and D. Cox, "An analaysis of transformations," *J. Royal Statistical Soc., Ser. B*, vol. 26, no. 2, pp. 211–252, 1964.

[7] J. Cain and C. Spanos, "Electrical linewidth metrology for systematic CD variation characterization and causal analysis," in *Proc. SPIE*, vol. 5038. 2003, pp. 350–361.
[8] Y. Cao, P. Gupta, A. B. Kahng, D. Sylvester, and J. Yang, "Design sensitivities to variability: Extrapolation and assessments in nanometer VLSI," in *Proc. IEEE ASIC/SoC Conf.*, 2002, pp. 411–415.
[9] C. Cho, D. Kim, J. Kim, J. Plouchart, D. Lim, S. Cho, and R. Trzcinski, "Decomposition and analysis of process variability using constrained principal component analysis," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 1, pp. 55–62, Feb. 2008.
[10] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statistical Soc., Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
[11] P. Friedberg, Y. Cao, J. Cain, R. Wang, J. Rabaey, and C. Spanos, "Modeling within-die spatial correlation effects for process-design co-optimization," in *Proc. Int. Symp. Quality Electron. Design Autom.*, 2005, pp. 516–521.
[12] A. Gattiker, "Unraveling variability for process/product improvement," in *Proc. Int. Test Conf.*, 2008, pp. 1–9.
[13] B. Hargreaves, H. Hult, and S. Reda, "Within-die process variations: How accurately can they be statistically modeled?" in *Proc. Asia South Pacific Design Autom. Conf.*, 2008, pp. 524–530.
[14] *International Technology Roadmap for Semiconductors*. (2007) [Online]. Available: http://public.itrs.net
[15] J. Jin and H. Guo, "ANOVA method for variance component decomposition and diagnosis in batch manufacturing processes," *Int. J. Flexible Manuf. Syst.*, vol. 15, no. 2, pp. 167–186, 2003.
[16] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Englewood Cliffs, NJ: Prentice Hall, 2007.
[17] H. R. Lindman, *Analysis of Variance in Experimental Design*. Berlin, Germany: Springer-Verlag, 1992.
[18] F. Liu, "A general framework for spatial correlation modeling in VLSI design," in *Proc. Design Autom. Conf.*, 2007, pp. 817–822.
[19] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.
[20] A. Nieuwoudt and Y. Massoud, "Assessing the implications of process variations on future carbon nanotube bundle interconnect solutions," in *Proc. Int. Symp. Quality Electron. Design*, 2007, pp. 119–126.
[21] M. Orshansky, L. Milor, and C. Hu, "Characterization of spatial intrafield gate CD variability, its impact on circuit performance, and spatial mask-level correction," *IEEE Trans. Semicond. Manuf.*, vol. 17, no. 1, pp. 2–11, Feb. 2004.
[22] M. Orshansky and S. Nassif, *Design or Manufacturability and Statistical Design: A Constructive Approach*. Berlin, Germany: Springer, 2007.
[23] S. Reda and S. Nassif, "Analyzing the impact of process variations on parametric measurements: Novel models and applications," in *Proc. Design, Autom. Test Eur. Autom.*, 2009, pp. 375–380.
[24] O. Schabenberger and C. A. Gotway, *Statistical Methods for Spatial Data Analysis*, 1st ed. London, U.K.: Chapman and Hall, 2004.
[25] B. E. Stine, E. Chang, D. S. Boning, and J. E. Chung, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," *IEEE Trans. Semicond. Manuf.*, vol. 10, no. 1, pp. 24–41, Feb. 1997.
[26] D. Sylvester, K. Agarwal, and S. Shah, "Variability in nanometer CMOS: Impact, analysis, and minimization," *VLSI J. Integr.*, vol. 41, no. 3, pp. 319–339, 2008.
[27] B. Wong, A. Mittal, Y. Cao, and G. W. Starr, *Nano-CMOS Circuit and Physical Design*. New York: Wiley-Interscience, 2004.
[28] J. Xiong, V. Zolotov, and L. He, "Robust extraction of spatial correlation," in *Proc. Int. Symp. Physical Design*, 2006, pp. 2–9.

Sherief Reda (M'06) received the B.S. (distinction with Honors) and M.S. degrees in electrical and computer engineering from Ain Shams University, Cairo, Egypt, in 1998 and 2000, respectively, and the Ph.D. degree in computer engineering and science from the University of California, San Diego, in 2006.

He is currently an Assistant Professor with the Division of Engineering, Brown University, Providence, RI. His current research interests include design automation and test of integrated circuits, reconfigurable computing and embedded systems, organization and architecture of computational systems, and synergies between biological and computational systems. He has about 50 refereed conference and journal papers in his research areas.

Dr. Reda served as a member of technical program committees for many IEEE/ACM conferences including ICCAD, ICCD, ASPDAC, GLSVLSI, and SLIP. He received a number of awards and acknowledgments, including the Best Paper Award in DATE 2002, a hot article in *Operations Research letters* in 2004, the First Place Award in ISPD VLSI Placement Contest in 2005, Best Paper Nominations in ICCAD 2005 and ASPDAC 2008, Brown's Salomon Award in 2008, and the NSF CAREER Award in 2010.

Sani R. Nassif (F'08) received the Bachelors degree from the American University of Beirut, Beirut, Lebanon, in 1980, and the Masters and Ph.D. degrees from Carnegie Mellon University, Pittsburgh, PA, in 1981 and 1986, respectively.

Until 1996, he was with Bell Laboratories, Allentown, PA. Since 1996, he has been with the Austin Research Laboratory, IBM Corporation, Austin, TX, where he currently manages the Department of Silicon Analytics. He has 44 patents, and has authored numerous conference and journal publications.

Dr. Nassif had received the IEEE Trans. CAD, ICCAD, DAC, ISQED, and ICCD Best Paper Awards. He has authored invited papers to the ISSCC, IEDM, ISLPED, HOTCHIPS, and CICC, and given keynote and plenary presentations at Sasimi, ESSCIRC, BMAS, SISPAD, SEMICON, and PATMOS. He is a member of the IBM Academy of Technology and ACM.