Analyzing the Impact of Process Variations on Parametric Measurements: Novel Models and Applications

Sherief Reda
Division of Engineering
Brown University
Providence, RI 02912
Email: sherief_reda@brown.edu

Sani R. Nassif
Austin Research Laboratory
IBM Corporation
Austin, TX 78758
Email: nassif@us.ibm.com

***Abstract*—In this paper we propose a novel statistical framework to model the impact of process variations on semiconductor circuits through the use of process sensitive test structures. Based on multivariate statistical assumptions, we propose the use of the expectation-maximization algorithm to estimate any missing test measurements and to calculate accurately the statistical parameters of the underlying multivariate distribution. We also propose novel techniques to validate our statistical assumptions and to identify any outliers in the measurements. Using the proposed model, we analyze the impact of the systematic and random sources of process variations to reveal their spatial structures. We utilize the proposed model to develop a novel application that signi cantly reduces the volume, time, and costs of the parametric test measurements procedure without compromising its accuracy. We extensively verify our models and results on measurements collected from more than 300 wafers and over 25 thousand die fabricated at a state-of-the-art facility. We prove the accuracy of our proposed statistical model and demonstrate its applicability towards reducing the volume and time of parametric test measurements by about $2.5 - 6.1\times$ at absolutely no impact to test quality.**

I. Introduction

Manufacturing process variations manifest themselves in gate length variations, line edge roughness, dopant fluctuations, and variations in the widths and heights of interconnects [2], [9]. These variations impact the key electrical parameters of semiconductor devices and interconnects, leading to variations in performance and power of fabricated chips. *Inter-die* variations account for variations that arise between different die in the same wafer or across different wafers, while *intra-die*, or *within-die*, variations account for variations that arise within the same die or more generally within a reticle field. Within-die variations are caused, for example, by lens aberrations and lithographic hot spots, while inter-die variations are caused by wafer-level physical phenomena as chemical-mechanical polishing and photoresist coating mechanisms. Process variations typically have *systematic* and *random* sources [2]. Systematic variations impact different die or wafers in a deterministic manner, while random variations are relatively unique to each die/wafer/lot. To cope with process variations during design and/or manufacturing, it is first important to analyze and develop models that explain process variability.

Given the parametric test measurements of process sensitive test structures, our overarching objectives are (1) to develop a statistical model that analyzes and summarizes the obtained measurements, and (2) to utilize the developed model to devise applications that are of benefit for both designers and process engineers. We summarize our contributions as follows.

- Based on multivariate statistical techniques, we propose using the Expectation-Maximization (EM) algorithm to estimate accurately any missing values in the test measurements. The EM algorithm enables maximum-likelihood estimation techniques to produce accurate statistical parameters for the proposed multivariate distribution. We also describe a procedure to verify the accuracy of our multivariate model and to detect any outliers in the measurements.
- We propose techniques to analyze the systematic and random sources of process variations and to reveal the spatial structure in these sources. We use variograms to model wafer-level random spatial variability trends.
- Based on the proposed model, we develop a new application to reduce significantly the time and costs required to carry out the parametric test measurements. The proposed approach carries out the test measurements on few selected sites and then uses the EM algorithm to estimate and classify accurately the measurements at all skipped sites. The proposed application drastically reduces the volume of the measurements that need to be conducted.
- Using thousands of measurements from process sensitive test structures, we validate the proposed techniques and demonstrate their applicability and accuracy. The proposed parametric test reduction application yields test cost reduction factors of about $2.5 - 6.1\times$ at absolutely no impact to variability characterization quality.

The organization of this paper is as follows. In Section II we provide background information on the test characterization procedure carried out to procure the parametric measurements. In Section III we develop the proposed multivariate statistical framework and describe the proposed expectation-maximization technique. In Section IV we decompose the observed variations into systematic and random sources and explore the spatial structure of each of them. Based on our model, we propose novel applications in Section V, and finally in Section VI we summarize the main conclusions of this work.

978-3-9810801-5-5/DATE09 © 2009 EDAA

II. Background on the Measurements

Parametric measurements occur throughout the fabrication process to track its quality as the manufacturing steps unfold [8], [4], [1]. These measurements are used to probe the various process characteristics such as critical dimensions, transistor thresholds, timing and leakage. The parametric data set we consider in this paper consists of frequency measurements from 65 nm Process Sensitive Ring Oscillators (PSROs). The PSROs are embedded within a production chip, and they are used as a quick 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]. To protect the sensitivity of our data, we do not disclose the number of stages of the PSRO. Each instance of the production chip holds 14 PSROs that are spatially organized along a grid that spans the die s area. Our data set consist of measurements from 348 wafers that span 23 wafer lots. Each wafer contains slightly more than 100 die, and the reticle field consists of only one die. In each wafer, a good number of the 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 number of wafers have received "special" processing steps in comparison to the rest of the population of the wafers.

III. The Proposed Statistical Model

In this section we describe the proposed multivariate statistical modeling framework.

- In Subsection III-A, we describe the foundations of our multivariate statistical techniques.
- To address the problem of missing measurements and its impact on the calculations of the statistical model parameters, we propose using the expectation-maximization algorithm 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 work we make only one fundamental assumption about the parametric measurements: we assume that the data set comes from a multivariate normal (MVN) distribution with potentially a number of outliers. The measurements obtained from a wafer are considered as an *observation* that is mathematically represented as a random vector $\mathbf{w}$. Each observation $\mathbf{w}$ 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 conducted within each die. In the MVN distribution, the probability density function for a random vector $\mathbf{w}$ is equal to

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

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. Let $\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_n}$ denote the parametric measurements from some n wafers, then the joint density function of all observations is the product of marginal normal densities:

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

Equation (2), considered 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 where

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{w_j} \tag{3}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n} (\mathbf{w_j} - \hat{\mu})(\mathbf{w_j} - \hat{\mu}) \tag{4}$$

are the *Maximum Likelihood Estimators (MLE)* of μ and Σ respectively [6]. Our objective is to estimate accurately the statistical parameters μ and Σ of the MVN $N_p(\mu, \Sigma)$. In our case, this estimation is complicated by the fact that each observation $\mathbf{w_j}$ could potentially have measurements with *missing values*. Ignoring the missing values will produce wrong estimates for $\hat{\mu}$ and $\hat{\Sigma}$ that could significantly deviate from the true μ and Σ of the process.

B. The Expectation-Maximization Algorithm

To compute the MLE distribution parameters from incomplete data, we propose utilizing the *Expectation Maximization (EM) algorithm* [3]. The EM algorithm enables parameter estimation in multivariate statistical models with incomplete data. The algorithm is an iterative procedure for estimating the 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 typically the *Expectation* step and the *Maximization* step. A general framework for the EM algorithm is as follows.

- Initialize the MLEs of distribution parameters ($\hat{\mu}$ and $\hat{\Sigma}$).
- Repeat until convergence:
  1) *E-Step*: estimate the **E**xpected value of the missing measurements, given the current MLEs ($\hat{\mu}$ and $\hat{\Sigma}$) of the distribution.
  2) *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

$$\text{mean/expectation of } \mathbf{w_j^u} = \hat{\mu}^{\mathbf{u}} + \hat{\Sigma}_{u,k}\hat{\Sigma}_{k,k}^{-1}(\mathbf{w_j^k} - \hat{\mu}^{\mathbf{k}}). \quad (5)$$

Equation (5) is the key method used for the E-Step, and it can be intuitively explained as follows (formal derivations can be found in [3], [6]). 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 missing measurements of wafer j from their mean $\hat{\mu}^{\mathbf{u}}$. This term is estimated to be equal to the deviations of the known measurements of wafer j from their mean $(\mathbf{w_j^k} - \hat{\mu}^{\mathbf{k}})$ multiplied by some weight. This weight is the product of:

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

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

$$\text{covariance of } \mathbf{w_j^u} = \hat{\Sigma}_{u,u} - \hat{\Sigma}_{u,k}\hat{\Sigma}_{k,k}^{-1}\hat{\Sigma}_{k,u}. \quad (6)$$

The diagonal elements of the covariance matrix of Equation (6) give the variances in the estimations of the missing measurements. 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.

Figure 1 shows an example where the EM algorithm fills the missing measurements of one of the wafers. The color of a measurement gives its value (or speed in this case). Visual inspection shows that predicted values seem to "fit" within the range of the rest of the measurements. Because the predicted values were missing in the first place, there is no way to verify

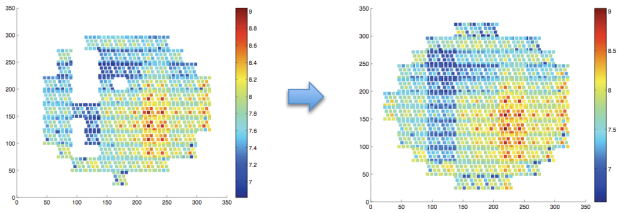

Fig. 1. An example of filling missing measurements on wafer using the EM algorithm.

the accuracy of the algorithm in this setting. However, we will thoroughly demostrate the accuracy of our approach in Section V by deleting existing measurements and then use the EM algorithm to predict them.

Our method is more powerful than spatial-based Kriging estimators [7] 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 some "adversary" has permuted the measurements on all wafers in the same manner, 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 various measurements on the different wafers to estimate the missing measurements. Furthermore, our model is more accurate and versatile that deterministic approaches [10] 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 variances for the estimations. These variance provide "safety nets", or formally speaking, *confidence intervals* for the estimations as we will further elucidate in Section V.

C. Verifying the MVN assumption

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

$$d_j^2 = (\mathbf{w_j} - \mu)\, \Sigma^{-1}(\mathbf{w_j} - \mu), \quad \mathrm{j} = 1, \ldots, n. \quad (7)$$

Essentially Equation (7) summarizes all the measurements of a wafer by a single number. It can be shown [6] that

$$d_j^2 = (\mathbf{w_j} - \mu)\, \Sigma^{-1}(\mathbf{w_j} - \mu) = \sum_{i=1}^{p} z_i^2, \quad (8)$$

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 $(\mathbf{w_j} - \mu)\, \Sigma^{-1}(\mathbf{w_j} - \mu)$ has a *chi-square distribution* with p-degrees of freedom. For a

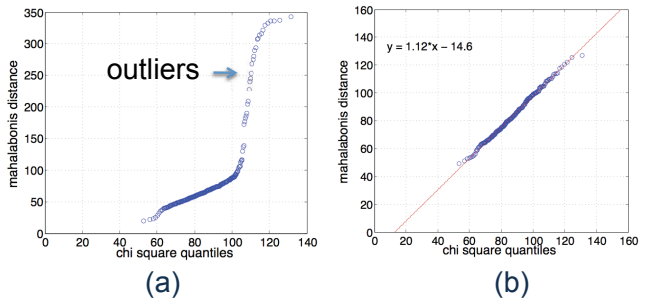

Fig. 2. Subfigure (a) shows the chi-square plot for the entire data set. Subfigure (b) gives the the chi-square plot after removing the few outlier wafers. The x-axis of the plots gives the chi-square quantiles, and the y-axis gives the Mahalanobis distance quantiles.

data set of $\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_n}$ observations, we can assess its multivariate normality by plotting the *chi-square plot*. We use the following procedure to construct the chi-square plot.

a. Use Equation (7) to calculate the squared Mahalanobis distances for all observation vectors. Then order the distances from smallest to largest $d^2_{(1)} \leq d^2_{(2)} \leq \ldots \leq d^2_{(n)}$.
b. Graph the pairs $(q_p((j-1/2)/n), d^2_j)$, where $q_p((j-1/2)/n)$ is the $100((j-1/2)/n)$ 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 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 parametric 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 Equation (7) for unusually large numbers. In a chi-square plot, the outliers would be the points farthest from the origin.

Figure 2.a gives the chi-square plot for our entire data set of measurements after filling the missing values and calculating the MLEs $\hat{\mu}$ and $\hat{\Sigma}$. The x-axis of the plot gives the chi-square quantiles, and the y-axis gives the Mahalanobis distance quantiles. As it clear from the figure, there are points with extreme values that are apart from the rest of the points. Checking these points against the manufacturing recipes of the different wafers confirms that the wafers corresponding to these points received special processing steps. After removing these outlier wafers (37 wafers from a total 348) from our data set, we re-calculate the estimated values of the missing measurements as well as the MLEs $\hat{\mu}$ and $\hat{\Sigma}$. We then re-plot the chi-square plot in Figure 2.b. In contrast to Figure 2.a, Figure 2.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. Analysis of the Systematic and Random Sources of Variations

Given the MLEs $\hat{\mu}$ and $\hat{\Sigma}$ of the MVN distribution of the data set using the methods of Section III, the objective of this section is carry out further analysis on the structure of both $\hat{\mu}$ and $\hat{\Sigma}$. $\hat{\mu}$ gives the mean value of the parameter under test for every location on the wafers; thus, one can think of $\hat{\mu}$ as the result of the *systematic* process variations sources. While the *residual* of each wafer, $\mathbf{w_i} - \hat{\mu}$, can be thought of as the result of the random process variations sources with the covariance that is given by $\hat{\Sigma}$.

Systematic Sources. The vector $\hat{\mu}$ gives the mean of the measurements at each location of the set of tested wafers. Figure 3.a shows the values of $\hat{\mu}$ as a function of their location on the wafer. One can observe the general trend that the measurements towards the center of the wafer are on the average higher in value than the measurements towards the periphery. Furthermore, one can observe the impact of systematic within-die variations which leads to a "repetitive" structure in $\hat{\mu}$. The spatial dependency in $\hat{\mu}$ can be confirmed by plotting the values of $\hat{\mu}$ as a function of the radius from the center of the wafer as shown in Figure 3.b. The plot shows that the average values of the measurements generally decrease in a linear fashion as the distance from the center increases.

Random Sources. One can think of the measurements of a wafer $\mathbf{w_j}$ as the result of the systematic variation sources $\hat{\mu}$ plus some random *field residual* $\mathbf{r_j} = \mathbf{w_j} - \hat{\mu}$ that gives the unique impact of the random lot/wafer/die variations on the measurements of wafer $\mathbf{w_j}$. Figure 4 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 $\mathbf{r_j}$ which is unique for every wafer. While the residuals are random in nature, they have correlations that are quantified by the covariance matrix Σ of the MVN distribution. The contribution of both systematic and random sources towards the total variability can be broken up as given in Figure 5. In our data set, we believe systematic

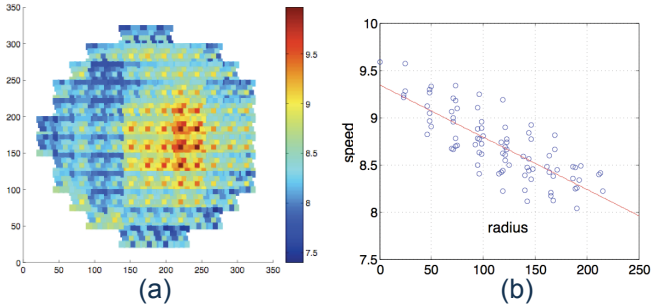

Fig. 3. Subfigure (a) gives the mean of the parametric measurements ($\hat{\mu}$) as a function of its location on the wafer. Subfigure (b) gives the average value (speed) of the parameter under test as a function of the distance (radius) from the center of the wafer.

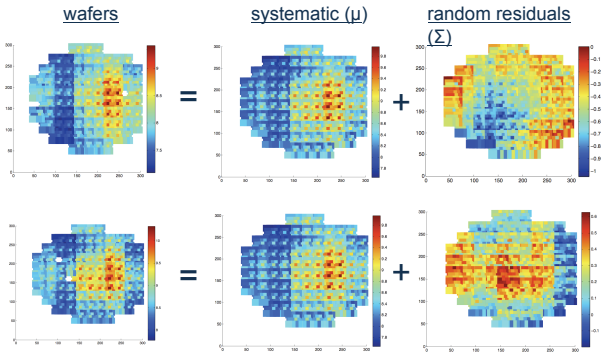

Fig. 4. Decomposition of the process variation into systematic and random residuals.

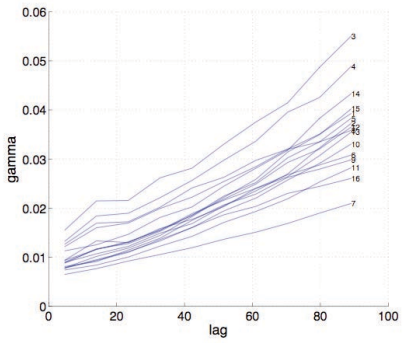

Fig. 6. Variograms for 16 wafers one lot.

effects contribute towards 48% of the total variability across all die, while random effects contribute 52%.

To reveal insights into the wafer-level spatial structure of the residuals, we propose using *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 [11], [7], [5]. In variograms 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

$$\gamma_j(h) = \frac{1}{2N_i(h)} \sum_{N_j(h)} (r_j(u) - r_j(u+h))^2, \qquad (9)$$

where $r_j(u)$ is the residual of wafer j at location u, $r_j(u+h)$ is the residual of wafer j at location $u+h$, and $N_i(h)$ gives the number of measurements that are at distance h from each other on wafer i. If there is a spatial structure in the data, then we would expect that 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). Figure 6 gives the variograms of 16 wafers in one of our lots. The variograms reveal a spatial correlation structure in the random residuals, where the independency in the measurements increases linearly (or equivalently the dependency between the measurements decays linearly) as the distance between the measurements increases. The trend in the decay of the dependency depends on the parameter being measured; different parameters might yield different variogram trends [11].

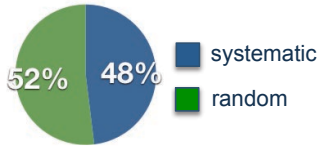

Fig. 5. Budgeting the contribution of process variations from systematic and random sources.

V. Applications: Reducing Parametric Test Volume and Costs

In this section we propose using the EM algorithm to reduce the volume, time and costs of parametric test measurements. Our basic idea is to measure only a few random or psuedo-random sites on the different wafers, and then use the EM algorithm to estimate the results of the measurements at all skipped test sites. For example, Figure 7.a shows the test results if all the sites of a wafer are measured; Figure 7.b shows the test results after skipping the measurements of 50 die; and Figure 7.d shows the test results after skipping the measurements of nearly 70 die. Given a subset of measurements, we execute the EM algorithm to estimate all skipped measurements as well as the MLEs $\hat{\mu}$ and $\hat{\Sigma}$ of the model. For the measurements in Figure 7.b and Figure 7.d, the estimation results are given in Figures 7.c and Figure 7.e respectively. There are two important points to stress:

- While we skip the measurements pseudo-randomly on all wafers, the skipped sites differ for different wafers.
- The EM algorithm does not make any use whatsoever of any earlier characterization results or from any μ and Σ that could have been calculated from previous characterizations. Making use of earlier characterization results can only strengthen the accuracy of our methods.

After the measurements of the skipped test sites on all wafers are estimated, the estimates are compared to the actual measurements which have been "hidden" throughout the estimation process. To assess the accuracy of the estimation, we calculate the absolute error between the estimations of the skipped sites and their actual measurement values. For a test volume and time reduction factor of $2.53\times$ (Figure Figure 7.b), the average absolute prediction error is equal to 0.69%, and for a test volume and time reduction factor of $6.05\times$ (Figure 7.d), the average absolute prediction error is equal to 1.17%.

One of the interesting aspects of the proposed approach is that not only it can estimate skipped measurements, but it also estimates the variances in the estimations as given by Equation (6). The variance of an estimate gives the confidence of the EM algorithm in its estimation of a skipped measurement. The variance enables the calculation of a *confidence interval* where the true measurement of a skipped site would likely

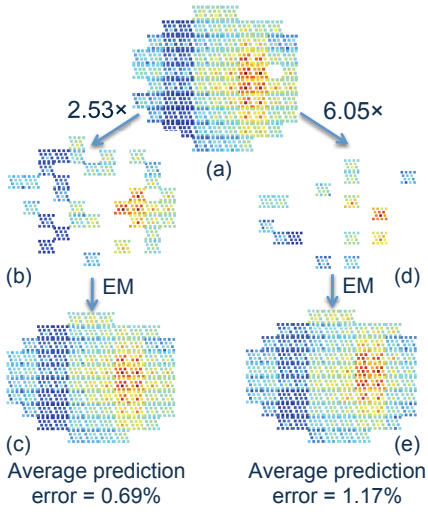

Fig. 7. Estimating skipped parametric test measurements.

to fall. 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 $m_s - 3\ \overline{\sigma_{s,s}}$ and $m_s + 3\ \overline{\sigma_{s,s}}$. Calculating the confidence interval can help in situations where die classification is required. For example, in many occasions after a test characterization measurement is conducted, the value of the measurement is evaluated to see whether the die parameters falls within some certain acceptable manufacturing range $[P_L, P_U]$. The proposed estimation technique can classify a die based on the following three cases:

a. If the estimated range $[m_s - 3\ \overline{\sigma_{s,s}}, m_s + 3\ \overline{\sigma_{s,s}}]$ falls entirely within $[P_L, P_U]$ then the proposed method can declare that the die parameter is acceptable without testing the die.
b. If the estimated range $[m_s - 3\ \overline{\sigma_{s,s}}, m_s + 3\ \overline{\sigma_{s,s}}]$ falls entirely outside $[P_L, P_U]$ then the proposed method can declare that the die parameter is unacceptable without testing the die.
c. If the estimated range $[m_s - 3\ \overline{\sigma_{s,s}}, m_s + 3\ \overline{\sigma_{s,s}}]$ partially overlaps with $[P_L, P_U]$ then the proposed method can *flag* the unmeasured die for measurement to accurately classify it.

Thus, our proposed application not only classifies a die (cases 1 and 2) but it also tells when it thinks that the classification cannot be possible (case 3). To assess the strength of our proposed classification approach, we apply it to 26310 die from our wafers. Instead of testing all die, we initially only test 5351 die for an average of 13 die per wafer. Then the proposed classification method is used to calculate the estimates and variances for all skipped test measurements. The results indicate that 18237 die are classified as either acceptable or unacceptable (cases 1 and 2), and 2722 die are flagged to be further tested (case 3). When we compared the classification results of our approach against the actual test measurements, we found that our approach manages to correctly classify (whether acceptable or unacceptable) all of the 18237 die. Furthermore, after counting all required tests (5351+2722), our proposed method yields a reduction in test measurements by a factor of $3.25\times$ at absolutely no impact to test quality.

VI. Conclusions

In this paper we have proposed a novel modeling technique to characterize process variability based on statistical multivariate techniques. We have proposed using the expectation-maximization algorithm to estimate the values of the missing test measurements, and to accurately estimate the statistical model parameters. Using chi-square plots, we have proposed techniques to verify the accuracy of our model and to detect any outliers. We have also carried out further analysis to reveal the spatial trends in the systematic and random sources of the variations. Based on our model, we have proposed two important applications to estimate the contribution of process variations sources and to reduce the costs and time of parametric tests. Our novel parametric test cost reduction approach uses the EM algorithm to give accurate estimation for any skipped test sites and to flag any sites that need to be measured. Thus, our method provide a natural mechanism to avoid any mischaracterization. Our results demonstrate that it is possible to reduce parametric test time and costs by $2.5 - 6.1\times$ at absolutely no impact to the accuracy of the parametric test procedure.

Acknowledgments. The authors would like to thank Henrik Hult of KTH, Sweden, on earlier discussions regarding the technical material of Section IV.

References

[1] M. Bhushan, A. Gattiker, M. Ketchen, and K. K. Das, "Ring Oscillators for CMOS Process Tuning and Variability Control," *IEEE Trans. Semiconductor Manufacturing*, vol. 19, pp. 10–18, 2006.
[2] D. Boning and S. Nassif, "Models of Process Variations in Device and Interconnect," in *Design of High-Performance Microporcessor Circuits*, 1st ed., A. Chandrakasan, W. J. Bowhill, and F. Cox, Eds. IEEE Press, 2001, pp. 98–115.
[3] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood From Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39(1), pp. 1–38, 1977.
[4] 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 *International Symposium on Quality Electronic Design Automation*, 2005, pp. 516–521.
[5] B. Hargreaves, H. Hult, and S. Reda, "Within-die Process Variations: How Accurately Can They Be Statistically Modeled?" in *Proc. Asia and South Pacific Design Automation Conference*, 2008, pp. 524–530.
[6] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Prentice Hall, 2007.
[7] F. Liu, "A General Framework for Spatial Correlation Modeling in VLSI Design," in *Proc. Design Automation Conference*, 2007, pp. 817–822.
[8] 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 Transactions on Semicondictor Manufacturing*, vol. 17(1), pp. 2–11, 2004.
[9] M. Orshansky and S. Nassif, *Design or Manufacturability and Statistical Design: A Constructive Approach*. Springer, 2007.
[10] 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, pp. 24–41, 1997.
[11] J. Xiong, V. Zolotov, and L. He, "Robust Extraction of Spatial Correlation," in *Proc. International Symposium on Physical Design*, 2006, pp. 2–9.