Handling Discontinuous Effects in Modeling Spatial Correlation of Wafer-level Analog/RF Tests

Ke Huang*, Nathan Kupp†, John M. Carulli Jr.‡, and Yiorgos Makris*
*Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75080
†Department of Electrical Engineering, Yale University, New Haven, CT 06511
‡Texas Instruments Inc., 12500 TI Boulevard, MS 8741, Dallas, TX 75243

***Abstract*—In an effort to reduce the cost of specification testing in analog/RF circuits, spatial correlation modeling of wafer-level measurements has recently attracted increased attention. Existing approaches for capturing and leveraging such correlation, however, rely on the assumption that spatial variation is smooth and continuous. This, in turn, limits the effectiveness of these methods on actual production data, which often exhibits localized spatial discontinuous effects. In this work, we propose a novel approach which enables spatial correlation modeling of wafer-level analog/RF tests to handle such effects and, thereby, to drastically reduce prediction error for measurements exhibiting discontinuous spatial patterns. The core of the proposed approach is a k-means algorithm which partitions a wafer into k clusters, as caused by discontinuous effects. Individual correlation models are then constructed within each cluster, revoking the assumption that spatial patterns should be smooth and continuous across the entire wafer. Effectiveness of the proposed approach is evaluated on industrial probe test data from more than 3,400 wafers, revealing significant error reduction over existing approaches.**

I. Introduction

Current industrial test practice for analog/RF Integrated Circuits (ICs) involves measuring the performances of each fabricated device and comparing to the design specifications prior to shipping it to customers, in order to detect spot defects or excessive process variations due to imperfections in the materials or excursions in the manufacturing process. The excessive cost of specification testing, which requires complex test equipment and elaborate measurement procedures, however, led to the development of various statistical approaches in an effort to reduce this cost. The common theme of such methods is that they employ *die-level* statistical models to approximate the original test set or predict pass/fail labels from a reduced or alternate low-cost set of measurements. For example, the specification test compaction approach described in [1] leverages the correlation amongst specification tests in order to perform only a subset of these tests during production and predict the values of the omitted ones and, thereby, reduce test cost. Similarly, the alternate test approach [2] replaces expensive specification tests by low-cost "alternate tests" specifically designed to be well correlated, through regression models, with specification tests. Along the same lines, in [3], [4], a machine learning-based approach is used to learn classification boundaries which separate passing and failing population of devices in a multi-dimensional space of low-cost measurements.

Recently, an orthogonal direction for leveraging statistical correlation towards reducing test cost has also attracted interest. This time, however, the sought after correlations are *wafer-level*. In other words, instead of completely eliminating some specification tests and predicting them from other low-cost tests, this approach performs these measurements on a sparse subset of die on each wafer, and subsequently uses statistical spatial correlation models in order to predict performance outcomes at unobserved die locations. For example, in [5], the expectation-maximization algorithm is used to estimate spatial wafer measurements by assuming that data comes from a multivariate normal distribution. Similarly, the "Virtual Probe" (VP) approach [6], [7], [8] models spatial variation via a Discrete Cosine Transform (DCT) that performs a frequency domain projection from spatially sampled measurements. Follow-on work described in [9] has demonstrated the utility of VP towards test time and test cost reduction. Along the same lines, recent work described in [10], [11] indicates that using Gaussian Process (GP) models, which were first explored in [12] for capturing spatial correlation, can significantly improve prediction results and computational time as compared to VP models.

While these methods promise high prediction accuracy through spatial correlation models, they rely on certain assumptions. Specifically, both the VP and GP approaches assume that spatial patterns of process variations are continuous and smooth. In reality, however, there exist various localized discontinuous effects which influence the spatial characteristics of wafer-level data and violate these assumptions. These effects may be caused by a variety of sources: multi-site testing, reticle size, chemical aging or crystalline non-uniformity resulting in individual spatial clusters on the wafer map, etc. As a result, in the presence of discontinuous effects, the prediction accuracy of the aforementioned methods for wafer-level spatial correlation modeling deteriorates drastically.

In this work, we propose a method for enhancing spatial correlation modeling of wafer-level tests so that it can handle discontinuous effects. Our method is based on a clustering approach, which partitions the wafer into k clusters, reflecting and capturing the collective impact of discontinuous effects on a measurement across the entire wafer. Subsequently, using the GP approach, spatial models are trained separately within each cluster, instead of a single model across the entire wafer as in [10]. Results on industrial probe test data from more than 3,400 wafers demonstrate significant prediction error improvement.

The remainder of this paper is organized as follows. Section II briefly reviews existing spatial correlation modeling approaches, focusing mainly on spatial interpolation using the

978-3-9815370-0-0/DATE13/©2013 EDAA

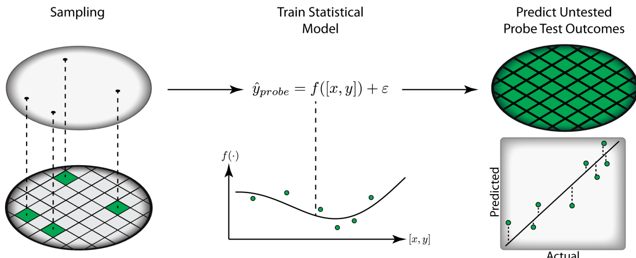

Fig. 1. Overview of wafer measurement spatial interpolation [10].

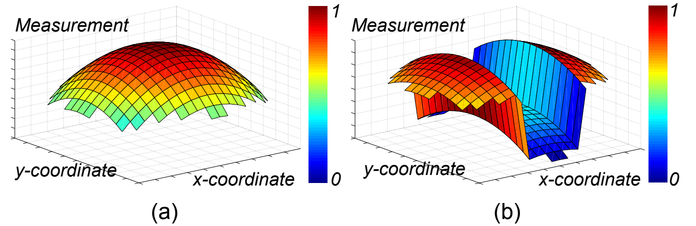

Fig. 2. Spatial wafer measurement data that (a) can be modeled efficiently by GP model and (b) cannot be modeled efficiently by GP model

GP model, which is the basis for our method. Section III discusses the origins of discontinuous effects in spatial wafer measurements. Section IV introduces the proposed method for handling discontinuous effects. Section V provides experimental results, and conclusions are drawn in Section VI.

II. Spatial interpolation of wafer measurements and Gaussian Process model

Recent research on modeling spatial measurement correlation has shown great promise in capturing wafer-level spatial variation and, thereby, reducing test cost [5], [6], [9], [10], [12]. The underlying idea is to collect measurements for a sparse subset of die on each wafer and subsequently train statistical spatial models to predict performance outcomes at unobserved die locations. Figure 1 shows an overview of the general wafer measurement spatial interpolation approach.

In [5], the expectation-maximization (EM) algorithm is used to estimate spatial wafer measurements, assuming that data comes from a multivariate normal distribution. The Box-Cox transformation is used in case data is not normally distributed. The "Virtual Probe" (VP) approach [6] models the spatial variation via a Discrete Cosine Transform (DCT) that projects spatial statistics into the frequency domain. The author of [12] lays the groundwork for applying Gaussian Process (GP) models to spatial interpolation of semiconductor data based on Generalized Least Square fitting and a structured correlation function. As recently shown in [10], [11], using such GP models can dramatically improve both prediction accuracy and computational time, as compared to the VP model.

The GP approach works by extrapolating a function over a Gaussian random field on limited observations [13]. Consider a training set of nt data points $\{m_1, \ldots, m_{nt}\}$ located at the Cartesian coordinate denoted by $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_{nt}\}$, $\mathbf{x} = [x, y]$. Using the GP approach, we define a Gaussian process as a collection of random variables $f(\mathbf{x}_i)$, $i = 1, \ldots, n_t$, for which any finite set of n_s function evaluations $f(\mathbf{x}_j)$, $j = 1, \ldots, n_s$, $n_s \leq n_t$ over the coordinates is jointly Gaussian-distributed. To derive a GP model for regression, we first consider a a noise-free linear model:

$$f(\mathbf{x}) = \phi(\mathbf{x})^\top \mathbf{w} \tag{1}$$

where $\phi(\mathbf{x})$ is a function of $\mathbf{x}$ mapping the input columns into some high dimensional feature space, and $\mathbf{w}$ is the coefficient of the linear model which can be assigned a Bayesian prior such that $\mathbf{w} \sim \mathcal{N}(0, \Sigma_p)$. By assuming the random variables $f(\mathbf{x}_j)$ have mean zero, we can then specify the GP with mean and covariance functions:

$$\mathbb{E}[f(\mathbf{x})] = \phi(\mathbf{x})^\top \mathbb{E}[\mathbf{w}] = 0, \tag{2}$$

$$\begin{aligned} \mathbb{E}[f(\mathbf{x})f(\mathbf{x}')] &= \phi(\mathbf{x})^\top \mathbb{E}[\mathbf{w}\mathbf{w}^\top]\phi(\mathbf{x}') \\ &= \phi(\mathbf{x})^\top \Sigma_p \phi(\mathbf{x}') \end{aligned} \tag{3}$$

It can be shown that the covariance function in (3) can be further written as a kernel function $k(\mathbf{x}, \mathbf{x}')$, which enables us to express the covariance between $f(\mathbf{x})$ and $f(\mathbf{x}')$ as a function of $\mathbf{x}$ and $\mathbf{x}'$, without explicitly computing $\phi(\mathbf{x})$. Any kernel function that satisfies Mercer's condition is a valid kernel function [14]. Once the covariance function is specified, for new input $\mathbf{x}_*$, we can readily predict $m_* = f(\mathbf{x}_*)$ by computing the conditional distributions of the joint Gaussian distribution.

The effectiveness of the GP approach in spatial correlation modeling of wafer measurements and its ability to handle non-linear variation has been established through the studies reported in [10], [11]. Its success, however, relies on the assumption that the generative function f underlying the spatial variation should be continuous, albeit possibly affected by noise. This assumption is explained by the way the GP models are constructed: proximal data points are modeled as being highly covariant, and distant points are modeled with low covariance. Similar assumptions are made when applying the VP and EM approaches. However, our experience with production data shows that these assumptions may not hold for certain measurements. For example, spatial features of certain test data on a wafer can involve localized effects such as discontinuous trends. We describe in detail these effects in the next section.

III. Discontinuous effects in spatial wafer tests

As discussed in the previous section, GP models are able to handle highly non-linear data even when the latter is affected by noise. An example of wafer measurement data that can be modeled efficiently by GP model is shown in Figure 2(a). As can be observed, despite the non-linear behavior of the measurements, the GP model is able to capture the spatial correlation. Similar continuity assumptions are made when applying the VP approach [6], wherein a small number of dominant DCT coefficients at low frequencies are assumed to represent most spatial patterns of process variations.

However, these assumptions may not always hold true: spatial features of wafer measurements can exhibit localized effects such as discontinuous trends. These localized discontinuous effects may be caused by a variety of origins: a reticle shot that produces several die patterns at the same time in the lithography process resulting in individual rectangular regions, multi-site testing strategy leading to systematic variations for

die that are tested at the same time, chemical aging or crystalline non-uniformity resulting in individual spatial clusters on the wafer, etc. Our experience with production test data shows that, for certain measurements, these localized discontinuous effects (a) dominate spatial variations on the wafer and (b) are stationary, i.e., most wafers have very similar spatial discontinuous patterns. Figure 2(b) illustrates an example of spatial discontinuous effects on a wafer measurement.

These discontinuous effects can result in violation of the assumptions made when learning spatial correlation models of wafer measurements. As a consequence, training with such data may result in spurious spatial models. In this work, we propose a novel approach to handle discontinuous effects, without making any assumption about the discontinuous forms caused by different origins during semiconductor production.

IV. Proposed Methodology

A. Overview

In this section, we describe in detail the proposed methodology for handling discontinuous effects in spatial correlation models. Figure 3 shows an overview of the method, which consists of two main stages, namely pre-training and training. The objective of the first stage is to partition the wafer into k clusters, which reflect the k "levels" of wafer measurements induced by discontinuous effects. For this purpose, a k-means clustering algorithm is applied on a wafer, on which all measurements for all die locations are explicitly collected. The left hand side of Figure 3 shows an example of k-means clustering where the wafer is partitioned into 2 clusters. The k-means clustering algorithm and the method for choosing an optimal value for k is presented in detail in Sections IV-B1 and IV-B2, respectively. We note that the clusters might be different for each measurement, yet we assume stationarity across wafers, so identifying the clusters is a one-time effort. Furthermore, the number of wafers needed to compute k-means clustering is typically very small.

Once the k clusters are identified on the wafer for each particular measurement, we proceed to the second stage, where we capture spatial correlation within each cluster. For this purpose, we employ the spatial interpolation methodology based on GP [10], which was summarized in Section II. In particular, for each new wafer, we collect a sparse subset of die from each individual cluster and subsequently train k spatial models in order to predict the test outcomes at unobserved die locations, as shown on the right hand side of Figure 3. Individually training and predicting measurements in each cluster allows us to avoid modeling spatial correlation on an entire wafer, which can lead to erroneous spatial models due to discontinuous effects.

B. Pre-training stage

1) k-means clustering algorithm: The k-means clustering algorithm aims to partition n observations into k clusters, in which each observation belongs to the cluster with the nearest mean. Formally, let the set

$$M^i = \{m_1, m_2, \ldots, m_n\} \quad (4)$$

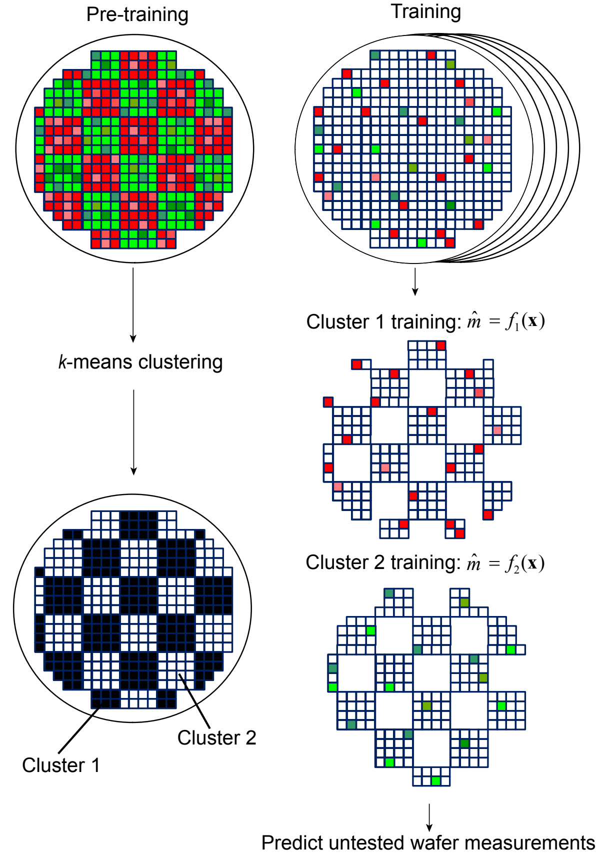

Fig. 3. Overview of proposed methodology

include the values of the i-th measurement on all die of a wafer, with m_j denoting the measurement on the j-th die and n denoting the total number of die which are to be clustered. The k-means clustering algorithm aims to partition M^i into k sets ($k \leq n$): $\{S_1, S_2, \ldots, S_k\}$ so as to minimize the expected *distortion D*, which is defined as the sum of squared distances between each observation and its dominating cluster mean:

$$D = \sum_j \|\bar{m}_{k(j)} - m_j\|^2 \quad (5)$$

where $\bar{m}_{k(j)}$ denotes the nearest cluster mean value for observation m_j. In this work, we use the most common iterative refinement technique to refine the choices of cluster means in order to reduce the distortion D. The technique involves the following steps [15]:

Step 1 - Initialization: Set k cluster means $\{\bar{m}_1, \bar{m}_2, \ldots, \bar{m}_k\}$ to random values.

Step 2 - Assignment: Each measurement in M^i is assigned to the cluster with the nearest cluster mean. The assigned p-th cluster is denoted by S_p:

$$S_p = \{m_j : \|m_j - \bar{m}_p\|^2 \leq \|m_j - \bar{m}_q\|^2, \forall 1 \leq q \leq k\} \quad (6)$$

Step 3 - Update: Compute the new cluster means.

$$\bar{m}_p = \frac{1}{n_p} \sum_{m_j \in S_p} m_j \quad (7)$$

where n_p is the number of observations in the p-th cluster.

Step 4 - Iterative Refinement: Repeat steps 2 & 3 until the

assignments do not change.

The k-means clustering algorithm is a simple, unsupervised learning approach which allows us to separate the die on a wafer into k different clusters caused by discontinuous effects. Subsequently, individual spatial correlation models can be learned on each cluster, without the need for *a priori* information on the form and origin of discontinuous effects.

2) Optimal choice of k: The question that naturally arises next concerns the choice of k. In this subsection, we discuss a method to determine the optimal value for k, since this choice is crucial in the k-means clustering algorithm. Underestimating k would result in clusters that still contain discontinuous patterns, while overestimating k would reduce the amount of available data in each cluster during training of spatial correlation models. The authors of [16] conducted a very comprehensive comparative study of 30 methods for determining the number of clusters in data. Among the variety of examined methods, the approach suggested in [17] generally outperformed the others. This approach consists of choosing an optimal value for k by maximizing the between-cluster dispersion and minimizing the within-cluster dispersion. Formally, the optimal value for k is defined as [17]

$$k = \underset{g}{\operatorname{argmax}}\, CH(g) \tag{8}$$

where $CH(g)$ is the Calinski and Harabasz index when g clusters are considered and is defined as

$$CH(g) = \frac{B(g)(g-1)}{W(g)(n-g)} \tag{9}$$

where n is the total number of die on the wafer, $B(g)$ and $W(g)$ are the between- and within-cluster sums of squared errors computed as

$$B(g) = \sum_{p=1}^{g} n_p(\bar{m}_p - \bar{m})(\bar{m}_p - \bar{m})^T \tag{10}$$

$$W(g) = \sum_{p=1}^{g} (\sum_{m_j \in S_p} (m_j - \bar{m}_p)(m_j - \bar{m}_p)^T) \tag{11}$$

where n_p denotes the number of samples in the p-th cluster, $\bar{m}_p$ denotes the cluster mean of the p-th cluster, and $\bar{m}$ denotes the mean of all measurement samples in M^i.

Equation (8) allows us to automatically choose an optimal value for k for a particular measurement without making any assumptions about its discontinuity trends.

C. Training stage

Once the wafer map is partitioned into k clusters, for each new wafer we can readily predict the test outcomes at unobserved die locations by collecting a sparse subset of die from each cluster and training k distinct models. Formally, let the set $S_p = \{m_1, \ldots, m_{nt_p}\}$ denote the observed measurement samples in the p-th cluster, and $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_{nt_p}\}$ denote the corresponding Cartesian coordinate of each sample, $\mathbf{x} = [x, y]$, where nt_p is the number of training samples in the p-th cluster. As described in Section II, we train a GP model f that maps the input parameter $\mathbf{x}$ to the target variable m, and we use $m_* = f(\mathbf{x}_*)$ to predict test outcome values at unobserved locations with new input $\mathbf{x}_*$. Finally, we combine prediction results from all k clusters to generate test outcome predictions (i.e. expected Test Escape and Yield Loss figures) for the entire wafer.

D. Prediction outcome evaluation

To evaluate effectiveness of the proposed methodology, we compute the mean absolute percent error across all predictions:

$$\epsilon_{jh} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left| (\hat{m}_i^{(h)} - m_i^{(h)}) / m_i^{(h)} \right| \tag{12}$$

where ϵ_{jh} represents the mean percent error of predicting the h-th measurement for all unmeasured die locations on a particular wafer j, and n_{test} denotes the number of predicted die locations on the j-th wafer. Then, we can summarize the mean prediction error over all considered wafers as:

$$\epsilon_h = \frac{1}{N_{wafers}} \sum_{i=1}^{N_{wafers}} \epsilon_{ih} \tag{13}$$

where N_{wafers} denotes the number of considered wafers.

In order to gain insight about the prediction outcome, it is also worthwhile to compute the Test Escape (TE) and Yield Loss (YL) incurred by applying the spatial correlation models. For a particular measurement, let the indicator functions $I_1^{(i)}$ and $I_2^{(i)}$ be equal to '1' if the predicted value of the i-th die location passes/fails its specification, while the actual value fails/passes the specification, and let $I_1^{(i)}$ and $I_2^{(i)}$ be equal to '0' otherwise. Then the overall TE and YL are defined as:

$$\hat{TE} = \frac{1}{N} \sum_{i=1}^{N} I_1^{(i)} \tag{14}$$

$$\hat{YL} = \frac{1}{N} \sum_{i=1}^{N} I_2^{(i)} \tag{15}$$

where N is the number of predicted die locations on all wafers.

V. Experimental results

We now demonstrate results of applying the proposed method on probe test data from high-volume semiconductor manufacturing. The device under consideration is an RF transceiver with multiple radios built in a 65nm technology. Our dataset contains a total of 3,406 wafers, each of which has approximately 2,000 devices, with 71 probe test measurements collected on each device. The number of clusters, k, and the shape of the clusters are computed by (8) during the pre-training stage, *using only the first wafer*. The minimum and maximum values of k obtained for the 71 measurements are 2 and 5, respectively, with an average of 3.9. Once the k clusters are known, we use a randomly chosen training sample of 20 devices from each cluster in order to train the spatial correlation models. Thus, a total number of 20×k devices are used for training in each wafer. The trained models are

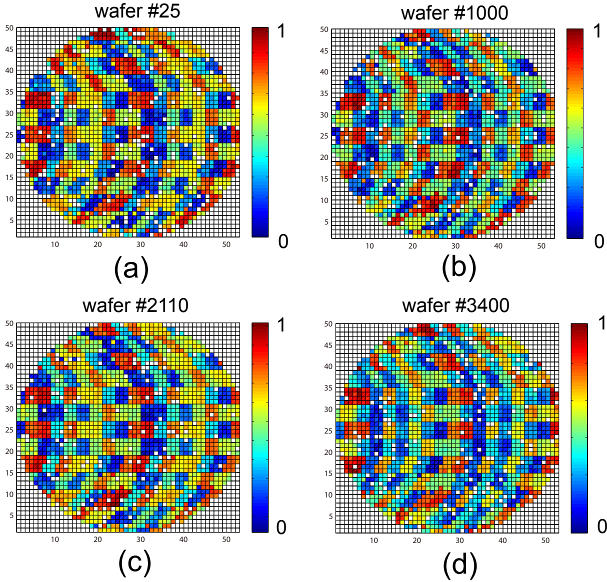

Fig. 4. Wafer maps of measurement 69 sampled randomly in the (a) 1st (b) 2nd (c) 3rd (d) 4th quarter of the 3,406 wafers

TABLE I
CLUSTER STATIONARITY VERIFICATION

	Wafer # 25		Wafer # 1000		Wafer # 2110		Wafer # 3400	
Clustering on wafer	#1	#25	#1	#1000	#1	#2110	#1	#3400
ϵ (%)	4.9	4	4.4	3.3	3.5	3.3	4.4	4.2
TE	0.03	0.03	0.02	0.02	0.03	0.03	0.04	0.03
YL	0	0	0.001	0.001	0	0	0.001	0

then used to predict the untested probe test outcomes at the remaining die coordinates, and the mean prediction errors across all wafers are computed through Equation (13).

A. Stationarity verification of discontinuous spatial patterns

The procedure outlined above relies on the assumption that the discontinuous spatial patterns are stationary over time, i.e. the clusters for a given measurement remain more or less the same across wafers. To illustrate that this is, indeed, the case, Figure 4 (a)/(b)/(c)/(d) shows the wafer maps of a measurement exhibiting discontinuous patterns, on randomly chosen wafers from the 1st/2nd/3rd/4th quarter of the 3,406 wafers. As can be observed, the discontinuous patterns remain very similar across wafers. To further investigate this issue, we performed the following experiment: for each of the four wafers shown in Figure 4, we ran the k-means clustering algorithm and used the obtained clusters, instead of the ones learned from the first wafer, to build the correlation models. What we found is that the number of clusters remained the same (i.e. 4) and that for the vast majority of die, the cluster that they ended up belonging to was the same. Most importantly, as shown in Table I, the differences in error rate, TE and YL were insignificant, corroborating stationarity.

We note that in cases where the discontinuous spatial patterns vary across wafers from different lots over different periods of time, a re-calibration of k-means clustering may be needed to adaptively learn the clustering over time.

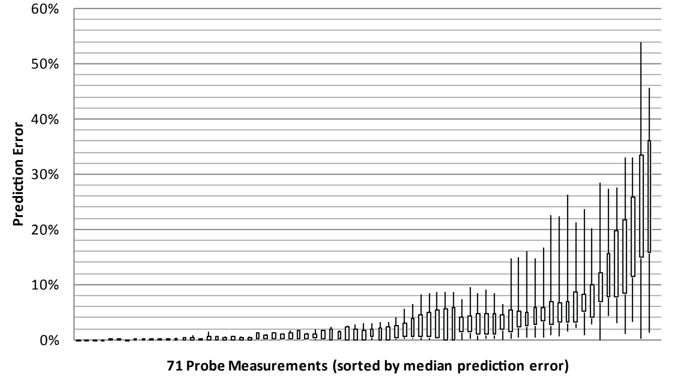

Fig. 5. Prediction error of proposed method for each probe measurement

B. Prediction errors using the proposed approach

Figure 5 summarizes the prediction error, with 10%-90% error bars shown, for each of the 71 measurements computed by (13). As can be observed in Figure 5, the prediction error for most measurements is below 5%, verifying the excellent capability of the proposed method in capturing spatial wafer correlation.

C. Comparison to existing approaches

In Figure 6, we compare the prediction error, as defined in (13), of the proposed approach to the error of the VP [6] and the GP [10] methods, using the same number of training samples per wafer and setting the VP error as the baseline at 0%. As can be observed, the GP method outperforms the VP approach, and the proposed approach achieves a significant further improvement in prediction error for several measurements. The measurements that enjoy a large improvement in prediction error are those who exhibit spatial discontinuous effects, yet we stress that, even for the measurements which show no discontinuous patterns, the proposed approach performs at least as well as the GP method.

Figure 7 shows the prediction wafer map of measurement 11 which exhibits discontinuous effect, using the proposed approach, the VP model and the GP model, respectively. Figure 7(a) shows the normalized actual wafer map, and Figures 7(b)/(c)/(d) show the prediction wafer map using the proposed approach/VP model/GP model, respectively. As can be observed, the discontinuous effect is correctly captured by the proposed approach, while the VP and GP models are very inaccurate when such discontinuous patterns exist on the wafer. This observation is further supported by the large improvement in percentile prediction error.

D. Test escape and yield loss improvement

To further elucidate the impact of the improved prediction error, in the 2nd and 3rd column of Table II we compare the Test Escape (TE) and Yield Loss (YL) computed using (14) and (15) for the VP, the GP, and the proposed method, for all 71 measurements and for all die of the 3,406 wafers. Evidently, the proposed approach achieves a significant TE improvement as compared to the VP and GP methods, while maintaining the same YL as the GP model.

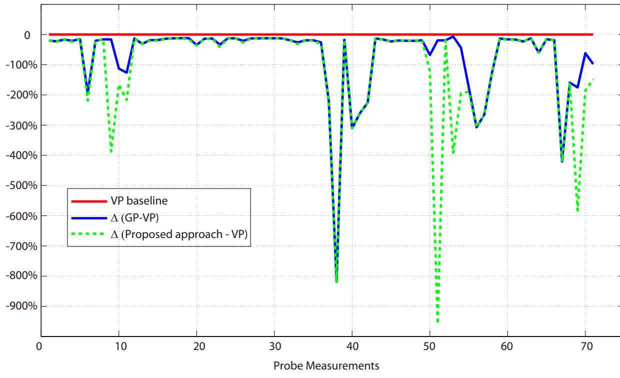

Fig. 6. Prediction error comparison of VP, GP and proposed approach

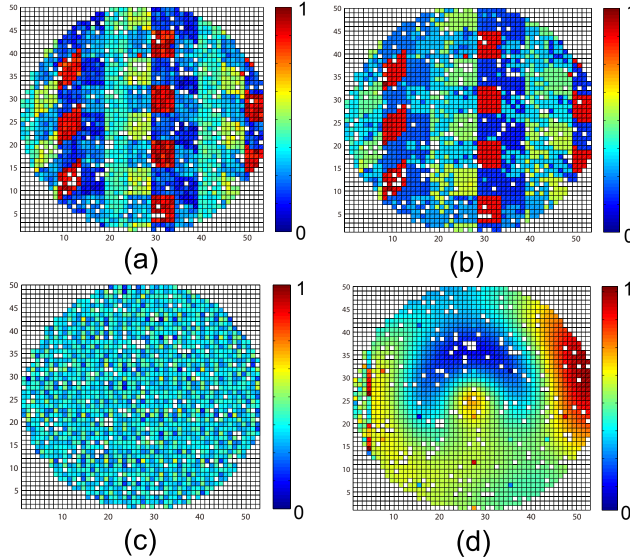

Fig. 7. Prediction of measurement 11: (a) Actual wafer map (b) predicted wafer map using the proposed methodology (c) predicted wafer map using VP (d) predicted wafer map using GP

The 4th to 7th columns of Table II report the TE and YL for two individual measurements, namely 53 and 69, where significant improvement in percentile prediction error is observed. The TE and YL are computed using (14) and (15) as before. As can be observed, the proposed approach consistently outperforms VP and GP, which is in-line with the percentile error improvement shown in Figure 6, and justifies the use of k-means algorithm in building spatial correlation models.

VI. Conclusions

In this work, we have demonstrated an approach to handle discontinuous effects in spatial correlation modeling of wafer measurements which drastically reduces prediction error for measurements exhibiting discontinuous patterns. The core of the proposed approach is a k-means algorithm which partitions a wafer into k clusters, as instigated by discontinuous effects. Effectiveness of the proposed approach is evaluated in comparison to existing approaches on industrial probe test data from over 3,400 wafers, demonstrating significant improvement in terms of percentile prediction error and, by extension, reduced Test Escape and Yield Loss rates.

TABLE II
Test Escape (TE) and Yield Loss (YL) comparison

	All measurements		Measurement 53		Measurement 69	
	TE	YL	TE	YL	TE	YL
VP	2.8e-2	8.5e-2	4.4e-6	4.8e-5	6.6e-6	8.1e-5
GP	2.8e-2	**4.4e-3**	3.9e-6	7e-6	**6.3e-6**	3.2e-5
Proposed	**1.4e-2**	**4.4e-3**	**1.9e-7**	**2.8e-6**	**6.3e-6**	**1.9e-7**

VII. Acknowledgement

This research has been carried out with the support of the National Science Foundation (NSF CCF-1149463) and the Semiconductor Research Corporation (SRC-1836.092).

References

[1] J. B. Brockman and S. W. Director, "Predictive subset testing: Optimizing IC parametric performance testing for quality, cost, and yield," *IEEE Transactions on Semiconductor Manufacturing*, vol. 2, no. 3, pp. 104–113, 1989.

[2] R. Voorakaranam, S. S. Akbay, S. Bhattacharya, S. Cherubal, and A. Chatterjee, "Signature testing of analog and RF circuits: Algorithms and methodology," *IEEE Transactions on Circuits and Systems - I*, vol. 54, no. 5, pp. 1018–1031, 2007.

[3] H.-G. Stratigopoulos and Y. Makris, "Error moderation in low-cost machine-learning-based Analog/RF testing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 2, pp. 339–351, 2008.

[4] H.-G. Stratigopoulos, P. Drineas, M. Slamani, and Y. Makris, "RF specification test compaction using learning machines," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 6, pp. 998–1002, 2010.

[5] S. Reda and S. R. Nassif, "Accurate spatial estimation and decomposition techniques for variability characterization," *IEEE Transactions on Semiconductor Manufacturing*, vol. 23, no. 3, pp. 345–357, 2010.

[6] W. Zhang, X. Li, F. Liu, E. Acar, R.A. Rutenbar, and R.D. Blanton, "Virtual probe: a statistical framework for low-cost silicon characterization of nanoscale integrated circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 12, pp. 1814–1827, 2011.

[7] W. Zhang, X. Li, E. Acar, F. Liu, and R. Rutenbar, "Multi-wafer virtual probe: Minimum-cost variation characterization by exploring wafer-to-wafer correlation," in *IEEE/ACM International Conference on Computer-Aided Design*, 2010, pp. 47–54.

[8] W. Zhang, K. Balakrishnan, X. Li, D. Boning, and R. Rutenbar, "Toward efficient spatial variation decomposition via sparse regression," in *IEEE/ACM International Conference on Computer-Aided Design*, 2011, pp. 162–169.

[9] H.-M. Chang, K.-T. Cheng, W. Zhang, X. Li, and K.M. Butler, "Test cost reduction through performance prediction using virtual probe," in *IEEE International Test Conference*, 2011, pp. 1–9.

[10] N. Kupp, K. Huang, J. Carulli, and Y. Makris, "Spatial correlation modeling for probe test cost reduction," in *IEEE/ACM International Conference on Computer-Aided Design*, 2012, pp. 23–29.

[11] N. Kupp, K. Huang, J. Carulli, and Y. Makris, "Spatial estimation of wafer measurement parameters using gaussian process models," in *IEEE International Test Conference*, 2012, pp. 5.1.1–5.1.8.

[12] F. Liu, "A general framework for spatial correlation modeling in VLSI design," in *Design Automation Conference*, 2007, pp. 817–822.

[13] C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

[14] V. Vapnik, *The Nature of Statistical Learning Theory*, John Wiley and Sons, Inc., 1995.

[15] D. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003.

[16] G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," *Psychometrica*, vol. 50, no. 2, pp. 159–179, 1985.

[17] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974.