Received: 7 December 2016 | Revised: 4 April 2017 | Accepted: 13 June 2017

DOI: 10.1002/qre.2192

RESEARCH ARTICLE

WILEY

Wafer yield prediction using derived spatial variables

Hang Dong[1] | Nan Chen[2] | Kaibo Wang[1]

[1]Department of Industrial Engineering, Tsinghua University, Beijing, China

[2]Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore

Correspondence
Kaibo Wang, Department of Industrial Engineering, Tsinghua University, Beijing, China.
Email: kbwang@tsinghua.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 71471096

Abstract

Unreliable chips tend to form spatial clusters on semiconductor wafers. The spatial patterns of these defects are largely reflected in functional testing results. However, the spatial cluster information of unreliable chips has not been fully used to predict the performance in field use in the literature. This paper proposes a novel wafer yield prediction model that incorporates the spatial clustering information in functional testing. Fused LASSO is first adopted to derive variables based on the spatial distribution of defect clusters. Then, a logistic regression model is used to predict the final yield (ratio of chips that remain functional until expected lifetime) with derived spatial covariates and functional testing values. The proposed model is evaluated both on real production wafers and in an extensive simulation study. The results show that by explicitly considering the characteristics of defect clusters, our proposed model provides improved performance compared to existing methods. Moreover, the cross-validation experiments prove that our approach is capable of using historical data to predict yield on newly produced wafers.

KEYWORDS
fused LASSO, semiconductor manufacturing, spatial, yield prediction

1 | INTRODUCTION

Semiconductor manufacturing is a complex and elaborate production process. Starting from wafer substrates, a typical semiconductor manufacturing process consists of hundreds of steps, through which small functional units called chips are produced for a broad range of applications. Figure 1 presents an outline of this process. To monitor and control the quality of chips on a wafer, a series of functional testing operations must be performed following the completion of wafer fabrication.[1] Functional testing, such as the walking I/O test and checkerboard test for memory chips, intends to verify and ensure the functionalities of chips. Different types of faults may affect the functionality of a chip. For example, a memory chip may suffer from an address decode fault, coupling fault, and so on. Consequently, each memory chip may be reported with a certain number of defects. A chip is then classified as either defective or nondefective based on its reparability, and only nondefective chips are further processed, packaged, and mounted in downstream products to perform designated functions, such as memory modules. During field use, an unreliable chip may fail before its expected lifetime. In this case, product quality and customer satisfaction are affected. Therefore, chips that are destined to fail in the future should be identified and discarded in early production. On the basis of domain knowledge and expertise, engineers believe that chips that perform poorly in functional testing are more likely to fail in field use.

Several works on modeling the production yield are introduced in Section 2. They treat yield as the ratio of functional chips after each test step. Here, we define the ratio of chips that do not fail before their expected lifetime in field use as yield, and the chips that fail before their expected lifetime as failed chips, which is not exactly the same as the common definition in the literature. In practice, such yield information is difficult to obtain from chip manufacturers and therefore not well studied in the literature. In our research, we collaborate with a memory manufacturer; experienced engineers provide the predicted status of each chip as pass or fail by combining various lab testing data or burn-in experiments

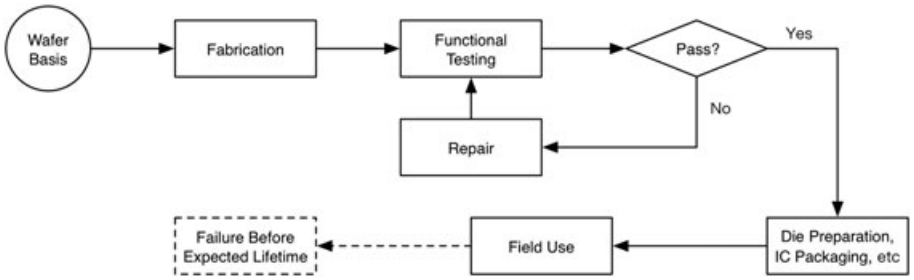


FIGURE 1 Outline of the semiconductor manufacturing process

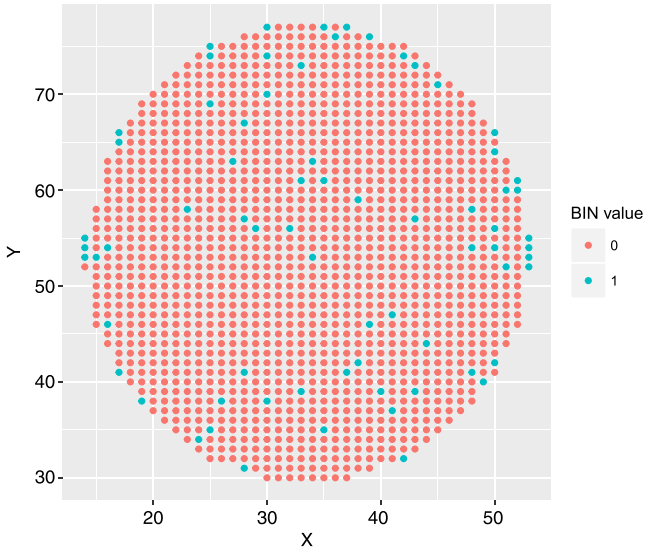


FIGURE 2 An example of a wafer bin map [Colour figure can be viewed at wileyonlinelibrary.com]

data[1,2] and failure mechanism analyses. Because such analyses are labor consuming, it is of great interest to develop a statistical model to predict final wafer yield based on functional testing results that are available in early production stages.

Yield data collected after field use (or provided by engineers in this study) are represented in the form of a wafer bin map. An example of a wafer bin map is presented in Figure 2; the BIN code is 0 if the chip lasts longer than its expected lifetime and 1 if it fails. Chips are spatially located on the wafer surface, and chip location information is also stored together with the BIN code. A wafer bin map can help engineers recognize embedded patterns of faulty chips and trace back to the process to locate possible process failures, thus helping to identify the root cause and improve quality. In addition to the binary data, discrete or continuous failure data that are collected at functional testing stations can be expressed in a similar map format. On such maps, each site shows the count of defects or values of critical dimensions rather than the binary value presented in Figure 2. These wafer maps are critical to the semiconductor industry for process analysis and yield improvement.

Defects generated on wafers can be classified into 2 categories: random defects and clustered defects (also called global defects and local defects).[3-5] Global defects are randomly distributed across a wafer, whereas local defects arise with different spatial patterns, such as lines and circles.[6] Engineering knowledge reveals that the spatial distribution patterns of functional defective chips on a wafer may be attributed to specific causes. For example, the zone pattern often arises from the thin film deposition stage, the edge ring is primarily due to etching problems, and the linear scratch is likely induced by machine handling issues. Although the development of integrated circuit (IC) fabrication technology has made it possible to produce increasingly smaller microchips with complex structures, defective chips cannot be completely avoided due to many known and unknown factors.[7] Therefore, having a good understanding of wafer defects and developing an efficient algorithm for yield prediction are of key importance to semiconductor manufacturers.

In the literature, different methods have been proposed for yield modeling, defect detection, and classification of IC chips. However, to the best of our knowledge, the spatial cluster information of defects has not been fully used for yield prediction. As we all know, one important feature that wafer

defects have is the presence of spatial clusters. Because adjacent chips on a wafer are fabricated under similar conditions, it is commonly observed that functional defective chips form certain spatial patterns. Therefore, a chip that is surrounded by defective chips is more likely to be defective, whereas a chip that is far from clustered defects is more likely to be nondefective. Defective chips shown in the form of clusters imply a certain spatial connection among them, and such information should be valuable for predicting the status of nearby chips.

This paper intends to propose a new model for defect cluster detection and yield prediction. In contrast to the existing methods that perform prediction based primarily on the statistical distribution of defects or the coordinate information of chips, we propose the use of both functional testing data and derived variables that are largely ignored in the existing models to improve the model performance. Given a wafer map obtained from functional test, we first identify clusters of defective chips and separate them from random defects; then, for each chip, apart from its corresponding functional testing information, we derive new variables based on the distribution of defect clusters and then incorporate these derived variables into a prediction model. The explicit use of such derived variables provides a new way to efficiently use spatial information and is thus expected to benefit model accuracy.

The remainder of this paper is organized as follows. In the following section, we provide a review of the literature that is relevant to wafer map data analysis and yield prediction modeling. Then, in Section 3, we present our prediction with derived variables. In Section 4, the proposed model is applied to real wafer samples to demonstrate its use and performance. In Section 5, a more extensive performance study is conducted based on simulated data. Finally, Section 5 discusses issues related to the implementation of the proposed model, and Section 6 concludes this work with suggestions for future research.

2 | LITERATURE REVIEW

Wafer map data are the main source for data-driven process management in semiconductor manufacturing. In the literature, progress in the following areas has been reported, namely, process monitoring for spatially clustered defect detection, wafer bin map defect pattern classification, and yield prediction.

As previously mentioned, a wafer may suffer from both random and clustered defects. The appearance of clustered defects is an indication of possible process shifts. Therefore, different control chart algorithms have been designed to monitor wafer maps and trigger early warnings if clustered defects are detected. Spanos[8] provided an overview of the application of statistical process control (SPC) techniques, such as X-R chart, in semiconductor manufacturing. Taam et al[9] proposed a measure for spatial randomness on wafers called the log odds ratio by counting the number of neighbors of defective chips and the number of neighbors of nondefective chips, and this measure is widely used for identifying the existence of spatial clusters in many works. Hansen et al[4] developed testing statistics for monitoring clustering defects on wafers. Friedman et al[10] proposed a model-free approach to estimate and monitor the probability of clustered failure and random failure separately. Ge et al[11] proposed an adaptive substatistical PCA model for process monitoring based on batches of multiple process variables. Wang et al[12,13] and Zhang et al[14] proposed control charts that consider cluster effects of defects or spatial correlations among sites to monitoring wafers with continuous readings. In contrast to the above SPC techniques that generally monitor across batches of wafers, we focus more on the use of detailed spatial distributions of defective chips on a wafer map to improve prediction accuracy.

By observing wafer maps, it is learned that defective chips form patterns with specific shapes or styles. Engineering knowledge can occasionally attribute certain types of patterns to particular process faults. Therefore, it is appealing to practitioners to classify wafer maps based on observed defect patterns. Defect pattern classification is closely related to the field of pattern recognition. Ooi et al[15] suggested that to develop an accurate defect cluster recognition system, one should determine (1) the optimal feature set, (2) the best feature extraction method and type of classifier, and (3) the classifier training method. Both supervised and unsupervised machine learning approaches have been adopted for this purpose. For example, Chen et al[6] proposed using a neural network architecture called adaptive resonance theory network1 (ART1) to recognize spatial patterns of clustered defects and showed that the method outperformed the unsupervised neural network of self-organizing map (SOM). DeNicolao et al[16] later argued that the ART1 technique is not adequate when applied to numerous data sets, but the SOM architecture performed extremely well. Moreover, Hsieh et al[17] represented defect clusters using minimum rectangle areas and used fuzzy rules to combine them and identify spatial defect patterns. Wang et al[18] proposed an online diagnosis system based on a composite clustering method combining both "fuzzy C means (FCM) partition" with "batch single linkage" recombination that could separate various types of defects on the same wafer. Yuan et al[19] used model-based clustering algorithms via Bayesian inferences to detect defect clusters and identify the pattern of each cluster. Jeong et al[20] used the dynamic time warping algorithm for spatial pattern classification based on spatial correlograms. There are also simulation-based works on defect pattern classification. For example, Hsieh et al[21] started from predefined defect patterns and conducted morphology-based simulations to help train support vector machine (SVM) classifiers for defect pattern identification. Note that none of these works on defect pattern

identification and classification used any measurement or pretest information, that is, only wafer bin maps presenting result of specific test procedures are used by these algorithms for model training and classification. Besides, these works on identification and classification of defect patterns do not provide detailed chip-level usability information. In this paper, we propose the use of both discrete-count test values and derived variables from the test values in model building, and predict chip-level performance in field use.

Yield modeling and prediction are important to practitioners in 2 aspects. First, yield modeling can help engineers obtain a better understanding of defect patterns and generation mechanisms; second, an accurate prediction model is helpful for efficient production planning. In the literature, different approaches have been proposed for yield modeling and prediction. Stapper[22] and Cunningham[23] modeled the number of defects in an area using a Poisson distribution. Stapper[24] proposed a yield prediction model based on a binomial distribution. Koren et al[25] proposed a unified negative binomial distribution-based yield prediction model for the yield analysis of defect-tolerant circuits. These early yield models were based on the type of data, where each chip may have multiple defects, but did not incorporate chip-level spatial information into the yield model. Kumar et al[26] presented a detailed review of yield modeling techniques from simple probabilistic yield models to the incorporation of critical features such as radial yield degradation. Longtin et al[27] considered a Markov random field model to capture the spatial clustering effects of defective chips. Wu et al[28] combined the key attribute parameters of defects, physical parameters of wafers, and wafer electrical test parameters in a fuzzy neural network to predict yield on wafers. However, neural network approaches could hardly provide interpretable results to help improve the process. Bae et al[29] and Yuan et al[30] incorporated the spatial locations of chips as covariates and modeled clustered defect counts across a wafer based on models of Poisson or negative binomial class. Using spatial coordinates as covariates implies one strong assumption, that is, local defects have a global trend that can be characterized along axial directions. However, this is not always true given that local defects occur randomly and in different shapes of clusters.

In the following, we will present a new yield prediction model that uses the local feature of functional defective chips. Considering the spatial connection among chips, we incorporate derived variables in the model and hope to improve the model performance.

3 | MODEL BUILDING USING DERIVED VARIABLES

In this section, we propose a model for yield prediction using functional testing data and derived variables. Functional testing data are obtained before packaging a chip; a chip that shows many defects during functional testing is an implication of unstable quality of the chip and chips around the location of the chip. Therefore, functional testing results can be viewed as an important indicator of production stability and quality and are thus helpful for yield prediction.

Another important observation that engineers have is the connection of chips. As previously mentioned, both functional defective chips and unreliable chips in field use tend to occur in clusters, and a chip that is surrounded by defective chips is more likely to be defective. However, such phenomenon has not been incorporated in yield prediction models. In this work, we propose to take such information into consideration and use derived variables to represent the relationship between a chip and a cluster. More specifically, we first identify chip cluster patterns on a wafer. Subsequently, each chip is assigned properties such as whether the chip belongs to a cluster and the size of the cluster. Then, these derived variables are used together with the functional testing result of each chip to predict yield. Therefore, the derivation of the proposed model includes 3 steps. Step 1, identify clusters from wafers based on functional testing results; step 2, assign values of derived variables to all chips based on cluster patterns; and step 3, build models for yield prediction using both functional testing results and derived variables.

3.1 | Cluster identification using fused LASSO

Figure 3 shows the defect map of a wafer after functional testing. Each dot represents a site, which corresponds to a chip. The test generates discrete values representing the number of defects found from the chip; the number of defects is heavily right skewed, with a peak on the left and a long tail on the

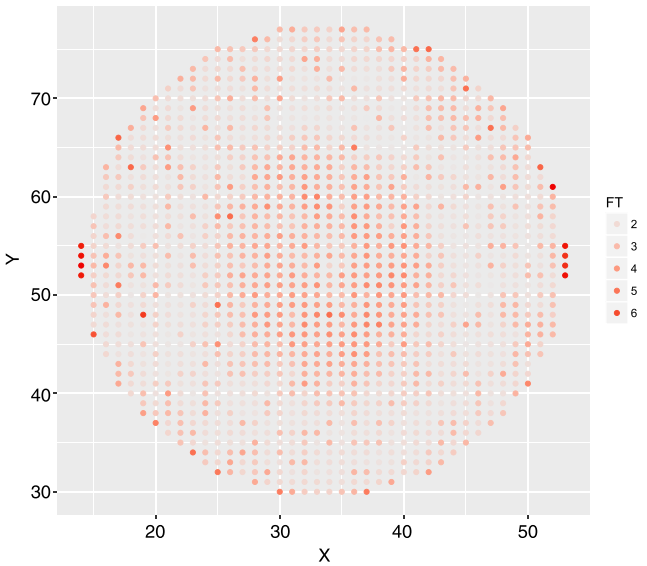


FIGURE 3 Example of logarithm of functional testing data on wafer [Colour figure can be viewed at wileyonlinelibrary.com]

right. A chip that has a large number of defects is an indication of unstable quality at that location. For easy handling, we take the logarithm of the defect data for further analysis.

It is clear from Figure 3 that defects occur in cluster patterns. So we need to smooth the data and identify possible clusters.

Note that the defect cluster identification step is different from the aforementioned methods in Section 2 in 2 aspects: Firstly, the type of input data in the proposed procedure is functionality test result of discrete values while the defect cluster detection or classification methods use binary wafer map as input. Secondly, most of these methods only provided classification result of each wafer, but did not provide chip-level information as output, namely, whether one specific chip is inside a cluster, the size of cluster, and the minimal distance to the cluster edge.

The cluster identification step here aims to extract the chip-level information in geometric measurements of spatial defect clusters. The fused LASSO procedure is a straightforward method to provide such information and remove the noise in discrete count data. So we adopt fused LASSO procedure to identify spatial defect clusters and derive variables for yield prediction. It should be noted that some existing algorithms that are capable of locating defect chips on wafers, after proper modification, can also be used to replace the fused LASSO algorithm in this work.

The LASSO method was first proposed by Tibshirani,[31] and it uses penalties on the regression coefficients to achieve sparseness. The penalized loss function of LASSO is

$$L(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta})^{(1)} = \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda_1 \sum_{i=1}^{p} |\beta_i|, \quad (1)$$

where $y \in \mathbb{R}^n$ is the response vector, $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the matrix of predictors, and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the coefficient vector. However, this L_1-penalty does not assume any intrinsic structure in the coefficients, eg, the underlying coefficients are in some way ordered, and thus, neighboring coefficients should have a smaller difference. Considering this situation, fused LASSO attempts to penalize both coefficients and their successive differences. The loss function of fused LASSO is

$$L(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta})^{(2)} = \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{i=1}^{p-1} |\beta_i - \beta_{i+1}|. \quad (2)$$

In addition to the penalty on the coefficients of variables, a second penalty term is added to smooth the overall estimates over the specific neighboring structure of coefficients, thus forming segments of zeros or nonzero but with equal values.

To smooth and segment the wafer map shown in Figure 3, we adopt the general spatial fused LASSO method proposed by Hoefling,[32] where the penalized loss function is

$$L(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta})^{(3)} = \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{(i,j)\in E, i<j} |\beta_i - \beta_j|, \quad (3)$$

where E is a set of edges in the graph $\mathcal{G} = (V, E)$ with $V = \{1, \ldots, p\}$ representing the variables. It is natural to assign physical adjacency to the graph structure, e.g. to make the immediate left, right, upper, and lower sites as neighbors of a specific site. Once the neighboring structure $\mathcal{G}$ is defined, the first penalty term in (3) would shrink small but nonzero elements to zero, and the second term in (3) would force an equal shift magnitude among neighbors, leaving clustered outliers standing out. Intuitively, this method balances the trade-off of sparsity and neighbor similarity.

In our application, measurement values on a wafer are treated as a high-dimensional random vector $\mathbf{y}$. Since the fault behind a cluster pattern tends to affect an area of adjacent chips, smoothing the measurement values may make clusters easier to stand out. A wafer has both local and global defects; in implementing the penalized estimation algorithm, we set $\mathbf{X}$ as an identity matrix $\mathbf{I}$ and $\lambda_1 \equiv 0$ to only force neighbors to have equal shifts but without a sparse penalty. Now that only λ_2 remains in the model, we denote λ_2 as λ for convenience. The neighborhood structure could be defined flexibly according to the specific application, such as a 4-neighbor structure (Rook-move neighbors) or an 8-neighbor structure (King-move neighbors).

The path algorithm for the generalized LASSO problem proposed by Tibshirani and Taylor[33] is adopted to calculate a series of λ values. This method is implemented in the R package *genlasso*.[34] It begins with a sufficiently large λ, where on all sites, the estimates are the same because a large λ forces neighboring sites to have equal estimates. Then, the value of λ decreases step by step to allow significant differences of neighboring sites to emerge. Note that when λ decreases to 0, we will obtain the ordinary least squares solution, where the estimate of each site is the original value itself. When λ is in between the sufficiently large value and 0, we can find proper values to obtain a map that only emphasizes significant jumps between several spatial parts on it.

3.2 | Derived variables for yield prediction

On the basis of the segmented and smoothed wafer map identified by the fused LASSO procedure, we now derive new variables that are potentially helpful for yield prediction.

As previously discussed, the main feature of wafer defects is clusters, and the relationship of a chip to a cluster is very important. Therefore, we consider using the following variables to better characterize the situation of a chip:

- Whether the chip belongs to a defect cluster. If a chip is within a defect cluster, it is more likely to also be defective unreliable in field use.
- Cluster size that a chip belongs to. If a chip belongs to a large defect cluster, it is more likely to be defective as a large cluster is possibly more influential.
- Minimal distance from the chip to the edge of a defect cluster. If a chip is close to a defect cluster, the chip is more likely to be affected by this cluster.

The above derived variables describe the relationship between a chip and a cluster. Assume that there are p chips on a wafer; then, the response vector $\mathbf{y}_i$ is a p-dimensional vector. We use $u_{ij}, i = 1, 2, \ldots, p; j = 1, 2, 3$ to denote the j^{th} derived variable on chip i. In the following, we provide a more detailed technical definition of these variables, and we test the statistical significance of these variables in a later section of this paper using real and simulated data.

3.2.1 | Whether the chip belongs to a defect cluster

This binary variable has a straightforward meaning: separating clustered suspicious chips from other normal chips. To construct this variable, we use an indicator function, as follows:

$$u_{i1} = 1(\beta_i > c), \quad i = 1, 2, \ldots, p, \tag{4}$$

which means that if β_i exceeds a threshold value c, we set u_{i1} to 1; otherwise, we set u_{i1} to 0. β_i is obtained in the previous section using the fused LASSO algorithm. u_{i1} is a direct indication of the location of a chip relative to a cluster. If $u_{i1} = 1$, we expect that the yield of this chip is adversely affected.

3.2.2 | Cluster size that a chip belongs to

To count the size of a cluster, we traverse all chips that belong to the cluster, ie, the set $S = \{i | u_{i1} = 1\}$, and take down the number of chips that belong to the same cluster. A more formal expression is

$$u_{i2} = \sum_{j \in D(i)} u_{j1}, \quad i = 1, 2, \ldots, p, \tag{5}$$

where $D(i)$ is the defect cluster district that has all its components connected with chip i. The intuition behind this derived variable is that the size of a cluster is a possible indication of the severity level of the root cause that led to the defect cluster; thus, we believe that the larger the size the suspicious cluster is, the more likely that chips inside the cluster would fail.

3.2.3 | Minimal distance from the chip to the edge of a defect cluster

For a chip in a defect cluster, it is more likely to fail if it remains close to a cluster. In other words, the failure probability of a in-cluster chip should increase with the distance to the edge of this cluster. Moreover, if a chip is not included in a clustering region recognized by fused LASSO, it is reasonable to consider it to be a good chip. Additionally, it is increasingly more likely to be good as the distance from the recognized edge of the cluster becomes longer. To construct a variable based on the above idea, we first develop an edge list M, which contains all chips that are on the edge of a cluster. Hence, the minimal distance between a chip and the edge of a defect cluster can be expressed as follows:

$$u_{i3} = b_i \min_{j \in M} \{ \|l_i - l_j\| \}, i = 1, 2, \ldots, p, \tag{6}$$

where l_i is the Cartesian coordinate of chip i. To impose on the variable, a trend that the normal chips should have less probability of failing if they are farther from the edges of defect clusters, we multiply this variable by a sign term $b_i = 2 \times u_{i1} - 1$. Thus, $b_i = 1$ if the chip is inside a cluster, and $b_i = -1$ if it does not belong to any cluster.

3.3 | A yield prediction model based on functional testing data and derived variables

In this section, we present a prediction model that incorporates functional testing data with the above derived variables from spatial wafer maps. In addition, spatially distributed measures typically exhibit certain spatial correlations. There are different ways to study spatial correlations in statistical models. For chip i, it is assumed to be directly affected by neighboring chips.[35] This relationship enables us to develop the following logistic regression model to predict whether a chip will fail:

$$\text{logit}(y_i) = \alpha + \gamma' x_i + \eta' \mathbf{u_i} + w^T I_{N(i)}(\mathbf{x}) + \epsilon_i \tag{7}$$

$$y_i = Pr(Y_i = 1 | model), \tag{8}$$

where

α – a constant;
$\mathbf{x}$ – a $p \times 1$ vector with functional testing data;
$\mathbf{u}$ – derived variable from fused LASSO, with dimensions $p \times 3$;
$I_{N(i)}(\cdot)$ – an indicator function of neighbors of chip i, where $I_{N(i)}(\mathbf{x})_j = 1$ if j is a neighbor of i and 0 otherwise;
w – a weight vector with dimensions $p \times 1$; all neighbors of a chip are assigned an equal weight. The weighted sum of functional testing data around chip i provides spatial support to the yield at this location.

We consider all the main effect and interactions of any 2 variables when selecting the model. And we evaluate each model based on the Akaike information criterion (AIC) using data from real wafer samples:

$$AIC = 2k - 2 \log(\hat{L}),$$

where k is the number of free parameters to be estimated, and $\hat{L}$ is the likelihood of the model given the data we use. It turns out that the form of Equation 7 obtains the smallest AIC value using real wafer data; thus, we adopt this model for yield prediction.

Because the proposed model considers information that is largely ignored in existing yield prediction models, we expect that the new model will provide better performance. Its performance will be studied extensively in the following section.

4 | PERFORMANCE STUDY

In this section, we conduct both simulation and empirical studies on the proposed method and study the performance of the proposed model.

4.1 | Experiments using real wafer maps

First, we apply the proposed approach to 3 real wafer examples. Following the proposed framework, we first conduct fused LASSO on wafer maps with functional testing results to derive cluster-related variables. Then, we use the proposed model to predict failed chips on the wafer and compare it with known results for performance evaluation.

4.1.1 | Cluster identification using fused LASSO

As shown in Figure 4, there are 1461 chips on the circular wafer. The original data show signs of mixed random effects and clustered effects. The fused LASSO procedure performs a task similar to noise filtering. It fuses the measurement data to make neighboring chips share an equal value. The fusing strength can vary by adjusting the parameter λ. In this step, we adopt the path algorithm for the generalized LASSO problem [33] and a 4-neighbor structure to calculate a series of λ values and choose an appropriate one.

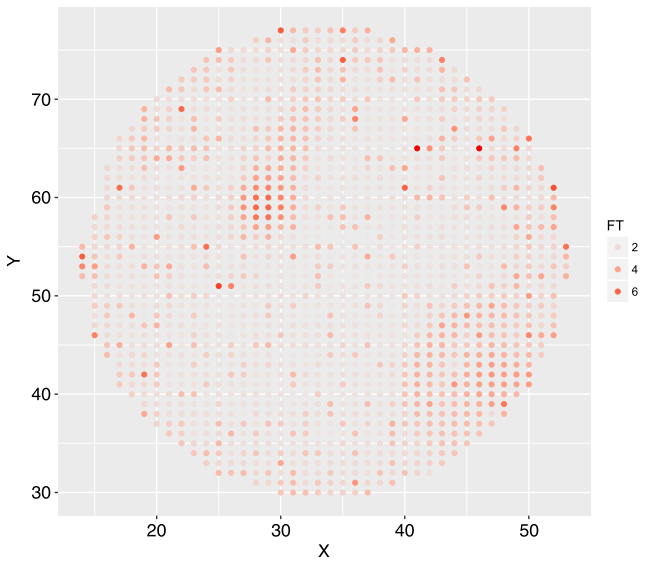
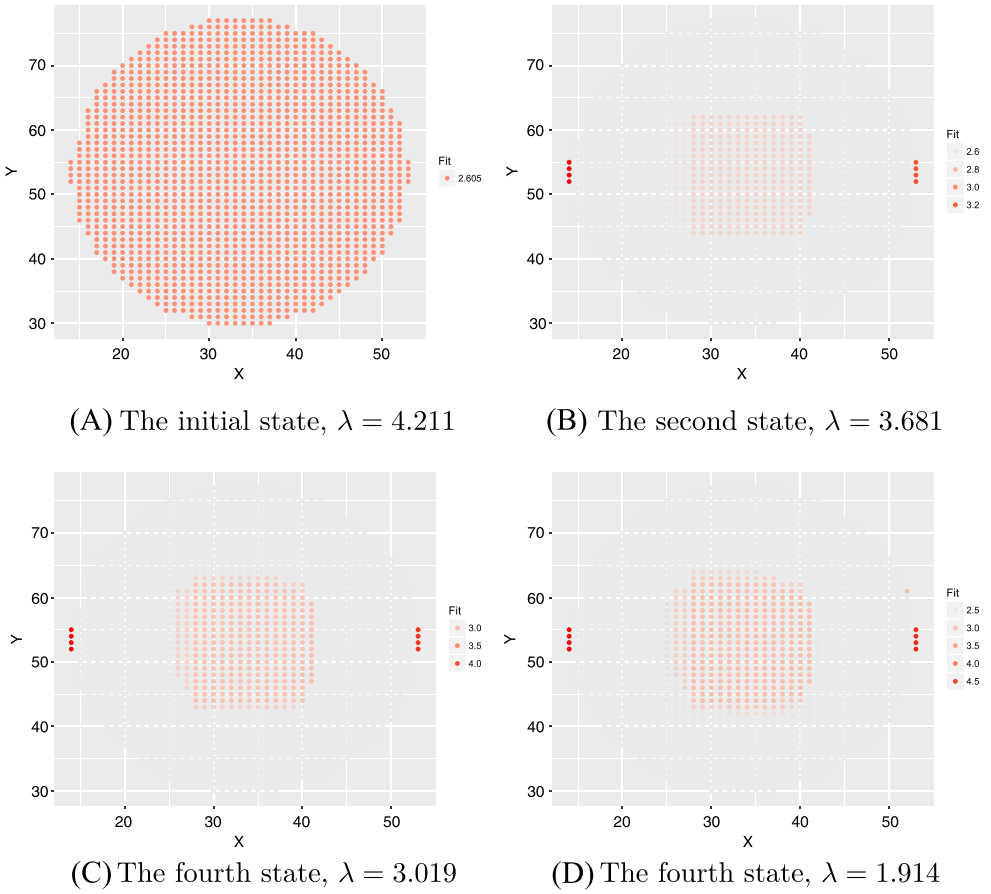


FIGURE 4 The functional testing data map on a wafer [Colour figure can be viewed at wileyonlinelibrary.com]

The solution path of fused LASSO begins with the initial state in Figure 5A, where $\lambda = 2.338$ is large enough to force the estimates to be equal on all chips. Then, as λ decreases, the central cluster emerges. Because we choose the 4-neighbor structure to conduct fused LASSO, a square first emerges. A different neighbor structure may result in a different emerging shape, thus slightly affecting the prediction performance, which we will discuss in Section 5. As shown in Figure 5, there are also chips with high values around the edge. The 4-chip lines on both sides could be defined as a cluster or just random defects according to the requirement to form a cluster. Under different production conditions, the definition of a cluster should vary. Engineers could choose an appropriate λ according to the specific situation. In this section, we simply set λ equal to the value nearest to half of the first value in the path of fused LASSO (the first value hitting the boundary). We will further discuss guidelines for selecting λ and other controllable parameters in Section 5.

4.1.2 | Logistic regression with derived variables

On the basis of the fused clusters, we now fit yield data with functional testing data and derived variables in the wafer to the logistic regression model defined in (7). The estimated coefficients of the proposed logistic regression model are presented in Table 1. The result shows that all the derived variables in the logistic regression model are statistically significant, which indicates that they could each explain part of the failure probability of chips on the wafer. Although there may exist differences in the estimated coefficients across different wafers, the scales of the coefficients are consistent. The functional testing data are found to be positively related to the failure probability of chips. The positive coefficient of the variable *distance from edge* indicates that if a chip belongs to a defect cluster, the farther it is from the cluster edge, the more likely it is to fail; if a chip does not belong to a defect cluster, the farther it is from the cluster edge, and the less likely it is to fail. This verifies our assumption of the underlying physical mechanism of clustered defects. Moreover, the *mean measurement of neighbors* explicitly adds the influence of functional testing values of neighboring chips under a specific neighbor structure: The higher the failure probability of neighboring chips is, the more likely the chip is to fail. The interaction term of the variables *in cluster* and *cluster size* indicates that a chip inside a defect cluster of a larger size is more likely to fail than a chip inside a defect cluster of a smaller size.



(A) The initial state, $\lambda = 4.211$ (B) The second state, $\lambda = 3.681$

(C) The fourth state, $\lambda = 3.019$ (D) The fourth state, $\lambda = 1.914$

FIGURE 5 The fused LASSO estimates under different λ values [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Estimation result for logistic regression model with derived variables

	Estimate	Std. Error	*p* Value
		Wafer1	
Intercept	−13.711344	0.665229	<2e-16
Measurement	2.434540	0.172998	<2e-16
Distance from edge	0.033162	0.011224	0.00313
Mean measurement of neighbors	2.128041	0.240530	<2e-16
In cluster: Cluster size	0.008060	0.001232	6.15e-11
		Wafer2	
Intercept	−10.911973	0.518863	<2e-16
Measurement	1.750133	0.135960	<2e-16
Distance from edge	0.039645	0.009853	5.73e-05
Mean measurement of neighbors	1.890055	0.204136	<2e-16
In cluster: Cluster size	0.010377	0.001612	1.22e-10
		Wafer3	
Intercept	−11.833445	0.612101	<2e-16
Measurement	2.097247	0.160819	<2e-16
Distance from edge	0.161589	0.021091	1.84e-14
Mean measurement of neighbors	2.011279	0.240508	<2e-16
In cluster: Cluster size	0.003981	0.001093	0.000271

4.1.3 | Cross-validation between wafers

In Section 4.1.2, we fit a model using both yield data and functional testing data. In practice, we want to predict the failure probabilities of chips using only functional testing data before knowing the state of a chip in field use. In other words, the practical situation requires historical data to train the model and then use the model to predict the failure probability of new wafers. To this end, we wish to validate the performance of our model trained and tested using different wafers. In this section, we conduct cross-validation of the proposed model on 3 wafers with clustered defects on them.

At each time, 2 wafers are used to fit the logistic regression model, and the model is evaluated using the data from the third wafer. The proposed model based on derived variables from the fused LASSO procedure (named FL for simplicity) is constructed according to (7) and (8), where the spatial clustering information of defects is explicitly represented using several derived variables. For comparison, we consider 2 other methods, namely, a basic logistic model with only functional testing data as covariates (Basic1) and a logistic model with both functional testing data and neighbor test values (Basic2). The model forms for Basic1 and Basic2 are shown in (9) and (10), where the notations are the same as those in Section 3.3.

$$\text{logit}(y_i) = \alpha + \gamma' x_i + \epsilon_i \tag{9}$$

$$\text{logit}(y_i) = \alpha + \gamma' x_i + w^T I_{N(i)}(\mathbf{x}) + \epsilon_i \tag{10}$$

In the cross-validation process, once we select a wafer for testing, the testing wafer is set aside during model training. The other 2 wafers are then used for parameter estimation, and the prediction performance is evaluated on this test wafer. This cross-validation procedure is conducted in all of the Basic1, Basic2, and FL models. Moreover, we also compare the cross-validation results with the fitness performance of all these 3 models fitted and tested both on the test wafer. In this way, we could expect to extend the use of the proposed model to the situation where only historical functional testing data are used to train a model, and the model is used to predict chip failures on newly produced wafers. Note that the fitness performance should theoretically be better than the cross-validation performance using the same model because the parameters used in cross-validation are trained using 2 other wafers while the parameters of fitness evaluation are trained using only the testing wafer.

We adopt F-score, area under the receiver operating curve (AUC), and the Akaike information criterion (AIC) to compare the performances of the above models. The F-score (also called F_1 score) and AUC are 2 widely used indices for evaluating and comparing prediction powers in machine learning and many other binary classification/prediction problems. By setting the prediction boundary as $p = 0.5$, we can calculate the F-score of the prediction:

$$\text{F-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

A high F-score indicates a good performance (weighting recall and precision equally) on binary classification prediction. AUC is the area of the plot of the true positive rate vs the false positive rate as the threshold value for classifying the chip as failed is increased from 0 to 1. A good property of AUC is that it is independent of the fraction of the test population, which is good or defective. Since the number of defective chips is significantly smaller than that of good chips on a wafer, AUC is considered to be a useful index for evaluating model performance on this unbalanced data set. In addition, the AIC values of the above models are also compared. Table 2 presents the results of the comparison. It is found that with respect to the prediction power (F-score and AUC), although the cross-validation prediction performance of the proposed model is not as good as the fitted model using the same data, it is generally better than the model using only functional testing data itself and its neighboring data. More importantly, the cross-validation result of our proposed model (FL-CV) uses no information of the testing wafer, and this feature makes it more powerful in practice. Regarding the relative modeling quality (AIC), our proposed approach has the lowest AIC values both in the fitting and cross-validation groups. To summarize, the cross-validation experiment between wafers shows that our proposed method

TABLE 2 Cross-validation results of logistic regression model on real wafers

Wafer		Methods					
		Basic1-Fitness	**Basic1-CV**	**Basic2-Fitness**	**Basic2-CV**	**FL-Fitness**	**FL-CV**
1	F-score	0.76716	0.76238	0.82396	0.82609	0.87065	0.82945
	AUC	0.91182	0.91182	0.93607	0.93554	0.94325	0.94049
	AIC	953.81	1462.04	779.71	1196.61	680.55	1076.46
2	F-score	0.70903	0.69504	0.81301	0.81605	0.85668	0.82692
	AUC	0.94537	0.94537	0.95893	0.95744	0.96816	0.96462
	AIC	740.47	1667.22	585.77	1381.42	498.85	1255.87
3	F-score	0.69214	0.71774	0.74899	0.75654	0.80478	0.73894
	AUC	0.93971	0.93971	0.95246	0.95180	0.95910	0.93999
	AIC	710.92	1716.55	602.20	1391.96	548.60	1203.83

Abbreviations: AIC, Akaike information criterion; AUC, area under curve; CV, cross-validation; FL, fused LASSO.

has satisfactory performance when learning from historical data to predict failures on new wafers.

4.2 | Experiments on simulated data

In this section, we conducted more extensive studies using simulated data to show situations in which our proposed method may work well.

As previously mentioned, defect chips on real wafers generally exist as a mixture of 2 spatial patterns: random defects and clustered defects. To consider this mixed-pattern situation in a practical semiconductor manufacturing process, we simulate both random defects and clustered defects on wafers in this section.

4.2.1 | Generation of simulated data

DeNicolao et al[16] simulated binary failure data on wafers assuming that the failure probability has relationships with the center of failure, and this method was also used by Jeong et al[20] and Xie et al[36]. However, this method directly generated the probability of failure according to physical locations and did not explicitly separate clustered defects and random defects. We propose a novel method of generating both functional testing data and the corresponding binary failure data. The probability of failure is generated according to a logistic model, with clustered defects and random defects being considered separately and combined later. Different shapes, including diamond, circular, ring, sector, line (scratch), and semi-ring, are considered under different defect intensities and densities.

To generate both the simulated discrete functional testing data and corresponding binary failure data, the underlying assumption is that the functional testing data follow a Poisson distribution, whose parameters follow a normal distribution. Then, the binary yield values are determined by a logistic model considering the spatial correlations of chips.

The discrete functional testing data are generated according to a Poisson distribution, and the parameter λ is obtained from the following model:

$$\log\lambda = z + \phi_0,$$

where z is a normal vector whose elements are independent normal random variables with mean μ_0 and variance σ_0^2. The spatial correlation term ϕ_0 follows a multivariate normal distribution with mean vector 0 and variance-covariance matrix $\Sigma = \{a_0 \times e^{-b_0 \times d}\}$, where d is the Euclidean distance of 2 chips on a wafer.

Then, we add a cluster with a specific shape on the wafer. The functional testing data on chips that belong to the cluster are also generated according to a Poisson distribution with parameter λ':

$$\log\lambda' = z' + \phi_1.$$

Again, z' is a vector with independent elements following $N(\mu_1, \sigma_1^2)$, and ϕ_1 follows a multivariate normal distribution with mean vector 0 and variance-covariance matrix $\Sigma = \{a_1 \times e^{-b_1 \times d}\}$. The parameters a_0, b_0, a_1, b_1 determine the spatial correlation of measurements. The parameters μ_0 and μ_1 represent the average failure level of chips out of cluster and inside cluster, respectively, and their difference $\mu_1 - \mu_0$ shows the signal strength of clustered defects.

4.2.2 | Performance on different shapes

We first discuss the performance of our proposed method when different shapes of defect clusters occur on wafers.

To investigate the performance of the proposed method when the cluster has different shapes, the parameters that are irrelevant to cluster shape are fixed as follows: $\mu_0 = -2$, $\sigma_0 = 1$, $a_0 = 0.1$, $b_0 = 0.05$, $\mu_1 = 1.5$, $\sigma_1 = 0.2$, $a_1 = 0.9$, and $b_1 = 0.4$. The center of the cluster is set at the site with coordinates $(X = 35, Y = 55)$, which is located near the physical center of the wafer.

Since the underlying defect cluster shape is known beforehand, the true values of the derived variables can be directly calculated in this procedure. Additionally, the failure probability is generated according to a logistic model:

$$\begin{aligned}\text{logit}(p) = \text{Intercept} &+ \gamma \times x + \eta_1 \times \text{In cluster} \times \text{Cluster size} \\ &+ \eta_2 \times \text{Distance from edge} \\ &+ \omega \times \text{Mean measurement of neighbors.}\end{aligned}$$

According to empirical practice, the failure probability is generated with $\gamma = 2$, $\eta_1 = 0.0065$, $\eta_2 = 0.1$, $\omega = 1$, and intercept $= -5$. The binary failure data (ie, response variables) are generated according to a binomial distribution with parameter p. To illustrate the performance of our proposed model on chip-level failure prediction, we compare our model using fused LASSO-derived variables (FL) with 3 other methods: the Basic1 model is simply the logistic model of binary failure value and functional testing data; the Basic2 model adds the variable of neighbors' mean measurement value as an independent variable based on the Basic1 model (these 2 models are explained in detail in Section 4.1.3); and the Real model is the true model that generates all the simulated data. Our proposed method, ie, the FL model, uses the same model structure as the Real model, but the FL model calculates derived variables only based on the fused LASSO results of functional testing data with no predefined information given.

We consider several shapes of failed chip clusters that occur in actual semiconductor manufacturing processes: circle, ring (donut), sector, line and semi-ring.[37] The functional testing data and binary failure values on simulated wafers are shown in Figure 6 and Figure 7. Table 3 summarizes the numerical results. Once again, the F-score, AUC, and AIC are used here to evaluate the performance of the predictive models. As shown in Table 3, for wafers with defects forming closed con-

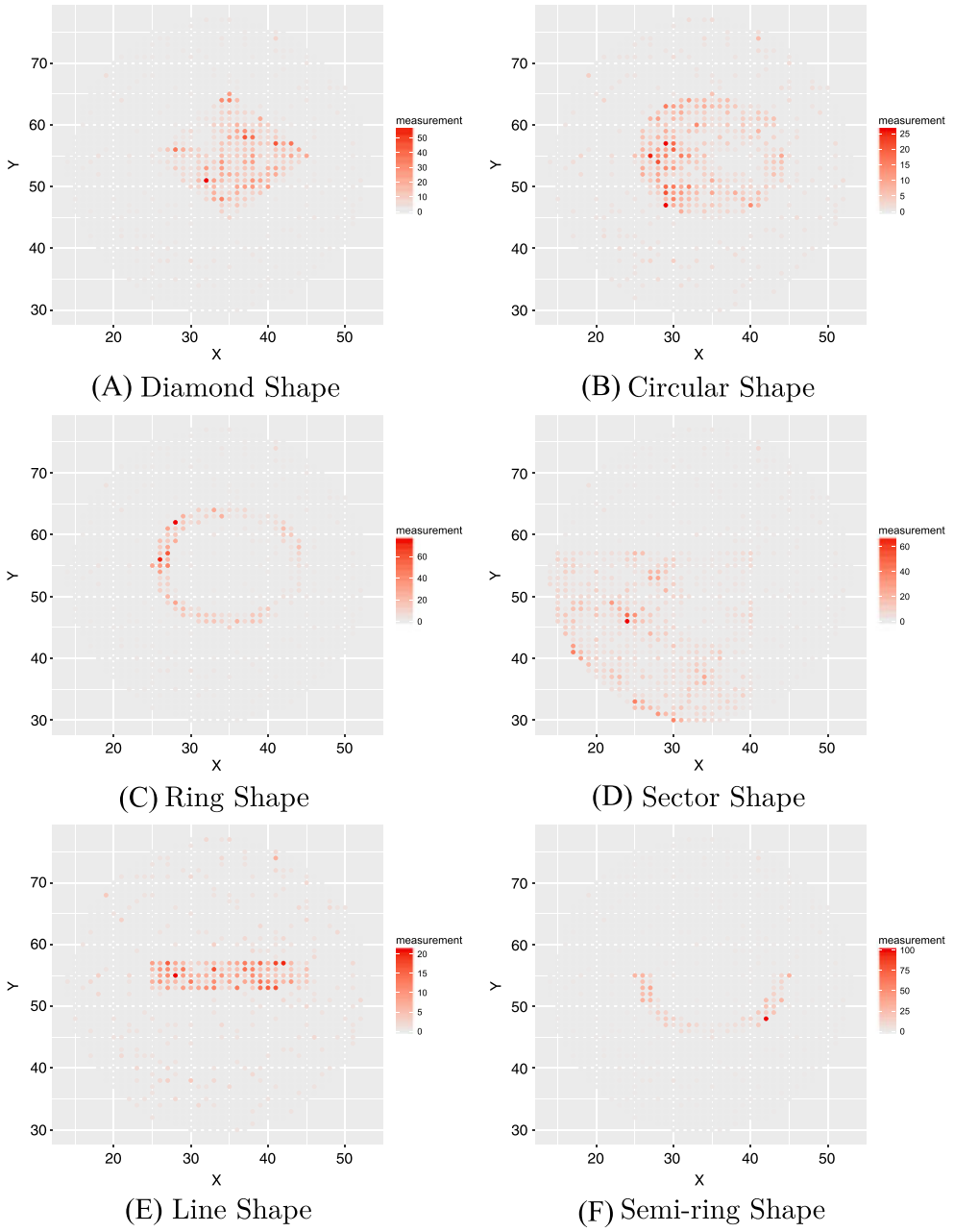


FIGURE 6 Simulated functional testing results on wafers with different defect cluster shapes [Colour figure can be viewed at wileyonlinelibrary.com]

vex shapes such as wide line, circular, and sector, the proposed method tends to provide a higher F-score and larger AUC. However, if the defect cluster is hollow, such as the shape of a ring (donut), our proposed method presents a result that is almost the same as that of a simpler model with only neighbor average measurement value. It is mainly because fused LASSO tends to fuse a convex block based on the neighborhood structure. Our proposed method generally performs between the true underlying model (Real) and the model that only considers neighbor measurements (Basic2).

4.2.3 | Performance on different cluster sizes

The simulation studies in Section 4.2.2 compared the performance of our proposed method when the defect clusters form different shapes and finds that the method works well, particularly when the cluster is in a closed convex shape such as wide lines and circles. To further investigate the impact of cluster size on the performance of our method, we design the following simulation study.

In this simulation, we set the cluster to different sizes when it is of 3 shapes: circle, wide-line, and semi-ring. Moreover, we vary the radius of the circle, edge length of the wide-line, and radius of the semi-ring between 2 levels, 5 and 15, to

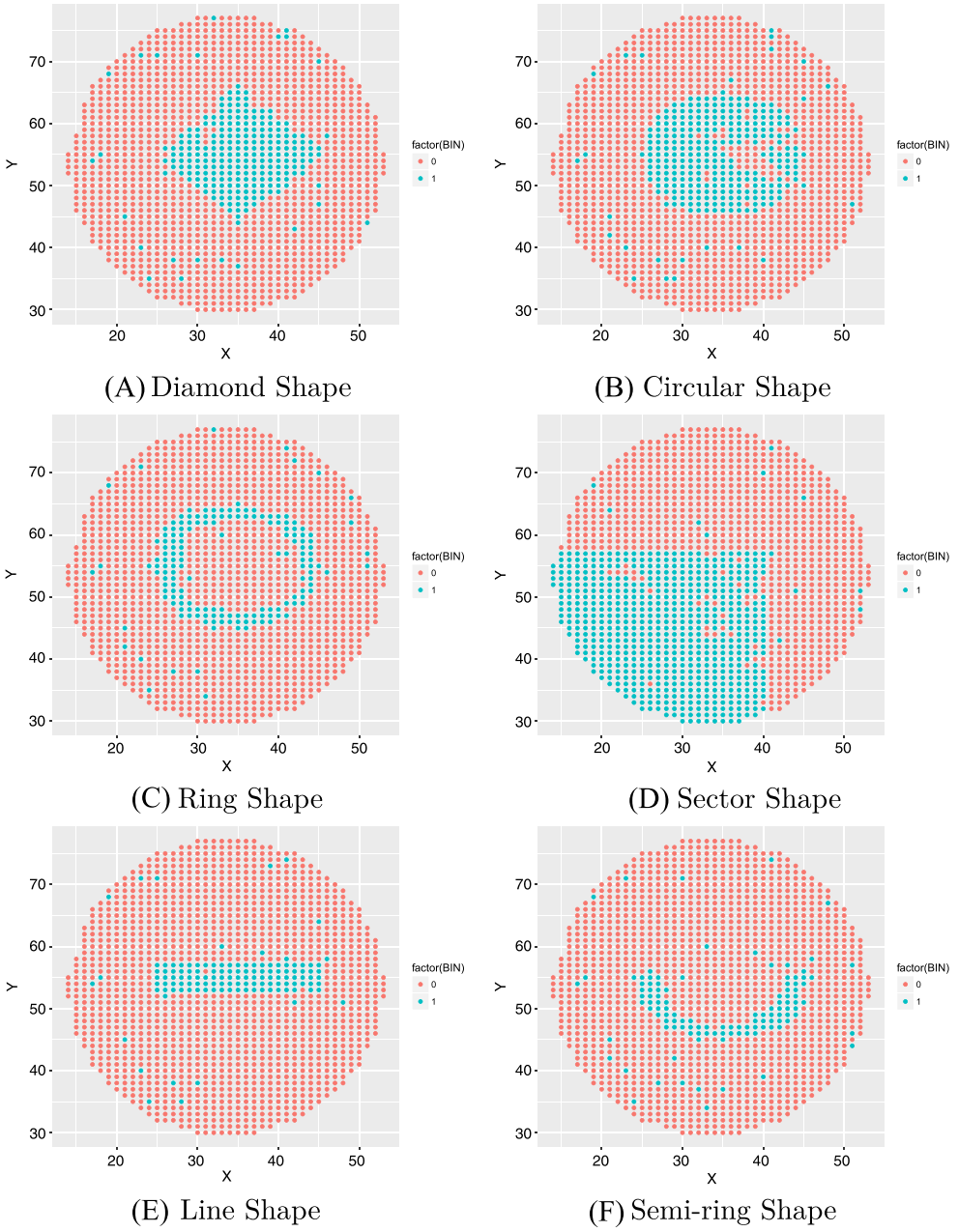


FIGURE 7 Simulated failure data on wafers with different defect cluster shapes [Colour figure can be viewed at wileyonlinelibrary.com]

test the prediction performance of our method and compare it with the **Basic1**, **Basic2**, and **Real** models. The performance is again evaluated in 3 aspects: F-score, AUC, and AIC. The results are shown in Table 4. Table 4 reveals that the effect of size depends on the specific shape of the defect cluster. For the circle shape, the proposed FL method is better when the circle size is not large. However, for the line shape, FL performs even better than the real generative model regarding AUC and AIC both when the line is long and short. For semi-circle, the proposed FL method performs slightly better than the Basic2 model when the shape is small, but it performs almost the same as the Basic2 model when the shape is large. In general, the proposed FL method performs between the Real model and the Basic2 model, and it tends to perform better when the cluster size is not very large.

4.2.4 | Performance on different failure-generating parameters

In this section, we further investigate the impact of the parameter settings on the performance of the proposed fused LASSO–based method. We consider $\sigma_0, \mu_1, \sigma_1$, representing the independent variation of non-clustered chips, mean functional testing results of clustered defective chips and independent variation of clustered defective chips, respectively. Each of the above parameters is set to 2 levels on wafers with circle-shaped defect clusters of the same size, and the results

TABLE 3 Logistic regression results of wafers with different defect cluster shapes

Wafer			Methods			
			Basic1	Basic2	Real	FL
1	Diamond	F-score	0.874	0.931	0.939	0.935
		AUC	0.958	0.990	0.994	0.992
		AIC	386	209	196	203
2	Circle	F-score	0.833	0.844	0.925	0.909
		AUC	0.950	0.983	0.991	0.986
		AIC	543	372	280	337
3	Ring	F-score	0.846	0.886	0.910	0.886
		AUC	0.925	0.985	0.987	0.986
		AIC	526	291	268	295
4	Sector	F-score	0.820	0.925	0.966	0.938
		AUC	0.919	0.979	0.995	0.989
		AIC	900	504	235	389
5	Line	F-score	0.851	0.895	0.902	0.912
		AUC	0.971	0.982	0.989	0.990
		AIC	246	187	178	180
6	Semi-ring	F-score	0.779	0.825	0.840	0.831
		AUC	0.945	0.976	0.980	0.977
		AIC	360	275	258	272

Abbreviations: AIC, Akaike information criterion; AUC, area under curve; FL, fused LASSO.

TABLE 4 Logistic regression results of wafers with different defect cluster sizes

Wafer			Methods			
			Basic1	Basic2	Real	FL
1	Circle (large)	F-score	0.863	0.914	0.969	0.928
		AUC	0.915	0.979	0.995	0.982
		AIC	930	555	271	514
2	Circle (small)	F-score	0.826	0.882	0.883	0.884
		AUC	0.941	0.987	0.989	0.989
		AIC	290	180	169	173
3	Line (long)	F-score	0.876	0.920	0.932	0.929
		AUC	0.974	0.990	0.992	0.992
		AIC	264	170	166	165
4	Line (short)	F-score	0.742	0.791	0.791	0.788
		AUC	0.957	0.971	0.976	0.977
		AIC	221	190	183	182
5	Semicircle (large)	F-score	0.797	0.856	0.891	0.855
		AUC	0.948	0.978	0.982	0.978
		AIC	401	284	251	288
6	Semicircle (small)	F-score	0.770	0.838	0.847	0.843
		AUC	0.918	0.973	0.975	0.975
		AIC	369	250	243	246

Abbreviations: AIC, Akaike information criterion; AUC, area under curve; FL, fused LASSO.

are shown in Table 5. This table shows that large σ_0, σ_1 could both make the AIC larger, which indicates that the logistic regression model generally performs worse if the independent variation increases, whether inside the defect cluster or outside the cluster. Moreover, when μ_1 is larger, the difference between the measurements on chips in cluster and out of cluster becomes larger, and it is easier to separate defective chips from the normal ones, which leads to a lower AIC.

We also investigate the performance of the proposed fused LASSO–based method when the variance and covariance structure of the spatial-correlated term varies. Recall that the variance-covariance matrix $\Sigma = \{a_0 \times e^{-b_0 \times d}\}$, we set a_1 at 4

TABLE 5 Logistic regression results of wafers with different failure strengths and spatial variations

Wafer		Methods			
		Basic1	Basic2	Real	FL
1	F-score	0.842	0.912	0.941	0.923
$(\sigma_0, \mu_1, \sigma_1)$	AUC	0.946	0.987	0.994	0.991
(0.5, 1.5, 0.2)	AIC	592	350	244	301
2	F-score	0.817	0.879	0.927	0.903
$(\sigma_0, \mu_1, \sigma_1)$	AUC	0.936	0.976	0.991	0.984
(1.5, 1.5, 0.2)	AIC	749	517	323	429
3	F-score	0.778	0.896	0.923	0.897
$(\sigma_0, \mu_1, \sigma_1)$	AUC	0.910	0.972	0.988	0.978
(1.0, 1.0, 0.2)	AIC	759	445	310	419
4	F-score	0.883	0.930	0.958	0.937
$(\sigma_0, \mu_1, \sigma_1)$	AUC	0.953	0.990	0.993	0.991
(1.0, 2.0, 0.2)	AIC	558	310	230	301
5	F-score	0.818	0.902	0.924	0.908
$(\sigma_0, \mu_1, \sigma_1)$	AUC	0.935	0.982	0.988	0.984
(1.0, 1.5, 0.5)	AIC	682	399	320	392
6	F-score	0.810	0.911	0.934	0.914
$(\sigma_0, \mu_1, \sigma_1)$	AUC	0.923	0.983	0.989	0.986
(1.0, 1.5, 0.8)	AIC	735	395	301	384

Abbreviations: AIC, Akaike information criterion; AUC, area under curve; FL, fused LASSO.

TABLE 6 Logistic regression results of wafers with different spatial variations

Wafer		Methods			
		Basic1	Basic2	Real	FL
1	F-score	0.910	0.963	0.976	0.974
(a_1, b_1)	AUC	0.964	0.993	0.994	0.994
(0.2, 0, 4)	AIC	487	217	176	183
2	F-score	0.881	0.948	0.970	0.960
(a_1, b_1)	AUC	0.956	0.993	0.995	0.994
(0.4, 0.4)	AIC	555	254	178	210
3	F-score	0.856	0.935	0.955	0.950
(a_1, b_1)	AUC	0.946	0.988	0.991	0.991
(0.6, 0.4)	AIC	619	337	254	273
4	F-score	0.856	0.930	0.954	0.945
(a_1, b_1)	AUC	0.941	0.987	0.991	0.991
(0.8, 0.4)	AIC	638	324	250	285
5	F-score	0.812	0.870	0.913	0.880
(a_1, b_1)	AUC	0.924	0.979	0.988	0.978
(0.9, 0.2)	AIC	694	472	326	453
6	F-score	0.848	0.936	0.961	0.954
(a_1, b_1)	AUC	0.936	0.988	0.991	0.991
(0.9, 0.8)	AIC	666	311	229	248

Abbreviations: AIC, Akaike information criterion; AUC, area under curve; FL, fused LASSO.

different levels, and set b_1 at 2 levels. The results are shown in Table 6. Since b_1 represents the decay rate of covariance over distance, when b_1 increases, the decay rate also increases, and this correlation structure is more similar to the neighbor structure that we adopt here. Thus, when b_1 is large, our proposed method is almost as good as the model that uses the true values of generating parameters.

5 | DISCUSSIONS

In this work, we proposed a model for yield prediction with derived variables. Compared to existing models, performance studies show that the proposed model has improved performance. However, the performance of the model is affected by other factors such as parameter setting and spatial cluster characteristics. In the following, we further discuss the selection of such parameters and the sensitivity of the method to cluster characteristics.

5.1 | Guidelines for parameter setting

The performance of the fused LASSO procedure depends on the setting of the penalty strength λ, and although not essentially, it is affected by the choice of neighborhood structure.

The fused LASSO step is used to derive variables that could characterize defect cluster information. Bearing in mind that the ultimate purpose of this procedure is to use the defect cluster information to make better predictions on yields, how to select an appropriate λ is primarily determined by the prior knowledge of the defect clusters. A rule-of-thumb is to set λ as half of the maximum value in the entire path of the fused LASSO algorithm (ie, the first value hitting the boundary). In fact, choosing λ between a specific interval from approximately 1/3 of the first value to 2/3 of the first value will not make a significant difference in the performance of the model. According to different data characteristics, an appropriate λ could be selected through training on several samples. Note that the insensitivity of the prediction result to the λ in the above interval helps to make the parameter setting work less technical. There is no need to stick to finding the optimal λ for the training samples.

The neighborhood structure that we consider in the experiment is the 4-neighbor structure. This structure forms a rectangle if there is a circular cluster with a decrease of λ in the fused LASSO procedure. The 8-neighbor structure could also be used, with which it would tend to form a rectangle first when there is a circular cluster. We have conducted experiments comparing different neighborhood structures, and the result shows no significant performance difference. However, the 4-neighbor structure is more computationally efficient; thus, generally adopting the 4-neighbor structure is recommended.

5.2 | Sensitivity to defect cluster

The proposed method does not actually classify which shape the cluster belongs to, but it is important to consider the types of defect clusters where our method is less helpful on the prediction performance. Since the main idea of the proposed method is to take advantage of spatial clusters on a wafer to derive additional informative variables, thus contribute to yield prediction, the usefulness (or potential contribution) of the derived variables is inevitably affected by the exact shape of a cluster. From the definition of our derived variables, a cluster that is convex and smooth in edge is more predictable using these derived variables (as in such cases, the status of each chip is similar to its surroundings and thus easier to predict if taking spatial information into consideration). In practice, as the exact shape and location of a cluster is unknown, the performance of the prediction model is also affected by the cluster identification algorithm. In this work, we adopted the fused LASSO procedure; if the shape of a cluster cannot be accurately identified by the algorithm, the contribution of the derived variables might be restricted.

Moreover, the performance of the proposed method is closely related to the size and the shift magnitude of the defect cluster. A strong shift signal of large area helps to clearly detect a cluster. Another important issue is the count of defect clusters. If more than one defect cluster exists, more details, such as the distance of clusters and relative strength of shift signals would also influence model performance. Intuitively, these factors would affect the cluster identification procedure, since the identification algorithm could only separately detect 2 clusters if they are far enough from each other or the shift signal is strong enough.

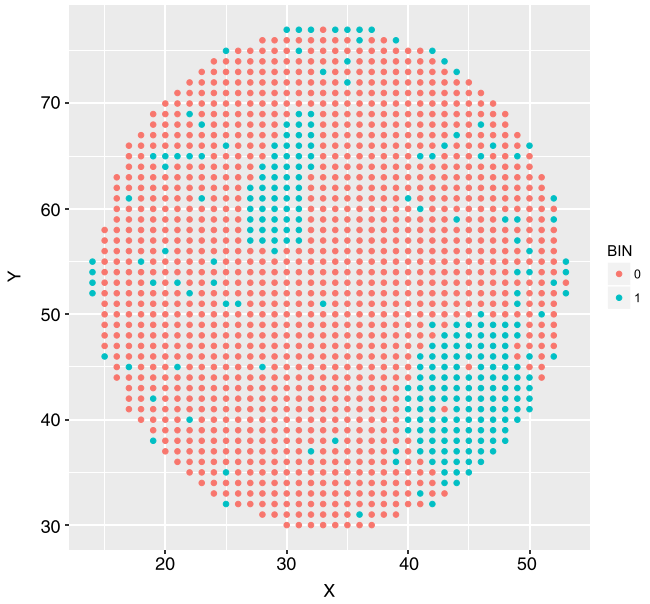


FIGURE 8 Diagnosis result of Wafer 3 used in Section 4.1 [Colour figure can be viewed at wileyonlinelibrary.com]

In Section 4.1, the proposed method is conducted on a wafer with 2 defect clusters(Wafer 3) as shown in Figure 8, and our method gives pretty good performance in this case. Although our method does not require that there cannot be more than one cluster, the performance of our method in various situations mentioned above has not been fully investigated. For example, if there are 2 clusters that are too close to each other, the fused LASSO procedure of our proposed method would probably identify them and the gap between them as a big cluster, and in this case, the yield prediction result of the chips between these 2 clusters would be largely affected, namely, good chips may probably be identified as failed chips in this area. This could be further analyzed in a future work.

6 | CONCLUSIONS

In this paper, a novel wafer yield prediction model based on a fused LASSO procedure is proposed. It incorporates spatial defect cluster information in indicative functional testing data. A fused LASSO procedure is adopted to derive several variables that characterize the clustering features of defect chips, and then these derived variables are included in a logistic regression model to predict the binary outcome of chip-level yield in field use.

The performance of our approach is evaluated both on real wafer data and on simulated examples. The results indicate that explicitly considering the spatial clustering information by adding variables derived from the fused LASSO procedure could improve the prediction results at both the chip level and wafer level.

The production of chips includes thousands of procedures, and vast amounts of functional testing data are collected during these procedures. This work provides a solution to learn from these procedure data to better understand the production process and to provide more accurate predictions on the future usability of chips. We believe that considerable work remains to be undertaken. First, how to combine spatial features of multiple measurement variables should be considered since different measurements may provide overlapped information with each representing some unique features. Second, additional studies on the settings of parameters and the detectability domain should be performed. As we have previously discussed, the choice of λ and the neighborhood structure could affect those derived variables. However, we still need theoretical guidelines on choosing optimal parameters. Additionally, regarding the strength of the shift signal and the sizes and locations of clusters, more extensive studies should be conducted to identify the situations where defect clusters could be characterized by the proposed cluster identification algorithm. Third, a "shape-specific" model could also be considered to use the information of a particular shape to further improve the performance of the proposed model. The fused LASSO algorithm or some methods that are capable of identifying exact cluster shape and location can be used to identify clusters first, then a directional model that takes such shape information into considering is potentially more efficient in prediction. As accurate prediction of wafer yield is a practically important while methodologically challenging problem, the above topics are promising in performance but all still deserve further research efforts.

ACKNOWLEDGMENTS

The authors would like to thank the editor and 2 anonymous referees for their valuable comments, which have helped us improve this work greatly. The authors are grateful to Samsung Electronics for providing financial support and sample data to this research. Dr Wang's research is supported by the National Natural Science Foundation of China under grant no. 71471096.

REFERENCES

1. Tong SH, Yum BJ. A dual burn-in policy for defect-tolerant memory products using the number of repairs as a quality indicator. *Microelectron Reliab*. 2008;48:471-480.
2. Tong SH, Yum BJ. Development of a dual burn-in policy for semiconductor products based on the number of defective neighborhood chips. *Int J Reliab Qual Saf Eng*. 2006;13:501-525.
3. Mirza AI, O'Donoghue G, Drake AW, Graves SC. Spatial yield modeling for semiconductor wafers. In: Advanced Semiconductor Manufacturing Conference and Workshop, Cambridge, Massachusetts, USA; 1995. ASMC 95 Proceedings IEEE/SEMI 1995; 1995:276-281.
4. Hansen MH, Nair VN, Friedman DJ. Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects. *Technometrics*. 1997;39:241-253.
5. Cunningham SP, MacKinnon S. Statistical methods for visual defect metrology. *IEEE Trans Semicond Manuf*. 1998;11:48-53.
6. Chen FL, Liu SF. A neural-network approach to recognize defect spatial pattern in semiconductor fabrication. *IEEE Trans Semicond Manuf*. 2000;13:366-373.
7. Cunningham SP, Shanthikumar JG. Empirical results on the relationship between die yield and cycle time in semiconductor wafer fabrication. *IEEE Trans Semicond Manuf*. 1996;9:273-277.
8. Spanos CJ. Statistical process control in semiconductor manufacturing. *Microelectron Eng*. 1991;10:271-276.
9. Taam W, Hamada M. Detecting spatial effects from factorial experiments: An application from integrated-circuit manufacturing. *Technometrics*. 1993;35:149-160.
10. Friedman DJ, Hansen MH, Nair VN, James DA. Model-free estimation of defect clustering in integrated circuit fabrication. *IEEE Trans Semicond Manuf*. 1997;10:344-359.
11. Ge Z, Song Z. Semiconductor manufacturing process monitoring based on adaptive substatistical PCA. *IEEE Trans Semicond Manuf*. 2010;23:99-108.
12. Wang A, Wang K, Tsung F. Statistical surface monitoring by spatial-structure modeling. *J Qual Technol*. 2014;46:359-376.
13. Wang K, Jiang W, Li B. A spatial variable selection method for monitoring product surface. *Int J Prod Res*. 2016;54:4161-4181.

14. Zhang L, Wang K, Chen N. Monitoring wafer geometric quality using additive Gaussian processes. *IIE Trans*. 2016;48:1-15.
15. Ooi MPL, Sok HK, Kuang YC, Demidenko S, Chan C. Defect cluster recognition system for fabricated semiconductor wafers. *Eng Appl Artif Intell*. 2013;26:1029-1043.
16. DeNicolao G, Pasquinetti E, Miraglia G. Unsupervised spatial pattern classification of electrical failures in semiconductor manufacturing. In: Proc. Artificial Neural Networks Pattern Recognition Workshop, Florence (Italy); 2003. http://www.dsi.unifi.it/ANNPR/
17. Hsieh HW, Chen FL. Recognition of defect spatial patterns in semiconductor fabrication. *Int J Prod Res*. 2004;42:4153-4172.
18. Wang CH, Wang SJ, Lee WD. Automatic identification of spatial defect patterns for semiconductor manufacturing. *Int J Prod Res*. 2006;44:5169-5185.
19. Yuan T, Kuo W. Spatial defect pattern recognition on semiconductor wafers using model-based clustering and Bayesian inference. *Eur J Oper Res*. 2007;190:228-240.
20. Jeong YS, Kim SJ, Jeong MK. Automatic identification of defect patterns in semiconductor wafer maps using spatial correlogram and dynamic time warping. *IEEE Trans Semicond Manuf*. 2008;21:625-637.
21. Hsieh TJ, Liao CS, Huang YS, Chien CF. A new morphology-based approach for similarity searching on wafer bin maps in semiconductor manufacturing. In: IEEE 16th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE; Wuhan, China; 2012:869-874.
22. Stapper CH. LSI yield modeling and process monitoring. *IBM J Res Dev*. 1976;20:228-234.
23. Cunningham JA. The use and evaluation of yield models in integrated circuit manufacturing. *IEEE Trans Semicond Manuf*. 1990;3:60-71.
24. Stapper CH. Defect density distribution for LSI yield calculations. *IEEE Trans Electron Devices*. 1973;20:655-657.
25. Koren I, Koren Z, Stepper CH. A unified negative-binomial distribution for yield analysis of defect-tolerant circuits. *IEEE Trans Comput*. 1993;42:724-734.
26. Kumar N, Kennedy K, Gildersleeve K, Abelson R, Mastrangelo CM, Montgomery DC. A review of yield modelling techniques for semiconductor manufacturing. *Int J Prod Res*. 2007;44:5019-5036.
27. Longtin MD, Wein LM, Welsch RE. Sequential screening in semiconductor manufacturing, I: exploiting spatial dependence. *Oper Res*. 1996;44:173-195.
28. Wu L, Zhang J. Fuzzy neural network based yield prediction model for semiconductor manufacturing system. *Int J Prod Res*. 2009;48:3225-3243.
29. Bae SJ, Hwang JY, Kuo W. Yield prediction via spatial modeling of clustered defect counts across a wafer map. *IIE Trans*. 2007;39:1073-1083.
30. Yuan T, Ramadan SZ, Bae SJ. Yield prediction for integrated circuits manufacturing through hierarchical bayesian modeling of spatial defects. *IEEE Trans Reliab*. 2011;60:729-741.
31. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Series B-Methodol*. 1996;58:267-288.
32. Hoefling H. A path algorithm for the fused LASSO signal approximator. *J Comp Graph Stat*. 2010;19:984-1006.
33. Tibshirani RJ, Taylor J. The solution path of the generalized Lasso. *Ann Stat*. 2011;39:1335-1371.
34. Arnold TB, Tibshirani RJ. genlasso: path algorithm for generalized lasso problems. https://CRAN.R-project.org/package=genlasso; 2014. Accessed on May 1, 2016
35. Zhang N, Yang Q. A random effect autologistic regression model with application to the characterization of multiple microstructure samples. *IIE Trans*. 2016;48:34-42.
36. Xie L, Huang R, Gu N, Cao Z. A novel defect detection and identification method in optical inspection. *Neural Comput Appl*. 2013;24:1953-1962.
37. Liu CW, Chien CF. An intelligent system for wafer bin map defect diagnosis: an empirical study for semiconductor manufacturing. *Eng Appl Artif Intell*. 2013;26:1479-1486.

Hang Dong is a PhD candidate in the Department of Industrial Engineering, Tsinghua University. At the meantime, she is a research engineer at National University of Singapore. She obtained her bachelor's degree in Industrial Engineering from Tsinghua University in 2015. Her research interests include statistical modeling and monitoring of complex data.

Nan Chen is an Assistant Professor in the Department of Industrial Systems Engineering and Management at National University of Singapore. He obtained his BS degree in Automation from Tsinghua University and MS degree in Computer Science, MS degree in Statistics, and PhD degree in Industrial Engineering from University of Wisconsin-Madison. His research interests include statistical modeling and surveillance of service systems, simulation modeling design, condition monitoring, and degradation modeling. He is a member of INFORMS, IIE, and IEEE.

Kaibo Wang is a Professor in the Department of Industrial Engineering, Tsinghua University, Beijing, China. He received his BS and MS degrees in Mechatronics from Xi'an Jiaotong University, Xi'an, China, and his PhD in Industrial Engineering and Engineering Management from the Hong Kong University of Science and Technology, Hong Kong. He is an Associate Editor of IISE Transactions and also on the Editorial Review Board of *Journal of Quality Technology*. His research focuses on statistical quality control and data-driven complex system modeling, monitoring, diagnosis, and control, with a special emphasis on the integration of engineering knowledge and statistical theories for solving problems from real industries. He is a senior member of ASQ and a member of INFORMS and IISE.

How to cite this article: Dong H, Chen N, Wang K. Wafer yield prediction using derived spatial variables. *Qual Reliab Engng Int*. 2017;33:2327–2342. https://doi.org/10.1002/qre.2192