

© 2015

BYUNGHOON KIM

ALL RIGHTS RESERVED

ADVANCED SPATIAL DATA MINING METHODOLOGY
AND ITS APPLICATIONS TO SEMICONDUCTOR
MANUFACTURING PROCESSES

by

BYUNGHOON KIM

A dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Industrial and Systems Engineering

Written under the direction of

Professor Myong K. Jeong

And approved by

New Brunswick, New Jersey

May, 2015

ABSTRACT OF THE DISSERTATION

Advanced Spatial Data Mining Methodology and Its Applications to Semiconductor Manufacturing Processes

By BYUNGHOON KIM

Dissertation Director: Dr. Myong K. Jeong

In this dissertation, we present several methodologies for mining data obtained in semiconductor manufacturing processes. We first present a new step-down spatial randomness test aiming at automatically detecting abnormal dynamic random access memory (DRAM) wafers with multiple spatial maps. Testing the spatial randomness of a DRAM wafer is challenging. A DRAM wafer includes multiple spatial maps, resulting to a more complex and lengthy testing process compared with that of a single wafer map of a flash memory. To monitor the spatial randomness of the multiple spatial maps, we propose a new step-down spatial randomness test to detect abnormal DRAM wafers. In the proposed methodology, we adopt nonparametric Gaussian kernel-density estimation to transform the original fail bit test (FBT) values into binary FBT values. We also propose a spatial local de-noising method to eliminate noisy defect chips to distinguish the random defect patterns from systematic ones.

Secondly, we propose a novel matrix factorization method, called regularized singular value decomposition (RSVD), which aims at the automated classification of chip level failure patterns on fail bit map (FBM) of each DRAM chip. The RSVD based approach decomposes a FBM into several binary eigen-images to extract features that can provide the characteristics of the failure patterns on the FBM. By employing the extracted features as input vectors, k -nearest neighbor (k -NN) classifier is applied to classify feature patterns on a FBM into either single bit failed one or non-single bit failed one.

Finally, we propose a new Bayesian classification model for uncertain data to classify abnormal DRAM wafers that include spatial features with uncertainty. Bayesian classifier has been extensively used for the classification of certain data. However, since every data object in the uncertain data is not represented by a point value, it is difficult to directly apply the Bayesian classifier for certain data. In the proposed approach, the multivariate kernel density estimate for uncertain data is proposed to estimate the class conditional probability density function (pdf). We then apply the Bayes theorem to calculate the posterior probability of a testing data object based on the estimated class conditional pdf.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere appreciation to my advisor, Professor Myong K. Jeong for his excellent guidance and support for my Ph.D. study. His guidance helped me to think critically and smartly in my research and writing of this thesis. I could not imagine a better advisor and mentor for my Ph.D. study. I must also thank my dissertation committee members, Professor Susan Albin, Professor Mohsen Jafari, and Professor Jingang Yi, for their encouragement, support, and insightful comments.

I would like to especially thank my wife, June Young Jeong, who have always trusted me and encouraged me with unlimited love and patience. I am also thankful to my son, Ethan Kim who always gives me such great pleasure. Without my wife and son, I could not have done this.

Finally, I must thank my friends in the Rutgers University and my families in South Korea-- they helped me to finalize my goal, and importantly, made my life most happy.

Table of Contents

ABSTRACT OF THE DISSERTATION	ii
ACKNOWLEDGEMENTS.....	iv
CHAPTER 1 Introduction.....	1
1.1 Overview.....	1
1.2 Thesis outline	4
CHAPTER 2 Step-Down Spatial Randomness Test for Detecting Abnormalities in DRAM Wafers with Multiple Spatial Maps	5
2.1 Introduction.....	5
2.2 Join count statistics and spatial correlograms for multiple spatial maps	11
2.3 Step-down randomness test for detecting abnormal DRAM wafers	14
2.3.1 Binarization of original FBT values for each FBT map based on Gaussian kernel- density estimation	14
2.3.2 New spatial local de-noising method for each binarized FBT map	15
2.3.3 New step-down spatial randomness test using N de-noised binarized FBT maps ..	20
2.4 Experimental results.....	22
2.4.1 Description of real DRAM wafers.....	22
2.4.2 Step-down randomness test for DRAM wafer	23
2.4.2.1 Binarization of original FBT values	23
2.4.2.2 Spatial local de-noising of the binarized FBT map	25
2.4.2.3. Spatial step-down randomness test using the spatial correlograms of de-noised FBT maps	26

2.4.3 Performance comparison of step-down randomness test	28
2.5 Conclusions and future research	31
CHAPTER 3 A Regularized Singular Value Decomposition-based Approach for Failure Pattern Classification on Fail Bit Map in a DRAM Wafer	32
3.1 Introduction	32
3.2 Fail bit map	35
3.3 Regularized singular value decomposition	37
3.4 Classification of failure patterns on FBMs using RSVD	42
3.4.1 Binarization of FBMs	42
3.4.2 Feature extraction from binarized FBMs using RSVD	44
3.4.3 Classification of the failure patterns on FBMs using k-NN classifier	45
3.5 Experimental results	46
3.6 Conclusion	53
CHAPTER 4 A New Multivariate Kernel Density Estimation of Uncertain Data for Defect Classification of Wafer Maps	54
4.1 Introduction	54
4.2 Bayesian classification model for uncertain data	58
4.2.1 Problem definition	59
4.2.2 Estimation of class conditional joint pdf based on multivariate KDE	60
4.2.2.1 Review of the traditional kernel density estimate for certain data	61
4.2.2.2 New multivariate kernel density (MKDE) estimation for uncertain data	62
4.2.2.2.1 Uncertain continuous data	62
4.2.2.2.2 Uncertain categorical data	65

4.2.2.2.3 Uncertain mixed data	67
4.2.2.3 Bandwidth selection of the proposed kernel density estimate	68
4.3 Case study: defect classification of DRAM wafers	69
4.3.1 Transformation of original FBT values into binary ones.....	69
4.3.2 Extraction of uncertain spatial feature	71
4.3.2.1 Uncertain feature of the critical map index	71
4.3.2.2 Uncertain feature related to the defect shape of the critical FBT map	74
4.3.2.3 Uncertain feature related to the defect location of the critical FBT map	76
4.3.3 An illustrative example.....	77
4.4 Performance comparison.....	79
4.5 Conclusion.....	85
CHAPTER 5 Concluding Remarks and Future Researches	86
5.1 Concluding remarks	86
5.2 Future researches	88
Appendix A. Derivation of Control Limits of Step-Down Spatial Randomness Test.....	89
Appendix B. Learning Algorithm for RSVD.....	91

List of Figures

Figure 2.1. Examples of spatial defect patterns on wafer maps.....	6
Figure 2.2 Original FBT maps for the DRAM wafer described in Table 2.1.....	9
Figure 2.3. Framework of our proposed procedure for detecting an abnormal DRAM wafer.	10
Figure 2.4 Neighbors in a rook-move neighborhood.....	12
Figure 2.5 3×3 spatial filter mask for the bottom edge of a defective chip.	18
Figure 3.1 Flowchart of the proposed approach for FBM pattern classifications.	35
Figure 3.2 An example of a fail bit map.	36
Figure 3.3 Example of failure patterns on FBMs.	37
Figure 3.4 Binarized FBMs of the single-bit and non-single bit failed chips.....	43
Figure 3.5 Eigen-images of the binarized FBMs for the defect types of failed chips.	45
Figure 3.6 Matrix norms of the eigen-images based on SVD and RSVD.	51
Figure 3.7 First eigen-images of the binarized FBMs based on SVD and RSVD.....	52
Figure 4.1 Multiple FBT maps Figure of a given DRAM wafer.	56
Figure 4.2 Contour plot of density estimates based on the bivariate kernel functions;	64
Figure 4.3 Framework of the defect classification of a DRAM wafer.	69
Figure 4.4 A spatial correlogram of an abnormal wafer map for each typical defect shape.	75
Figure 4.5 Three subareas of a binarized FBT map.....	77
Figure 4.6 Estimation of joint probability based on extracted uncertain features.	78
Figure 4.7 Binarized FBT maps of a DRAM wafer.....	83

List of Tables

Table 2.1 Data structure of test result from a DRAM wafer	8
Table 2.2 Threshold of each FBT for binarization	24
Table 2.3 Sensitivity of the parameter of each process in the proposed step-down spatial randomness test.....	27
Table 2.4 Sensitivity of each class for the different overall false alarm rates of the proposed step-down randomness test.....	28
Table 2.5 Comparison of the accuracy of various randomness tests	30
Table 3.1 Classification accuracy of failed chips using different methods of features extraction.....	47
Table 3.2 Classification accuracy between neural networks and 1-NN classifier with RSVD	50
Table 4.1 Uncertain categorical data extracted from DRAM wafers	57
Table 4.2 Performance comparison of the different classification models.....	80
Table 4.3 Joint probability distributions of uncertain features of two DRAM wafers that belong to the different classes.....	81
Table 4.4 Joint probability distribution of uncertain features representing the spatial characteristics of the binarized FBT maps in Figure 4.7	83

CHAPTER 1

Introduction

1.1 Overview

Several hundred integrated circuits (ICs) are simultaneously fabricated on a single wafer (Fenner *et al.* 2005). A semiconductor wafer includes rich dataset that are categorized wafer level and chip level data. At the wafer level analysis, it is critical to locate defective ICs chips on the wafer. Defective chips on wafer map commonly occur in clusters or display systematic patterns due to process problems (Hansen *et al.* 1997). To this end, the wafer maps are usually employed to investigate spatial dependence of defective chips across wafers. At the chip level analysis, a fail bit map is commonly used as an important diagnosis tool for testing memory devices (Nakamae et al, 2001; Wang, 2009). The predominant failure modes of a memory device could quickly be discovered by the characteristics of the failure patterns on a fail bit map (FBM) (Hamada et al, 1997; Zanon et al, 2003; Jung, 2011).

Typically, engineers classify and summarize the defect information from a number of wafer maps and FBMs manually, as well as many yield related professionals are charged with verifying defect types of every device. However, due to dense memory devices and massive production, it becomes challenging to analyze the maps and to predict the causes. In this dissertation, we focus on the development of algorithms that automatically detect and classify abnormal wafer and defective chips.

To automatically detect an abnormal wafer, a great deal of research has been conducted to develop methodologies that capture the spatial pattern shown on a wafer map (Chen and Liu, 2000; Liu *et al.*, 2002; Hsieh and Chen, 2004; Tong *et al.*, 2005; Jeong *et al.*, 2008). However, most of the research on detection of defect pattern of a semiconductor wafer has been limited to flash memory or SRAM. There is only little literature on the analysis of dynamic random-access memory (DRAM) data with multiple wafer maps. We have developed a new step-down spatial randomness test for detecting abnormalities on multiple spatial maps of a DRAM wafer. Based on the proposed randomness test, we sequentially check the spatial randomness of each wafer map for a DRAM wafer.

At the chip level, some researches on analyzing failure patterns on fail bit maps (FBMs) have been conducted in the literature (Hsieh *et al.*, 2010; Zanon *et al.*, 2003; Han *et al.*, 2005). FBMs represent the failed cell count during wafer functional tests and have been popularly used as one of diagnosis tools in semiconductor manufacturing for process monitoring, root cause analysis, and yield improvements. However, the visual inspection process is costly and time-consuming. Therefore, we propose an automated classification procedure for failure patterns on FBMs in dynamic random access memory (DRAM) wafers. The novel matrix factorization approach, called regularized singular value decomposition (RSVD), is proposed to decompose binarized FBMs into several binary eigen-images to extract features that can provide the characteristics of the failure patterns on FBMs. By using the extracted features, k-nearest neighbor (k-NN) classifier is employed to classify feature patterns on FBMs into single bit failed maps and non-single bit failed ones.

Many spatial features that describe defect characteristics on a wafer map or FBM have been extracted for classifying the defect pattern on a semiconductor wafer. However, due to the complexity of the wafer map or FBM, spatial features are always incomplete for representing the characteristics of the images. The incomplete spatial features generate the uncertainty in representing the spatial pattern, which may deteriorate the performance of the defect classification. Therefore, it is unquestionable that handling the uncertainty of the spatial features should be adeptly conducted in order to improve the accuracy of the defect classification.

To handle the uncertainty, we propose a new Bayesian classification model that considers the uncertainty of spatial features. In the proposed approach, the multivariate kernel density estimate for uncertainty is proposed to estimate the class conditional probability density function. Based on the class conditional density estimates and class prior probability, the posterior class probability of a test data with unknown class label is calculated by using Bayes theorem. After a classification model for uncertain data is constructed, the test data is assigned to the class with the maximum posterior class probability.

1.2 Thesis outline

This thesis is organized as follows. Chapter 2 presents the detection methodology of spatial defect patterns on multiple spatial data with the specific application to a DRAM wafer map analysis in a semiconductor manufacturing process. A new step-down randomness test is developed to sequentially monitor the spatial defect patterns on the multiple spatial maps of a DRAM wafer. Chapter 3 proposes a regularized singular value decomposition algorithm for fault pattern classification on fail bit map of a failed DRAM chip. By using the RSVD algorithm, we can extract novel features of the fail bit map that provide the good classification accuracy of the failed DRAM chip. Chapter 4 presents a new Bayesian classifier for uncertain data to identify the fault type of an abnormal DRAM wafer. Finally, Chapter 5 summarizes the research results and describes several research problems for future investigations.

CHAPTER 2

Step-Down Spatial Randomness Test for Detecting Abnormalities in DRAM Wafers with Multiple Spatial Maps

2.1 Introduction

Fabricating integrated circuits (ICs) on a single semiconductor wafer encompasses hundreds of steps (Fenner *et al.*, 2005). After being fabricated, each IC is tested and classified either functional or defective, creating a wafer map that shows the locations of all defective chips on the wafer. Chips may be defective at a random pattern or in several patterns (Hansen *et al.* 1997). The spatial pattern of defective chips provides information useful for improving the quality of a wafer (Taam and Hamada, 1993; Chou *et al.*, 1997; Tong and Yum, 2006).

Figure 2.1 shows four wafer maps with different spatial defect patterns. Figure 2.1 (a) shows a random pattern from an in-control process. In contrast, Figure 2.1 (b), (c), and (d) show some spatial defect patterns from out-of-control processes. Such defect patterns contain much information useful for tracking the causes of the out-of control process (Wang, 2008; Li and Huang, 2009; Chang *et al.*, 2012). For example, elliptical zones of failure are often caused by issues in thin-film deposition. Circular failure can be caused by etching issues. Repetitive patterns result mainly from stepper (or probe) malfunctions and sawing imperfections. Linear scratches are often caused by machine-handling problems (Hansen *et al.*, 1997; Chen and Liu, 2000; Liu *et al.*, 2002).

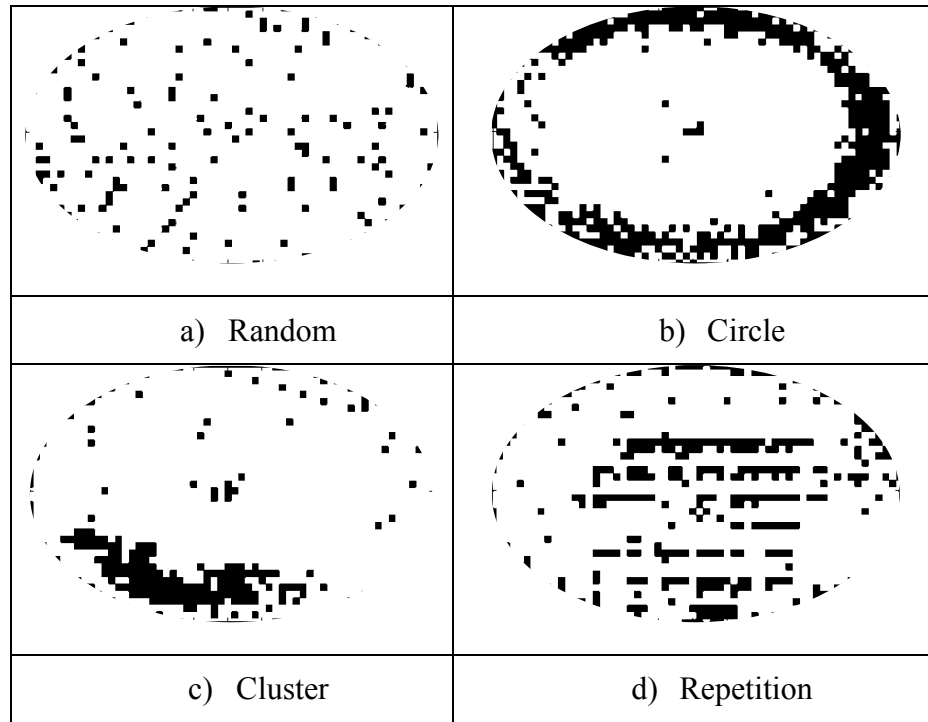


Figure 2.1 Examples of spatial defect patterns on wafer maps. Black dots indicate defective chips; white dots indicate functional chips.

Much work has gone towards identifying and classifying the spatial defect patterns of wafer maps. Some examples include the following research: Wang (2008) proposed a hybrid method, combining entropy fuzzy c -means and a decision tree. Chen and Liu (2000) presented a neural-network architecture named adaptive resonance theory network 1 (ART1). Liu *et al.* (2002) developed a hybrid, self-organizing map-support vector machine approach by using binary wafer maps. Wang *et al.* (2006) proposed a hybrid method combining hierarchical clustering with K -means partitioning; they used an estimation module, such as a Gaussian EM algorithm, to predict each defect pattern. Jeong *et al.* (2008) proposed a methodology that integrated spatial correlograms and

dynamic time warping (DTW) distance; they used spatial correlograms to capture the spatial non-random patterns on a wafer map and adopted a DTW distance-based k-nearest neighbor (KNN) classifier to automatically classify the defect patterns based on the spatial correlogram. Chang *et al.* (2012) proposed a method that used a Hough transformation with the ratio approach to distinguish diverse defect patterns on semiconductor wafers. Liu and Chen (2013) proposed a clustering method that combined spatial statistics, a cellular neural network, an ART1 neural network, and a moment invariant to detect defect patterns.

However, all these studies examined flash memory, which generates a single binary wafer map from functional testing results, as shown in Figure 2.1. Dynamic random-access memory (DRAM), the most common computer memory is more complicated to test than flash memory because it produces non-negative integer values of multiple fail bit test (FBT) results (Liu *et al.*, 2010). Table 2.1 shows a data structure of test results from a DRAM wafer; the second column (i, j) indicates the location of each chip on a DRAM wafer (ID: 188G1W02), and the other columns show the number of failed unit cells on each chip from different FBTs.

Table 2.1 Data structure of test result from a DRAM wafer

Wafer Index	(<i>i</i> , <i>j</i>)	FBT 00	FBT 01		FBT <i>N</i> -2	FBT <i>N</i> -1
188G1W02	14,54	6828717	454942	172	8073979
188G1W02	14,55	3387648	448173	163	4650002
.....
188G1W02	15,46	2061199	273893	69	2910799
188G1W02	15,47	7	124	0	192
188G1W02	15,48	10	71	109	251
188G1W02	15,49	378	84	0	527
188G1W02	15,50	1490	210	51	2103
188G1W02	15,51	1	84	0	141

To check whether a DRAM wafer is normal or abnormal (i.e., has a spatially non-random defect pattern), multiple FBT results should be considered simultaneously. Testing a DRAM wafer produces a map from each FBT, called an FBT map, producing N FBT maps for N FBTs. Figure 2 shows four FBT maps (FBT00–FBT03) for the DRAM wafer described in Table 2.1. FBT 0 and FBT 1 found a spatially non-random cluster of defective chips in the right-center area, whereas the remaining FBTs found spatially random patterns of defective chips.

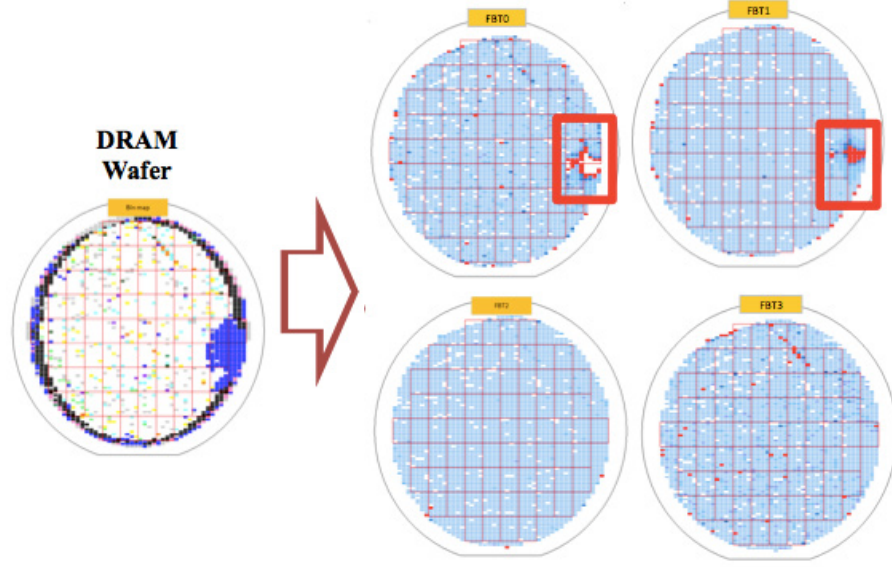


Figure 2.2 Original FBT maps for the DRAM wafer described in Table 2.1. Darker colors denote more failed cells.

As shown in Figure 2.2, considering multiple FBT maps is essential to detecting abnormal DRAM wafers. To our best knowledge, no procedures are available to detect the abnormal DRAM wafers using multiple FBT maps even though there are some studies on spatial defect analysis of DRAM wafers using a single wafer map (Chang et al., 2012; Yuan et al., 2010). In this chapter, we propose a new step-down spatial randomness test, which monitors the spatial randomness of multiple FBT maps of a DRAM wafer (Kim et al., 2015a). To transform the original FBT values into binarized ones, we use kernel-density estimation to estimate the probability distribution of the original FBT values for the functional chips. Based on these binarized FBT maps, we develop a step-down spatial randomness test with a local de-noising method to detect abnormal DRAM wafers. Figure 2.3 shows the framework of our proposed method for a DRAM wafer.

The chapter is organized as follows: Section 2.2 presents the brief background on join count statistics and spatial correlograms for multiple spatial maps. Section 2.3 proposes the step-down spatial randomness test with multiple maps based on a local denoising method. Section 2.4 presents our experimental results. Section 2.5 summarizes the contributions of this section and suggests future research.

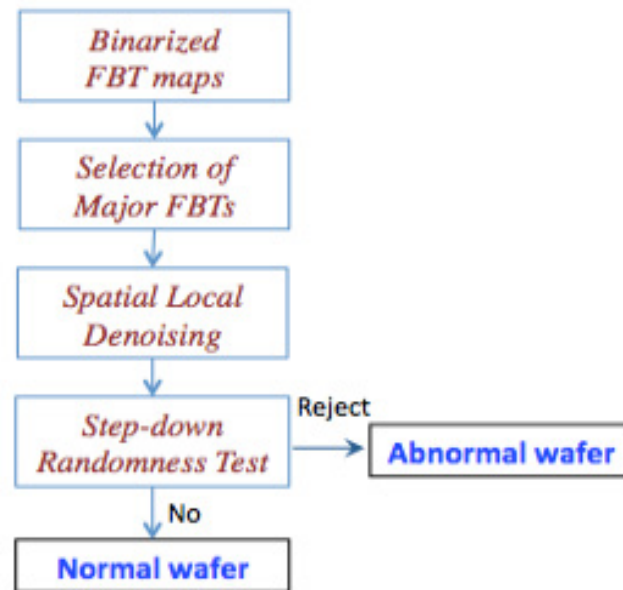


Figure 2.3 Framework of our proposed procedure for detecting an abnormal DRAM wafer.

2.2 Join count statistics and spatial correlograms for multiple spatial maps

Our proposed methodology can be implemented by extending the generalized join count (JC) statistics to represent the distribution of spatial autocorrelation for the multiple spatial maps. Suppose that, for a given DRAM wafer, we have N binarized FBT maps: $M_k, k = 1, 2, \dots, N$. For each binarized FBT map, a join forms when two chips on the binary map are near each other. Let $H^{(k)}(g)$ denote a set of the g th order neighbors, defined as chips with a distance g from each other on the k th FBT map, and let n denote the total chips on the wafer. The notation $(i, j) \in H^{(k)}(g)$ implies that chips i and j on the k th FBT map are g th-order neighbors with each other. Then, based on a neighborhood construction rule, the number of possible joins is given by $c^{(k)}(g) = \sum_{i < j} w_{ij}^{(k)}(g)$ for $i, j = 1, 2, \dots, n$, where $w_{ij}^{(k)}(g) = \begin{cases} 1, & (i, j) \in H^{(k)}(g) \\ 0, & \text{elsewhere} \end{cases}$.

In spatial statistics, there are three types of join: 0-to-0 join (between functional chips), 0-to-1 join (between functional and defective chips), and 1-to-1 join (between defective chips). Existing spatial non-randomness measures are defined based on the single lag to define the neighborhood of a chip. We extended this concept to define the neighborhood at lag (or distance) g as shown in Figure 2.4. Suppose that $c_{00}^{(k)}(g)$, $c_{01}^{(k)}(g)$, and $c_{11}^{(k)}(g)$ denote the numbers of 0-to-0, 0-to-1, and 1-to-1 joins with a spatial lag g on the k th FBT map, respectively. Then,

$$c_{00}^{(k)}(g) = \sum_{i < j} w_{ij}^{(k)}(g)(1 - x_i^{(k)})(1 - x_j^{(k)})$$

$$c_{01}^{(k)}(g) = \sum_{i < j} w_{ij}^{(k)}(g)(x_i^{(k)} - x_j^{(k)})^2$$

$$c_{11}^{(k)}(g) = \sum_{i < j} w_{ij}^{(k)}(g) x_i^{(k)} x_j^{(k)},$$

where $x_i^{(k)}$ is an indicator variable for the i th chip on M_k such that

$$x_i^{(k)} = \begin{cases} 1, & \text{defective chip} \\ 0, & \text{functional chip} \end{cases} \text{ and } w_{ij}^{(k)}(g) = \begin{cases} 1, & (i, j) \in H^{(k)}(g) \\ 0, & \text{elsewhere} \end{cases}.$$

From the definitions of $c_{00}^{(k)}(g)$, $c_{01}^{(k)}(g)$, and $c_{11}^{(k)}(g)$, then $c^{(k)}(g) = c_{00}^{(k)}(g) + c_{01}^{(k)}(g) + c_{11}^{(k)}(g)$. In practice, $(c_{00}^{(k)}, c_{01}^{(k)}, c_{11}^{(k)})$ depends on which neighborhood construction rule is applied. To represent the relationship among chips, we adopted a rook-move neighborhood (RMN) construction rule, which forms joins orthogonally, as shown in Figure 2.4. Because $H^{(k)}(g)$ is considered as a set of joins with length g , the join length corresponds exactly to the Manhattan distance, $d((x_i^{(k)}, y_i^{(k)}), (x_j^{(k)}, y_j^{(k)})) = |x_i^{(k)} - x_j^{(k)}| + |y_i^{(k)} - y_j^{(k)}|$ between the two chips involved, consistent with the RMN construction rule.

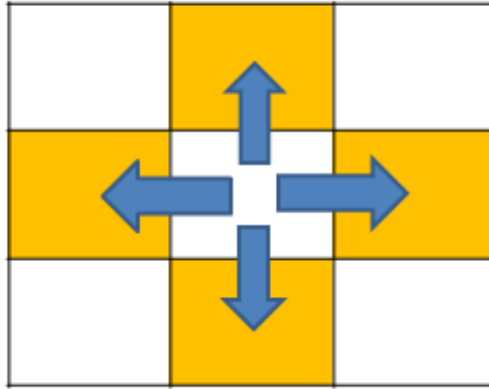


Figure 2.4 Neighbors in a rook-move neighborhood.

As a measure of spatial autocorrelation, a generalized JC-based statistic with the g th-order neighbors can be written as follows:

$$T^{(k)}(g) = p^{(k)}c_{00}^{(k)}(g) + (1 - p^{(k)})c_{11}^{(k)}(g),$$

where $p^{(k)}$ is the defect rate of the k th FBT map, which is equal to the fraction of defective chips on the k th map. $T^{(k)}(g)$ is the convex combination of 0-to-0 and 1-to-1 joins where $p^{(k)}$ provides the minimum variance among all possible combinations (Jeong et al., 2008). Higher values of $T^{(k)}(g)$ imply that many 1-1 joins (or 0-0 joins) with distance g are formed on the k th FBT map. Therefore, the higher $T^{(k)}(g)$ indicates any cluster patterns of defective chips on the k th wafer map because cluster patterns on the k th wafer maps produce large $c_{00}^{(k)}$ and $c_{11}^{(k)}$, whereas smaller $T^{(k)}(g)$ can be interpreted as indicators of no evidence of spatial dependence on the k th wafer map. By extending the work in Jeong *et al.* (2008) for a single wafer map, we can derive the following results for multiple maps:

$$E[T^{(k)}(g)] = c^{(k)}(g)p^{(k)}(1 - p^{(k)}),$$

$$Var[T^{(k)}(g)] = c^{(k)}(g)(p^{(k)})^2(1 - p^{(k)})^2.$$

Using the central limit theorem, the test statistic corresponding to the spatial lag g approximates to the standard normal distribution.

$$Z_T^{(k)}(g) = \frac{T^{(k)}(g) - c^{(k)}(g)p^{(k)}(1 - p^{(k)})}{\sqrt{c^{(k)}(g)(p^{(k)})^2(1 - p^{(k)})^2}} \sim N(0,1) \quad \text{as } c^{(k)}(g) \rightarrow \infty, \quad (2.1)$$

where g is the g th-order neighbor.

2.3 Step-down randomness test for detecting abnormal DRAM wafers

In this section, we propose a step-down randomness test that considers multiple binary maps to detect abnormal DRAM wafers. The proposed procedure has three steps: 1) binarization of original FBT values for each FBT map, 2) spatial local de-noising for each binarized FBT map, and 3) step-down spatial randomness testing, using N de-noised binarized FBT maps.

2.3.1 Binarization of original FBT values for each FBT map based on Gaussian kernel-density estimation

To apply the generalized join count statistics presented in Section 2.2 to the original FBT values, we need to first transform the non-negative integer values into binary ones. To perform this transformation, we investigated the probability distribution of FBT values of the functional chips in normal wafers. Because the distribution of FBT values is highly skewed, it is problematic to apply a parametric approach to estimate the probability density function of functional chips on normal wafers. Thus, we adopt a nonparametric kernel-density estimation approach.

Let us suppose $u_1^{(k)}, u_2^{(k)}, u_3^{(k)}, u_4^{(k)}, \dots, u_{n_f}^{(k)} \sim f^{(k)}(u)$, where $u_i^{(k)}$ is the original FBT value of the i th functional chip of the k th FBT map, and n_f is the total number of functional chips on a normal wafer. Then, the estimated probability density function is given by

$$\hat{f}_h^{(k)}(u) = \frac{1}{n_f} \sum_{i=1}^{n_f} K_{h^{(k)}}(u^{(k)} - u_i^{(k)}) = \frac{1}{n_f * h^{(k)}} \sum_{i=1}^{n_f} K\left(\frac{u^{(k)} - u_i^{(k)}}{h^{(k)}}\right),$$

where $K\left(\frac{u^{(k)} - u_i^{(k)}}{h^{(k)}}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(u^{(k)} - u_i^{(k)})^2}{2h^{(k)^2}}}$ is a Gaussian kernel. The length of the bandwidth, $h^{(k)}$, can be chosen by the following equation (Higgins, 2003):

$$h^{(k)} = \left(\frac{4\hat{\sigma}^{(k)^5}}{3n_f}\right)^{\frac{1}{5}} \approx 1.06\hat{\sigma}^{(k)}(n_f)^{-\frac{1}{5}},$$

where $\hat{\sigma}^{(k)}$ is the standard deviation of the original FBT values of functional chips on the k th FBT map on a normal wafer.

Based on the estimated kernel probability density function, a threshold value for the binarization of the k th FBT map, $\delta_{\gamma,k}$, can be chosen as its $100(1-\gamma)$ percentile. For a given threshold, the original FBT values can be binarized as follows:

$$x^{(k)} = \begin{cases} 1, & u^{(k)} > \delta_{\gamma,k} \\ 0, & u^{(k)} \leq \delta_{\gamma,k} \end{cases}.$$

In an abnormal DRAM wafer, some of the binarized FBT maps will show a spatially non-random pattern.

2.3.2 New spatial local de-noising method for each binarized FBT map

After transforming the original FBT values into binary ones, we de-noise the binarized FBT maps to clarify the spatial patterns and improve the accuracy of the spatial randomness test. To remove random noise and identify the systematic defects on a wafer

map, Wang et al. (2006) and Yuan et al. (2010) used a spatial filter. To apply the spatial filter to the binarized FBT map, we must first obtain the weighted sum of the values of the binarized chips that surround each chip. If this weighted sum is greater than a threshold, the chip is regarded as a systematic defect. However, the spatial filter uses the same weighted sum for de-noising the chips on the edge of the wafer map, which have fewer neighboring chips than those on the interior of the wafer map, so this approach may remove non-random patterns near the edges of the binarized FBT map. Thus, we propose a new spatial filter that uses different weights for de-noising the edges of the binarized FBT maps. This de-noising procedure has two cases: i) de-noising the interior of the binarized FBT map and ii) de-noising the edge of the binarized FBT map.

Case I: De-noising the interior of the binarized FBT map

To de-noise the interior of the k th binarized FBT map, we consider the filter mask with an $l \times l$ grid structure, centered on a chip with a value of 1. If the fraction of chips that have a value of 1 in the filter mask is small, the binarized FBT value of the centering chip is updated with 0. The weight of each individual die in the $l \times l$ filter mask (e.g. 3×3) is to be $1/l^2$. Then, the weighted sum for the central die at the spatial coordinate (i, j) can be obtained as follows:

$$R^{(k)}(i, j) = \frac{1}{l^2} \sum_{m=-t}^t \sum_{n=-t}^t x_{i,j}^{(k)} x_{i+m,j+n}^{(k)},$$

where $t = \frac{l-1}{2}$, and $x_{i,j}^{(k)}$ is the binarized FBT value obtained from the previous section.

If $R^{(k)}(i, j)$ is less than a pre-determined de-noising threshold, then the (i, j) th chip is considered a random defect, so we update its binarized FBT value to 0 (i.e., $x_{i,j}^{(k)} = 0$). If the weighted sum is greater than the de-noising threshold, this chip remains on the binarized FBT map (i.e., $x_{i,j}^{(k)} = 1$). Empirically, setting the de-noising threshold to the defect rate of the binarized FBT map may improve the spatial randomness test.

Case 2: De-noising the edges of the binarized FBT map

While de-noising the edges of the k th binarized FBT map, there are fewer than l^2 chips in the $l \times l$ filter mask in which the weight of each individual die is greater than $\frac{1}{l^2}$; thus, we propose a different weighting system for the edge than for the interior. If the fraction of binarized chips with a binarized FBT value of 1 among all the chips in the $l \times l$ filter mask is smaller than the pre-defined threshold, then the defective chip is removed from the edge. The weighted sum for the central dies on the edges can be defined as:

$$R^{(k)}(i, j) = \frac{1}{N(i, j)} \sum_{m=-t}^t \sum_{n=-t}^t x_{i,j}^{(k)} x_{i+m,j+n}^{(k)} A(i+m, j+n)$$

where $t = \frac{l-1}{2}$ and $N(i, j)$ is the total number of chips included in the filter mask whose centering chip is the (i, j) th one, and

$$A(i, j) = \begin{cases} 1, & \text{there is the } (i, j)\text{th chip on the FBT map} \\ 0, & \text{there is no } (i, j)\text{th chip on the FBT map} \end{cases}.$$

For example, in Figure 2.5, the 3×3 filter mask of the (7, 4)th chip includes only six chips; thus, the weight of each individual die is set to $1/6$ (i.e. $N(7,4) = 1/6$). The remaining de-noising procedures are the same as those in Case 1.

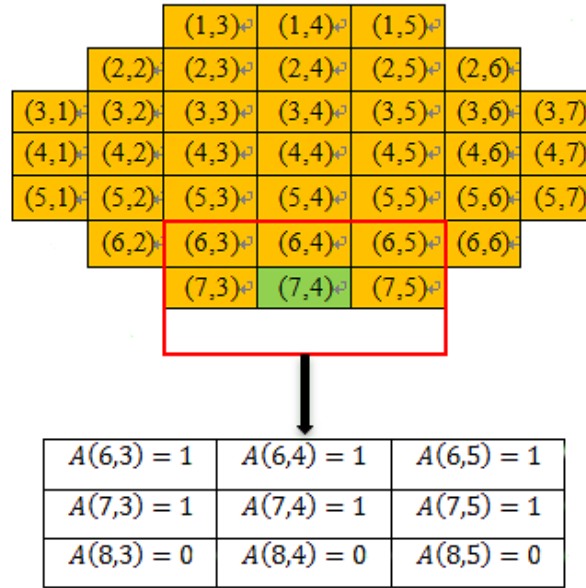


Figure 2.5 3×3 spatial filter mask for the bottom edge of a defective chip.

Figure 2.6 shows the de-noised maps of the original FBT ones shown in Figure 2.2 by applying the proposed approach where darker colors denote bad chips.

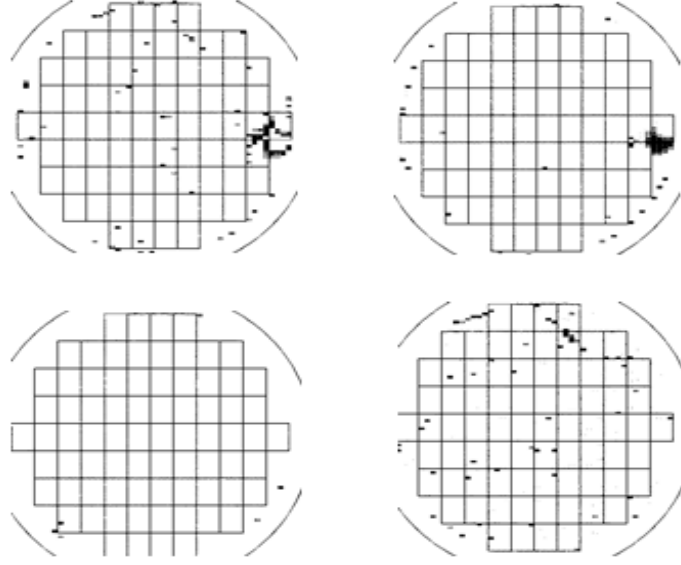


Figure 2.6 Denoised FBT maps for the FBT maps described in the Figure 2.2.

After de-noising all the binarized FBT maps of a DRAM wafer, we obtain the de-noised FBT maps for the DRAM wafer. By applying the procedure introduced in Section 2.3, we generate a spatial correlogram of the k th de-noised binarized FBT map as follows:

$$Z_T^{(k)}(g) = \begin{cases} \frac{T^{(k)}(g) - c^{(k)}(g)p^{(k)}(1 - p^{(k)})}{\sqrt{c^{(k)}(g)(p^{(k)})^2(1 - p^{(k)})^2}}, & p^{(k)} > 0 \\ 0, & p^{(k)} = 0, \end{cases}$$

where $T^{(k)}(g) = p^{(k)}c_{00}^{(k)}(g) + (1 - p^{(k)})c_{11}^{(k)}(g)$,

$c^{(k)}(g) = c_{00}^{(k)}(g) + c_{01}^{(k)}(g) + c_{11}^{(k)}(g)$, and $p^{(k)}$ is the defect rate of the k th de-noised binarized FBT map.

2.3.3 New step-down spatial randomness test using N de-noised binarized FBT maps

The goal is to test whether the given wafer is spatially random based on N spatial maps. When all of these N spatial maps show the spatially random patterns, the given wafer can be concluded to have a spatially random pattern. The challenging problems are that how to capture the spatial characteristics of data and consider the spatial correlation among these multiple maps for a given wafer. Because existing spatial randomness tests were developed for a single wafer map, such as those produced by flash memory wafers, we develop a new step-down spatial randomness test, which sequentially test the randomness of multiple binarized FBT maps of a DRAM wafer.

The main idea of the step-down randomness test is to consider the correlation among different FBT maps by using spatial correlograms, whereas the existing approaches do not consider these correlations. Suppose $\mathbf{y}_r^{(k)} = [Z_T^{(k)}(1), Z_T^{(k)}(2), \dots, Z_T^{(k)}(r)]^T$ is the first r lags of the correlogram of the k th de-noised binarized FBT map. Then, $\mathbf{y}_r^{(k)}$ is approximately normally distributed with a mean vector of length r , $\boldsymbol{\theta}_r^{(k)}$, and covariance matrix $\boldsymbol{\Sigma}_r^{(k)}$, where $k=1,2,\dots,N$. Here, N is the number of de-noised binarized FBT maps used in the step-down test. The spatial randomness test with N de-noised FBT maps can be formulated as the following sequential test hypothesis problem:

$$H_0^{(k)}: \boldsymbol{\theta}_r^{(k)} = \mathbf{0} \text{ vs. } H_1^{(k)}: \boldsymbol{\theta}_r^{(k)} \neq \mathbf{0}, \boldsymbol{\theta}_r^{(j)} = \mathbf{0}, 2 \leq j \leq k-1.$$

Let $T_k^2 = \begin{pmatrix} \mathbf{y}_r^{(1)} \\ \vdots \\ \mathbf{y}_r^{(k)} \end{pmatrix}^T \mathbf{S}_r(k)^{-1} \begin{pmatrix} \mathbf{y}_r^{(1)} \\ \vdots \\ \mathbf{y}_r^{(k)} \end{pmatrix}$, where $\mathbf{S}_r(k)$ is the sample covariance matrix of $\begin{pmatrix} \mathbf{y}_r^{(1)} \\ \vdots \\ \mathbf{y}_r^{(k)} \end{pmatrix}$ using n samples. In the first step, we test whether the de-noised wafer map of the first FBT has a non-random spatial pattern by checking whether $\boldsymbol{\theta}_r^{(1)}$ is equal to a zero vector. We can use the following test statistic:

$$Z_1 = \mathbf{y}_r^{(1)T} \mathbf{S}_r^{(1)-1} \mathbf{y}_r^{(1)}.$$

The control limit of Z_1 is given by

$$CL_1 = \frac{(n-1)r}{(n-r)} F_{\alpha_1}(r, n-r),$$

where α_1 is the false-alarm rate of the first test and $F(r, n-r)$ represents the F distribution with r and $(n-r)$ degrees of freedom. If Z_1 is greater than CL_1 in the first step, the DRAM wafer is spatially non-random (i.e., abnormal). Otherwise, we test the spatial randomness of the second FBT map. Given the first $(k-1)$ FBT maps are spatially random, the test statistic for the spatial randomness test of the k th FBT map and its corresponding control limit are described as follows (see the Appendix for a detailed derivation):

$$Z_k = \frac{T_k^2 - T_{k-1}^2}{1 + T_{k-1}^2/(n-1)}$$

and

$$CL_k = \frac{(n-1)r}{(n-kr)} F_{\alpha_k}(r, n-kr), k=2, 3, \dots, N,$$

where α_k is the false-alarm rate of the k th test. Z_k is large when $\mathbf{y}_r^{(k)}$ does not have zero mean vector and the covariance structure of $(\mathbf{y}_r^{(1)} \dots \mathbf{y}_r^{(k)})^T$ is different from $\mathbf{S}_r(k)$. In this way, we consider not only the mean vector of the spatial correlogram but also the correlation between the spatial correlograms of FBT maps.

If at least one test statistic is greater than a given control limit, the DRAM wafer is considered spatially non-random (i.e., abnormal). The overall false-alarm rate of the step-down randomness test can be set to $\alpha = 1 - \prod_{k=1}^N (1 - \alpha_k)$. In this study, we have selected some FBTs for spatial randomness test based on the prior knowledge of experts, which prevent N from being large. Thus, each step has a quite reasonable false alarm rate under the overall false alarm rate α .

2.4 Experimental results

In this section, we implement the proposed step-down randomness test to detect an abnormal DRAM wafer with multiple maps by using experimental data. Then, we compare our approach with existing methods.

2.4.1 Description of real DRAM wafers

Real DRAM wafer maps provided by a semiconductor manufacturing company were analyzed. We analyzed 45 normal wafers and 41 abnormal wafers, with 1804 chips per wafer. Table 2.1 shows the group information, wafer index, results of 28 FBT tests, bin numbers, and (i, j) spatial coordinates for each chip.

2.4.2 Step-down randomness test for DRAM wafer

2.4.2.1 Binarization of original FBT values

To transform the original FBT values into binary ones, we applied the Gaussian kernel-density estimation procedure described in Section 2.3.1. We set γ to 0.05 for each FBT, so the threshold value for binarizing each FBT is its 95th percentile. Table 2.2 summarizes the thresholds of 28 FBTs whereas transforming the original FBT values into binary ones. However, if we apply all FBTs, the false alarm rate of each step gets extremely small. As for the DRAM data, among 28 FBT maps, a few of them (which are called as the major FBTs) are important for the purpose of fault diagnosis. We selected four FBTs (00, 04, 05, and 16) based on the recommendation of an engineer because they capture spatial patterns well and prevent N from being large. Thus, each step has a quite reasonable false alarm rate under the overall false alarm rate α of 0.05.

Table 2.2 Threshold of each FBT for binarization

FBT No.	Thresholds ($\delta_{0.05,k}$)	FBT No.	Thresholds ($\delta_{0.05,k}$)
FBT00	163.11	FBT14	185.61
FBT01	79.61	FBT15	27.74
FBT02	7.25	FBT16	26.81
FBT03	9.59	FBT17	884.21
FBT04	20.06	FBT18	176.07
FBT05	4.99	FBT19	1834.66
FBT06	6.23	FBT20	248.14
FBT07	2.00	FBT21	86.82
FBT08	0.41	FBT22	4.42
FBT09	0.06	FBT23	7.02
FBT10	1.86	FBT24	63.95
FBT11	0.95	FBT25	0.11
FBT12	9.08	FBT26	1031.70
FBT13	83.46	FBT27	5.27

2.4.2.2 Spatial local de-noising of the binarized FBT map

After binarizing the original FBT values based on the estimated thresholds shown in Table 2.2, we de-noised the binarized FBT maps. Figure 2.7 shows the four de-noised FBT maps and their corresponding spatial correlograms for the abnormal and normal DRAM wafers. The abnormal DRAM wafer exhibited spatially non-random patterns on the at least one of the denoised maps (FBT04 and FBT05), and the corresponding spatial correlogram smoothly changes along the spatial lags. In contrast, the normal wafers exhibited spatially random patterns on the de-noised maps of all FBTs, and their corresponding spatial correlograms fluctuated around zero. In addition, the FBT maps of an abnormal DRAM wafer might be much more correlated than those of a normal wafer.

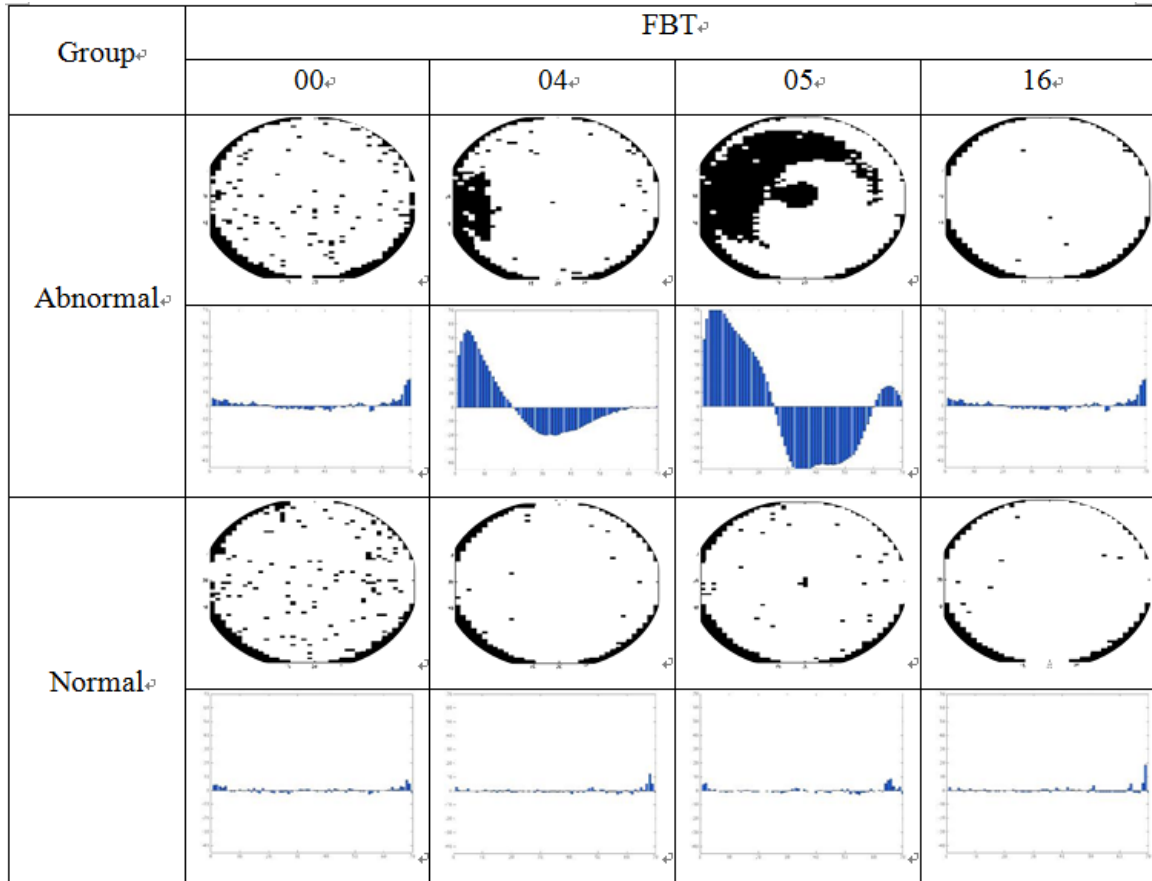


Figure 2.7. Selected de-noised FBT maps and their corresponding correlograms.

2.4.2.3. Spatial step-down randomness test using the spatial correlograms of de-noised FBT maps

Before implementing the step-down randomness test to automatically detect abnormal wafers, we must determine the optimal lag of the spatial correlograms and the testing order of the FBTs. In our empirical study, the randomness test performed well with four lags for each FBT, and the testing order of the FBTs did not significantly affect its performance. Thus, we implemented the randomness test in FBT order (i.e., 00, 04, 05, and 16).

To show the effect of the processing, we have implemented the sensitivity analysis by

using 5-fold cross validation. Table 2.3 (a) presents the effect of γ for binarization, showing that the value of 0.05 produces the best performance. The selection of the denosing threshold (τ) can be made based on the defective rate. Table 2.3 (b) shows that the threshold value based on the defective rate produces the best performance. Thus, based on the result of sensitivity analysis, the parameters are selected, respectively.

Table 2.3 Sensitivity of the parameter of each process in the proposed step-down spatial randomness test

γ	<i>Normal</i>	<i>Abnormal</i>	τ	<i>Normal</i>	<i>Abnormal</i>
0.01	0.82	0.76	0.00	0.85	0.77
0.05	0.91	0.90	p^*	0.91	0.90
0.10	0.87	0.76	0.50	0.60	0.69
			p^* : defective rate		

To study the effect of the overall false alarm rates (α), Table 2.4 shows the accuracy of the spatial randomness test for normal wafers, abnormal wafers, and all wafers. As the false alarm rate gets larger, the accuracy of the randomness test for normal (abnormal) wafers becomes smaller (larger).

Table 2.4 Sensitivity of each class for the different overall false alarm rates of the proposed step-down randomness test

α	<i>Normal</i>	<i>Abnormal</i>	<i>All</i>
0.01	0.96	0.76	0.86
0.05	0.91	0.90	0.91
0.10	0.79	0.96	0.87

2.4.3 Performance comparison of step-down randomness test

To validate each procedure of the proposed spatial randomness test, we used 5-fold cross validation method to assess the test accuracy of our proposed spatial randomness test, and compared the performance of the proposed procedure with those of other approaches: Method A used a wafer bin map (*i.e.*, single wafer map) to test the spatial randomness of the DRAM wafer. Method B binarized the original FBT values based on a normal distribution instead of using kernel-density estimation; after extracting the correlograms of the de-noised FBT map, we implemented the step-down randomness test. Method C applied Otsu’s method for the binarization (Otsu, 1979) instead of the kernel based binarization. Method D did not de-noise the binarized FBT maps for the step-down randomness test. Method E used one of the traditional noise reduction method, median filter (Lim, 1990), rather than using the proposed de-noising method. Method F used the Bonferroni method for multiple testing, which independently tests the spatial randomness of each binarized FBT map with an overall false alarm rate of 0.05.

Table 2.5 summarizes the experimental results for the comparisons in terms of

accuracies. All the parameters (thresholds) are determined at the training set and apply the pre-defined parameters to the test set. As we can see the results in Table 2.5, the test accuracy reveals our proposed step-down randomness test has the best performance overall. In particular, the proposed method of each procedure outperforms the existing ones for the following reasons. At first step, for the binarization, the traditional Otsu's method in method C has a bad performance. This is because Otsu's method has the different goal with the kernel based binarization. Otsu's method obtains the optimum threshold of binarization separating the chips into two groups so that their intra-group variance is minimal whereas the proposed method aims at selecting the outlier chips with extremely large FBT values among all the chips (Otsu, 1979). At second step, median filter in method E considers all the chips for de-noising, not removing only bad chips (Lim, 1990). Then, the median filter makes a good chip with many neighboring bad chips has the value of one (i.e. the value of a bad chip), which can make the spatially random FBT maps into non-random maps. At the last step, Bonferroni multiple testing in method F does not consider the correlation between the different FBT maps whereas Step-down test does. In this way, each of the proposed procedure is verified.

Table 2.5 Comparison of the accuracy of various randomness tests

<i>Test Method</i>	<i>Procedure of the randomness test</i>			<i>Accuracy</i>		
	Binarization	De-noising	Test	Normal	Abnormal	All
A	No	No	Single test	0.50	0.56	0.52
B	Normal	Proposed denoising	Step-down	0.84	0.79	0.81
C	Otsu	Proposed denoising	Step-down	0.17	0.88	0.51
D	KDE	No	Step-down	0.78	0.79	0.78
E	KDE	Median Filter	Step-down	0.83	0.75	0.79
F	KDE	Proposed denoising	Bonferroni	0.92	0.71	0.81
G	KDE	Proposed denoising	Step-down	0.91	0.90	0.91

2.5 Conclusions and future research

We proposed a new step-down spatial randomness test for detecting abnormal DRAM wafers based on multivariate sequence tests. The proposed approach can identify abnormal DRAM wafers with binarized FBT maps using step-down spatial randomness testing. Furthermore, we used a spatial local de-noising method to eliminate noise of defect chips, letting us better distinguish systematic defect patterns on DRAM wafers from random ones. Applying this test to experimental data, we found that it more accurately detected abnormal DRAM wafers than other methods.

The proposed approach can be extended to classify abnormal DRAM wafers with multiple binary maps based on the spatial defect patterns, which is critical for rapidly detecting the cause of defects. Further studies are needed to develop automated tools to effectively classify spatial defect patterns on DRAM wafers. Also, in order to remove the effect of the processing parameters and simplify the abnormality detection process of DRAM wafers, we may develop the spatial randomness test based on the continuous spatial data.

CHAPTER 3

A Regularized Singular Value Decomposition-based Approach for Failure Pattern Classification on Fail Bit Map in a DRAM Wafer

3.1 Introduction

In semiconductor manufacturing industry, wafer probe tests are conducted on each chip to identify good and defective ones, recognize defective patterns on a wafer, and also identify the corresponding process problems (Fenner et al., 2005; Jeong et al., 2008; Tong et al., 2008). A bit map analysis is necessary to improve the quality of manufacturing processes based on the failure patterns shown on bit maps (Hamada and Sugimoto, 1997; Vollrath et al., 2001). However, an inspection of failure patterns on massive fail bit maps data have practically depended on manual review by experts, which could lead to time-consuming and expensive cost. In addition, inspectors are not able to keep consistent inspections and concentrate for long time periods due to subjective judgments and fatigues. Thus, it is important to develop an automated procedure for failure pattern analysis on bit maps to quickly detect failure conditions.

The fail bit maps (FBMs), which represent the failed cell count during wafer functional tests, have been popularly used as one of diagnosis tools in semiconductor manufacturing for process monitoring, root cause analysis, and yield improvements (Muthumalai et al., 2014; Nakamae et al., 2001). Depending on the number of failed cells,

unit blocks of FBMs are categorized as functional or defective elements. Failing blocks are usually localized and formed specific failure patterns (Hsieh et al., 2010; Li et al., 2012; Yin et al., 2013). The failure patterns on FBMs can be classified into a single bit failure pattern and non-single bit failure pattern depending on the distribution type of failed blocks. The classification between single- and non-single bit failure patterns is a fundamental, but significant task to quickly detect failure occurrences and identify failure causes on a dynamic random access memory (DRAM) manufacturing process. For instance, a single bit failure pattern leads to a failure over cell areas while a non-single bit failure pattern can cause a failure over core or periphery areas, which play a role as a bridge to deliver signals to cell areas. The single bit failure patterns tend to have the spatially random pattern with lower grade levels while the non-single bit failure typed chips tend to have the *spatially non-random patterns* based on the higher grade levels, which are needed for analyzing the cause of failure.

Some researches on analyzing failure patterns on FBMs have been conducted in the literature. Hsieh et al. (2010) presented the method of the principal component analysis (PCA)-based approach to cluster failure events and explore IC failure mechanisms. Zanon et al. (2003) analyzed FBM data of static random access memory (SRAM) and presented a neural networks based approach to cluster failure patterns. In addition, Han et al. (2005) proposed a tree-based classifier for an automatic method of failure patterns on FBMs, but they did not present the details of experimental analysis and data set they used.

As shown in previous literatures (Collica et al., 1995; Jung, 2011; Vollrath et al., 2001), even though analysis of failure patterns on FBMs has been performed, there are only a few automatic procedures available in the literatures to distinguish between single-

bit failure patterns and non-single bit failure ones. Because failure patterns on non-single failed maps can be connected to specific process failures, accurate automatic classification between single- and non-single failure patterns can help engineers to aid the understanding of the cause of failures,

In this chapter, we propose an automated classification procedure for failure patterns (single bit failure vs. non-single bit failure) on FBMs from DRAM wafers (Kim et al., 2015b). For this, the regularized singular value decomposition (RSVD) method is developed for extracting features (*e.g.* eigen-images) from binarized FBMs to distinguish between single-bit failure patterns and non-single bit failure ones. The benefit of the proposed RSVD is to obtain binary eigen-images from a binary matrix so that the eigen-images from single bit failed maps and non-single bit failed maps can be clearly discriminated. By using the extracted features, k -nearest neighbor (k -NN) classifier distinguishes between single bit failures pattern and non-single bit failures pattern on FBMs. Figure 3.1 illustrates the flowchart of the proposed procedure for failure pattern classification. More details on each step of the proposed classification procedure are presented in Section 3.4.

This chapter is organized as follows; a brief review of FBMs is given in Section 3.2. The proposed RSVD is presented in Section 3.3, which is followed by the details of the proposed procedure for failure pattern classification on FBMs in Section 3.4. In Section 3.5, experimental studies using real-life DRAM wafers are presented. Finally, Section 3.6 concludes with a brief summary as well as suggestions of future research directions.

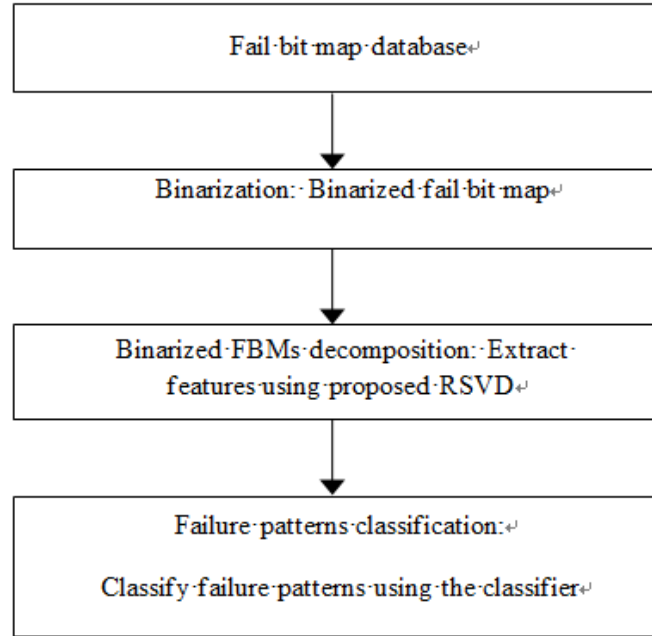


Figure 3.1 Flowchart of the proposed approach for FBM pattern classifications.

3.2 Fail bit map

Fail bit data represent the location of failing bits in chips, and form specific failure patterns, which can be analyzed to find certain process failures. Figure 3.2 shows an example of a FBM, and the corresponding grade level. As a grade level is smaller, the number of failed cell is smaller. See Section 3.4 for the detailed description of each grade level in FBMs.

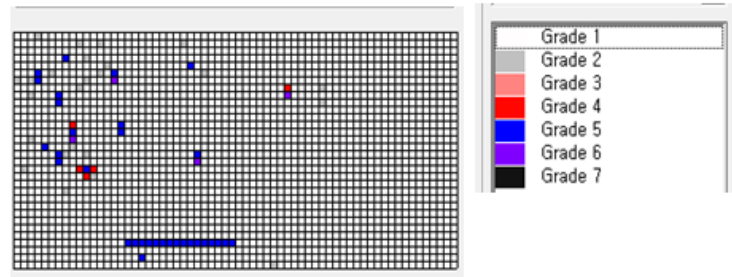


Figure 3.2 An example of a fail bit map.

In FBMs, failure patterns can be categorized as 1) Single bit failure pattern and 2) Non-single bit failure pattern. A single bit failure map does not have a specific failure pattern while the non-single bit failure map represents specific patterns such as a vertical or horizontal line pattern caused by high grades (greater than or equal to Grade 3) as shown in Figure 3. In Figure 3(a), long vertical failure line and short horizontal failure line are shown on right side and upper- middle side, respectively. On the other hand, there is no specific pattern on single bit failure map as shown in Figure 3(b). As the memory size gets larger, manual inspection of massive FBMs is not practical. In addition, because quick detection of failure patterns on FBMs can be connected to yield improvement, it is significant to develop the automatic classification method of failure patterns on FBMs. Therefore, regularized singular value decomposition (RSVD) approach is developed to extract important features that characterize spatial patterns of each defect type.

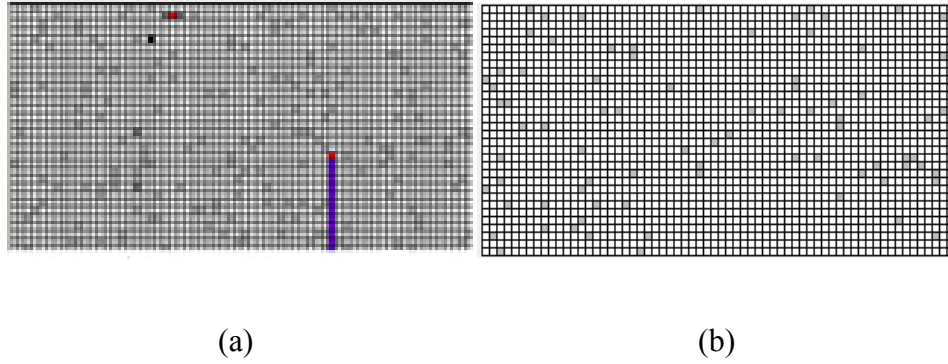


Figure 3.3 Example of failure patterns on FBMs; (a) Non-single bit failure map. (b) Single bit failure map.

3.3 Regularized singular value decomposition

The singular value decomposition (SVD) is an attractive transform approach for image processing applications. There have been many researches that apply SVD transformation to digital image processing (Amirshahi and Torkamaniazar, 2012; Andrews and Patterson, 1976; Sadek, 2012).

Let X be the m -by- n matrix ($m \geq n$) in which $X_{i,j}$ represents the grade of the (i, j) -th unit block. The SVD of X is defined as (Sadek, 2012; Hastie et al., 2009)

$$X = U\Sigma V^T = [u_1 u_2 \dots u_m] \begin{bmatrix} s_{11} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & s_{mm} & 0 & \dots & 0 \end{bmatrix} [v_1 v_2 \dots v_n]^T \quad (3.1)$$

where U is the m -by- m matrix of the eigenvectors of XX^T , V is the n -by- n matrix of the eigenvectors of $X^T X$, and Σ is m -by- n matrix in which s_{ii} is the i -th singular value of X for $i=1,2,\dots,m$. If the matrix Σ is written in the following form

$$\mathbf{\Sigma} = \begin{bmatrix} s_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_{mm} \end{bmatrix} = \begin{bmatrix} s_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} + \dots + \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_{mm} \end{bmatrix},$$

then the Eq. (3.1) can be re-written by

$$\mathbf{X} = \sum_{i=1}^m s_{ii} \mathbf{u}_i \mathbf{v}_i^T \quad (3.2)$$

where $s_{11} \geq s_{22} \geq \dots \geq s_{mm}$. The singular value s_{ii} is equivalent to the matrix norm of the corresponding eigen-image (Andrews and Patterson, 1976).

However, when the original matrix \mathbf{X} is binary, the eigen-images could contain negative values. In this case, the interpretation of the images is difficult, and important characteristics of data may not be captured well (Liu et al., 2008; Miettinen et al., 2008; Zhang et. al, 2007).

In order to overcome the drawback of conventional SVD, regularized SVD (RSVD) adds the constraints on binary eigen-images. The proposed RSVD restricts $s_{kk} \mathbf{u}_k \mathbf{v}_k^T$ to be binary by adding the constraint $(s_{kk} \mathbf{u}_k \mathbf{v}_k^T)_{ij}^2 - (s_{kk} \mathbf{u}_k \mathbf{v}_k^T)_{ij} = (s_{kk} u_{ik} v_{jk})^2 - (s_{kk} u_{ik} v_{jk}) = 0$, which generates binary eigen-images. In this way, RSVD has some advantages over conventional SVD when the data are binary. Given a binary matrix \mathbf{X} , RSVD is to find binary matrix $s_{ii} \mathbf{u}_i \mathbf{v}_i^T$ for which \mathbf{u}_i , \mathbf{v}_i , and s_{ii} are nonnegative. Then, RSVD solves the following mathematical programming.

$$\min \frac{1}{2} \|\mathbf{X} - \mathbf{U} \mathbf{S} \mathbf{V}^T\|^2$$

$$\text{s.t. } \sum_{k=1}^m \sum_{i=1}^m \sum_{j=1}^n \left((s_{kk} u_{ik} v_{jk})^2 - (s_{kk} u_{ik} v_{jk}) \right)^2 = 0.$$

Alternatively, the above problem can be represented as follows;

$$\min \frac{1}{2} \|\mathbf{X} - \mathbf{USV}^T\|^2 + \frac{\lambda}{2} \sum_{k=1}^m \sum_{i=1}^m \sum_{j=1}^n \left((s_{kk} u_{ik} v_{jk})^2 - (s_{kk} u_{ik} v_{jk}) \right)^2,$$

where u_{ik} is the (i, k) -th element of \mathbf{U} and v_{jk} is the (j, k) -th element of \mathbf{V} , and s_{kk} is the k -th diagonal element of \mathbf{S} . In addition, λ is the penalty coefficient, which is increased at each iteration to converge the solution (here, increased by a factor of $\sqrt{10}$ at each iteration). We chose the value of penalty coefficient λ based on the preliminary experiments and literatures (Zhang et al., 2007). The following multiplicative updating procedure is developed to solve the above optimization problem (see Appendix B for the detailed derivation).

Multiplicative update algorithm for RSVD

$$u_{ik}^{(t+1)} \leftarrow u_{ik}^{(t)} \frac{(\mathbf{XV}^{(t)}(\mathbf{S}^{(t)})^T)_{ik} + \lambda \sum_{j=1}^n 3(s_{kk}^{(t)})^3 (u_{ik}^{(t)})^2 (v_{jk}^{(t)})^3}{(\mathbf{U}^{(t)}\mathbf{S}^{(t)}(\mathbf{V}^{(t)})^T \mathbf{V}^{(t)}(\mathbf{S}^{(t)})^T)_{ik} + \lambda \sum_{j=1}^n (2(s_{kk}^{(t)})^4 (u_{ik}^{(t)})^3 (v_{jk}^{(t)})^4 + (s_{kk}^{(t)})^2 (u_{ik}^{(t)}) (v_{jk}^{(t)})^2)} \quad (3.3)$$

$$s_{kl}^{(t+1)} \leftarrow s_{kl}^{(t)} \frac{((\mathbf{U}^{(t+1)})^T \mathbf{XV}^{(t)})_{kl} + 1(k=l)\lambda \sum_{i=1}^m \sum_{j=1}^n 3(s_{kk}^{(t)})^2 (u_{ik}^{(t+1)})^3 (v_{jk}^{(t)})^3}{((\mathbf{U}^{(t+1)})^T \mathbf{U}^{(t+1)}\mathbf{S}^{(t)}(\mathbf{V}^{(t)})^T \mathbf{V}^{(t)})_{kl} + 1(k=l)\lambda \sum_{i=1}^m \sum_{j=1}^n (2(s_{kk}^{(t)})^3 (u_{ik}^{(t+1)})^4 (v_{jk}^{(t)})^4 + s_{kk}^{(t)} (u_{ik}^{(t+1)})^2 (v_{jk}^{(t)})^2)} \quad (3.4)$$

$$v_{jl}^{(t+1)} \leftarrow v_{jl}^{(t)} \frac{(\mathbf{X}^T \mathbf{U}^{(t+1)}\mathbf{S}^{(t+1)})_{jl} + 1(k=l)\lambda \sum_{i=1}^m 3(s_{kk}^{(t+1)})^3 (u_{ik}^{(t+1)})^3 (v_{jk}^{(t)})^2}{(\mathbf{V}^{(t)}\mathbf{S}^{(t+1)}(\mathbf{U}^{(t+1)})^T \mathbf{U}^{(t+1)}\mathbf{S}^{(t+1)})_{jl} + 1(k=l)\lambda \sum_{i=1}^m (2(s_{kk}^{(t+1)})^4 (u_{ik}^{(t+1)})^4 (v_{jk}^{(t)})^3 + (s_{kk}^{(t+1)})^2 (u_{ik}^{(t+1)})^2 v_{jk}^{(t)})} \quad (3.5)$$

After solving the optimization problem, we obtain the k -th column vector of new \mathbf{U} and \mathbf{V} (i.e. \mathbf{u}_k and \mathbf{v}_k) and the k -th diagonal element of new \mathbf{S} (i.e. s_{kk}). We then have the new k -th eigen-image $s_{kk}\mathbf{u}_k\mathbf{v}_k^T$. The sum of the eigen-images approximates the original binary matrix \mathbf{X} .

$$\mathbf{X} \approx \sum_{k=1}^m s_{kk}\mathbf{u}_k\mathbf{v}_k^T \quad (3.6)$$

The proposed RSVD algorithm is summarized as follows.

RSVD Algorithm:

Input: \mathbf{X} : Binary matrix with m -by- n size

Step 1: Initialize non-negative matrices \mathbf{U} , \mathbf{S} , and \mathbf{V} by taking absolute values of the resulting SVD matrices.

Step 2: For \mathbf{U} , \mathbf{S} , and \mathbf{V} , use the multiplicative updating procedure to solve the following optimization problem:

$$\min \frac{1}{2} \|\mathbf{X} - \mathbf{USV}^T\|^2 + \frac{\lambda}{2} \sum_{k=1}^m \sum_{i=1}^m \sum_{j=1}^n \left((s_{kk}u_{ik}v_{jk})^2 - (s_{kk}u_{ik}v_{jk}) \right)^2$$

Step 3: If $\sum_{k=1}^m \sum_{i=1}^m \sum_{j=1}^n \left((s_{kk}u_{ik}v_{jk})^2 - (s_{kk}u_{ik}v_{jk}) \right)^2 < \epsilon$, break

else set $\lambda = \sqrt{10}\lambda$, and return to Step 2.

In Step 1, the *non-negative* matrices U , S , and V are initialized *by taking absolute values of the resulting matrices* from *SVD* of the binary matrix X . In Step 2, we compute the quadratic programming problem using Lagrangian multipliers. Next, we repeatedly update U , S , and V until the convergence criterion is satisfied. Note that RSVD algorithm is converged because the non-negative optimization problems are non-increasing under the multiplicative update algorithms (Lee and Seung, 2001; Yang and Laaksonen, 2007). By reversing the roles of U , S , and V , the algorithm can obtain the updated matrices using Eqs. (3) - (5).

Remarks: In RSVD algorithm, the matrix U and V are not necessarily orthogonal. This can restrict to use RSVD for getting a pseudo-inverse matrix or solving a system of linear equations in which an orthogonal property is used. To impose an orthogonal property to RSVD, the formulation can be modified as follows;

$$\begin{aligned} & \min \frac{1}{2} \|X - USV^T\|^2 \\ & \text{s.t. } \sum_{k=1}^m \sum_{i=1}^m \sum_{j=1}^n \left((s_{kk} u_{ik} v_{jk})^2 - (s_{kk} u_{ik} v_{jk}) \right)^2 = 0 \\ & \mathbf{U}^T \mathbf{U} = \mathbf{I}, \mathbf{V}^T \mathbf{V} = \mathbf{I}. \end{aligned}$$

Exploring the orthogonal property by dealing with the modified formulation is out of scope of this study and is one of good future research directions.

3.4 Classification of failure patterns on FBMs using RSVD

This section will present the classification procedure of failure patterns on FBMs by using our proposed RSVD method. The failure pattern classification procedure is composed of 1) binarization of FBMs; 2) feature extraction from FBMs using the RSVD; 3) classification of failure patterns on FBMs using the k -NN classifier.

The FBMs characterize the defect type of the DRAM chips. As shown in Figure 3.2, each chip on FBMs are categorized based on grade levels, which represent the number of failed cells on each chip. Grade 0 indicates that there is no failed cell on chips whereas Grade 7 implies a large number of failed cells on chips. A detailed description of each grade is not presented due to the confidentiality issue of company data. A single bit failed FBM does not have specific failure patterns of the unit blocks whereas FBM with non-single bit failed chips has any specific failure patterns. Because non-single bit failure patterns on FBMs could be connected to specific process faults, semiconductor industry has focused on automated classification procedure of failure patterns on FBMs.

3.4.1 Binarization of FBMs

Suppose that there is the FBM of a failed chip, \mathbf{M} , that is represented by $m \times n$ matrix whose element is the grade of unit block on the FBM. We can binarize the real value matrix \mathbf{M} based on the grade level of each chip. That is,

$$\mathbf{B}(i, j) = \begin{cases} 1, & \text{if } \mathbf{M}(i, j) \geq 3 \\ 0, & \text{if } \mathbf{M}(i, j) < 3 \end{cases} \text{ for } i = 1, 2, \dots, m, \text{ and } j = 1, 2, \dots, n. \quad (3.7)$$

Here, based on engineering knowledge, if the grade level is less than 3, the corresponding chips are set to 0, otherwise assigned to 1.

Figure 3.4 shows an example of the binarized FBMs of the single-bit and non-single bit failed maps. In Figure 3.4 (b), the black unit box represents the cell with greater than or equal to grade level 3 whereas the white one represents the cell with less than grade level 3. The FBMs with non-single bit failed chips tend to have the block or spatial failure patterns based on the black unit box, whereas single-bit failed chips tend to have spatial random patterns on the FBMs (Jung, 2011).

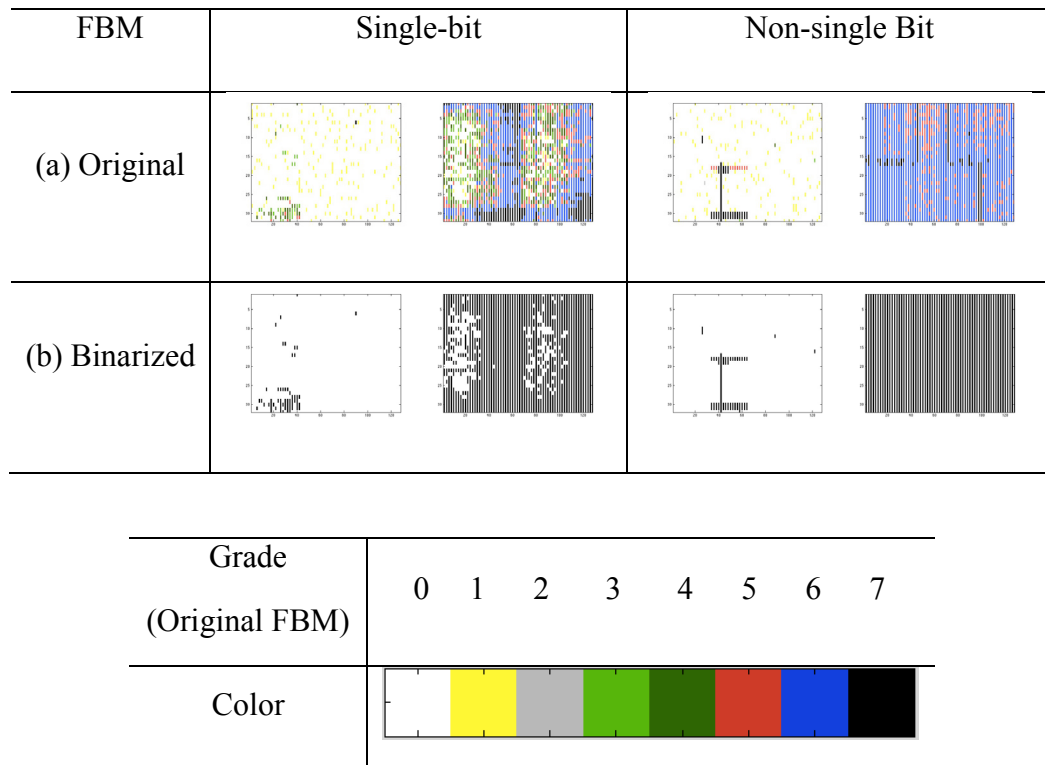


Figure 3.4 Binarized FBMs of the single-bit and non-single bit failed chips.

3.4.2 Feature extraction from binarized FBMs using RSVD

To classify failure patterns on FBMs, this subsection describes the procedure of feature extraction from FBMs that can provide the characteristics of the failure types at the chip level. With discrete representation of FBMs, the proposed RSVD decomposes the binarized matrix \mathbf{B} into m eigen-images such that $E_k = s_{kk} \mathbf{u}_k \mathbf{v}_k^T$, for $k = 1, 2, \dots, m$

Since each eigen-image displays *separate spatial patterns* of the failed chip on the FBM, the matrix norm of the eigen-image is equivalent to the contribution that the failure patterns of each eigen-image make to the binarized FBM. For calculating matrix norm, we employ the Frobenius norm. The Frobenius norm of an eigen-image E_k is defined as follows.

$$\|E_k\| = \left(\sum_i \sum_j e_{ij(k)}^2 \right)^{\frac{1}{2}} = s_{kk} \sqrt{\text{tr}(\mathbf{u}_k \mathbf{v}_k^T \mathbf{v}_k \mathbf{u}_k^T)} \quad (3.8)$$

where $e_{ij(k)}$ is the (i, j) -th element of the eigen-image E_k (i.e. $s_{kk} \mathbf{u}_k \mathbf{v}_k^T$). $\|E_k\|$ captures the contribution of spatial failure patterns that are shown on the eigen-image E_k for FBM of the failed chips. Note that $\|E_k\| \geq \|E_{k+1}\|$ for all $k (\leq m)$.

Figure 3.5 shows the eigen-images (E_k) of the binarized FBMs generated by RSVD with the matrix norm $\|E_k\|$ for the different defect types of the failed chips. In Figure 3.5 (a), none of the binarized FBM has the first eigen-image with the zero norm. In Figure 3.5 (b), for the non-single-bit failed chip of type II, the binarized FBM has the fifth eigen-image with zero norm. In Figure 3.5 (c), for the non-single-bit failed chip of type I, the binarized FBM has the tenth eigen-image with zero norm. For the single failed chips, the binarized FBMs do not have the eigen-images with zero norm. In other words, the

different types of the failed chips have the zero norm for the different eigen-images of the binarized FBM.

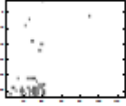
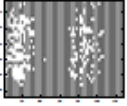
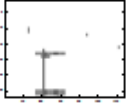
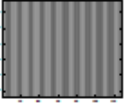
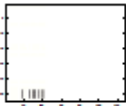
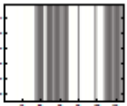

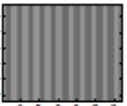
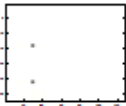
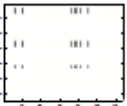
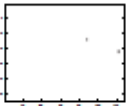
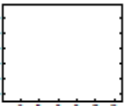
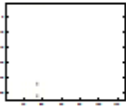
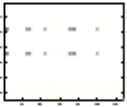
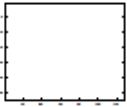
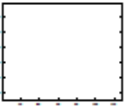
FBM ϕ	Single bit		Non-single bit	
	Type I	Type II	Type I	Type II
Binarized				
(a) E_1	 (4.23)	 (30.98)	 (6.90)	 (45.25)
(b) E_5	 (2.00)	 (5.91) ϕ	 (1.00)	 (0.00)
(c) E_{10}	 (1.41)	 (4.69)	 (0.00)	 (0.00)

Figure 3.5 Eigen-images of the binarized FBMs for the defect types of failed chips.

3.4.3 Classification of the failure patterns on FBMs using k-NN classifier

As shown in Eq. (3.8), $\|E_k\|$ is equivalent to the matrix norm of the k -th eigen image. In other words, $\|E_k\|$ of the binarized FBM represents the contribution of spatial failure patterns shown on the eigen-image k . Thus, those matrix norms of FBMs can be used as key features to discriminate the failure patterns between single bit and non-single bit maps.

In this study, we use k-nearest neighbor (k-NN) classifier because the k-NN classifier has been widely used for classification problem (Theodoridis and Koutroumbas, 2009). Suppose there are N training FBMs including single bit and non-single bit maps. The matrix norms $\|E_k\|$'s of each training FBM are extracted by using the proposed RSVD. For the testing FBM, \mathbf{x} , k training FBMs nearest to \mathbf{x} in Euclidean distance are selected, with k being an odd number. Out of the selected k FBMs, the testing FBM is assigned to one of the failure types (*i.e.* single bit failure pattern or non-single bit failure pattern) that is most represented in the k selected FBMs.

3.5 Experimental results

Real-life fail bit maps, provided by a partner semiconductor manufacturing industry, are analyzed to evaluate the performance of the proposed approach. The total number of the FBMs used in this experiment is 200 including 100 single-bit failed patterns and 100 non-single bit failed ones.

Because our main points of interests are features extracted from the data set for the classification, we compare the proposed approach with the SVD-based approach in terms of the classification accuracy. By using 5-fold cross validation, we randomly partitioned the 200 FBMs into 5 sub-datasets and apply k-NN classifier to evaluate the classification accuracy of each method. For each method, we use the different number of the features (*i.e.* matrix norm of an eigen-image) to classify the failed DRAM chips. Table 3.1 summarizes the classification accuracy depending on the number of features extracted from the SVD and RSVD, respectively, for each testing dataset. As shown in Table 3.1,

our proposed procedure shows a better performance than SVD-based approaches regardless of how many features we use for classifying failed DRAM chips. In addition, it is recommended to use more than 10 features of the RSVD based approach for classifying the DRAM chips. Regarding to the number of k in a NN classifier, in addition, Table 3.1 shows that the classification accuracy of 1-NN classifier is overall slightly better than that of 3-NN and 5-NN classifier, but the accuracy of 1-NN classifier is not always the best one.

Table 3.1 Classification accuracy of failed chips using different methods of features extraction. Classification accuracy of (a) 1-NN, (b) 3-NN, and (c) 5-NN

(a)

Testing set	(3) [*]		(7)		(12)		(16)	
	SVD	RSVD	SVD	RSVD	SVD	RSVD	SVD	RSVD
{1}	0.78	0.88	0.88	0.95	0.90	0.98	0.90	0.98
{2}	0.85	0.85	0.90	0.98	0.93	1.00	0.93	1.00
{3}	0.85	0.85	0.93	1.00	0.93	0.98	0.93	0.98
{4}	0.90	0.93	0.90	0.93	0.95	0.95	0.95	0.95
{5}	0.75	0.95	0.88	0.95	0.90	0.95	0.90	0.93
Average	0.83	0.89	0.90	0.96	0.92	0.97	0.92	0.97

(^{*}) : Number of the eigen-images used for classification

(b)

Testing set	(3) [*]		(7)		(12)		(16)	
	SVD	RSVD	SVD	RSVD	SVD	RSVD	SVD	RSVD
{1}	0.83	0.85	0.90	0.98	0.90	0.98	0.90	0.98
{2}	0.83	0.85	0.95	0.93	1.00	0.98	1.00	0.95
{3}	0.88	0.88	0.98	1.00	0.95	0.98	0.95	0.98
{4}	0.85	0.88	0.98	0.95	1.00	0.95	1.00	0.95
{5}	0.80	0.88	0.90	0.88	0.90	0.93	0.90	0.93
Average	0.84	0.87	0.94	0.95	0.95	0.96	0.95	0.96

(^{*}) : Number of the eigen-images used for classification

(c)

Testing set	(3) [*]		(7)		(12)		(16)	
	SVD	RSVD	SVD	RSVD	SVD	RSVD	SVD	RSVD
{1}	0.83	0.80	0.93	0.98	0.93	0.98	0.93	0.98
{2}	0.90	0.88	0.90	0.93	0.95	0.95	0.95	0.95
{3}	0.83	0.88	0.95	0.95	0.95	0.95	0.95	0.95
{4}	0.83	0.85	0.98	0.95	0.98	0.95	0.98	0.95
{5}	0.80	0.88	0.90	0.88	0.93	0.93	0.93	0.93
Average	0.84	0.86	0.93	0.94	0.95	0.95	0.95	0.95

(^{*}) : Number of the eigen-images used for classification

To explore the comparison between the proposed method and a popular existing classification technique in term of the classification accuracy, in this experiment, binary and RSVD-based supervised multilayer perceptron neural network are compared with the proposed method. TABLE 3.2 presents the classification performance of the 1-NN classifier with the RSVD (16 eigen-images), binary neural networks (Binary NNets), and RSVD-based neural networks (RSVD NNets). Binary NNets has an input vector of 4096 ($=32 \times 128$) binary elements while RSVD-based NNets has that of the matrix norms of the first 16th eigen-images. The NNets include one hidden layer with ten neurons and one output layer neuron. Log sigmoid function and linear transfer function are used for activation functions of the hidden and output layer. TABLE 3.2 shows the classification accuracy of each approach, and indicates that the accuracy of 1-NN classifier with RSVD is much better than that of binary NNets and competitive with the NNets with eigen-images.

Table 3.2 Classification accuracy between neural networks and 1-NN classifier with
RSVD

Testing set	Binary NNets	RSVD NNets	RSVD 1-NN
{1}	0.58	0.93	0.98
{2}	0.73	0.88	1.00
{3}	0.68	0.93	0.98
{4}	0.58	0.90	0.95
{5}	0.55	0.90	0.93
Average	0.62	0.91	0.97

Figure 3.6 presents the rationale why the features extracted from the RSVD provide the better discrimination between the single bit failed maps and non-single failed ones compared to the features extracted from the SVD. Since the matrix norms are sorted in descending order, the first matrix norm implies the contribution of the first eigen-image to the Figure 3.6 (b) binarized FBM. As shown in Figure 3.6 (c), the RSVD based matrix norms of the eigen-images tend to have the smaller decreasing rate for the binarized FBM of single-bit failed map. On the other hand, by using the conventional SVD, as shown in Figure 3.6 (d), it is unclear to discriminate failure patterns between a single-bit and non-single bit failure pattern based matrix norms of the eigen-images.

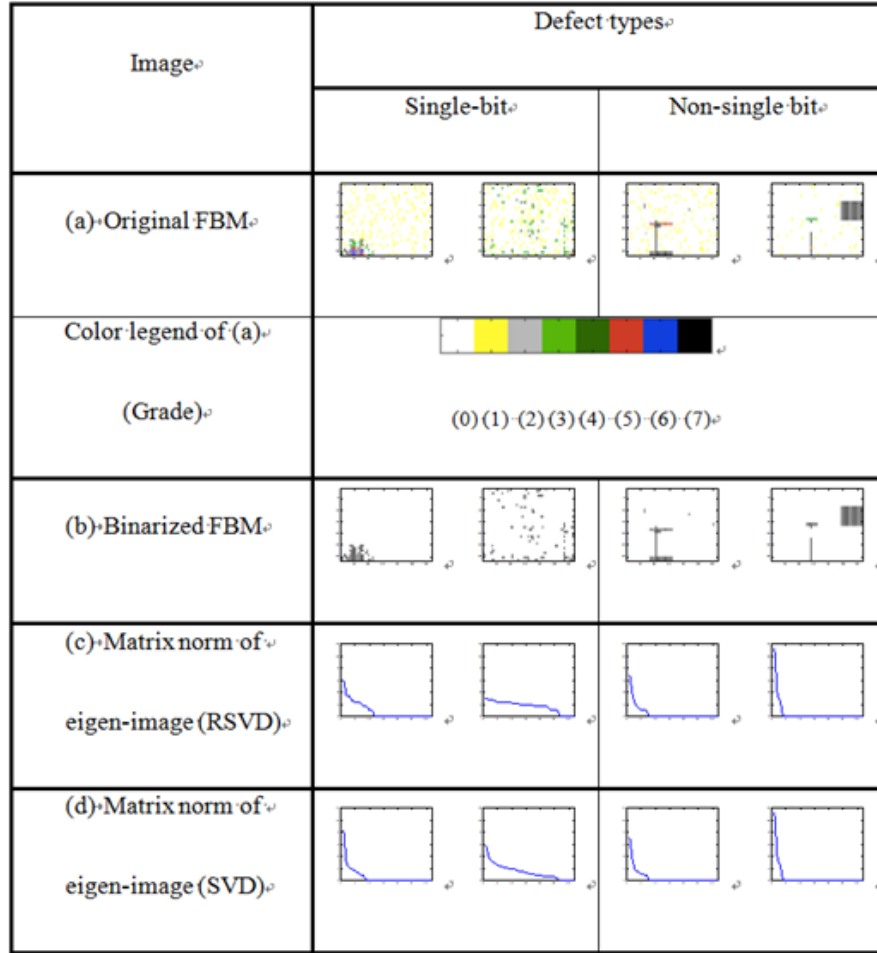


Figure 3.6 Matrix norms of the eigen-images based on SVD and RSVD.

In addition, Figure 3.7 presents the first eigen-images based on SVD and RSVD for a non-single FBMs. This figure shows that the first eigen-image of RSVD can clearly capture an unique pattern on non-single FBMs, while the first eigen-image of SVD has some unit blocks with the value between zero and one because SVD may not produce binary eigen-images. This can make the interpretation of an original image be difficult.

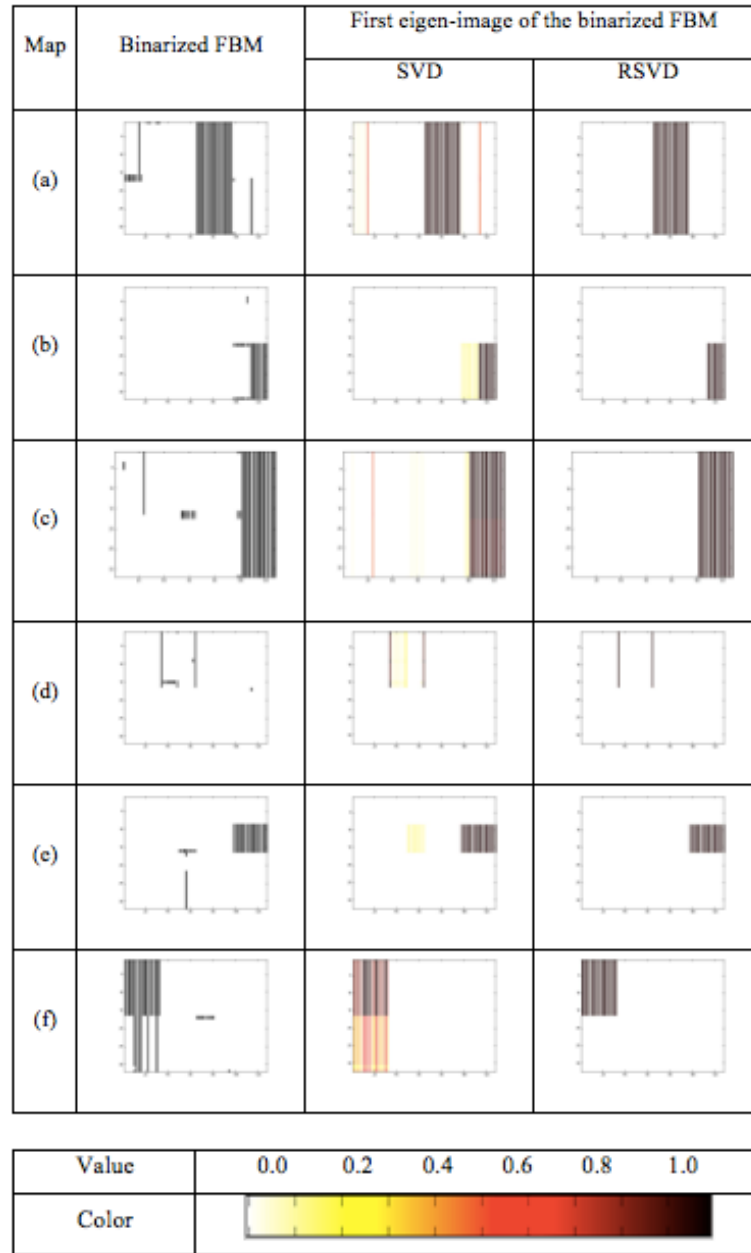


Figure 3.7 First eigen-images of the binarized FBMs based on SVD and RSVD.

3.6 Conclusion

In this chapter, the automated classification procedure for failure patterns on fail bit maps in DRAM wafers is proposed. Key features for the classification are extracted based on the novel regularized singular value decomposition (RSVD), which decomposes a binary matrix X into binary eigen-images satisfying $X \approx USV^T$. In addition, multiplicative updating procedure of RSVD has been derived to solve the formulation of RSVD. Experimental results with real-life DRAM wafers show the competitiveness of the proposed procedure for the automatic classification of single bit failed maps and non-single bit failed ones. The automated classification analysis not only helps an engineer to monitor manufacturing processes automatically, but also provides a valuable tool for yield improvement.

Future work is needed to explore other advanced classifiers such as recent variants of support vector machine (SVM) and Bayesian relevance vector machine to improve classification accuracy, and to extend the proposed approach to other applications such as bioinformatics.

CHAPTER 4

A New Multivariate Kernel Density Estimation of Uncertain Data for Defect Classification of Wafer Maps

4.1 Introduction

In the semiconductor manufacturing process, each integrated circuit (IC) chip is fabricated by a complex and costly process that involves hundreds of steps (Fenner et al., 2005). The fabricated ICs are then classified as either a functional or defective one. Defective chips commonly occur in clusters or display some systematic patterns on the wafer, which holds important information that can assist process engineers in their understanding of the ongoing manufacturing processes (Jeong et al., 2008).

A wafer map is a useful image data that displays the spatial pattern of the defective chips on a wafer. The human operators have reviewed a wafer map to detect the abnormal defect patterns from out-of-control process and to identify the root causes (Segal *et al.*, 2001). Engineers then adjust the manufacturing process to remove the defects based on the results of the inspection and their knowledge. However, this manual process is extensively time-consuming to the point where the need for an automated defect classification system inevitably arises. Subsequently, there have been many studies on the automatic retrieval of spatial features of the defects displayed on a wafer map (Chou *et al.*, 1997; Kameyama and Kosugi, 1999; Wang, 2008).

In semiconductor wafer inspections, a complex spatial pattern of a wafer map is characterized by several spatial features such as shape, location, and size of defects (Fukushima et al, 1999). For example, the shapes of the spatial defect patterns are typically represented by categorical features such as random, cluster, and circle (Jeong et al, 2008). Jung (2011) proposed categorical features, such as edge, middle, and center, to indicate the defect location of a wafer map. However, in practice, it is very rare that a limited number of categorical features perfectly capture the spatial defect patterns on a wafer.

Figure 4.1 considers the example of the defect classification of a dynamic random-access memory (DRAM) wafer. For a given DRAM wafer, there are multiple spatial maps, called fail bit test (FBT) maps (Liu *et al.*, 2010). As shown in Figure 4.1, the FBT maps include the distinct defect patterns; therefore, it is uncertain which pattern is the most critical among the multiple patterns. In addition, the defect pattern of the 5th FBT map in Figure 4.1 may not be exactly described by some categorical features due to its complexity. The incomplete spatial features generate the uncertainty in representing the spatial pattern, which may deteriorate the performance of the defect classification. Therefore, it is unquestionable that handling the uncertainty of the spatial features should be adeptly conducted in order to improve the accuracy of the defect classification.

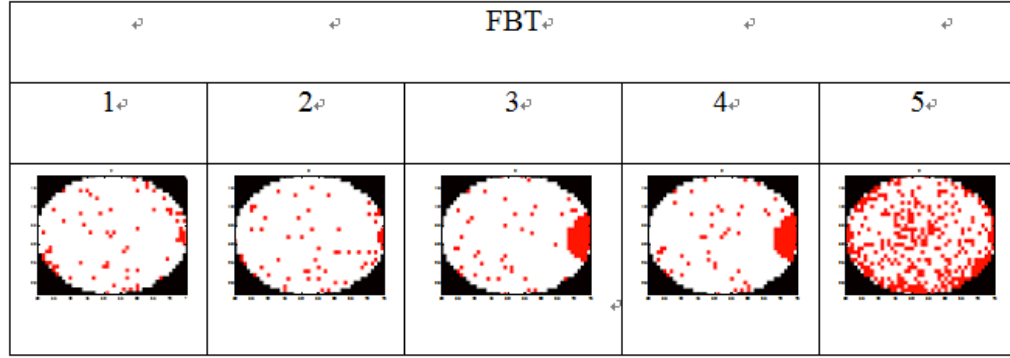


Figure 4.1 Multiple FBT maps Figure of a given DRAM wafer.

In uncertain data, it is assumed that a single feature is ideally captured not by a point value (i.e. a certain value) but by a range of values with a probability distribution (Tsang et al., 2011). Table 4.1 shows an example of uncertain categorical data extracted from a DRAM wafer. Following engineers' troubleshooting knowledge, a single FBT map that includes the most significant defect pattern among all the FBT ones is firstly indicated by a categorical feature. Several uncertain features that describe the spatial defect pattern on the selected FBT map are extracted. As shown in Table 4.1, the spatial feature vector of the i -th DRAM wafer is represented by the joint probability mass function $P(FBT_i = m, Shape_i = s, Location_i = l)$. In practice, the uncertain categorical data is common in the defect classification of DRAM wafers. Under this data structure, the traditional classifier gradually loses its momentum when the uncertainty of features increases. To tackle this problem, several traditional classifiers have been extended for uncertain data (Bi and Zhang, 2005; Qin et al., 2009; Angiulli and Fassetti, 2013).

Table 4.1 Uncertain categorical data extracted from DRAM wafers

Wafer	Joint probability mass function
1	$P(FBT_1 = m, Shape_1 = s, Location_1 = l)$
...	...
i	$P(FBT_i = m, Shape_i = s, Location_i = l)$
...	...
N	$P(FBT_N = m, Shape_N = s, Location_N = l)$

As shown in Table 4.1, it is required to deal with the uncertain categorical or mixed data for the defect classification of DRAM wafers. Among all the existing approaches, some of them are able to classify uncertain categorical or mixed data. Tsang et al. (2011) proposed a classification algorithm that construct decision tree out of uncertain data. Qin et al., (2010) developed a naïve Bayes classifier for uncertain continuous and mixed data. However, those methods have a drawback in that they assume the uncertain features are mutually independent, which is rare in real-life applications. As shown in Table 4.1, the critical FBT map index (i.e. FBT_i) and the shape and location related features (i.e. $Shape_i, Location_i$) are not mutually independent. The violation of this condition may affect the performance of the defect classification.

This study aims at classifying uncertain data that are extracted from DRAM wafers. We propose a new Bayesian classifier for uncertain multivariate data. For this, we

develop a new multivariate kernel function to estimate a class conditional probability density (or mass) function (pdf) of uncertain multivariate continuous, categorical, and mixed data, respectively. Based on the class conditional density estimates and class prior probability, the posterior class probability of a test data with unknown class label is calculated by using Bayes theorem. After a classification model for uncertain data is constructed, the test data is assigned to the class with the maximum posterior class probability. We also discuss rationale why the proposed classifier outperforms existing ones by illustrating how the features used each classifier capture the spatial defect patterns of real-life DRAM wafer data provided by a semiconductor industry.

The remainder of this chapter is organized as follows. We briefly review the traditional Bayesian classifier in Section 4.2. In Section 4.3, we propose the novel Bayesian classifier for uncertain data and describe the details. Section 4.4 is devoted to the application of the proposed classifier for the defect classification of DRAM wafers. The conclusion and future researches are discussed in Section 4.5.

4.2 Bayesian classification model for uncertain data

In this section, we propose a new Bayesian classifier that aims at classifying uncertain data. At first, we formally define the problem of the Bayesian classifier on uncertain data. Then, we propose a new multivariate kernel density estimation to handle the uncertainty of the data. Finally, we discuss the parameter optimization of the proposed kernel density estimate.

4.2.1 Problem definition

Suppose that a data set consists of N data objects with an s -dimensional feature vector and \mathbf{D} denotes the domain of the feature vector. A feature vector of a certain data object is an element of \mathbf{D} whereas that of an uncertain data object is a random variable with a pdf defined over the domain \mathbf{D} . The classification problem is to train a model that maps an uncertain feature vector \mathbf{Y} , distributed with pdf $g(\mathbf{x})$, to a posterior probability distribution $P(c_g|\mathbf{X} = \mathbf{Y})$ where c_g denotes the class g that belongs to a set of class $\mathbf{C} = \{c_1, c_2, \dots, c_G\}$. The class label of \mathbf{Y} is assigned as an element of \mathbf{C} that maximizes the posterior probability as follows.

$$Y(\mathbf{U}_*) = \operatorname{argmax}_{c_g \in \mathbf{C}} P(c_g|\mathbf{X} = \mathbf{Y}).$$

The posterior probability is estimated by using the class conditional probability of \mathbf{Y} , $P(\mathbf{X} = \mathbf{Y}|c_g)$, and the class prior probability, π_g , as follows:

$$P(c_g|\mathbf{X} = \mathbf{Y}) = \frac{P(\mathbf{X}=\mathbf{Y}|c_g)\pi_g}{\sum_{i=1}^G P(\mathbf{X}=\mathbf{Y}|c_i)\pi_i}$$

where class prior probability π_g can be estimated by the fraction of the data objects with the class label c_g in the set \mathbf{C} .

In the estimation of the class conditional probability $P(\mathbf{X} = \mathbf{Y}|c_g)$, we note that the feature vector \mathbf{Y} is a random one distributed with the joint pdf $g(\mathbf{y})$. By using the law of unconscious statistician (Casella and Berger, 2002), $P(\mathbf{X} = \mathbf{Y}|c_g)$ is equal to the expected value of the class conditional pdf of $P(\mathbf{X} = \mathbf{y}|c_g) = f_{\mathbf{X}}(\mathbf{y}|c_g)$ with respect to the uncertain feature vector \mathbf{Y} as follows:

$$P(\mathbf{X} = \mathbf{Y}|c_g) = \int_{\mathbf{y}} f_{\mathbf{x}}(\mathbf{y}|c_g)g(\mathbf{y}) d\mathbf{y}. \quad (4.1)$$

$P(\mathbf{X} = \mathbf{Y}|c_g)$ is the probability that the feature vector \mathbf{X} of a data object in class c_g and that of a data with an unknown class has exactly same values. Therefore, as $g(\mathbf{y})$ is similar to $f(\mathbf{x}|c_g)$, $P(\mathbf{X} = \mathbf{Y}|c_g)$ gets larger. Accordingly, the feature vector \mathbf{Y} has the larger posterior probability of $P(c_g|\mathbf{X} = \mathbf{Y})$.

In Eq. (4.1), the most challenging task is the estimation of the unknown class conditional pdf $f(\mathbf{x}|c_g)$ based on the uncertain data objects. Ren et al. (2009) extended a univariate kernel density estimation to estimates $f(x_j|c_g)$ for the each component of the feature vector \mathbf{x} and obtained $\hat{f}(\mathbf{x}|c_g)$ by the product of the estimated class conditional pdfs for all the components (i.e. $\hat{f}(\mathbf{x}|c_g) = \prod_{j=1}^s \hat{f}(x_j|c_g)$). However, this approach assumes that all the components of \mathbf{x} are mutually independent. Thus, it is inappropriate for the direct application of the existing approach to the density estimation of an uncertain feature vector when the components of the feature vector are not mutually independent. To this end, we develop a new multivariate KDE method on uncertain data to estimate $f(\mathbf{x}|c_g)$, which is introduced in the following subsection.

4.2.2 Estimation of class conditional joint pdf based on multivariate KDE

We propose a new multivariate KDE (MKDE) to estimate the class conditional pdf based on the probability distributions of the uncertain data objects. In the first sub-section, we introduce the multivariate kernel density estimation for certain data. We then describe new MKDE methods for uncertain continuous, categorical, and mixed data, respectively.

4.2.2.1 Review of the traditional kernel density estimate for certain data

The traditional kernel density estimation is a non-parametric approach aiming at estimating an unknown probability distribution function for *certain* data. It is assumed that x_1, x_2, \dots, x_n are independent and identically distributed with the univariate *pdf* of $f(x)$. The kernel density estimate of the pdf is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)$$

where $h>0$ is a smoothing parameter called bandwidth and $K(\cdot)$ is usually chosen to be a univariate probability density function. A typical kernel function is Gaussian kernel with mean zero and unit variance as follows.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Now we turn our attention to the multivariate kernel density estimate for certain. Extension to the multivariate kernel density estimation generally relies on the multivariate kernel function, $K_H(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x})$. Suppose that there are multivariate random samples of $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ that are independent and identically distributed with a multivariate *pdf* of $f(\mathbf{x})$. Based on the kernel function, the multivariate kernel density estimate is defined by (Scott, 2009)

$$\hat{f}(\mathbf{x}; \mathbf{H}) = \frac{1}{n} \sum_{i=1}^n K_H(\mathbf{x} - \mathbf{x}_i) \quad (4.2)$$

where \mathbf{H} is the smoothing parameter. When we use the standard multivariate normal kernel (i.e. $K(\mathbf{x}) = (2\pi)^{-s/2} \exp(-\frac{1}{2}\mathbf{x}^T\mathbf{x})$), the multivariate kernel density estimate in Eq. (4.2) leads to

$$\hat{f}(\mathbf{x}; \mathbf{H}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{(2\pi)^{\frac{s}{2}} |\mathbf{H}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_i)^T \mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_i)\right). \quad (4.3)$$

4.2.2.2 New multivariate kernel density (MKDE) estimation for uncertain data

4.2.2.2.1 Uncertain continuous data

The extension of the multivariate kernel function, described in Eq. (4.2), for uncertain data, is not straightforward due to some challenging issues. At first, the feature vectors are uncertain ones distributed with multivariate probabilities distributions. Suppose that there is an uncertain data object whose feature vector \mathbf{U}_i is distributed with $f_i(\mathbf{u})$. The function of uncertain feature vector, $K_H(\mathbf{x} - \mathbf{U}_i)$, is also uncertain. In order to handle the uncertainty of \mathbf{U}_i , the expected value of the kernel function was utilized as follows (Ren et al., 2009);

$$E[K_H(\mathbf{x} - \mathbf{U}_i)] = \int_{\mathbf{u}} K_H(\mathbf{x} - \mathbf{u}) f_i(\mathbf{u}) d\mathbf{u}. \quad (4.3)$$

Under the assumption that $K(\cdot)$ is the multivariate Gaussian kernel function and \mathbf{U}_i is distributed with multivariate normal with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$, the Eq. (4.3) leads to

$$\int_{\mathbf{u}} (2\pi)^{-\frac{s}{2}} |\mathbf{H}|^{-\frac{1}{2}} e^{-\frac{(x-u)^T \mathbf{H}^{-1} (x-u)}{2}} (2\pi)^{-\frac{s}{2}} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} e^{-\frac{(u-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (u-\boldsymbol{\mu}_i)}{2}} d\mathbf{u}. \quad (4.4)$$

By the convolution of two multivariate Gaussian distributions $N(\mathbf{u}, \mathbf{H})$ and $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, the Eq. (4.4) leads to

$$(2\pi)^{-\frac{s}{2}} |\mathbf{H} + \boldsymbol{\Sigma}_i|^{-\frac{1}{2}} e^{-\frac{(\mathbf{x}-\boldsymbol{\mu}_i)^T (\mathbf{H}+\boldsymbol{\Sigma}_i)^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)}{2}}. \quad (4.5)$$

Note that $E[K_{\mathbf{H}}(\mathbf{x} - \mathbf{U}_i)]$ converges to $f_i(\mathbf{x}) = (2\pi)^{-\frac{s}{2}} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} e^{-\frac{(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)}{2}}$ as the matrix norm of \mathbf{H} is close to zero. Therefore, the performance of the kernel function depends crucially on the matrix norm of \mathbf{H} .

However, the matrix \mathbf{H} has too many parameters (i.e. $s(s+1)/2$) to optimize the matrix norm. Due to the large number of parameters, it is difficult to control the matrix norm of \mathbf{H} . In order to overcome this issue, we propose to use the product kernel whose bandwidth matrix \mathbf{H} is diagonal. The product kernel function of the Eq. (4.3) is given by

$$E[K_{\mathbf{H}}(\mathbf{x} - \mathbf{U}_i)] = E \left[\prod_{j=1}^s h_j^{-1} K \left(\frac{x_j - U_{i,j}}{h_j} \right) \right].$$

where x_j (or $U_{i,j}$) is the j th component of \mathbf{x} (or \mathbf{U}_i). Thus, it is much easier to optimize since there are only s parameters. Another advantage of the product kernel is that this method has a good estimation of a density as well as a general kernel function. In Figure 4.2, we compare the product kernel function with the general kernel function by using the contour plots of density estimates based on the Eq. (4.5). We can see that the two contours look similar, which verifies the good approximation of the product kernel based density estimation for uncertain data.

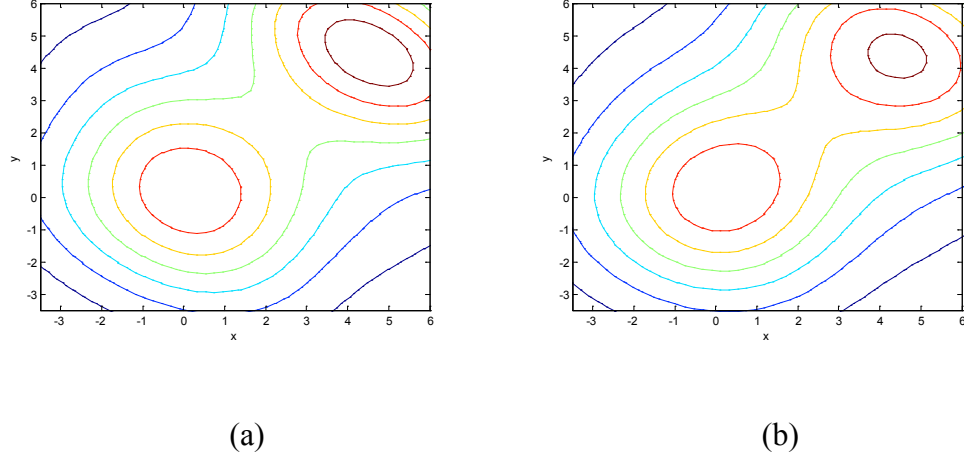


Figure 4.2 Contour plot of density estimates based on the bivariate kernel functions;

(a) General kernel ($\mathbf{H} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$) and (b) Product kernel ($\mathbf{H} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$).

Now we estimate $f(\mathbf{u}|c_g)$ based on the proposed kernel function. Suppose that there are independent uncertain continuous feature vectors $\{\mathbf{U}_i\}_{i=1}^{N_g}$ with the class label c_g where \mathbf{U}_i is distributed with the joint pdf of $f_i(\mathbf{u})$. In order to handle the uncertainty, we take the expected value of the KDE as follows.

$$\hat{f}(\mathbf{x}|c_g; h_1, \dots, h_s) = \frac{1}{N_g} \sum_{i=1}^{N_g} E \left[\prod_{j=1}^s h_j^{-1} K \left(\frac{x_j - U_{i,j}}{h_j} \right) \right] \quad (4.6)$$

where x_j and $U_{i,j}$ denote the j -th components of \mathbf{x} and \mathbf{U}_i . If $K(\cdot)$ is the univariate Gaussian kernel function and $h_j = h$ for all j , the expected value of $\prod_{j=1}^s h_j^{-1} K \left(\frac{x_j - U_{i,j}}{h_j} \right)$ can be obtained as follows.

$$E \left[\prod_{j=1}^s h_j^{-1} K \left(\frac{x_j - U_{i,j}}{h_j} \right) \right] = \int_{\mathbf{u}} (2\pi h^2)^{-\frac{s}{2}} e^{-\frac{\sum_{j=1}^s (x_j - u_j)^2}{2h^2}} f_i(\mathbf{u}) d\mathbf{u} \quad (4.7)$$

where u_j denote the j -th components of \mathbf{u} . When \mathbf{U}_i is distributed with $\mathbf{MVN}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, we use the Eq. (4.5) to obtain

$$(2\pi|h^2\mathbf{I} + \boldsymbol{\Sigma}_i|)^{-\frac{s}{2}} e^{-\frac{(\mathbf{x} - \boldsymbol{\mu}_i)^T (h^2\mathbf{I} + \boldsymbol{\Sigma}_i)^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)}{2}}$$

where \mathbf{I} is the identity matrix. Therefore, $\hat{f}(\mathbf{x}|c_g; h)$ is obtained by

$$\hat{f}(\mathbf{x}|c_g; h) = \frac{1}{N_g} \sum_{i=1}^{N_g} (2\pi|h^2\mathbf{I} + \boldsymbol{\Sigma}_i|)^{-\frac{s}{2}} e^{-\frac{(\mathbf{x} - \boldsymbol{\mu}_i)^T (h^2\mathbf{I} + \boldsymbol{\Sigma}_i)^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)}{2}}.$$

4.2.2.2.2 Uncertain categorical data

We consider the estimation of class conditional probability mass function (pmf) for uncertain categorical data. Similar with the continuous one, we use the multivariate *kernel* density estimation method in order to estimate the class conditional pmf. However, the general kernel function shown in the Eq. (4.2) may not be available for categorical data since it computes the distance between categorical features, which is not meaningful for categorical data. A suitable counter part of the kernel function for certain categorical data has been proposed as follows (Li and Racine, 2003).

Suppose that there are two independent certain categorical feature vectors \mathbf{x} and \mathbf{u} . Now we define a univariate kernel function $l(x_j, u_j, \lambda)$ for the certain categorical data. $l(x_j, u_j, \lambda)$ is equal to $1 - \lambda_j$ if $x_j = u_j$, and it is λ otherwise where the smoothing

parameter λ is between 0 and 0.5. For the multivariate data, the univariate kernel function is extended by

$$L(\mathbf{x}, \mathbf{u}; \lambda) = \prod_{j=1}^s l(x_j, u_j, \lambda) = (1 - \lambda)^{s-d(\mathbf{x}, \mathbf{u})} \lambda^{d(\mathbf{x}, \mathbf{u})}.$$

where the dissimilarity $d(\mathbf{x}, \mathbf{u})$ is ‘the number of disagreement’ in the corresponding components of two categorical feature vectors \mathbf{x} and \mathbf{u} , which can be defined by

$$d(\mathbf{x}, \mathbf{u}) = \sum_{j=1}^s 1(x_j \neq u_j)$$

where x_j and u_j are the j -th components of \mathbf{x} and \mathbf{u} .

Now we consider the kernel function of uncertain categorical data. Suppose that \mathbf{U}_i is an uncertain categorical feature vector with the pmf of $P(\mathbf{U}_i = \mathbf{u})$. In order to handle the uncertainty of \mathbf{U}_i , we take the expected value of the kernel function as follows.

$$E[L(\mathbf{x}, \mathbf{U}_i; \lambda)] = \sum_{\mathbf{u}} (1 - \lambda)^{s-d(\mathbf{x}, \mathbf{u})} \lambda^{d(\mathbf{x}, \mathbf{u})} P(\mathbf{U}_i = \mathbf{u}). \quad (4.8)$$

Note that the Eq. (4.8) is more generalized average of pdfs. To be specific, the kernel function $(1 - \lambda)^{s-d(\mathbf{x}, \mathbf{u})} \lambda^{d(\mathbf{x}, \mathbf{u})}$ gets larger when \mathbf{x} and \mathbf{u} have large number of ‘agreement components’. For example, $E[L(\mathbf{x}, \mathbf{U}_i; \lambda)]$ is equal to $P(\mathbf{U}_i = \mathbf{x})$ when $\lambda=0$. Based on the Eq. (4.8), $f(\mathbf{x}|c_g)$ for uncertain categorical data is driven by

$$\hat{f}(\mathbf{x}|c_g; \lambda_1, \dots, \lambda_s) = \frac{1}{N_g} \sum_{i=1}^{N_g} \sum_{\mathbf{u}} (1 - \lambda)^{s-d(\mathbf{x}, \mathbf{u})} \lambda^{d(\mathbf{x}, \mathbf{u})} P(\mathbf{U}_i = \mathbf{u}).$$

4.2.2.2.3 Uncertain mixed data

Now, we consider s -dimensional uncertain mixed continuous and categorical data. Recall that the continuous and categorical data requires the different types of kernel function. In order to estimate the mixed data, we can use both types of the kernel function by using product kernel (Li and Racine, 2003). Without loss of generality, we assume that the first $q(< s)$ components are continuous and the remaining components are categorical. Given N_g independent objects with an uncertain feature vector $\mathbf{U}_i = (\mathbf{U}_i^c, \mathbf{U}_i^d) \in c_g$, we estimate the class conditional pdf. Since \mathbf{U}_i 's have the uncertainty, we take the expected value of the product kernel of $K(\cdot)$ and $L(\cdot)$ as follows:

$$\hat{f}(\mathbf{x}|c_g; h_1, \dots, h_q, \lambda_{q+1}, \dots, \lambda_s) = \frac{1}{N_g} \sum_{i=1}^{N_g} E \left[\prod_{j=1}^q h_j^{-1} K\left(\frac{x_j - U_{i,j}^c}{h_j}\right) \prod_{j=q+1}^s L(x_j, U_{i,j}^d, \lambda_j) \right]$$

where \mathbf{U}_i is distributed with the pdfs of $f_i(\mathbf{u})$. Therefore, the expected value of the product kernel can be calculated by integrating out the pdf of \mathbf{U}_i as follows.

$$\int_{\mathbf{u}} \prod_{j=1}^q h_j^{-1} K\left(\frac{x_j - u_j^c}{h_j}\right) \prod_{j=q+1}^s L(x_j, u_j^d, \lambda_j) f_i(\mathbf{u}) d\mathbf{u}$$

where $K(\cdot)$ and $L(\cdot)$ are the kernel functions used in Eqs. (4.3) and (4.6) and $\int_{\mathbf{u}} d\mathbf{u}$ is equivalent to $\sum_{\mathbf{u}^d} \int_{\mathbf{u}^c} d\mathbf{u}^c$.

Remark: The expected value of the product kernel function is equivalent to the weighted integral (or sum) of the joint pdf $f_i(\mathbf{u})$. The product kernel puts the different weights on the pdfs based on the similarity between vectors \mathbf{x} and \mathbf{u} . For example, the product kernel of continuous features described in Eq. (4.4), $(2\pi h^2)^{-\frac{s}{2}} e^{-\frac{(\mathbf{x}-\mathbf{u})^T(\mathbf{x}-\mathbf{u})}{2h^2}}$, becomes large as the

Euclidean distance between \mathbf{x} and \mathbf{u} gets close to zero given the smoothing parameter h .

4.2.2.3 Bandwidth selection of the proposed kernel density estimate

In KDE, the selection of the smoothing parameters primarily determines the quality of the density estimation (Scott, 2009). There have been many existing researches conducted on the selection of the smoothing parameter for KDE of certain data. Generally, the optimal bandwidth of KDE is the one that minimizes the integrated square error (ISE) or the mean integrated square error (MISE) for density estimation (Silverman, 1986). However, Ghosh and Hall (2004) observed that the bandwidth that minimizes ISE or MISE may lead to the poor performance of the traditional Bayesian classifier for certain data. Despite the extensive literatures, there has been few work focused on the optimal bandwidth selection of KDE for uncertain data. Thus, we evaluate the performance of the proposed classifier by estimating the misclassification rate and choose the optimal bandwidth that minimizes the misclassification rate.

The misclassification rate of the proposed classifier is calculated by using the following equation.

$$\widehat{MR}(\mathbf{h}) = \frac{1}{n} \sum_{j=1}^n 1\{c_i \neq UB(\mathbf{U}_i; \mathbf{h})\}$$

where $UB(\mathbf{U}_i; \mathbf{h})$ denotes the proposed classifier with the parameter \mathbf{h} . We estimate the misclassification rate by using the training data set. Based on the misclassification rate from the training data, we select an optimal one of \mathbf{h}^* that minimizes the misclassification rate as follows.

4.3 Case study: defect classification of DRAM wafers

The procedure is composed of three steps: 1) the binarization of original FBT values, 2) the extraction of uncertain features that characterize the defects of a DRAM wafer, and 3) the defect classification of DRAM wafers based on the extracted features (see Figure 4.3).

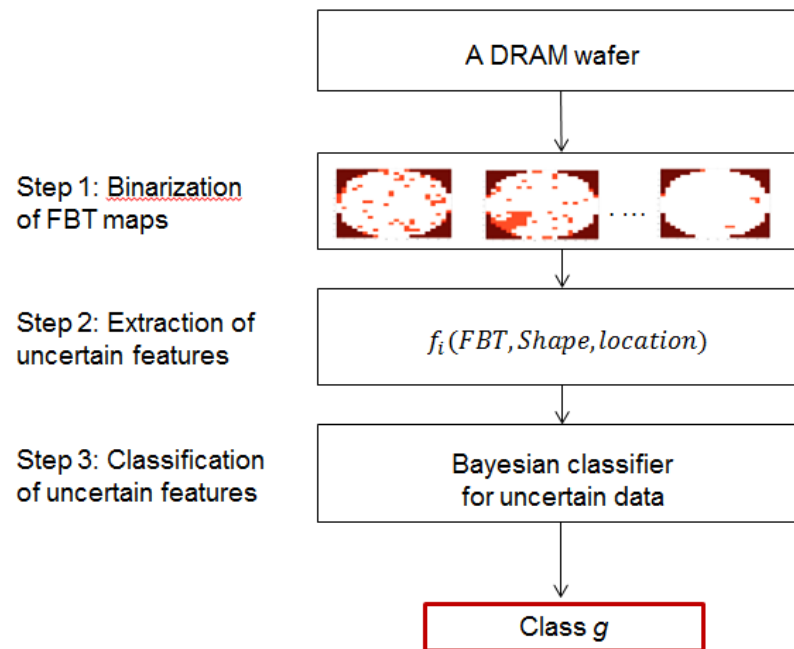


Figure 4.3 Framework of the defect classification of a DRAM wafer.

4.3.1 Transformation of original FBT values into binary ones

For a DRAM wafer, there are multiple FBT maps to be analyzed for the defect classification. Each of FBT maps consists of chips including a non-negative integer value.

Much work has gone towards identifying and classifying the spatial defect pattern of a wafer map (Chen and Liu, 2000; Liu *et al.*, 2002; Wang *et al.*, 2006; Jeong *et al.*, 2008; Chang *et al.* 2012; Liu and Chen, 2013). However, all these existing studies analyzed a single binary wafer map from functional testing results. For the remedy, we need to first transform the non-negative integer values into binary ones.

Before extracting features of DRAM wafers, we need to transform the original FBT values into binary ones. For binarizing the original FBT values, we investigate the probability distribution of original FBT values of functional chips in normal wafers by using the KDE for certain data. Suppose that $y_1^{(k)}, y_2^{(k)}, \dots, y_{n_f}^{(k)} \sim f^{(k)}(y)$, where $y_i^{(k)}$ is the original FBT value of the i -th functional chip of the k -th FBT map and n_f is the total number of functional chips on a normal wafer, respectively. Then, we choose the $100(1-\gamma)$ percentile ($\delta_{\gamma,k}$) as the threshold value of the binarization of the k -th FBT map. Based on the threshold value for the binarization of the k -th FBT map, the original FBT values can be binarized as follows:

$$x_i^{(k)} = \begin{cases} 1, & y_i^{(k)} > \delta_{\gamma,k} \\ 0, & y_i^{(k)} \leq \delta_{\gamma,k} \end{cases} \text{ for } i=1, \dots, K$$

where K is the total number of FBT maps for a given DRAM wafer. In order to obtain $\delta_{\gamma,k}$, we estimate the 95 percentile of $y_i^{(k)}$ based on Gaussian kernel-density estimation of $f^{(k)}(y)$.

4.3.2 Extraction of uncertain spatial feature

In this subsection, we propose several uncertain features that characterize the spatial defect patterns of binarized FBT maps for the given DRAM wafer. Based on the engineer's troubleshooting knowledge, we select i) an FBT map with the most significant defect pattern (hereafter critical map) among K maps, and identify ii) the defect shape and iii) the defect location on the selected critical map. Based on this information, we extract the three features: the critical map index, the defect shape, and location of the critical map. The uncertainty of each feature is quantified by estimating the probability distribution of the feature.

4.3.2.1 Uncertain feature of the critical map index

We extract the feature indicating the critical map index and quantify its uncertainty. At first, we detect FBT maps with a spatial defect pattern among the K binarized FBT maps. We then estimate the probability of each FBT map being the critical map.

A popular way of detecting the binarized FBT maps with spatial defect patterns is by measuring the spatial autocorrelation of defective chips. Many researches have been conducted for detecting a binary map with spatial autocorrelation (Cliff and Ord, 1981; Hansen and Thregod, 1998; Taam and Hamada, 1993). We assess the spatial autocorrelation of each binarized FBT map via spatial correlogram (Jeong *et al.*, 2008), which examines how many neighboring defective chips (or functional chips) are to be found around defective ones (or functional ones).

Suppose that we have K binarized FBT maps: $M_k, k = 1, 2, \dots, K$. For computing the spatial correlogram of the k -th FBT map M_k , $T^{(k)}(g)$ is obtained as follows.

$$T^{(k)}(r) = p^{(k)}c_{00}^{(k)}(r) + (1 - p^{(k)})c_{11}^{(k)}(r) \quad \text{for } k=1,2,3,\dots,K$$

where $c_{00}^{(k)}(g)$ and $c_{11}^{(k)}(g)$ denote the number of 0-to-0 and 1-to-1 *joins* (or *neighbors*) at a spatial lag r on the k -th binarized FBT map, and $p^{(k)}$ is the defective rates on M_k . Here, “0” indicates a functional chip while “1” presents a defective one. The higher value of $T^{(k)}(r)$, the more zeroes or ones are joined at the lag r . As proven by Jeong *et al.* (2008),

$$E[T^{(k)}(r)] = c^{(k)}(r)p^{(k)}(1 - p^{(k)})$$

$$\text{Var}[T^{(k)}(r)] = c^{(k)}(r)(p^{(k)})^2(1 - p^{(k)})^2$$

where $c(r) = c_{00}(r) + c_{01}(r) + c_{11}(r)$, and $c_{01}(r)$ denote the numbers of 0-to-1 joins at a spatial lag r on M_k . Then, the spatial correlogram of M_k is obtained by standardizing $T^{(k)}(r)$ for each lag r .

$$Z_T^{(k)}(r) = \frac{T^{(k)}(r) - c^{(k)}(r)p^{(k)}(1 - p^{(k)})}{\sqrt{c^{(k)}(r)(p^{(k)})^2(1 - p^{(k)})^2}}. \quad (4.9)$$

Now we use the spatial correlogram obtained in Eq. (4.9) to extract the feature M that indicates which FBT map is the most critical one among K ones. Let $\mathbf{Z}_r^{(k)} = [Z_T^{(k)}(1), \dots, Z_T^{(k)}(r)]$ be the first r lags of spatial correlogram for the k -th FBT map. Under the assumption of spatial independence, $\mathbf{Z}_r^{(k)}$ follows the approximate multivariate Normal distribution with mean zero vector of length r and covariance $\mathbf{\Sigma}_r^{(k)}$ (Jeong *et al.*, 2008). By using the following statistic, we can measure the spatial independence of M_k

$$T^{(k)2} = \mathbf{Z}_r^{(k)T} \hat{\Sigma}_r^{(k)-1} \mathbf{Z}_r^{(k)}$$

where $\hat{\Sigma}_r^{(k)}$ is the estimated covariance matrix using n spatially independent maps. The stronger spatial dependence M_k has, the larger $T^{(k)2}$ gets.

For identifying the critical map M , we test whether there is a large spatial autocorrelation on each FBT map by using the statistical hypothesis testing. For the testing the spatial independence of M_k , we use the probability distribution of $T^{(k)2}$ (Jeong *et al.*, 2008). Under the assumption of spatial independence, the k -th FBT map M_k is critical if

$$T^{(k)2} > \frac{r(n-1)}{n-r} F_{r,n-r}. \quad (4.10)$$

Based on an empirical study, $r = 4$ and $\alpha = 0.05$ is set.

We quantify the uncertainty in the critical FBT index M by estimating the probability that the k -th map M_k becomes the critical FBT map M ; thus, $P(M = k)$. The detail of the procedure is described as follows. First, we test that the k -th binarized FBT is critical by checking the condition shown in the Eq. (4.10). Then, the probability that the k -th FBT map being most critical is the reciprocal of the number of FBT maps satisfying Eq. (4.10) among K FBT maps if $T^{(k)2} > \frac{r(n-1)}{n-r} F_{r,n-r}$ and zero otherwise. Thus,

$$P(M = k) = \begin{cases} \frac{1}{N_c}, & T^{(k)2} > \frac{r(n-1)}{n-r} F_{r,n-r} \\ 0, & T^{(k)2} \leq \frac{r(n-1)}{n-r} F_{r,n-r} \end{cases} \quad (4.11)$$

where N_c is the number of FBT maps satisfying Eq. (4.10) among K binarized FBT maps.

4.3.2.2 Uncertain feature related to the defect shape of the critical FBT map

To extract an uncertain feature (S_M) that indicates the defect shape of a critical FBT map M , we classify the defect shape of the critical map using the corresponding spatial correlogram. Then, the uncertainty of S_M is quantified by estimating the probability that the correlogram of the map M is classified to a defect shape.

The spatial correlogram produces the unique shape for each spatial pattern of the binarized FBT map (Jeong *et al.*, 2008). Figure 4.4 presents several examples of binary maps with typical spatial defect patterns and the corresponding spatial correlograms. On the binary maps displayed in Figure 4.4, black and white colors indicate defective chips and functional ones, respectively. In Figure 4.4 a), the correlogram of the cluster pattern map smoothly changes along spatial lag r , and its absolute $Z_T(r)$ values are relatively larger than that of other spatial patterns. In Figure 4.4 b), the spatial correlogram of the circle pattern map also shows large $Z_T(r)$ absolute values, but with a soft cosine waveform. In Figure 4.4 c), the $Z_T(r)$ values of the repetitive pattern correlogram are consistently small, and no special pattern appears except for a frequent crossing around zero. Therefore, we can employ the spatial correlogram to select the defect shape of the critical FBT map among the typical defect patterns: cluster, circle, and repetition.

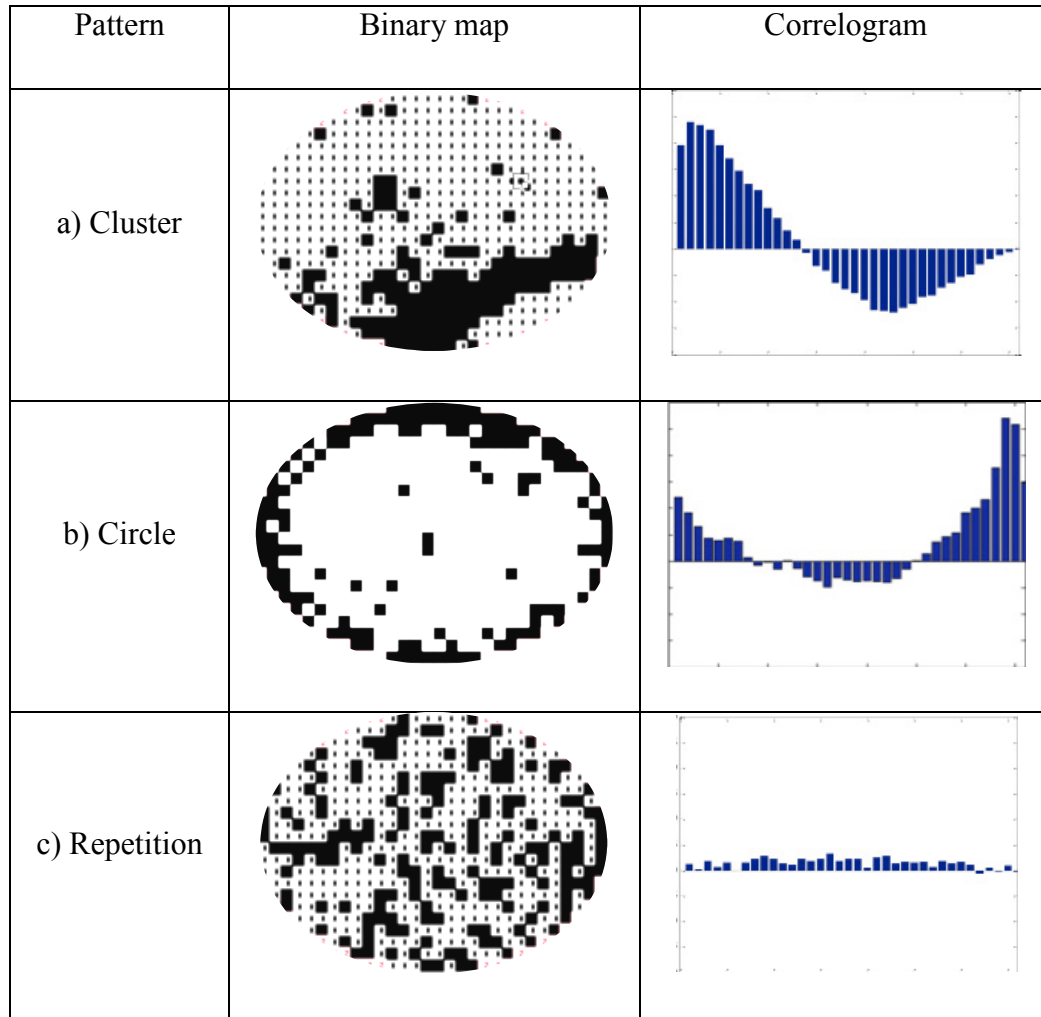


Figure 4.4 A spatial correlogram of an abnormal wafer map for each typical defect shape.

However, in the real life data, defect shapes of some critical maps do not belong to the four defect shape shown in Figure 4.4 (e.g. mixed one). In order to quantify such an uncertainty, we estimate the probability that S_M belongs to a defect shape among the four ones shown in Figure 4.4. For this, we adopt the linear discriminant analysis to compute the probability that the correlogram of the map $M (\mathbf{Z}^{(M)})$ belongs to $s \in \{\text{cluster, circle, repetition}\}$. In this study, the linear discriminant model is trained

based on 100 reference maps including cluster, circle, repetitive, and random patterns. Assuming the $M = k$, the probability that the map k has a defect shape can be obtained by estimating the posterior probability that the correlogram of the map i (i.e. $\mathbf{Z}^{(k)}$) belongs to a defect shape s , which is given by (Hastie et al. 2009)

$$P(S_M = s|M = k) = \frac{f(s|M=k)\hat{\pi}_s}{\sum_{s'} f(s'|M=k)\hat{\pi}_s} \quad (4.12)$$

where $f(s|M = k) = \frac{1}{(2\pi)^{r/2} |\hat{\Sigma}|^{1/2}} e^{-\frac{1}{2}(x-\hat{\mu}_s)^T \hat{\Sigma}^{-1}(x-\hat{\mu}_s)}$ and (define s')

$\hat{\pi}_s$ = the fraction of the defect shape s among all the reference maps

$\hat{\mu}_s$ = the mean of reference correlograms belonging defect shape s

$\hat{\Sigma}$ = the estimated covariance matrix of all the reference correlograms.

4.3.2.3 Uncertain feature related to the defect location of the critical FBT map

To extract the location related feature, we here define several subareas on the critical map M . In this study, we use the three different subareas: Edge, Middle, and Center (see Figure 4.5). We then calculate the defective rate of each subarea of the binarized critical FBT map. In order to quantify the uncertainty of the defect location related feature, we estimate the probability that the defect location of the critical map M , A_M , is a subarea a (i.e. $P(A_M = a|M)$) by normalizing the defect rate of each sub-area as follows.

$$P(A_M = a|M) = \frac{p_a^{(M)}}{p^{(M)}} \text{ for } a \in \{edge, middle, center\} \quad (4.13)$$

where $p^{(M)}$ is the defective rate of the map M and $P_a^{(M)}$ is the defective rate of the subarea a of the map M .

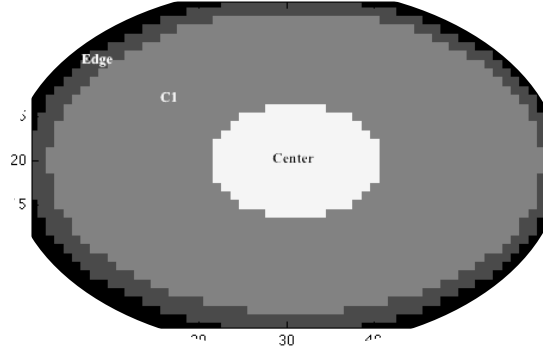


Figure 4.5 Three subareas of a binarized FBT map.

4.3.3 An illustrative example

Based on the extracted uncertain features from the Sections 4.3.1-4.3.2, we classify DRAM wafers by applying the proposed Bayesian classifier for uncertain data. In order to construct the uncertain data represented by joint probability distributions, we estimate the joint probability distribution of each DRAM wafer based on pdfs of Eqs. (4.11)-(4.13). Since the shape and location features S_M and A_M depend on the map M , we calculate the probability that the defect shape S_M is s and the defect location A_M is a conditioned on $M = k$.

Figure 4.6 shows a typical example that present how to estimate the joint pdf of an uncertain object. Among K binarized FBT maps, binarized FBT 03 maps have the probability being critical FBT maps of 0.33. Under the condition that $M = 3$, the FBT 3 map has the cluster pattern with the probability one and the defect location of ‘Edge’ with

the probability 0.47. Then, the joint probability of $(M = 3, S_M = Clu., A_M = Edge)$ is given by $P(M = 3, S_M = Clu., A_M = Edge) = 0.34 \times 0.98 \times 0.47 = 0.16$.

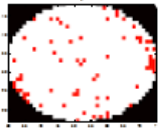
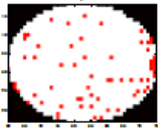
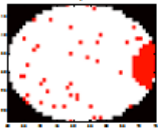
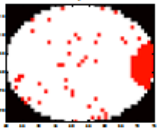
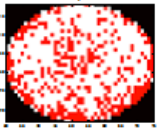
FBT				
00	01	05	07	08
				
Detection				
$P(M = 1)$ $= 0.00$	$P(M = 2)$ $= 0.00$	$P(M = 3)$ $= 0.33$	$P(M = 4)$ $= 0.33$	$P(M = 5)$ $= 0.33$
Shape			Location	
$P(S_M = Clu. M = 5) = 0.98$			$P(A_M = Edge M = 5) = 0.47$	
$P(S_M = Cir. M = 5) = 0.02$			$P(A_M = Mid. M = 5) = 0.41$	
$P(S_M = Rep. M = 5) = 0.00$			$P(A_M = Cen. M = 5) = 0.12$	

Figure 4.6 Estimation of joint probability based on extracted uncertain features.

4.4 Performance comparison

Real-life dynamic random access memory (DRAM) wafer maps provided by a semiconductor manufacturing company were analyzed to extract uncertain features and evaluate the performance of the proposed classifier. The data set used in this experiment includes 45 DRAM wafer maps that have three different classes. For each of the DRAM wafers, there are multiple fail bit tests (FBTs) maps. Based on engineering knowledge, we selected five FBTs (01, 02, 03, 04, and 05) because they capture spatial defects well.

We have compared the proposed method with three existing approaches, which include the traditional Bayesian classifier for certain data, uncertain naïve Bayes model, and decision tree for uncertain data, in terms of the misclassification probability. However, since the existing classifiers cannot be directly applied, we transformed the extracted uncertain feature vectors. At first, in order to apply the traditional naïve Bayes model, we extract a categorical feature vector without uncertainty for a DRAM wafer map. We estimate the feature vector by taking mode of the joint pdf $P(M = k, A_M = a, S_M = s)$ because the vector includes only the categorical feature components. Secondly, to apply the uncertain naïve Bayes classifier, we obtained a new feature vector with uncertainty whose components are mutually independent. For this, we estimated the marginal pdfs of the shape and location related uncertain features by marginalizing out the critical map index related feature M from the Eqs. (4.11)-(4.13), which removes the dependence between the critical map index related feature (M) and the defect location (A_M) and shape (S_M) related features. Table 4.2 compares the three different models in terms of the estimated misclassification rates using 5-fold cross validation. The proposed

model outperforms the traditional Bayesian classifier, the uncertain naïve Bayes, and decision tree for uncertain data.

Table 4.2 Performance comparison of the different classification models

Testing set	Traditional Bayes	Uncertain naïve Bayes	Uncertain decision tree	Proposed method
{1}	0.78	0.78	0.78	0.89
{2}	0.67	1.00	0.78	1.00
{3}	0.44	0.89	0.89	1.00
{4}	0.67	0.67	0.67	1.00
{5}	1.00	1.00	0.89	1.00
Average	0.71	0.87	0.80	0.98

Now, we discuss the rationale why the proposed classifier outperforms others by illustrating how the features used each classifier capture the spatial defect patterns on the binarized FBT maps. We have compared the uncertain features of the proposed Bayesian classifier with the certain features of the traditional Bayesian classifier, which was obtained by taking mode of the joint probability distribution of the uncertain features. Table 4.3 shows the joint probability distribution of uncertain features that are presented on the binarized FBT maps of two DRAM wafers. Since the wafers belong to the different classes and include the different defect patterns on the binarized FBT maps,

their uncertain features have the different probability distributions, which enables us to discriminate the wafers by applying the proposed classification model. On the other hand, both of the DRAM wafers (i.e. $M = 3, S_M = \text{Cluster}, A_M = \text{Edge}$) have the same modes of the joint pdfs in spite of the different probability distributions. Based on the modes of pdfs, the traditional Bayesian classifier cannot discriminate the DRAM wafers that belong to the different classes.

Table 4.3 Joint probability distributions of uncertain features of two DRAM wafers that belong to the different classes

Features			Probability distributions	
M	S_M	A_M	Wafer I	Wafer II
3	Cluster	Edge	0.17	0.37
		Middle	0.13	0.13
		Center	0.03	0.00
4	Cluster	Edge	0.16	0.00
		Middle	0.11	0.00
		Center	0.06	0.00
5	Cluster	Edge	0.05	0.00
		Middle	0.10	0.00
		Center	0.15	0.00
	Circle	Edge	0.00	0.27
		Middle	0.00	0.20
		Center	0.00	0.03

In addition, we illustrate why we need to consider the dependence between features by comparing the joint pdf of uncertain features of a DRAM wafer for the proposed Bayesian classifier with the product of marginal pdf of the features for uncertain naïve Bayes classifier. Figure 4.7 shows binarized FBT maps of a single DRAM wafer whose spatial features are represented by the pdfs presented in Table 4.4. Table 4.4 a) shows the original joint probability distribution that considers the dependence between features, which is estimated by using the pdfs of Eq. (4.11)-(4.13). In Table 4.4 b), the probability distribution was estimated by the product of the marginal pdfs of the uncertain features. As shown in Table 4.4, the joint probability in a) distribution captures the spatial defect patterns of the binarized FBT maps shown in Figure 4.7 better than the product of the marginal distributions in b). For example, the FBT 08 shows circle pattern on the corresponding binarized FBT map. Based on the probability distribution in Table 4.4 a), the DRAM wafer has high probabilities that $(M, S_M, A_M) = (3, \text{Cluster}, \text{Edge})$ and $(5, \text{Circle}, \text{Edge/Middle})$, which corresponds the defect patterns shown in Figure 4.7. On the other hand, based on the probabilities in Table 4.4 b), all binarized FBT maps are most likely to have cluster pattern on the edge, which contradicts the spatial pattern of the binarized FBT (08) map. This is because we have removed the effect of the selection of critical map from the defect and location features by estimating the marginal distributions of the two features.

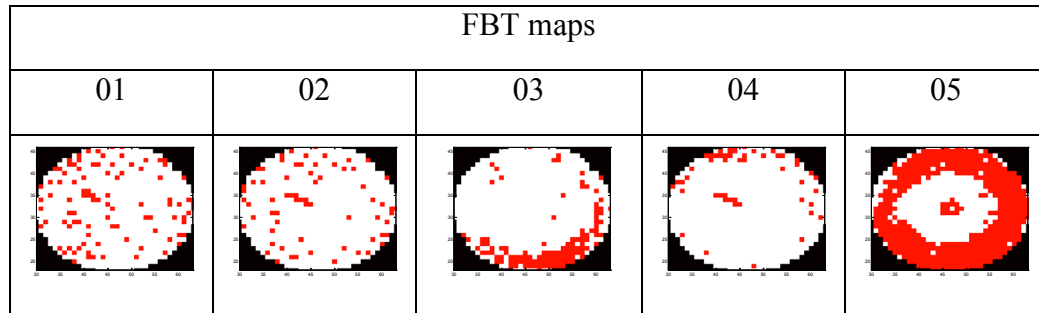


Figure 4.7 Binarized FBT maps of a DRAM wafer.

Table 4.4 Joint probability distribution of uncertain features representing the spatial characteristics of the binarized FBT maps in Figure 4.7

Features			a) Joint probability distribution	b) Product of marginal distributions
M	S_M	A_M		
3	Cluster	Edge	0.21	0.10
		Middle	0.11	0.06
		Center	0.01	0.03
	Circle	Edge	0.00	0.06
		Middle	0.00	0.03
		Center	0.00	0.02
	Repetition	Edge	0.00	0.01
4	Cluster	Edge	0.13	0.10
		Middle	0.04	0.06
		Center	0.10	0.03
	Circle	Edge	0.00	0.06
		Middle	0.00	0.03
		Center	0.00	0.02

5	Repetition	Edge	0.02	0.01
		Middle	0.01	0.00
		Center	0.02	0.00
	Random	Edge	0.01	0.00
		Center	0.01	0.00
	Cluster	Edge	0.00	0.10
		Middle	0.00	0.06
		Center	0.00	0.03
	Circle	Edge	0.15	0.06
		Middle	0.15	0.03
		Center	0.03	0.02
	Repetition	Edge	0.00	0.01

4.5 Conclusion

We have developed a new Bayesian classifier that considers the dependence between uncertain features based on extended multivariate kernel density estimation (KDE) approach for uncertain data. We have proposed the multivariate-KDE based Bayesian classifier by taking the expected values of multivariate kernel functions. We then applied the proposed classifier to the defect classification of a dynamic random access memory (DRAM) wafer. For this, we have constructed the uncertain categorical dataset by transforming the original fail bit test (FBT) maps into the binary ones, extracting the features characterizing defect patterns of the binary maps, and quantifying the uncertainty of the spatial features. Real-life DRAM wafer maps were analyzed to extract uncertain features and evaluate the performance of the proposed classifier. The experimental results have showed that the proposed approach outperforms the existing ones.

Future research directions include exploring other advanced classifiers such as Bayesian relevance vector machine and random forest to improve classification accuracy, and to extend the proposed approach to other applications such as bioinformatics. In addition, an interesting line of models for uncertain data can be developed by extending lots of traditional data mining techniques, such as feature selection (or ranking), regression analysis, and anomaly detection method for uncertain data.

CHAPTER 5

Concluding Remarks and Future Researches

5.1 Concluding remarks

In this dissertation, we have proposed and subsequently implemented several methodologies for spatial data mining with applications to semiconductor manufacturing process. In Chapter 2, we proposed a methodology for detecting abnormal DRAM wafers based on multivariate sequence tests. The proposed approach can detect abnormal DRAM wafers with multiple spatial maps using step-down spatial randomness testing. In addition, we used a spatial local de-noising method to eliminate noise of defect chips, letting us better distinguish systematic defect patterns on DRAM wafers from random ones. The experimental results show that the proposed method is more accurately detected abnormal DRAM wafers than other methods.

In Chapter 3, the automated classification procedure for failure patterns on fail bit maps in DRAM wafers was proposed. For this, we propose the novel regularized singular value decomposition (RSVD), which decomposes a binary matrix X into binary eigen-images. To decompose the matrix X , we have solved the formulation of RSVD by multiplicative updating procedure of RSVD. The experimental results show the proposed procedure has a great potential for the automatic classification of single bit failed maps

and non-single bit failed ones.

In Chapter 4, we proposed a uncertain data mining methodology for classifying abnormal DRAM wafers. For this, we extracted the categorical features characterizing defect patterns of the DRAM wafers, and quantified the uncertainty of the features by estimating the probability distribution. We then applied Bayesian classification model to classify the DRAM wafers based on the extracted uncertain categorical features. To estimate the conditional pdf of each class by handling the uncertainty of the features, we proposed the new multivariate KDE. Real-life DRAM wafer maps were analyzed to extract uncertain features and evaluate the performance of the proposed classifier. The experimental results have showed that the proposed approach outperforms the existing classification models for uncertain data.

5.2 Future researches

Future studies are needed that focus on improving and simplifying the proposed abnormality detection process of DRAM wafers. There are several parameters to be optimized in the detection process such as those for binarization and de-noising that may affect the performance of the proposed detection process. For this, we may develop the spatial randomness test based on the continuous spatial data. A meaningful line of research would be to extend spatial correlogram-based approach with the aim of identifying defect patterns on wafer maps.

In addition, we have focused on extracting features that capture the failure patterns shown on the fail bit maps for classifying DRAM chips. Future work is needed to explore more advanced classifiers such as recent variants of support vector machine (SVM) and Bayesian relevance vector machine to improve classification accuracy, and to extend the proposed approach to other applications such as bioinformatics. Also, we may extend the kernel singular value decomposition method to capture non-linearity in FBM.

Finally, the development of anomaly detection approach on uncertain data is also open for the future research aiming at detecting an abnormal DRAM wafer with an anomalous (i.e. unknown) defect pattern. We believe that more methodologies will bring improvements in this research area.

Appendix A. Derivation of Control Limits of Step-Down Spatial Randomness Test

Here we derive the control limits of step-down spatial randomness test. The combined correlogram $\mathbf{y}_r(k) = (\mathbf{y}_r^{(0)}, \mathbf{y}_r^{(1)}, \dots, \mathbf{y}_r^{(k)})$ is approximately normally distributed $N(\boldsymbol{\theta}_r(k), \boldsymbol{\Sigma}_r(k))$ where $\boldsymbol{\theta}_r(i) = (\boldsymbol{\theta}_r^{(0)}, \boldsymbol{\theta}_r^{(1)}, \dots, \boldsymbol{\theta}_r^{(k)})$, and the covariance matrix $\boldsymbol{\Sigma}_r(k)$ can be partitioned as

$$\boldsymbol{\Sigma}_r(k) = \begin{bmatrix} \boldsymbol{\Sigma}_r^{(1,1)} & \dots & \boldsymbol{\Sigma}_r^{(1,k)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_r^{(k,1)} & \dots & \boldsymbol{\Sigma}_r^{(k,k)} \end{bmatrix}.$$

Similarly, we can partition the sample covariance matrix $\mathbf{S}_r(k)$ of a combined correlogram $\mathbf{y}_r(k)$ from a given DRAM wafer into

$$\mathbf{S}_r(k) = \begin{bmatrix} \mathbf{S}_r^{(1,1)} & \dots & \mathbf{S}_r^{(1,k)} \\ \vdots & \ddots & \vdots \\ \mathbf{S}_r^{(k,1)} & \dots & \mathbf{S}_r^{(k,k)} \end{bmatrix}.$$

Note that $(n-1)$ times the sample covariance matrix $\mathbf{S}_r(k)$ has a Wishart distribution with $(n-1)$ degrees of freedom, and the parameter $\boldsymbol{\Sigma}_r(k)$, where $n (\geq k * r + 1)$, is the number of wafer samples. Then, the step-down randomness test statistic Z_k for $H_0^{(k)}$ is defined as (Marden and Perlman, 1990)

$$Z_k = \frac{T_k^2 - T_{k-1}^2}{1 + T_{k-1}^2 / (n-1)},$$

where $T_k^2 = \begin{pmatrix} \mathbf{y}_r^{(1)} \\ \vdots \\ \mathbf{y}_r^{(k)} \end{pmatrix}^T \mathbf{S}_r(k)^{-1} \begin{pmatrix} \mathbf{y}_r^{(1)} \\ \vdots \\ \mathbf{y}_r^{(k)} \end{pmatrix}.$

Then, based on the result from Marden and Perlman (1980), we show that the conditional distribution of Z_k has a non-central F distribution with the noncentral parameter

$$\frac{\Delta_k}{1+T_{k-1}^2/(n-1)}.$$

$$Z_k | T_{k-1}^2 \sim \frac{(n-1)r}{n-kr} F(r, n-kr; \frac{\Delta_k}{1+T_{k-1}^2/(n-1)}),$$

where $\Delta_k = \tau_k^2 - \tau_{k-1}^2$, $\tau_k^2 = \boldsymbol{\theta}_r(k)^T \boldsymbol{\Sigma}_r(k)^{-1} \boldsymbol{\theta}_r(k)$. If $\Delta_k = 0$ (i.e., $\boldsymbol{\theta}_r^{(k)} = \mathbf{0}$), then Z_k and T_{k-1}^2 are independent. Therefore, the control limit of Z_k with a false-alarm rate α_k can be established as

$$CL_k = \frac{(n-1)r}{(n-kr)} F_{\alpha_k}(r, n-kr).$$

Appendix B. Learning Algorithm for RSVD

Here we present derivations on the multiplicative updating procedure for RSVD based on the gradient decent algorithm. As mentioned in Section 3.3, we need to minimize the following objective function;

$$f(\mathbf{U}, \mathbf{S}, \mathbf{V}) = \frac{1}{2} \|\mathbf{X} - \mathbf{USV}^T\|^2 + \frac{\lambda}{2} \sum_{k=1}^m \sum_{i=1}^m \sum_{j=1}^n \left((s_{kk} u_{ik} v_{jk})^2 - (s_{kk} u_{ik} v_{jk}) \right)^2.$$

The first term in the objective function can be expressed as follows.

$$\begin{aligned} \|\mathbf{X} - \mathbf{USV}^T\|^2 &= \text{tr}[(\mathbf{X} - \mathbf{USV}^T)^T(\mathbf{X} - \mathbf{USV}^T)] \\ &= \text{tr}[(\mathbf{X}^T \mathbf{X} - \mathbf{VS}^T \mathbf{U}^T \mathbf{X} - \mathbf{X}^T \mathbf{USV}^T + \mathbf{VS}^T \mathbf{U}^T \mathbf{USV}^T)]. \end{aligned}$$

Firstly taking derivative of the objective function w.r.t u_{ik} , we have

$$\begin{aligned} \frac{\partial}{\partial u_{ik}} f(\mathbf{U}, \mathbf{S}, \mathbf{V}) &= (-\mathbf{XVS}^T + \mathbf{USV}^T \mathbf{VS}^T)_{ik} \\ &\quad + \lambda \sum_{j=1}^n \left((s_{kk} u_{ik} v_{jk})^2 - (s_{kk} u_{ik} v_{jk}) \right) (2(s_{kk} u_{ik} v_{jk}) - 1)(s_{kk} v_{jk}) \\ &= -(\mathbf{XVS}^T - \mathbf{USV}^T \mathbf{VS}^T)_{ik} \\ &\quad + \lambda \sum_{j=1}^n \left(2(s_{kk} u_{ik} v_{jk})^3 - 3(s_{kk} u_{ik} v_{jk})^2 + (s_{kk} u_{ik} v_{jk}) \right) (s_{kk} v_{jk}). \end{aligned}$$

For u_{ik} , using gradient decent method with the step size η_{ik} , we have

$$\begin{aligned} u_{ik} &\leftarrow u_{ik} - \eta_{ik} \frac{\partial f}{\partial u_{ik}} \\ &= u_{ik} + \eta_{ik} [(XVS^T - USV^T VS^T)_{ik} \\ &\quad - \lambda \sum_{j=1}^n \left(2(s_{kk} u_{ik} v_{jk})^3 - 3(s_{kk} u_{ik} v_{jk})^2 + (s_{kk} u_{ik} v_{jk}) \right) (s_{kk} v_{jk})]. \end{aligned}$$

By setting $\eta_{ik} = \frac{u_{ik}}{(USV^T VS)_{ik} + \lambda \sum_{j=1}^n (2s_{kk}^4 u_{ik}^3 v_{jk}^4 + s_{kk}^2 u_{ik} v_{jk}^2)}$, we get the following equation:

$$u_{ik} \leftarrow u_{ik} \frac{(XVS^T)_{ik} + 3\lambda \sum_{j=1}^n s_{kk}^3 u_{ik}^2 v_{jk}^3}{(USV^T VS)_{ik} + \lambda \sum_{j=1}^n (2s_{kk}^4 u_{ik}^3 v_{jk}^4 + s_{kk}^2 u_{ik} v_{jk}^2)}.$$

Secondly, taking derivative of the objective function w.r.t s_{kl} , we have

$$\begin{aligned} &\frac{\partial}{\partial s_{kl}} f(\mathbf{U}, \mathbf{S}, \mathbf{V}) \\ &= (-\mathbf{U}^T \mathbf{XV} + \mathbf{U}^T \mathbf{USV}^T \mathbf{V})_{kl} \\ &\quad + 1(k \\ &= l) \lambda \sum_{i=1}^m \sum_{j=1}^n \left(2(s_{kk} u_{ik} v_{jk})^3 - 3(s_{kk} u_{ik} v_{jk})^2 + (s_{kk} u_{ik} v_{jk}) \right) (u_{ik} v_{jk}) \end{aligned}$$

where $1(k = l)$ is one if $k = l$, and zero otherwise.

For s_{kl} , using gradient decent method with the step size ρ_{kl} , we have

$$s_{kl} \leftarrow s_{kl} - \rho_{kl} \frac{\partial f}{\partial s_{kl}}$$

$$= s_{kl} + \rho_{kl} [(U^T X V - U^T U S V^T V)_{kl}$$

$$- 1(k=l)\lambda \sum_{i=1}^m \sum_{j=1}^n (2s_{kk}^3 u_{ik}^4 v_{jk}^4 - 3s_{kk}^2 u_{ik}^3 v_{jk}^3 + s_{kk} u_{ik}^2 v_{jk}^2)].$$

By setting $\rho_{kl} = \frac{s_{kl}}{(U^T U S V^T V)_{kl} + 1(k=l)\lambda \sum_{i=1}^m \sum_{j=1}^n (2s_{kk}^3 u_{ik}^4 v_{jk}^4 + s_{kk} u_{ik}^2 v_{jk}^2)}$, we have

$$s_{kl} \leftarrow s_{kl} \frac{(U^T X V)_{kl} + 1(k=l)\lambda \sum_{i=1}^m \sum_{j=1}^n 3s_{kk}^2 u_{ik}^3 v_{jk}^3}{(U^T U S V^T V)_{kl} + 1(k=l)\lambda \sum_{i=1}^m \sum_{j=1}^n (2s_{kk}^3 u_{ik}^4 v_{jk}^4 + s_{kk} u_{ik}^2 v_{jk}^2)}$$

Finally, taking the derivative of the objective function w.r.t v_{jl} , we have

$$\frac{\partial}{\partial v_{jl}} f(\mathbf{U}, \mathbf{S}, \mathbf{V})$$

$$= (-X^T U S + V S U^T U S)_{jl}$$

$$+ 1(k$$

$$= l)\lambda \sum_{i=1}^m \left(2(s_{kk} u_{ik} v_{jk})^3 - 3(s_{kk} u_{ik} v_{jk})^2 + (s_{kk} u_{ik} v_{jk}) \right) (u_{ik} s_{kk}).$$

For v_{jl} , using gradient decent method with the step size ξ_{jl} , we have

$$v_{jl} \leftarrow v_{jl} - \xi_{jl} \frac{\partial f}{\partial v_{jl}}$$

$$= v_{jl} + \xi_{jl} [(X^T U S - V S U^T U S)_{jl}$$

$$-1(k=l)\lambda \sum_{i=1}^m (2s_{kk}^4 u_{ik}^4 v_{jk}^3 - 3s_{kk}^3 u_{ik}^3 v_{jk}^2 + s_{kk}^2 u_{ik}^2 v_{jk})].$$

Setting $\xi_{jl} = \frac{v_{jl}}{(\mathbf{VSU}^T \mathbf{US})_{jl} + 1(k=l)\lambda \sum_{i=1}^m (2s_{kk}^4 u_{ik}^4 v_{jk}^3 + s_{kk}^2 u_{ik}^2 v_{jk})}$, we have

$$v_{jl} \leftarrow v_{jl} \frac{(\mathbf{X}^T \mathbf{US})_{jl} + 1(k=l)\lambda \sum_{i=1}^m 3s_{kk}^3 u_{ik}^3 v_{jk}^2}{(\mathbf{VSU}^T \mathbf{US})_{jl} + 1(k=l)\lambda \sum_{i=1}^m (2s_{kk}^4 u_{ik}^4 v_{jk}^3 + s_{kk}^2 u_{ik}^2 v_{jk})}.$$

References

- Aggarwal, C. C., and Yu, P. S. (2009). A survey of uncertain data algorithms and applications. *Knowledge and Data Engineering, IEEE Transactions on*, 21(5), 609-623.
- Amirshahi, S. A., and Torkamani-Azar, F. (2012). Human optic sensitivity computation based on singular value decomposition. *Optica Applicata*, 42(1), 137-146.
- Andrews, H., and Patterson, C. (1976). Singular value decompositions and digital image processing. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 24(1), 26-53.
- Angiulli, F., and Fasseti, F. (2013). Nearest Neighbor-Based Classification of Uncertain Data. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(1), 1.
- Bi, J. and Zhang, T. (2005). Support vector classification with input data uncertainty. *Advances in neural information processing systems*, 17, 161-169.
- Casella, G., and Berger, R. L. (2002). *Statistical inference 2nd edition*. Pacific Grove, CA: Duxbury.
- Chang, C. W., Chao, T. M., Horng, J. T., Lu, C. F., and Yeh, R. H. (2012). Development Pattern Recognition Model for the Classification of Circuit Probe Wafer Maps on Semiconductors. *Components, Packaging and Manufacturing Technology, IEEE Transactions on*, 2(12), 2089-2097.
- Chen, F. L., and Liu, S. F. (2000). A neural-network approach to recognize defect spatial pattern in semiconductor fabrication. *Semiconductor Manufacturing, IEEE Transactions on*, 13(3), 366-373.
- Chien, C. F., Wang, W. C., and Cheng, J. C. (2007). Data mining for yield enhancement in semiconductor manufacturing and an empirical study. *Expert Systems with Applications*, 33(1), 192-198.
- Chou, P. B., Rao, A. R., Sturzenbecker, M. C., Wu, F. Y., and Brecher, V. H. (1997). Automatic defect classification for semiconductor manufacturing. *Machine Vision and Applications*, 9(4), 201-214.
- Cliff, A. D., and Ord, J. K. (1981). *Spatial processes: models and applications*. London: Pion.
- Collica, R. S., Card, J. P., and Martin, W. (1995). SRAM bitmap shape recognition and sorting using neural networks. *Semiconductor Manufacturing, IEEE Transactions on*, 8(3), 326-332.

- Duong, T. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software*, 21(7), 1-16.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to Bootstrap*. Chapman and Hall, New York.
- Fenner, J. S., Jeong, M. K., and Lu, J. C. (2005). Optimal automatic control of multistage production processes. *Semiconductor Manufacturing, IEEE Transactions on*, 18(1), 94-103.
- Fukushima, T., Narazaki, H., and Konishi, M. (1999). A method of feature extraction from an image for quality analysis. In *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on* (vol. 2, pp. 952-957).
- Ghosh, A. K., and Chaudhuri, P. (2004). Optimal smoothing in kernel discriminant analysis. *Statistica Sinica*, 14(2), 457-484.
- Hall, P., and Kang, K. H. (2005). Bandwidth choice for nonparametric classification. *Annals of statistics*, 284-306.
- Hamada, T., and Sugimoto, M. (1997). Application of a bitmap analysis system to the forefront of DRAM devices development. In *Advanced Semiconductor Manufacturing Conference and Workshop, 1997. IEEE/SEMI* (pp. 222-227). IEEE.
- Han, Y., Kim, J., and Lee, C. (2005). Automatic detection of failure patterns using data mining. In *Knowledge-Based Intelligent Information and Engineering Systems* (pp. 1312-1316). Springer Berlin Heidelberg.
- Hansen, M. H., Nair, V. N., and Friedman, D. J. (1997). Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects. *Technometrics*, 39(3), 241-253.
- Hansen, C. K., and Thyregod, P. (1998). Use of wafer maps in integrated circuit manufacturing. *Microelectronics Reliability*, 38(6), 1155-1164.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*. New York: Springer.
- Higgins, J. J. (2003). *Introduction to modern nonparametric statistics*. Duxbury.
- Hsieh, H. W. and Chen, F. L. (2004). Recognition of defect spatial patterns in semiconductor fabrication. *International journal of production research*, 42(19), 4153-4172.
- Hsieh, Y. L., Tzeng, G. H., Lin, G. R., and Yu, H. C. (2010). Wafer sort bitmap data analysis using the PCA-based approach for yield analysis and optimization. *Semiconductor Manufacturing, IEEE Transactions on*, 23(4), 493-502.

- Hülsmann, J., and Brockmann, W. (2012). Classification of Uncertain Data: An Application in Nondestructive Testing. In *Advances in Computational Intelligence* (pp. 231-240). Springer Berlin Heidelberg.
- Jeong, Y. S., Kim, S. J., and Jeong, M. K. (2008). Automatic identification of defect patterns in semiconductor wafer maps using spatial correlogram and dynamic time warping. *Semiconductor Manufacturing, IEEE Transactions on*, 21(4), 625-637.
- Jung, J. J. (2011). Ontology-based decision support system for semiconductors EDS testing by wafer defect classification. *Expert Systems with Applications*, 38(6), 7425-7429.
- Kameyama, K., and Kosugi, Y. (1999). Semiconductor defect classification using hyperellipsoid clustering neural networks and model switching. In *Neural Networks, 1999. IJCNN'99. International Joint Conference on* (pp. 3505-3510). IEEE.
- Kim, B., Jeong, Y. S., Tong, S. H., Chang, I. K., and Jeong, M. K. (2015a). Step-down spatial randomness test for detecting abnormalities in DRAM wafers with multiple spatial maps, to appear in *Semiconductor Manufacturing, IEEE Transactions on*.
- Kim, B., Jeong, Y. S., Tong, S. H., Chang, I. K., and Jeong, M. K. (2015b). A Regularized Singular Value Decomposition-Based Approach for Failure Pattern Classification on Fail Bit Map in a DRAM Wafer. *Semiconductor Manufacturing, IEEE Transactions on*, 28(1), 41-49.
- Lee, D. D., and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems* (pp. 556-562).
- Liu, S. F., Chen, F. L., and Lu, W. B. (2002). Wafer bin map recognition using a neural network approach. *International journal of production research*, 40(10), 2207-2223.
- Li, Q., and Racine, J. (2003). Nonparametric estimation of distributions with categorical and continuous data. *journal of multivariate analysis*, 86(2), 266-292.
- Li, Q., Maasoumi, E., and Racine, J. S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics*, 148(2), 186-200.
- Li, T. S., and Huang, C. L. (2009). Defect spatial pattern recognition using a hybrid SOM-SVM approach in semiconductor manufacturing. *Expert Systems with Applications*, 36(1), 374-385.
- Li, J., Huang, Y., Cheng, W. T., Schuermyer, C., Xiang, D., Faehn, E., and Farrugia, R. (2012). A Hybrid Flow for Memory Failure Bitmap Classification. In *Test Symposium (ATS), 2012 IEEE 21st Asian* (pp. 314-319). IEEE.
- Lim, J. S. (1990). Image restoration. *Two-Dimensional Signal and Image Processing*, 524-588.

- Liu, W., Tang, A., Ye, D., and Ji, Z. (2008). Nonnegative singular value decomposition for microarray data analysis of spermatogenesis. In *Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on* (pp. 225-228). IEEE.
- Liu, C. H., Lin, S. J., and Lewis, C. (2010). Life cycle assessment of DRAM in Taiwan's semiconductor industry. *Journal of Cleaner Production*, 18(5), 419-425.
- Liu, C. W., and Chien, C. F. (2013). An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing. *Engineering Applications of Artificial Intelligence*, 26(5), 1479-1486.
- Marden, J., and Perlman, M. D. (1980). Invariant tests for means with covariates. *The Annals of Statistics*, 25-63.
- Marden, J. I., and Perlman, M. D. (1990). On the inadmissibility of step-down procedures for the Hotelling T² problem. *The Annals of Statistics*, 172-190.
- Miettinen, P., Mielikainen, T., Gionis, A., Das, G., and Mannila, H. (2008). The discrete basis problem. *Knowledge and Data Engineering, IEEE Transactions on*, 20(10), 1348-1362.
- Muthumalai, V., Ling, Y. E., Ross, R., Lockwood, R., Wu, M., & White, S. (2014, May). Methodologies for fast yield ramp with limited engineering resources utilizing Inline Defect data overlay to SRAM Bitmap failure and Logic Diagnostics. In *Advanced Semiconductor Manufacturing Conference (ASMC), 2014 25th Annual SEMI* (pp. 341-344). IEEE.
- Nakamae, K., Itoh, A., and Fujioka, H. (2001). Fail pattern classification and analysis system of memory fail bit maps. In *Modeling and Simulation of Microsystems*. (pp. 598-601).
- Otsu, N. (1979). A threshold selection method from gray-level histogram. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1), 62-66.
- Pei, J., Jiang, B., Lin, X., and Yuan, Y. (2007). Probabilistic skylines on uncertain data. In *Proceedings of the 33rd international conference on Very large data bases* (pp. 15-26). VLDB Endowment.
- Qin, B., Xia, Y., Prabhakar, S., and Tu, Y. (2009). A rule-based classification algorithm for uncertain data. In *Data Engineering, 2009. ICDE'09. IEEE 25th International Conference on* (pp. 1633-1640). IEEE.
- Qin, B., Xia, Y., and Li, F. (2010). A Bayesian classifier for uncertain data. In *Proceedings of the 2010 ACM Symposium on Applied Computing* (pp. 1010-1014). ACM.

- Ramaswamy, S., Rastogi, R., and Shim, K. (2000). Efficient algorithms for mining outliers from large data sets. In *ACM SIGMOD Record* (Vol. 29, No. 2, pp. 427-438). ACM.
- Ren, J., Lee, S. D., Chen, X., Kao, B., Cheng, R., and Cheung, D. (2009). Naive bayes classification of uncertain data. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on* (pp. 944-949). IEEE.
- Sadek, R. A. (2012). SVD Based Image Processing Applications: State of The Art, Contributions and Research Challenges. *International Journal of Advanced Computer Science and Applications*, 3(7), 26-34.
- Schafer, J. L. (2010). *Analysis of incomplete multivariate data*. CRC press.
- Scott, D. W. (2009). *Multivariate density estimation: theory, practice, and visualization*. John Wiley and Sons.
- Segal, J., Jee, A., Lepejian, D., and Chu, B. (2001). Using electrical bitmap results from embedded memory to enhance yield. *IEEE Design and Test of Computers*, 18(3), 28-39.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. CRC press.
- Sun, L., Cheng, R., Cheung, D. W., and Cheng, J. (2010). Mining uncertain data with probabilistic guarantees. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 273-282). ACM.
- Taam, W., and Hamada, M. (1993). Detecting spatial effects from factorial experiments: An application from integrated-circuit manufacturing. *Technometrics*, 35(2), 149-160.
- Theodoridis, S. and Koutroumbas, K. (2009). *Pattern Recognition 4th edition*. Academic Press.
- Tong, L. I., Wang, C. H., & Huang, C. L. (2005). Monitoring defects in IC fabrication using a Hotelling T² control chart. *Semiconductor Manufacturing, IEEE Transactions on*, 18(1), 140-147.
- Tong, S. H., and Yum, B. J. (2006). Development of a dual burn-in policy for semiconductor products based on the number of defective neighborhood chips. *International Journal of Reliability, Quality and Safety Engineering*, 13(06), 501-525.
- Tong, S. H., and Yum, B. J. (2008). A dual burn-in policy for defect-tolerant memory products using the number of repairs as a quality indicator. *Microelectronics Reliability*, 48(3), 471-480.
- Tsang, S., Kao, B., Yip, K. Y., Ho, W. S., and Lee, S. D. (2011). Decision trees for uncertain data. *Knowledge and Data Engineering, IEEE Transactions on*, 23(1), 64-78.

- Vollrath, J., Lederer, U., and Hladschik, T. (2001). Compressed bit fail maps for memory fail pattern classification. *Journal of Electronic Testing*, 17(3-4), 291-297.
- Wang, C. H., Kuo, W., and Bensmail, H. (2006). Detection and classification of defect patterns on semiconductor wafers. *IIE Transactions*, 38(12), 1059-1068.
- Wang, C. H. (2008). Recognition of semiconductor defect patterns using spatial filtering and spectral clustering. *Expert Systems with Applications*, 34(3), 1914-1923.
- Yang, Z., and Laaksonen, J. (2007). Multiplicative updates for non-negative projections. *Neurocomputing*, 71(1), 363-373.
- Yin, J., Yu, J., Xie, S., Sun, D., Ling, Y. E., Song, Z., Schiller, C. E., Teverskoy, G., and Villalobos, M. J. (2013). Case studies of fault isolation for the global failing patterns on SRAM bitmap caused by the defects in peripheral logic regions. In *Advanced Semiconductor Manufacturing Conference (ASMC), 2013 24th Annual SEMI*(pp. 226-230). IEEE.
- Yuan, T., Bae, S. J., and Park, J. I. (2010). Bayesian spatial defect pattern recognition in semiconductor fabrication using support vector clustering. *The International Journal of Advanced Manufacturing Technology*, 51(5-8), 671-683.
- Zanon, T., Ferdman, M., Maly, W., and Komeyli, K. (2003). Analysis of IC manufacturing process deformations: An automated approach using SRAM bit fail maps. In *International Symposium for Testing and Failure Analysis* (pp. 232-241).
- Zhang, Z., Ding, C., Li, T., and Zhang, X. (2007). Binary matrix factorization with applications. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on* (pp. 391-400). IEEE.