Wafer Defect Patterns Recognition Based on OPTICS and Multi-Label Classification

Mengying Fan, Qin Wang, Ben van der Waal
School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong Unversity
Texas Instruments, Shanghai, China
Email: {fiona_fanmy, qinqinwang}@sjtu.edu.cn

***Abstract*—In the industry of integrated circuits, defect patterns shown on a wafer map contain crucial information for quality engineers to find the cause of defect to increase yield. This paper proposes a method for wafer defect pattern recognition which could recognize more than one defect patterns based on Ordering Point to Identify the Cluster Structure(OPTICS) and Support Vector Machine(SVM). The effectiveness of the proposed method has been verified from following three aspects from a real-world data set of wafer maps(WM-811K): salient defect pattern recognition accuracy up to 94.3% and the accuracy of some types has an obvious improvement, multi-patterns recognition accuracy(82.0%), and computation time has a significantly reduction.**

***Index Terms*—pattern recognition, wafer defect patterns, OPTICS.**

I. Introduction

For manufacturing of integrated circuits, wafer maps are created after wafer electrical-test which provide visual details that are crucial for identifying the possible cause of failures. In general, wafer defects are classified into two categories, global defects and local defects [9]. Global defects are scattered over the whole wafer and do not have significant patterns. The causes of global defects is always difficult to address. And as to local defects, there are usually some obvious patterns to be inspected [6]. Experienced engineers will go back to check the production line to determining possible causese such as human mistakes, particles from equipment, and chemical stains, etc [10].For example, a curvilinear shape is probably caused by scratch. But if a wafer's global defects almost cover the whole wafer, the cause also may be considered as human mistakes.

With the development of semiconductor manufacturing technology, the size and cost of wafer increased a lot. Therefore, analysis of the causes of wafer defects become important to increase yield and maximize throughput of good dies. However, the traditional method to analysis wafer map is time consuming and low accuracy. Based on the study conducted by A.Freeman [4], the accuracy of human-expert based methods are less than 45%. Many studies have investigated wafer map failure pattern recognition problem in recent years. In [9]- [11], pattern identification is considered to be a model-selection process. Like multivariate normal distributions, principal curves and spherical shells are used to describe amorphous pattern, curvilinear pattern and edge-ring shaped pattern respectively.

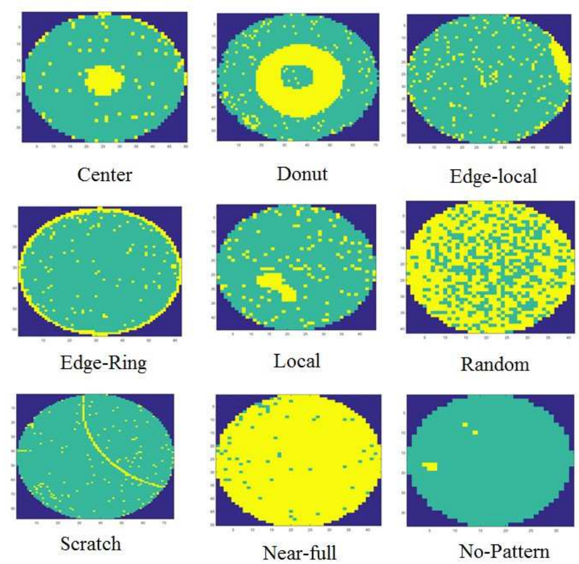

Fig. 1. Typical examples of wafer failure types

The region-based modeling methods mentioned above could only recognize three classes regular patterns and their methods just depends on synthetic wafer data. In [8], wafer map failure pattern recognition (WMFPR) proposed by Ming-Ju Wu combined support vector machines classifier and extracted radon-based features and geometry-based features to predict failure patterns. The benefit is that it could get similarity ranking for large-scale data sets but it not able to recognize more than one pattern on a wafer map. Chih-Hsuan use spectral clustering method to recognize the defect patterns [7] but they did not do a classification and determining the possible number of clusters is also a big problem.

By combining the Ordering Point to Identify the Cluster Structure(OPTICS) clustering method with SVM classifier, the proposed method is able to label the wafer map which might has not only one defect pattern and verified by real wafer maps' data with acceptable accuracy and high efficiency.

The rest of this paper is organized as follows. Section II describes the proposed method to recognition of defect patterns. Section III presents the simulation results of proposed method with performance comparison of accuracy and computation time with WMFPR. Section IV concludes the paper.

978-1-4673-9613-4/16/$31.00 ©2016 IEEE

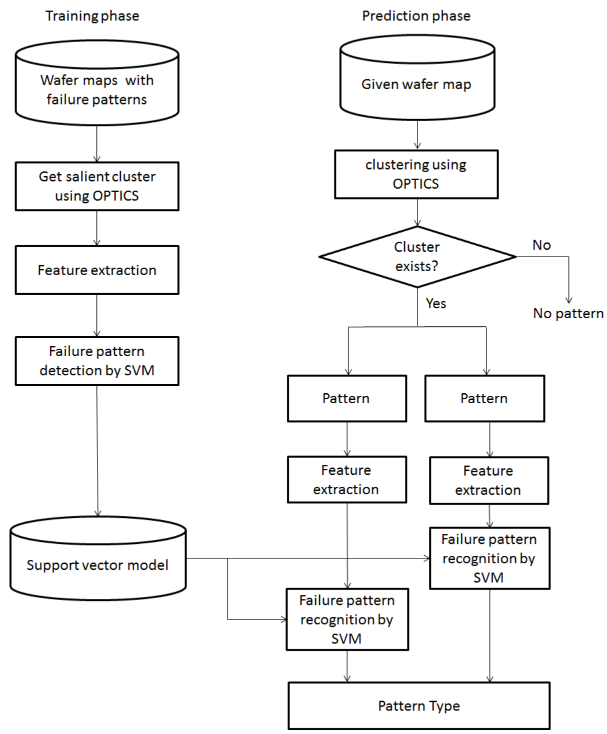

Fig. 2. Flowchart of the proposed method

II. Proposed Method

This paper proposed a new method to recognize the defect pattern on wafer maps. The wafer maps dataset are classified into 9 classes shown in Fig. 1 : Center, Donut, Edge-Local, Edge-Ring, Local, Near-full, Random, Scratch and No-pattern could cover almost all situations. As shown in the flowchart in Fig. 2, the proposed method is divided into 3 steps: clustering, feature extraction and pattern recognition. For the training set, we derive the salient cluster of wafer map by OPTICS clustering method to extract features as the training feature vector. On the other hand, for the testing set, first we also cluster the wafer map. A wafer would be categorized to the type No-Pattern if there is no cluster at all, and then SVM will be used to classify each cluster respectively. All the classification result of each cluster of a specific wafer will become a label of this wafer. Rest of this section will introduce the three steps as details.

A. Defect Clustering

OPTICS is a density based clustering method which is an improvement of Density-Based Spatial Clustering of Application with Noise(DBSCAN) [5]. It inherits the some advantages the same as DBSCAN that it can detect arbitrarily shaped clusters and does not need to predict the number of clusters. Moreover, it addressed DBSCAN's major weakness of being so sensitive to parameters that perform bad on data of varying density. The key idea of density-based clustering is that for each object of a cluster the neighborhood within a given radius (ϵ) contains at least a minimum number of objects (MinPts). The goal of OPTICS is to output an ordered set of data and two attributes for each element: core-distance and reachability-distance [1].

$$cd(x) = \begin{cases} Undefined & \text{, if } N_\epsilon(x) \leq MinPts \\ d(x, N_\epsilon^M(x)) & \text{, otherwise} \end{cases} \tag{1}$$

$$rd(x,y) = \begin{cases} Undefined & \text{, if } N_\epsilon(x) \leq MinPts \\ max\{cd(x), d(x,y)\} & \text{, otherwise} \end{cases} \tag{2}$$

(1) present the core-distance of x which is an object from a database X where ϵ is a distance and $N_\epsilon^M(x)$ is the ϵ-neiborhood of x. The core-distance of x is the minimum distance ϵ between x and an object in its ϵ-neighborhood such that x would be a core object with respect to ϵ if this neighbor is contained in $N_\epsilon(x)$. Otherwise, the core-distance is undefined. The reachability-distance (2) presents the minimum neighborhood radius where x is the core object and y is directly density-reachable from x.

After getting the order of core-distance and reachability-distance of data set, we can extract clusters easily from these data. In wafer pattern clustering, we set $MinPts$ equals 2 and ϵ equals 3. At first, we set all dies on wafer map as noise and then through all data by the order to determine to which cluster each elements belongs. If the reachability-distance of one object is smaller than ϵ then it belongs to current cluster. If it's reachability-distance is larger than ϵ and the core-distance is smaller than ϵ then it is recognized to be in a new cluster. Otherwise, this die is a noise. Fig.3 shows some examples using OPTICS.

(a)

(b)

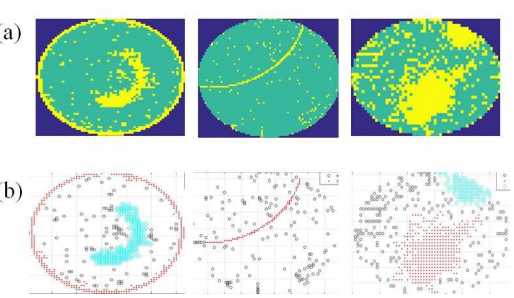

Fig. 3. Examples of some wafer maps using OPTICS

B. Feature Extraction

1) Density-based Features: The proposed density-based features extraction extraction process is displayed as in Fig. 4. We divide the wafer map into 13 parts and then compute fail density of each part. We can image that these eight class patterns will have their corresponding density distribution as follows:

Center: part 9 will have high fail density.

Donut: part 5 to part 13 but part 9 will have high fail density.

Edge-local: One or two of part 1 to part 4 will have high fail density.

Edge-ring: part 1 to part 4 will have high and similar fail density.

Local: servarl part of part 5 to part 9 will have high fail density.
Random: each part has similar fail density.
Near-full: each part density is almost 100%.
Scrach: random density distribution.

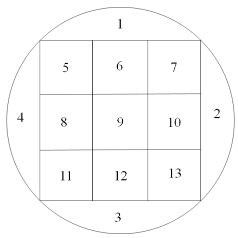

Fig. 4. Density based feature extraction

2) Geometry-Based Features: Geometry-based features are used to measure the geometric attributes of each pattern. Four representative attributes are chosen and detailed as follows:

- Convexity, indicating the the convexity of defect pattern. Computed as perimeter of estimated convex hull divid perimeter of defect pattern.
- Form factor, computed as area/perimeter.
- Solidity, indicating the proportion of defective dice in the estimated convex hull of cluster of defect die.
- Eccentricity, indicating the shape of the estimated eclipse which is the ratio of the distance between the foci of the ellipse and its major axis length.

C. Failure Pattern Recognition

SVM is a supervised learning model with associated learning algorithms that uesed to classify data. An SVM is designed to determine the hyperplane at which the margin between two classeds of data is maximized [3]. The set of labeled training patterns is considered to be linearly separable if there exists a vector w and a scalar b satisfies the constraints in (3)

$$\begin{cases} w^T x_i + b \geq 1 & , \forall y_i \in 1 \\ w^T x_i + b \leq -1 & , \forall y_i \in -1 \end{cases} \tag{3}$$

where x_i is a d-dimensional feature vector of wafer map, and y_i is the label of corresponding wafer map. Consider the constraint it too strict to training data be separated without error, let us introduce some non-negative variables ξ_i. Then the optimal hyperplane can be formulated as: minmize the function

$$\frac{1}{2}\left\|w\right\|^2 + C\sum_{i=1}^{n}\xi_i \quad s.t., y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, \ldots, n \tag{4}$$

where C is a predefined penalty factor. Then we need construct Lagrangian

$$L_p = \frac{1}{2}\left\|w\right\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left(y_i(w^T x_i + b) - 1\right) \tag{5}$$

where α_i represent the Lagrange multipiler. The corresponding Karush-Kuhn-Tucker conditions are

$$\begin{cases} \frac{\partial L}{\partial w} = 0 & \Rightarrow w = \sum_{i=1}^{n}\alpha_i y_i x_i \\ \\ \frac{\partial L}{\partial b} = 0 & \Rightarrow \sum_{i=1}^{n}\alpha_i y_i = 0 \end{cases} \tag{6}$$

So the hyperplane can be expressed as

$$f(x) = (\sum_{i=1}^{n}\alpha_i y_i x_i)^T x + b = \sum_{i=1}^{n}\alpha_i y_i\langle x_i, x\rangle + b \tag{7}$$

A mapping function can be applied to transform each x_i into a high dimensional space to make x_i linear separarity. According to the kernel trick, we denote K as the similarity function in transformed space, then

$$S(x_i, x_j) = \langle K(x_i), K(x_j)\rangle \tag{8}$$

where x_i and x_j are the feature vectors. Consequently, the optimal hyperplane also has a similar representation

$$f(x) = \sum_{i=1}^{n}\alpha_i y_i\langle K(x_i), K(x)\rangle + b = \sum_{i=1}^{n}\alpha_i y_i S(x_i, x) + b \tag{9}$$

Here, sigmod function kernel is adopted in (10)

$$S(x_i, x) = \tanh(gamma * x^T * x_i + coef0) \tag{10}$$

Function (9) and (10) can be used to predict the class of a wafer map by feature vecor x depending on value of f(x). To meet multi-class classificaiton, building a set of one-versus-one classifiers, and choosing the class that is selected by the most classifiers.

III. Performance evaluation and comparison

We use WM-811K dataset which has 811457 wafer maps collected from 46393 lots in real-world fabrication [8]. The dataset were divided into training set and test set by hold out. We get each wafer map's largest cluster of training set using OPTICS and extract both density-based features and geometry features as the training feature vector. For the training and classification, the well-know SVM toolbox LIBSVM [2] was adopted where sigmod function with the gamma(0.9) was used as the kenerl function.

Prediction \ Annotation	Center	Donut	Edge-Loc	Edge-Ring	Loc	Near-Full	Ran-dom	Scratch
Center	95.6%	5.3%	0.2%	0	4.3%	0	1.1%	2.8%
Donut	0	77.5%	0	0	1.7%	0	0	0
Edge-Loc	0.5%	0.9%	92.3%	0.8%	1.1%	0	6.6%	6.0%
Edge-Ring	0.3%	0	2.1%	98.4%	0	0	0.3%	8.0%
Loc	3.2%	12.8%	4.2%	0	81.8%	0	4.6%	6.8%
Near-Full	0	0	0	0	0	100%	0.7%	0
Random	0.2%	2.5%	0.7%	0	0.2%	0	85.8%	2.0%
Scratch	0.2%	0.9%	0.5%	0.6%	0.9%	0	0.9%	74.4%

Fig. 5. Salient Defect Pattern Recognition Rate

A. Salient Defect Pattern Recognition Rate

Dataset WM-811K just only have one label each wafer map, and the result of recognition rate of salient cluster shown in Fig. 5. In this figure, the annotations are shown in the top row, and the predictions by proposed method are in the left column. The diagonal elements present the recognition rate of each type. It can be observed that our method have a good performance for the wafer map classification especially on Center, Edge-Loc, Edge-Ring and Near-Full. And the Loc type compare with whose recognition rate only up to 68.5% in [8] is also acceptable.

B. Multi Defect Patterns Recognition Rate

We labeled 100 wafer maps which have two defect patterns and with another 200 wafer maps that only have one defect pattern as test set to validate our multi-label classification. Tabel I shows the result.

TABLE I
MULTI-DEFECT PATTERNS RECOGNITION RATE

	Rate
One pattern recognition accuracy	94.3%
Both patterns recognition accuracy	82.0%
One pattern is mistakenly recognized as two	27.5%
Two patterns is mistakenly recognized as one	6.0%

C. Computation Time

Our evaluation enviroment is a personal computer with an Intel Core i7 2.50GHz, 6GB RAM, and MATLAB R2015a. The computation time comparison between proposed method and WMFPR [8] shown in table II. In order to facilitate comparison, the test time of proposed method is the actual test time minus time of clustering and feature extraction beacuse WMFPR's test time does not contain the time of feature extraction. It shows that WMFPR spends more than three times of feature extraction than proposed method. Although have one more step, proposed method also faster than WMFPR.

TABLE II
COMPUTATION TIME OF THE PROPOSED METHOD COMPARE WITH WMFPR

	Proposed method (ms/waer)	WMFPR (ms/waer)
Clustering time	9.1	-
Feature extraction time	20.6	73.7
Training time	0.08	2.5
Test time	0.07	0.05

IV. CONCLUSION

Compared with the current approaches applied in wafer defect recognition, the proposed method not only have a higher accuracy up to 94.3% for classification of wafer maps but also is able to recognize more than one fail pattern on a single wafer map. In addition, the efficiency of proposed method is also validated. It just take 30.4ms to recognize the patterns per wafer. For the future research, more features recognition are targeted. For example, a center type pattern might have grid structure simultaneously.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant No.61176037

REFERENCES

[1] Mihael Ankerst, Markus M Breunig, Hans-Peter Kriegel, and Jörg Sander. Optics: ordering points to identify the clustering structure. In *ACM Sigmod Record*, volume 28, pages 49–60. ACM, 1999.

[2] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[3] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[4] Andrew Drozda-Freeman, Mike McIntyre, Mike Retersdorf, Christopher Wooten, Xin Song, and Adam Hesse. The application and use of an automated spatial pattern recognition (spr) system in the identification and solving of yield issues in semiconductor manufacturing. In *2007 IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, pages 302–305. IEEE, 2007.

[5] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[6] Jung Yoon Hwang and Way Kuo. Model-based clustering for integrated circuit yield enhancement. *European Journal of Operational Research*, 178(1):143–153, 2007.

[7] Chih-Hsuan Wang. Recognition of semiconductor defect patterns using spectral clustering. In *2007 IEEE International Conference on Industrial Engineering and Engineering Management*, pages 587–591. IEEE, 2007.

[8] M. J. Wu, J. S. R. Jang, and J. L. Chen. Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Transactions on Semiconductor Manufacturing*, 28(1):1–12, Feb 2015.

[9] T. Yuan, W. Kuo, and S. J. Bae. Detection of spatial defect patterns generated in semiconductor fabrication processes. *IEEE Transactions on Semiconductor Manufacturing*, 24(3):392–403, Aug 2011.

[10] Tao Yuan, Suk Joo Bae, and Jong In Park. Bayesian spatial defect pattern recognition in semiconductor fabrication using support vector clustering. *The International Journal of Advanced Manufacturing Technology*, 51(5):671–683, 2010.

[11] Tao Yuan and Way Kuo. A model-based clustering approach to the recognition of the spatial defect patterns produced during semiconductor fabrication. *IIE Transactions*, 40(2):93–101, 2007.