Available online at www.sciencedirect.com

ScienceDirect

Int. J. Production Economics 107 (2007) 88–103

international journal of production economics

www.elsevier.com/locate/ijpe

Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing

Shao-Chung Hsu, Chen-Fu Chien*

Department of Industrial Engineering & Engineering Management, National Tsing Hua University, Hsinchu 30043, Taiwan, ROC

Received 16 September 2005; accepted 25 May 2006
Available online 4 December 2006

Abstract

Semiconductor manufacturing involves lengthy and complex processes, and hence is capital intensive. Companies compete with each other by continuously employing new technologies, increasing yield, and reducing costs. Yield improvement is increasingly important as advanced fabrication technologies are complicated and interrelated. In particular, wafer bin maps (WBM) that present specific failure patterns provide crucial information to track the process problems in semiconductor manufacturing, yet most fabrication facility (fabs) rely on experienced engineers' judgments of the map patterns through eye-ball analysis. Thus, existing studies are subjective, time consuming, and are also restricted by the capability of human recognition. This study proposes a hybrid data mining approach that integrates spatial statistics and adaptive resonance theory neural networks to quickly extract patterns from WBM and associate with manufacturing defects. An empirical study of WBM clustering was conducted in a fab for validation. The results showed practical viability of the proposed approach and now an expert system embedded with the developed algorithm has been implemented in a fab in Taiwan. This study concludes with a discussion on further research.

Keywords: Data mining; Yield improvement; Wafer bin map; Spatial randomness test; ART1; Neural networks; Semiconductor manufacturing; Quality engineering

1. Introduction

The semiconductor manufacturing processes are lengthy and complex. Thus, the capital investments in the semiconductor industry are huge. The manufacturing usually contains 100–200 process steps, in which the wafers move from step to step in groups of 25 or 24 identical wafers in a fabrication facility (fab). After wafer fabrication, circuit probe (CP) testing is performed on each die on the wafer. Then, the wafers are died up, and the good dies are packaged into chips and shipped to the customer after final testing (FT).

The critical factors maintaining competitive advantages for semiconductor wafer fabs include lowering die costs via lean production and increasing yield via quick response to yield excursions (Leachman and Hodges, 1996; Peng and Chien, 2003). In particular, the defect problems should be detected in time and the assignable causes should then be resolved to reduce the loss of hundreds of thousands of dollars of scraped wafers as soon as

*Corresponding author. Tel.: +886 3 5742648; fax: +886 3 5722685.
E-mail address: cfchien@mx.nthu.edu.tw (C.-F. Chien).

0925-5273/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpe.2006.05.015

possible. Four yield definitions are used in semiconductor manufacturing: CP test yield (CP yield, Y_{CP}), fabrication line yield (Y_L), assembly yield (Y_{AS}) and final test yield (Y_{FT}). Among them, the CP yield is the most critical (Cunningham et al., 1995). CP yield improvement is divided into two major categories: (1) based line yield improvement, and (2) low yield trouble shooting. The based line yield improvement is based on tuning the process recipes to improve device performance and reduce defects, while low yield trouble shooting involves monitoring and diagnosing the failures caused by abnormal events such as mis-operation, trouble tool and contamination. Wafer bin map (WBM) is the result of CP inspection of dies on the wafer at the end of fabrication. WBM patterns can provide information to monitor the process and product.

This study aims to develop a hybrid data mining approach that integrates spatial statistics and adaptive resonance theory (ART) neural networks to rapidly extract patterns from WBM and associate them with manufacturing defects. An empirical study of WBM clustering was conducted in a fab for validation. Mining large amounts of data can help the engineers make the right decision of classifying patterns. During the manufacturing process, data are collected for various purposes. In particular, Wafer In Process (WIP) data refers to the data collected while processing wafers; metrology data refers to the data collected from in-line inspections; electrical test (E-test) data refers to data collected to measure device performance in chips, and the CP test data records each chip' functional test result after wafer fabrication. Since modern fabs are equipped with the computer integrated manufacturing (CIM) system, data collection is no longer a major issue. Furthermore, an engineering data analysis (EDA) system that is an off-line analysis-oriented system with data warehouse is generally developed to support data analysis activities (Peng and Chien, 2003; Chien et al., 2007). A remaining issue is to sieve out relevant data from a massive pool to derived useful information that can assist engineers in timely trouble shooting and yield enhancement.

WBMs are multi-dimensional and have complex structures, can provide essential information for engineers to identify problems in the manufacturing process. Fig. 1 shows a typical WBM where the different symbols denote chips failing in different functional tests. To assist visualization and analysis, WBM is usually transformed into a binary map that represents it using binary code or two different colors. This study uses yellow squares or "1" to denote defective chips, and red squares or "0" to denote functional chips.

The failure patterns of WBM can be classified into three major categories (Taam and Hamada, 1993; Stapper, 2000):

(1) Random defect: No spatial clustering or pattern exists, and the defective chips are randomly distributed in the two-dimensional map. Random defects are usually caused by the manufacturing environmental factors. Even in a in a near-sterile environment, particles cannot be removed completely. However, reducing the level of random defects can improve the overall productivity of wafer fabrication.
(2) Systematic defect: The positions of defective chips in the wafer show the spatial correlation, for example, ring, edge-fail, checkerboard. Fig. 2 shows 10 systematic patterns that are frequently seen in fab, as defined by domain experts.

a

```
        w d %* * * * *
      %%* * S a d d %* S
    m%w%mmm%%%G%* %
  * ww* %o %S * * * * %S u S S
%%* * S %%%* * * * I * * * %* *
* S * %%%* %%%%* * * d * %* *
* * %%* * * c * * * d * %%* * * *
* * * %%* * * %* %%* * G* s * *
    %%* * %* * f * * %* * %* *
      * * %* * * * %* * * * %*
        * * * d %* * * * %*
```

b

```
        * * %* * * * *
      %%* * * * * * %* *
    * %* %* * * %%%* %* %
  * * * * %* %* * * * * %* * * *
% %%* * * %%%* * * * * * * * %* * %
*  * * * %%%* %%%%* * * * * %* * *
%  * * %%* * * * * * * * * %%* * * * %
  * * * %%* * * %* %%* * * * * * *
    %%* * %* * * * * %* * %* *
      * * %* * * * %* * * * %*
        * * * * %* * * * %*
```

Fig. 1. Example of WBM: (a) WBM with each failure bin denoted by a different symbol; (b) WBM with specific bin denoted by the symbol "%".

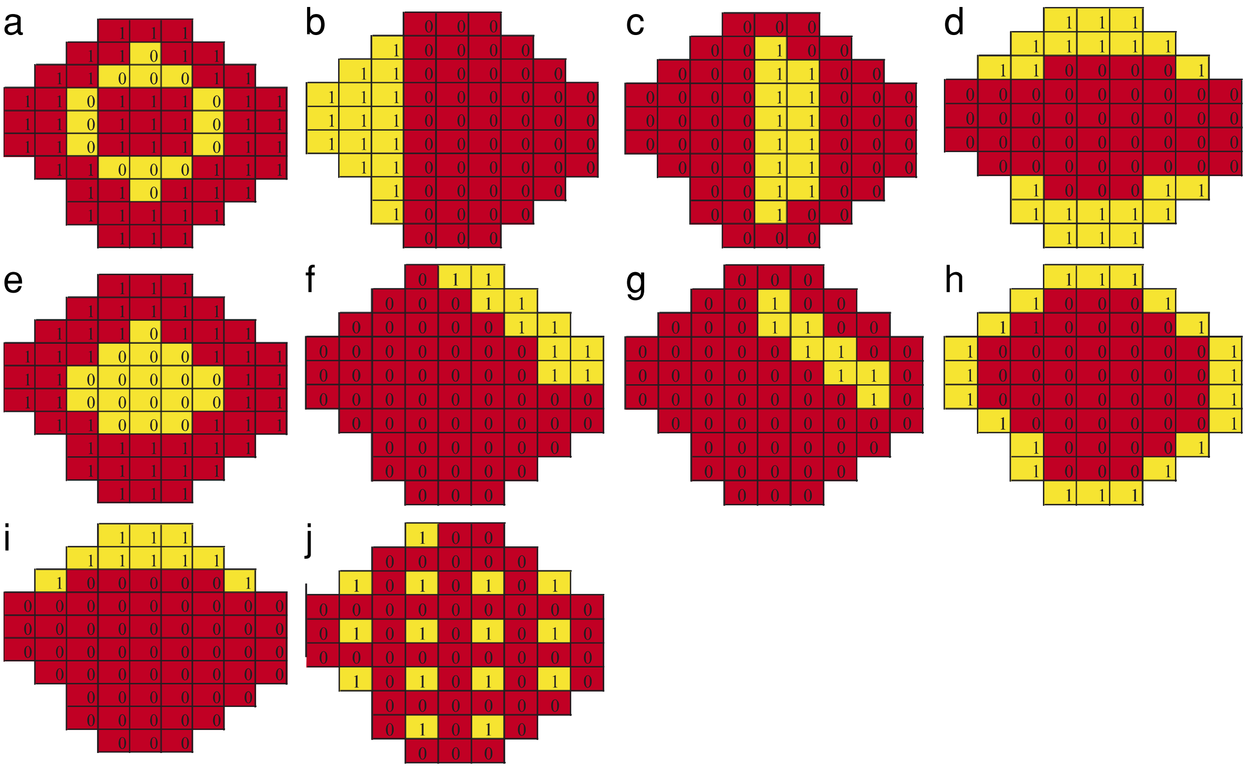

Fig. 2. Systematic defect patterns in semiconductor manufacturing.

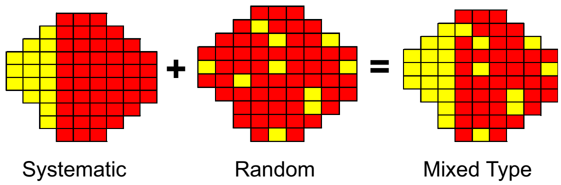

Fig. 3. Mixed defect map.

(3) Mixed defect: Consisting of a random defect and a systematic defect in one map. Most wafer maps are of this type, as shown in Fig. 3. Engineer needs to separate random and systematic defects in the WBM, since the systematic defect's signature can reveal the process problem (Friedman et al., 1997).

The rest of this study is organized as follows: Section 2 formulates the WBM clustering problems in yield improvement and reviews related studies. Section 3 outlines the proposed approaches. Section 4 details a research framework for pattern extraction from WBM. Section 5 describes an empirical study for validation. Finally, Section 6 concludes this study with discussing the results and the direction of future research.

2. Problem formulation

This study addresses the problem of finding patterns from WBMs to monitor semiconductor manufacturing process and provide pattern information for trouble shooting. Engineers have to integrate and analyze large amounts of data during analysis. Data mining can efficiently find hidden yet potentially valuable information in terms of specific patterns from massive data (Keki et al., 1993; Kusiak, 2000; Peng et al., 2004). Several studies have applied data mining techniques to solve manufacturing and yield problems. For example, the decision tree method can be used to identify the faulty equipment and the corresponding dates of failure (Bergeret and Gall, 2003; Braha and Shmilovici, 2002, 2003; Chien et al., 2007). Self-Organization Map (SOM) clustering has been applied to cluster E-test, CP fail bins and metrology data to detect the failure patterns (Chien et al., 2003). In addition, Braha and Shmilovici (2003) apply three classification-based data mining methods (decision tree induction, neural networks, and composite classifiers) to refine dry-cleaning technology for process improvement.

However, few studies have been conducted to develop WBM clustering methods, though WBM spatial patterns contain useful information about potential manufacturing problems. For example, mask misalignment in the lithographic process generates a checkerboard pattern; the abnormal temperature control in the rapid thermal annealing process (RTP) can generate a ring of failing chips around the edge of the wafer. By reviewing WBM patterns, experienced engineers can quickly clarify the root causes and identify the assignable causes. Most existing studies focus on diagnosing systematic defects or patterns in wafer maps (Mallory et al., 1983; Hansen and Nair, 1995; Kaempf, 1995; Friedman et al., 1997), but give no information about pattern types.

The lack of adequate tools has also led engineers to only use data relating to simple measures such as overall yield to track process problems. Experienced engineers review the WBM to possibly identify spatial patterns with only visual analysis based on printed maps. Furthermore, most effort in existing analysis is devoted to data integration from testers and manufacturing process. The existing practice based on human judgments causes low analysis efficiency and the results varied among the experts since they had different mental models.

3. Fundamental

3.1. Spatial randomness test

The spatial correlation of two groups of data can be tested by the odd ratio hypothesis test (Agresti, 1990; Taam and Hamada, 1993). The revised estimator is as Eq. (1).

$$\hat{\theta} = \frac{(N_{\mathrm{GG}}+0.5)(N_{\mathrm{BB}}+0.5)}{(N_{\mathrm{GB}}+0.5)(N_{\mathrm{BG}}+0.5)}. \tag{1}$$

In Eq. (1), N_{GG}, N_{BB}, N_{BG} and N_{GB} are defined as the counts of relations of adjacent chips in the two-way contingency table, as described in Table 1. The relation is defined by the King–Move neighborhood in two-dimensional space as described in Fig. 4.

Let Y_i represents the CP test result of chip in the position i of map. If $Y_i = 1$, then the chip in position i fails in the CP test; conversely, $Y_i = 0$ means the chip passes the CP test. Then, N_{GG}, N_{BB}, N_{BG} and N_{GB} can be calculated as follows:

Table 1
Two-way contingency table of adjacent chips

Position i	Position j	
	Good	Bad
Good	N_{GG}	N_{GB}
Bad	N_{BG}	N_{BB}

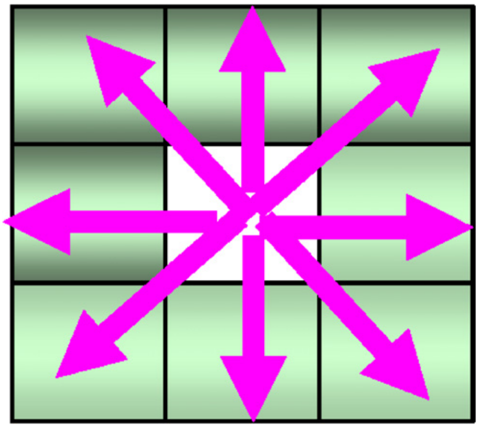

Fig. 4. King–Move neighborhood.

$$N_{\mathrm{GG}} = \sum\sum_{i<j} \delta_{ij}(1-Y_i)(1-Y_j), \tag{2}$$

$$N_{\mathrm{GB}} = \sum\sum_{i<j} \delta_{ij}(1-Y_i)Y_j, \tag{3}$$

$$N_{\mathrm{BG}} = \sum\sum_{i<j} \delta_{ij}Y_i(1-Y_j), \tag{4}$$

$$N_{\mathrm{BB}} = \sum\sum_{i<j} \delta_{ij}Y_iY_j, \tag{5}$$

where

$$\delta_{ij}\begin{cases} 1 & \text{if } Y_i \text{ and } Y_j \text{ are in King} - \text{Move neightbor}, \\ 0 & \text{otherwise}, \end{cases}$$

The steps of spatial randomness testing are described as follows:

Step 1: Establish hypotheses:

Two alternative hypotheses are built for the spatial randomness test.

- H0: The distribution of failure chips or good chips in the wafer is spatially random.

- H1: The distribution is not spatially random and the map exhibits special clusters of fail chips or repeat patterns.

Step 2: Select the test statistics:

When the sample size is large, the distribution of estimator, $\log\hat{\theta}$ is asymptotic to normal distribution with the parameters (μ, σ),where $\mu = 0$, and $\sigma = \left(\frac{1}{N_{GG}+0.5} + \frac{1}{N_{BB}+0.5} + \frac{1}{N_{GB}+0.5} + \frac{1}{N_{BG}+0.5}\right)^{1/2}$

Step 3: Test WBM with following rules:

- Rule 1: If $\log\hat{\theta} \approx 0$, then the map is spatially random.
- Rule 2: If $\log\hat{\theta} \gg 0$, then the map shows special clusters.
- Rule 3: If $\log\hat{\theta} \ll 0$, then the map shows a repeating pattern.

3.2. ART for map clustering

The ART (Carpenter and Grossberg, 1988; Freeman and Skapura, 1991) has been applied in many areas including pattern recognition and spatial analysis. ART derives fundamentally from the adaptive resonant feedback between two layers of neurons. To manage the variety input, ART has the following characteristics: (1) balance on stability and plasticity, (2) match and reset, and (3) balance on search and direct access. ART solves the stability–plasticity dilemma, which is caused by learning new data leading to unstable conditions and loss of data. Several algorithms are derived from the original ART, including ART1 (Carpenter and Grossberg, 1988), ART2 (Carpenter and Grossberg, 1987), ART3 (Carpenter and Grossberg, 1990), ARTMAP (Carpenter et al., 1991), and Fuzzy ART (Carpenter et al., 1991).

This study employs the ART1 algorithm for WBM clustering, since the input data form a binary map.

3.3. Decision tree

Decision tree can be used to extract models to describe important data classes or to predict future data trends. A decision tree is a flow-chart-like tree structure where the root at the top and the leaves at the bottom. Through serial tests on the attributes of the root node containing the entire dataset, the branches representing the outcomes of the tests and the leaves indicating the classes are thus constructed. Each path from the root node to a leaf can be interpreted as a rule.

Decision tree construction can be separated to three basic elements: first, growing the tree; second, pruning the tree; third, extracting rules from the tree. Several algorithms have been developed to construct decision tree model. Chi-squared automatic interaction detection (CHAID) is a non-binary decision tree that determines the best multi-way partitions of the data on the basis of significance tests (Kass, 1980). CHAID is designed specifically to deal with categorical variables. classification and regression tree (CART) is a binary decision tree with the Gini-index of diversity as the splitting criterion, and pruning by minimizing the true misclassification error estimate (Breiman et al., 1984). CART can deal with categorical and continuous variable. C4.5 is a variant and extension of a well-known decision tree algorithm, ID3 (Quinlan, 1993). The splitting criterion of C4.5 algorithm is gain ratio that expresses the proportion on information generated by a split. The error-based pruning is used to C4.5 for pruning. Lim et al. (2000) compared the prediction accuracy, complexity and training time of 33 classification algorithms, and indicated that decision tree provides good accuracy and data interpretation.

4. The research framework

This study proposes a framework integrated with spatial statistics and ART1 network with domain knowledge to improve the efficiency of WBM clustering. Then, the extracted spatial patterns will be correlated with process data by applying decision tree to identify the root causes. Fig. 5 illustrates the research framework.

4.1. Data pre-processing

Before clustering, the WBM data are pre-processed in three stages: data integration, data cleaning and data transformation.

(1) Data integration: Engineers usually use a specific bin or combined multi-bins to query WBM data. Two options are available for integrating the WBM data: Analysis by wafer where each map represents a wafer, the other is analysis by lot where the map is generated from all wafers within the lot. Each map represents the integrated

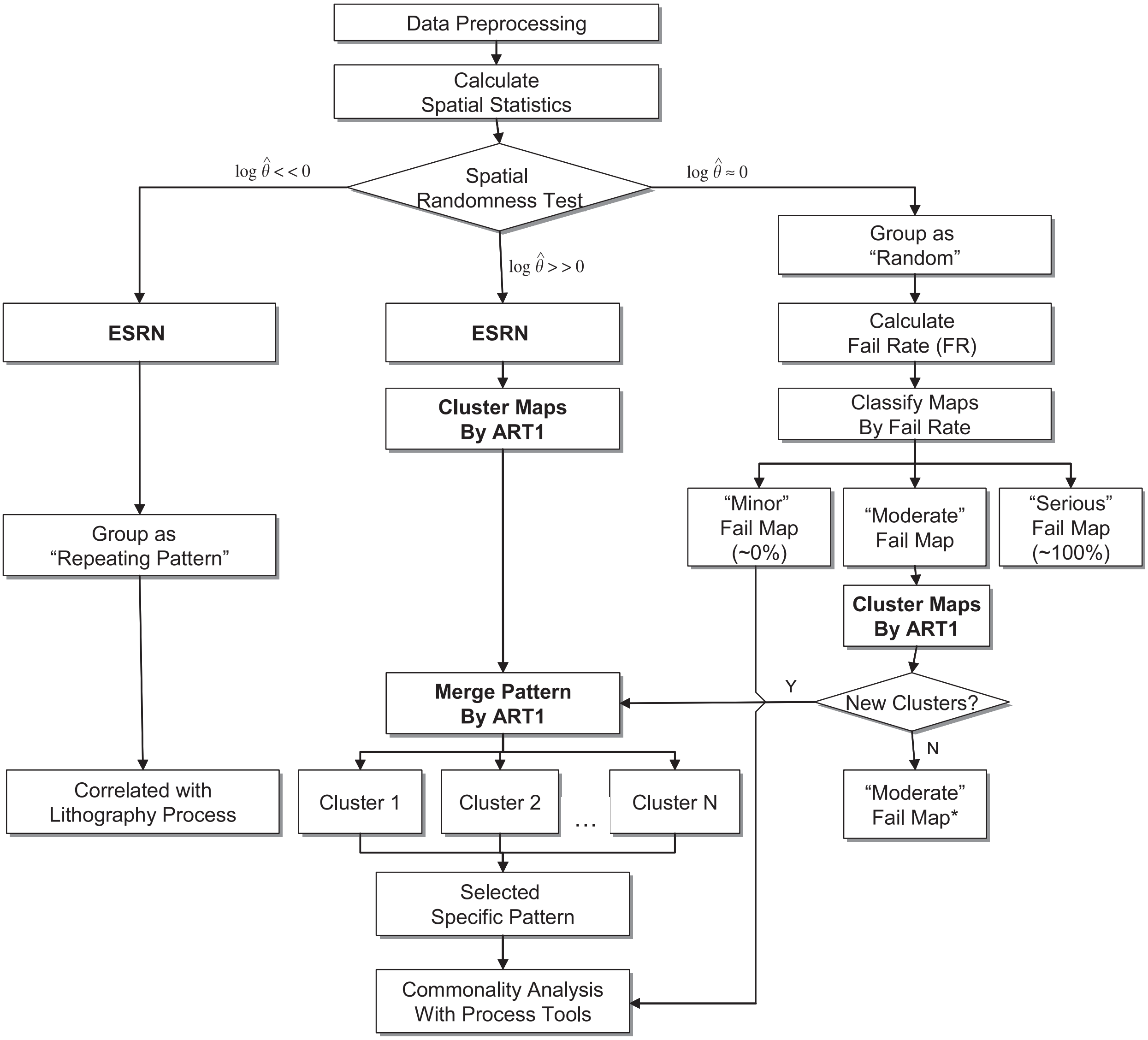

Fig. 5. Research framework for WBM clustering.

spatial result of a lot. In practice, engineers will try to correlate the spatial patterns with process data, like tools, date and operator, yet without systematic methodologies to support effective analyses. The data resolution of process data is usually with lot-based in 200 mm wafer fabs. Thus, the WBM data has to integrate from wafers into lot-based map for data preparation to support analysis with the lot-based process data. In this study, we applied a threshold value to translate the accumulated percentage by wafers into binary data to represent the lot-based WBM.

(2) Data cleaning: Missing data often arise in the WBM because the probe might not work correctly in some chips during the CP test. This study deletes missing data. Therefore, if a position of has no test data in one map, then the data in this position in other maps are not included in the calculation.

(3) Data transformation: The data are transformed into several formats for different purposes. The proposed framework defines two formats to be transformed:
   (a) Binary map: Based on the selected bin, BIN_i, the "binary" map is created using 1 to denote the defective chips, and 0 to denote passing chips. The binary map is used for to visualize the chip, and to test for spatial randomness.

(b) Binary vector: Each WBM is converted from the two-dimensional map into an one-dimensional vector using the index map. The index map is defined by coding the position of each chip in the map with a sequence sorted from left to right and top to bottom. Then, each map can be converted into a vector in terms of this sequence. Fig. 6 illustrates this transformation process. The binary vector is applied to generate the ART1 network.

4.2. Spatial randomness testing

Each map is tested for spatial randomness testing and classified into three types: "Checkerboard Defect", "Clustered Defect" and "Random Defect". The "Checkerboard Defect" (or "Repeating Defect") group can be further analyzed by being directly correlated with the lithographic process data. The "Clustered Defect" maps are clustered using ART1. The "Random Defect" maps are further classified by the WBM fail rate that represents the degree of trouble in manufacturing. "Random Defect" maps are classified into three sub-groups: "Minor Fail", "Moderate Fail" and "Serious Fail". The maps in the "Moderate Fail" group are also clustered by ART1 to ensure that no information has been lost in pattern extraction, since the spatial randomness testing only tests the spatial independence between "Good" and "Bad" chips, not the pattern itself. Additionally, the definition of 'neighborhood' affects the test results.

4.3. Enhance the signal and remove the noise

To obtain good clustering results, we developed a data preparation procedure to enhance the signal and remove the noise (ESRN) as follows:

First, the King–Move neighborhood is applied to define the weighs of the adjacent chips. The positions on the four nearest neighbors (left, right, top and down) are weighted 1. The corner positions are weighted 0.5.

Second, two rules are applied to each chip in the map. For spatial randomness testing, Y_i is defined as the CP test result of the chip in the position i of the map. If $Y_i = 1$, then the chip in position i fails the CP test; while $Y_i = 0$ indicates that the chip passes the CP test.

For a given position i, let

$$N_{i+} = \sum_{j=1}^{n} \delta_{ij}(1 - Y_i)Y_j, \tag{6}$$

$$N_{i-} = \sum_{j=1}^{n} \delta_{ij} Y_i(1 - Y_j), \tag{7}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } Y_i \text{ and } Y_j \text{ are in King} - \text{Move neightbor}, \\ 0 & \text{otherwise.} \end{cases}$$

Rule 1: If $Y_i = 0$ and $N_{i+} > \rho_1$, then $Y_i = 1$ (enhance signal).

Rule 2: If $Y_i = 1$ and $N_{i-} > \rho_2$, then $Y_i = 0$ (remove noise),

where ρ_1 and ρ_2 represent the threshold value of enhancing signal and removing noise.

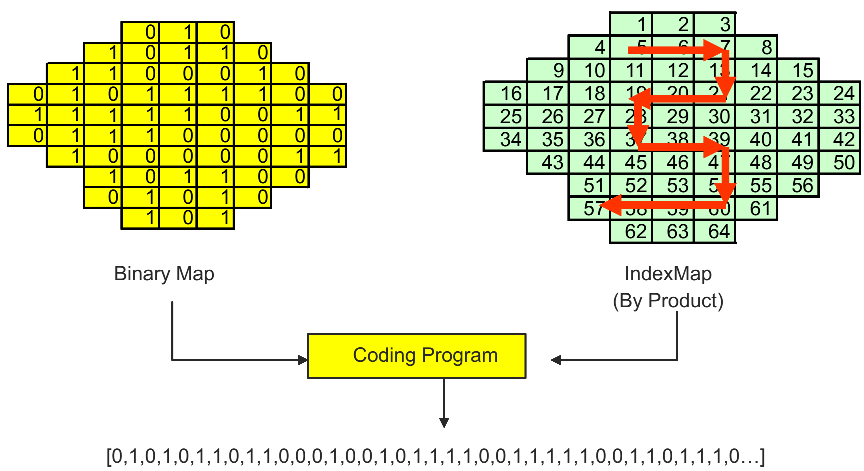

Fig. 6. Transfer map data (2D) format into vector (1D).

Fig. 7 illustrates two rules applied to chips in a King–Move neighborhood. An experiment with different ρ_1 and ρ_2 setting was performed to identify which types of signals are enhanced and which are degraded. We select three frequent-seen patterns and one composite pattern for ESRN test. The setting of ρ_1 changes from 4 to 5, while ρ_2 changes from 5 to 6. Fig. 8 displays the experiment result of ESRN. As shown in the results, some random defective chips have been removed from original maps of the "Right-Down Edge" pattern and "Composite" pattern and the rule 2 has enhanced the pattern in the maps of these patterns. However, the pattern with "Checkerboard" defect that is caused by mask defects is totally destroyed and the "Ring" defect is affected by the setting of threshold value. These results are due to the King–Cross neighbor can handle the blocked pattern more effectively. To solve this issue, we apply the spatial randomness testing before ESRN to separate the "Checkerboard" defect from others in our proposed framework. For the "Ring" defect, we set the

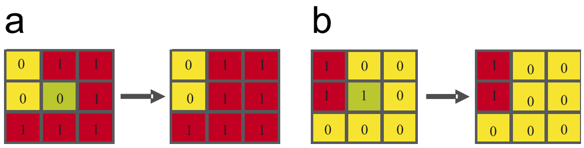

Fig. 7. ESRN graph of a chip: (a) Enhance signals; (b) Remove the noise.

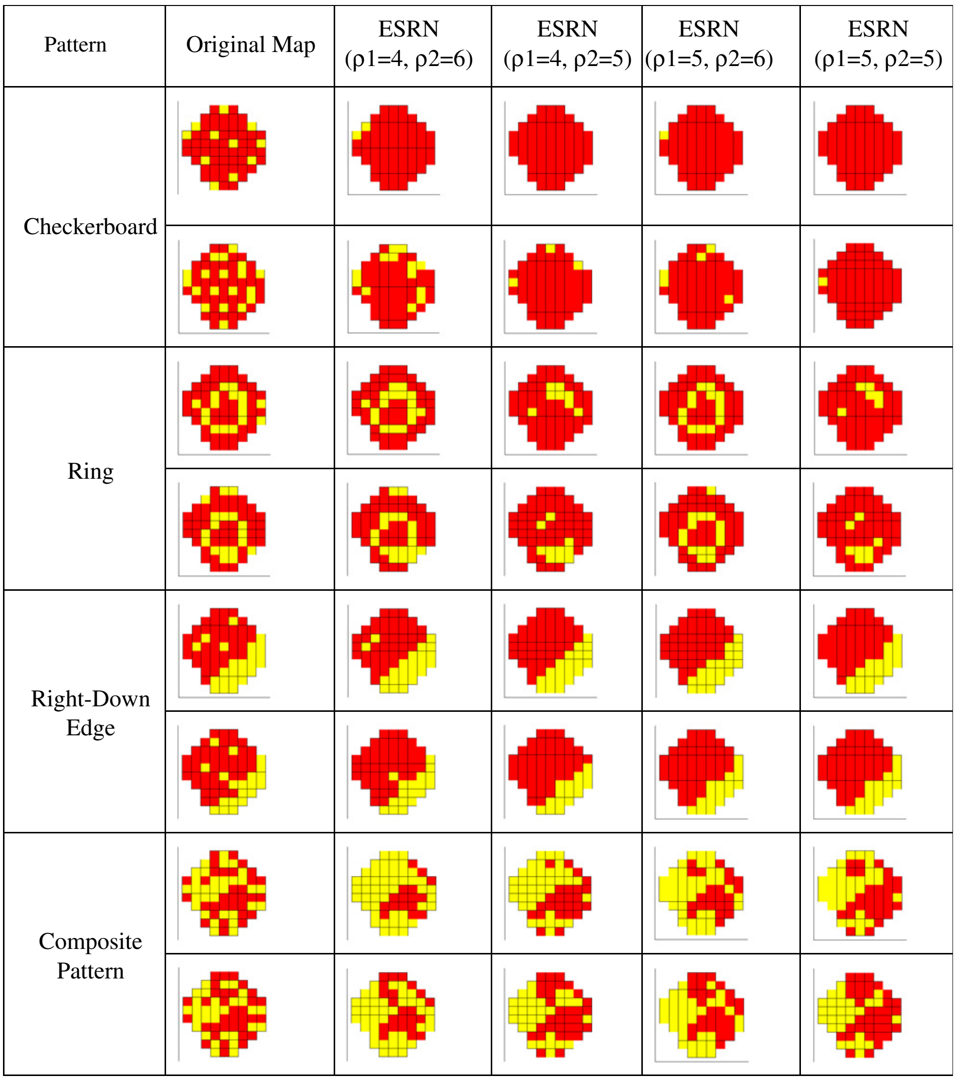

Fig. 8. Comparison of maps before and after ESRN with different threshold settings.

threshold value of ρ_1 and ρ_2 to 4 and 6, respectively, in our algorithm, which are better setting in the experiment.

4.4. Map clustering

The objective of map clustering is to maximize the spatial similarity within clusters and minimize the number of clusters. This study applies the ART1 method to cluster the similar maps. The other objective is to extract the common pattern from each cluster. The common pattern is defined as the intersection of maps within a cluster. With this information, engineers do not need to examine the patterns map by map, and can correlate the patterns with the specific process problem mentioned in Section 2. Two groups of maps are clustered by ART1 after spatial randomness testing: the "Clustered Defect" group and the "Moderate Fail" sub-group of "Random Fail".

4.5. Merger pattern

The patterns that are generated from the "Clustered Defect" group and the "Moderate Fail" group are collected and clustered using ART1 to merge the patterns. This step serves two purposes: (1) to minimize the number of clusters as few as possible since many clusters make patterns hard to interpret, and (2) to integrate the patterns, since some maps in the "Random Fail" group may still have patterns of spatial cluster, which may be similar to the "Clustered Defect" patterns.

4.6. Interpret results

Engineers can select patterns to analyze further. For example, they can conduct correlation analysis to correlate the special patterns with the process information. If the result is unacceptable, engineers can return to the steps of Map Clustering or Merge Pattern to adjust the parameters in the framework to generate a new result.

4.7. Root cause identification

The step of root cause identification is employed to identify the suspected process tools or periods that cause the fault or abnormal pattern of selected component. Classification techniques, i.e., decision tree, are applied to clarify the relations of the selected component to clarify the relations and related process information systematically. The first step is to integrate these data together and transfer into a new variable with two categories, "With Pattern" and "Without Pattern". This new variable is the target variable of decision tree. Then, the related process data including tools, date, and operator are used as the input for decision tree analysis to identify the suspected process steps for further investigations. Finally, engineer can make judgments on the suspected causes of extracted clustered WBMs.

4.8. A numerical example

To demonstrate the effectiveness of the proposed framework, a case of 25 maps was generated from seven groups with special patterns plus random defects (see Fig. 9 for some of the patterns). Each map contained 63 chips and the fail rate varied from 0% to 100%.

(1) Spatial Randomness Testing: Table 2 summarizes the spatial randomness results. Three maps were placed in the "Clustered Defect" group with the criterion of p-value <0.00005 (maps 8, 13 and 16). Three maps with negative log odd ratio and p-value <0.05 were classified as the "Checkerboard Defect" (maps 2, 7 and 18). Others were classified as "Random Defect", and were further classified into three sub-groups by the fail rate. The "Minor Fail" sub-group comprised the maps with fail rate $\leqq 5\%$ (maps 4, 12 and 22), while the "Serious Fail" sub-group had fail rate $\geqq 95\%$ (map 6, 17 and 21). The remaining maps (total 13 maps) were placed into the "Moderate Fail" group. Further examination of the maps in the "Moderate Fail" group indicated 11 maps with two special patterns, "Ring" (maps 3, 5,10,11 and 14) and "Composite"(maps 9,15, 20, 24 and 25). The ART1 clustering was applied to the "Moderate Fail" group later.

(2) Map Clustering: With the vigilance threshold, one cluster was extracted from the "Clustered Defect" group, and two were extracted from the "Moderate Fail" group. After merging the patterns, the three groups had not changed. Fig. 10 summarizes the final results. Comparing Figs. 9 and 10, shows that the maps with the special patterns defined in Fig. 10 are in the same group after clustering. For example, the maps with "Right-Down Edge" pattern are

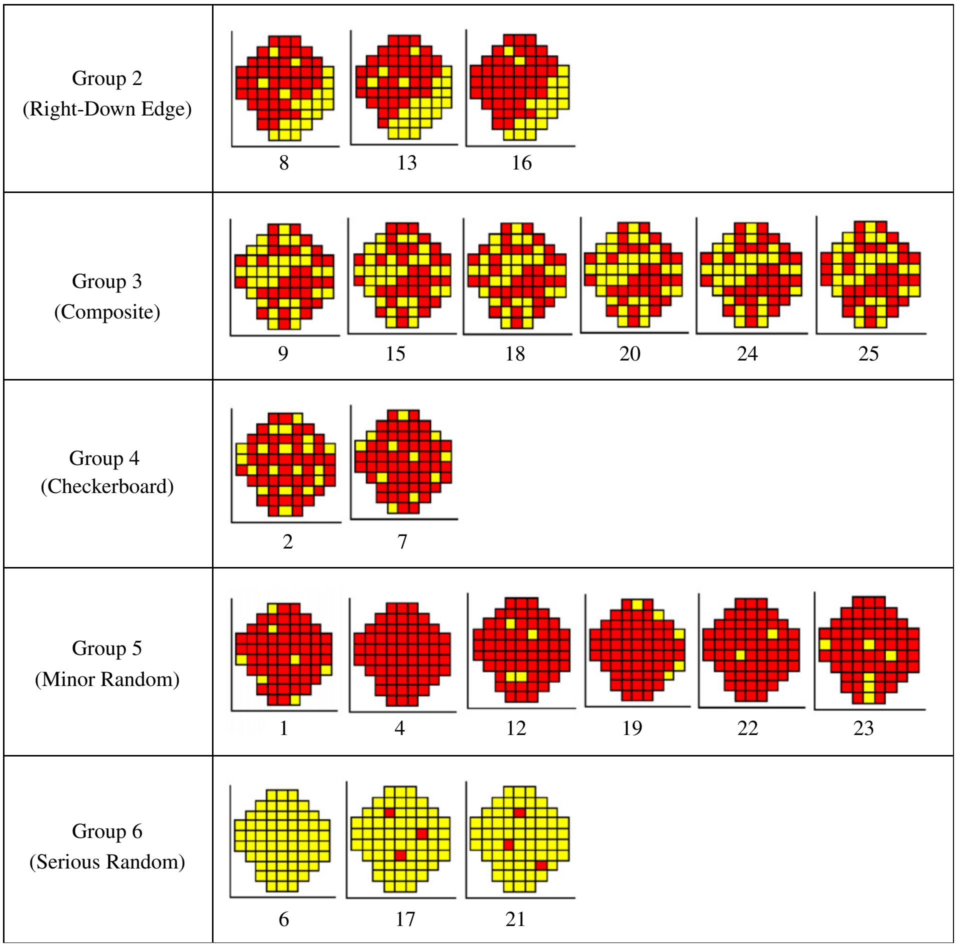

Fig. 9. Groups and patterns used to generate maps.

clustered in the C1 group before and after clustering. The only mis-classified map is the 18th map, which belongs to the "Composite" group but is classified into the "Checkerboard" cluster due to this map's overall mis-classification rate is 4% (1/25). Therefore, the clustering results using the proposed framework are fairly consistent with the true groups.

5. Empirical study

For validation, an empirical study was conducted to verify the effectiveness of the proposed framework over conventional methods in a wafer fab. The study used a product with high yield variation in CP testing over a period of 2 months. A total of 138 lots completed the CP test during this period.

5.1. Data pre-processing

The raw data were selected from database and transformed into 138 maps, each of which represented an integrated result of lot accumulated by wafers within the lot. The original map consisted of 301 chips, but 33 chips had at least one map with missing data. Excluding these chips, the location of 268 chips was used for clustering and spatial randomness testing.

5.2. Spatial randomness testing

After testing, 59 maps were found to have spatial correlation under significant level $\alpha = 0.00005$, and were placed in the "Clustered Defect" group, while the other 79 maps were placed in the "Random

Table 2
Summary of spatial randomness testing of 25 maps

Id	N_{BB}	N_{GB}	N_{BG}	N_{GG}	Log OR	Std.	p-value	Fail_rate (%)
1	0	20	18	175	−1.464	1.452	0.15673	11
2	7	56	57	93	−1.533	0.423	0.00015	19
3	18	48	44	103	−0.120	0.327	0.35707	27
4	213	0	0	0	6.057	2.450	0.00672	0
5	21	42	45	105	0.160	0.319	0.30826	28
6	0	0	0	213	6.057	2.450	0.00672	100
7	1	30	33	149	−1.517	0.858	0.03856	17
8	33	40	19	121	1.640	0.338	0.00000	31
9	52	56	56	49	−0.206	0.273	0.22577	50
10	13	35	34	131	0.371	0.373	0.15957	18
11	23	44	47	99	0.101	0.310	0.37238	28
12	1	15	15	182	0.131	0.895	0.44203	6
13	51	38	16	108	2.174	0.339	0.00000	37
14	19	41	41	112	0.242	0.329	0.23114	25
15	46	57	54	56	−0.176	0.274	0.25980	47
16	37	32	11	133	2.595	0.390	0.00000	30
17	189	12	12	0	−0.500	1.471	0.36695	95
18	38	61	59	55	−0.538	0.278	0.02631	45
19	1	11	10	191	0.867	0.924	0.17419	8
20	52	56	56	49	−0.206	0.273	0.22577	50
21	190	11	12	0	−0.412	1.474	0.39003	95
22	0	8	8	197	0.312	1.497	0.41732	3
23	2	20	18	173	0.134	0.713	0.42533	9
24	45	52	52	64	0.063	0.275	0.40972	45
25	40	59	58	56	−0.419	0.276	0.06443	45

Defect" group. The "Random Defect" group was divided into three sub-groups using the definition in Section 6. The "Minor Defect" group contained three maps with fail rate ≦5%, while the "Serious Defect" contained eight maps with fail rate ≧95%. The other maps (68 in total) were placed in the "Moderate Defect" group.

5.3. Map clustering

A total of 29 patterns were generated from the "Clustered Defect" group with vigilance threshold. The same algorithm of ART1 clustering was also applied to the "Moderate Defect" group. A total of 6 patterns (20 maps) were extracted from the "Moderate Defect" group. After merging the patterns, 21 patterns were finally obtained from 79 maps. After excluding the patterns with single map, there are 14 patterns with more than one map within cluster. Fig. 11 presents the patterns with more than one map within cluster.

5.4. Result and discussion

In summary, the WBM patterns from 138 maps were extracted and organized in several groups by using the proposed framework. Fig. 12 summarizes the final results of this study. The results reveal some specific patterns defined in Fig. 2, such as the half-moon (P2), center (P8) and ring patterns (P11). However, some maps had apparent clustering patterns and were thus placed in the "Random Defect" group with no special patterns, which was due to the vigilance threshold setting in the ART1 network. The lower the vigilance threshold value, the more patterns are extracted from the maps. However, a low vigilance threshold value may also cause dissimilar maps to group in the same cluster.

To estimate the effectiveness of the proposed framework, nine senior integration engineers who are domain experts for WBM analysis in this empirical study were asked to cluster the 138 WBMs by themselves. The results were recorded and the time to complete the clustering was measured. Table 3 shows the summary of the WBM clustering result by testers. The results showed that most of the experts clustered the maps into 12–28 patterns and the identified maps with spatial patterns varied from 72 to 103. These results revealed that the sensitivity of identifying patterns by experienced engineers is higher than our algorithm with the current setting of parameters. In addition, the cluster dissimilarity that measures the variability of tester's clustering result to the result of our model is defined as follows:

$$\text{Dissimilarity} = 1 - \frac{\text{Max}(\#\text{ of matched map between tester's cluster and selected cluster})}{\#\text{ of map of selected cluster}}. \tag{8}$$

For specific patterns such as ring, half-moon, and center defect, the engineers can easily cluster the maps together. That is, the maps with blocked defective pattern can be easily grouped by both human and computer algorithm, while the unblocked patterns were recognized from person to person with high variation.

As for the efficiency, the fastest engineer took 15.5 min to complete the clustering (9 patterns of 72 maps) and the time taken varied from 15 to 28 min, while the developed WBM clustering and

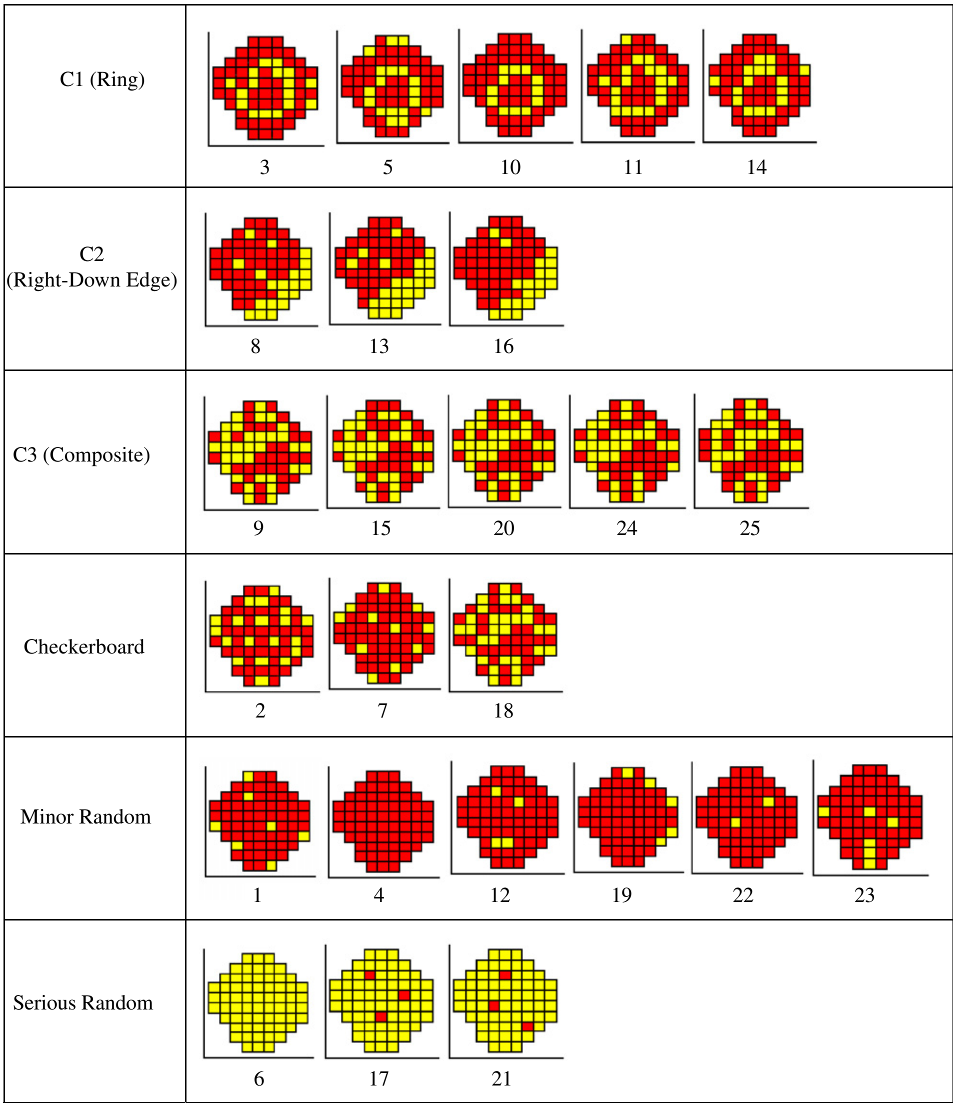

Fig. 10. Patterns extracted from 25 maps.

classification system embedded with the developed hybrid algorithm only took 3 min to generate the result (21 patterns of 79 maps). The algorithm was executed on a PC with a Pentium 4 processor, with a 2.4 GHz clock speed and 512 MB RAM. Furthermore, the computational requirements of our proposed algorithm depend on the training parameters. However, the computational complexity of the algorithm is essentially $O(gm)$, where g is the gross die number and m is the number of map.

5.5. Root cause analyses and identification

This study employed decision tree to identify root cause of specific patterns (Chien et al., 2007). In particular, Patterns P1 and P2 were selected as the target variables, in which Pattern P1 contains 12 maps and Pattern P2 contains 5 maps. For commonality analysis, we selected 10 lots from "Minor Defect" group and "Moderate Defect" group as the "Without Pattern" category of target variable, where three lots from "Minor Defect" group and seven lots from "Moderate Defect" group with top seven highest CP yield. A total of 226 process steps with tool data of these lots are selected as the explained variables. For Pattern P1, the decision tree shows the "ADO_Photo" step has the highest impact factor at the first branch of derived decision tree by excluding the metrology and clean process steps, as shown in Fig. 13. The

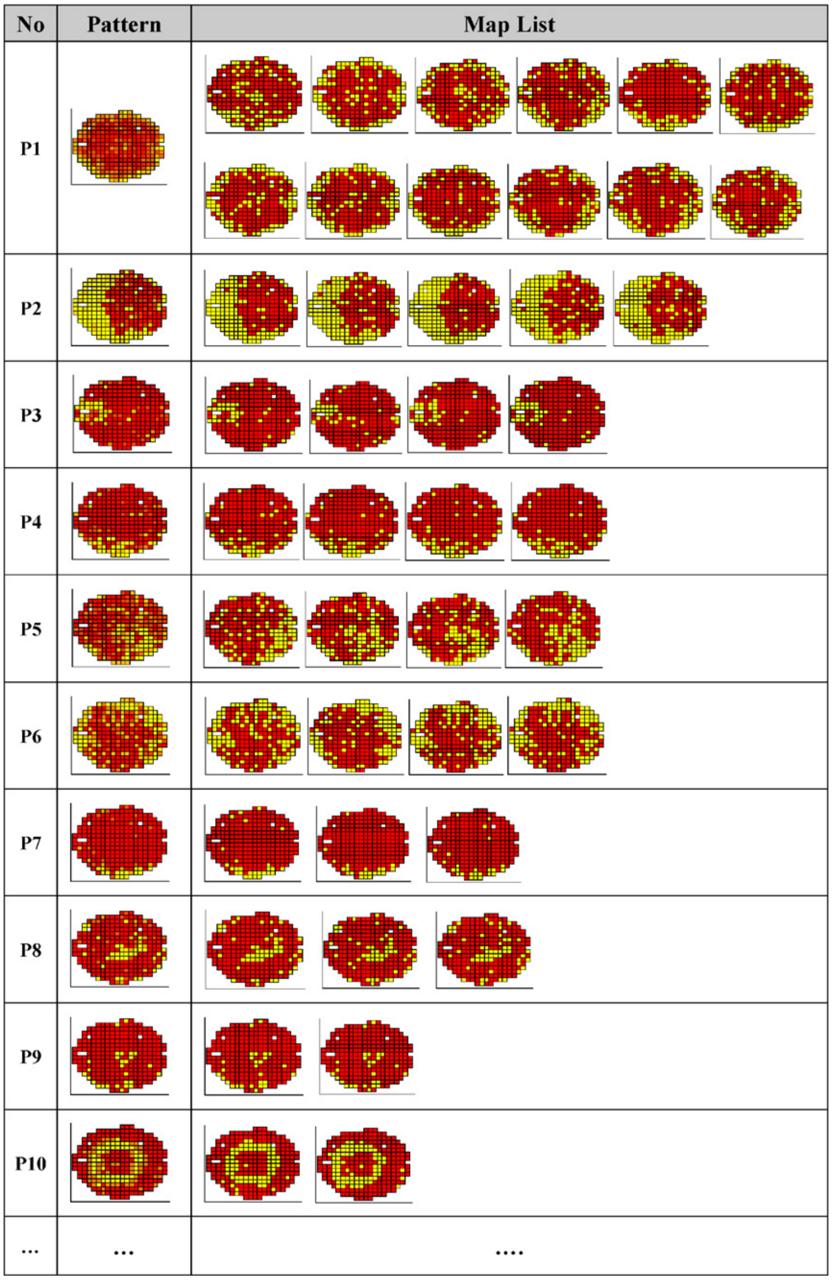

Fig. 11. Partial pattern list after merging patterns.

branch of this tree reveals the higher percentage of lots that were processed by PHO_13 tool had significant P1 pattern than those processed by the other tools. This implies that the lots with P1 pattern may be damaged by "ADO Photo" step and the suspected abnormal tool is "PHO_13". The derived result was then presented to domain experts for interpretation. The process engineers found that the PHO_13 tool did have abnormal events. Similarly, decision tree analysis was employed for Pattern P2 and the result showed the MTE_13 tool in ML1_Etch process caused the lots to have P2 failure pattern than those by the other tools, as shown in Fig. 14. The domain experts who are process engineers have confirmed the results.

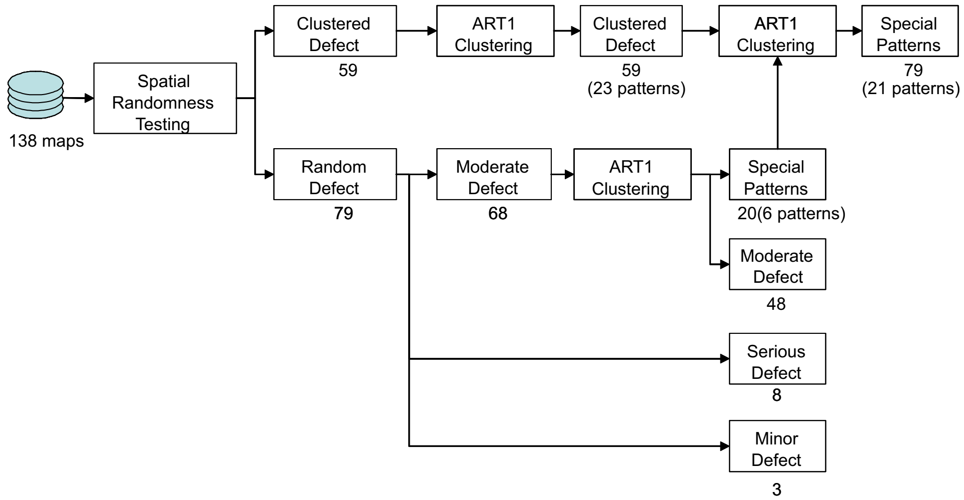

Fig. 12. Summary of clustering results of empirical data.

Table 3
WBM clustering result by nine domain experts

Pattern dissimilarity	Domain experts									Statistics	
	♯1	♯2	♯3	♯4	♯5	♯6	♯7	♯8	♯9	Mean	SD
P1	0.58	0.58	0.08	0.17	0.42	0.67	0.58	0.25	0.67	0.44	0.22
P2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P3	0.00	0.00	0.00	1.00	0.00	0.50	0.00	0.00	0.00	0.17	0.35
P4	0.25	0.00	0.00	0.25	0.50	1.00	0.25	0.00	0.00	0.25	0.33
P5	0.00	0.00	0.00	0.25	0.00	1.00	0.25	0.00	0.00	0.17	0.33
P6	0.25	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.25	0.44	0.11
P7	0.00	0.00	0.00	0.00	0.33	1.00	1.00	0.00	0.00	0.26	0.43
P8	0.00	0.00	0.00	0.00	0.67	0.67	0.00	0.00	0.00	0.15	0.30
P9	0.00	0.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	0.33	0.50
P10	0.00	1.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.44	0.53
P11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P12	0.00	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.11	0.22
P13	0.00	0.00	0.00	1.00	0.00	1.00	1.00	1.00	0.00	0.44	0.53
P14	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	0.00	0.67	0.50
♯ of Patterns	11	11	17	24	16	18	21	9	13	15.6	5.00
♯ of maps with Patterns	91	79	75	79	103	91	99	72	86	86.1	10.74
Time (min)	20.0	22.8	22.5	20.5	23.0	23.8	26.5	15.5	28.5	22.6	22.85

5.6. Implement

The results of the above empirical study show practical viability of the developed algorithm. Indeed, a WBM clustering and classification (WBMCC) expert system embedded with the developed algorithm has now been implemented in a fab in Taiwan. Through the developed WBMCC, engineers can directly query data, view maps, perform WBM clustering and view the results. Alternatively, engineers can also classify an unknown WBM into specific pattern group via WBMCC. Furthermore, engineers can dynamically adjust the ART1 parameters to see the change in result. The system also permits merging patterns by ART1 clustering or by manually selecting several patterns. The clustering results can also link to decision tree analysis to correlate with process tools automatically.

6. Conclusions

This study presented a hybrid algorithm that integrates spatial statistics and ART1 neural

Name	Global Impact
PL2 WETSTRIP	84.6
SPR CLN	71.4
STI CMP CLN	60.0
AOD PHOTO	60.0
CO1 DRY STRIP	56.6
STI DRY STRIP	50.0
SPR ASHING	50.0
ONO DRY ETCH	45.7
CO1 SPR ETCH	44.4
ILD BPSG FLOW	41.7
ZER WET STRIP	41.2
PL4 BARC ETCH	41.2
PL2 PHOTO	41.2
WL SPR ETCH	41.2

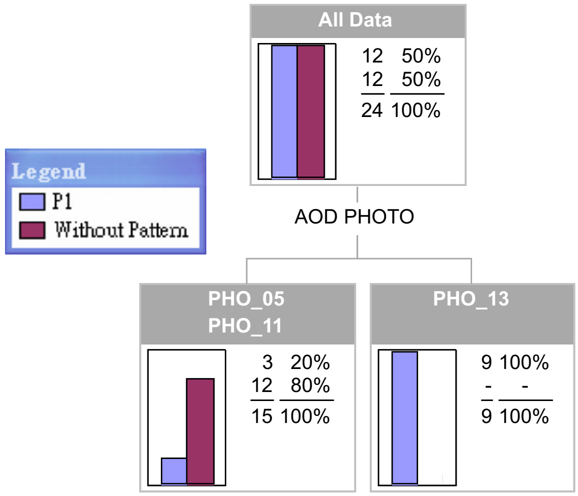

Fig. 13. Commonality analysis result of P1 pattern.

Name	Global Impact
PL1 LIN-OX CLN	100
PL4 BARC ETCH	100
WL SPR ETCH	100
ML1 ETCH	100
STI SW-OX	76.4
STI DRY STRIP	73.8
PL2 WETSTRIP	73.8
SPR ASHING	73.8
PS1 ETCH	73.8
STI ETCH	59.5
RDF WET STRIP	59.5
PW1 WETSTRIP	59.5
NW1 DESCUM	59.5
PL3 WETSTRIP	59.5
SOFT ETCH	59.5

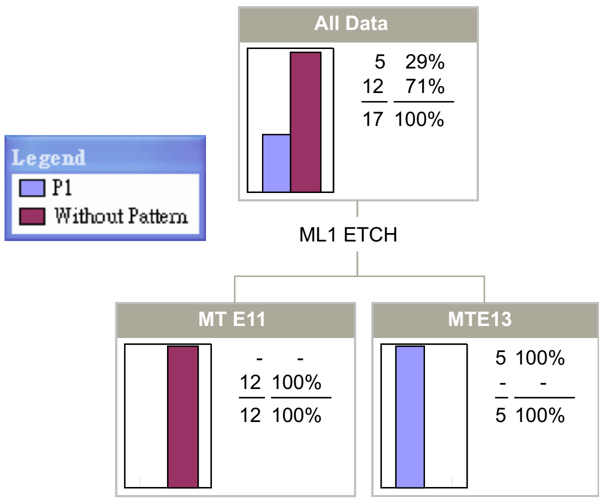

Fig. 14. Commonality analysis result of P2 pattern.

networks to automatically extract patterns from WBM. The framework can identify the maps with spatial correlation using the spatial randomness testing. Furthermore, the patterns can be specified by the ART1 clustering method and extended to link with trouble shooting procedure to provide useful information to support yield improvement decisions and activities in modern fabs. An empirical study showed that the proposed framework can effectively improve the efficiency of WBM clustering, save WBM clustering time consistently and provide the essential information for root causes analysis.

However, the proposed framework that employed ART1 as the clustering method can handle only binary map. Other clustering methods such as ART2 can be applied in the framework for clustering the data with continuous type in the future study. We found that the parameter settings in ESRN and ART1 are sensitive to our framework. For example, the number of clusters extracted in the input maps by ART1 is sensitive to the vigilance parameter. Thus, further studies are needed to fine-tune the parameters in various contexts to increase the effectiveness of WBM clustering. In addition, some systematic patterns such as ring pattern are difficult to detect them by neither spatial randomness testing nor ART. Further research is needed to develop different methodologies to effectively identify specific WBM patterns.

Acknowledgements

This study is partially supported by National Science Council, Taiwan (NSC94-2213-E-007-050) and Macronix International Co., Ltd. Special thanks go to Smith Peng, Top Lin, Qiao-Wen Liu, and Rachel Huang for their assistance.

References

Agresti, A., 1990. Categorical Data Analysis. Wiley, New York.

Bergeret, F., Gall, C.L., 2003. Yield improvement using statistical analysis of process dates. IEEE Transactions on Semiconductor Manufacturing 16 (3), 535–542.

Braha, D., Shmilovici, A., 2002. Data mining for improving a cleaning process in the semiconductor industry. IEEE Transactions on Semiconductor Manufacturing 15 (1), 91–101.

Braha, D., Shmilovici, A., 2003. On the use of decision tree induction for discovery of interactions in a photolithographic process. IEEE Transactions on Semiconductor Manufacturing 16 (4), 644–652.

Breiman, L., Friedman, J.H., Olshen, R.J., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont, CA.

Carpenter, G.A., Grossberg, S., 1988. The ART of adaptive pattern recognition by a self-organization neural network. Computer 21 (3), 77–88.

Carpenter, G.A., Grossberg, S., 1987. ART 2: Stable self-organization of pattern recognition codes for analog input patterns. Applied Optics 26, 4919–4930.

Carpenter, G.A., Grossberg, S., 1990. ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. Neural Networks 3, 129–152.

Carpenter, G.A., Grossberg, S., Reynolds, J.H., 1991. ARTMAP: Supervised real-time learning and classification of non-stationary data by a self-organizing neural network. Neural Networks 4, 565–588.

Carpenter, G.A., Grossberg, S., Rosen, D.B., 1991. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Networks 4, 759–771.

Chien, C., Lee, P., Peng, C., 2003. Semiconductor manufacturing data mining for clustering and feature extraction. Journal of Information Management 10 (1), 63–84.

Chien, C., Wang, W., Cheng, J., 2007. Data mining for yield enhancement in semiconductor manufacturing and an empirical study. Expert Systems with Applications 33 (1), 1–7.

Cunningham, S.P., Spanos, C.J., Voros, K., 1995. Semiconductor yield improvement: Results and best practices. IEEE Transactions on Semiconductor Manufacturing 8 (2), 103–109.

Freeman, J.A., Skapura, D.M., 1991. Neural Networks: Algorithm and Programming Techniques. Addison-Wesley, Reading, MA.

Friedman, D.J., Hansen, M.H., Nair, V.N., James, D.A., 1997. Model-free estimation of defect clustering in integrated circuit fabrication. IEEE Transactions on Semiconductor Manufacturing 10 (3), 344–359.

Hansen, M.H., Nair, V.N., 1995. Monitoring wafer map from integrated circuit fabrication processes for spatially clustered defects. Technometrics 39 (3), 241–253.

Kaempf, U., 1995. The binomial test: A simple tool to identify process problems. IEEE Transactions on Semiconductor Manufacturing 8 (2), 160–166.

Kass, G.V., 1980. An exploratory technique for investigating large quantities of categorical data. Applied Statistics 292 (2), 119–127.

Keki, B., Cheng, J., Fayyad, U., Qian, Z., 1993. Applying machine learning to semiconductor manufacturing. IEEE Expert, 41–47.

Kusiak, A., 2000. Decomposition in data mining: An industrial case study. IEEE Transactions on Electronic Packaging Manufacturing 23 (4), 345–352.

Leachman, R.C., Hodges, D.A., 1996. Benchmarking semiconductor manufacturing. IEEE Transactions on Semiconductor Manufacturing 9 (2), 158–169.

Lim, T.S., Loh, W.Y., Shih, Y.S., 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine Learning 40, 203–229.

Mallory, C.L., Perloff, D.S., Hasan, T.F., Stanley, R.M., 1983. Spatial yield analysis in integrated circuit manufacturing. Solid State Technology November, 121–127.

Peng, C., Chien, C., 2003. Data value development to enhance yield and maintain competitive advantage for semiconductor manufacturing. International Journal of Service Technology and Management 4 (6), 365–383.

Peng, J., Chien, C., Tseng, B., 2004. Rough set theory for data mining for fault diagnosis on distribution feeder. In: IEE Proceedings-Generation, Transmission, and Distributions 151 (6), 689–697.

Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco, CA.

Stapper, C.H., 2000. LSI yield modeling and process monitoring. IBM Journal of Research and Development 44 (2), 112–118.

Taam, W., Hamada, M., 1993. Detecting spatial effects from factorial experiments: an application from integrated-circuit manufacturing. Technometrics 35 (2), 149–160.