Research Article

Clustering Ensemble for Identifying Defective Wafer Bin Map in Semiconductor Manufacturing

Chia-Yu Hsu

Department of Information Management and Innovation Center for Big Data & Digital Convergence, Yuan Ze University, Chungli, Taoyuan 32003, Taiwan

Correspondence should be addressed to Chia-Yu Hsu; cyhsu@saturn.yzu.edu.tw

Wafer bin map (WBM) represents specific defect pattern that provides information for diagnosing root causes of low yield in semiconductor manufacturing. In practice, most semiconductor engineers use subjective and time-consuming eyeball analysis to assess WBM patterns. Given shrinking feature sizes and increasing wafer sizes, various types of WBMs occur; thus, relying on human vision to judge defect patterns is complex, inconsistent, and unreliable. In this study, a clustering ensemble approach is proposed to bridge the gap, facilitating WBM pattern extraction and assisting engineer to recognize systematic defect patterns efficiently. The clustering ensemble approach not only generates diverse clusters in data space, but also integrates them in label space. First, the mountain function is used to transform data by using pattern density. Subsequently, k-means and particle swarm optimization (PSO) clustering algorithms are used to generate diversity partitions and various label results. Finally, the adaptive response theory (ART) neural network is used to attain consensus partitions and integration. An experiment was conducted to evaluate the effectiveness of proposed WBMs clustering ensemble approach. Several criterions in terms of sum of squared error, precision, recall, and F-measure were used for evaluating clustering results. The numerical results showed that the proposed approach outperforms the other individual clustering algorithm.

1. Introduction

To maintain their profitability and growth despite continual technology migration, semiconductor manufacturing companies provide wafer manufacturing services generating value for their customers through yield enhancement, cost reduction, on-time delivery, and cycle time reduction [1, 2]. The consumer market requires that semiconductor products exhibiting increasing complexity be rapidly developed and delivered to market. Technology continues to advance and required functionalities are increasing; thus, engineers have a drastically decreased amount of time to ensure yield enhancement and diagnose defects [3].

The lengthy process of semiconductor manufacturing involves hundreds of steps, in which big data including the wafer lot history, recipe, inline metrology measurement, equipment sensor value, defect inspection, and electrical test data are automatically generated and recorded. Semiconductor companies experience challenges integrating big data from various sources into a platform or data warehouse and lack intelligent analytics solutions to extract useful manufacturing intelligence and support decision making regarding production planning, process control, equipment monitoring, and yield enhancement. Scant intelligent solutions have been developed based on data mining, soft computing, and evolutionary algorithms to enhance the operational effectiveness of semiconductor manufacturing [4–7].

Circuit probe (CP) testing is used to evaluate each die on the wafer after the wafer fabrication processes. Wafer bin maps (WBMs) represent the results of a CP test and provide crucial information regarding process abnormalities, facilitating the diagnosis of low-yield problems in semiconductor manufacturing. In WBM failure patterns the spatial dependences across wafers express systematic and random effects. Various failure patterns are required; these pattern types facilitate rapidly identifying the associate root causes of low yield [8]. Based on the defect size, shape, and location on the wafer, the WBM can be expressed as specific patterns

such as rings, circles, edges, and curves. Defective dies caused by random particles are difficult to completely remove and typically exhibit nonspecific patterns. Most WBM patterns consisted of a systematic pattern and a random defect [8–10].

In practice, thousands of WBMs are generated for inspection and engineers must spend substantial time on pattern judgment rather than determining the assignable causes of low yield. Grouping similar WBMs into the same cluster can enable engineers to effectively diagnose defects. The complicated processes and diverse products fabricated in semiconductor manufacturing can yield various WBM types, making it difficult to detect systematic patterns by using only eyeball analysis.

Clustering analysis is used to partition data into several groups in which the observations are homogeneous within a group and heterogeneous between groups. Clustering analysis has been widely applied in applications such as grouping [11] and pattern extraction [12]. However, most conventional clustering algorithms influence the result based on the data type, algorithm parameter settings, and prior information. For example, the k-means algorithm is used to analyze substantial amount of data that exhibit time complexity [13]. However, the results of the k-means algorithm depend on the initially selected centroid and predefined number of clusters. To address the disadvantages of the k-means algorithm evolutionary methods have been developed to conduct data clustering such as the genetic algorithm (GA) and particle swarm optimization (PSO) [14]. PSO is particularly advantageous because it requires less parameter adjustment compared with the GA [15].

Combining results by applying distinct algorithms to the same data set or algorithm by using various parameter settings yields high-quality clusters. Based on the criteria of the clustering objectives, no individual clustering algorithm is suitable for whole problem and data type. Compared with individual clustering algorithms, clustering ensembles that combine multiple clustering results yield superior clustering effectiveness regarding robustness and stability, incorporating conflicting results across partitions [16]. Instead of searching for an optimal partition, clustering ensembles capture a consensus partition by integrating diverse partitions from various clustering algorithms. Clustering ensembles have been developed to improve the accuracy, robustness, and stability of clustering; such ensembles typically involve two steps. The first step involves generating a basic set of partitions that can be similar to or distinct from those of various parameters and cluster algorithms [17]. The second step involves combining the basic set of partitions by using a consensus function [18]. However, with the shrinking integrated circuit feature size and complicated manufacturing process, the WBM patterns become more complex because of various defect density, die size, and wafer rotation. It is difficult to extract defect pattern by single specific clustering approach and needs to incorporate different clustering aspects for various complicated WBM patterns.

To bridge the need in real setting, this study proposes a WBM clustering ensemble approach to facilitate WBM defect pattern extraction. First, the target bin value is categorized into binary value and the wafer maps are transformed from two-dimensional to one-dimensional data. Second, k-means and PSO clustering algorithms are used to generate various diversity partitions. Subsequently, the clustering results are regarded as label representations to facilitate aggregating the diversity partition by using an adaptive response theory (ART) neural network. To evaluate the validity of the proposed method, an experimental analysis was conducted using six typical patterns found in the fabrication of semiconductor wafers. Using various parameter settings, the proposed cluster ensembles that combine diverse partitions instead of using the original features outperform individual clustering methods such as k-means and PSO.

The remainder of this study is organized as follows. Section 2 introduces a fundamental WBM. Section 3 presents the proposed approach to the WBM clustering problem. Section 4 provides experimental comparisons, applying the proposed approach to analyze the WBM clustering problem. Section 5 offers a conclusion and the findings and future research directions are discussed.

2. Related Work

A WBM is a two-dimensional failure pattern. Based on various defects types, random, systematic, and mixed failure patterns are primary types of WBMs generated during semiconductor fabrication [19, 20]. Random failure patterns are typically caused by random particles or noises in the manufacturing environment. In practice, completely eliminating these random defects is difficult. Systematic failure patterns show the spatial correlation across wafers such as rings, crescent moon, edge, and circles. Figure 1 shows typical WBM patterns which are transformed into binary values for visualization and analysis. The dies that pass the functional test are denoted as 0 and the defective dies are denoted as 1. Based on the systematic patterns, domain engineers can rapidly determine the assignable causes of defects [8]. Mixed failure patterns comprise the random and systematic defects on a wafer. The mixed pattern can be identified if the degree of the random defect is slight.

Defect diagnosis of facilitating yield enhancement is critical in the rapid development of semiconductor manufacturing technology. An effective method of ensuring that the causes of process variation are assignable is analyzing the spatial defect patterns on wafers. WBMs provide crucial guidance, enabling engineers to rapidly determine the potential root causes of defects by identifying patterns. Most studies have used neural network and model-based approaches to extract common WBM patterns. Hsu and Chien [8] integrated spatial statistical analysis and an ART neural network to conduct WBM clustering and associated the patterns with manufacturing defects to facilitate defect diagnosis. In addition to ART neural network, Liu and Chien [10] applied moment invariant for shape clustering of WBMs. Model-based clustering algorithms are used to construct a model for each cluster and compare the likelihood values between clusters to identify defect patterns. Wang et al. [21] used model-based clustering, applying a Gaussian expectation maximization algorithm to estimate defect patterns. Hwang and Kuo [22] modeled global defects
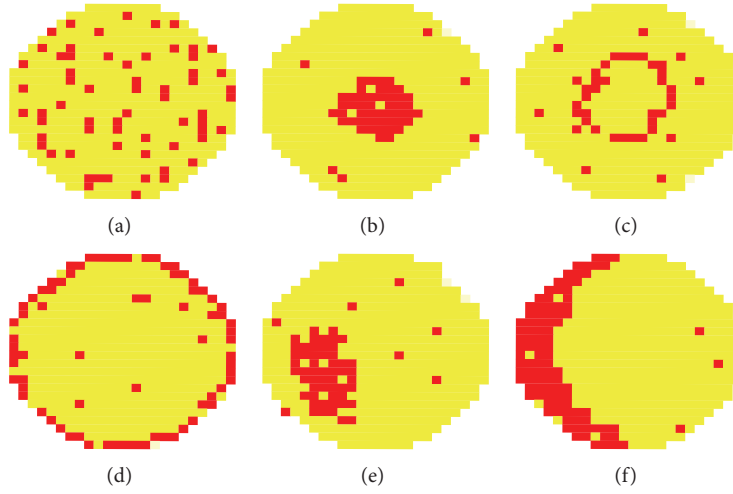
FIGURE 1: Typical WBM patterns.

and local defects in clusters exhibiting ellipsoidal patterns and local defects in clusters exhibiting linear or curvilinear patterns. Yuan and Kuo [23] used Bayesian inference to identify the patterns of spatial defects in WBMs. Driven by continuous migration of semiconductor manufacturing technology, the more complicated types of WBM patterns have been occurred due to the increase of wafer size and shrinkage of critical dimensions on specific aspect of complex WBM pattern and little research has evaluated using the clustering ensemble approach to analyze WBMs and extract failure patterns.

3. Proposed Approach

The terminologies and notations used in this study are as follows:

N_g: number of gross dies;

N_w: number of wafers;

N_p: number of particles;

N_c: number of clusters;

N_b: number of bad dies;

i: wafer index, $i = 1, 2, \ldots, N_w$;

j: dimension index, $j = 1, 2, \ldots, N_g$;

k: cluster index, $k = 1, 2, \ldots, N_c$;

l: particle index, $l = 1, 2, \ldots, N_p$;

q: clustering result index, $q = 1, 2, \ldots, M$;

r: bad die index, $r = 1, 2, \ldots, N_b$;

s: clustering subobjective in PSO clustering, $s = 1, 2, 3$;

U: uniform random number in the interval $[0, 1]$;

ω_v: inertia weight of velocity update;

ω_s: weight of clustering subobjective;

c_p: personal best position acceleration constants;

c_g: global best position acceleration constants;

β: a normalization factor;

m: a constant for approximate density shape in mountain function;

y_r: the rth bad die on a wafer;

n_k: the number of WBMs in the kth cluster;

n_{lk}: the number of WBMs in the kth cluster of lth particle;

C_k: subset of WBMs in the kth cluster;

$x^{\max}$: maximum value in the WBM data;

$\mathbf{m}_k$: vector of the kth cluster centroid, $\mathbf{m}_k = [m_{k1}, m_{k2}, \ldots, m_{kN_g}]$;

$\mathbf{m}_{lk}$: vector centroid of the kth cluster of lth particle;

$\mathbf{p}_l$: vector centroids of the lth particle, $\mathbf{p}_l = [m_{l1}, m_{l2}, \ldots, m_{lk}]$;

θ_{lj}: position of the lth particle at the jth dimension;

V_{lj}: velocity of the lth particle at the jth dimension;

ψ_{lj}: personal best position ($p$best) of the lth particle at jth dimension;

ψ_{gj}: global best position ($g$best) at the jth dimension;

$\mathbf{x}_i$: vector of the ith WBM, $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{iN_g}]$;

$\boldsymbol{\Theta}_l$: vector position of the lth particle, $\boldsymbol{\Theta}_l = [\theta_{l1}, \theta_{l2}, \ldots, \theta_{lN_g}]$;

$\mathbf{V}_l$: vector velocity of the lth particle, $\mathbf{V}_l = [V_{l1}, V_{l2}, \ldots, V_{lN_g}]$;

$\boldsymbol{\psi}_l$: vector personal best of the lth particle, $\boldsymbol{\psi}_l = [\psi_{l1}, \psi_{l2}, \ldots, \psi_{lN_g}]$;

$\boldsymbol{\psi}_g$: vector global best position, $\boldsymbol{\psi}_g = [\psi_{g1}, \psi_{g2}, \ldots, \psi_{gN_g}]$.
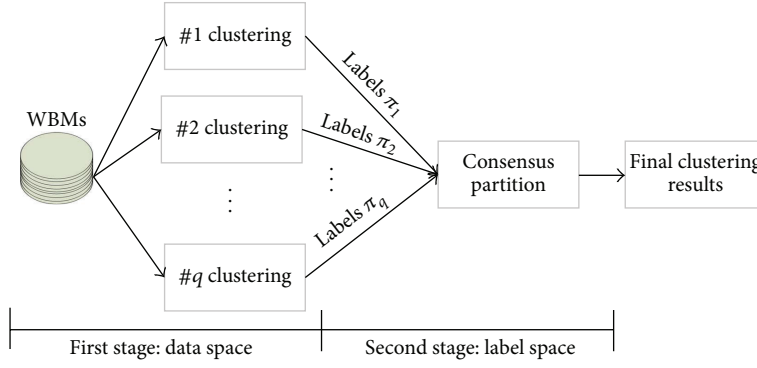
Figure 2: A framework for WBMs clustering ensemble.

3.1. Problem Definition of WBM Clustering Ensemble.

Clustering ensembles can be regarded as two-stage partitions, in which various clustering algorithms are used to assess the data space at the first stage and consensus function is used to assess the label space at the second stage. Figure 2 shows the two-stage clustering perspective. Consensus function is used to develop a clustering combination based on the diversity of the cluster labels derived at the first stage.

Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_w}\}$ denote a set of N_w WBMs and $\Pi = \{\pi_1, \pi_2, \ldots, \pi_M\}$ denote a set of partitions based on M clustering results. The various partitions of $\pi_q(x_i)$ represent a label assigned to x_i by the qth algorithm. Each label vector π_q is used to construct a representation Π, in which the partitions of $\mathbf{X}$ comprise a set of labels for each wafer $\mathbf{x}_i$, $i = 1, \ldots, N_w$. Therefore, the difficulty of constructing a clustering ensemble is locating a new partition Π that provides a consensus partition satisfying the label information derived from each individual clustering result of the original WBM. For each label π_q, a binary membership indicator matrix $H^{(q)}$ is constructed, containing a column for each cluster. All values of a row in the $H^{(q)}$ are denoted as 1 if the row corresponds to an object. Furthermore, the space of a consensus partition changes from the original N_g features into N_w features. For example, Table 1 shows eight WBMs grouped using three clustering algorithms (π_1, π_2, π_3); the three clustering results are transformed into clustering labels that are transformed into binary representations (Table 2). Regarding consensus partitions, the binary membership indicator matrix $H^{(q)}$ is used to determine a final clustering result, using a consensus model based on the eight features ($v_1, v_2, \ldots, v_8$).

3.2. Data Transformation.

The binary representation of good and bad dies is shown in Figure 3(a). Although this binary representation is useful for visualisation, displaying the spatial relation of each bad die across a wafer is difficult.

To quantify the spatial relations and increase the density of a specific feature, the mountain function is used to transform the binary value into a continuous value. The mountain method is used to determine the approximate cluster center by estimating the probability density function of a feature [24]. Instead of using a grid node, a modified mountain

Table 1: Original label vectors.

	π_1	π_2	π_3
$\mathbf{x}_1$	1	1	1
$\mathbf{x}_2$	1	1	1
$\mathbf{x}_3$	1	1	1
$\mathbf{x}_4$	2	2	1
$\mathbf{x}_5$	2	2	2
$\mathbf{x}_6$	3	1	2
$\mathbf{x}_7$	3	1	2
$\mathbf{x}_8$	3	1	2

Table 2: Binary representation of clustering ensembles.

Clustering results		v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
$H^{(1)}$	h_{11}	1	1	1	0	0	0	0	0
	h_{12}	0	0	0	1	1	0	0	0
	h_{13}	0	0	0	0	0	1	1	1
$H^{(2)}$	h_{21}	1	1	1	0	0	1	1	1
	h_{22}	0	0	0	1	1	0	0	0
$H^{(3)}$	h_{31}	1	1	1	1	0	0	0	0
	h_{32}	0	0	0	0	1	1	1	1

function can employ data points by using a correlation self-comparison [25]. The modified mountain function for a bad die r on a wafer $M(y_r)$ is defined as follows:

$$M(y_r) = \sum_{r=1}^{N_b} e^{-m\beta d(y_r, y_s)}, \quad r = 1, 2, 3, \ldots, N_b, \quad (1)$$

where

$$\beta = \left(\frac{d(y_r - y_{wc})}{N_b} \right)^{-1} \quad (2)$$

and $d(y_r, y_s)$ is the distance between dices r and s. Parameter β is the normalization factor for the distance between bad die r and the wafer centroid y_{wc}. Parameter m is a constant. Parameter $m\beta$ determines the approximate density shape of the wafer. Figure 3(b) shows an example of WBM transformation. Two types of data are used to generate a basic set of partitions. Moreover, each WBM must sequentially transform

(1) Randomly select k data as the centroid of cluster
(2) Repeat
    For each data vector, assign each data into the group with respect to the closest centroid by
        minimum Euclidean distance.
        recalculate the new centroid based on all data within the group.
    end for
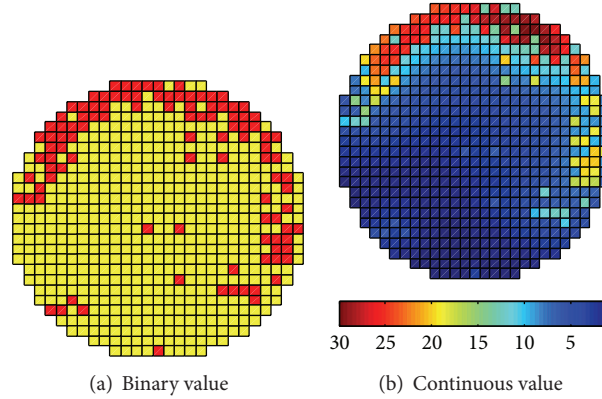(3) Steps 1 and 2 are iterated until there is no data change.

PROCEDURE 1: k-means algorithm.

(a) Binary value   (b) Continuous value

FIGURE 3: Representation of wafer bin map by binary value and continuous value.

from a two-dimensional map into a one-dimensional data vector [8]. Such vectors are used to conduct further clustering analysis.

3.3. Diverse Partitions Generation by k-Means and PSO Clustering.

Both k-means and PSO clustering algorithms are used to generate basic partitions. To consider the spatial relations across a wafer, both the binary and continuous values are used to determine distinct clustering results by using k-means and PSO clustering. Subsequently, various numbers of clusters are used for comparison.

K-means is an unsupervised method of clustering analysis [13] used to group data into several predefined numbers of clusters by employing a similarity measure such as the Euclidean distance. The objective function of the k-means algorithm is to minimize the within-cluster difference, that is, the sum of the square error (SSE) which is determined using (3). The k-means algorithm consists of the following steps as shown in Procedure 1:

$$\text{SSE} = \sum_{k=1}^{N_c} \sum_{\mathbf{x}_i \in C_k} \left(\mathbf{x}_i - \mathbf{m}_k \right)^2. \tag{3}$$

Data clustering is regarded as an optimisation problem. PSO is an evolutionary algorithm [14] which is used to search for optimal solutions based on the interactions amongst particles; it requires adjusting fewer parameters compared with using other evolutionary algorithms. van der Merwe and Engelbrecht [26] proposed a hybrid algorithm for clustering data, in which the initial swarm is determined using the k-means result and PSO is used to refine the cluster results.

A single particle $\mathbf{p}_l$ represents the k cluster centroid vectors: $\mathbf{p}_l = [m_{l1}, m_{l2}, \ldots, m_{lk}]$. A swarm defines a number of candidate clusters. To consider the maximal homogeneity within a cluster and heterogeneity between clusters, a fitness function is used to maximize the intercluster separation and minimize the intracluster distance and quantisation error

$$f(\mathbf{p}_i, \mathbf{Z}_l) = \omega_1 \times J_e + \omega_2 \times \overline{d}_{\max}(\mathbf{p}_l, \mathbf{Z}_l) + \omega_3 \\ \times (X_{\max} - d_{\min}(\mathbf{p}_l)), \tag{4}$$

where $\mathbf{Z}_l$ is a matrix representing the assignment of the WBMs to the clusters of the lth particle. The following quantization error equation is used to evaluate the level of clustering performance:

$$J_e = \frac{\sum_{k=1}^{N_c} \left\lfloor \sum_{\forall \mathbf{x}_i \in C_k} d(\mathbf{x}_i, m_k)/n_k \right\rfloor}{K}. \tag{5}$$

In addition,

$$\overline{d}_{\max}(\mathbf{p}_i, \mathbf{Z}_l) = \max_{k=1,2,\ldots,N_c} \left\lfloor \sum_{\forall \mathbf{x}_i \in C_{lk}} \frac{d(\mathbf{x}_i, \mathbf{m}_{lk})}{n_{lk}} \right\rfloor \tag{6}$$

is the maximum average Euclidean distance of particle to the assigned clusters and

$$d_{\min}(\mathbf{p}_l) = \min_{\forall u, v, u \neq v} \left[d(\mathbf{m}_{lu}, \mathbf{m}_{lv}) \right] \tag{7}$$

is the minimum Euclidean distance between any pair of clusters. Procedure 2 shows the steps involved in the PSO clustering algorithm.

(1) Initialize each particle with k cluster centroids.
(2) For iteration $t = 1$ to $t = $ max do
    For each particle l do
    For each data pattern $\mathbf{x}_i$
        calculate the Euclidean distance to all cluster centroids and assign pattern $\mathbf{x}_i$ to cluster c_k
        which has the minimum distance
    end for
    calculate the fitness function $f(\mathbf{p}_i, \mathbf{Z}_l)$
    end for
    find the personal best and global best positions of each particle.
    update the cluster centroids by the update velocity equation (i) and update coordinate equation (ii).
    $\mathbf{V}_i(t + 1) = \omega_v \mathbf{V}_i(t) + c_p u(\psi_l(t) - \Theta_l(t)) + c_g u(\psi_g(t) - \Theta_l(t))$ (i)
    $\Theta_l(t + 1) = \Theta_l(t) + \mathbf{V}_l(t + 1)$ (ii)
    end for
(3) Step 2 is iterated until these is no data change

PROCEDURE 2: PSO clustering algorithm.

3.4. Consensus Partition by Adaptive Response Theory. ART has been used in numerous areas such as pattern recognition and spatial analysis [27]. Regarding the unstable learning conditions caused by new data, ART can be used to address stability and plasticity because it addresses the balance between stability and plasticity, match and reset, and search and direct access [8]. Because the input labels are binary, the ART1 neural network [27] algorithm is used to attain a consensus partition of WBMs.

The consensus partition approach is as follows.

Step 1. Apply k-means and PSO clustering algorithms and use various parameters (e.g., various numbers of clusters and types of input data) to generate diverse clusters.

Step 2. Transform the original clustering label into binary representation matrix H as an input for ART1 neural network.

Step 3. Apply ART1 neural network to aggregate the diverse partitions.

4. Numerical Experiments

In this section, this study conducts a numerical study to demonstrate the effectiveness of the proposed clustering ensemble approach. Six typical WBM patterns from semiconductor fabrication were used such as moon, edge, and sector. In the experiments, the percentage of defective dies in six patterns is designed based on real cases. Without losing generality of WBM patterns, the data have been systematically transformed for proprietary information protection of the case company. Total 650 chips were exposed on a wafer. Based on various degrees of noise, each pattern type was used to generate 10 WBMs for estimating the validity of proposed clustering ensemble approach. The noise in WBM could be caused from random particles across a wafer and test bias in CP test, which result in generating bad die randomly on a wafer and generating good die within a group of bad dies. It means that some bad dices are shown as good dice and the
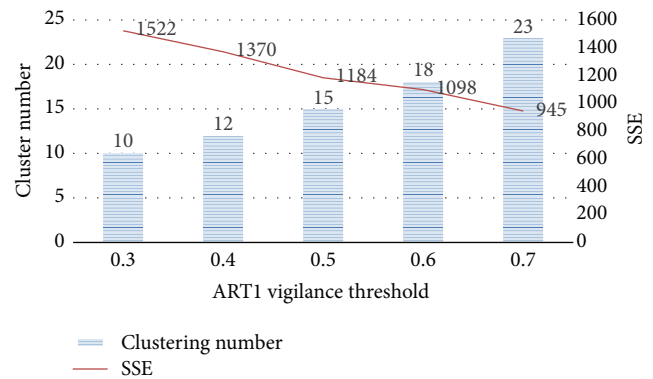

FIGURE 4: Comparison of various ART1 vigilance threshold.

density of bad die could be sparse. For example, the value of degree of noise is 0.02 which represents total 2% good die and bad dies are inverse.

The proposed WBM clustering ensemble approach was compared with k-means, PSO clustering method, and the algorithm proposed by Hsu and Chien [8]. Six numbers of clusters were used for single k-means methods and single PSO clustering algorithms. Table 3 showed the parameter settings for PSO clustering. The number of clusters extracted by ART1 neural network is sensitive to the vigilance threshold value. The high vigilance threshold is used to produce more clusters and the similarity within a cluster is high. In contrast, the low vigilance threshold results in fewer numbers of clusters. However, the similarity within a cluster could be low. To compare the parameter setting of ART1 vigilance threshold, various values were used as shown in Figure 4. Each clustering performance was evaluated in terms of the SSE and number of clusters. The SSE is used to compare the cohesion amongst various clustering results, and a small SSE indicates that the WBM within a cluster is highly similar. The number of clusters represents the effectiveness of the WBM grouping. According to the objective of clustering is to group

TABLE 3: Parameter settings for PSO clustering.

Parameter	Value	Parameter	Value
m	20	ω	1
$X^{\max}$	1	a_1	0.4
c_p	2	a_2	0.3
c_g	2	a_3	0.3
Iteration	500		

TABLE 4: Results of clustering methods by SSE.

Methods		Noise degree				
		0.02	0.04	0.06	0.08	0.10
Hsu and Chien [8]		1184	1192	1203	1248	1322
Individual clustering	KB	2889	3092	3003	4083	3570
	KC	3331	2490	2603	3169	2603
	PB	5893	3601	6566	5839	6308
	PC	4627	4873	3330	3787	6112
Clustering ensemble	KB and PB	1827	1280	1324	1801	2142
	KC and PC	2272	2363	2400	1509	1718
	KB and PC	1368	1459	2400	1509	2597
	KC and PB	2100	2048	1421	1928	2043
	KB and PB and KC and PC	1586	1550	1541	1571	1860

the WBM into few clusters in which the similarities among the WBMs within a cluster are high as possible. Therefore, the setting of ART1 vigilance threshold value is used as 0.50 in the numerical experiments.

WBM clustering is to identify the similar type of WBM into the same cluster. To consider only six types of WBMs that were used in the experiments, the actual number of clusters should be six. Based on the various degree of noise in WBM generation as shown in Table 4, several individual clustering methods including ART1 [8], k-means clustering, and PSO clustering were used for evaluating clustering performance. Table 4 shows that the ART1 neural network yielded a lower SSE compared with the other methods. However, the ART1 neural network separates the WBM into 15 clusters as shown in Figure 5. The ART1 neural network yields unnecessary partitions for the similar type of WBM pattern. In order to generate diverse clustering partitions for clustering ensemble method, four combinations with various data scale and clustering algorithms including k-means by binary value (KB), k-means by continuous value (KC), PSO by binary value (PB), and PSO by continuous value (PC) are used. Regardless of the individual clustering results based on six numbers of clusters, using K-means clustering and PSO clustering individually yielded larger SSE values than using ART1 only.

Table 4 also shows the clustering ensembles that use various types of input data. For example, the clustering ensemble method KB&PB integrates the six results including the k-means algorithm by three kinds of clusters (i.e., $k = 5, 6, 7$) and PSO clustering by three kinds of clusters (i.e., $k = 5, 6, 7$), respectively, to form the WBM clustering via
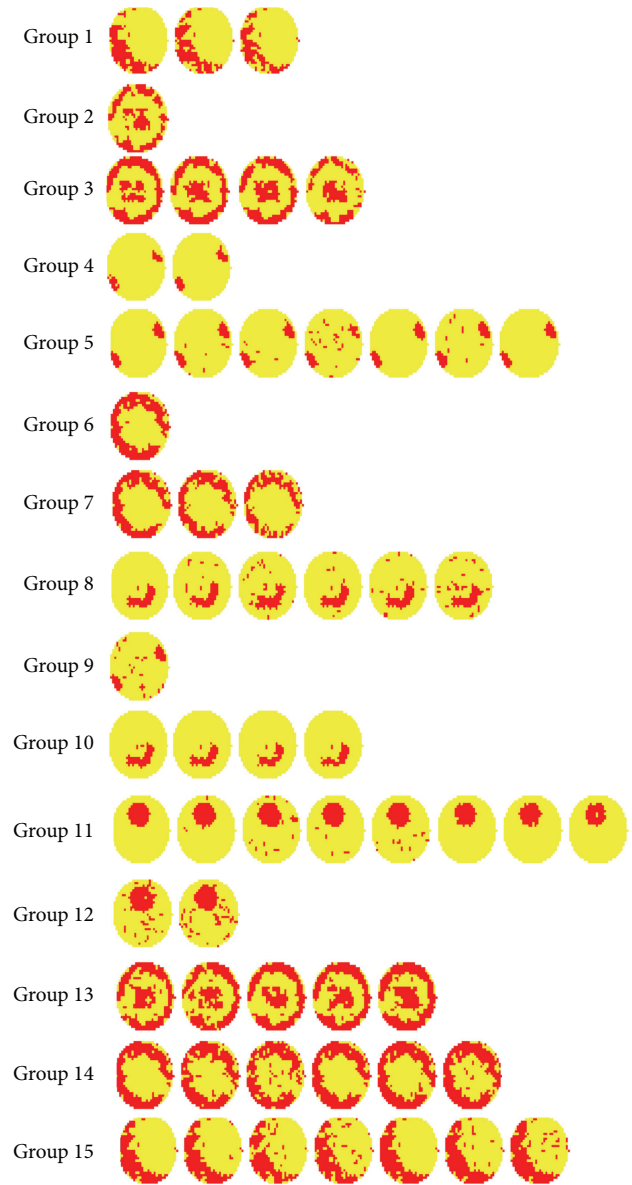

FIGURE 5: Clustering result by ART1 (15 clusters).

label space. In general, the clustering ensembles demonstrate smaller SSE values than do individual clustering algorithms such as the k-means or PSO clustering algorithms.

In addition to compare the similarity within the cluster, an index called specificity was used to evaluate the efficiency of the evolved cluster over representing the true clusters [28]. The specificity is defined as follows:

$$\text{specificity} = \frac{t_c}{T_e}, \qquad (8)$$

where t_c is the number of true WBM patterns covered by the number of evolved WBM patterns and T_e is the total number of evolved WBM patterns. As shown in the ART1 neural network clustering results, the total number of evolved WBM clusters is 15 and number of true WBM clusters is 6. Then, the specificity is 0.4. Table 5 shows the results of specificity
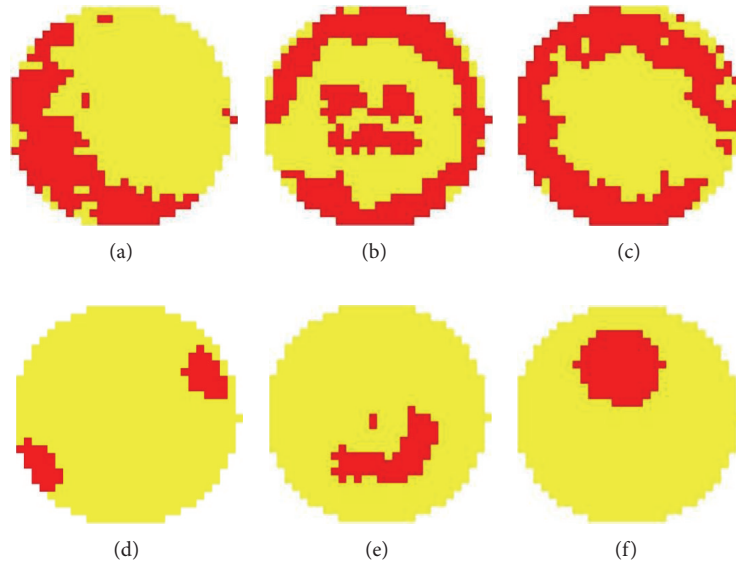
FIGURE 6: Six types of WBM patterns.

TABLE 5: Results of clustering methods by specificity.

Methods		Noise degree				
		0.02	0.04	0.06	0.08	0.10
Hsu and Chien [8]		0.4	0.4	0.4	0.4	0.4
Individual clustering	KB	1.0	1.0	1.0	1.0	1.0
	KC	1.0	1.0	1.0	1.0	1.0
	PB	1.0	1.0	1.0	1.0	1.0
	PC	1.0	1.0	1.0	1.0	1.0
Clustering ensemble	KB and PB	0.7	0.5	0.5	0.5	0.8
	KC and PC	0.5	0.8	0.9	0.8	0.6
	KB and PC	0.5	0.7	0.9	0.8	0.7
	KC and PB	0.9	0.5	0.5	0.6	0.7
	KB and PB and KC and PC	1.0	0.9	0.9	0.9	1.0

among clustering methods. The ART1 neural network has the lowest specificity due to the large number of clusters. The specificity of individual clustering is 1 because the number of evolved WBM patterns is fixed as 6. Furthermore, compared with individual clustering algorithms, combining various clustering ensembles yields not only smaller SSE values, but also smaller numbers of clusters. Thus, the homogeneity within a cluster can be improved using proposed approach. The threshold of ART1 neural network yields maximal cluster numbers. Therefore, the proposed clustering ensemble approach considering diversity partitions has better results regarding the SSE and number of clusters than individual clustering methods.

To evaluate the results among various clustering ensembles and to assess cluster validity, WBM class labels are employed based on six pattern types as shown in Figure 6.

Thus, the indices including precision and recall are two classification-oriented measures [29] defined as follows:

$$
\begin{aligned}
\text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\
\text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}},
\end{aligned}
\tag{9}
$$

where TP (true positive) is the number of WBMs correctly classified into WBM patterns, FP (false positive) is the number of WBMs incorrectly classified, and FN (false negative) is the number of WBMs that need to be classified, but not to be determined incorrectly. The precision measure is used to assess how many WBMs classified as Pattern (a) are actually Pattern (a). The recall measure is used to assess how many samples of Pattern (a) are correctly classified.

However, a trade-off exists between precision and recall; therefore, when one of these measures increases, the other decreases. The F-measure is a harmonic mean of the precision and recall which is defined as follows:

$$
F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{2\text{TP}}{\text{FP} + \text{FN} + 2\text{TP}}.
\tag{10}
$$

Specifically, the F-measure represents the interaction between the actual and classification results (i.e., TP). If the classification result is close to the actual value, the F-measure is high.

Tables 6, 7, and 8 show a summary of various metrics among six types of WBM in precision, recall, and F-measure, respectively. As shown in Figure 6, Patterns (b) and (c) are similar in the wafer edge, demonstrating smaller average precision and recall values compared with the other patterns. The clustering ensembles which generate partitions by using k-means make it difficult to identify in both Patterns (b) and (c). Using a mountain function transformation enables

TABLE 6: Clustering result on the index of precision.

		Hsu and Chien [8]	Clustering ensemble				
			KB and PB	KC and PC	KB and PC	KC and PB	KB and PB and KC and PC
Precision	A	0.70	0.84	0.92	0.92	0.92	0.98
	B	0.50	0.66	0.96	0.92	0.62	0.96
	C	0.60	0.64	1.00	1.00	0.60	1.00
	D	0.70	0.98	0.92	0.92	0.98	1.00
	E	0.60	0.94	0.82	0.82	0.98	0.98
	F	0.80	0.98	0.76	0.76	0.98	0.98
	Avg.	**0.65**	**0.84**	**0.90**	**0.89**	**0.85**	**0.98**

TABLE 7: Clustering result on the index of recall.

		Hsu and Chien [8]	Clustering ensemble				
			KB and PB	KC and PC	KB and PC	KC and PB	KB and PB and KC and PC
Recall	A	1.00	1.00	1.00	0.93	1.00	1.00
	B	1.00	0.97	0.7	0.78	0.83	1.00
	C	1.00	0.94	0.67	0.84	0.67	0.97
	D	1.00	0.81	1.00	1.00	1.00	1.00
	E	1.00	0.79	1.00	1.00	1.00	1.00
	F	1.00	1.00	1.00	1.00	1.00	1.00
	Avg.	**1.00**	**0.92**	**0.90**	**0.93**	**0.92**	**1.00**

TABLE 8: Clustering result on the index of F-measure.

		Hsu and Chien [8]	Clustering ensemble				
			KB and PB	KC and PC	KB and PC	KC and PB	KB and PB and KC and PC
F-measure	A	0.82	0.91	0.96	0.92	0.96	0.99
	B	0.67	0.79	0.81	0.84	0.71	0.98
	C	0.75	0.76	0.8	0.91	0.63	0.98
	D	0.82	0.89	0.96	0.96	0.99	1.00
	E	0.75	0.86	0.90	0.90	0.99	0.99
	F	0.89	0.99	0.86	0.86	0.99	0.99
	Avg.	**0.78**	**0.87**	**0.88**	**0.90**	**0.88**	**0.99**

considering the defect density of the spatial relations between the good and bad dies across a wafer. Based on the F-measure, the clustering ensembles obtained using all generated partitions exhibit larger precision and recall values and superior levels of performance regarding each pattern compared with the other methods. Thus, the partitions generated by using k-means and PSO clustering in various data types must be considered.

The practical viability of the proposed approach was examined. The results show that the ART1 neural network performing into data space directly leads to worse clustering performance in terms of precision. However, the true types of WBM can be identified through transforming original data space into label space and performing consensus partition by ART1 neural network. The proposed cluster ensemble approach can get better performance with fewer numbers of clusters than other conventional clustering approaches including k-means, PSO clustering, and ART1 neural network.

5. Conclusion

WBMs provide important information for engineers to rapidly find the potential root cause by identifying patterns correctly. As the driven force for semiconductor manufacturing technology, WBM identification to the correct pattern becomes more difficult because the same type of patterns is influenced by various factors such as die size, pattern density, and noise degree. Relying on only engineers' experiences of visual inspections and personal judgments in the map patterns is not only subjective, and inconsistent, but also very time-consuming and inefficient. Therefore, grouping similar WBM quickly helps engineer to use more time to diagnose the root cause of low yield.

Considering the requirements of clustering WBMs in practice, a cluster ensemble approach was proposed to facilitate extracting the common defect pattern of WBMs, enhancing failure diagnosis and yield enhancement. The advantage of the proposed method is to yield high-quality clusters by applying distinct algorithms to the same data

set and by using various parameter settings. The robustness of clustering ensemble is higher than individual clustering method because the clustering from various aspects including algorithms and parameter setting is integrated into a consensus result.

The proposed clustering ensemble has two stages. At the first stage, diversity partitions are generated using two types of input data: various cluster numbers and distinct clustering algorithms. At the second stage, a consensus partition is attained using these diverse partitions. The numerical analysis demonstrated that the clustering ensemble is superior to using individual k-means or PSO clustering algorithms. The results demonstrate that the proposed approach can effectively group the WBMs into several clusters based on their similarity in label space. Thus, engineers can have more time to focus the assignable cause of low yield instead of extracting defect patterns.

Clustering is an exploratory approach. In this study, we assume that the number of clusters is known. Evaluating the clustering ensemble approach, prior information is required regarding the cluster numbers. Further research can be conducted regarding self-tuning the cluster number in clustering ensembles.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] R. C. Leachman, S. Ding, and C.-F. Chien, "Economic efficiency analysis of wafer fabrication," *IEEE Transactions on Automation Science and Engineering*, vol. 4, no. 4, pp. 501–512, 2007.

[2] C.-F. Chien and C.-H. Chen, "A novel timetabling algorithm for a furnace process for semiconductor fabrication with constrained waiting and frequency-based setups," *OR Spectrum*, vol. 29, no. 3, pp. 391–419, 2007.

[3] C.-F. Chien, W.-C. Wang, and J.-C. Cheng, "Data mining for yield enhancement in semiconductor manufacturing and an empirical study," *Expert Systems with Applications*, vol. 33, no. 1, pp. 192–198, 2007.

[4] C.-F. Chien, Y.-J. Chen, and J.-T. Peng, "Manufacturing intelligence for semiconductor demand forecast based on technology diffusion and product life cycle," *International Journal of Production Economics*, vol. 128, no. 2, pp. 496–509, 2010.

[5] C.-J. Kuo, C.-F. Chien, and J.-D. Chen, "Manufacturing intelligence to exploit the value of production and tool data to reduce cycle time," *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 1, pp. 103–111, 2011.

[6] C.-F. Chien, C.-Y. Hsu, and C.-W. Hsiao, "Manufacturing intelligence to forecast and reduce semiconductor cycle time," *Journal of Intelligent Manufacturing*, vol. 23, no. 6, pp. 2281–2294, 2012.

[7] C.-F. Chien, C.-Y. Hsu, and P.-N. Chen, "Semiconductor fault detection and classification for yield enhancement and manufacturing intelligence," *Flexible Services and Manufacturing Journal*, vol. 25, no. 3, pp. 367–388, 2013.

[8] S.-C. Hsu and C.-F. Chien, "Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing," *International Journal of Production Economics*, vol. 107, no. 1, pp. 88–103, 2007.

[9] C.-F. Chien, S.-C. Hsu, and Y.-J. Chen, "A system for online detection and classification of wafer bin map defect patterns for manufacturing intelligence," *International Journal of Production Research*, vol. 51, no. 8, pp. 2324–2338, 2013.

[10] C.-W. Liu and C.-F. Chien, "An intelligent system for wafer bin map defect diagnosis: an empirical study for semiconductor manufacturing," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 5-6, pp. 1479–1486, 2013.

[11] C.-F. Chien and C.-Y. Hsu, "A novel method for determining machine subgroups and backups with an empirical study for semiconductor manufacturing," *Journal of Intelligent Manufacturing*, vol. 17, no. 4, pp. 429–439, 2006.

[12] K.-S. Lin and C.-F. Chien, "Cluster analysis of genome-wide expression data for feature extraction," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3327–3335, 2009.

[13] J. A. Hartigan and M. A. Wong, "A K-means clustering algorithm," *Applied Statistics*, vol. 28, no. 1, pp. 100–108, 1979.

[14] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, December 1995.

[15] D. W. Boeringer and D. H. Werner, "Particle swarm optimization versus genetic algorithms for phased array synthesis," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 3, pp. 771–779, 2004.

[16] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *The Journal of Machine Learning Research*, vol. 3, no. 3, pp. 583–617, 2002.

[17] A. L. V. Coelho, E. Fernandes, and K. Faceli, "Multi-objective design of hierarchical consensus functions for clustering ensembles via genetic programming," *Decision Support Systems*, vol. 51, no. 4, pp. 794–809, 2011.

[18] A. Topchy, A. K. Jain, and W. Punch, "Clustering ensembles: models of consensus and weak partitions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1866–1881, 2005.

[19] C. H. Stapper, "LSI yield modeling and process monitoring," *IBM Journal of Research and Development*, vol. 20, no. 3, pp. 228–234, 1976.

[20] W. Taam and M. Hamada, "Detecting spatial effects from factorial experiments: an application from integrated-circuit manufacturing," *Technometrics*, vol. 35, no. 2, pp. 149–160, 1993.

[21] C.-H. Wang, W. Kuo, and H. Bensmail, "Detection and classification of defect patterns on semiconductor wafers," *IIE Transactions*, vol. 38, no. 12, pp. 1059–1068, 2006.

[22] J. Y. Hwang and W. Kuo, "Model-based clustering for integrated circuit yield enhancement," *European Journal of Operational Research*, vol. 178, no. 1, pp. 143–153, 2007.

[23] T. Yuan and W. Kuo, "Spatial defect pattern recognition on semiconductor wafers using model-based clustering and Bayesian inference," *European Journal of Operational Research*, vol. 190, no. 1, pp. 228–240, 2008.

[24] R. R. Yager and D. P. Filev, "Approximate clustering via the mountain method," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, no. 8, pp. 1279–1284, 1994.

[25] M.-S. Yang and K.-L. Wu, "A modified mountain clustering algorithm," *Pattern Analysis and Applications*, vol. 8, no. 1-2, pp. 125–138, 2005.

[26] D. W. van der Merwe and A. P. Engelbrecht, "Data clustering using particle swarm optimization," in *Proceedings of the Congress on Evolutionary Computation (CEC '03)*, pp. 215–220, December 2003.

[27] G. A. Carpenter and S. Grossberg, "The ART of adaptive pattern recognition by a self-organization neural network," *Computer*, vol. 21, no. 3, pp. 77–88, 1988.

[28] C. Wei and Y. Dong, "A mining-based category evolution approach to managing online document categories," in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, January 2001.

[29] L. Rokach and O. Maimon, "Data mining for improving the quality of manufacturing: a feature set decomposition approach," *Journal of Intelligent Manufacturing*, vol. 17, no. 3, pp. 285–299, 2006.