A Neural-Network Approach To Recognize Defect Spatial Pattern In Semiconductor Fabrication

Fei-Long Chen and Shu-Fan Liu

***Abstract*—Yield enhancement in semiconductor fabrication is important. Even though IC yield loss may be attributed to many problems, the existence of defects on the wafer is one of the main causes. When the defects on the wafer form spatial patterns, it is usually a clue for the identification of equipment problems or process variations. This research intends to develop an intelligent system, which will recognize defect spatial patterns to aid in the diagnosis of failure causes. The neural-network architecture named adaptive resonance theory network 1 (ART1) was adopted for this purpose. Actual data obtained from a semiconductor manufacturing company in Taiwan were used in experiments with the proposed system. Comparison between ART1 and another unsupervised neural network, self-organizing map (SOM), was also conducted. The results show that ART1 architecture can recognize the similar defect spatial patterns more easily and correctly.**

***Index Terms*—ART1, defects, semiconductor, SOM, spatial pattern recognition, yield.**

I. INTRODUCTION

SEMICONDUCTOR manufacturing has emerged as one of the most important world industries. Even with the highly automated and precisely monitored facilities used to process the complex manufacturing steps in a near particle free environment, processing variations in wafer fabrication still exist. The causes of these variations may arise from equipment malfunctions, delicate and difficult processing steps, or human mistakes. In order to be competitive in the semiconductor manufacturing industry, the detection of these problems becomes a critical issue because yield performance is closely related to the control and efficiency of the wafer manufacturing process.

Today, yield enhancement engineering usually focuses on the investigation of low-yield lots, the elimination of defects, process excursions, the correlation between electrical and functional experiment results, and the improvement of baseline product yield [2]. In general, the main cause of IC yield loss can be attributed to defects on the wafers. A defect is defined as anything that may cause a product to fail, whereas a fault is any form of defect that induces product failure. Defect and fault density requirements vary substantially with the maturity of a process and the minimum feature sizes of the associated product [1]. The occurrence of defects on a wafer may result in the yield loss of a single wafer or, more seriously, an entire wafer lot must be discarded. Usually semiconductor fabs use control charts to monitor the total number of defects found on a wafer. However, this approach is not adequate for efficient process variation detection and the yield may be underestimated [2].

Since all of the yield enhancement tasks require that engineers digest a tremendous amount of data, defect pattern recognition is usually conducted through statistical data analysis. Cunningham [3] classified the common statistics for visual defect metrology into three types.

1) *Quadrate Statistics:* Defects distributed on a wafer are analyzed to predict the yield model. Spatial pattern and defect clustering phenomena are ignored. The occurrence of a defect in any location is usually assumed to be independent of the occurrence of other defects at different locations. Many models [4]–[10] have been based on this type of statistics.
2) *Cluster Statistics:* The data values are the location coordinates of the defects. Since the occurrence of defects may violate the random assumption of the predictive yield model, some research works have focused on the recognition of the defect-clustering phenomenon to enhance the accuracy of yield prediction [11]–[17].
3) *Spatial Point Pattern Statistics:* In addition to defect clusters, the spatial pattern of the defects usually provides a good direction for problem solving. Ken [18] pointed out that special process signatures appearing on the defect map pattern might come from machines or processes. Past experience has also pointed out that when there were problems with machines or products, the clustered defects on the wafer would be distributed in certain patterns. Thus, spatial pattern recognition algorithms are therefore necessary for detecting cluster signatures.

Typical spatial patterns include ring, semiring, scratch, repeat, centralized, radiated, and die-edge defect types. Traditionally, these patterns are recognized by visually reviewing the defects and classifying them according to some predetermined patterns. Disadvantages of this approach include the substantial effort invested in training the defect review/classification task and the high possibility of recognition variability even when inspected by the same operator. For this reason, the development of an automated system is highly desirable.

According to Cunningham's survey [3], most of the existing spatial pattern recognition algorithms can only detect scratch patterns based on the collinear concept. For instance, the defect classification system (DCS-1) developed by ADE cooperation was the first commercially available automated defect

Manuscript received January 6, 1999; revised March 23, 2000. This work was supported by the National Science Council of R.O.C.. under Grant NSC88-2213-E-007-027.

The authors are with the Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan, R.O.C. (e-mail: flchen@IE.nthu.edu.tw).

Publisher Item Identifier S 0894-6507(00)06371-5.

0894–6507/00$10.00 © 2000 IEEE

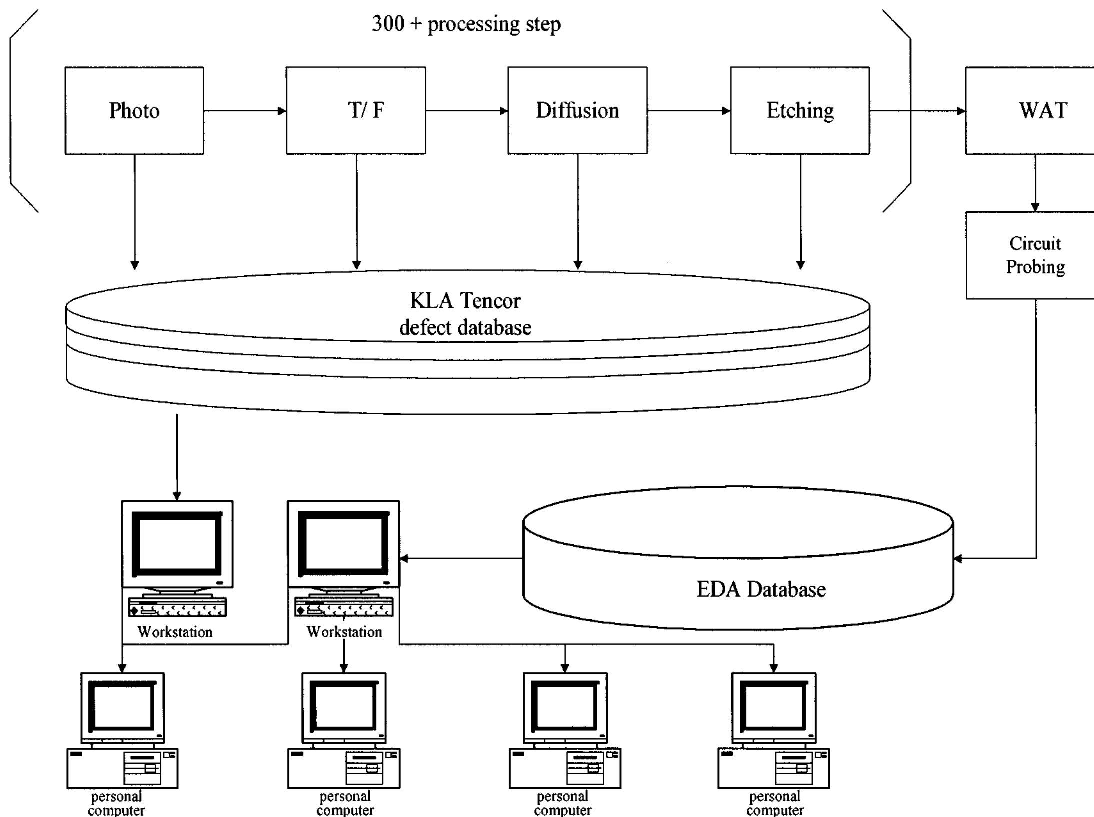

Fig. 1. Flow of data analysis.

classification tool [19]. This system combined the image processing techniques and fuzzy logic expert system to recognize the scratch pattern or other patterns described by users. Duvivier [20] developed a statistical method to detect and classify spatial defect patterns. In his methods, all the wafermaps were examined to generate the so-called random ratios (RR). If all the failing die were caused by a spatial signature, then RR= 0. Otherwise, if they were all spatially independent, then RR= 1. After that, a segmentation technique was applied to further describe the nature of the detected signature. The major limitation of this approach is that different predetermined criteria will be required for the detection of different defect patterns. Lee *et al.* [21] presented a computer-based pattern matching algorithm for the defect pattern detection. In the pattern matching procedure, a supervised learning concept was adopted. Enough standard or representative training templates must be provided in order to obtain good defect pattern recognition. This became the main shortcoming of their method. Knights Technology has announced a software tool named "spatial pattern recognition (SPaR) using the defect map pattern analysis on semiconductor wafers. A major component of this software is a signature classifier, which can be trained by users to build up the knowledge base. The algorithm behind the software was developed at Oak Ridge National Laboratories. The major shortcoming of this algorithm is the tremendous amount of time consuming in training new patterns. The NeuralNet^{TM} Engineering Data Analysis (NEDA) developed by Defect & Yield Management (DYM), Inc. applied neural-network techniques to detect the similar patterns. Again, enough templates must be provided to train the knowledge base.

In viewing of the limitations of the above methods, this research intends to develop an intelligent algorithm for detecting a greater number of differing spatial patterns on a wafer. In order to speed up the detection process, neural-network architecture named adaptive resonance theory (ART1) was adopted in this research.

II. Defect Spatial Pattern Recognition

A. Data Collection and Transformation

There are over 300 steps in the semiconductor manufacturing process. For the purposes of maintaining quality and yield, some inspection stations are established along some of the steps in this process. Usually, the most critical processing steps or the steps processed by machines with a higher probability of causing problems receive the highest priority for inspection. Normally, about ten inspection stations are established on most product lines. The usual machines used for defect inspection include KLA, Tencor, and Orbot. These machines can detect visual defects as small as 0.20 μm. Proper inspection machines are installed to collect defect data according to each machine inspection properties and capabilities. Fig. 1 shows the flow of data analysis.

Before analysis of the collected defect data can be performed, transformation of the data coordinates is necessary. Though the data collection format is the same, different inspection machines

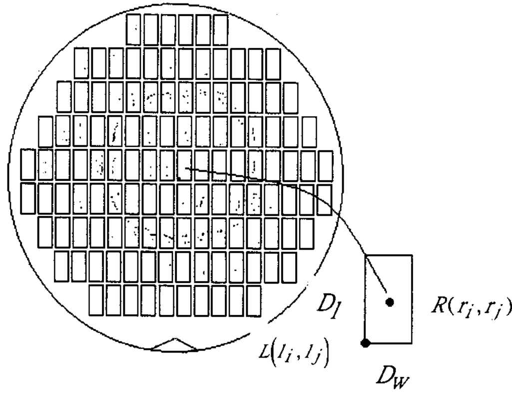

Fig. 2. Representation of defect data relative location.

produce different original coordinates. Before storing the collected data into a neutral data set, it is necessary to find out the relative defect location and transform it into a unique coordinate system. Take Fig. 2 as an example. In this figure, $L(l_i, L_j)$ means the original x, y-axis coordinates of a die on the wafer and $R(r_i, r_j)$ is the actual x, y axis location in that die. The coordinate transformation can be executed using the following process:

$$T(i, j) = T(l_i * D_l + r_i, l_j * D_w + r_j)$$

where

- l_l original x axis coordinates of a die on the wafer;
- l_j original y axis coordinates of a die on the wafer;
- D_l length of the die dimension;
- D_w width of the die dimension;
- r_i actual x axis locations on the die;
- r_j actual y axis locations on the die.

B. Design the Input Vector

The input vector of the training samples is also named the characteristic vector. The number of processing units depends upon the type of problems to be studied. A linear transformation function is usually used to pass the input vector into the next layer. The design of the input vector differs for every product type. The number of dies in a specific product type determines the number of nodes in the input layer. The detailed notations are explained below and represented in Fig. 3.

- N number of dies per wafer;
- X_i input vector of the ith sample data (wafer);
- x_{ij} jth element of the input vector.

where $X_i = (x_{i1}, x_{i2}, x_{i3}, \cdots x_{iN})$

$$\begin{aligned} x_ij &= 1, \quad \text{if defect occurs on a die;} \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

After the sample training data needed for the unsupervised neural network has been provided the number of nodes in the input layer and their corresponding values must be defined to start the training process. In this research, the unsupervised neural network was trained by product type. The reasons are as follows.

1) The number of input processing units is the total number of dies for a wafer. Different product types have different numbers of dies for the wafer. The collection of weights must be prepared by product type.
2) When the network is trained by product type, this research is extendable to a correlation with the circuit probe (CP) data.
3) Even a single input pattern can be classified. Insufficient data was not a concern during the neural-network architecture.
4) Even though the life cycles of certain products may not be long, a considerable number of wafers will be produced in fabrication. Pattern type is defined as the key field in the knowledge base design, therefore the limits of training the network by product type were eliminated.

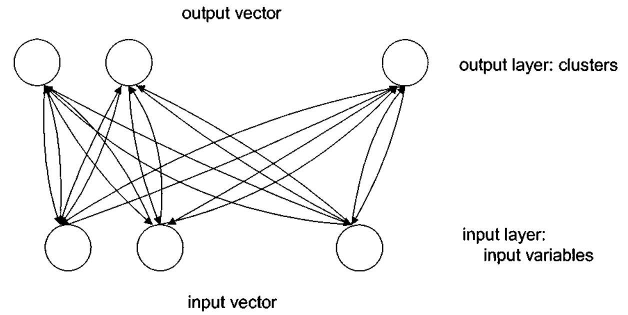

Fig. 3. ART I network connections.

C. ART1 Network Model

For the huge amount of defect map data, it is difficult to decide how many clusters of defect spatial patterns in the semiconductor manufacturing. For this reason, learning was accomplished by the input data alone since the number of output patterns is unpredictable. This type of learning is so-called **unsupervised learning**.

ART1 is an unsupervised network that accepts binary inputs [22]. A good knowledge-based system has to satisfy two characteristics: stability and plasticity. ART1 uses a vigilance test to learn new patterns without forgetting old knowledge and thus can solve the contradiction between stability and plasticity. The concept of the vigilance test is described as follows.

1) If the characteristic of a new pattern is quite similar to a previously stored pattern (vigilance test passed), only a slight modification of the knowledge contained in the old patterns will be executed. The characteristics of the old and new patterns can be satisfied and the old knowledge can be properly retained. Stability of the system can be maintained.
2) If the characteristics of a new pattern are not similar to all of the previously stored patterns (vigilance test failed), new knowledge for the new pattern will be created. This implies quick learning of a new pattern, or the so-called plasticity.

Because of the above two characteristics, ART1 was adopted in this research to detect and recognize spatial defect patterns. The construction of ART1 architecture includes an input layer, network connection, and output layer (see Fig. 4).

There are two types of weight connections between every input unit and output unit. The matched weight is from the input

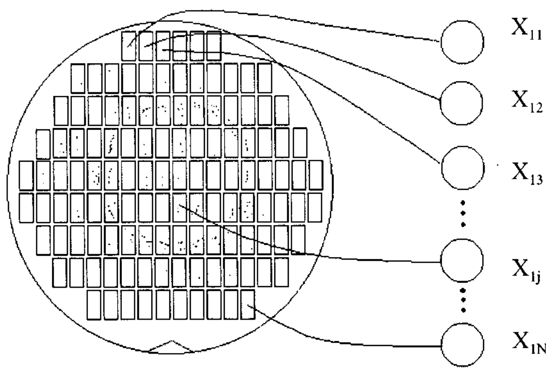

Fig. 4. Relationship between input vector and dies.

layer to the output layer while the similar weight is from the output layer to the input layer.

ART 1 uses an output-processing unit to present a certain cluster. Every connection weight between the input layer and the output units indicates the characteristic of a specific cluster. The number of output processing units passing the vigilance test may exceed one so the network utilizes the match value to control the output processing units. The vigilance test is first applied to the output processing units with the highest match value. In general, the higher the match value possessed by an output-processing unit, the higher its similarity. The output-processing unit with the highest similarity is not always the one with the highest match value.

The major characteristic of ART1 algorithm is the vigilance value, which can be used to distinguish the similar patterns. The vigilance test is first applied to the output processing units with the highest matching value. However, the output-processing unit with the highest similarity is not always the one with the highest matching value. When a high vigilance value is assigned, few output units will pass the test and more output units will therefore be created. On the contrary, the lower the assigned vigilance value is, the fewer output units will be. So the ART1 network is capable of detecting similar but different types of clusters. The implementation procedure for ART1 algorithm is listed in Appendix A.

III. Data Generation and Network Training

After the conceptual design of the intelligent defect recognition system, a practical software system was developed for system implementation and verification. This system was developed using Borland C++ and SAS version 6.12, under a Microsoft Windows 95 operating platform. Actual data from a product with 294 dies for a wafer were provided by a semiconductor company and tested through this system. Before the ART1 network can be used to identify the spatial pattern types on the wafers, the network must be trained. Due to the difficulty in collecting sufficient defective data, sample data was created for ART1 neural-network training. Since ART1 is an unsupervised network, there was no need to link the input and output vectors to attain good recognition. Instead, the input vectors were designed to represent a symbolic pattern. This makes it possible to train the network even without actual defect

TABLE I
Patterns of the 35 Samples

Pattern	*Type*	*Numbers of sample data*
Ring	3.5 cm	5
	4.5 cm	5
	5.5 cm	5
Scratch	Right to up	5
	Right to down	5
	Left to up	5
	Left to down	5

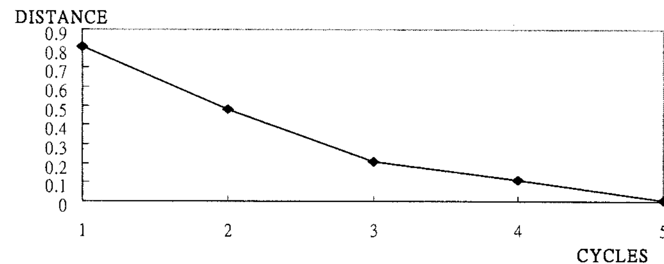

Fig. 5. Converge situation in ART 1 network.

patterns. The recognition performance can be further enhanced when actual defective data are collected and used for training.

At the current stage, the training samples contain only the two most frequent patterns, i.e., ring and scratch. The ring type patterns can be divided into three types of different sizes and the scratch type patterns can be divided into four types. The system is trained on each type of defect using five data samples. These samples are summarized in Table I.

The distance between the input nodes and output nodes evaluated the convergence of the network. The adjustment of the vigilance value helps control the number of output nodes. In this research, the vigilance value was set at 0.11 and the output nodes were equal to seven. During the training processes, the ART1 network converged in five cycles as Fig. 5 shows. The time utilized for training 35 samples on a PC with an Intel Pentium 166 and 48MB RAM was approximately 3 s.

To evaluate the training performance of ART1 network, another unsupervised network, Kohonen self-organizing map (SOM), is selected for comparison. SOM accepts continuous inputs and its goal is to map an n-dimension input space into a one or two-dimension output layer such that a meaningful topology exists within the output nodes [23]. The procedure for generating this network is summarized in Appendix B. After training with the same data set, the convergence condition of the SOM network is depicted in Fig. 6. From this comparison, ART1 obviously converges much faster than SOM in terms of data training.

After training the two networks, 35 simulated testing data were then applied to test whether the defect maps can be correctly recognized. The results showed that ART1 required less learning time than SOM. The time consuming in training 35 samples on a PC with an Intel Pentium 166 and 48 MB RAM

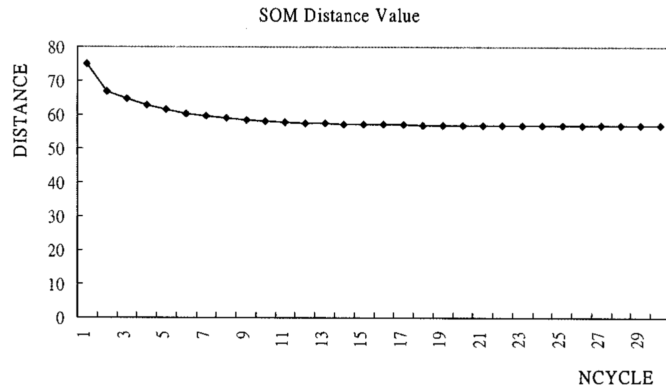

Fig. 6. Converge situation in SOM network

was approximately 3 s for ART1 and 30 s for SOM. Though SOM can recognize different defect maps, it is not capable of detecting similar defect maps such as the ring type with different radii. ART1 can not only correctly recognize different defect maps, but also be capable of distinguishing similar but different defect maps. Therefore, ART1 is more adequate for the recognition of defects.

IV. Experimental Results

In this section, the pretrained ART1 network is used to recognize real defect maps. A semiconductor company provided 14 actual data from a DRAM product with 294 dies for a wafer. Twelve of these were visually judged to have ring type patterns and the other two exhibited random type patterns. Because the total number of dies for this DRAM product type was 294, the number of input nodes was 294 in the ART1 network. It is expected that the number of outputs corresponding to the number of patterns would be seven. The adjustment of the vigilance value in the network learning stage helps control the number of output nodes. In this research, the vigilance value was set at 0.11 and the output nodes were equal to seven.

With the trained ART1 network, every new pattern was classified according to the maximum match value. The match value indicated the degree of match with the recognized spatial patterns. When all of the matched values were smaller than a predetermined threshold, P_m, the input pattern could not be classified into any specific cluster. If this value were set high, the maps classified into the same cluster would have a very similar pattern. When the value was set to one, only the completely identical maps would be classified into one cluster. In this research, the threshold was empirically determined to be 0.3.

In the following are the four possible situations in which an input pattern would be fed into the ART1 network (see Fig. 7).

1) In maps 1 and 2, both ring and scratch type defects are recognized. The scratch type defect received a higher match value (0.6524). Fig. 8 depicts the two inspected maps and the trained scratch pattern. It can be observed from this figure that the scratch-type defect is usually part of the ring-type defect. So when a map is recognized to have significant match values for both types of defect, it is classified as a ring type.

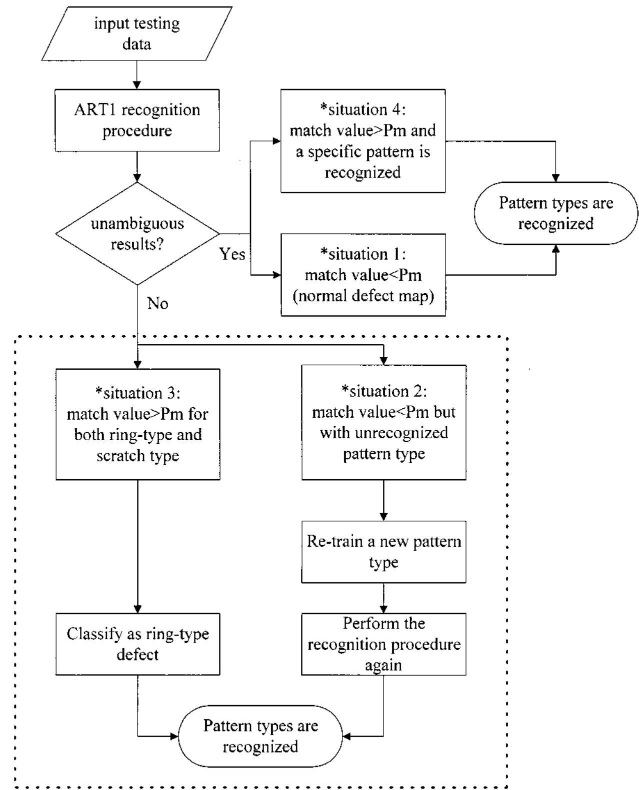

Fig. 7. Possible results of ART 1 analysis.

2) For maps 3 and 4, special signatures were detected, but could not be recognized by the ART1 network. The most possible reason is insufficient sample data. In other words, this particular ring size had not been trained into the ART1 procedure. For this reason, map 3 was treated as sample data and sent into the ART1 pretraining procedure to create a new pattern type. After this retraining procedure, map 4 could then be successfully recognized with a match value of 0.9873.

Case 1: Match value $<P_m$ (normal defect map)

No signature for clustered defects existing in a wafer. Defects fall on the wafer randomly or cluster in a small area without any significant pattern.

Case 2: Match value $<P_m$ (unrecognized pattern type)

There are situations when special signatures exist but cannot be recognized by the ART1 network. This is because the system has not been trained on certain important defect patterns. When an unrecognized pattern type is encountered, the ART1 network should be retrained.

Case 3: Match value $>P_m$ and more than one pattern is recognized

It is possible that two or more match values are greater than P_m, and there is no significant difference existing between these match values. Again, the input pattern can be sent back to the ART 1 and the output-processing unit readjusted for further analysis.

Case 4: Input pattern match values $>P_m$ and a specific pattern type is recognized.

After inputting the 14 test maps into the trained ART1 network, the testing results can be generated within seconds.

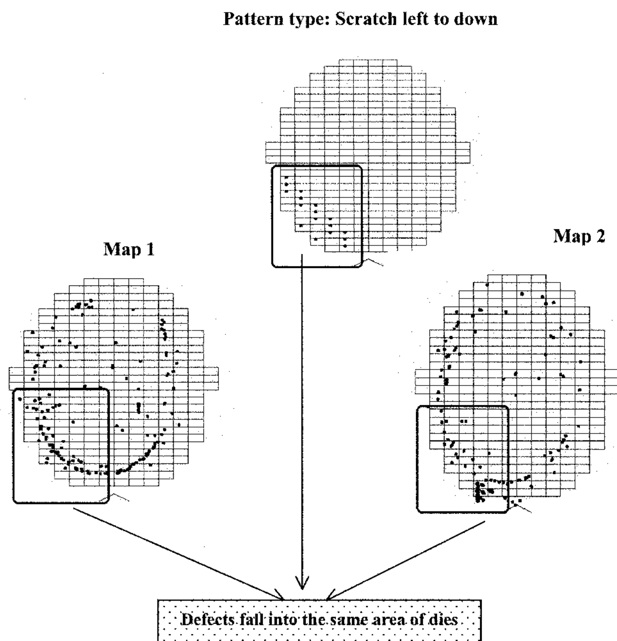

Fig. 8. Misjudgment of defect pattern.

TABLE II
ART I RECOGNITION RESULTS

MAP NO.	*MAX. MATCH VALUE*	*RECOGNIZED SITUATION*	*RECOGNIZED RESULTS*
1	0.65240	Situation 3	Scratch left to down
	0.57142		Ring 5.5cm
2	0.65240	Situation 3	Scratch left to down
	0.57142		Ring 5.5cm
3	0.09231	Situation 2	New pattern
4	0.06154	Situation 2	New pattern
5	0.57142	Situation 4	Ring 5.5cm
6	0.49232	Situation 4	Ring 3.5cm
7	0.40010	Situation 4	Ring 3.5cm
8	0.38571	Situation 4	Ring 5.5cm
9	0.85713	Situation 4	Ring 5.5cm
10	0.71269	Situation 4	Ring 4.5cm
11	0.46155	Situation 4	Ring 3.5cm
12	0.57142	Situation 4	Ring 5.5cm
13	0.09196	Situation 1	Random Type
14	0.28571	Situation 1	Random Type

Table II summarizes the results obtained. It can be observed from this table that maps 5–14 were correctly recognized. The reason why maps 1–4 could not be correctly recognized is explained as below.

From the experimental results in the 14 defect maps above, the method developed achieved the expectation of automatically recognizing the spatial patterns of clustered defects. The diagnosed results can help engineers determine which processing steps or machines in the fabrication process induced such spatial patterns.

V. CONCLUSIONS

In the semiconductor industry, the primary cause of IC yield loss can be attributed to defects on the wafer. In practice, engineers usually spend much time checking entire defect maps in lots and choose the maps having clustered defect spatial signatures. When these defects are clustered, the size and shape of the spatial pattern indicates specific process problems. Because the patterns are not well defined, the similarities between these patterns are difficult to decide. Without an automated approach, however, gathering and analyzing the defect data can take days or even weeks in some cases. In view of this, this research developed an intelligent system, which can recognize the spatial patterns of clustered defects to help in the diagnosis of possible failure causes. The system features a modular structure and incorporates the ART1 technique. The experimental results show that this approach provides not only the automated classification of known patterns but also the detection of new unknown patterns. When training the new patterns, ART1 consumes less time in comparison with the SOM architecture. The major restriction of the ART1 network is its limited capability for die-level only, i.e., it can classify patterns of defective die, but not patterns of defects.

Actual data obtained from a semiconductor manufacturing company were tested through this system. All of the actual defect maps could be recognized and discrimination was accomplished between the systematic and random type defects. Due to the difficulties in collecting actual data from semiconductor manufacturing companies, the inclusion of other types of spatial pattern defects will be the future extension of this research. Another possible extension is the incorporation of CP (circuit probe) maps and defect knowledge to increase the capability for recognizing defect spatial patterns and determining the possible causes in the process.

APPENDIX A

The ART I algorithm can be expressed in the following steps [21]:

Step 1. Initially the weight bij are initialized to the same low value which should be

$$b_{ij} < \frac{L}{(l-1+m)}$$

where m is the number of components in the input vector and L is a constant, typically $L = 2$.

Step 2. When an input pattern, $\boldsymbol{X}$, is presented to the network, the recognition layer selects the winner as the maximum of all the net outputs:

$$net_j = \sum_{i=1}^{N} b_{ij} c_i$$

where N is the number of neurons in the comparison layer.

```
Step 3.  Perform the vigilance test. A
  neuron j is declared to pass the vigi-
  lance test, if and only if,
```

$$\sum_{i=1}^{((net_j)/N)>\rho} X_i$$

```
  where ρ is the vigilance threshold.
Step 3a.  If the winner fails the test,
  mask the current winner and go to Step 2
  to select another winner.
Step 3b.  Repeat the cycle (Step 1 through
  2a) until a winner is determined that
  passes the vigilance test, then go to
  Step 5.
Step 4.  If no neuron passes the vigilance
  test, create a new neuron to accommodate
  the new pattern.
Step 5.  Adjust the feedforward weights
  for the winner neuron. Update the feed-
  back weights from the winner neuron to
  its inputs.
```

APPENDIX B

The Kohonen SOM algorithm can be expressed in the following steps [21]:

```
Step 1.  Initialization:
  Initialize the weight vectors Wj(O), the
  leaming rate η(O) and the neighborhood
  function Ai(x)(0). Both learning rate and
  neighborhood function should be large
  initially.
Step 2.  For each vector x in the samples,
  perform steps 2a, 2b and 2c.
Step 2a.  Place the sensory stimulus
  vector, x onto the input layer of the
  network.
Step 2b.  Similarity matching:
  Select the neuron whose weight vectors
  best matches x as the winning neuron.
  Using the Euclidean criteria, the index
  of the winning neuron will be
```

$$i(x) = k \quad \text{where } \|W_k - x\| < \|W_j - x\| \qquad j = 1, 2, \cdots, n$$

```
Step 2c.  Training:
  Train the weight vectors such that neu-
  rons within the activity bubble are
  moved toward the input vector as fol-
  lows:
```

$$W_j(n+1) = \{W_j(n) + \eta(n)[x - W_j(n)]\} \qquad j \in \Lambda_{i(x)}(n).$$

```
Step 3.  Update the learning rate, η(n):
  A linear decrease of the leaming rate
  should produce satisfactory results.
Step 4.  Reduce the neighborhood function,
  Ai(x)(n).
Step 5.  Check stopping condition:
  Exit when no noticeable change to the
  feature map has occurred.
  Otherwise go to Step 2.
```

REFERENCES

[1] C. Weber, B. Moslehi, and M. Dutta, "An integrated framework For yield management and defect/fault reduction," *IEEE Trans. Semiconduct. Manufact.*, vol. 8, pp. 110–120, May 1995.

[2] W. Shindo, E. H. Wang, R. Akella, and A. J. Strojwas, "Effective excursion detection by defect type grouping in in-line inspection and classification," *IEEE Trans. Semiconduct. Manufact.*, vol. 12, pp. 3–9, Feb. 1999.

[3] S. P. Cunningham and S. MacKinnon, "Statistical methods for visual defect metrology," *IEEE Trans. Semiconduct. Manufact*, vol. 11, pp. 48–53, Feb. 1998.

[4] C. N. Berglund, "A unified yield model incorporating both defect and parametric effects," *IEEE Trans. Semiconduct. Manufact*, vol. 9, pp. 447–454, Aug. 1996.

[5] R. S. Collica, "The physical mechanisms of defect clustering and its correlation to yield model parameters for yield improvement," in *1990 IEEE/SEMI Advanced Semiconduct. Manufact. Conf.*, 1990, pp. 91–94.

[6] ——, "The effect of the number of defect mechanisms on fault clustering and its detection using yield model parameters," *IEEE Trans. Semiconduct. Manufact.*, vol. 5, pp. 189–195, Aug. 1992.

[7] P. Wang, R. Chan, R. Goodner, F. Lee, and R. Ceton, "Development of the yield enhancement system of a high-volumes 8-in wafer fab," in *IEEE Int. Symp. Semiconductor Manufact.*, 1995, pp. 51–52.

[8] C. Weber, B. Moslehi, and M. Dutta, "An integrated framework for yield management and defect/fault reduction," *IEEE Trans. Semiconduct. Manufact.*, vol. 8, pp. 110–120, May 1995.

[9] R. K. Nurani, A. J. Strojwas, W. P. Maly, C. Ouyang, W. Shindo, R. Akella, M. G. Mcintyre, and J. Derrett, "In-line yield prediction methodologies using patterned wafer inspection information," *IEEE Trans. Semiconduct. Manufact.*, vol. 11, pp. 40–47, Feb. 1998.

[10] P. Fang, "Yield modeling in a custom IC manufacturing line," in *1990 IEEE/SEMI Advanced Semiconduct. Manufact. Conf.*, 1990, pp. 88–90.

[11] J. A. Cunningham, "The use and evaluation of yield models in integrated circuit manufacturing," *IEEE Trans. Semiconduct. Manufact.*, vol. 3, pp. 60–71, May 1990.

[12] A. V. Ferris-Prabhu, "A clustered-modified poisson model for estimating defect density and yield," *IEEE Trans. Semiconduct. Manufact.*, vol. 3, pp. 54–59, May 1990.

[13] D. J. Friedman, M. H. Hansen, V. N. Nair, and D. A. James, "Model-free estimation of defect clustering in integrated circuit fabrication," *IEEE Trans. Semiconduct. Manufact*, vol. 10, pp. 344–359, Aug. 1997.

[14] F. J. Meyer and D. K. Pradhan, "Modeling defect spatial distribution," *IEEE Trans. Comput.*, vol. 38, pp. 538–546, Apr. 1989.

[15] P. R. Pukite and C. L. Berman, "Defect cluster analysis for wafer-scale integration," *IEEE Trans. Semiconduct. Manufact.*, vol. 3, pp. 128–135, Aug. 1990.

[16] C. H. Stapper, "The effects of wafer to wafer defect density variations on integrated circuit defect and fault distributions," *IBM J. Res. Development*, vol. 29, no. 1, pp. 87–97, Jan. 1985.

[17] A. Tyagi and M. A. Bayoumi, "Defect clustering viewed through generalized poisson distribution," *IEEE Trans. Semiconduct. Manufact.*, vol. 5, pp. 196–206, Aug. 1992.

[18] R. Ken, S. Brain, and H. Neil, "Using full wafer defect maps as process signatures to monitor and control yield," in *IEEE/SEMI Semiconductor Manufacturing Sci. Symp.*, 1991, pp. 129–135.

[19] L. Breaux and B. Singh, "Automatic defect classification system for patterned semiconductor wafers," in *1995 Int. Symp. Semiconduct. Manufact.*, 1995, pp. 68–73.

[20] F. Duvivier, "Automatic detection of spatial signature on wafermaps in a high volume production," in *1999 Int. Symp. Defect Fault Tolerance VLSI Syst.*, 1999, pp. 61–66.

[21] F. Lee, A. Chatterjee, and D. Croley, "Advance yield enhancement: Computer-based spatial pattern analysis—Part 1," in *1996 IEEE/SEMI Advance Semiconduct. Manufact. Conf.*, 1996, pp. 409–415.

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

[23] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.
[24] J. A. Freeman and D. M. Skapura, *Neural Networks Algorithms, Applications, and Programming Techniques*. New York: Addison-Wesley, 1992.
[25] A. S. Pandya and R. B. Macy, *Pattern Recognition with Neural Networks in C++*. Piscataway, NJ: IEEE Press, 1996, pp. 319–371.

Fei-Long Chen received the B.S. degree in industrial engineering from National Tsing-Hua University (NTHU), Taiwan, R.O.C., in 1982, and the M.S. and Ph.D. degrees in industrial engineering from Auburn University, Auburn, AL, in 1988 and 1991, respectively.

He has been with the Department of Industrial Engineering, NTHU, since 1991. His current research interests include manufacturing automation, computer-integrated manufacturing, automated inspection, engineering data analysis for semiconductor manufacturing, and enterprise resource planning.

Shu-Fan Liu received the B.S. degree in industrial engineering and management from National Yunlin University of Science and Technology, Yunlin, Taiwan, R.O.C. She received the M.S. degree in industrial engineering from National Tsing-Hua University, Hsinchu, Taiwan, where she is pursuing the Ph.D. degree.

Her research interests include artificial neural network and semiconductor yields analysis.