INT. J. PROD. RES., 2002, VOL. 40, NO. 10, 2207–2223

Taylor & Francis
Taylor & Francis Group

Wafer bin map recognition using a neural network approach

S. F. LIU†, F. L. CHEN†* and W. B. LU‡

Although the fabrication of modern integrated circuits uses highly automatic and precisely controlled operations, equipment malfunctions or process drifts are still inevitable owing to the high complexity involved in the hundreds of processing steps. To detect the existence of these problems at the earliest stage, some important analytical tools must be applied. Among them is wafer bin map analysis. When the bin map exhibits specific patterns, it is usually a clue that equipment problems or process variations have occurred. The aim was to develop an intelligent system that could automatically recognize wafer bin map patterns and aid in the diagnosis of failure causes. A neural network architecture named Adaptive Resonance Theory Network 1 was adopted for the purpose. Actual data collected from a semiconductor manufacturing company in Taiwan were used for system verification. Experimental results show that with an adequate parameter, the neural network can successfully recognize and distinguish random and systematic wafer bin map patterns.

1. Introduction

In many countries, such as the USA, Japan, South Korea and Taiwan, semiconductor manufacturing has emerged as one of the most important industries. However, complexity and cost continue to increase with each new generation of semiconductor products. Semiconductor companies are being forced to improve their manufacturing capability and develop their own techniques to create process steps with tighter tolerances and smaller geometrics on larger dies. To compete in this world-wide market, quality improvement and yield enhancement have received increasing attention. For these reasons, every semiconductor manufacturing company makes an effort to monitor and control the manufacturing processes to reduce variations and enhance yield (Mirza *et al.* 1995).

Typically, all dies on a wafer after the completion of the wafer manufacturing process, must go through the so-called circuit probe (CP) test. The purpose of the CP is to test the electrical functions of a die to determine the inferior dies on each wafer. CP consists of serial pass-or-fail functional tests. The CP results determine the grade of each die on a wafer. Each grade will be presented using a unique bin code. In general, for different product types and different companies, the number of die grades differs. Five, 10 or even 75 die grades are possible (Cox and Reynolds 1996). A good die can pass the functional tests thoroughly and be assigned with the best grade. Other bin code assignments usually indicate different degrees of quality inferiority. Table 1 illustrates an example of CP results. Here, all of the

Revision received July 2001.

† Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

‡ Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, Taiwan, R.O.C.

* To whom correspondence should be addressed. e-mail: flchen@IE.nthu.edu.tw

International Journal of Production Research ISSN 0020–7543 print/ISSN 1366–588X online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00207540210122275

Bin	Count	Yield (%)
DB 1 GOOD	875	78
DB 5 OPEN/SHORT	4	0
DB 6 I/P LEAKAGE HIGH	2	0
⋮		
DB 16 GROSS FUNCTION (6.5V)	5	0
DB 18 CKB FUNCTION (4V)	22	1
⋮		

Table 1. Example of CP results.

dies are bad except for those with Bin Code 1. After all of the dies in a wafer have been CP-tested, a so-called wafer bin map (WBM) can be generated to depict the distribution of quality grades. Since the occurrence of any quality inferiority can usually be attributed to some specific causes, it is believed that WBM analysis can help determine the possible causes of failures and help to devise solutions to prevent the reoccurrence of these failures.

Patterns existing in WBMs may be classified into random and systematic types (Kaempf 1995) (figure 1). Typical examples of systematic WBMs include cluster, circle, arc, doughnut and Bull's-eye patterns. Traditionally, the analysis of these patterns relies heavily upon the engineers' experience and domain knowledge. Since this 'pocket' knowledge is difficult to accumulate, several advanced techniques have been developed to aid in the diagnosis of WBM patterns. These techniques can be categorized into statistic and artificial intelligence approaches.

1.1. *Statistic approach*

It presented a join-count statistics based spatial clustering method to analyse WBMs in two-dimensional space. Suppose that the ith die is the centre of a WBM, the relationship between this die and the four neighbouring dies could be defined as follows:

$$J(\mathrm{GG}) = \sum \sum_{i<j} \delta_{ij} Y_i Y_j$$

$$J(\mathrm{GB}) = \sum \sum_{i<j} \delta_{ij} (Y_i(1 - Y_j) + (1 - Y_i) Y_j)$$

$$J(\mathrm{BB}) = \sum \sum_{i<j} \delta_{ij} (1 - Y_i)(1 - Y_j),$$

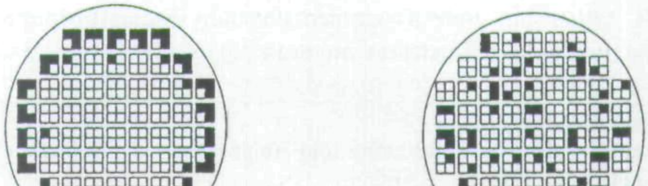

Figure 1. Systematic (left) and random (right) WBMs (Kaempf 1995).

where, $J(\mathrm{GG})$, $J(\mathrm{GB})$ and $J(\mathrm{BB})$ are the number of times that two neighbouring dies are good, one is good while the other is bad, and all are bad, respectively. If $Y = 1$, die i is good. If $Y = 0$, die i is bad. δ_{ij} is a neighbouring index, and $= 1$ if dies i and j are neighbours. Otherwise, $\delta_{ij} = 0$. The index below evaluates whether the dies on WBMs are scattered randomly:

$$\hat{\theta} = \log\left(\frac{(J(\mathrm{GG}) + 0.5)(J(\mathrm{BB}) + 0.5)}{(J(\mathrm{GB})/2 + 0.5)^2}\right).$$

When $\hat{\theta}$ approaches 0, the dies are randomly scattered on the WBMs and no systematic patterns exist. When $\hat{\theta} > 0$, the dies are scattered systematically (Taam and Hamada 1993). A similar approach was also adopted by Kaempf (1995) to determine the distribution of dies on WBMs using a binomial test.

Mirza *et al.* (1995) divided the patterns on WBMs into gross (systematic) and local (random) types and proposed methods to distinguish these two types of patterns. Gross failure means that a large die area or even all of the dies on a wafer are bad. This kind of failure may be caused by temperature variations during processing or other serious reasons. For Local failure, small die areas are bad. Pinholes from particles or spot defects are usually the main causes. Mirza *et al.* used a Gibbs/Markov random field (G/MRF) model to develop a Spatial Randomness Test methodology. Because the G/MRF model may be used to determine object independence or dependence, it can recognize gross failure and local failure on WBMs.

Friedman *et al.* (1997) developed a two-stage spatial pattern analysis approach. In the first stage, good or bad die images (white and black; 0 and 1) were preceded with an averaging operation where each die was averaged by its 3×3 or 5×5 neighbourhoods. The original die is then replaced by the obtained average value. A smoothed grey-level WBM could be obtained after this stage. A threshold value must be selected in the second stage. If the grey level of a die exceeds the threshold value, it is determined to be bad; otherwise, it is good.

It must be noted that although the statistical based approaches can detect the existence of systematic patterns, they lack the capability to identify the correct pattern types. It is therefore difficult to devise solutions for the detected failures.

1.2. *Artificial intelligent approach*

Lin (1998) used machine vision techniques to preprocess WBMs. A supervised neural network architecture (Back Propagation) was then applied to recognize the systematic patterns. In his research, three limitations were determined, which are discussed as follows.

- The image-based approach requires much storage space. It also requires long processing time in the preprocessing stage.
- In the back propagation (BP) architecture, the number of output nodes must be defined in advance. In fact, a skilled manufacturing engineer cannot be sure how many systematic patterns actually exist. If new patterns occur, they will not be detected and recognized.
- Sufficient symbolic training patterns must be provided. A tremendous amount of time will be consumed in training new patterns.

Similar to Lin's work, software named NeuralNet$^{\mathrm{TM}}$ Engineering Data Analysis (NEDA), developed by Defect & Yield Management, Inc. (DYM), was applied to neural network techniques to detect similar patterns. Enough templates must be provided to train the knowledge base. Meanwhile, users cannot retrain new WBM patterns and hence the flexibility is limited.

In view of the limitations in the above methods, this research was intended to develop an intelligent algorithm for detecting a greater number of different WBMs. To speed up the detection process and to provide flexibility for new pattern learning, a neural network architecture named Adaptive Resonance Theory (ART1) was adopted in this research. Details about the development of this algorithm are presented below.

2. Recognition of patterned wafer bin maps

2.1. *Design the input vector*

Before the collected wafer bin maps can be analysed, the input data format design is crucial for sample pretraining and future recall procedures. The input training sample vector 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. In solving a problem, in which the recognition performance may be influenced by too much noise (Lin 1998), a noise reduction procedure was first applied to all of the training samples. The input vector design 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 shown in figure 2:

N number of dies per wafer,
X_i input vector of the ith sample data, and
x_{ij} jth element of the input vector,

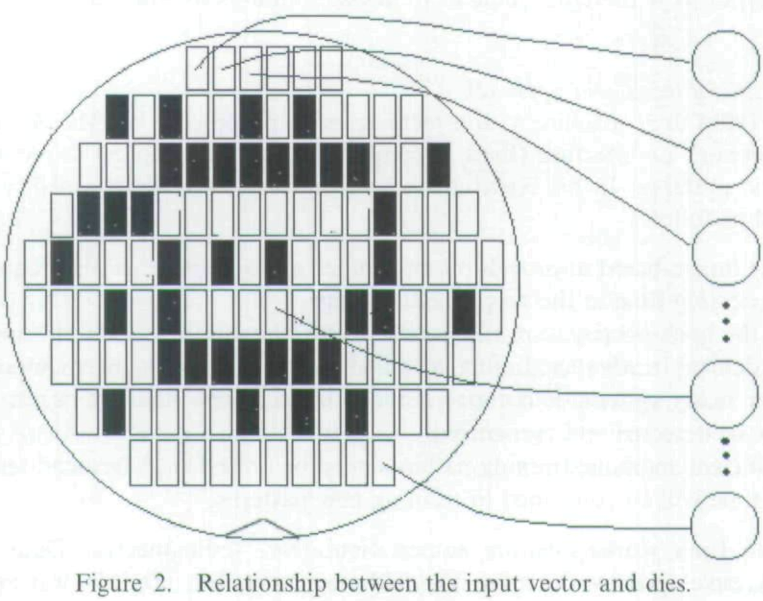

Figure 2. Relationship between the input vector and dies.

where

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

$$x_{ij} = \begin{cases} 1, & \text{if the bin code is not equal to 0;} \\ 0, & \text{otherwise.} \end{cases}$$

After the required sample training data for the 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 according to different product types. The reasons are as follows.

- The number of input processing units is the total number of dies for a wafer. Different product types have different numbers of dies for each type of wafer. The collection of weights must be prepared by product type.
- Since the unsupervised neural network can also classify a single input pattern, data insufficiency is not a critical concern in the neural network architecture. After training with the available samples, any type of pattern can be recognized.
- Although the life cycles of certain products may not be long, a considerable number of wafers will be produced in fabrication. Pattern recognition training is also required for these types of products.

2.2. *Selection of network model*

Because of the huge number of WBMs, it is difficult and time consuming to decide how many types of patterns may appear. That is, the number of output patterns is unpredictable. For this reason, network learning can only be accomplished using the input data alone. This is the reason why unsupervised learning was adopted. Chen and Liu (2000) used ART1 and SOM architectures to compare the pattern recognition performance on defect maps (under wafer level consideration). The results showed that ART1 performed better than SOM. Therefore, among the many kinds of unsupervised neural networks architectures, the Adaptive Resonance Theory (ART) was considered in this research. The ART network accepts input vectors that are classified according to the stored pattern they most resemble. The ART network also provides a mechanism allowing adaptive expansion of the neuron output layers until an adequate size is reached based on the number of classes. The ART network can adaptively create a new neuron corresponding to an input pattern if that pattern is determined to be 'sufficiently' different from the existing clusters. This determination, called the vigilance test, is incorporated into the adaptive backward network. The ART network modifications are ART1 and ART2 (Pandya and Macy 1996). The ART1 model clusters binary input patterns and ART2 clusters analogue input patterns. Considering the significant processing timesaving, ART1 was selected here. A good knowledge-based system must 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 vigilance test concept is described as follows.

- 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 pattern will be executed. The characteristics of the old and new patterns can be satisfied and the old knowledge can be properly retained. The stability of the system can be maintained.

- If the characteristics of a new pattern are not similar to any 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.

The ART1 architecture construction includes an input layer, network connection and output layer (figure 3). It uses an output-processing unit to present a certain cluster. Every connection weight between the input layer and the output units indicates the characteristics of a specific cluster. The number of output processing units passing the vigilance test may exceed one, so the network utilizes a 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 vigilance value plays an important role in the ART1 algorithm. This value can be used to distinguish the similar patterns. The higher the assigned vigilance value, the fewer the number of output units that will pass the vigilance test. On the contrary, the lower the assigned vigilance value, the greater the number of output units that will pass the vigilance tests. In other words, this system can detect similar but different types of clusters.

The detailed steps of the ART1 algorithm developed in this research are stated below and shown in figure 4.

(1) Set network parameter
Nout = 1

(2) Set initial weighting matrix, W

$$W^t[i][1] = 1 \text{ (a similar value)}$$

$$W^b[i][1] = \frac{1}{N+1} \text{ (a match value)}$$

(3) Input a training sample vector X

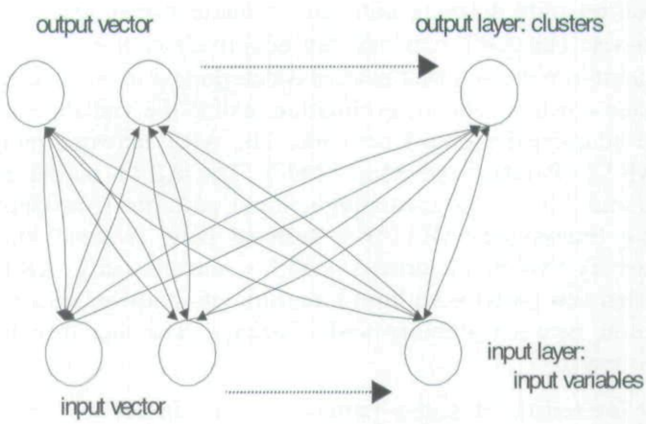

Figure 3. ART1 network.

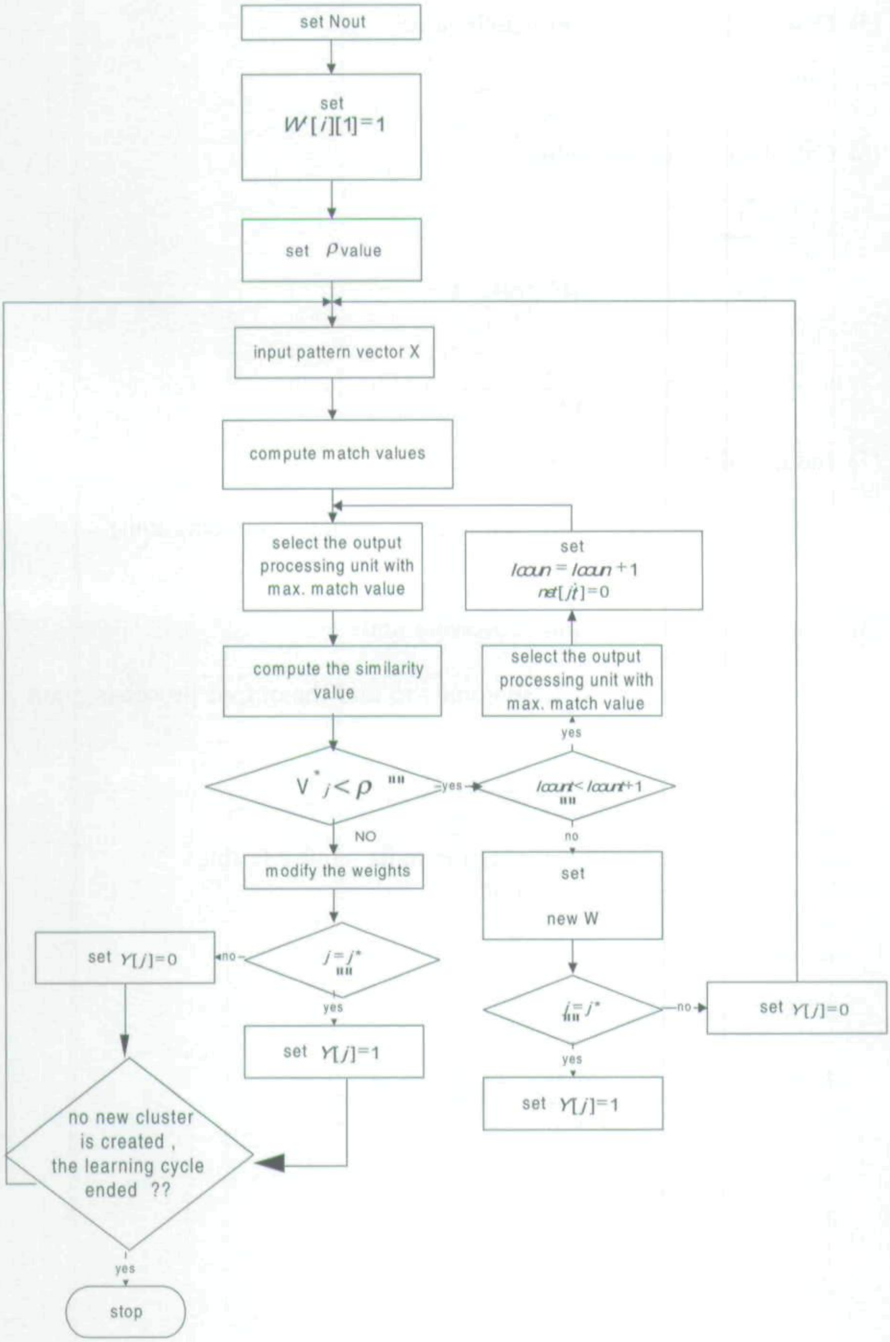

Figure 4. ART1 network procedure.

(4) Calculate the match value

$$net[j] = \sum_i W^b[i][j] \cdot X[i]$$

$$lcount = 0$$

(5) Determine the maximum match value

$$net[j^*] = \max_j(net[j])$$

(6) Calculate the similar values

$$\|X\| = \sum_i X[i]$$

$$\|W_j^t \cdot X\| = \sum_i W^t[i][j^*] \cdot X[i]$$

$$j^* = \frac{\|W_j^t \cdot X\|}{\|X\|}$$

(7) Judge and test the vigilance value
if $j^* < \rho$(vigilance value)
go to step (8) (not similar, test another output processing unit)
or else
go to step (9) (similar enough, modify weights)

(8) Test any available output processing units
if $lcount < Nout$
test another output processing units to find the output processing unit with max. match value.
set $lcount = lcount + 1$
set $net[j^*] = 0$
go to step (5)
or else (no other output processing units require testing)
(a) create a new cluster output
set $Nout = Nout + 1$
set new weight

$$W^t[i][Nout] = X[i]$$

$$W^b[i][Nout] = \frac{X[i]}{0.5 + \sum_i X[i]}$$

(b) set output unit value
if $j = j^*$
$Y[j] = 1$
else
$Y[j] = 0$
(c) go to step (3)

(9) Modify the weight
(a) modify weights

$$W^t[i][j^*] = W^t[i][j^*] \cdot X[i]$$

$$W^b[i][j^*] = \frac{W^t[i][j^*] \cdot X[i]}{0.5 = \sum_i W^t[i][j^*] \cdot X[i][j^*]}$$

(b) set output unit value
if $j = j^*$
$Y[j] = 1$
or else
$Y[j] = 0$
(c) if a learning cycle is finished and no new clusters are created, the final results are obtained and the network ceases to learn situations.
or else go to step (3)

3. Network training

There are two phases in the ART1 architecture. One is the training process and the other is the recognition process (figure 5). After the intelligent patterned WBMs recognition system was conceptually designed, a practical software system was developed for system implementation and verification. This system was developed using Borland C++, under a Microsoft Windows 98 operating platform. Actual data from a product with 387 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 systematic pattern types on the wafers, the network must be trained. Since ART1 is an unsupervised network, it is not necessary to link the input and output vectors to attain good recognition. That is, we do not need to send the training output results back into the network.

Because the learning is unsupervised, the ARTI network convergent process cannot be evaluated using error rate or square error. Total distance is used in this kind of network and it can be defined as:

$$\text{Total distance} = \sum_p \left(\min_j d_j^p \right).$$

where d_j^p is the distance between the pth samples and the jth output processing unit = 1 – similarity between the pth samples and the jth output processing unit.

During our training processes, the ART1 network converged in three cycles. The time utilized for training 20 samples on a PC with an Intel Pentium 166 and 64 MB RAM was approximately 3 s.

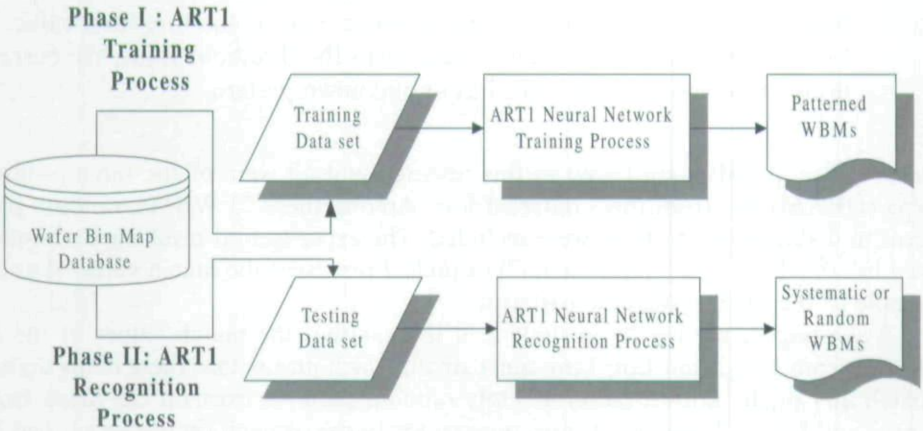

Figure 5. Two phases in the ART1 network.

4. Experimental results

To evaluate the performance of the developed system, the ART1 network must be trained and tested using actual samples. Owing to the business security consideration and the fact that ART1 network does not need many samples for training. A semiconductor manufacturing company in Taiwan provided only 20 WBM samples. Each of these map samples consisted of 387 dies. For these 20 WBMs samples, skilled manufacturing engineers identified the following three types of systematic patterns:

(1) Type I: ring-type with small thickness (figure 6).
(2) Type II: ring-type with large thickness (figure 7).
(3) Type III: Bull's eye (figure 8).

The ART1 network was trained and adjusted to recognize the above three systematic patterns. It adjusted the number of output nodes (number of pattern types) according to the vigilance value. With an adequate vigilance value, the training WBMs will be assigned to the right group. According to the pilot runs in this study, when the vigilance value is set in the range (0.330, 0.340), the classification of all of the training WBMs would be the same as the results from skilled manufacturing engineers. The exact vigilance value used in the ART1 training was set to 0.335. The ART1 network converged after three cycles.

With the trained ART1 network, the weight parameters were used to test the other WBMs. Test sample classification was determined using the match and threshold values. The match value (ranging from 0 to 1) indicates the similarity between a test sample and a certain pattern type. The higher the match value, the more likely a test sample will pass the threshold and be allocated to a group. If the match value is smaller than the threshold value, the current test sample will not be allocated to any group. In this situation, the test sample was either a random pattern or an unrecognized pattern. If the match value is greater than the threshold value, the test sample will be recognized as one of the three known systematic patterns or unknown patterns and further analysis will be required. Figure 9 illustrates the procedures for testing WBMs and the possible test results are summarized below.

(1) Match value $<$ threshold value. In this situation, the test sample is either recognized as a random or new pattern that has not yet been recognized.
(2) Match value $\geq$ threshold value. The test sample is classified as one of the three pattern types if only one match value exceeds the threshold value. If two or more match values are greater than the threshold value, the current testing sample will be identified as an unknown pattern.

Seventy-five WBMs were tested in this research and all were of the same product type (DRAM) but from three different lots. Among these 75 WBMs, random patterns and systematic patterns were included. The experimental results are summarized in table 2. Match [1] to match [3] in table 2 represent the match value of a test sample to the three systematic patterns.

According to the results in table 2, it is clear that the match values of the 50 WBMs from Lot_2 and Lot_3 are quiet small, which means that these maps do not match any of the known patterns. Only random patterns exist on the maps from Lot_2 and Lot_3. These results are consistent with the recognition results verified by the skilled manufacturing engineers. The WBM test results from Lot_1 are of more

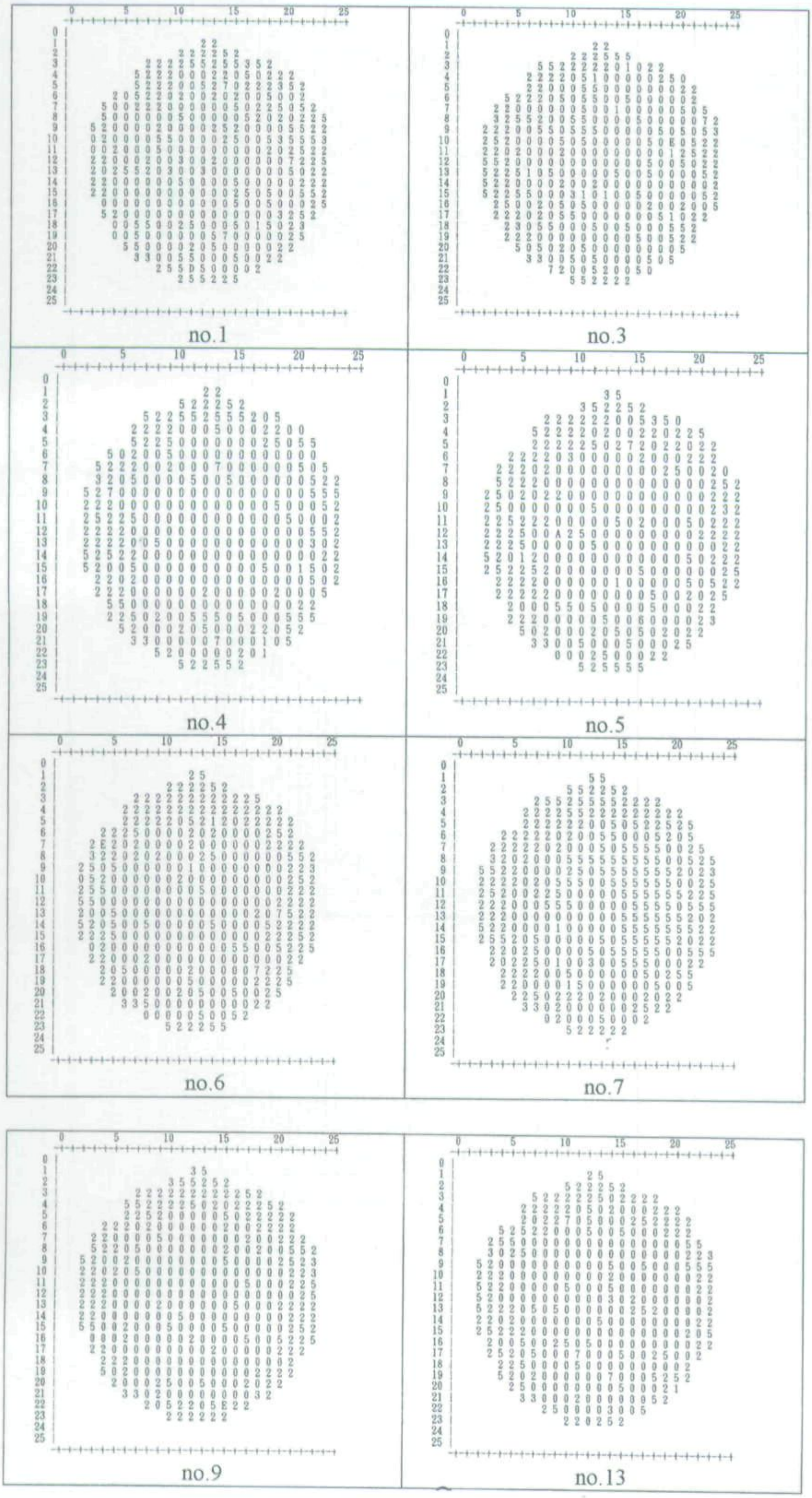

Figure 6. Type I defect — ring-type with a small thickness.

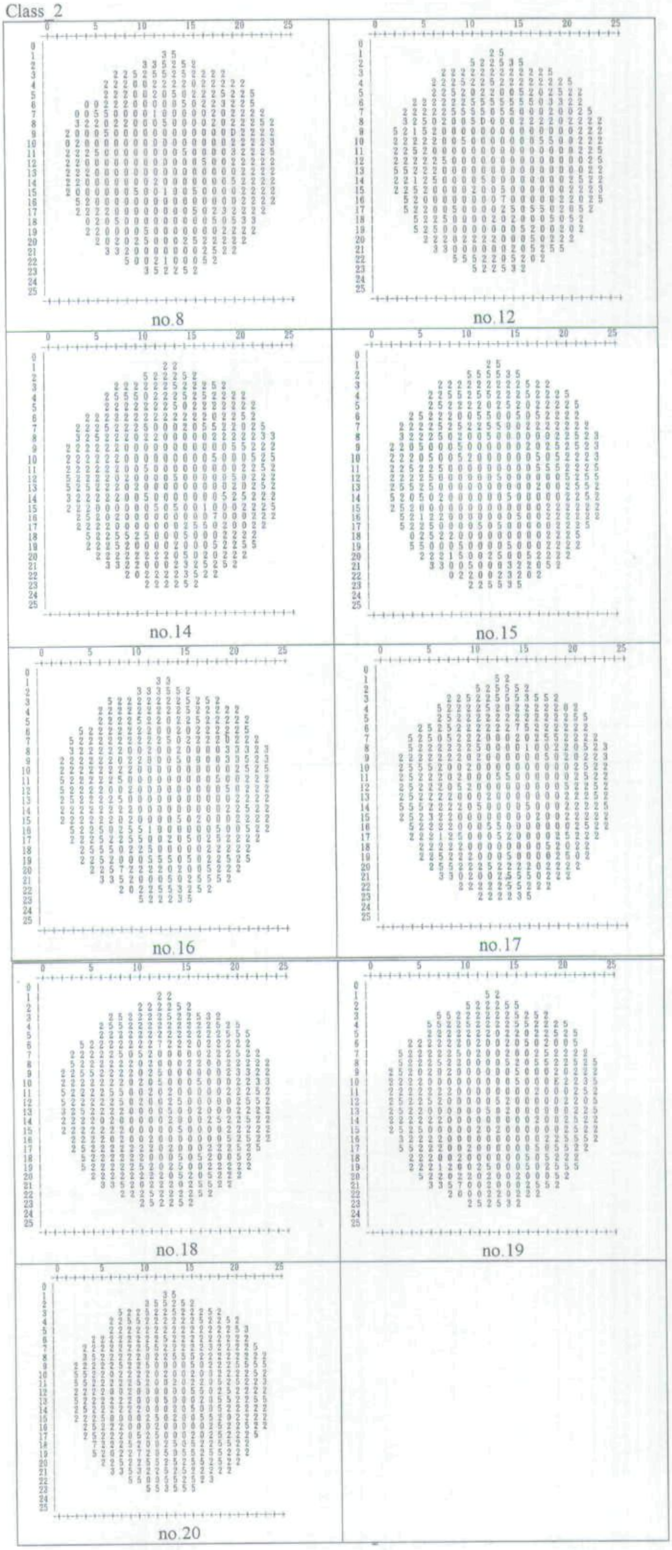

Figure 7. Type II — ring-type with a large thickness.

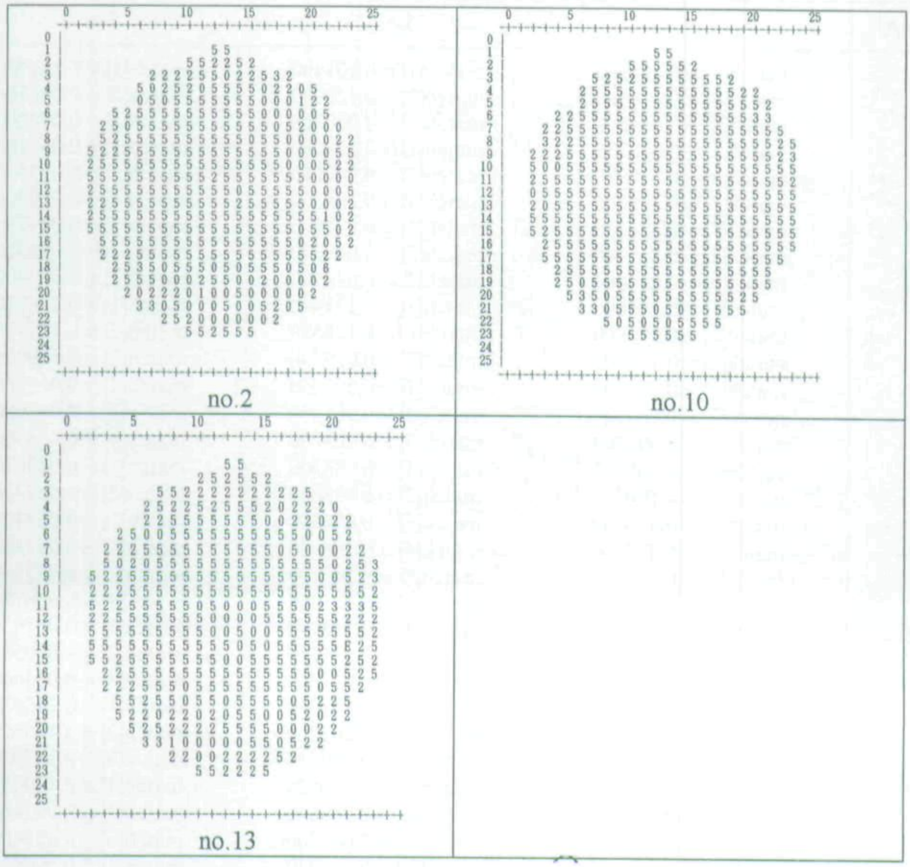

Figure 8. Type III defect — Bull's-eye WBMs.

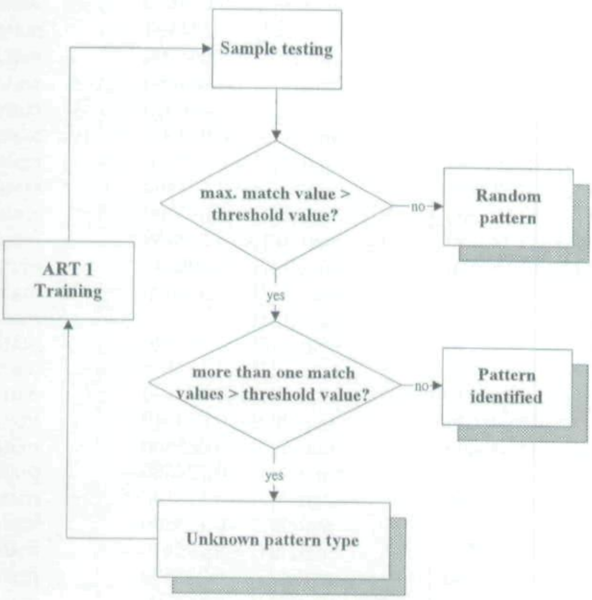

Figure 9. Sample testing procedure.

Map	Lot_1	Lot_2	Lot_3
1	match[1] = 0.134200 match[2] = 0.110980 match[3] = 0.054600	match[1] = 0.101400 match[2] = 0.120900 match[3] = 0.070200	match[1] = 0.121700 match[2] = 0.086160 match[3] = 0.070200
2	match[1] = 0.638739 match[2] = 0.510920 match[3] = 0.425100	match[1] = 0.128080 match[2] = 0.143600 match[3] = 0.105300	match[1] = 0.068160 match[2] = 0.053540 match[3] = 0.050700
3	match[1] = 0.337230 match[2] = 0.251740 match[3] = 0.234000	match[1] = 0.071000 match[2] = 0.049640 match[3] = 0.042900	match[1] = 0.047730 match[2] = 0.036520 match[3] = 0.023400
4	match[1] = 0.246600 match[2] = 0.170900 match[3] = 0.120900	match[1] = 0.131490 match[2] = 0.126580 match[3] = 0.089700	match[1] = 0.178030 match[2] = 0.117720 match[3] = 0.120900
5	match[1] = 0.374500 match[2] = 0.264140 match[3] = 0.304200	match[1] = 0.161550 match[2] = 0.101760 match[3] = 0.109200	match[1] = 0.066300 match[2] = 0.081900 match[3] = 0.058500
6	match[1] = 0.668520 match[2] = 0.504580 match[3] = 0.335400	match[1] = 0.098300 match[2] = 0.088640 match[3] = 0.074100	match[1] = 0.085670 match[2] = 0.067720 match[3] = 0.062400
7	match[1] = 0.473960 match[2] = 0.371940 match[3] = 0.249600	match[1] = 0.168730 match[2] = 0.147500 match[3] = 0.081900	match[1] = 0.097320 match[2] = 0.087580 match[3] = 0.050700
8	match[1] = 0.077870 match[2] = 0.078360 match[3] = 0.054600	match[1] = 0.112970 match[2] = 0.125160 match[3] = 0.081900	match[1] = 0.051630 match[2] = 0.070560 match[3] = 0.054600
9	match[1] = 0.035100 match[2] = 0.039000 match[3] = 0.046800	match[1] = 0.147090 match[2] = 0.101760 match[3] = 0.097500	match[1] = 0.218970 match[2] = 0.166660 match[3] = 0.117000
10	match[1] = 0.581160 match[2] = 0.451000 match[3] = 0.370500	match[1] = 0.134820 match[2] = 0.088640 match[3] = 0.089700	match[1] = 0.120640 match[2] = 0.097860 match[3] = 0.085800
11	match[1] = 0.321530 match[2] = 0.259180 match[3] = 0.175500	match[1] = 0.047730 match[2] = 0.040420 match[3] = 0.023400	match[1] = 0.090630 match[2] = 0.075520 match[3] = 0.081900
12	match[1] = 0.194820 match[2] = 0.118780 match[3] = 0.105300	match[1] = 0.058500 match[2] = 0.074460 match[3] = 0.066300	match[1] = 0.079860 match[2] = 0.073040 match[3] = 0.042900
13	match[1] = 0.203030 match[2] = 0.110620 match[3] = 0.109200	match[1] = 0.087660 match[2] = 0.089000 match[3] = 0.042900	match[1] = 0.117180 match[2] = 0.060280 match[3] = 0.046800
14	match[1] = 0.667099 match[2] = 0.587499 match[3] = 0.382200	match[1] = 0.205820 match[2] = 0.151400 match[3] = 0.093600	match[1] = 0.077380 match[2] = 0.053540 match[3] = 0.031200
15	match[1] = 0.237150 match[2] = 0.160260 match[3] = 0.120900	match[1] = 0.143060 match[2] = 0.099640 match[3] = 0.054600	match[1] = 0.128520 match[2] = 0.112400 match[3] = 0.093600
16	match[1] = 0.094890 match[2] = 0.086160 match[3] = 0.035100	match[1] = 0.155250 match[2] = 0.117720 match[3] = 0.113100	match[1] = 0.215300 match[2] = 0.209540 match[3] = 0.156000
17	match[1] = 0.866849 match[2] = 0.691039 match[3] = 0.643499	match[1] = 0.093470 match[2] = 0.084740 match[3] = 0.085800	match[1] = 0.060360 match[2] = 0.051060 match[3] = 0.027300
18	match[1] = 0.317810 match[2] = 0.238980 match[3] = 0.163800	match[1] = 0.124800 match[2] = 0.113100 match[3] = 0.124800	match[1] = 0.059430 match[2] = 0.062760 match[3] = 0.058500
19	match[1] = 0.464690 match[2] = 0.321240 match[3] = 0.323700	match[1] = 0.292420 match[2] = 0.236500 match[3] = 0.144300	match[1] = 0.036030 match[2] = 0.020920 match[3] = 0.031200

(*continued opposite*)

Map	Lot_1	Lot_2	Lot_3
20	match[1] = 0.109950	match[1] = 0.138720	match[1] = 0.055530
	match[2] = 0.095740	match[2] = 0.108140	match[2] = 0.045740
	match[3] = 0.050700	match[3] = 0.085800	match[3] = 0.046800
21	match[1] = 0.168370	match[1] = 0.231080	match[1] = 0.076890
	match[2] = 0.138280	match[2] = 0.145020	match[2] = 0.060280
	match[3] = 0.097500	match[3] = 0.117000	match[3] = 0.031200
22	match[1] = 0.051630	match[1] = 0.219430	match[1] = 0.095020
	match[2] = 0.071980	match[2] = 0.152460	match[2] = 0.062400
	match[3] = 0.042900	match[3] = 0.124800	match[3] = 0.058500
23	match[1] = 0.595720	match[1] = 0.103260	match[1] = 0.220800
	match[2] = 0.536480	match[2] = 0.128360	match[2] = 0.106020
	match[3] = 0.315900	match[3] = 0.078000	match[3] = 0.089700
24	match[1] = 0.249420	match[1] = 0.390720	match[1] = 0.139650
	match[2] = 0.210280	match[2] = 0.256000	match[2] = 0.114880
	match[3] = 0.097500	match[3] = 0.198900	match[3] = 0.085800
25	match[1] = 0.287950	match[1] = 0.130920	match[1] = 0.124180
	match[2] = 0.231180	match[2] = 0.063820	match[2] = 0.113460
	match[3] = 0.187200	match[3] = 0.085800	match[3] = 0.085800

Table 2. Testing results of 75 WBMs.

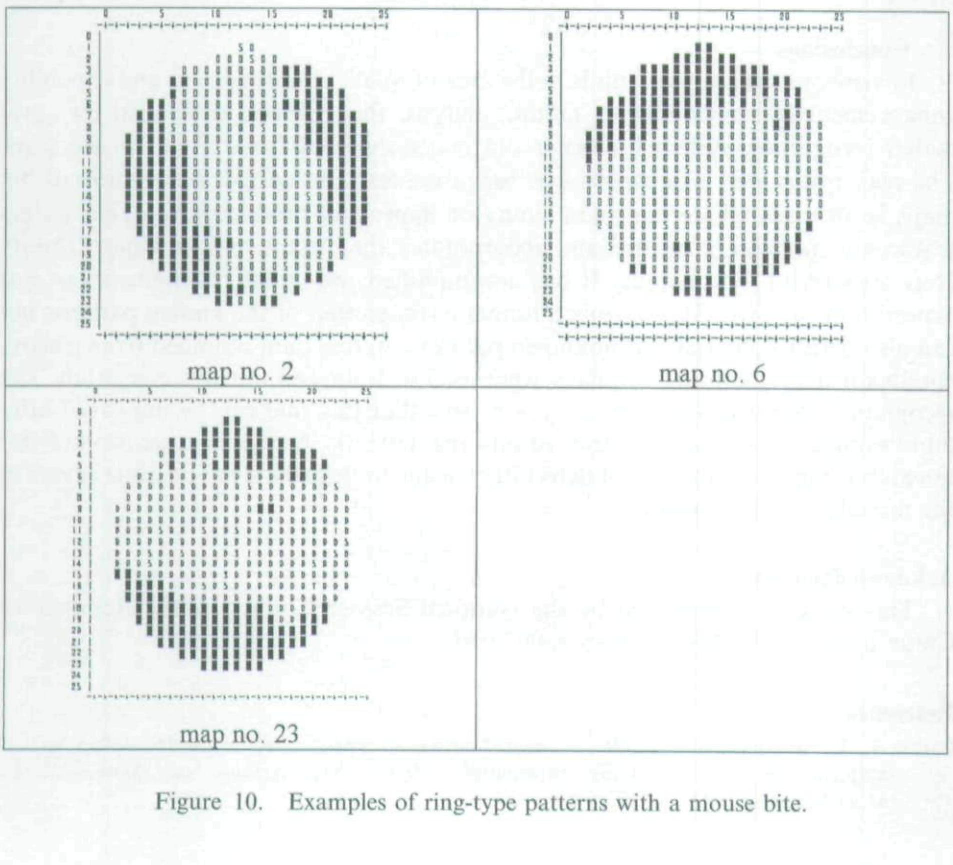

Figure 10. Examples of ring-type patterns with a mouse bite.

interest. Two of the match values (match [1] and match [2]) from maps 2, 6, 14 and 23 are greater than the threshold value. According to the predetermined rule, if two or more match values are greater than the threshold value, the tested sample will be classified as having an unknown pattern. This does not agree with the preknown answers provided by the manufacturing engineers. The algorithm failed to correctly recognize the four maps. The recognition rate can therefore be claimed to be about 95% (71/75) at this moment.

To improve the recognition rate, the aforementioned four maps (maps 2, 6, 14 and 23) from Lot_1 deserve further analysis. The systematic pattern of map 2 is a ring type pattern with a mouse bite on the top of the ring, while the patterns of maps 6 and 23 have a mouse bite on the left and right sections of the ring (figure 10). A ring-like pattern is a common characteristic of these maps and the known ring type patterns. Consequently, the test result shows that these maps are similar to the two ring type patterns. To increase the recognition capability of these pattern types, these three maps should be trained into the network to become a new pattern.

In summary, the approach presented in this paper helps skilled manufacturing engineers to screen out the systematic patterns. The recognition rate can be up to 95%, which can further be improved by training with more data in the future. This approach also successfully replaces the human recognition and improves the processing speed. In real situations, there may be plenty of possible situations in which systematic patterns will occur. Although the developed system can only recognize three types of patterns due to the limited number of available samples, the ART1 network can be retrained whenever new patterns are presented to the system. This flexibility makes it a valuable tool in wafer bin map recognition.

5. Conclusions

In view of increasing demands in the area of quality improvement and capability enhancement in semiconductor failure analysis, this research developed an automated recognition system for wafer bin maps collected from circuit probe tests. The reason this area was chosen was because a fast and correct recognition of bin maps is an essential step in maintaining or improving product yield. This system features a modular structure and incorporates the Adaptive Resonance Theory Network1 (ART1) technique. It has accomplished the initial requirements of our expectations. It provides not only automatic recognition of the known patterns but can also detect potential and unknown patterns. Actual data obtained from a semiconductor manufacturing company were used to train and test this algorithm. The recognition rate was about 95%. It is believe that this rate can be improved after more sample maps have been trained into the network. A possible extension of this research is the incorporation of defect knowledge to determine the possible causes in the manufacturing processes.

Acknowledgement

This work was supported by the National Science Council of the Republic of China under Grant NSC90-8218-E-007-010.

References

Chen, F. L. and Liu, S. F., 2000, A neural-network approach to recognize defect spatial patterns in semiconductor fabrication. *IEEE Transactions on Semiconductor Manufacturing*, **13**, 366–373.

Cox, I. and Reynolds, A., 1996, *Manufacturing Digital's Alpha AXP: Rapid Applications Development to Assure Critical Success Through Quality* (SAS Co.).

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

Kaempf, U., 1995, The binominal test: a simple tool to identify process problems. *IEEE Transactions on Semiconductor Manufacturing*, **8,** 160–166.

Lin, G. T., 1998, Patterned recognition of wafer bin maps. Master thesis, Taiwan University.

Mirza, A. I., Donoghue, G. O., Drake, A. W. and Graves, S. C., 1995, Spatial yield modeling for semiconductor wafers. Proceedings of the IEEE/SEMI Advanced Semiconductor Manufacturing Conference, pp. 276–281.

Pandya, A. S. and Macy, R. B.,1996, *Pattern Recognition with Neural Networks in C++* (IEEE Press).

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