Expert Systems with Applications 39 (2012) 2856–2864

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Constructing a yield model for integrated circuits based on a novel fuzzy variable of clustered defect pattern

Jun-Shuw Lin

Department of Industrial Engineering and Management, National Chiao Tung University, 1001 Dah-Hsei Road, Hsin-Chu 300, Taiwan, ROC

ARTICLE INFO

Keywords:
Yield model
Clustered defects
Fuzzy logic control
Fuzzy variable of clustering pattern (FVCP)
Ant colony optimization
Back-propagation neural network

ABSTRACT

As the wafer size increases, the clustering phenomenon of defects becomes significant. In addition to clustered defects, various clustering patterns also influence the wafer yield. In fact, the recognition of clustering pattern usually exists fuzziness. However, the wafer yield models in previous studies did not consider the fuzziness of clustering pattern belonging to which shape in recognition. Therefore, the objective of this study is to develop a new fuzzy variable of clustering pattern (FVCP) by using fuzzy logic control, and predict the wafer yield by using back-propagation neural network (BPNN) incorporating ant colony optimization (ACO). The proposed method utilizes defect counts, cluster index (*CI*), and FVCP as inputs for ACO-BPNN. A simulated study is utilized to demonstrate the effectiveness of the proposed model.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The integrated circuits (IC) manufacture has become the major industry worldwide, and all electrical appliances are closely linked with the IC. Wafer yield is a key index for evaluating the process capability of IC manufacturers, and it can also reflect the problem of process. An accurate yield prediction model is very useful to predict manufacturing costs for products still under development (Kumar et al., 2006), which can offer a reasonable and acceptable price to customers. Therefore, it is a very important task to manage the wafer yield.

When the clustering phenomenon of defects is not significant and chip size is small (Cunningham, 1990), the Poisson yield model can estimate the wafer yield reasonably. As the wafer size increases and the clustering phenomenon of defects is significant, the Poisson yield model becomes inappropriate (Stapper, 1985).

The Negative Binomial yield model (Stapper, 1973) includes a clustering index (α), but the value of α can be very scattered and negative that leads to unhandy analysis (Cunningham, 1990). Many mathematical models have been developed to predict wafer yield in the last 40 years (Cunningham, 1990; Stapper, 1991; Stapper & Rosner, 1995), but these models are very complicated in practice. Neural networks are also utilized to construct the wafer yield models, but these models have certain problems such as setting parameters (e.g., the number of neurons in the hidden layers, the momentum, and the learning rate) and local optimal solution (Tong & Chao, 2008).

When these defects cluster, the size and shape of clustered defect pattern can indicate the specific cause (e.g., diffusion problems, photo spin anomalies, etch discordance, and handling damage such as scratches) of process problems (Neyer & Hafner, 2004). For this reason, it is very important for managers to recognize the clustering pattern in order to monitor and control the wafer yield. In fact, the recognition of clustering pattern belonging to which shape usually exists fuzziness. However, the wafer yield models in previous studies did not consider the fuzziness of clustering pattern belonging to which shape in recognition.

In order to improve the above drawbacks, this study develops a new fuzzy variable of clustering pattern (FVCP), which considers the fuzziness of clustering pattern belonging to which shape in recognition by using fuzzy logic control, and predict the wafer yield by using back-propagation neural network (BPNN) incorporating ant colony optimization (ACO). In this study, ACO (Dorigo & Stutzle, 2004) algorithm is adopted to determine the Neural Network's parameters (e.g., the number of neurons in the hidden layers, the momentum, and the learning rate), and ACO algorithm can overcome the drawbacks (e.g., setting parameters, overtraining, untimely convergence, and local optimal solution) in BPNN (Rumelhart, Hinton, & Williams, 1986). The proposed method utilizes defect counts, cluster index (*CI*) (Jun, Hong, Kim, Park, & Park, 1999), and FVCP as the inputs for ACO-BPNN (Liu, Wu, & Qian, 2006; Wei, 2007). A simulated study is utilized to demonstrate the effectiveness of the proposed model.

2. Related literature

This section reviews literature related to the wafer yield model, clustering index, ACO-BPNN model scheme, and fuzzy logic inference.

E-mail address: junsoon1@hotmail.com

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.08.144

2.1. Wafer yield model

The Poisson yield model assumes that the defects on a chip follow a Poisson probability distribution. Under this assumption, the probability that a chip has k number of defects is

$$P(k) = \frac{e^{-\lambda_0}\lambda_0^k}{k!}, \quad k = 0, 1, 2\ldots \tag{1}$$

where λ_0 is the average number of defects per chip, and k is the number of defects per chip. The Poisson yield model can be obtained as

$$Y = P(k = 0) = e^{-\lambda_0} \tag{2}$$

Cunningham (1990) indicated that, when the chip size is less than 0.25 cm^2, the Poisson yield model is appropriate. However, as the chip size increases, the conventional Poisson yield model will frequently underestimate the actual wafer yield.

The Negative Binomial yield model proposed by Stapper (1973) is a widely applied yield model, which employs a gamma function for the distribution of defect density. The Negative Binomial yield model can be expressed as

$$Y = \frac{1}{(1 + D_0A/\alpha)^\alpha} \tag{3}$$

where D_0 is the average number of defects per unit area, A is the chip area, and α is the cluster parameter. The value of α is calculated by the following equation:

$$\alpha = \bar{\lambda}^2/(\sigma^2 - \bar{\lambda}) \tag{4}$$

where $\bar{\lambda}$ is the mean number of defects per chip, and σ^2 is the variance. Cunningham (1990) indicated that, the value of α can be quite scattered and sometimes negative when the Negative Binomial yield model is used to predict yield.

Other yield models are summarized in Stapper and Rosner (1995). Tong, Lee, and Su (1997) proposed a neural network-based approach to predict the wafer yield. Langford, Liou, and Raghavan (2001) presented a simple robust windowing method for the Poisson yield model to extract the systematic and random components of yield from wafer probe bin map data. Liou et al. (2002) presented a statistical modeling of MOS devices for parametric yield prediction. Meyer and Park (2003) presented a center-satellite model to predict defect-tolerant yield in the embedded core context. Dupret and Kielbasa (2004) presented the partial least square (PLS) regression model to predict the yield from measurements obtained during the production. Kim and Baldwin (2005) presented a theoretical yield model for assembly processes of area array solders inter connect process. Tong and Chao (2008) proposed a general regression neural network (GRNN) to predict the wafer yield with clustered defects.

2.2. Defect cluster index

In this study, I use the clustering index (CI) proposed by Jun et al. (1999) to measure the clustering phenomenon of defects. Suppose that CI is the clustering parameter, shown in Eq. (5).

$$CI = \min\left\{\frac{S_v^2}{\overline{V}^2}, \frac{S_W^2}{\overline{W}^2}\right\} \tag{5}$$

where

$$\overline{V} = \sum V_i/n \tag{6}$$
$$S_v^2 = \sum (V_i - \overline{V})^2/(n-1) \tag{7}$$
$$\overline{W} = \sum W_i/n \tag{8}$$
$$S_w^2 = \sum (W_i - \overline{W})^2/(n-1) \tag{9}$$

where V_i and W_i are a sequence of defect intervals on the X-axis and Y-axis defined as

$$V_i = X_{(i)} - X_{(i-1)} \quad i = 1, 2, \ldots, n, \tag{10}$$
$$W_i = Y_{(i)} - Y_{(i-1)} \quad i = 1, 2, \ldots, n, \tag{11}$$

where $X_{(i)}$ and $Y_{(i)}$ denote the ith smallest defect coordinates on the X-axis and Y-axis respectively, $X_{(0)} = Y_{(0)} = 0$, and n is the number of defects on a wafer. The value of CI is close to 1 if the defects are randomly scattered, and the value of CI is expected to be greater than 1 if clustering of defects appears.

2.3. ACO-BPNN model scheme

The architecture of BPNN includes the input layer, output layer, and hidden layer. The number of process element (PE) within the input layer and the output layer depend on the analyzed problem, but the number of hidden layer and the number of PE within the hidden layer are uncertain. The learning process of BPNN adopts the gradient steepest descent method to adjust the connection weights and reduce the inaccuracy of neural networks. The BPNN has been successfully applied to many research fields, such as engineering management, climatology, and economics. Fig. 1 (Chen, Chen, & Kuo, 2010) shows the architecture of the BPNN.

ACO (Dorigo & Stutzle, 2004) algorithm draws its inspiration from the behavior of real ants as they move from their nest towards a food source (Colorni, Dorigo, & Maniezzo, 1991). ACO has been successfully applied to solve some complex combinatorial optimization problems with NP-hard characteristic, such as traveling salesman problems (TSP) (Dorigo & Gambardella, 1997), quadratic assignment problems (Maniezzo & Colorni, 1999), and scheduling problems (Blum & Sampels, 2004). ACO is based on the observation of the behavior of real ant colonies searching for food sources. Real ants deposit an aromatic essence, called pheromone, on the path that they walk. Other ants searching for food sense that pheromone and use this information in selecting their path. The quantity of pheromone deposited on a path is based on the length of the path and the quality of the food source. As more ants follow a path the level of pheromone on that path will increase, increasing its selection probability by other ants. In ACO, artificial ants are used for searching the optimal solutions to an optimization problem.

The scheme of ACO-BP (Liu et al., 2006; Wei, 2007) neural network can be depicted as follows. After the architecture of a neural network is selected, it needs to be trained before being used. Given D parameters in the network, which consist of all the weights and biases, the evolution of network parameters can be regarded as the process of searching for the optimal combination of the D parameters in their solution spaces. Several candidacy groups of combination of network parameters can be provided by the ACO scheme. The BP algorithm initializes the weights of the network with these values and begins to train the network. Since ACO provides the BP with several groups of good initial values, the risk of being trapped in the local optima sharply decreases. Consequently, both the training effectiveness and evolving speed can be enhanced. In other words, the basic idea of the hybrid algorithm of

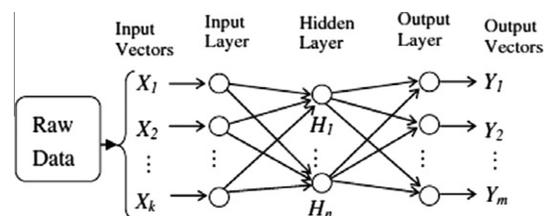

Fig. 1. The architecture of BPNN (Chen et al., 2010).

ACO and BP is to use ACO to search the optimal combination of all the network parameters, and then use the BP algorithm to train the data through the optimal combination of network parameters.

2.4. Fuzzy logic inference

Fuzzy theory (Zadeh, 1965) is a special method to handle the problem of blurred uncertainty phenomenon. Because the concept of uncertainty can not be described by crisp set, it is usually represented by fuzzy set (Zadeh, 1965). Fuzzy logic control (Mamdani & Assilian, 1981) is an important application of fuzzy logic inference. The following two sub sections introduce the brief concept of fuzzy set and fuzzy logic control.

2.4.1. Fuzzy set

In crisp set, the value of characteristic function is either 0 or 1, but the dichotomy is usually inadequate in human mind. In a real world, it is full of blurred uncertainty. Therefore, an appropriate value is taken between 0 and 1 to represent the degree of one element belonging to a set. The membership function is denoted by $\mu_A(x)$, and it can be expressed as $\mu_A : \mathrm{X} \to [0,1]$ (i.e.,$0 \le \mu_A(\mathrm{x}) \le 1, \mathrm{x} \in \mathrm{X}$), where A is fuzzy set, x is element, X is universal set. When the value of μ_A (x) is closer to 1, the degree of x belonging to A is larger.

2.4.2. Fuzzy logic control

In fuzzy logic control, some fuzzy rules (e.g., IF…,THEN…) are employed to establish the inference mechanism. The input value of fuzzy logic controller is often a crisp value, and then fuzzificate the input value into the inference mechanism. Finally, the output value of fuzzy logic controller is defuzzificated to infer the actual situation. Fig. 2 shows the architecture of fuzzy logic controller.

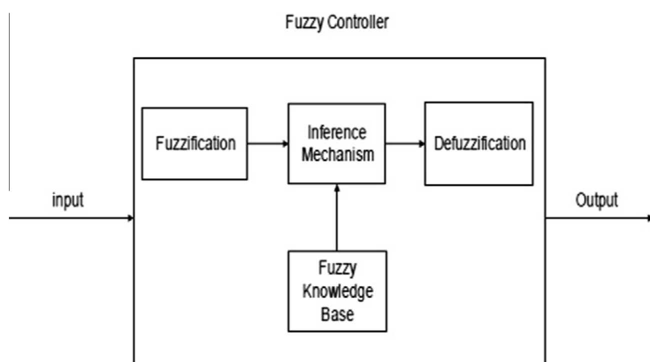

Fig. 2. The architecture of fuzzy logic controller.

3. Proposed approach

The constructing procedure of the proposed wafer yield model can be described as following sections:

3.1. Develop a new FVCP

In fact, the recognition of clustering pattern belonging to which shape usually exists fuzziness. Fig. 3 illustrates what is the fuzziness of clustering pattern belonging to which shape in recognition. For example, Fig. 3(a) is a typical bottom pattern, Fig. 3(c) is a typical crescent moon pattern, and Fig. 3(b) is a real clustering pattern. When the managers want to recognize the clustering pattern, the Fig. 3(b) should belong to bottom pattern or crescent moon pattern? In order to consider the fuzziness of clustering pattern recognition, this study develops a new numerical FVCP by using fuzzy logic control to recognize the clustering pattern. The variation interval of an included angle between each defect coordinate and the first quadrant on X-axis per wafer, the distance interval variation between each defect coordinate and the origin per wafer are both used to infer the degree of clustering pattern belonging to which shape.

In this study, one random pattern and four clustering patterns (i.e., bull eye pattern, edge pattern, bottom pattern, and crescent moon pattern) (Friedman, Hansen, Nair, & James, 1997) are considered. The five clustering patterns are shown in Fig. 4. Therefore, the range of FVCP is from 0 to 5. The value 1 represents random pattern, 2 represents bull eye pattern, 3 represents edge pattern, 4 represents bottom pattern, and 5 represents crescent moon pattern. Because the value of FVCP is a fuzzy inference number, it can be a decimal. For example, the output value 4.7 (FVCP = 4.7) of fuzzy logic controller implies that the recognized clustering pattern has 70% degree belonging to crescent moon pattern and 30% degree belonging to bottom pattern. The detailed descriptions of the proposed FVCP developing procedure in this study are listed as the following six steps:

Step 1: To calculate the positive included angle (θ_i) between each defect coordinate and the first quadrant on X-axis per wafer, where $i = 1,2,\ldots,n$, $\theta_i = \tan^{-1}(Y_i/X_i)$,$X_i$ is the X-axis coordinate of the ith defect point, Y_i is the Y-axis coordinate of the ith defect point, and n is the number of defects. After that, to arrange the positive included angle (θ_i) in order, where $\theta_{(i)}$ is the ith smallest included angle in order. Then the interval of the positive included angle (θ_i) is denoted as A_i, where $A_i = \theta_{(i)} - \theta_{(i-1)}$, $i = 1,2,\ldots,n$, and $\theta_{(0)}$=0. Fig. 5 shows the angle of each defect coordinate on a wafer.

Step 2: To calculate the distance (L_i) between each defect coordinate and the origin per wafer, where $L_i = \sqrt{X_i^2 + Y_i^2}$, $i = 1,2,\ldots,n$, X_i is the X-axis coordinate of the ith defect point, Y_i is the Y-axis coordinate of the ith defect point, and n is the number of defects. After that, to arrange the distance (L_i) between each defect coordinate and the origin in order, where $L_{(i)}$ is the ith smallest distance (L_i) between each defect coordinate and the origin in order. Then the interval of the distance (L_i) between each defect coordinate and the origin is denoted as D_i, where $D_i = L_{(i)} - L_{(i-1)}$, $i = 1,2,\ldots,n$, and $L_{(0)} = 0$. Fig. 6 shows the distance of each defect coordinate on a wafer.

Step 3: To calculate $\left\{\frac{S_A^2}{\bar{A}^2}, \frac{S_D^2}{\bar{D}^2}\right\}$ per wafer, where $\bar{A} = \sum A_i/n$, $\bar{D} = \sum D_i/n$, $S_A^2 = \sum(A_i - \bar{A})^2/(n-1)$, and $S_D^2 = \sum(D_i - \bar{D})^2/(n-1)$. This study utilizes $\frac{S_A^2}{\bar{A}^2}$ and $\frac{S_D^2}{\bar{D}^2}$ as input1 and input2 for fuzzy logic controller respectively, and then the output of fuzzy logic controller is the FVCP.

Step 4: The computer software, fuzzy TECH 4.22, is used to perform fuzzy inference. The input1 ($\frac{S_A^2}{\bar{A}^2}$) and input2 ($\frac{S_D^2}{\bar{D}^2}$) are respectively divided into four levels as "low", "medium_low", "medium_high", and "high". The output is divided into five levels as "term1", "term2", "term3", "term4", and "term5". The membership functions of input1, input2, and output are respectively shown from Figs. 7–9.

Step 5: Write the rules of fuzzy logic controller. Fig. 10 shows the architecture of the fuzzy pattern variable controller. Fig. 11 shows the rules of fuzzy logic controller.

Step 6: Perform the fuzzy logic controller to obtain the output (FVCP). For example, we let (input1,input2) = (20.12,18.09) into the fuzzy logic controller, and then the output defuzzification value by Center-of-Maximum (CoM) is 1.67 (FVCP = 1.67). Fig. 12 shows the outcome of fuzzy logic controller. Fig. 13 shows the 3-D plot among input1, input2, and output.

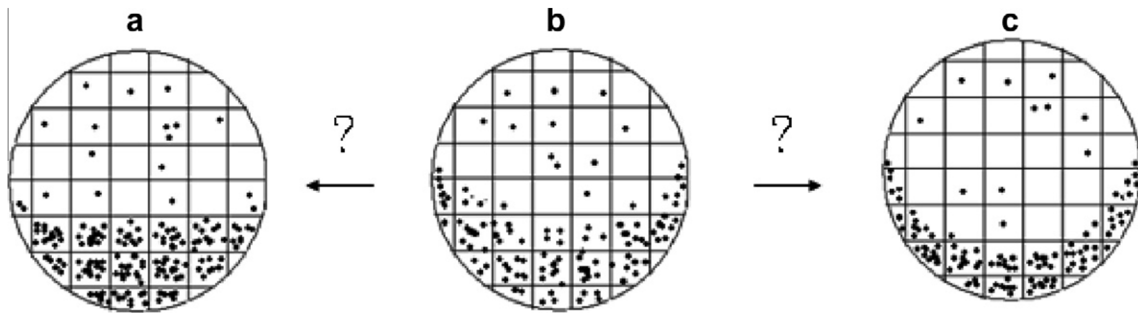

Fig. 3. The fuzziness of clustering pattern recognition.

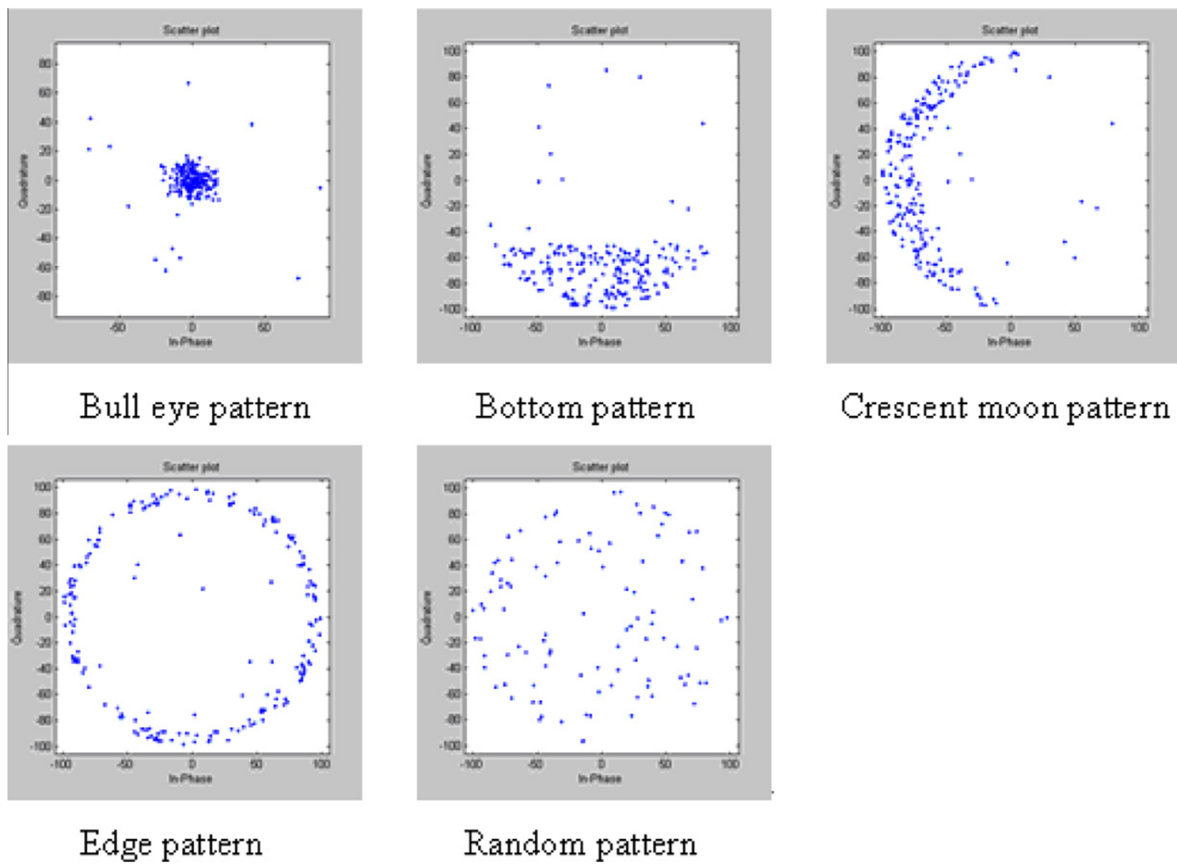

Fig. 4. The five clustering patterns.

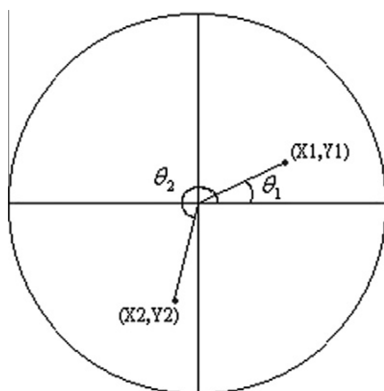

Fig. 5. Angle of each defect coordinate on a wafer.

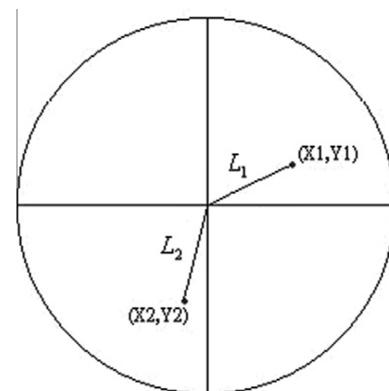

Fig. 6. Distance of each defect coordinate on a wafer.

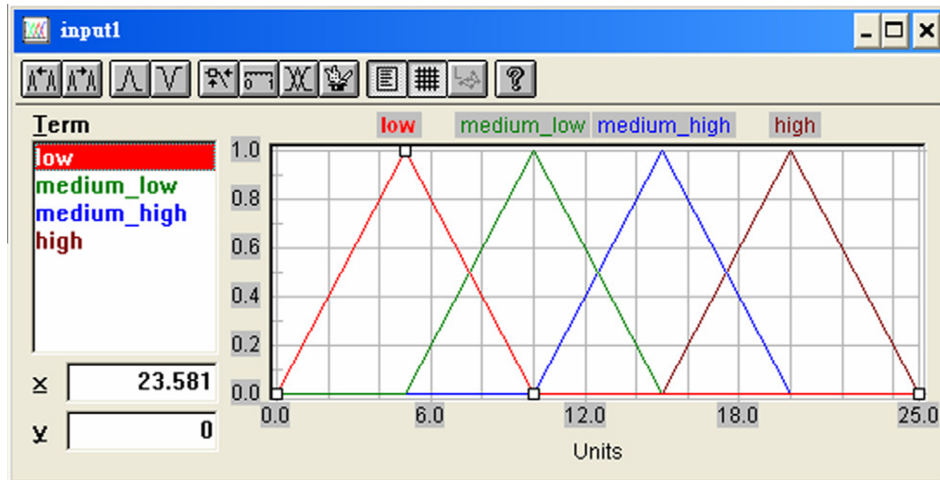

Fig. 7. The membership function of input1.

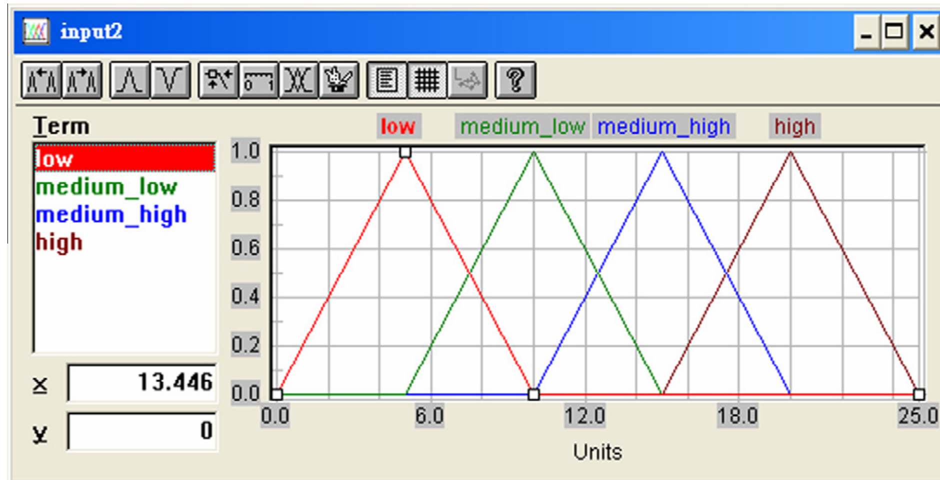

Fig. 8. The membership function of input2.

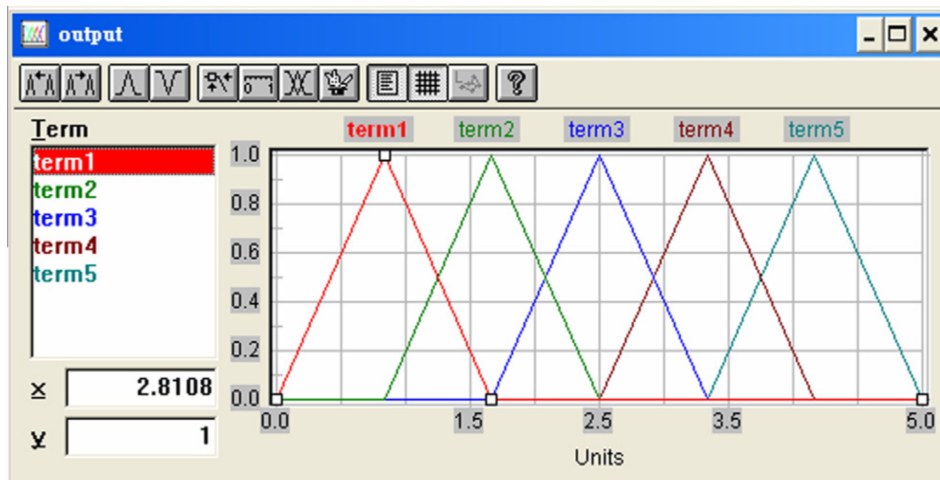

Fig. 9. The membership function of output.

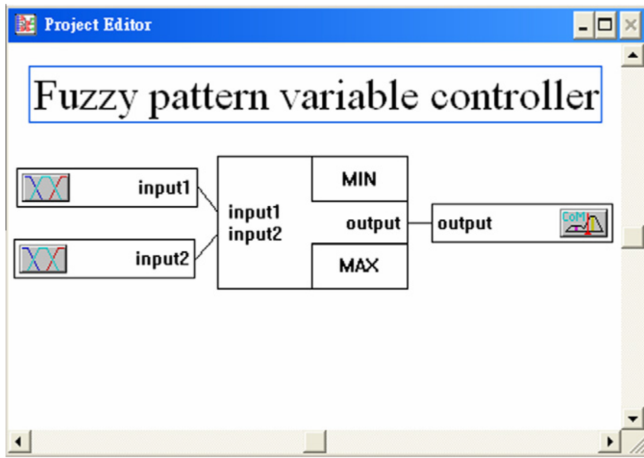

Fig. 10. The architecture of fuzzy pattern variable controller.

3.2. Prepare the relative data per wafer

In this study, defect counts per wafer, the value of *CI* per wafer, and the value of FVCP per wafer are utilized as the input variables for ACO-BPNN. The value of actual yield per wafer is the output variable for ACO-BPNN. Follows are brief descriptions for the obtainment of *CI*, FVCP, and the actual wafer yield.

(1) Calculate the value of *CI*.
The clustering phenomenon of defects on a wafer influences the accuracy of a wafer yield model, and the *CI* can effectively measure the clustering phenomenon on a wafer. The *CI* can be obtained by the calculating manner introduced in Section 2.2.

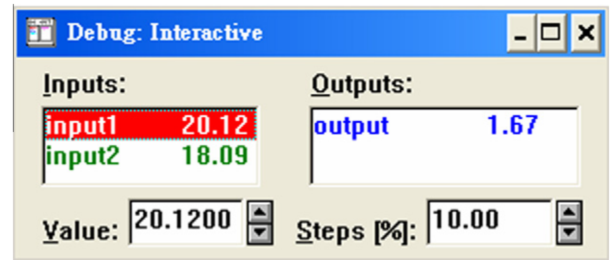

Fig. 12. The outcome of fuzzy logical inference.

(2) Obtain the value of FVCP.
In order to consider the fuzziness of clustering pattern recognition, this study utilizes the new FVCP to recognize the clustering pattern per wafer. The FVCP can be obtained by the developing procedure of six steps introduced in Section 3.1.

(3) Calculate the value of actual wafer yield.
The actual yield value can be obtained by the number of non-defective chips divided by the total number of chips on a wafer.

3.3. Construct the ACO-BPNN model

In this study, ACO is adopted to determine the parameters of BPNN (e.g., the number of neurons in the hidden layers, the momentum, and the learning rate). Because ACO can intelligently find out the optimal parameters of BPNN by pheromone principle, it can make BPNN model to avoid local optimum and has high accuracy.

First, the optimal parameters of BPNN are determined by ACO, and then construct the proposed wafer yield model using BPNN

File Edit Debug Analyzer Compile Neuro Options Window Help

	IF		THEN	
	input1	input2	DoS	output
1	low	low	1.00	term1
2	low	medium_low	1.00	term1
3	low	medium_high	1.00	term4
4	low	high	1.00	term4
5	medium_low	low	1.00	term5
6	medium_low	medium_low	1.00	term5
7	medium_low	medium_high	1.00	term4
8	medium_low	high	1.00	term4
9	medium_high	low	1.00	term5
10	medium_high	medium_low	1.00	term5
11	medium_high	medium_high	1.00	term1
12	medium_high	high	1.00	term1
13	high	low	1.00	term3
14	high	medium_low	1.00	term3
15	high	medium_high	1.00	term2
16	high	high	1.00	term2

Ready. Design Mode

Fig. 11. The rules of fuzzy logic controller.

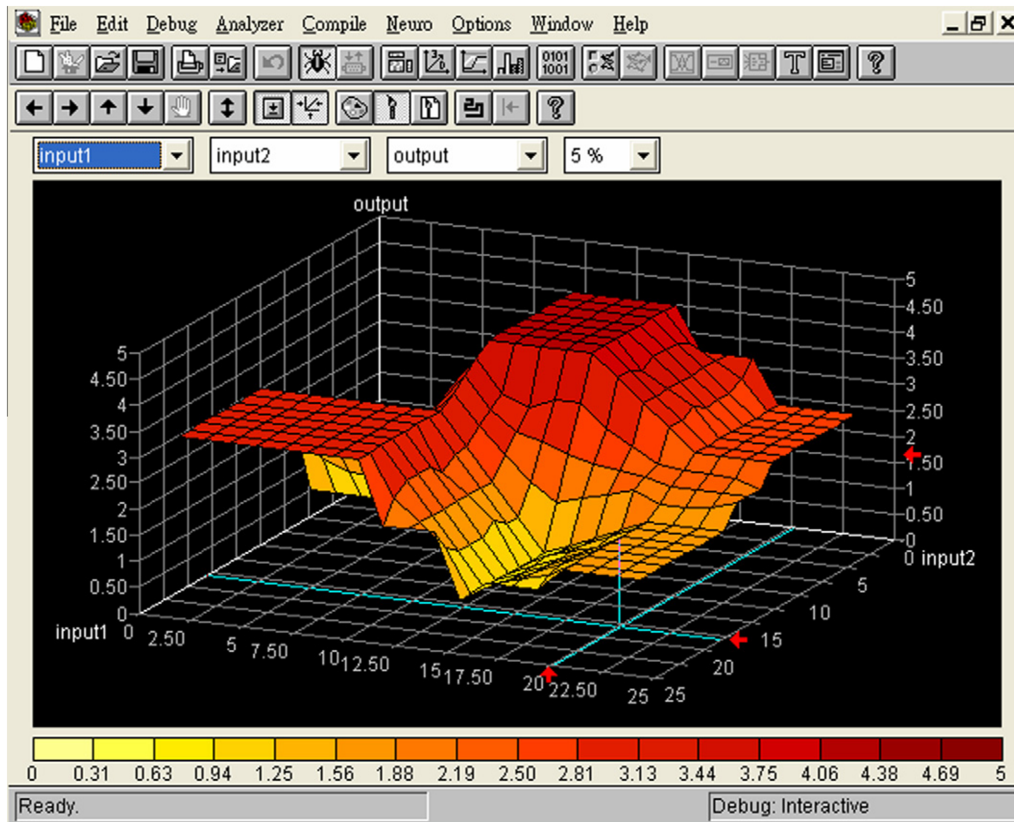

Fig. 13. The 3-D plot among input1, input2, and output.

to train and test the samples simulated by Matlab 7.0 programming language.

3.4. Verify the proposed model

The accuracy of neural networks can be measured by a root-mean squared error (RMSE). When the value of RMSE is smaller, the accuracy of neural networks is higher. The RMSE can be calculated as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(A_i - O_i)^2}{n}} \tag{12}$$

where n represents the number of data, A_i represents the actual value of output, and O_i represents the predicted value. The general indicator for measuring the strength of the relationship between the actual and predicted outputs is the Pearson's linear correlation coefficient r. In this study, RMSE and r are both used to evaluate the performance of wafer yield model.

4. Implementation

In this study, Matlab 7.0 programming language is used to simulate the coordinates of defect points and the five clustering patterns (i.e., bull eye pattern, edge pattern, bottom pattern, crescent moon pattern, and random pattern) on 8-in. wafer are generated. The proposed wafer yield model is constructed by using the procedure introduced from Sections 3.1–3.4.

Comparisons are also made among the Negative Binomial yield model, the BPNN (without FVCP) yield model, and the proposed ACO-BPNN (with FVCP) yield model to demonstrate that the proposed model is superior.

4.1. Simulation study

This section presents a simulation study to demonstrate the effectiveness of the proposed approach. The followings are brief descriptions of these four design factors for this simulation study:

(1) The four kinds of clustering patterns (i.e., bull eye pattern, edge pattern, bottom pattern, and crescent moon pattern) are designed to have three levels (50%, 70%, and 90%) of clustering degree respectively. Therefore, 12 ($4 \times 3 = 12$) kinds of the simulated clustering patterns are generated.
(2) The defect counts of one random and four kinds of clustering patterns are all designed to have five levels (50, 100, 150, 200, and 250), thus 65 ($(1 \times 5) + (12 \times 5) = 65$) kinds of simulated wafer data are generated.
(3) Each kind of simulated wafer data is simulated by ten times repeatedly, so it generates 650 ($65 \times 10 = 650$) simulated wafer data totally. The defect counts, *CI*, FVCP, and the actual wafer yield value are obtained on each simulated wafer data respectively.
(4) Five hundred and twenty simulated wafer data are randomly selected as training samples, and the rest 130 simulated wafer data are the testing samples.

4.2. The results of simulation study

Assumes that each wafer is divided into 400 chips, and there are 650 simulated wafer data generated by Matlab 7.0 programming language totally. These simulated wafer data are listed in Table 1.

Five hundred and twenty simulated wafer data are randomly selected as training samples, and the rest 130 simulated wafer

Table 1
The 650 simulated wafer data in this simulation study.

No.	Defect counts	*CI*	FVCP	Actual wafer yield	Degree of clustering pattern
1	48	1.12	1.59	0.9002	Bull eye pattern (50%)
2	102	1.27	1.81	0.9112	Bull eye pattern (70%)
3	137	2.51	2.12	0.9265	Bull eye pattern (90%)
4	95	1.23	2.64	0.8523	Edge pattern (50%)
5	203	1.29	2.88	0.8017	Edge pattern (70%)
6	234	2.39	3.13	0.7996	Edge pattern (90%)
...	...	...	...	...	...
648	140	0.79	0.91	0.7231	Random pattern
649	171	0.85	0.96	0.7004	Random pattern
650	266	0.73	1.14	0.6605	Random pattern

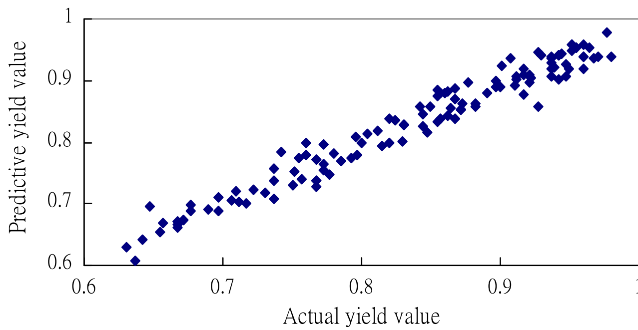

Fig. 14. The scatter plot in the Negative Binomial yield model.

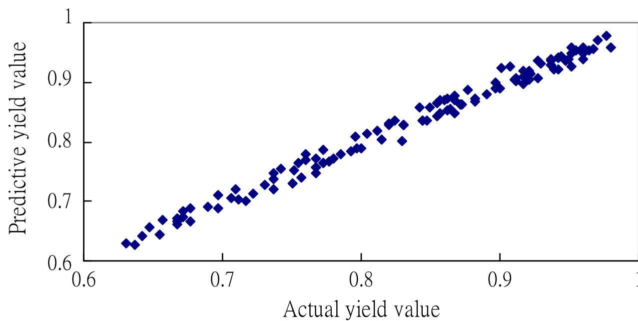

Fig. 15. The scatter plot in the BPNN (without FVCP) yield model.

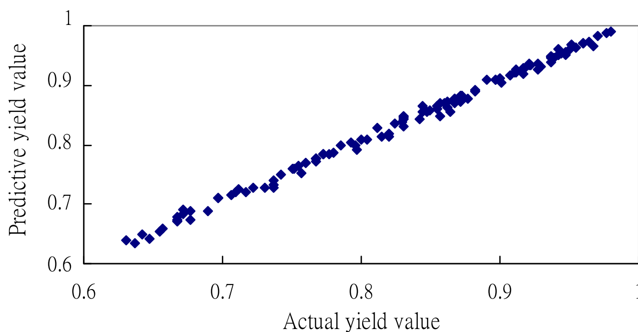

Fig. 16. The scatter plot in the proposed ACO-BPNN (with FVCP) yield model.

data are the testing samples. The RMSE is fitness function in ACO. The network architecture determined by ACO is 3-2-2-1 (i.e., the number of neurons in the input layer is 3, the number of neurons in the first hidden layer is 2, the number of neurons in the second hidden layer is 2, and the number of neurons in the output layer is 1). The network parameters determined by ACO are given: learning rate is 0.25, momentum is 0.88, and train the data through 2500 times.

Table 2
Comparisons of RMSE and *r* between predictive and actual yield value.

Yield model	RMSE	*r*
Negative Binomial yield model	0.0437	0.9159
BPNN (without FVCP) yield model	0.0259	0.9332
Proposed ACO-BPNN (with FVCP) yield model	0.0106	0.9538

The scatter plots in the Negative Binomial yield model, the BPNN (without FVCP) yield model, and the proposed ACO-BPNN (with FVCP) yield model are shown from Figs. 14–16.

Finally, comparisons made among the Negative Binomial yield model, the BPNN (without FVCP) yield model, and the proposed ACO-BPNN (with FVCP) yield model are listed in Table 2. From Table 2, we can see that the proposed model in this study has the smallest value of RMSE and the largest value of correlation coefficient. Therefore, the predictive accuracy of the proposed model in this study is indeed superior.

5. Conclusions

In fact, the recognition of clustering pattern usually exists fuzziness. However, the wafer yield models in previous studies did not consider the fuzziness of clustering pattern belonging to which shape in recognition. Therefore, this study develops a new FVCP to recognize the clustering pattern, and construct a wafer yield model based on FVCP. The proposed FVCP considers the fuzziness of clustering pattern recognition, and it can make the clustering pattern recognition to match the actual situation closely.

The proposed model is an intelligent kind of neural networks combining evolutionary computation and neural networks theory. Because in the proposed model, the auto-adaptability of evolutionary computation and learning capability of neural networks can be combined effectively, it has become the inevitable tendency of neural networks.

The merits of the proposed approach are summarized as follows:

(1) The proposed FVCP is helpful in improving the accuracy of the wafer yield prediction model.
(2) ACO can intelligently find out the optimal parameters of neural networks using pheromone principle. Furthermore, it can make the neural network model to avoid achieving convergence untimely leading to local optimum and thus the neural network model can has high accuracy stably.

(3) The proposed model can help the IC manufacturers to manage the wafer yield and evaluate their process capability in relation to profit and loss.

Acknowledgements

The author would like to thank the National Chiao Tung University for its resourceful support.

References

Blum, C., & Sampels, M. (2004). An ant colony optimization algorithm for shop scheduling problems. *Journal of Mathematical Modelling and Algorithms, 3*(3), 285–308.

Chen, F. L., Chen, Y. C., & Kuo, J. Y. (2010). Applying moving back-propagation neural network and moving fuzzy-neuron network to predict the requirement of critical spare parts. *Expert Systems with Applications, 37*, 6695–6704.

Colorni, A., Dorigo, M., Maniezzo, V. (1991). Distributed Optimization by Ant Colonies. In *Proceedings of the first European conference on artificial life*.

Cunningham, J. A. (1990). The use and evaluation of yield models in integrated circuit manufacturing. *IEEE Transactions on Semiconductor Manufacturing, 3*(2), 60–71.

Dorigo, M., & Gambardella, L. M. (1997). Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation, 1*(1), 53–66.

Dorigo, M., & Stutzle, T. (2004). *Ant colony optimization*. Cambridge: MIT Press.

Dupret, Y., Kielbasa, R. (2004). Modeling semiconductor manufacturing yield by test data and partial least squares. In *Proceedings of 16th international conference on microelectronics* (pp. 404–407).

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

Jun, C. H., Hong, Y., Kim, S. Y., Park, K. S., & Park, H. (1999). A simulation-based semiconductor chip yield model incorporating a new defect cluster index. *Microelectronics Reliability, 39*(4), 451–456.

Kim, C., & Baldwin, D. F. (2005). A theoretical yield model for assembly process of area array solder interconnects packages with experimental verification. *IEEE Transactions on Electronics Packaging Manufacturing, 28*(4), 344–354.

Kumar, N., Kennedy, K., Gildersleeve, K., Abelson, R., Mastrangelo, C. M., & Montgomery, D. C. (2006). A review of yield modeling techniques for semiconductor manufacturing. *International Journal of Production Research, 44*(23), 5019–5036.

Langford, R. E., Liou, J. J., & Raghavan, V. (2001). The application and validation of a new robust windowing method for the Poisson yield model. In *Advanced Semiconductor Manufacturing Conference, IEEE/ SEMI* (pp. 157–160).

Liou, J. J., Zhang, Q., McMacken, J., Thomson, J. R., Stiles, K., & Layman, P. (2002). Statistical modeling of MOS devices for parametric yield prediction. *Microelectronics Reliability, 42*(4), 787–795.

Liu, Y. P., Wu, M. G., & Qian, J. X. (2006). Predicting coal ash fusion temperature based on its chemical composition using ACO-BP neural network. *Thermochimica Acta, 454*, 64–68.

Mamdani, E. H., & Assilian, S. (1981). An experiment in linguistic synthesis with a fuzzy logic controller. *Fuzzy Reasoning and its Applications*.

Maniezzo, V., & Colorni, A. (1999). The ant system applied to the quadratic assignment problem. *IEEE Transactions on Knowledge and Data Engineering, 11*(5), 769–778.

Meyer, F. J., & Park, N. (2003). Predicting defect-tolerant yield in the embedded core context. *IEEE Transactions on Computers, 52*(11), 1470–1479.

Neyer, T., & Hafner, M. (2004). Yield learning using the defect reticle method. *Advanced Semiconductor Manufacturing Conference*, 110–114.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536 (London).

Stapper, C. H. (1973). Defect density distribution for LSI yield calculations. *IEEE Transactions on Electron Devices, 20*(7), 655–657 (Correspondence).

Stapper, C. H. (1985). The effects of wafer to wafer defect density variations on integrated circuit defect and fault distributions. *IBM Journal of Research Development, 29*(1), 87–97.

Stapper, C. H. (1991). On Murphy's yield integral. *IEEE Transactions on Semiconductor Manufacturing, 4*(4), 294–297.

Stapper, C. H., & Rosner, R. J. (1995). Integrated circuit yield management and yield analysis: Development and implementation. *IEEE Transactions on Semiconductor Manufacturing, 8*(2), 95–102.

Tong, L. I., Lee, W. I., & Su, C. T. (1997). Using a neural network-based approach to predict the wafer yield in integrated circuit manufacturing. *IEEE Transactions on Components, Packaging, and Manufacturing Technology – Part C, 20*(4), 288–294.

Tong, L. I., & Chao, L. C. (2008). Novel yield model for integrated circuit with clustered defects. *Expert Systems with Applications, 34*, 2334–2341.

Wei, G. (2007). Study on evolutionary neural network based on ant colony optimization. In *international conference on computational intelligence and security workshops, DEC* (pp. 15–19).

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.