Applied Soft Computing 12 (2012) 1733–1740

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

A systematic estimation model for fraction nonconforming of a wafer in semiconductor manufacturing research

Jun-Shuw Lin

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

ARTICLE INFO

Article history:
Received 15 February 2011
Received in revised form 22 January 2012
Accepted 23 January 2012
Available online 21 February 2012

Keywords:
Semiconductor manufacturing
Estimation for fraction nonconforming
Back-propagation neural network
Genetic algorithm
Self-organized map

ABSTRACT

The clustering phenomenon of defects usually occurs in semiconductor manufacturing. However, previous studies did not pay much attention to the influence of clustering phenomenon for estimating fraction nonconforming of a wafer. Thus, this paper presents a systematic estimation model with considering relevant variables about clustering defects for fraction nonconforming of a wafer. The method combines back-propagation neural network (BPNN) with genetic algorithm (GA) to obtain an estimation model. In this study, GA aims to optimize the parameters of BPNN. Five relevant variables: number of defects (*ND*), squared coefficient of angle variation (SCV_A) for defects, squared coefficient of distance variation (SCV_D) for defects, defect cluster index (CI_M), and the number of cluster groups (*NCG*) for defects by self-organized map (SOM) are utilized as inputs for GA–BPNN. Finally, a simulation case and a real-world case are used to confirm the effectiveness of proposed method.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Semiconductor manufacturing has become the major industry worldwide, and all electrical appliances are closely linked with integrated circuits (ICs). The fraction nonconforming of a wafer is a key index for IC manufacturers to evaluate their process capability. The fraction nonconforming of a wafer is defined as the probability that a chip on a wafer has defects. Accurate estimation for fraction nonconforming of a wafer is very useful to decrease manufacturing costs for products still under development [1], which can offer a reasonable and acceptable price to customers. In many performance technologies [2], estimating for fraction nonconforming of a wafer is one of the most widely researched approaches in semiconductor manufacturing.

As the wafer size increases, the clustering phenomenon of defects becomes significant. In previous literatures, the negative binomial yield model [3] includes a clustering index (α), but the value of α can be scattered and negative that leads to unhandy analysis [4]. Tyagi and Bayoumi [5,6] proposed a variance/mean ratio (V/M) to measure the strength of defects clustered. The values of (V/M) depend on how the grids are selected and cannot indicate the gradualness of cross-wafer defect density variations [5,6]. Jun et al. [7] proposed a cluster index (*CI*) to evaluate the strength of defects clustered on a wafer. In some cases, *CI* values calculated from different defect patterns may be insensitive. For the purpose of overcoming above-mentioned drawbacks, this study uses the cluster index (CI_M) of defects to measure the strength of clustering defects. The advantage of CI_M is sensitive to the percentage of defects in the clustering area and can be generalized to any wafer size.

In addition to cluster index of defects, various clustering patterns of defects and the number of cluster groups (*NCG*) for defects can also influence the fraction nonconforming of a wafer, respectively. Therefore, this study considers one random pattern and three common clustering patterns (i.e., bull eye pattern, bottom pattern, and crescent moon pattern) [8] for estimating fraction nonconforming of a wafer. The self-organized map (SOM) [9,10] is used to determine the *NCG* for defects. Concurrent consideration for clustering defects, clustering patterns, and cluster groups is the novelty of this study different from others.

Previous studies did not pay much attention to the influence of clustering phenomenon for estimating fraction nonconforming of a wafer. Thus, this paper presents a systematic estimation model with considering relevant variables about clustering defects for fraction nonconforming of a wafer. The methodology combines back-propagation neural network (BPNN) [11,12] with genetic algorithm (GA) [13] to obtain an estimation model. In this study, the GA aims to optimize the parameters of BPNN. The advantages of this methodology and outperforming other approaches are that the GA–BPNN model may enhance the likelihood of global optimum solution to have better accuracy for estimating fraction nonconforming of a wafer. Five relevant variables: number of defects (*ND*), squared coefficient of angle variation (SCV_A) for defects, squared coefficient of distance variation (SCV_D) for defects, defect cluster

E-mail address: junsoon1@hotmail.com

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2012.01.019

index [14], and the *NCG* are utilized as inputs for the GA–BPNN. The actual value of fraction nonconforming of a wafer is utilized as output for the GA–BPNN.

In summary, the research aim of this paper is to construct a systematic estimation model with high accuracy to support IC manufacturers monitoring the production status of wafers. Finally, a simulation case and a real-world case are used to confirm the effectiveness of proposed method. Comparisons are also made among the BPNN model, radial basis function neural network (RBFNN) model, and the proposed GA–BPNN model to demonstrate that this proposed method is indeed superior.

2. Related literature

The strength of clustering phenomenon for defects on a wafer can be measured by a cluster index of defects. Therefore, the cluster parameter (α) of the negative binomial model, the variance/mean ratio (V/M), and the non-parameters assumption cluster index are respectively introduced. The negative binomial yield model can be expressed as follows:

$$Y = \frac{1}{(1+\bar{\lambda}/\alpha)^{\alpha}} \tag{1}$$

where α is the cluster parameter and $\bar{\lambda}$ is the mean number of defects per chip. Cunningham [4] reported that, cluster parameter α in the negative binomial model may be quite scattered and may even have a negative value when the model is used to forecast yield.

Tyagi and Bayoumi [5,6] utilized various grid sizes superimposed on a wafer map to measure the strength of defects distributed on a wafer. The defects contained within each grid can be used to judge the spatial distribution of defects. The distribution of defects follows a Poisson distribution if the defects are randomly distributed. Because both variance (V) and mean (M) are equal in the Poisson distribution, the value of (V/M) equals 1 if the wafer defects are randomly scattered. The value of (V/M) exceeds 1 if the defects distributed on a wafer are clustered. The (V/M) values can be proven to possess a t distribution with $n-1$ degrees of freedom, and it can be expressed as follows:

$$t_{n-1} = \frac{(V/M-1)}{\sqrt{2/(n-1)}} \tag{2}$$

where n denotes the number of squares, V and M represent the variance and mean, respectively. The values of (V/M) depend on how the grids are selected and cannot indicate the gradualness of cross-wafer defect density variations.

Jun et al. [7] proposed a cluster index of defects by the projected x and y coordinates of defect locations on a wafer. The clustering of defects tends to show clumps in the x and the y coordinates, which result in a large variance in defect intervals. However, showing clumps either on the x-axis or on the y-axis does not necessarily represent the clustering defects. The value of *CI* is close to 1 when the defects are randomly scattered, and the value of *CI* is expected to be greater than 1 when clustering of defects appears.

3. The proposed method

3.1. GA–BPNN model scheme

The GA [13,15–17] is probabilistic heuristic search processes based on natural selection. And its modern form is derived mainly from Holland's work [13] and the second edition of Holland's classic 1992 book "Adaptation in Natural and Artificial Systems" [18]. GA is capable of solving wide range of complex optimization problems only using three simple genetic operations (selection, crossover and mutation) on coded solutions (strings, chromosomes) for the parameter set, not the parameters themselves in an iterative fashion. GA considers several points in the search space simultaneously, which reduces the chance of convergence to a local optimum [19,20]. Adopting the GA to create the initial value of parameters for BPNN and select the initialized BP weights, the hybrid of GA–BPNN model improves the accuracy of estimation.

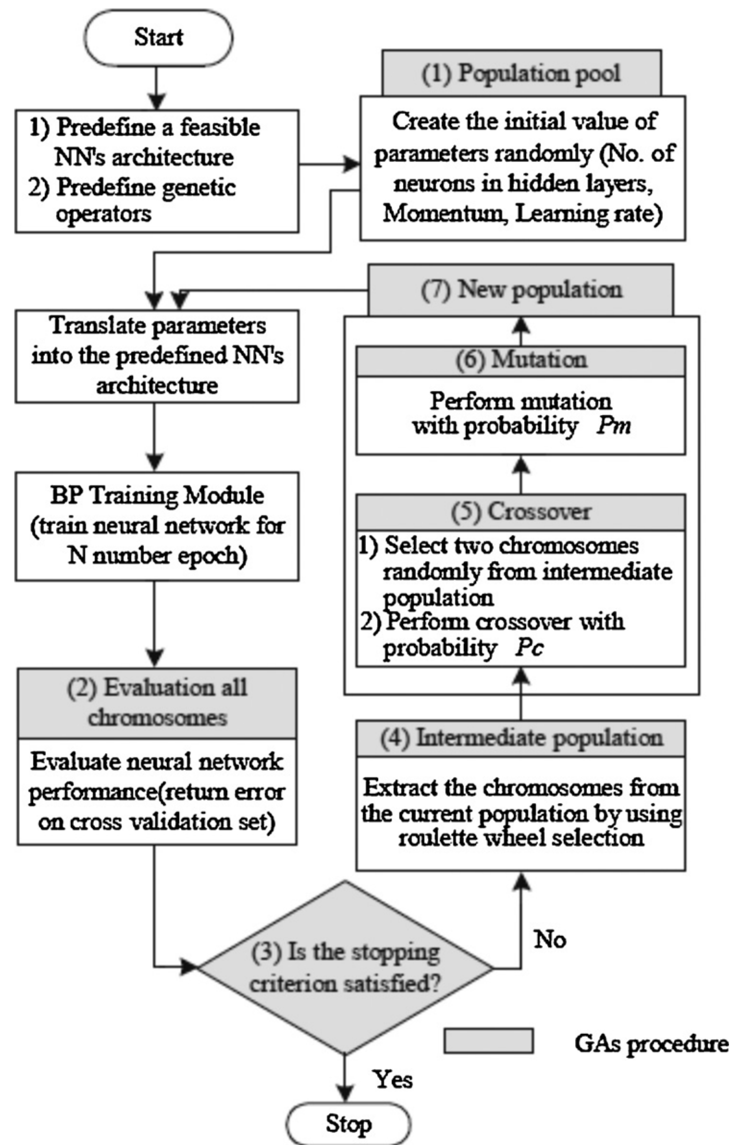

Fig. 1. Using the GA to optimize parameters of BPNN.

The hybrid network learning process consists of two stages: firstly employing GA to search for optimal or approximate optimal connection weights and thresholds for the BPNN, then using the BP to adjust the final weights, in which the sigmoid function is used as the activation function. The steps of learning optimal value for network weights are achieved using the hybrid of GA–BP algorithm as shown in Fig. 1. At first, the populations initialization is done; then performance of tournament selection; followed by crossover with probability P_c and mutation with probability P_m. Inverse value of the learning error is taken as the fitness function that is calculated to find the best fitness population member. If the GA terminating condition is false, the program returns for tournament selection; otherwise, it continues to select potential candidates and compute holdout sample weights.

3.2. Cluster index (CI_M) of defects

This study uses the cluster index [14] of defects to measure the strength of the clustering phenomenon on a wafer. The CI_M is developed by the rotating axis technique from multivariate analysis to overcome the drawbacks of CI. Some points are located in a two-dimensional space (that is, x_1 and x_2) and a new coordinate x_1^* is obtained by rotating the x_1 axis counterclockwise using θ angle, where $0^\circ \le \theta \le 180^\circ$. Accordingly, these points in two-dimensional space can be projected into the new axis x_1^*. The corresponding coordinates are determined as follows.

$$x_1^* = \cos\theta x_1 + \sin\theta x_2 \tag{3}$$

The detailed descriptions of obtaining CI_M are listed as the following five steps.

Step 1: Project the defect coordinates (X_i, Y_i) into a new axis obtained by rotating the x-axis counterclockwise using θ°. Suppose that a wafer has n defects, and (X_i, Y_i) denotes the x and y coordinates of the ith defect location in a two-dimensional space, $i = 1, \ldots, n$. These n defects then can be projected onto a new axis $X_{i,\theta}^*$ obtained by rotating the x-axis counterclockwise using θ°. The new coordinates for the ith defect with respect to θ then can be calculated as follows.

$$X_{i,\theta}^* = \cos\theta X_i + \sin\theta Y_i \tag{4}$$

where i denotes the ith defect and θ represents a rotating angle, where $0^\circ \le \theta \le 180^\circ$.

Step 2: Sort the $X_{i,\theta}^*$ values in ascending order and calculate the intervals between each adjacent coordinate value $X_{i,\theta}^*$. The intervals between each adjacent coordinate value $X_{i,\theta}^*$ then can be calculated as follows.

$$V_{i,\theta} = X_{(i,\theta)}^* - X_{(i-1,\theta)}^* \tag{5}$$

where $V_{i,\theta}$ represents the ith interval between $X_{(i,\theta)}^*$ and $X_{(i-1,\theta)}^*$.

Step 3: Calculate the squared coefficient of variation (SCV) for $V_{i,\theta}$. The SCV for $V_{i,\theta}$ can be determined as follows.

$$SCV_\theta = \frac{S_{v,\theta}^2}{\bar{V}_\theta^2} \tag{6}$$

where SCV_θ represents the squared coefficient of variation for $V_{i,\theta}$, $\bar{V}_\theta = \left(\sum_{i=1}^{n} V_{i,\theta}\right)/n$, and $S_{V,\theta}^2 = \left(\sum_{i=1}^{n} (V_{i,\theta} - \bar{V}_\theta)^2\right)/(n-1)$.

Step 4: Change the angle of θ and calculate the corresponding $\theta = 1^\circ$ value. The number of 180 SCV_θ values with respect to θ, increased by $\theta = 1^\circ$, can be obtained through Steps 1–3.

Step 5: According to the SCV_θ values obtained from Step 4, the average SCV_θ value determines the clustering index, as follows:

$$CI_M = \frac{\sum_{\theta=0^\circ}^{180^\circ} SCV_\theta}{180^\circ} \tag{7}$$

where CI_M represents defect cluster index. A larger CI_M value indicates a stronger degree of defect clustering formed on a wafer.

3.3. The variation of angle and distance for defects

Further, squared coefficient of angle variation and squared coefficient of distance variation are also utilized as relevant variables. The SCV_A and SCV_D can be derived as follows:

(1) Calculate the positive included angle (θ_i) between each defect coordinate and the first quadrant on X-axis. $\theta = \tan^{-1}(Y_i/X_i)$, $i = 1, 2, \ldots, n$; where X_i is the coordinate for the ith defect on X-axis, Y_i is the coordinate for the ith defect on Y-axis, and n is the number of defects. Then, arrange θ_i in increasing order and define $A_i = \theta_{(i)} - \theta_{(i-1)}$, $i = 1, 2, \ldots, n$; where $\theta_{(i)}$ is the ith smallest included angle, n is the number of defects, and $\theta_{(0)} = 0$. Finally, the SCV_A can be obtained by follows:

$$SCV_A = \left(\frac{S_A}{\bar{A}}\right)^2 \tag{8}$$

where S_A and $\bar{A}$ denote the sample standard deviation and mean value of A_i, respectively.

(2) Calculate the distance (L_i) between each defect coordinate and the origin on the coordinate axis. $L_i = \sqrt{X_i^2 + Y_i^2}$, $i = 1, 2, \ldots, n$; where X_i is the coordinate for the ith defect on X-axis, Y_i is the coordinate for the ith defect on Y-axis, and n is the number of defects. Then, arrange L_i in increasing order and define $D_i = L_{(i)} - L_{(i-1)}$, $i = 1, 2, \ldots, n$; where $L_{(i)}$ is the ith smallest distance, n is the number of defects, and $L_{(0)} = 0$. Finally, the SCV_D can be obtained by follows:

$$SCV_D = \left(\frac{S_D}{\bar{D}}\right)^2 \tag{9}$$

where S_D and $\bar{D}$ denote the sample standard deviation and mean value of D_i, respectively.

3.4. Verify the proposed model

The performance of neural networks can be evaluated by a root-mean squared error (RMSE). When the value of RMSE is smaller, the prediction accuracy of neural networks is higher. The RMSE can be obtained as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(A_i - O_i)^2}{n}} \tag{10}$$

where n represents the number of data, A_i represents the actual value of output, and O_i represents the predicted value. Another indicator for measuring the strength of the relationship between the actual and predicted outputs is the Pearson's linear correlation coefficient r. In addition to the RMSE and the Pearson's linear correlation coefficient r, the mean absolute deviation (MAD) and the mean absolute percentage error (MAPE) are also considered for comparing the results of neural networks. The MAD can be obtained as follows:

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|A_i - O_i| \tag{11}$$

where n represents the number of data, A_i represents the actual value of output, and O_i represents the predicted value. Subsequently, the MAPE can be obtained as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|A_i - O_i|}{A_i} \times 100\% \tag{12}$$

where n represents the number of data, A_i represents the actual value of output, and O_i represents the predicted value. In summary, there are four measures (RMSE, r, MAD, and MAPE) to evaluate the estimation performance of neural network models in this study.

4. Simulation

The computer software, Matlab 7.0, is used to simulate the coordinates of defect points and the four defect patterns (i.e.,

(a) Random defect-pattern

(b) Bull eye defect-pattern

(c) Bottom defect-pattern

(d) Crescent moon defect-pattern

Fig. 2. The four defect patterns.

random pattern, bull eye pattern, bottom pattern, and crescent moon pattern) on 8 in. wafer are generated. The four defect patterns are shown in Fig. 2. Five relevant variables are utilized as inputs for the GA–BPNN. The actual value of fraction nonconforming of a wafer is utilized as output for the GA–BPNN. The actual value of fraction nonconforming of a wafer can be obtained by the number of defective chips divided by the total number of chips.

For example, there are hypothetically 104 defect counts on a simulation wafer, shown in Table 1. This study uses the computer

Fig. 3. The SOM plot of cluster groups for these 104 defect counts.

Table 1
The 104 defect counts on a simulation wafer.

No.	*X* coordinate	*Y* coordinate	Distance	Angle
1	−3	−3	4.24	45.00
2	−3	2	3.61	−33.69
3	−2	2	2.83	−45.00
...	...	...	...	...
102	4	1	4.12	14.04
103	4	2	4.47	26.57
104	4	3	5.00	36.87

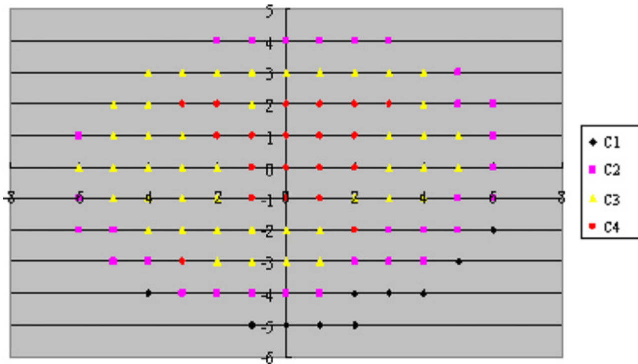

Fig. 4. The scatter plot of 4 cluster groups for 104 defect counts.

software, SOMine, to obtain the SOM of these 104 defect counts on this simulated wafer. Fig. 3 shows the SOM plot of cluster groups for these 104 defect counts on this simulated wafer. According to Fig. 3, the number of cluster groups for defects by SOM is equal to 4. Fig. 4 shows the scatter plot of 4 cluster groups for these 104 defect counts on this simulated wafer. This study uses the same way to obtain the *NCG* of 500 simulation wafer data in Section 4.2.

4.1. Simulation study

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

(1) The three clustering patterns (i.e., bull eye pattern, bottom pattern, and crescent moon pattern) are designed to have three levels (50%, 70%, and 90%) of clustering percentage degree, respectively. Therefore, 9 ($3 \times 3 = 9$) kinds of the simulated clustering patterns are generated.
(2) The defect counts for one random pattern and 9 kinds of the simulated clustering patterns are designed to have five levels (50, 100, 150, 200, and 250), respectively. Therefore, 50 ($(1 \times 5) + (9 \times 5) = 50$) kinds of simulated wafer data are generated.

Table 2
The 500 simulated wafer data in this simulation study.

No.	*ND*	SCV_A	SCV_D	CI_M	*NCG*	Clustering pattern
1	48	3.24	1.12	1.59	2	Bull eye pattern (50%)
2	102	3.58	1.27	2.81	2	Bull eye pattern (70%)
3	137	3.93	2.01	3.12	3	Bull eye pattern (90%)
4	95	2.02	1.23	2.24	2	Bottom pattern (50%)
5	203	2.87	1.29	3.13	3	Bottom pattern (70%)
6	234	2.98	2.39	3.88	4	Bottom pattern (90%)
...	...	...	...	...	...	...
498	140	1.18	0.79	0.13	1	Random pattern
499	171	1.52	0.85	0.10	1	Random pattern
500	266	1.81	0.73	0.11	1	Random pattern

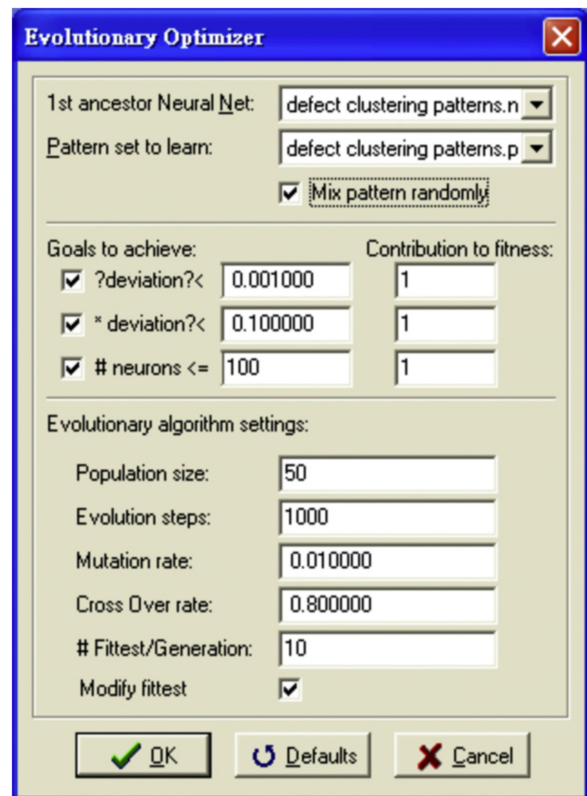

Fig. 5. The setting parameters of GA.

(3) Each simulated wafer data is replicated ten times. Therefore, it generates 500 ($50 \times 10 = 500$) simulated wafer data totally. In each replication, I change the setting of random number seed. Then, 400 simulation wafer data are randomly selected as training samples, the remaining 100 simulation wafer data are the testing samples.

4.2. Simulation result

Assume that each wafer is divided into 396 chips, and there are 500 simulated wafer data generated by Matlab 7.0 totally. These 500 simulated wafer data are listed in Table 2. The computer software, Pythia 1.02, is used to perform GA–BPNN. According to the study of Ting et al. [21], it is reported that the better setting for mutation rate is between 0.01 and 0.2; cross over rate is between 0.8 and 1.0. Further, by try and error method, we find that the parameters (population size is 50, evolution steps are under 1000, mutation rate is 0.01, and cross over rate is 0.8) can obtain the best outcome in this study. Fig. 5 shows the setting parameters of GA. Fig. 6 shows the evolutionary optimization of BPNN topology. From Fig. 6, the BPNN architecture determined by GA is 5–7–5–1 (i.e., the number of neurons in the input layer is 5, the number of neurons in the first hidden layer is 7, the number of neurons in the second hidden layer is 5, and the number of neurons in the output layer is 1). The network

Table 3
Comparisons (simulation study) of RMSE, MAD, MAPE, and *r* between estimative and actual value.

Estimation method	RMSE	MAD	MAPE	*r*
BPNN model	0.0412	0.0205	0.13%	0.8124
RBFNN model	0.0283	0.0097	0.08%	0.8526
Proposed GA–BPNN model	0.0102	0.0013	0.04%	0.9132

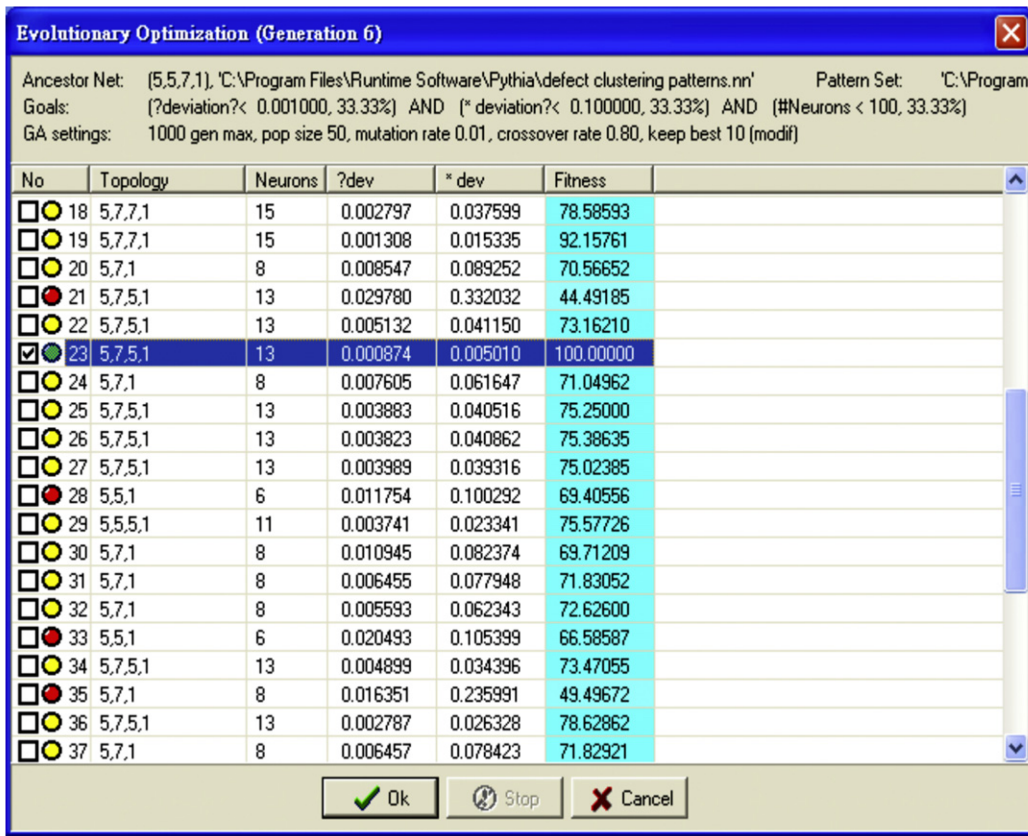

Fig. 6. The evolutionary optimization of BPNN topology.

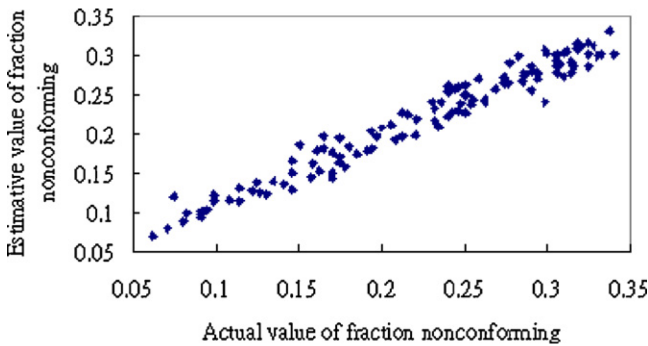

Fig. 7. The scatter plot of BPNN model.

parameters of BPNN are given: learning rate is 0.25; momentum is 0.88, and train the data through 2500 times. Besides, the sigmoid function is used as activation function in the training process.

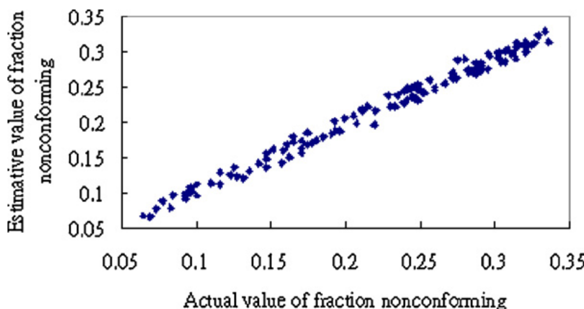

Fig. 8. The scatter plot of RBFNN model.

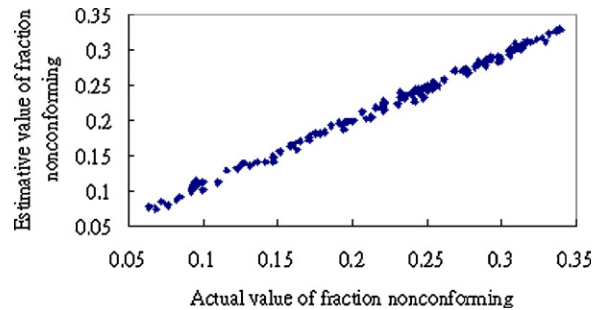

Fig. 9. The scatter plot of GA–BPNN model.

The scatter plots for the BPNN model, the RBFNN model, and the proposed GA–BPNN model are shown from Figs. 7–9. Finally, comparisons (simulation study) made among the BPNN model, the RBFNN model, and the proposed GA–BPNN model are listed in Table 3. From Table 3, it shows that the proposed GA–BPNN model in this study has the smallest value of RMSE, MAD, and MAPE; the largest value of correlation coefficient r. Therefore, the estimation accuracy of the proposed method in this study is indeed superior.

5. Case study

In this section, a real-world case (without simulation) obtained in a DRAM company in Taiwan is utilized to demonstrate the effectiveness of proposed method. There are totally 113 data of 8 in. wafer in this case, and each wafer is divided into 400 chips. In this case, the four defect patterns (i.e., random pattern, bull eye pattern, bottom pattern, and crescent moon pattern) are also considered, and the five relevant variables are also calculated by the same way

Table 4
The 113 real wafer data with relevant variable values in case study.

No.	ND	SCV_A	SCV_D	CI_M	NCG	Clustering pattern
1	92	3.07	1.35	1.13	2	Bull eye pattern
2	85	3.01	1.63	1.91	2	Bull eye pattern
3	104	3.24	2.27	3.03	2	Bull eye pattern
4	156	2.34	1.46	2.56	2	Bottom pattern
5	197	2.56	1.98	3.52	3	Bottom pattern
6	215	2.89	2.54	3.78	3	Bottom pattern
...	...	...	...	...	...	...
111	167	1.25	0.77	0.11	1	Random pattern
112	188	1.34	0.61	0.13	1	Random pattern
113	254	1.68	0.54	0.06	1	Random pattern

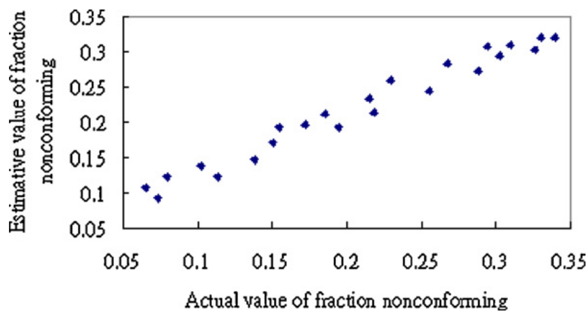

Fig. 10. The scatter plot of BPNN model (case study).

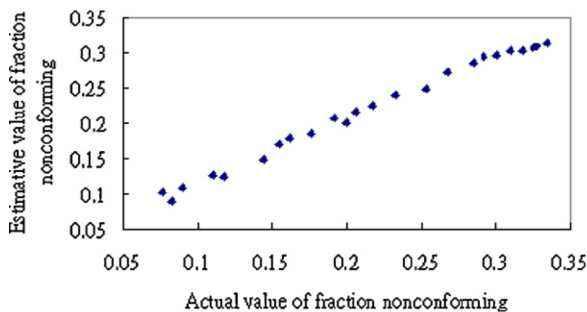

Fig. 11. The scatter plot of RBFNN model (case study).

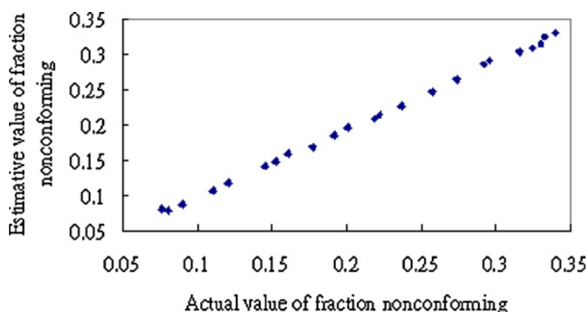

Fig. 12. The scatter plot of GA–BPNN model (case study).

of Section 4. The 113 data with relevant variable values are listed in Table 4. Comparisons are also made among the BPNN model, RBFNN model, and the proposed GA–BPNN model to demonstrate that the proposed method is indeed superior for estimating fraction nonconforming of a wafer.

These 113 wafer data are divided into two parts: one part contains 90 wafers which are used to train the GA–BPNN model; the second part contains 23 wafers which are employed to test the accuracy of the GA–BPNN model. The scatter plots (case study) for the BPNN model, the RBFNN model, and the proposed GA–BPNN model are shown from Figs. 10–12. Finally, comparisons (case study) made among the BPNN model, the RBFNN model, and the proposed GA–BPNN model are listed in Table 5. From Table 5, it shows that the proposed GA–BPNN model in this study has the smallest value of RMSE, MAD, and MAPE; the largest value of correlation coefficient r. Therefore, the estimation accuracy of the proposed method in this study is indeed superior.

Table 5
Comparisons (case study) of RMSE, MAD, MAPE and r between estimative and actual value.

Estimation method	RMSE	MAD	MAPE	r
BPNN model	0.1013	0.0876	0.15%	0.7938
RBFNN model	0.0627	0.0438	0.12%	0.8436
Proposed GA–BPNN model	0.0214	0.0102	0.08%	0.9254

6. Results and discussions

According to the outcome of Sections 4 and 5, the estimation accuracy of GA–BPNN model is better than BPNN (without GA) model or RBFNN model. Although the BPNN and RBFNN are prominent on nonlinear forecasting and pattern recognition, they still have the problems of setting parameters and local optimum on convergence. Therefore, this study uses GA to optimize the parameters (e.g., the connection weight for each layer; the best topology of neural networks) of BPNN. Because the GA is a global optimum searching algorithm, it may make BPNN to enhance the likelihood of global optimum solution, and so this GA–BPNN model can have better accuracy of estimation.

In the aspect of estimating fraction nonconforming of a wafer, the Poisson yield model is inaccurate when the clustering phenomenon of defects is significant. Moreover, the negative binomial model also lacks the accuracy when the clustering phenomenon of defects occurs on a wafer. Thus, it can be seen that the clustering phenomenon of defects is significantly relative to the fraction nonconforming of a wafer. In order to give careful consideration to the influence of clustering phenomenon for estimating fraction nonconforming of a wafer, this study synchronously utilizes the variation of angle-distance, defect cluster index, and number of cluster groups to feature the characteristics of clustered defects and patterns. These relevant variables may aid this GA–BPNN model to estimate fraction nonconforming of a wafer more closely.

The self-organized map is used to determine the NCG for defects in this study. SOM is very suitable for the analysis and visualization of high dimensional data. It converts complex nonlinear relationships between high dimensional input data into simple geometric relationships. SOM learns to find regularities and correlations of input, and roughly preserves the most important relationships and topological of the original data, and clusters future input data accordingly. SOM includes two layers: the input layer and the output layer. The input layer is fully connected to the output layer of map nodes. The learning process is competitive and unsupervised. When an input is presented the output nodes compete to represent the pattern. The node whose vector of weights is closest to the input pattern wins the competition. The winner is then updated by moving its weight vector closer to the input pattern. Units near the winner are also moved, as training progresses units that are neighbors tend to come to represent similar patterns, while nodes far from each other in the map represent dissimilar patterns. The nods within a cluster tend to activate the same output unit, while nods from other clusters will be represented by separate units. Therefore, SOM can effectively identify the NCG for defects.

For the purpose of improving drawbacks of previous studies, this paper presents a systematic estimation model with considering relevant variables about clustering defects for fraction nonconforming of a wafer. By the demonstration of simulation and case

study, the five relevant variables (ND, SCV_A, SCV_D, CI_M, and NCG) tied in with the GA–BPNN model may indeed have superior accuracy for estimating fraction nonconforming of a wafer. Therefore, the methodology proposed in this study may be suggested to be used in semiconductor manufacturing industry.

7. Conclusions

As the wafer size increases, the defects on a wafer become clustered and form specific patterns. Traditional statistical models still appropriately estimate the fraction nonconforming of a wafer when the chip size is small, but they are failed to be used when the clustering phenomenon of defects is significant. In addition to clustering defects, various clustering patterns of defects can also influence the fraction nonconforming of a wafer. In this situation, it is difficult for engineers to estimate the fraction nonconforming of a wafer accurately.

In order to overcome the above-mentioned difficulty, this paper presents a systematic estimation model with considering relevant variables about clustering defects for fraction nonconforming of a wafer. In addition to cluster index of defects, clustering patterns of defects are also considered simultaneously. In this study, five relevant variables: ND, SCV_A, SCV_D, CI_M, and NCG are considered to feature the characteristics of clustered defects and patterns. Then, they are tied in with the GA–BPNN to construct an estimation model for fraction nonconforming of a wafer.

This paper is different from previous studies and contributes to semiconductor manufacturing as follows:

(1) Concurrent consideration for clustering defects, clustering patterns, and cluster groups is the uniqueness of this paper different from others.
(2) The five relevant variables can help the GA–BPNN model to estimate fraction nonconforming of a wafer more effectively.
(3) The proposed methodology may support the semiconductor manufacturing industry to evaluate the process capability in relation to profit and loss.

Acknowledgment

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

References

[1] N. Kumar, K. Kennedy, K. Gildersleeve, R. Abelson, C.M. Mastrangelo, D.C. Montgomery, A review of yield modelling techniques for semiconductor manufacturing, International Journal of Production Research 44 (23) (2006) 5019–5036.
[2] R.C. Leachman, The competitive semiconductor manufacturing survey, in: Proceedings of the IEEE Conference on Semiconductor Manufacturing, 1993, pp. 359–381.
[3] C.H. Stapper, Defect density distribution for LSI yield calculations, IEEE Transactions on Electron Devices 20 (7) (1973) 655–657 (Correspondence).
[4] J.A. Cunningham, The use and evaluation of yield models in integrated circuit manufacturing, IEEE Transactions on Semiconductor Manufacturing 3 (2) (1990) 60–71.
[5] A. Tyagi, M.A. Bayoumi, Defect clustering viewed through generalized Poisson distribution, IEEE Transactions on Semiconductor Manufacturing 5 (3) (1992) 196–206.
[6] A. Tyagi, M.A. Bayoumi, The nature of defect patterns on integrated circuit wafer maps, IEEE Transactions on Reliability 43 (1) (1994) 22–29.
[7] C.H. Jun, Y. Hong, S.Y. Kim, K.S. Park, H. Park, A simulation-based semiconductor chip yield model incorporating a new defect cluster index, Microelectronics Reliability 39 (4) (1999) 451–456.
[8] D.J. Friedman, M.H. Hansen, V.N. Nair, D.A. James, Model-free estimation of defect clustering in integrated circuit fabrication, IEEE Transactions on Semiconductor Manufacturing 10 (3) (1997) 344–359.
[9] S.B. Cho, Ensemble of structure-adaptive self-organizing maps for high performance classification, Information Sciences 123 (2000) 103–114.
[10] T. Kohonen, Self-Organizing Map, Springer, Berlin, 1995.
[11] W.C. Chen, S.W. Hsu, A neural-network approach for an automatic LED inspection system, Expert Systems with Applications 33 (2) (2006) 531–537.
[12] H.R. Maier, G.C. Dandy, Understanding the behavior and optimizing the performance of back-propagation neural networks: an empirical study, Environmental Modeling and Software 13 (1998) 179–191.
[13] J. Holland, Adaptation in Natural and Artificial System, University of Michigan Press, Ann Arbor, Michigan, 1975.
[14] L.I. Tong, C.H. Wang, D.L. Chen, Development of a new cluster index for wafer defects, International Journal of Advanced Manufacturing Technology 31 (2007) 705–715.
[15] M. Gen, R. Cheng, Genetic Algorithms and Engineering Optimization, John Wiley & Sons, New York, 2000.
[16] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, New York, 1989.
[17] G.E. Liepins, M.R. Hilliard, Genetic algorithms: foundations and applications, Annals of Operation Research 21 (1989) 31–58.
[18] J. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology Control and Artificial Intelligence, 2nd ed., MIT Press, 1992.
[19] F.Y. Partovi, M. Anandarajan, Classifying inventory using an artificial neural network approach, Computers & Industrial Engineering 41 (4) (2002) 389–404.
[20] N. Jiang, Z. Zhao, L. Ren, Design of structural modular neural networks with genetic algorithm, Advances in Engineering Software 34 (1) (2003) 17–24.
[21] C.K. Ting, S.T. Li, C.N. Lee, TGA: a new integrated approach to evolutionary Algorithms, in: Proceedings of Congress on Evolutionary Computation, 2001, pp. 917–924.