Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 34 (2008) 2334–2341
Expert Systems with Applications
www.elsevier.com/locate/eswa

Novel yield model for integrated circuits with clustered defects

Lee-Ing Tong, Li-Chang Chao *

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

Abstract

As wafer sizes increase, the clustering phenomenon of defects increases. Clustered defects cause the conventional Poisson yield model underestimate actual wafer yield, as defects are no longer uniformly distributed over a wafer. Although some yield models, such as negative binomial or compound Poisson models, consider the effects of defect clustering on yield prediction, these models have some drawbacks. This study presents a novel yield model that employs General Regression Neural Network (GRNN) to predict wafer yield for integrated circuits (IC) with clustered defects. The proposed method utilizes five relevant variables as input for the GRNN yield model. A simulated case is applied to demonstrate the effectiveness of the proposed model.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Clustered defects; General regression neural network; IC; Pattern; Yield model

1. Introduction

Wafer yield is an important index of success used in integrated circuits (IC) manufacturing. Wafer yield is defined as the probability that a chip on a wafer has no defect. Defects are physical anomalies which result in circuit faults; dirt particles are the primary source of defects in IC manufacturing (Ferris-Prabhu, 1992).

Numerous mathematical models have been developed for predicting wafer yield in the last 40 years (Cunningham, 1990; Stapper, 1991; Stapper & Rosner, 1995; Tyagi & Bayoumi, 1992). Most of these models treat wafer yield as a function of chip size, mean number of defects per chip and the average number of defects per unit area; the Poisson model, compound Poisson models and negative binomial model are examples such models (Cunningham, 1990). The Poisson model is the simplest model to use; however, to successfully predict wafer yield, defect must occur independently with constant probability of occurring in any small area on a wafer (Albin & Friedman, 1991). If these assumptions hold, defects are uniformly scattered over a wafer. However, Stapper (1985) reported that defects are typically clustered rather than dispersed randomly over a wafer, and this distribution becomes more evident as wafer size increases. Clustered defects usually violate the independence assumption of the Poisson model. The Poisson model, therefore, underestimates actual yield when defects cluster.

Under this scenario, numerous yield models obtain more accurate yield predictions than the Poisson model. Compound Poisson yield models are complicated and only evaluate the relationship between chip size and yield (Cunningham, 1990). The cluster parameter α of the negative binomial model can be very scattered and negative when the model is applied to predict yield (Cunningham, 1990). Consequently, these mathematical yield models have particular problems in predicting wafer yield. Dupret and Kielbasa (2004) use the partial least square (PLS) regression methods to model the yield from measurements obtained during the production. However, an advanced statistics is needed to use the PLS regression methods. Neural networks can handle problems such as recognizing complicated patterns and fitting nonlinear functions. Back-Propagation Neural Network (BPNN), known for its general pattern-mapping capability, can be applied to numerous prediction problems and always performs well (Bishop, 1994; Fausett, 1994). However, obtaining good

* Corresponding author. Tel.: +886 35 731 896; fax: +886 35 733 873.
E-mail address: lichang.iem91g@nctu.edu.tw (L.-C. Chao).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.03.013

prediction network requires substantial effort to identify BPNN's parameters, such as the number of hidden layers, number of hidden units, learning rate and momentum. Furthermore, the BPNN model has certain problems such as local optimal solution, overtraining and undertraining. Compared with BPNN, the General Regression Neural Network (GRNN) model has numerous advantages: learning is fast; only one parameter is required; there are no overtraining or undertraining problems; and, the likelihood of obtaining a global optimal solution is higher than with BPNN.

This study presents a novel yield model which employs GRNN to predict wafer yield with clustered defects. The proposed GRNN yield model utilizes five relevant variables as input variables to predict the wafer yield: number of defects; chip size; mean number of defects per chip; mean number of defects per unit area; and, clustering index. The prediction accuracy of the proposed approach is compared with those of the negative binominal yield model and the BPNN yield model. A simulated case is presented to demonstrate the effectiveness of the proposed approach.

2. Yield models

The Poisson yield model (Ferris-Prabhu, 1992), which is based on the Poisson distribution, is

$$Y_1 = P(k = 0) = e^{-\lambda_0}, \tag{1}$$

where k represents the number of defects in a chip and λ_0 represents the mean number of defects per chip. The Poisson yield model was sufficiently effective for small chip sizes and tended to underestimate yields for larger chip sizes (Cunningham, 1990). To identify the clustering properties of defects in the yield model, some spatial distributions, including compound Poisson distributions, have been considered (Raghavachari, Srinivasan, & Sullo, 1997). The compound Poisson yield model replaces defect density, which is assumed to be a constant in the Poisson yield model, with a probability density function. The compound Poisson yield model can be described as

$$Y = \int_0^\infty e^{-DA} f(D)\, dD, \tag{2}$$

where D represents the defect density, A represents the chip size and $f(D)$ represents the probability density function of defects. The compound Poisson yield model is complex and only considers relations between chip size and yield.

The negative binomial yield model, which is a widely applied yield model, employs a gamma function for the distribution of defect density (Okabe, Nagata, & Shimada, 1972; Stapper, 1973). The negative binomial distribution can be described as

$$P(k) = \frac{\Gamma(k+\alpha)(\bar{\lambda}/\alpha)^k}{k!\Gamma(\alpha)(1+\bar{\lambda}/\alpha)^{k+\alpha}}, \quad k = 0, 1, 2, \ldots, \tag{3}$$

where $\bar{\lambda}$ and α are parameters of the negative binomial distribution. The negative binomial yield model is

$$Y_2 = \frac{1}{(1+\bar{\lambda}/\alpha)^\alpha}. \tag{4}$$

Parameter α, called the cluster parameter, can be calculated as

$$\alpha = \frac{\bar{\lambda}^2}{(\sigma^2 - \bar{\lambda})}, \tag{5}$$

where $\bar{\lambda}$ is the mean number of defects per chip and σ^2 is the variance. The negative binomial model has been shown to be a powerful prediction model in IC manufacturing (Cunningham, 1990). However, reports also show that the cluster parameter α in the negative binomial model can be very scattered and negative when the model is used to predict yield (Cunningham, 1990).

Langford, Liou, and Raghavan (2001) presents 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) presents a statistical modeling of MOS devices for parametric yield prediction. Skinner et al. (2002) discuss two classes of traditional multivariate statistical methods and a classification and regression tree (CART) method for modeling and analysis of wafer probe test data to determine the cause of low yield wafers. Meyer and Park (2003) present a center-satellite model to Predicting defect-tolerant yield in the embedded core context. Dupret and Kielbasa (2004) presents partial least square (PLS) regression methods to model the yield from measurements obtained during the production. Hong, Milor, Choi, and Lin (2005) utilize two models which are derived from the Poisson yield model and the negative binomial yield model for the effect of area scaling on IC reliability. Kim and Baldwin (2005) present a theoretical yield model for assembly process of area array solder interconnect process. Other yield models used in various companies are summarized in Stapper and Rosner (1995).

In summary, existing wafer yield models have significant limitations: clustered defects cause the conventional Poisson yield model to underestimate wafer yield; the compound Poisson yield model is too complex; the cluster parameter α of the negative binomial model can be substantially scattered and sometimes negative; and, many parameters must be set when applying the BPNN model. Such drawbacks affect performance when these models are employed to predict yield. The most accurate models are the negative binomial (Cunningham, 1990; Stapper, 1973) and BPNN network (Bishop, 1994; Fausett, 1994) yield models. Only these two models, therefore, were selected for comparison in this study.

3. General regression neural network implementation

The major difference between GRNN and other supervised neural networks is that GRNN can treat continuous valued outputs and categorize data, and there are fewer training parameters are required, such as the number of hidden layers, number of hidden units, learning rate and

momentum, than in BPNN. Moreover, GRNN can be used for any regression problem in which a linearity assumption is violated, and it converges fast on the optimal regression surface as the number of samples becomes substantially large. The GRNN model, then, is used in this study to predict wafer yield.

Fig. 1 shows the three-layer network of the GRNN model (Specht, 1991). Input units are merely distribution units which forward measurement variables to the pattern units in the second (hidden) layer. This hidden layer consists of one neuron for each pattern in the training pattern. The GRNN is essentially trained after one pass of the training patterns and its activation function normally uses an exponential function. The unique parameter of GRNN is the smoothing factor σ which influences the output value; that is, high smoothing factors produce increased relaxed surface fits throughout the data.

Unlike the conventional regression model, GRNN can be defined through its joint continuous probability density function, rather than utilizing a specified function that must be determined in advance. Assume that $f(\mathbf{x}, y)$ represents the known joint continuous probability density function of a vector variable, $\mathbf{x}$, and a scalar random variable, y; the regression of y on $\mathbf{X}$, then, is

$$E[y|\mathbf{x}=\mathbf{X}] = \frac{\int_{-\infty}^{\infty} y f(\mathbf{X}, y)\,\mathrm{d}y}{\int_{-\infty}^{\infty} f(\mathbf{X}, y)\,\mathrm{d}y}. \tag{6}$$

When the density $f(\mathbf{x}, y)$ is unknown, it must be estimated by observations of $\mathbf{x}$ and y. The GRNN model utilizes a Parzen (1962) window, which is a nonparameter approach to estimating the joint continuous probability density function $f(\mathbf{x}, y)$. The estimator can be represented as

$$\hat{f}(\mathbf{X}, Y) = \frac{1}{(2\pi)^{(p+1)/2}\sigma^{(p+1)}} * \frac{1}{n}\sum_{i=1}^{n} \exp\left[-\frac{(\mathbf{X}-\mathbf{X}^i)^{\mathrm{T}}(\mathbf{X}-\mathbf{X}^i)}{2\sigma^2}\right] * \exp\left[-\frac{(Y-Y^i)^2}{2\sigma^2}\right], \tag{7}$$

where p is the dimension of $\mathbf{x}$, σ is the smoothing parameter, $\mathbf{X}^i$ and Y^i are sample values of observations $\mathbf{x}$ and y; and n is the number of sample observations. Combining Eq. (6) and (7), Eq. (8) can be obtained as

$$\widehat{E}(y|\mathbf{X}) = \widehat{Y}(\mathbf{X}) = \frac{\sum_{i=1}^{n} \exp\left[-\frac{(\mathbf{X}-\mathbf{X}^i)^{\mathrm{T}}(\mathbf{X}-\mathbf{X}^i)}{2\sigma^2}\right] \int_{-\infty}^{\infty} y \exp\left[-\frac{(y-Y^i)^2}{2\sigma^2}\right] \mathrm{d}y}{\sum_{i=1}^{n} \exp\left[-\frac{(\mathbf{X}-\mathbf{X}^i)^{\mathrm{T}}(\mathbf{X}-\mathbf{X}^i)}{2\sigma^2}\right] \int_{-\infty}^{\infty} \exp\left[-\frac{(y-Y^i)^2}{2\sigma^2}\right] \mathrm{d}y}. \tag{8}$$

Eq. (8) can be further simplified as Eq. (9), which is given by

$$\widehat{Y}(\mathbf{X}) = \frac{\sum_{i=1}^{n} Y^i \exp\left[-\frac{D_i^2}{2\sigma^2}\right]}{\sum_{i=1}^{n} \exp\left[-\frac{D_i^2}{2\sigma^2}\right]}, \tag{9}$$

where $D_i^2 = (\mathbf{X}-\mathbf{X}^i)^{\mathrm{T}}(\mathbf{X}-\mathbf{X}^i)$.

The GRNN model utilizes Eq. (9) to estimate y.

Typically, the activation function of GRNN network is exponential, as shown in Eq. (10):

$$f(D_i^2) = \exp\left[-\frac{D_i^2}{2\sigma^2}\right]. \tag{10}$$

A new vector $\mathbf{X}$ is subtracted from the stored pattern vector when it enters the network. The squares of the difference are summed and input into the activation function in Eq. (10). Those values which pass through the activation function are the pattern unit outputs and are forwarded to the summation units. The summation units proceed to sum the dot product between a weight vector and the pattern unit outputs to generate an estimate as shown in Eq. (11).

$$\sum_{i=1}^{n} \exp\left[-\frac{D_i^2}{2\sigma^2}\right]. \tag{11}$$

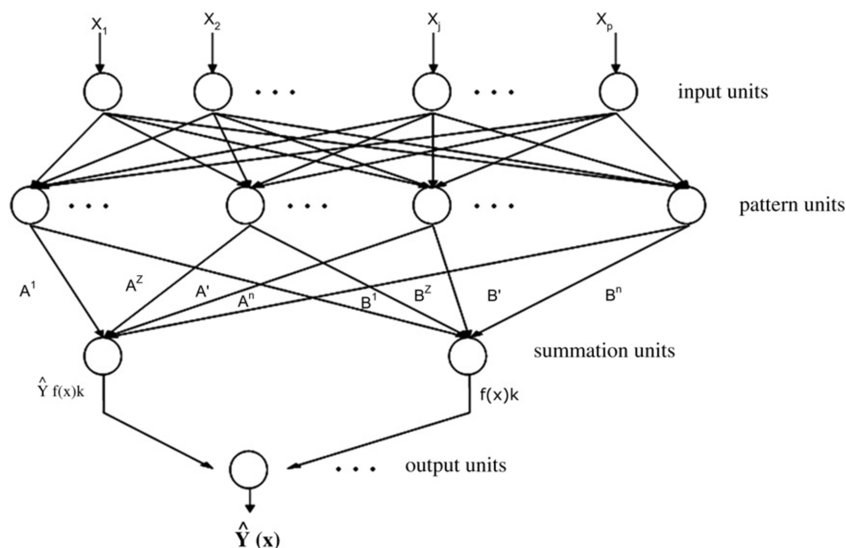

Fig. 1. GRNN block diagram (Specht, 1991).

Conversely, the summation units sum the dot product between the samples Y^i and the pattern unit outputs to generate an estimator as shown in Eq. (12).

$$\sum_{i=1}^{n} y^i \exp\left[-\frac{D_i^2}{2\sigma^2}\right]. \tag{12}$$

Finally, the output unit divides Eq. (12) by Eq. (11) to obtain the desired estimate of y, which is the same as Eq. (9).

GRNN measures how far a given sample pattern is from patterns in the training set. When a new pattern is presented to the network, the input pattern is compared to all of the patterns in the training set to determine how far it is from those patterns. The output that is predicted by the network is a proportional amount of all of the outputs in the training set. The proportion is based upon how far the new pattern is from the given patterns in the training set. GRNN uses an algorithm to find appropriate individual smoothing factors for each input as well as an overall smoothing factor. The algorithm proceeds in two parts. The first part trains the network with the data in the training set. The second part tests a whole range of smoothing factors. The method will produce networks which work much better on the test set.

The performance of neural networks can be measured by a root-mean squared error (RMSE), which can be calculated as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(A_i - O_i)^2}{n}}, \tag{13}$$

where n represents the number of patterns, 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 this study, RMSE and r are applied to evaluate the performance of the negative binomial, BPNN and the proposed GRNN yield model.

4. Proposed approach

4.1. Defect clustering patterns

A major cause affecting yield is the degree to which defects are clustered (Friedman, Hansen, Nair, & James, 1997; Stapper, Armstrong, & Saji, 1983). Hence, the defects clustering phenomenon must be integrated when constructing a yield model. In this study, Borland Delphi programming language is employed to simulate a variety of defect clustering patterns for 8-in. wafers. Fig. 2 presents a simple representation of defect clustering patterns. Three design factors are employed in this study to simulate defect clustering patterns: the cluster pattern; percentage of defects located on grey regions; and, chip size. The following is a brief description of these three design factors.

(1) *Cluster pattern*: Fig. 2 presents one random pattern and four clustering patterns (Friedman et al., 1997). The defects in a random pattern are distributed randomly over the entire wafer. Distribution of defects in the four clustering patterns depends on the percentage of defects located in grey region. Grey region represents the defect-dense areas.
(2) *Percentage of defects located on grey regions*: In the four clustering patterns, four percentages, 60%, 70%, 80% and 90%, of the total number of defects are located in grey regions, and the remaining defects are distributed randomly.

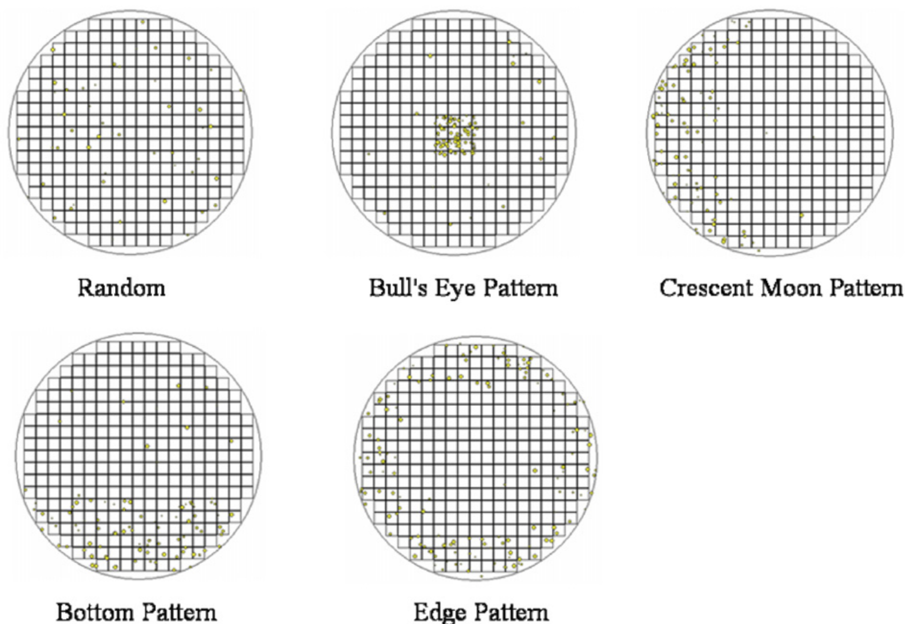

Fig. 2. A simple representation of the defect clustering patterns.

(3) *Chip size*: Six chip sizes are considered: $1(1\times 1)$, $1.44(1.2\times 1.2)$, $1.96(1.4\times 1.4)$, $2.56(1.6\times 1.6)$, $3.24(1.8\times 1.8)$, and $4(2\times 2)$ cm^2.

From these three design factors, a total of 102 simulation trials can be obtained. For each simulation trial, up to 292 defects are obtained randomly. The number of defects obtained randomly creates simulations that are significantly close to real IC manufacturing conditions. The maximum number of chip that can be cut in an 8-in. wafer is 292. Therefore, the number of defects are simulated up to 292 defects to reduce the possible influence of outliers. Each simulation trial represents one defect clustering pattern.

4.2. Procedure of the proposed approach

Models of predicting yield can be classified into macro yield modeling and micro yield modeling. Macro yield modeling uses die size, device density, and other large-scale factors to predict yields for new designs. Micro yield modeling uses critical device area, parametric sensitivity, redundancy effect, and other factors to predict yields (Mullenix, Zalnoski, & Kasten, 1997). Gruber's general yield model (Gruber, 1994) is the most recognized model in Macro yield modeling, and can be described as

$$Y = Y_0(D, A, \theta)L(Y), \tag{14}$$

where Y_0 represents the asymptotic yield, which is a real-valued function of D, A, θ, and; D represents point defect density per unit area, A represents chip area, θ represents a set of parameters unique to the specific yield model, $L(Y)$ represents a real-valued function describing learning effects. In this study, there are numerous attributes that can be obtained from each defect clustering pattern by simple calculation: number of defects; chip size; mean number of defects per chip; mean number of defects per unit area; and, clustering index CI (Jun, Hong, Kim, Park, & Park, 1999). The clustering index CI can be calculated as

$$\mathrm{CI} = \min\left\{\frac{s_v^2}{\bar{v}^2}, \frac{s_w^2}{\bar{w}^2}\right\}, \tag{15}$$

where v_i and w_i are a sequence of defect intervals on the x axis and y axis defined as

$$v_i = x_{(i)} - x_{(i-1)}, \quad i = 1, 2, \ldots, n,$$
$$w_i = y_{(i)} - y_{(i-1)}, \quad i = 1, 2, \ldots, n,$$

where $x_{(i)}$ and $y_{(i)}$ denote the ith smallest defect coordinates on the x axis and y axis, respectively; $\bar{v}$ and s_v^2 represent the sample mean and the sample variance of v_i, respectively; $\bar{w}$ and s_w^2 denote the sample mean and the sample variance of w_i, respectively. 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 defects are clustered. These attributed values are input into the GRNN yield model, whereas the only output of GRNN yield model is the actual wafer yield. The number of replications for each simulation is 10; hence, a total of 1020 pairs of input–output data are obtained to train and test the GRNN yield model.

The Poisson yield model is easy to apply; however, the effect of defect clustering is not considered by the conventional model. This study, proposes a GRNN yield model to predict wafer yield. The proposed approach assumes that wafer yield is affected by each wafer defect. Under this assumption, the proposed approach for the wafer yield prediction in IC manufacturing can be described as follows:

Step 1: Determine the defect clustering pattern. Obtain the simulated defect wafer map. Utilize Borland Delphi programming language to simulate all possible defect clustering patterns for 8-in. wafers.

Step 2: Calculate all attributed values for each pattern. For each defect clustering pattern on a wafer, calculate the following attributed values of patterns: number of defects; chip size; mean number of defects per chip; mean number of defects per unit area; and, clustering index CI.

Step 3: Build a GRNN yield model. Input the attributed values in Step 2 into the GRNN yield model. The actual yield of the wafer is the only output of the GRNN yield model. The percentage of the chip without defects on a wafer is used as the actual yield value of the wafer. In this study, the neural networks package NeuroShell 2 is employed to train and test the GRNN network. A trained GRNN network can be obtained after a few training patterns have been input. Finally, the trained GRNN network produces the network's prediction for each pattern in the test set.

Step 4: Calculate predicted yields. Input the attributed values in Step 2 into the negative binomial yield model to derive the predicted yields of the model. Then build the BPNN yield model as in Step 3 to obtain the predicted yields for the BPNN yield model.

Step 5: Predict and analyze the wafer yield. Utilize the predicted yields obtained by the negative binomial yield model, BPNN yield model and the proposed GRNN yield model to predict the actual yields for the wafer and compare these three yields.

5. Implementation

5.1. A simulation study

This section presents a simulation study to demonstrate the effectiveness of the proposed approach. The data required in this simulation study are obtained by employing the Borland Delphi programming language to simulate a variety of defect clustering patterns for 8-in. wafers. A total of 102 combinations for simulation trials are obtained by combining the three design factors outlined in Section 4.

The number of replications is 10 in each simulation; hence, a total of 1020 pairs of input–output wafer attributed values are obtained and used to train and test the GRNN network. These wafers are divided into two parts: one part contains 816 wafers which are used to train the GRNN network; and, the second part contains 204 wafers which are employed to test the accuracy of the GRNN network.

These attributed values and the actual yields of 1020 wafers are, respectively, utilized as the inputs and output for the proposed GRNN network. The percentage of the chip without defects on a wafer is used as the actual yield value of the wafer. NeuroShell 2 is utilized to train and test the GRNN network. Trained GRNN network are obtained after inputting the 1020 training patterns. Finally, the trained GRNN network is utilized to produce the network's prediction of yields for these 204 test patterns. In this study, the unique parameter of GRNN network, that is, the smoothing factor σ, is set at 0.076, and training is terminated when the error with no improvement of 1%.

Substitute the attributed values of these identical 204 simulated wafers, respectively, into the negative binomial yield model to calculate the predicted yields of the model. Then build a BPNN yield model to obtain the predicted yields for the same wafers. The BPNN network in this study are constructed as three layers (Hornick, Stinchcombe, & White, 1990), with input and output units the same as those in the proposed GRNN network and 17 hidden units (Widrow, Winter, & Baxter, 1987). The learning rate and momentum are 0.6 and 0.9, respective, which are the defaults in NeuroShell 2. The learning epochs are 10,000. To evaluate the performance of these yield models, the relationships between the predicted and actual yield are evaluated. The proposed GRNN yield model (Fig. 3) effectively estimates the actual yield. Table 1 presents the comparisons of RMSE and correlation coefficients r for these three predicted yields and the actual yields. A low RMSE value and high r value indicates that the yield model performs better than the other models. The RMSE of the proposed approach is 0.0914, which is the smallest value of these three RMSEs, and the correlation coefficient is 0.9127, which is the largest value of these three correlation coefficients. These findings reveal that the proposed approach precisely estimates wafer yield, and more accurately predicts yield than the other models.

The influences of each of the following three designed factors on the wafer yield are analyzed: cluster pattern; percentage of defects located on grey regions; and, chip size. The RMSE and correlation coefficients r are applied to analyze the performance of the negative binomial, BPNN and the proposed GRNN yield model. Table 2 shows the RMSE and correlation coefficients of these three yield models for three designed factors. Table 2 reveals that the proposed approach produces the best prediction for wafer yield of the three yield models.

This study varies the simulated wafer sizes from 6 to 12 in., and compares the prediction results for wafer yield from the three yield models. The simulation procedure is the same as that in the previous simulation. Table 3 summarizes the prediction results for three wafer sizes. Table 3 reveals that the proposed approach performs best of these three yield models regardless of wafer size. This simulation obtains the same result as obtained in the 8-in. wafer simulation; that is, the proposed model performs best, and is followed by the BPNN yield model and the negative binomial yield model. The performance of the proposed approach and the BPNN yield model are very close and are better than that of the negative binomial yield model.

Although the performance of the proposed GRNN yield model is slightly better than the BPNN yield model, the proposed GRNN yield model has numerous advantages over the BPNN yield model. The GRNN yield model learns quickly and requires only one parameter.

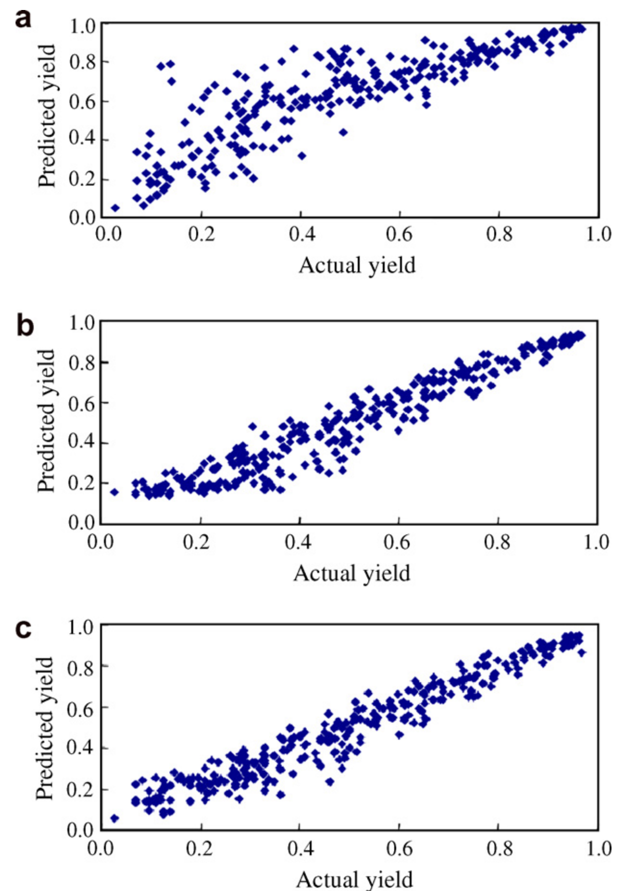

Fig. 3. The relationships between the predicted and actual yields for the negative binomial, BPNN and proposed yield models.

Table 1
The comparisons of RMSE and correlation coefficients between predicted and actual yields

Yield model	RMSE	Correlation coefficient
Negative binomial yield model	0.1203	0.8838
BPNN yield model	0.0960	0.9030
Proposed yield model	0.0914	0.9127

Table 2
The RMSE and correlation coefficients of these three yield models for three design factors

Design factor	Level	Negative binomial		BPNN		Proposed GRNN	
		RMSE	Correlation coefficient	RMSE	Correlation coefficient	RMSE	Correlation coefficient
Cluster pattern	Random	0.0989	0.9465	0.0506	0.9716	0.0214	0.9947
	Bull's eye	0.1225	0.7785	0.0622	0.927	0.0386	0.9732
	Crescent Moon	0.1315	0.8221	0.0884	0.847	0.0848	0.8599
	Bottom	0.1218	0.7833	0.0977	0.7951	0.0977	0.8066
	Edge	0.1463	0.831	0.0722	0.9371	0.0691	0.9466
Percentage	60%	0.1136	0.9239	0.0433	0.9801	0.0464	0.9807
	70%	0.1312	0.8686	0.0511	0.9687	0.0496	0.9701
	80%	0.1287	0.8332	0.0481	0.9511	0.0462	0.9558
	90%	0.1359	0.9009	0.0622	0.9256	0.0577	0.9324
Chip size	1	0.1626	0.9529	0.0671	0.9191	0.0626	0.9255
	1.44	0.098	0.9018	0.1191	0.7168	0.084	0.8766
	1.96	0.0628	0.9503	0.1075	0.8212	0.0839	0.9002
	2.56	0.0819	0.9197	0.1381	0.7185	0.0939	0.8824
	3.24	0.1526	0.8155	0.1118	0.8047	0.0848	0.9039
	4	0.153	0.9078	0.1476	0.7409	0.1128	0.8616

Table 3
The prediction results for three wafer sizes

Wafer size	The actual yield		Yield model	The predicted yield		RMSE	Correlation coefficient
	Average	Std. dev.		Average	Std. dev.		
6-in.	0.6610	0.2070	Negative binomial	0.6350	0.2348	0.1225	0.8597
			BPNN	0.6538	0.1939	0.0854	0.9114
			Proposed GRNN	0.6593	0.1906	0.0836	0.9145
8-in.	0.5662	0.2207	Negative binomial	0.5784	0.2561	0.1203	0.8838
			BPNN	0.5599	0.2142	0.0960	0.9030
			Proposed GRNN	0.5666	0.1852	0.0914	0.9127
12-in.	0.7007	0.1584	Negative binomial	0.7517	0.1772	0.0902	0.9073
			BPNN	0.6955	0.1365	0.0667	0.9085
			Proposed GRNN	0.7045	0.1393	0.0529	0.9448

Furthermore, the proposed GRNN yield model has no overtraining or undertraining problems and the likelihood of obtaining a global optimal solution is higher than that of a BPNN yield model.

6. Conclusion

As wafer size increases, the clustering of defects increases. Under this scenario, the conventional Poisson yield model cannot predict wafer yield. In this study, a proposed neural network-based approach is presented for defect clustering patterns to predict the wafer yield in IC manufacturing. The GRNN network is used to construct the yield model that can accurately predict wafer yield.

The merits of the proposed approach are as follows:

1. The proposed approach utilizes five relevant variables as input variables to predict the wafer yield, rather than utilizing only some of those variables as do the Poisson yield model, compound Poisson yield models and the negative binomial yield model. Therefore, the proposed model is more accurate than both the negative binomial and BPNN yield model.
2. The influences of each of the three designed factors on the wafer yield are analyzed. The RMSE and correlation coefficients of these three yield models for three designed factors reveals that the proposed approach produces the best prediction for wafer yield of the three yield models.
3. This study varies the simulated wafer sizes from 6 to 12 in., and compares the prediction results for wafer yield from the three yield models. This simulation obtains the same result as obtained in the 8-in. wafer simulation regardless of wafer size.
4. The proposed GRNN yield model is fast learning and requires only one parameter to identify for the learning.
5. The proposed approach does not need to construct a complex mathematical yield model and more advanced statistics skill – it only requires a neural network package to predict wafer yield.

References

Albin, S. L., & Friedman, D. J. (1991). Clustered defects in IC fabrication: impact on process control charts. *IEEE Transactions on Semiconductor Manufacturing, 4*(1), 36–42.

Bishop, C. M. (1994). Neural networks and their applications. *Review of Scientific Instrumentation, 65*(6), 1803–1832.

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

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). France.

Fausett, L. (1994). *Fundamentals of neural networks architectures, algorithms, and applications*. Englewood CLiffs, NJ: Prentice Hall.

Ferris-Prabhu, A. V. (1992). *Introduction to semiconductor device yield modeling*. Boston: Artech House.

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.

Gruber, H. (1994). *Learning and strategic product innovation: Theory and evidence for the semiconductor industry*. Amsterdam, Netherlands: Elsevier.

Hong, C., Milor, L., Choi, M., & Lin, T. (2005). Study of area scaling effect on integrated circuit reliability based on yield models. *Microelectronics Reliability, 45*(9–11), 1305–1310.

Hornick, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks, 3*(5), 551–560.

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 interconnect packages with experimental verification. *IEEE Transactions on Electronics Packaging Manufacturing, 28*(4), 344–354.

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). Germany.

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, 9.

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

Mullenix, P., Zalnoski, J., & Kasten, A. J. (1997). Limited yield estimation for visual defect sources. *IEEE Transactions on Semiconductor Manufacturing, 10*(1), 17–23.

Okabe, T., Nagata, M., & Shimada, S. (1972). Analysis of yield of integrated circuits and a new expression of the yield. *Electrical Engineering in Japan, 92*(12), 135–141.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics, 33*(3), 1065–1076.

Raghavachari, M., Srinivasan, A., & Sullo, P. (1997). Poisson mixture yield models for integrated circuits: A critical review. *Microelectronics Reliability, 37*(4), 565–580.

Skinner, K. R., Montgomery, D. C., Runger, G. C., Fowler, J. W., McCarville, D. R., Rhoads, T. R., et al. (2002). Multivariate statistical methods for modeling and analysis of wafer probe test data. *IEEE Transactions on Semiconductor Manufacturing, 15*(4), 523–530.

Specht, D. F. (1991). A general regression neural network. *IEEE Transactions Neural Networks, 2*(6), 568–576.

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

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., Armstrong, F. M., & Saji, K. (1983). Integrated circuit yield statistics. *Proceedings of the IEEE, 71*(4), 453–470.

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.

Tyagi, A., & Bayoumi, A. M. (1992). Defect clustering viewed through generalized Poisson distribution. *IEEE Transactions on Semiconductor Manufacturing, 5*(3), 196–206.

Widrow, B., Winter, R.G., & Baxter, R.A. (1987). Learning phenomena in layered neural networks. In *Proceedings of the First IEEE International Conference on Neural Networks* (pp. 411–429). San Diego.