Expert Systems with Applications 39 (2012) 6665–6671

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

A novel design of wafer yield model for semiconductor using a GMDH polynomial and principal component analysis

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
Defect cluster index
Group method of data handling (GMDH)
Principal component analysis (PCA)

ABSTRACT

According to previous studies, the Poisson model and negative binomial model could not accurately estimate the wafer yield. Numerous mathematical models proposed in past years were very complicated. Furthermore, other neural networks models can not provide a certain equation for managers to use. Thus, a novel design of this paper is to construct a new wafer yield model with a handy polynomial by using group method of data handling (GMDH). In addition to defect cluster index (CI_M), 12 critical electrical test parameters are also considered simultaneously. Because the number of input variables for GMDH is inadvisable to be too many, principal component analysis (PCA) is used to reduce the dimensions of 12 critical electrical test parameters to a manageable few without much loss of information. The proposed approach is validated by a case obtained in a DRAM company in Taiwan.

1. Introduction

For integrated circuits (IC) manufacturers, the wafer yield is a key index to evaluate their profit. Semiconductor manufacturing companies strive to achieve defect-free products and increase profit rate by adopting advanced manufacturing, planning, and evaluating technologies. In these performance technologies (Leachman, 1993), wafer yield prediction is one of the most widely researched approaches in the complicated semiconductor manufacturing system. Wafer yield prediction is very important for a semiconductor manufacturing factory in improving yield, decreasing cost and maintaining a good relationship with customers (Kumar et al., 2006). For this reason, it is an essential task for engineers to manage the wafer yield.

As the wafer size increases, the clustering phenomenon of defects becomes pronounced. Although the Poisson model is the simplest model to use, the essential assumption is that defects must occur independently with constant probability of occurring in small area on a wafer (Albin & Friedman, 1991). 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). Numerous 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 those models (Tong & Chao, 2008; Tong, Lee, & Su, 1997) must set several parameters (e.g., the number of neurons in the hidden layers, the momentum, and the learning rate) and can not provide a certain equation for managers to use. Thus, those neural networks models are often difficult for managers without profound profession knowledge to use in performing wafer yield prediction.

On the basis of practicability, a novel design of this paper is to construct a new wafer yield model with a handy polynomial by using group method of data handling (GMDH) (Ivakhnenko, 1968, 1971). This proposed GMDH model does not need any statistical assumption and can be friendly to use. In addition to defect cluster index (CI_M) (Tong, Wang, & Chen, 2007), 12 critical electrical test parameters are also considered simultaneously. Because the number of input variables for GMDH is inadvisable to be too many, principal component analysis (PCA) (Pearson, 1901) is used to reduce the dimensions of 12 critical electrical test parameters to a manageable few without much loss of information for convenient analysis.

Finally, a case of a DRAM company in Taiwan is utilized to demonstrate the effectiveness of the proposed approach. Comparisons are also made among negative binomial yield model, back-propagation neural network (BPNN) yield model, general regression neural network (GRNN) yield model (Tong & Chao, 2008), and the proposed GMDH yield model to demonstrate that the proposed approach is indeed superior.

2. Literature review

2.1. Yield models

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

E-mail address: junsoon1@hotmail.com

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.09.146

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

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

$$Y = \frac{1}{(1 + D_0 A/\alpha)^{\alpha}} \tag{3}$$

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 et al. (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

The intensity of defects clustered on a wafer can be depicted by a defect cluster index. The cluster parameter (α) of the negative binomial model, the variance/mean ratio (V/M) and the non-parameters assumption cluster index (CI) are commonly used. The negative binomial yield model is as follows:

$$Y = \frac{1}{(1 + \bar{\lambda}/\alpha)^{\alpha}} \tag{5}$$

where α is the cluster parameter and $\bar{\lambda}$ is the mean number of defects per chip. Earlier reports show 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 (Cunningham, 1990).

Tyagi and Bayoumi (1992, 1994) utilized various grid sizes superimposed on a wafer map to measure the intensity 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 values of V/M depend on how the grids are selected and cannot indicate the gradualness of cross-wafer defect density variations.

Jun, Hong, Kim, Park, and Park (1999) proposed a cluster index based on the projected x and y coordinates of defect locations on a wafer. Defect clustering 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 clustered defects. The clustering index CI can be calculated as

$$CI = \min\left\{\frac{S_v^2}{\overline{V}^2}, \frac{S_W^2}{\overline{W}^2}\right\} \tag{6}$$

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

$$V_i = X_{(i)} - X_{(i-1)}, \quad i = 1, 2, \ldots, n \tag{7}$$

$$W_i = Y_{(i)} - Y_{(i-1)}, \quad i = 1, 2, \ldots, n \tag{8}$$

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.

3. Proposed approach

The constructing of the proposed wafer yield model is described in the following subsections.

3.1. Group method of data handling (GMDH)

The GMDH (Ivakhnenko, 1968, 1971) is a special model, and it can be expressed as a set of neurons in which different pairs of them in each layer are connected through a polynomial and, so produce new neurons in the next layer. For instance, the training set is divided into two parts: model learning set E_1 and model selecting set E_2 in GMDH. Let $X = (X_1, X_2, \ldots, X_n)$ and y be the input vector and actual output, respectively. Given M observations of multi-input, single-output data pairs $\{y_i, X_{i1}, X_{i2}, \ldots, X_{in}, i = 1, 2, \ldots, M\}$ in set E_1, I train a GMDH-type neural network to predict the output values $\hat{y}_i$:

$$\hat{y}_i = \hat{f}(X_{i1}, X_{i2}, \ldots, X_{iM}), \quad i = 1, 2, \ldots, M \tag{9}$$

The problem transforms to construct a GMDH-type neural network so that

$$\min \sum_{i=1}^{M} [\hat{f}(X_{i1}, X_{i2}, \ldots, X_{iM}) - y_i]^2 \tag{10}$$

The connection between the inputs and the output variables can be expressed by a complicated discrete form of the Volterra functional series in the form of

$$y = a_0 + \sum_{i=1}^{M} a_i X_i + \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij} X_i X_j + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M} a_{ijk} X_i X_j X_k + \cdots \tag{11}$$

which is also called as the Kolmogorov–Gabor (K–G) polynomial (Madala & Ivakhnenko, 1994; Muller & Lemke, 2000), in particular by the K–G polynomial of degree 2 consisting of only two variables (neurons) in the form of

$$\hat{y} = G(X_i, X_j) = a_0 + a_1 X_i + a_2 X_j + a_3 X_i X_j + a_4 X_i^2 + a_5 X_j^2, i \neq j \tag{12}$$

In this manner, such a partial quadratic description is recursively used in a network of connected neurons to construct the general mathematical relation of the inputs and output variables given in Eq. (11). The coefficients a_i in Eq. (12) are calculated with least squares (LS) (Madala & Ivakhnenko, 1994; Muller & Lemke, 2000). In this manner, the coefficients of each quadratic function G_i are given to optimally fit the output y_i in the whole set E_1, that is

$$\min\left[\frac{\sum_{i=1}^{M}(y_i - G_i)^2}{M}\right] \tag{13}$$

By the GMDH algorithm, all the possibilities of two independent variables out of the total n input variables are taken in order to construct the polynomial in the form of Eq. (12) that best fits the dependent observations (y_i, $i = 1,2,\ldots,M$) with LS. Therefore, $C_n^2 = n(n-1)/2$ neurons will be constructed in the first hidden layer of the feed-forward network from the observations $\{(y_i, X_{ip}, X_{iq})\}$ for different $p,q \in \{1,2,\ldots,n\}$. Likewise, it is now possible to construct M data triples $\{(y_i, X_{ip}, X_{iq})\}$ from observations with such $p,q \in \{1,2,\ldots,n\}$ in the form $\begin{pmatrix} X_{1p} & X_{1q} & y_1 \\ X_{2p} & X_{2q} & y_2 \\ X_{Mp} & X_{Mq} & y_M \end{pmatrix}$.

By the quadratic sub-expression in the form of Eq. (12) for each row of M data triples, the following matrix equation can be given as $A\mathbf{a} = Y$, where $\mathbf{a}$ is the vector of unknown coefficients of the quadratic polynomial in Eq. (12), $\mathbf{a} = \{a_0, a_1, a_2, a_3, a_4, a_5\}^{\mathrm{T}}$ and $Y = \{y_1, y_2, \ldots, y_M\}^{\mathrm{T}}$ is the vector of the output's value from observation. It can be shown in the following

$$\begin{pmatrix} 1 & X_{1p} & X_{1q} & X_{1p}X_{1q} & X_{1p}^2 & X_{1q}^2 \\ 1 & X_{2p} & X_{2q} & X_{2p}X_{2q} & X_{2p}^2 & X_{2q}^2 \\ 1 & X_{Mp} & X_{Mq} & X_{Mp}X_{Mq} & X_{Mp}^2 & X_{Mq}^2 \end{pmatrix}$$

The LS obtains the solution of the equations in the form of

$$a = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}Y \tag{14}$$

which determines the vector of the best coefficients of Eq. (12) for the whole set of M data triples. It should be paid attention to that this procedure is repeated for each neuron of the next hidden layer according to the connectivity topology of the network. In each layer, it uses LS to estimate the parameters of candidate models in set E_1, and uses the external criterion to evaluate and select the candidate models in set E_2. The process continues and should be stopped when we find the optimal model by the termination principle, which is presented by the theory of optimal complexity (Madala & Ivakhnenko, 1994): along with the increase of model complexity, the value of external criterion will decrease first and then increase, and finally the global extreme value agrees with the optimal complexity.

3.2. Principal component analysis (PCA)

Given a set of centered input vectors x_t ($t = 1,2,\ldots,l$ and $\sum_{t=1}^{l} x_t = 0$), each of which is of m dimension $x_t = [x_t(1), x_t(2),\ldots,x_t(m)]^T$ usually $m < l$, PCA (Pearson, 1901) linearly transforms each vector x_t into a new one s_t by

$$s_t = U^{\mathrm{T}}x_t \tag{15}$$

where U is the $m \times m$ orthogonal matrix whose ith column, u_i is the eigenvector of the sample covariance matrix

$$C = \frac{1}{l}\sum_{t=1}^{l} x_t x_t^{\mathrm{T}} \tag{16}$$

In other words, PCA firstly solves the eigenvalue problem

$$\lambda_i u_i = Cu_i, \quad i = 1,2,\ldots,m \tag{17}$$

where λ_i is one of the eigenvalues of C, u_i is the corresponding eigenvector. Based on the estimated u_i, the components of s_t are then calculated as the orthogonal transformations of x_t

$$s_t(i) = u_i^{\mathrm{T}}x_t, \quad i = 1,2,\ldots m \tag{18}$$

The new components are called principal components. By using only the first several eigenvectors sorted in descending order of the eigenvalues, the number of principal components in s_t can be reduced. So PCA has the dimensional reduction characteristic. The principal components of PCA have the following properties: $s_t(i)$ are uncorrelated, has sequentially maximum variances and the mean squared approximation error in the representation of the original inputs by the first several principal components is minimal (Jolliffe, 1986).

3.3. Defect cluster index (CI_M)

In this study, I use the clustering index (CI_M) proposed by Tong et al. (2007) to measure the clustering phenomenon of defects. 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:

$$X_{i,\theta}^* = \cos\theta \times X_i + \sin\theta \times Y_i \tag{19}$$

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

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:

$$V_{i,\theta} = X_{(i,\theta)}^* - X_{(i-1,\theta)}^* \tag{20}$$

where $X_{(0,\theta)}^* = 0$ and $V_{i,\theta}$ 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:

$$SCV_\theta = \frac{S_{v,\theta}^2}{\overline{V}_\theta^2} \tag{21}$$

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

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 (CI_M), as follows:

$$CI_M = \frac{\sum_{\theta=0}^{180} SCV_\theta}{180} \tag{22}$$

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

3.4. Prepare the relative data per wafer

In this study, defect counts, the value of CI_M, and the value of principal component scores are utilized as the input variables for

Table 1
The partial 12 critical electrical test parameters of 111 wafer data in this case.

No.	Parameter x_1	Parameter x_2	Parameter x_3	...	Parameter x_{12}
1	1.1727	0.0658	322.9658	...	2.1098
2	1.1662	0.0615	315.9007	...	2.6533
3	1.1774	0.0682	322.2566	...	2.3412
4	1.1695	0.0632	310.9355	...	2.2145
5	1.1834	0.0595	313.5132	...	2.4589
6	1.1607	0.0659	320.4290	...	2.5098
...	...	...	...	...	...
109	1.1808	0.0614	319.6768	...	2.3325
110	1.1762	0.0604	321.3472	...	2.5977
111	1.1849	0.0632	317.0603	...	2.1104

GMDH. The value of actual wafer yield is the output variable for GMDH. Follows are brief descriptions for the obtainment of CI_M, principal component scores, and the actual wafer yield.

3.4.1. Calculate the value of CI_M

The clustering phenomenon of defects on a wafer influences the accuracy of a wafer yield model, and the CI_M can effectively measure the clustering phenomenon on a wafer. The CI_M can be obtained by the five calculating steps introduced in Section 3.3.

3.4.2. Obtain the value of principal component scores

Use the principal component analysis (PCA) to form new variables that are linear combinations of the original variables (i.e., 12 critical electrical test parameters). Then let the standardized data of original variables into the linear equations of new variables to obtain the value of principal component scores.

3.4.3. 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.

3.5. 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

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(A_i - O_i)^2}{n}} \tag{23}$$

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.

Value number	Eigenvalues of correlation matrix, and related statistics (P Active variables only Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	5.129247	42.74373	5.12925	42.7437
2	4.629295	38.57746	9.75854	81.3212
3	1.110889	9.25741	10.86943	90.5786
4	0.370103	3.08419	11.23953	93.6628
5	0.266142	2.21785	11.50568	95.8806
6	0.151061	1.25884	11.65674	97.1395
7	0.116160	0.96800	11.77290	98.1075
8	0.093370	0.77808	11.86627	98.8856
9	0.057011	0.47509	11.92328	99.3607
10	0.051976	0.43313	11.97525	99.7938
11	0.017018	0.14182	11.99227	99.9356
12	0.007727	0.06439	12.00000	100.0000

Fig. 2. The eigenvalues of PCA.

4. Implementation

In this section, a case of a DRAM company in Taiwan is utilized to demonstrate the effectiveness of the proposed approach. Comparisons are also made among negative binomial yield model, back-propagation neural network (BPNN) yield model, general regression neural network (GRNN) yield model (Tong & Chao, 2008), and the proposed GMDH yield model to demonstrate that the proposed approach is indeed superior.

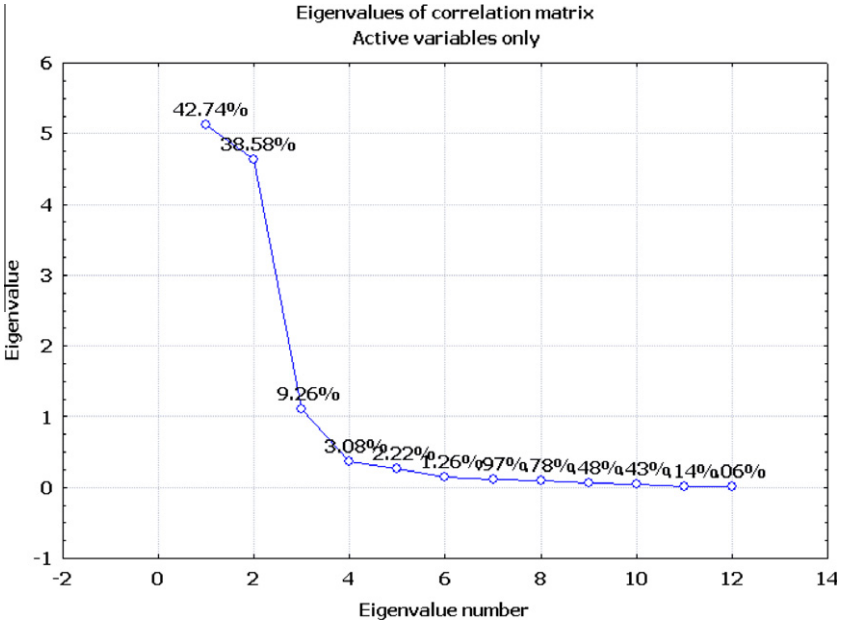

Fig. 1. The scree plot of PCA.

Variable	Eigenvectors of correlation matrix (PCA) Active variables only											
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Fact.10	Fact.11	Fact.12
x_1	-0.026680	0.007736	0.938526	0.127459	-0.157406	-0.123953	-0.133264	0.018572	-0.017007	0.013427	-0.139705	0.154656
x_2	-0.328483	0.282971	0.065546	-0.168876	-0.074959	-0.425082	0.120689	0.502514	0.206744	0.137985	0.330287	-0.393623
x_3	-0.284337	0.301425	-0.178336	-0.480361	-0.226507	-0.262102	-0.362571	-0.139032	-0.059997	0.053464	-0.175944	0.507484
x_4	0.359187	0.237424	-0.019702	0.130635	0.001284	-0.449703	-0.361268	-0.448067	-0.245931	-0.121084	0.214231	-0.377637
x_5	-0.272282	-0.339456	0.078521	-0.231748	0.310367	0.008429	-0.204747	-0.313331	0.083133	0.586170	-0.198445	-0.352669
x_6	0.347641	0.251945	0.043786	-0.107546	-0.087213	0.465810	-0.482921	0.113897	0.502033	0.172310	0.219985	-0.035977
x_7	-0.270017	0.330309	-0.070557	0.332557	0.328266	0.202867	-0.407528	0.349552	-0.328406	-0.091201	-0.354217	-0.153614
x_8	-0.290553	0.309489	0.032273	0.388395	0.422376	-0.007296	0.129798	-0.319868	0.190088	0.177209	0.409165	0.367875
x_9	0.336855	0.272771	-0.009069	-0.050891	0.361627	-0.302759	0.233266	0.001897	0.456505	-0.058673	-0.566241	0.011414
x_{10}	0.319876	-0.266016	0.077999	-0.286445	0.563069	-0.182851	-0.168444	0.362845	-0.242944	0.019753	0.283913	0.297071
x_{11}	0.018506	-0.408228	-0.244049	0.530754	-0.205661	-0.386598	-0.320535	0.208367	0.279295	0.180354	-0.075028	0.194091
x_{12}	-0.336256	-0.277897	0.053399	-0.119851	0.195159	-0.014010	-0.249226	-0.127720	0.380170	-0.716863	0.083764	-0.101691

Fig. 3. The eigenvectors of PCA.

Table 2
The standardized principal component scores of Prin1, Prin2, and Prin3.

No.	Scores of Prin1	Scores of Prin2	Scores of Prin3
1	1.2071	0.6119	0.1203
2	0.6368	1.5014	0.0992
3	1.6939	0.1562	1.3042
...	...	...	...
109	0.8735	1.3814	0.5404
110	0.6523	0.8559	1.4322
111	0.2593	1.6555	1.4796

4.1. PCA of 12 critical electrical test parameters

There are totally 111 data of 8-in. wafer in a case of a DRAM company in Taiwan, and 12 critical electrical test parameters per wafer are considered in this case. The partial 12 critical electrical test parameters of 111 wafer data are listed in Table 1.

The computer software, STATISTICA 6.0, is used to perform PCA. Fig. 1 shows the scree plot of PCA. Fig. 2 shows the eigenvalues of PCA. Fig. 3 shows the eigenvectors of PCA. By Kaiser's rule, we retain only those components whose eigenvalues are greater than 1. Therefore, there are 3 principal components which should be retained. According to Fig. 3, the 3 principal components can be calculated as follows:

$$\text{Prin1} = -0.0267x_1 - 0.3285x_2 - 0.2843x_3 + \cdots - 0.3363x_{12} \quad (24)$$

$$\text{Prin2} = 0.0077x_1 + 0.2830x_2 + 0.3014x_3 + \cdots - 0.2779x_{12} \quad (25)$$

$$\text{Prin3} = 0.9385x_1 + 0.0655x_2 - 0.1783x_3 - \cdots + 0.0534x_{12} \quad (26)$$

The standardized principal component scores of Prin1, Prin2, and Prin3 are partially listed in Table 2.

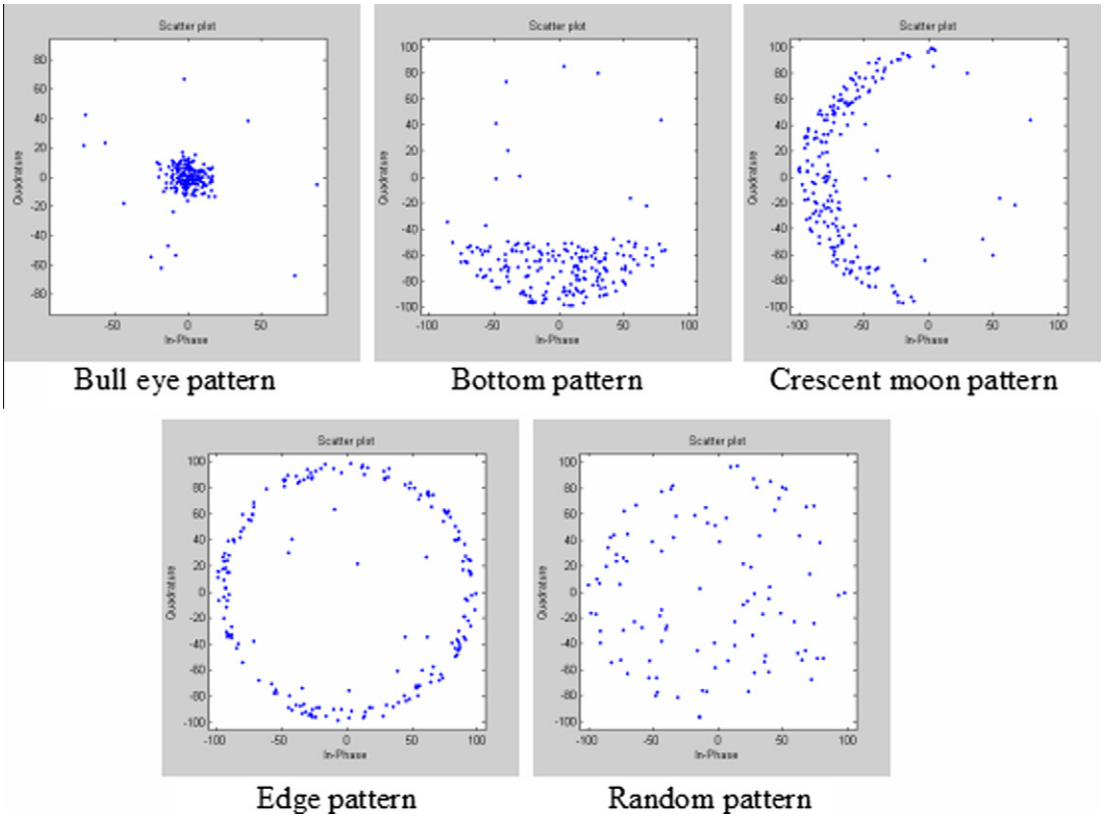

Fig. 4. The five clustering patterns.

4.2. Construct a new wafer yield model using GMDH

In this study, the computer software, NeuroShell 2.0, is used to construct the proposed GMDH yield model. In this case of a DRAM company in Taiwan, 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, and these five clustering patterns are shown in Fig. 4.

Eighty-nine wafer data are randomly selected as training samples, and the rest 22 wafer data are the testing samples. The result of GMDH learning is shown in Fig. 5. By Fig. 5, the GMDH polynomial of proposed yield model is shown in Eq. (27)

$$\begin{aligned} Y = {} & 8.8E-002 * X5 - 0.15 * X2 + X3 - 0.11 * X4 - 3.8E \\ & - 002 - 7.3E - 002 * X1 - 1.1 * X3\text{^}2 - 6.9 * X4\text{^}2 \\ & - 0.61 * X3\text{^}3 + 1.8 * X4\text{^}3 + 8.5 * X3 * X4 - 1.3 * X3 \\ & * X5 + 7 * X4 * X5 + 7.9 * X3 * X4 * X5 + 0.1 * X2\text{^}2 \\ & + 1.1 * X2 * X3 + 0.28 * X2 * X5 - 1.4 * X2 * X3\text{^}2 - 6.5 \\ & * X2 * X4\text{^}2 - 0.77 * X2 * X3\text{^}3 + 1.7 * X2 * X4\text{^}3 + 8.1 \\ & * X2 * X3 * X4 - 1.3 * X2 * X3 * X5 + 6.6 * X2 * X4 * X5 \\ & + 7.5 * X2 * X3 * X4 * X5 - 0.26 * X1\text{^}2 + 0.12 * X1\text{^}3 \\ & + 0.19 * X5\text{^}2 + 0.15 * X2\text{^}3 + 0.21 * X5\text{^}3 \end{aligned} \quad (27)$$

where Y denotes the predictive wafer yield value, $X1$ is defect counts, $X2$ is the value of CI_M, $X3$ is scores of Prin1, $X4$ is scores of Prin2, and $X5$ is scores of Prin3. The value of RMSE $= \sqrt{\text{MSE}} = \sqrt{0.010540} = 0.1027$, and the value of correlation coefficient is 0.9784.

4.3. Compare with other wafer yield models

Finally, the comparisons made among negative binomial yield model, back-propagation neural network (BPNN) yield model, general regression neural network (GRNN) yield model (Tong & Chao, 2008), and the proposed GMDH yield model are listed in Table 3. The scatter plots in the negative binomial yield model, BPNN yield model, GRNN yield model, and the proposed GMDH yield model are shown from Figs. 6–9.

From Table 3, it can be seen that the proposed GMDH 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.

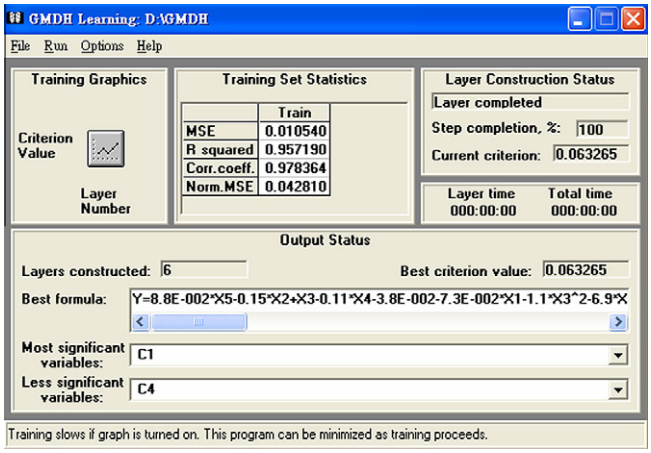

Fig. 5. The result of GMDH learning.

Table 3
Comparisons of RMSE and r between predictive and actual yield value.

Yield model	RMSE	r
Negative binomial yield model	0.1443	0.9159
BPNN yield model	0.1224	0.9308
GRNN yield model	0.1189	0.9496
Proposed GMDH yield model	0.1027	0.9784

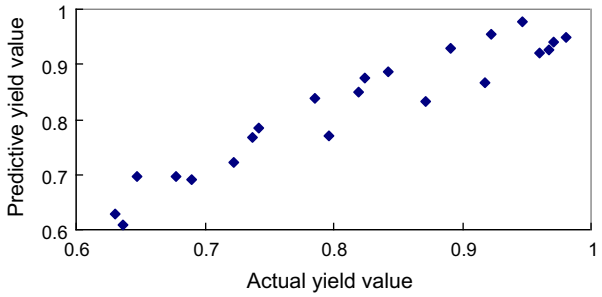

Fig. 6. The scatter plot in negative binomial yield model.

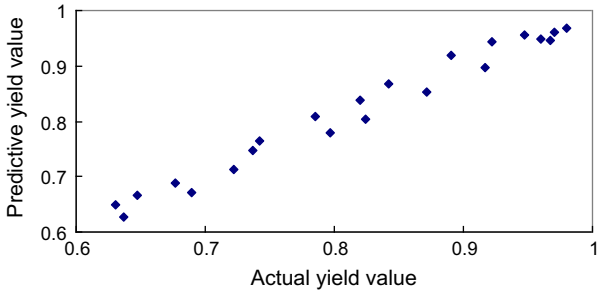

Fig. 7. The scatter plot in BPNN yield model.

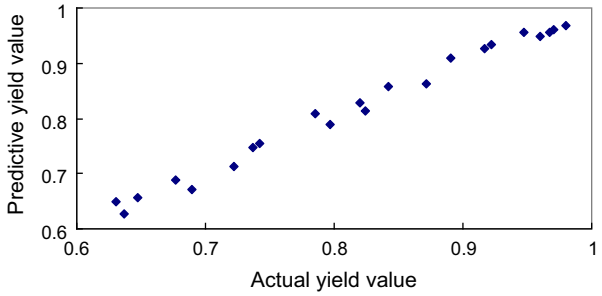

Fig. 8. The scatter plot in GRNN yield model.

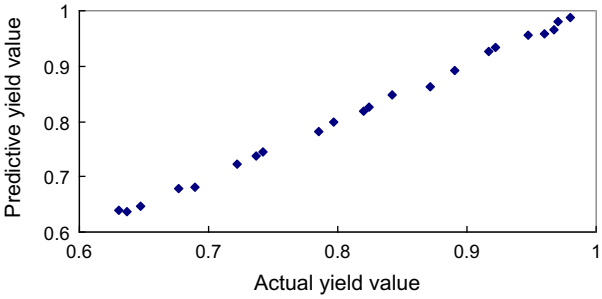

Fig. 9. The scatter plot in GMDH yield model.

5. Conclusions

When the clustering phenomenon of defects is pronounced, the conventional Poisson yield model can not reasonably estimate the wafer yield. Other neural networks models have the problem of setting parameters and can not provide a certain equation for managers to use.

On the basis of practicability, a novel design of this paper is to construct a new wafer yield model with a handy polynomial by using GMDH, and it can accurately predict the wafer yield. In addition to defect cluster index (CI_M), 12 critical electrical test parameters are also considered simultaneously. In this study, the PCA is used to reduce the dimensions of 12 critical electrical test parameters to a manageable few without much loss of information for convenient analysis.

The merits of the proposed approach are summarized as follows:

(1) The proposed GMDH yield model can provide a handy polynomial for managers to use, and this model does not require setting parameters of neural networks.
(2) This study employs PCA to reduce the dimensions of 12 critical electrical test parameters to a manageable few without much loss of information, and it can effectively simplify the constructions of variables.
(3) The proposed GMDH yield model is fast learning and has high accuracy of prediction.
(4) The proposed GMDH yield model does not need any statistical assumption and can be friendly to use.
(5) The proposed GMDH yield model can help the IC manufacturers to manage the wafer yield and evaluate their process capability in relation to profit and loss.

Acknowledgement

The author thanks the National Chiao Tung University for its resourceful support.

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.

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).

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.

Ivakhnenko, A. G. (1968). The group method of data handling; a rival of the method of stochastic approximation. *Soviet Automatic Control, 13*(3), 43–55.

Ivakhnenko, A. G. (1971). Polynomial theory of complex systems. *IEEE Transactions on Systems, Man and Cybernetics, 1*(4), 364–378.

Jolliffe, I. J. (1986). *Principal component analysis*. New York: Springer.

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).

Leachman, R. C. (1993). The competitive semiconductor manufacturing survey. In *IEEE international symposium on semiconductor manufacturing conference, Austin, Texas, USA* (pp. 359–381). Piscataway, NJ: IEEE.

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.

Madala, H. R., & Ivakhnenko, A. G. (1994). *Inductive learning algorithms for complex systems modeling*. Boca Raton, FL: CRC Press.

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

Muller, J. A., & Lemke, F. (2000). *self-organising data mining: An Intelligent approach to extract knowledge from data*. Hamburg: Libri.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2*, 559–572.

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

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

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., Wang, C. H., & Chen, D. L. (2007). Development of a new cluster index for wafer defects. *International Journal of Advanced Manufacturing Technology, 31*, 705–715.

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

Tyagi, A., & Bayoumi, M. A. (1994). The nature of defect patterns on integrated circuit wafer maps. *IEEE Transactions on Reliability, 43*(1), 22–29.