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

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 (CIM), 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 (z), but the value of z 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 (CIM) (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
The number of defects per chip. The Poisson yield model can be obtained as
\[ Y = P(k=0) = e^{-\bar{x}} \] (2)

Cunningham (1990) indicated that, when the chip size is less than 0.25 cm², 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
\[ Y = \frac{1}{(1 + D_0 A/\bar{x})^a} \] (3)

where \( D_0 \) is the average number of defects per unit area, \( A \) is the chip area, and \( \bar{x} \) is the cluster parameter. The value of \( \bar{x} \) can be quite scattered and sometimes negative when the negative binomial yield model is used to predict yield.


2.2. Defect cluster index

The intensity of defects clustered on a wafer can be depicted by a defect cluster index. The cluster parameter \( \bar{x} \) of the negative binomial model, the variance/mean ratio \( \bar{V}/\bar{M} \) and the non-parameters assumption cluster index \( CI \) are commonly used. The negative binomial yield model is as follows:
\[ Y = \frac{1}{(1 + \bar{x}/\bar{X})^a} \] (5)

where \( \bar{x} \) is the cluster parameter and \( \bar{x} \) is the mean number of defects per chip. Earlier reports show that cluster parameter \( \bar{x} \) 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 granularity 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 \( y \) 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
\[ CI = \min \left\{ \frac{S_x^2}{S_y^2}, \frac{S_y^2}{S_x^2} \right\} \] (6)

where \( S_x \) and \( S_y \) are a sequence of defect intervals on the \( x \)-axis and \( y \)-axis defined as
\[ V_i = X_{(i)} - X_{(i-1)}, \quad i = 1, 2, \ldots, n \] (7)
\[ W_i = Y_{(i)} - Y_{(i-1)}, \quad i = 1, 2, \ldots, n \] (8)

where \( X_{(i)} \) and \( Y_{(i)} \) denote the \( i \)th 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 inputs \( \bar{x}_i \), \( \bar{y}_i \) and mean \( \bar{X}_i \), \( \bar{Y}_i \) are equal in GMDH. Let \( X = (X_1, X_2, \ldots, X_M) \) and \( y \) be the input vector and actual output, respectively. Given M observations of multi-input, single-output data pairs \( \{y_i, X_{i1}, X_{i2}, \ldots, X_{im}\}, \quad i = 1, 2, \ldots, M \) in set \( E_1 \), I train a GMDH-type neural network to predict the output values \( \hat{y}_i \):
\[ \hat{y}_i = \bar{f}(X_{i1}, X_{i2}, \ldots, X_{im}), \quad i = 1, 2, \ldots, M \] (9)

The problem transforms to construct a GMDH-type neural network so that
\[ \min \sum_{i=1}^{M} \left| \bar{f}(X_{i1}, X_{i2}, \ldots, X_{im}) - y_i \right|^2 \] (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
\[ y = a_0 + \sum_{i=1}^{M} a_i X_i + \sum_{i=1}^{M} \sum_{j=1}^{M} a_{ij} X_i X_j + \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{k} a_{ijk} X_i X_j X_k + \cdots \] (11)

which is also called 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
\[ \hat{y} = G(X_i, X_j) = a_0 + a_1 X_i + a_2 X_j + a_3 X_i X_j + a_4 X_i^2 + a_5 X_j^2, \quad i \neq j \] (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 \( E_i \), that is

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

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, \( c_i^2 - n(n - 1)/2 \) 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

\[
\begin{pmatrix}
X_{ip} & X_{iq} & y_i \\
X_{ip} & X_{iq} & y_2 \\
\vdots & \vdots & \vdots \\
X_{ip} & X_{iq} & y_M
\end{pmatrix}
\]

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

\[
Aa = Y, \quad \text{where } a = \{a_0, a_1, a_2, a_3, a_4\}^T \text{ and } Y = [y_1, y_2, \ldots, y_M]^T \text{ is the vector of the output's value from observation. It can be shown in the following}
\]

\[
\begin{pmatrix}
1 & X_{ip} & X_{ip}X_{iq} & X_{ip}^2 & X_{iq}^2 \\
1 & X_{ip} & X_{ip}X_{iq} & X_{ip}^2 & X_{iq}^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & X_{ip} & X_{ip}X_{iq} & X_{ip}^2 & X_{iq}^2
\end{pmatrix}
\]

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

\[
a = (A^TA)^{-1}A^TY \tag{14}
\]

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 lower 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 \( x_t(1), x_t(2), \ldots, x_t(m) \) usually \( m < l \), PCA (Pearson, 1901) linearly transforms each vector \( x_t \) into a new one \( s_t \) by

\[
s_t = U^T x_t \tag{15}
\]

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

\[
C = \frac{1}{l} \sum_{t=1}^{l} x_t x_t^T 
\]

In other words, PCA firstly solves the eigenvalue problem

\[
\lambda u_i = C u_i, \quad i = 1, 2, \ldots, m 
\]

where \( \lambda_i \) 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 \( x_t \)

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

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: \( s_t(i) \) is 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 (CIM)

In this study, I use the clustering index (CIM) proposed by Tong et al. (2007) to measure the clustering phenomenon of defects. The detailed descriptions of obtaining CIM 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 \( \theta \). Suppose that a wafer has \( n \) defects, and \((X_i, Y_i)\) denotes the \( x \) and \( y \) coordinates of the \( i \)-th defect location in a two-dimensional space, \( i = 1, \ldots, n \). These \( n \) defects can then be projected onto a new axis \( X_{i}', Y_{i}' \) obtained by rotating the \( x \)-axis counterclockwise using \( \theta \). The new coordinates for the \( i \)-th defect with respect to \( \theta \) then can be calculated as follows:

\[
X_i' = \cos \theta \times X_i + \sin \theta \times Y_i 
\]

where \( i \) denotes the \( i \)-th defect and \( \theta \) represents a rotating angle, where \( 0 \leq \theta \leq 180 \).

Step 2: Sort the \( X_i' \) values in ascending order and calculate the intervals between each adjacent coordinate value \( X_i' \), then can be calculated as follows:

\[
V_i = X_i' - X_{i-1}' 
\]

where \( X_{i-1}' = 0 \) and \( V_i \) represents the \( i \)-th interval between \( X_i' \) and \( X_{i-1}' \).

Step 3: Calculate the squared coefficient of variation (SCV) for \( V_i \). The SCV for \( V_i \) can be determined as follows:

\[
SCV_i = \frac{S_i^2}{\bar{V}_i} \tag{21}
\]

where \( SCV_i \) represents the squared coefficient of variation for \( V_i \), \( \bar{V}_i = \frac{\sum_{i=1}^{n} V_i}{n} \) and \( S_i^2 = \frac{\sum_{i=1}^{n} (V_i - \bar{V}_i)^2}{(n - 1)} \).

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

Step 5: According to the \( SCV_i \) values obtained from Step 4, the average \( SCV_i \) value determines the clustering index (CIM), as follows:

\[
CIM = \frac{\sum_{i=1}^{180} SCV_i}{180} \tag{22}
\]

where \( CIM \) represents defect cluster index. A larger \( CIM \) 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 \( CIM \), and the value of principal component scores are utilized as the input variables for
GMDH. The value of actual wafer yield is the output variable for GMDH. Follows are brief descriptions for the obtainment of \( C_{\text{IM}} \), principal component scores, and the actual wafer yield.

3.4.1. Calculate the value of \( C_{\text{IM}} \)

The clustering phenomenon of defects on a wafer influences the accuracy of a wafer yield model, and the \( C_{\text{IM}} \) can effectively measure the clustering phenomenon on a wafer. The \( C_{\text{IM}} \) 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

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(A_i - O_i)^2}{n}}
\]

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

4. Implementation

In this 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.
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:

\[
\text{Prin}_1 = -0.0267x_1 - 0.3285x_2 - 0.2843x_3 + \cdots - 0.3363x_{12}
\]

\[
\text{Prin}_2 = 0.0077x_1 + 0.2830x_2 + 0.3014x_3 + \cdots - 0.2779x_{12}
\]

\[
\text{Prin}_3 = 0.9385x_1 + 0.0655x_2 - 0.1783x_3 + \cdots + 0.0534x_{12}
\]

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

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

<table>
<thead>
<tr>
<th>No.</th>
<th>Scores of Prin1</th>
<th>Scores of Prin2</th>
<th>Scores of Prin3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2071</td>
<td>0.6119</td>
<td>0.1203</td>
</tr>
<tr>
<td>2</td>
<td>0.6369</td>
<td>1.5014</td>
<td>0.0992</td>
</tr>
<tr>
<td>3</td>
<td>1.0039</td>
<td>0.1562</td>
<td>1.3042</td>
</tr>
<tr>
<td>109</td>
<td>0.8735</td>
<td>1.3814</td>
<td>0.5404</td>
</tr>
<tr>
<td>110</td>
<td>0.8523</td>
<td>0.8559</td>
<td>1.4322</td>
</tr>
<tr>
<td>111</td>
<td>0.2593</td>
<td>1.6555</td>
<td>1.4796</td>
</tr>
</tbody>
</table>

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.
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)

\[
Y = 8.8E - 002 * X5 - 0.15 * X2 + X3 - 0.11 * X4 - 3.8E
- 002 - 7.3E - 002 * X1 - 1.1 * X3'2 - 6.9 * X4'2
- 0.61 * X3'3 + 1.8 * X4'3 + 8.5 * X3 * X4 - 1.3 * X3
* X5 + 7 * X4 + X5 + 7.9 * X3 * X4 + X5 + 0.1 * X2'2
+ 1.1 * X2 * X3 + 0.28 * X2 * X5 - 1.4 + X2 * X3'2 - 6.5
* X2 * X4'2 - 0.77 * X2 * X3'3 + 1.7 * X2 * X4'3 + 8.1
* X2 * X3 + X4 - 1.3 * X2 * X3 + X5 + 6.6 * X2 * X4 * X5
+ 7.5 * X2 * X3 + X4 + X5 - 0.26 + X1'2 + 0.12 * X1'3
+ 0.19 * X5'2 + 0.15 * X2'3 + 0.21 * X5'3
\]

(27)

where Y denotes the predictive wafer yield value, X1 is defect counts, X2 is the value of Clsh, X3 is scores of Prin1, X4 is scores of Prin2, and X5 is scores of Prin3. The value of RMSE = \(\sqrt{MSE = \sqrt{0.010540 = 0.1027}}\), 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.

<table>
<thead>
<tr>
<th>Yield model</th>
<th>RMSE</th>
<th>(r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative binomial yield model</td>
<td>0.1443</td>
<td>0.9159</td>
</tr>
<tr>
<td>BPNN yield model</td>
<td>0.1224</td>
<td>0.9308</td>
</tr>
<tr>
<td>GRNN yield model</td>
<td>0.1189</td>
<td>0.9496</td>
</tr>
<tr>
<td>Proposed GMDH yield model</td>
<td>0.1027</td>
<td>0.9784</td>
</tr>
</tbody>
</table>

**Fig. 5.** The result of GMDH learning.

**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 \((C_{\text{MDH}})\), 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 constrictions 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


