ELSEVIER

Computers in Industry 34 (1997) 285–294

COMPUTERS IN INDUSTRY

A neural network-based procedure for the process monitoring of clustered defects in integrated circuit fabrication

Chao-Ton Su *, Lee-Ing Tong

Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan

Received 25 June 1996; accepted 7 March 1997

Abstract

In integrated circuit (IC) fabrication, a wafer's defects tend to cluster. As the wafer size increases, the clustering phenomenon of the defects becomes increasingly apparent. When the conventional control chart (c chart) is used, the clustered defects frequently cause many false alarms. In this study, we propose a neural network-based procedure for the process monitoring of clustered defects in IC fabrication. The proposed procedure can reduce the phenomenon of the false alarms caused by the clustered defects. A case study is also presented to show the effectiveness of the proposed procedure.
© 1997 Elsevier Science B.V.

Keywords: Neural networks; Clustering analysis; c chart; Defect; Cluster

1. Introduction

An integrated circuit (IC) design, even when error-free, is subject to defects during manufacturing and could ultimately produce a faulty chip. Global defects damaging a large portion of the wafer are frequently caused by mishandling and can be easily detected. However, small defects such as minute spots of extra material or missing material are often difficult to detect. This paper examines small defects having dimensions comparable to the size of transistors. An increasing defect number implies a decreasing IC yield. Therefore, the defect number must be decreased to ensure that the more complex IC is still functioning. Consequently, defect counts should be carefully monitored during the IC fabrication process.

To monitor the defect of IC products, most manufacturers use control charts for their process control. Control charts (c charts) are simple graphs based on basic statistical theories to distinguish variations in a process due to common causes with variations due to special causes. During IC fabrication, the c chart for the number of defects per sample is used to monitor the manufacturing process. The c chart is constructed under the assumption that the number of defects in a sample follows the Poisson distribution. The Poisson assumption implies that (i) occurrence of a defect in any location is independent of the occurrence of defects in other locations and (ii) for all locations in the sample, the likelihood of a defect occurring is the same [1]. If the Poisson assumption holds, defects are uniformly scattered over a sample.

Stapper [2] reported that a wafer's defects tend to

* Corresponding author. Tel.: +886-3-5731857, Fax: +886-3-5722392, E-mail: ctsu@cc.nctu.edu.tw.

0166-3615/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved.
PII S0166-3615(97)00050-X

cluster. This clustering phenomena becomes more evident as the wafer size increases. Consequently, a defect in any location is no longer independent of defects in other locations, i.e., the Poisson assumption may be violated. Therefore, applying the Poisson-based c chart may be inappropriate when defects tend to cluster. Albin and Friedman [1] indicated that the c chart frequently produces a false alarm. That is, sample measurements falling out of the upper control limit appear to be a process out-of-control; however, the process is actually in control and the IC product's yield still attains its normal value. Albin and Friedman [1,3] recommended the use of the Neyman type-A distribution for defect data. The control limits derived from the Neyman type-A distribution are wider than the control limits calculated with the Poisson distribution. Consequently, applying the Neyman-based c chart may reduce the false alarms for a process that yields clustered defects. However, the Neyman-based c chart cannot detect the variation of the clustered defects within a wafer. Furthermore, despite the increasing size of wafers used today, the Neyman type-A model can only be applied to small wafers, thereby limiting the use of the c chart.

Defect clustering frequently causes a yield prediction problem. When clustering occurs, defects are no longer uniformly scattered over a wafer. Under this circumstance, a large number of defects may not necessarily lower the product's yield and the conventional c chart for monitoring defects may produce too many false alarms. This paper presents a novel approach to enhance the analysis of defect data. Neural networks can be used to group observations into clusters at a high computational rate. This capability enables their applications in manufacturing. Although many efficient statistical clustering algorithms are available, applying neural networks to clustering problems has recently attracted much attention [4]. Neural networks pose advantages over classical statistical techniques since the former utilize the parallel architecture of a neural network. In this study, we propose a neural network-based procedure to monitor the process of clustered defects in IC fabrication. The proposed procedure can reduce the phenomenon of the false alarms caused by the clustered defects. A case study shows the proposed procedure's effectiveness.

2. Background information

2.1. Construction of the c chart

The conventional c chart is constructed on the basis of the assumption that the number of defects in a sample follows the Poisson distribution. The Poisson distribution is defined as follows:

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

where X represents the number of defects in a sample and λ represents the mean number of defects per sample. Notably, the mean and variance of the Poisson distribution are the same value. Given λ, the three sigma limits for the c chart can be constructed by

$$\begin{aligned}\mathrm{UCL}&=\lambda+3\sqrt{\lambda}\\ \mathrm{LCL}&=\lambda-3\sqrt{\lambda}\end{aligned}, \tag{2}$$

where UCL is the upper control limit and LCL is the lower control limit. These control limits are for a c chart such that the probability of a sample point is outside the limits when the process that is in-control is equal to 0.0027. Restated, an incorrect out-of-control signal (or false alarm) is generated in only 27 out of 10,000 points.

2.2. Neyman-based control chart

When the clustering phenomenon exists, the assumption for Poisson-based c chart is not satisfied. The consequence of incorrectly using a c chart leads to a situation in which several out-of-control sample points would appear. Albin and Friedman [1] listed two possible causes for those out-of-control points: (i) the process is indeed out of control or (ii) the control limits are incorrectly calculated because they are based on a Poisson assumption. Restated, the process yields clustered defects. In their study, they also proposed a procedure to distinguish the two types of causes. Their procedure entailed removing the outliers from the defective data and then testing the hypothesis that the remaining data follow the Poisson distribution. If the data follow the Poisson distribution, the Poisson-based c chart is used. Otherwise, a c chart based on the Neyman-type A distribution is used.

The Neyman distribution is a member of the family of compound Poisson distributions. The underlying assumptions for Neyman-type A distribution are: (i) the number of defect clusters is Poisson-distributed with a mean λ, and (ii) the number of defects within each cluster is Poisson-distributed with a mean ϕ. Let X be the number of defects in a sample. The Neyman distribution for X is expressed as:

$$P(X=k) = \sum_{j=1} \frac{e^{-\lambda}\lambda^j}{j!} \cdot \frac{e^{-j\phi} j\phi^k}{k!} \quad k=1,2,\ldots \tag{3}$$

The parameters λ and ϕ can be estimated from a set of independent and identically Neyman-distributed data, $x_1, \ldots, x_n$, as

$$\hat{\lambda} = \frac{(\bar{X})^2}{S^2 - \bar{X}} \text{ and } \hat{\phi} = \frac{S^2 - \bar{X}}{\bar{X}}, \tag{4}$$

where $\bar{X}$ and S^2 are the sample mean and sample variance, respectively. With the estimated parameters of λ and ϕ, the control limits for a c chart on the Neyman distribution can be computed by selecting UCL and LCL to satisfy

$$\sum_{k=0}^{\mathrm{UCL}} P(X=k) = 1 - \frac{\alpha}{2} \text{ and } \sum_{k=0}^{\mathrm{LCL}} P(X=k) = \frac{\alpha}{2}, \tag{5}$$

where $P(X=k)$ is given in Eq. (3) and α is the significance level.

For the Neyman distribution, the mean of defects is $\lambda\phi$ and the variance of defects is $\lambda\phi(1+\phi)$. The variance-to-mean ratio in the Neyman distribution is $1+\phi$, in contrast to that in the Poisson distribution, which is 1. Therefore, the control limits from the Neyman distribution are $1+\phi$ times as wide as the Poisson-based control limits. Therefore, a Neyman-based c chart can reduce some of the false alarms.

However, the Neyman-based c chart has the following limitations.

(i) Neyman-based c chart can only monitor the variation of defects between wafers. It cannot detect the variation of defects within a wafer. For instance, two wafers have the same number of defects but different defect distributions; one is random and the other is clustered. The yield of the clustered defect wafers might be higher than that of the other wafers.

(ii) Stapper [2] reported that the Neyman model has a lower limit, $e^{-\lambda}$, thereby limiting its use only to small wafers.

Increasing the wafer's size and complexity cause the number of defects to increase and the defect clustering phenomena to become more pronounced. Therefore, a Neyman-based model may be inappropriate in some cases.

2.3. Test for randomness

Two methods are commonly used to analyze the distribution of sample points on a surface: quadrat method and distance method [5]. The quadrat method divides the surface area into random or contiguous quadrats of the same size. Since selecting random quadrats affects the judgment of the distribution of points on a surface, contiguous quadrats are used herein for analyzing the samples. By using the points in a quadrat as samples, the mean and variance can be calculated for all the samples. Consequently, a t-test statistic [6] developed by Greig-Smith can be obtained as follows:

$$t = \frac{V/M - 1}{\sqrt{2/(n-1)}}, \tag{6}$$

where M and V are the mean and variance for n selected samples, respectively. The t-test statistic follows a t-distribution with $n-1$ degrees of freedom. If a t value is greater than the right-side critical value, $t_{\alpha,n-1}$, then the distribution of points on surface is not random, i.e., the points tend to cluster.

3. Neural networks

Neural networks are composed of processing elements and connections. Each processing element (node) has an output signal that fans out along connections to each of the other processing elements. The connections are characterized by their weights. A node's output depends on the threshold specified and the transfer function. Common configurations of neural networks are fully interconnected. The two types of learning are supervised and unsupervised learning. For supervised learning, a set of training

input vectors with a corresponding set of target vectors is trained to adjust the weights in a neural net. For unsupervised learning, a set of input vectors is proposed; however, no target vectors are specified.

Our approach towards the clustering problem is based on unsupervised neural networks. Several well-known unsupervised learning neural network models are competitive learning, self-organizing maps and adaptive resonance theory (ART). Those models are applied primarily in the area where patterns are grouped into clusters. The major difference between ART and other unsupervised neural networks is the former's so-called vigilance parameter (ρ). In an ART, a degree of similarity between a new pattern and the stored pattern is defined. This similarity, which is compared to ρ, is a measure to ensure whether the new pattern is properly classified or not. The other unsupervised learning neural networks, which do not implement vigilance, may cause a significantly different input pattern to be forced into an inappropriate cluster. Accordingly, ART is used in this study for the clustering problem.

ART has several variations. The fuzzy ART, which incorporates fuzzy computations into the ART-based neural network, is proposed in this work for the clustering problem. Fuzzy ART can be used for both binary and continuous valued input patterns, and has a relative simpler architecture than other variations of ART. Fig. 1 shows a simple representation of the structure for fuzzy ART [7]. The outline of the fuzzy ART is briefly described as follows [8].

Fuzzy ART consists of two layers: the input (F1) and the output (F2) layers. Each input $\boldsymbol{I}$ is presented by an M-dimensional vector $\boldsymbol{I} = (I_1, I_2, \dots, I_M)$, where each component I_i is in the interval [0,1]. Each output category (output node) j is represented by one set of weight vectors $W_j = (w_{j1}, w_{j2}, \dots, w_{jM})$. Initially, $w_{j1} = w_{j2} = \cdots w_{jM} = 1$, for all j.

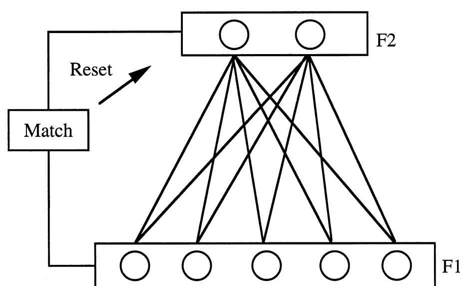

Fig. 1. A simple representation of fuzzy ART [7].

To categorize input patterns, the output nodes receive net input in the form of the following choice function:

$$T_j = \frac{|\boldsymbol{I} \wedge \boldsymbol{W}_j|}{\alpha_0 + |\boldsymbol{W}_j|}, \tag{7}$$

where $\wedge$ is the fuzzy MIN operator defined as:

$$(X \wedge Y)_i \equiv \min(x_i, y_i), \tag{8}$$

and the norm $|\cdot|$ is defined by

$$|X| \equiv \sum_{i=1}^{M} |x_i|. \tag{9}$$

The output node, θ, with the highest value of T_j is selected to claim the current input pattern. That is,

$$T_\theta = \max\{T_j\} \tag{10}$$

For node θ to cluster the pattern, the match function should exceed the vigilance parameter, i.e.,

$$\frac{|I \wedge W_\theta|}{|I|} \geq \rho \tag{11}$$

If node θ does not pass the similarity test, T_θ is set to -1 to prevent the system from choosing the same category again. The weight vector of the winning node is updated as follows:

$$\boldsymbol{W}_\theta^{\text{new}} = \beta\left(\boldsymbol{I} \wedge \boldsymbol{W}_\theta^{\text{old}}\right) + (1-\beta)\boldsymbol{W}_\theta^{\text{old}} \tag{12}$$

Fuzzy ART has three parameters: (i) the choice parameter $\alpha_0 > 0$, (ii) the learning rate $\beta \in [0,1]$, and (iii) the vigilance parameter $\rho \in [0,1]$. Depending on the characteristic of the problem to be categorized, these parameters are specified by the users.

4. Proposed procedure

Defect clustering invalidates the assumptions of the Poisson-based c chart. Therefore, defect clustering must be analyzed when the c chart is constructed. In this study, we use fuzzy ART to find the number of defect clusterings. Next, the location of the defect clustering's center is determined. Accordingly, the total number of defects on a wafer is adjusted. If the adjusted data adhere to the assump-

tion of the Poisson distribution, the c chart is constructed (by using the adjusted data) for process control. The neural network-based procedure for process monitoring of clustered defects in IC fabrication is described in the following.

Step 1: Obtain the wafer map by using the wafer inspection system (such as KLA 2110).

Step 2: Identify the possible outliers and find the causes of these possible outliers. Discard the outliers.

Step 3: Use the Greig-Smith's t-statistic to test whether wafer's defect clustering exists. If a significant clustering situation arises, go to Step 4; otherwise, go to Step 7.

Step 4: Specify the vigilance parameter ρ and use the fuzzy ART to find the number of clusters. Treat all defects in a cluster as one defect and the location of this defect as the location of the cluster's center.

Step 5: Recalculate the number of defects on each wafer. Repeat Step 3 to check whether the adjusted (reduced) number of defects satisfies the assumptions for the Poisson distribution. If it does, go to Step 6; otherwise, go to Step 4.

Step 6: Set $\lambda =$ mean of the adjusted (reduced) number of defects in a wafer.

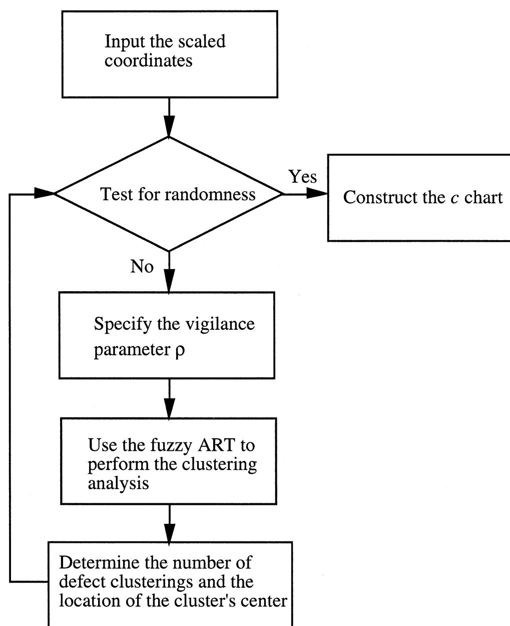

Fig. 2. Framework of the proposed procedure.

Step 7: Construct the c chart by using the following limits:

$$\begin{aligned} \mathrm{UCL} &= \lambda + 3\sqrt{\lambda} \\ \mathrm{LCL} &= \lambda - 3\sqrt{\lambda} \end{aligned} \tag{13}$$

Because input attributes for fuzzy ART must lie between 0 and 1, the normalization of input patterns is necessary. Let (x_i, y_i) be the coordinates of defects in the wafer. In this study, x_i and y_i are scaled between their minimum value and maximum value as follows [7]:

$$\begin{aligned} u_i &= \frac{x_i - \min(x)}{\max(x) - \min(x)} \\ v_i &= \frac{y_i - \min(y)}{\max(y) - \min(y)} \end{aligned} \tag{14}$$

where (u_i, v_i) are the scaled coordinates. Fig. 2 illustrates the proposed procedure's framework.

5. Implementation

In this section, a case study is presented to demonstrate the effectiveness of the proposed procedure. The required data in this case study are obtained from an IC manufacturing company in Taiwan. There are over 100 sequential process steps in IC fabrication. When the wafer goes through the process of 'Metal 2 Etch,' the coordinates of the wafer's defects are collected by using the KLA 2110 wafer inspection system. Once the wafer maps are obtained, outliers should be evaluated. In this study, we use the F-spread method [9] to identify the possible outliers. Let F_L and F_U be the first quartile and the third quartile of the number of defects, respectively. Any defect number greater than $D = F_U + 1.5(F_U - F_L) = 93 + 1.5(93 - 21) = 201$ can be treated as a possible outlier. After analyzing the causes of the possible outliers, we decided to discard them and 111 wafers are left for further analysis.

Greig-Smith's t-statistic is used to test whether the wafer's defects are randomly distributed. To reduce the probability of rejecting a wafer with randomly distributed defects, the significance level is set at 0.01. Since 396 dies are contained on each wafer, we have $n - 1 = 395$. The critical value of t is $t_{0.01,395} = 2.337$. If the t value is > 2.337, the

Table 1
The original and reduced number of defects and their corresponding t values for the 111 wafers

No.	Number of defects (before clustering)	t value	Number of defects (after clustering)	t value	No.	Number of defects (before clustering)	t value	Number of defects (after clustering)	t value
1	17	4.4	15	−0.50	57	19	3.8	16	−0.53
2	23	16	13	−0.43	58	29	20	19	−0.64
3	7	3.8	6	−0.18	59	35	73	20	−0.68
4	34	20	21	0.63	60	197	910	23	1.7
5	7	−0.21	random		61	48	51	27	1.2
6	40	12	29	0.95	62	34	14	25	1.4
7	50	13	38	0.91	63	56	130	12	−0.39
8	22	6.9	18	0.96	64	39	94	14	1.6
9	27	26	13	1.7	65	27	30	19	2.3
10	9	−0.28	random		66	61	220	15	1.4
11	33	30	19	−0.64	67	15	22	8	−0.25
12	12	−0.39	wcrandom		68	15	13	10	−0.32
13	13	1.7	random		69	20	2.1	random	
14	9	2.8	8	−0.25	70	19	8.3	17	1.1
15	32	5.9	29	0.95	71	48	39	25	1.4
16	35	0.4	random		72	16	4.7	14	1.6
17	48	220	17	−0.57	73	107	91	35	−0.4
18	6	4.5	5	−0.14	74	54	140	25	−0.85
19	23	45	13	−0.43	75	108	440	22	0.53
20	10	2.5	9	−0.28	76	9	2.8	8	−0.25
21	42	3.9	39	2.3	77	21	0.63	random	
22	25	0.27	random		78	184	16	135	0.45
23	18	2.5	16	−0.53	79	30	2.7	27	−0.93
24	9	−0.28	random		80	31	8	25	−0.85
25	28	82	10	−0.32	81	59	4.6	49	−1.10
26	10	2.5	9	−0.28	82	21	4.7	18	−0.6
27	30	−1	random		83	79	50	50	1.1
28	135	26	81	1.7	84	73	1.3	random	
29	85	150	44	1	85	50	5	42	−0.12
30	97	380	22	0.53	86	45	2.8	42	0.55
31	45	14	34	0.48	87	49	61	29	−0.025
32	68	340	23	0.44	88	79	220	22	0.53
33	8	3.3	7	−0.21	89	12	−0.39	random	
34	78	98	27	0.12	90	68	150	29	0.95
35	36	76	14	1.6	91	24	12	16	1.2
36	66	20	43	−0.18	92	13	6.1	11	−0.36
37	103	130	58	−0.57	93	19	−0.64	random	
38	124	430	52	0.35	94	19	2.3	random	
39	50	13	40	−0.68	95	21	0.63	random	
40	79	200	20	0.73	96	23	1.7	random	
41	47	180	18	−0.6	97	31	83	17	−0.57
42	72	7.6	63	1.4	98	16	−0.53	random	
43	169	58	101	−0.77	99	18	10	14	1.6
44	93	6.1	83	0.14	100	25	4.8	21	−0.71
45	46	14	37	0.24	101	30	4.6	27	1.2
46	30	1.8	random		102	41	58	21	−0.71
47	24	8.6	19	−0.64	103	26	11	21	−0.71
48	26	6.7	24	1.5	104	52	130	19	−0.64
49	105	66	34	2.1	105	19	6.8	16	1.2
50	40	140	9	−0.28	106	44	55	25	0.27
51	24	18	18	−0.6	107	37	79	23	0.44
52	120	150	39	2.3	108	18	0.96	random	

Table 1 (continued)

No.	Number of defects (before clustering)	t value	Number of defects (after clustering)	t value	No.	Number of defects (before clustering)	t value	Number of defects	t value
53	29	8.7	22	0.53	109	16	1.2	random	
54	108	370	26	1.3	110	20	6.4	16	−0.53
55	81	88	33	2.3	111	33	29	20	−0.68
56	30	20	22	−0.75					

defects tend to cluster. When a wafer exhibits a significant clustering phenomenon, the clustering analysis should be performed. For instance, wafer #32 has 68 defects and its t value is 340 (>2.337). Fig. 3 displays the results of applying the fuzzy ART for wafer #32 to form the clusters. According to this figure, four new clusters are formed. The first cluster contains 41 defects, the second cluster contains four defects, and the third and fourth clusters have two defects. Each of the four clusters is treated as one defect, thereby reducing the number of defects from 68 to 24. The t value for the reduced defects is 0.44 (<2.337), indicating that they are randomly distributed. Table 1 lists the original and reduced number of defects and their corresponding t values for the 111 wafers. Twenty wafers have randomly distributed defects.

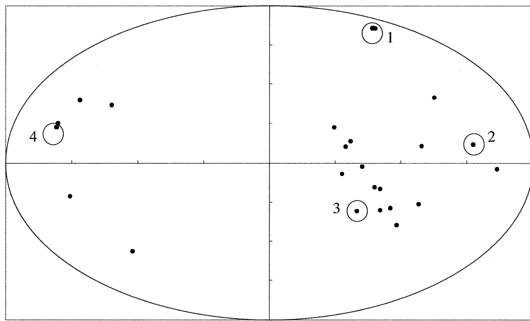

Fig. 3. Clustering result for wafer #32.

When the phenomenon of the defect clustering is removed, the average of the reduced number of defects, 25.74, can be used to construct the c chart. By using Eq. (13), UCL $= 40.96$. Fig. 4 plots the adjusted data, i.e., the reduced number of defects, in the c chart. Fourteen points above the UCL indicate 'out-of-control'. To verify the effectiveness of the proposed procedure, Fig. 4 also plots the estimated yield for each wafer. The estimated yield is defined by the percentage of the chip without defects in a

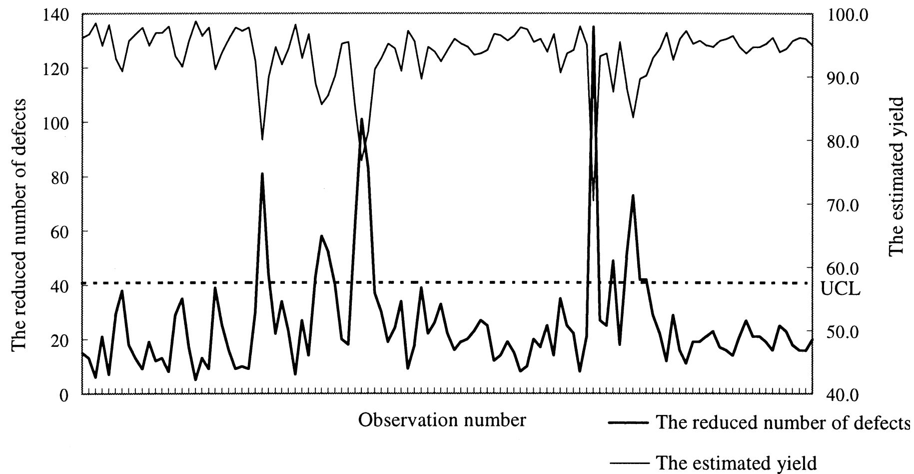

Fig. 4. Control chart for the proposed procedure.

wafer. This figure reveals that the smaller (larger) the reduced number of defects, the larger (smaller) the estimated yield of a wafer. No false alarms occur in Fig. 4.

6. Analysis and comparison

6.1. Analysis

Fuzzy ART is a network attempting to group patterns into the proper number of clusters. In fuzzy ART, all the parameters α_0, β and ρ may affect the number of clusters formed. The relationship between the reduced number of defects and the estimated yield of a wafer can be used to recognize the performance of the clustering result. The stronger the 'negative correlation' between the reduced number of defects and the estimated yield, the better the performance of the clustering result. Through several pilot runs, the values of α_0 and β seem to have only a slight effect for the clustering result; therefore, they are set at 0.01 and 0.5, respectively. To study the vigilance parameter's behavior on the performance of the clustering result, we vary ρ from 0.7 to 1.0 and apply the fuzzy ART for 111 wafers to form the clusters. Table 2 computes and lists the simple correlation coefficients for the number of clusters and the estimated yield. According to this table, reducing the vigilance parameter influences the clustering result. When ρ ranges between 0.95 and 0.99, a good clustering performance can be obtained. In this study, we recommend initially setting $\rho = 0.99$. If the wafer's defects are not randomly distributed, we progressively subtract the vigilance parameter 0.01 and repeat the clustering analysis by fuzzy ART until the desired situation is reached. In this study, the 111 wafers all pass the randomness test at $\rho \geq 0.95$.

Table 2
Sensitivity analysis of the vigilance parameter

ρ	Correlation coefficient
1.00	−0.751
0.99	−0.992
0.98	−0.994
0.97	−0.977
0.96	0.971
0.95	−0.947
0.90	−0.780
0.85	−0.613
0.80	−0.466
0.75	−0.353
00.70	−0.305

6.2. Comparison

When the conventional *c* chart is used, the mean of the number of defects for the 111 wafers is 44.50 and UCL = 64.51. Fig. 5 plots the defect data in the conventional *c* chart, showing that 24 points exceed the Poisson UCL. As indicated in Fig. 5, 10 false alarms occur: wafers #30, #32, #34, #40, #54,

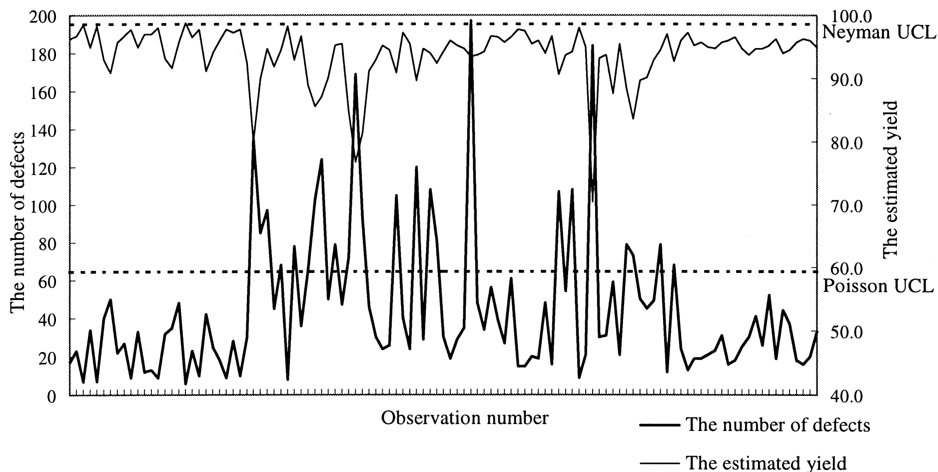

Fig. 5. Control chart for defect data with Poisson and Neyman upper control limits.

#55, #60, #75, #88 and #90. When the Neyman-based c chart suggested by Albin and Friedman [1,3] is constructed, the mean and the variance for the number of defects for the 111 wafers are 44.496 and 1391.107, respectively, and the estimated parameters for Neyman's type-A distribution are $\hat{\lambda} = 1.470$ and $\hat{\phi} = 30.264$. By using $\alpha = 0.27\%$, UCL = 194. Fig. 5 shows that one point exceed the Neyman UCL. The Neyman-based control limits are wider than the Poisson-based control limits and, therefore, can reduce some false alarms caused by the defect clustering. However, the Neyman-based c chart still cannot avoid the contradiction with the yield. For instance, the yields for wafers #28, #37, #42, #78 and #84 are extremely low; however, they are within the control limits. Furthermore, the simple correlation coefficients for the number of defects and the estimated yield (by using the two existing approaches and the proposed procedure) are -0.75 and -0.99, respectively, showing that the proposed procedure has the larger negative correlation coefficient. Above illustrations reveal that the proposed procedure yields a better result than those of the other two approaches.

7. Conclusion

IC products have changed from large scale (LSI) to very large scale (VLSI) since the mid-1970s. As the wafer size increases, a wafer's defects are no longer randomly distributed; instead, they tend to cluster. In this study, we propose a neural network-based procedure for process monitoring of clustered defects in IC fabrication. The fuzzy ART network for the clustering analysis is used to adjust the number of wafer defects, thereby allowing, not only the adjusted number of defects to satisfy the assumption of Poisson distribution, but also the conventional c chart to be used also. The proposed procedure can reduce the phenomenon of the false alarm caused by the clustered defects. The effectiveness of the proposed procedure is shown by using a case study. Owing to the fuzzy ART's high computational rate, the average execution time for each wafer (from Step 3 to Step 5 in the proposed procedure) is less than 2 s on a 486 PC. With these advantages, the proposed procedure can be written as a computer software that can be used for the online process control in IC fabrication.

Defects on nonpatterned wafers are frequently monitored to assess equipment performance. On the other hand, patterned wafer defects are monitored for online control of the manufacturing process. Monitoring defects on a patterned wafer may, in the future, become more important and advanced than monitoring defects on a nonpatterned wafer. Nevertheless, the proposed procedure can be applied to either patterned or nonpatterned wafers without any modifications. In addition, the proposed procedure can be applied to liquid crystal display (LCD) or printed circuit board (PCB) production.

Acknowledgements

The authors would like to thank Professor Wei-I Lee, Department of Electrophysics, National Chiao Tung University, Taiwan, for providing the necessary information in IC manufacturing and the anonymous reviewers for their helpful suggestions in improving the quality of the final manuscript. The National Science Council of the ROC is also appreciated for financial support of this manuscript under Contract No. NSC-86-2213-E-009-035.

References

[1] S. Albin, D.J. Friedman, Clustered defects in IC fabrication: impact on process control charts, IEEE Transactions on Semiconductor Manufacturing 4 (1) (1991) 36–42.
[2] C.H. Stapper, The effects of wafer to wafer density variations on integrated circuit defect and fault density, IBM J. Res. Dev. 29 (1) (1985) 87–97.
[3] S. Albin, D.J. Friedman, The impact of clustered defect distributions in IC fabrication, Management Science 35 (9) (1989) 1066–1078.
[4] C.M. Bishop, Neural networks and their applications, Rev. Sci. Instrum. 65 (6) (1994) 1803–1832.
[5] J.A. Ludwig, J.F. Reynolds, Statistical ecology: a primer on methods and computing, Wiley, 1988.
[6] A. Rogers, Statistical analysis of spatial dispersion: the quadrat method, Pion, 1974.
[7] L. Buke, S. Kamal, Neural networks and the part family/machine group formation problem in cellular manufacturing: a framework using fuzzy ART, Journal of Manufacturing Systems 14 (3) (1995) 148–159.

[8] G.A. Carpenter, S. Grossberg, D.B. Rosen, Fuzzy ART: faster stable learning and categorization of analog patterns by an adaptive resonance systems, Neural Netw. 4 (1991) 759–771.
[9] D.C. Hoaglin, B. Iglewicz, J.W. Tukey, Performance of some resistant rules for outlier labelling, Journal of American Statistical Association 81 (1986) 991–999.

Chao-Ton Su is an associate professor in the Department of Industrial Engineering and Management at National Chiao Tung University, Taiwan, ROC. He received his BS and MS from Chung Yuan Christian University, Taiwan, and PhD from University of Missouri-Columbia, USA, all in Industrial Engineering. His current research activities include quality engineering, production management and neural networks in industrial applications.

Lee-Ing Tong is an associate professor in the Department of Industrial Engineering and Management at National Chiao Tung University, Taiwan, ROC. She received her BS from National Cheng Kung University, Taiwan, and MS and PhD from University of Kentucky, USA, all in Statistics. She is formerly an associate professor in the Department of Mathematics at University of Massachusetts-Dartmouth, USA. Her current research activities include quality control, experimental design and statistics in industrial applications.