Spatial Correlogram-Based Dynamic Time Warping for Automatic Detection of Defect Patterns in Wafer Maps

Young-Seon Jeong*, Seong-Jun Kim+, Myong K. Jeong*
*Department of Industrial & Systems Engineering, Rutgers University, Piscataway, NJ 08854
+Department of Industrial & Systems Engineering, Kangnung National University, Kangnung, South Korea

Abstract
This paper proposes a new methodology which combines spatial correlogram with dynamic time warping (DTW) for the automatic detection of defect patterns in semiconductor wafer maps. Spatial correlogram is used for the detection of the presence of spatial autocorrelations while DTW distance is adopted for the automatic classification of defect patterns. The experimental results show that our method is robust to random noise, defect location and size.

1. Introduction

A wafer is an elementary unit in semiconductor manufacturing. Several hundred integrated circuits (ICs) are simultaneously fabricated on a single wafer (Fenner *et al.* 2005). After the completion of IC fabrication, each chip is classified as either functional or defective. A wafer map is used to display the locations of defective ICs chips on the wafer. A wafer map is likely to exhibit a spatial dependence across the wafer. As explained in Hansen *et al*. (1997), defective chips commonly occur in clusters or display some systematic patterns. Such defect patterns contain useful information about manufacturing process conditions (Cunningham and McKinnon 1998). For example, uneven temperatures or chemical aging lead to spatial cluster on the wafer map. Clusters also can be the result of crystalline nonuniformity, photo-mask misalignment or particles caused by mechanical vibration. Stepper and/or probe malfunctioning and sawing imperfections also are major causes of repetitive patterns. Material shipping and handling also can leave a scratch on the wafer map (Cunningham and McKinnon 1998, Hansen and Tyregod 1998, Hansen *et al*. 1997, and Taam and Hamada 1993).

The defect patterns represented on the wafer map hold important information that can assist process engineers in their understanding of the ongoing manufacturing processes. Consequently, wafer maps have been widely used in the semiconductor industry for process monitoring and yield enhancement. Chen *et al.* (2000) and Liu *et al.* (2002) developed intelligent systems that use wafer maps and wafer bin maps, respectively, to recognize defect spatial patterns and aid in the diagnosis of causes of failures. They adapted a neural network called as adaptive resonance theory network 1 (ART1) for this purpose. Hsieh (2004) developed an analytical structure made up of a fuzzy rule-based inference system to help identify defect spatial patterns. Tong *et al.* (2005) used the multivariate Hotelling T^2 control chart that indexes the number of defects and defect clusters as a way to monitor the wafer manufacturing process. The merit of this method is that it simultaneously monitors the number of defects and the presence of the cluster of defects.

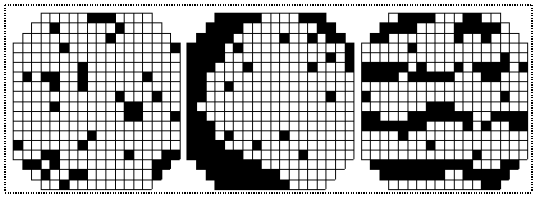
As a wafer gets larger, a spatial inhomogeneity frequently occurs. According to the literature (Bailey and Gatrell 1995), analysis of spatial inhomogeneity is also one of the promising approaches for detecting defective clustering. However, there is very little in the literature about the use of spatial correlogram to analyze defect patterns on the wafer map. This paper proposes a new methodology based on spatial correlogram to detect the presence of spatial autocorrelations and classify defect patterns. This paper is the first attempt to develop a methodology to detect spatial autocorrelation and to classify defect patterns automatically based on a spatial correlogram of a wafer map. After detecting the presence of defect patterns, dynamic time warping (DTW) is adopted to classify defect patterns into one of known patterns automatically. Spatial correlogram based on the proposed method is very robust to random noise, defect location, and defect size on the wafer map.

The remainder of this paper is organized as follows. Section 2 presents defect patterns on wafer map and spatial randomness test methods. Section 3 describes a spatial correlogram and proposes generalized joint-count based statistic with optimal weights. Section 4 contains a visual illustration that uses simulated and real life examples and presents a new spatial randomness test. In Section 5, we present the new automatic defect classification methodology and compare its performance with that of neural network. Section 6 presents conclusions and some future research topics.

2. Spatial dependences in wafer map

2.1. Defect patterns on wafer map

The spatial patterns formed by defect chips are broadly categorized into two classes: Random pattern and nonrandom pattern. Figure 1 illustrates typical examples of defect pattern in wafer map. Figure 1(a) is spatially random pattern which is assumed to be spatially homogeneous Bernoulli process (SHBP). Figures 1(b) and (c), respectively, show wafer maps with a clustered effect and with a regular pattern.



(a) SHBP (b) Clustered effect (c) Regular pattern

Figure 1. Typical spatial patterns of wafer map

The proposed methodology in this paper is to deal with chip-level wafer maps by comparing how many functional chips are around a defect chip and how many defective chips are around a functional chip.

2.2. Spatial randomness test

Spatial associative effects on the wafer map can be tested by log odds ratio (LOR) as follows (Taam and Hamada 1993).

$$LOR = \log \frac{c_{00} c_{11}}{(c_{01}/2)^2}$$

where c_{00}, c_{01} and c_{11} denote the counts of relation of adjacent chips. To discriminate between the three joins, indicator variable $x_i = 1$ for chip i when chip is defective and $x_i = 0$ when chip is functional. Then, c_{00}, c_{01} and c_{11} can be expressed as follows:

$$c_{00} = \sum_{i<j} w_{ij}(1-x_i)(1-x_j)$$

$$c_{01} = \sum_{i<j} w_{ij}(x_i - x_j)^2$$

$$c_{11} = \sum_{i<j} w_{ij} x_i x_j$$

where

$$w_{ij} = \begin{cases} 1, & x_i \text{ and } x_j \text{ are in Rook - Move neighbor} \\ 0, & \text{elsewhere} \end{cases}$$

Positive LOR indicates a clustered effect on the wafer map while a negative LOR indicates a regular effect like a chessboard pattern. Small LOR values around zero can be interpreted as indicators of no evidence of spatial dependence. Hansen and Thyregod (1998) have described a LOR test as a procedure to identify the statistical significance of spatial patterns. Because the standard deviation of LOR is approximately

$$\hat{\sigma}_{LOR} = \sqrt{c_{00}^{-1} + c_{11}^{-1} + 4c_{01}^{-1}}$$

for a large sample size (Agresti 1990), the following test statistic was proposed under an SHBP null hypothesis.

$$Z_{LOR} = \frac{LOR}{\hat{\sigma}_{LOR}} \sim N(0,1)$$

They concluded through numerical experiments that LOR test works well in detecting the presence of spatial dependences, but it is incapable of identifying spatial patterns.

3. Proposed methodology

This paper proposes hybrid approach that integrates spatial correlogram and DTW classification to automatically classify defect patterns on wafer map.

3.1. Spatial statistic for correlogram

As pointed out by Hansen and Thyregod (1998), a single monitoring statistic is insufficient to represent a variety of widespread patterns across the wafer map. To overcome this drawback, this paper presents new methodology to identify spatial patterns on the wafer map by using a spatial correlogram. Although spatial correlogram has been widely used in diverse fields of science such as geography, ecology, and the environment (Cliff and Ord 1981, Pierre and Louis 1998), to our knowledge, no study has reported the use of spatial correlogram for analysis of wafer maps.

Let $H(g)$ denote a set of gth-order neighbors, defined as chips that are g distant from each other. When the distance between two chips is denoted by $d(i,j)$, $H(g)$ can be written as

$H(g) = \{(i,j) \in W \mid d(i,j) = g\}$ for $g = 1,2,\cdots,m$

where W is a collection of all possible joins within the wafer map and m is a maximum length of join. Therefore the number of gth-order joins is

$$c(g) = \sum_{i<j} w_{ij}(g)$$

where

$$w_{ij}(g) = \begin{cases} 1, & (i,j) \in H(g) \\ 0, & \text{elsewhere} \end{cases}.$$

Based on previous literature (Cliff and Ord 1981), generalized spatial statistic can be expressed as

$$T(g) = \alpha_0 f(c_{00}(g)) + \alpha_1 f(c_{11}(g))$$

where α_0 and α_1 are weights to be chosen after consideration of the degree of spatial clustering. By minimizing the variance of *T(g)* subject to $\alpha_0 + \alpha_1 = 1$, the following lemma presents the optimal weight value.

Lemma1 1: For a generalized spatial statistic with *g*th-order neighbors, *i.e.*

$$T(g) = \alpha_0 f(c_{00}(g)) + \alpha_1 f(c_{11}(g)),$$

the optimal weights that minimize the variance of *T(g)* when $f(\cdot)$ is identity function are given as follows:

$$(\alpha_0, \alpha_1) = (p, q) \text{ subject to } \alpha_0 + \alpha_1 = 1.$$

where *p* is defective rate and *q*=1-*p*.
See the Appendix for the proof of Lemma 1.

Therefore, the gth-order *T(g)* can be simplified as follows:

$$T(g) = pc_{00}(g) + qc_{11}(g).$$

In addition, expectation and variance of statistic *T(g)* are obtained as follows (See the Appendix for their detailed derivation):

$$\mathrm{E}[T(g)] = c(g)pq$$
$$\mathrm{V}[T(g)] = c(g)p^2q^2$$

Standardized statistic *T(g)* approximates the standard normal distribution based on Central Limit Theorem, i.e.,

$$Z_T(g) = \frac{T(g) - c(g)pq}{\sqrt{c(g)p^2q^2}} \sim N(0,1) \quad \text{as } c(g) \to \infty, \qquad (1)$$

where $c(g) = c_{00}(g) + c_{11}(g) + c_{01}(g)$.

4. Illustrative Case Study

This section presents the case study using simulated and real-life wafer maps.

4.1. Simulated wafer maps

Before investigating spatial correlogram for defect pattern classification, we present formal randomness test that combines the test statistic with multiple spatial lags. There are several test statistics for spatial randomness testing such as LOR test and CR test as mentioned in section 2.2. However, they are not applicable for spatial randomness test using multiple spatial lags because test statistic using multiple spatial lags should take into account all test values with different lags simultaneously. As we mentioned in Section 3, $Z_T(g)$ is approximately normally distributed when $c(g)$ is large. If we let $X_r = [Z_T(1), Z_T(2), \ldots, Z_T(r)]$, which is a collection of $Z_T(g)$ for the first *r* spatial lags, then X_r follows approximate multivariate normal distribution with mean of zero vector of length *r* and covariance $\sum_r$ under SHBP condition. For spatial randomness test using the first *r* spatial lags, we can use the following test statistic:

$$T_H^2(r) = \mathbf{x}_r' \hat{\Sigma}_r^{-1} \mathbf{x}_r,$$

where $\hat{\Sigma}_{rr}$ is the estimated covariance matrix using *m* samples of wafer maps. Because the samples are individual observations (i.e., individual wafer maps), an approximate critical limit is given by (Montgomery, 2005)

$$\mathrm{CL} = \frac{r(m-1)}{m-r} F_{\alpha,r,m-r}$$

Table 1.Spatial randomness testing for SHBP

P	0.1	0.2	0.3	0.4	0.5
LOR	98%	97%	92%	95%	93%
CR	98%	97%	93%	95%	94%
$T_H^2(1)$	96%	97%	95%	97%	97%
$T_H^2(2)$	97%	98%	98%	96%	94%
$T_H^2(3)$	97%	96%	98%	98%	94%
$T_H^2(4)$	98%	97%	97%	97%	93%
$T_H^2(5)$	95%	97%	96%	98%	96%
$T_H^2(6)$	95.%	97%	97%	97%	97%
$T_H^2(7)$	98%	98%	96%	97%	97%
$T_H^2(8)$	97%	97%	96%	96%	97%
$T_H^2(9)$	96%	95%	94%	94%	95%
$T_H^2(10)$	97%	95%	96%	93%	95%

Table 1 shows the comparison results of the proposed spatial randomness test for 150 SHBP

wafer maps in terms of test accuracy with popular LOR and CR tests ($\alpha = 0.05$). We have used the total 150 of SHBP wafer maps for each defective rate (p) ranging from 0.1 to 0.5. Overall, existing test procedures performed better for low defective rate (p=0.1) while our proposed procedure showed the improved performance for larger defective rates ($p \geq 0.2$). In our testing procedure, as defective rate (p) is getting larger, test statistic using larger spatial lags produced better accuracy. Therefore, 2~4 spatial lags is recommended for spatial randomness test for smaller defective rate ($p \leq 0.3$) whereas 5~7 is recommended for larger defective rates.

However, as mentioned earlier, previous randomness test is insufficient to recognize a variety of widespread defect patterns across the wafer map. This paper proposes new approach to detect spatial defect patterns using spatial correlogram. We construct 20x20-sized wafer maps as specified SHBP, cluster, circle, repetition, and mixed patterns. The generation of the simulated wafer maps is based on the previous literature (DeNicolao *et al.* 2003) except the SHBP. SHBP wafer maps are produced using random number generator.
Figure 2 shows spatial correlograms for each spatial defect pattern in wafer map.

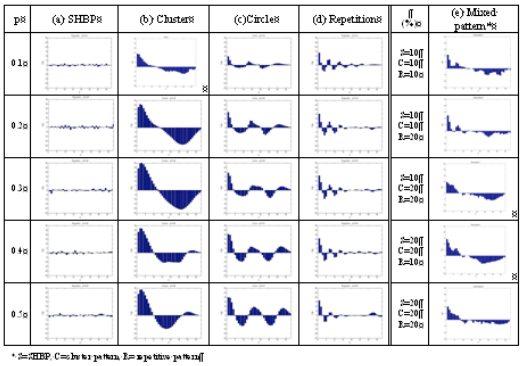


Figure 2. Spatial correlograms of simulated wafer maps

In Figure 2, the $Z_T(g)$ values computed from different detective rate p using equation (1) are displayed along the spatial lag g. In case of SHBP as shown in Figure 2(a), most of the $Z_T(g)$ values fluctuate around zero. SHBP has no unique shape of the correlogram. In case of the cluster patterns, $Z_T(g)$ smoothly changes along spatial lag g, and its absolute values are relatively larger. The characteristic of a spatial correlogram of a circle pattern is that the absolute value of $Z_T(g)$ is also large like that of cluster pattern, but the circle pattern contains a soft cosine waveform. $Z_T(g)$ values of repetitive pattern are consistently small and no special pattern appears except for a frequent crossing around zero. Finally, we attempt to simulate mixed effects by superimposing three types of wafer maps: SHBP, cluster pattern, and repetitive pattern. Figure 3 shows some examples of wafer maps with mixed effects of cluster pattern and repetitive pattern and their corresponding correlograms are shown in Figure 2(e). Interestingly, the distinctive patterns produced by the cluster and repetitive patterns are preserved under the superimposed models. It is indicated that global shape of spatial correlogram is similar to cluster's one and at the same time it locally contains the characteristic of repetitive pattern's one.

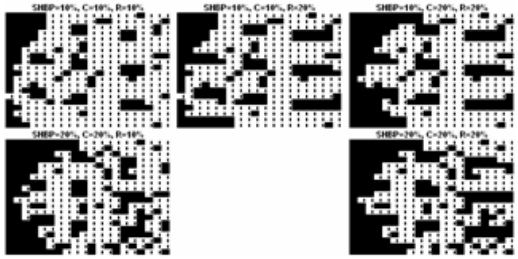
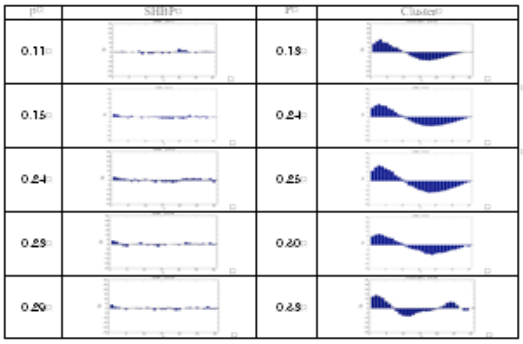


Figure 3. Wafer map with mixed effects
(C: cluster pattern, R: repetitive pattern)

4.2. Real-life wafer maps

Real-life wafer maps provided by a semiconductor manufacturing company were analyzed using the proposed approach. Each of wafer maps consists of 268 chips as shown in Figure 4(a). The defective rate p in Figure 4 is calculated as followed: $p=b/N$ where b is the number of defective chips and N is the number of total chips on the wafer. In Figure 4(b), it is observed that there is a distinction between spatial correlograms of cluster pattern and those of SHBP. In case of the clustered effects, $Z_T(g)$ changes smoothly along spatial lag g, and its absolute values are relatively larger. Not so with SHBP in which a frequent crossing around zero occurs.

(a) Real-life wafer maps



(b) Spatial correlograms

Figure 4. Real-life wafer maps and their corresponding spatial correlograms

Illustrative examples show that a spatial correlogram has the potential to identify defect patterns in semiconductor wafers. Information drawn by a spatial correlogram includes not only simple examination of spatial dependence for defective chips, but also recognition of a defect pattern in a wafer. Moreover, we can discover several benefits of our approach. A spatial correlogram is robust to defect location, robustness to defect size, and robustness to random noise. More details on the advantage of the proposed method will be explained in Section 5.

5. Automatic Classification Defect Pattern

This section presents new classification methodology based on dynamic time warping (DTW) using spatial correlogram and compares its performance with that of popular neural network approach (Hsu *et al*. 2007, Huang 2007, and Palma *et al*. 2005).

5.1. Automatic defect pattern classification

As seen earlier, defect patterns of same class produce similar shapes of correlogram. In order to classify defect patterns based on correlogram, we have to first calculate the distance among different correlograms and then use the classification techniques that use distance measures. In this paper, we use the 1-nearest neighbor classifier. However, since they are not aligned in the lag axis, linear mapping technique such as Euclidean distance that assumes *i*th point in one correlogram is aligned with the *i*th point in the other may produce higher misclassification rate. To accurately classify each defect pattern using correlogram, DTW distance, which finds an optimal match between two sequences by allowing a non linear mapping of the one sequence to another by minimizing the distance between the two, is applied for defect classification. Please see the DTW-related references for more details (Ratanamahatana and Keogh, 2004a and 2004b).

5.2. Experiment and results

In order to evaluate the classification performance of the proposed algorithms, we generated a total of 400 wafer maps with 400 chips per wafer (20 by 20-sized map), i.e., 80 wafer maps for each of five patterns such as SHBP, circle, cluster, repetition and spot. We have eight level of random noise ranging from 0.05, 0.1, 0.15,⋯ to 0.4. For each combination of noise level and pattern (total 8x5=40 combinations), we generated 10 wafer maps. Dataset {1} consists of wafer maps with the noise level of 0.05, dataset {2} with the noise level of 0.1, and so on. In this experiment, we divided 400 wafer maps into four different data sets as shown in Table 2.

4-fold cross validation (CV) is implemented for the comparison of classification accuracy of different procedures. Binary and multi-lags based supervised multilayer perceptron neural network (Huang 2007) is selected for comparison with the proposed method. The difference between the binary neural network (B-NNet) and multi-lags neural network (M-NNet) is the input vector. The type of input vector of binary neural network is "1" or "0" while that of multi-lags neural network is the $Z_T(g)$ values along spatial lag *g*. For instance, in case of wafer map with 20 by 20 size, binary neural networks have 400 (=20x20) binary values ("0" or "1") as input vector. On the other hand, multi-lags neural networks have a total 38 of $Z_T(g)$ as its input vector because a total number of spatial lag under rook-move neighborhood (RMN) rule of 20 by 20 sized wafer map is 38.

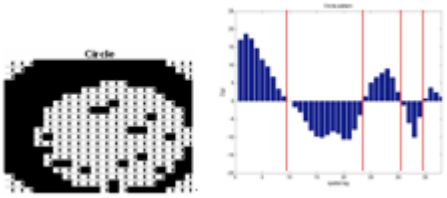
The architecture of neural network is composed as follows: 400 neurons in the input layer for binary neural network and 38 neurons for multi-lags neural network, single hidden layer with 10 neurons, and 1 output neurons.

Tangent sigmoid function and linear transfer function are used for activation function in the hidden and output layer. On the other hand, multi-lags DTW utilizes a number of 38 of $Z_T(g)$ because a maximum number of spatial lag under RMN rule of 20 by 20 sized wafer map is 38 which shows best performance.

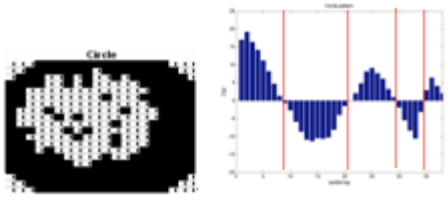
Table 2. Summary of classification performance

Testing set	B-NNet	M-NNet	Euclidean distance	DTW
{1},{2}	81.3%	78.8%	92.5%	98.8%
{3},{4}	73.8%	71.3%	86.3%	92.5%
{5},{6}	58.8%	62.5%	77.5%	82.5%
{7},{8}	58.8%	71.3%	83.8%	88.8%
Average	68.2%	71.0%	85.0%	90.6%

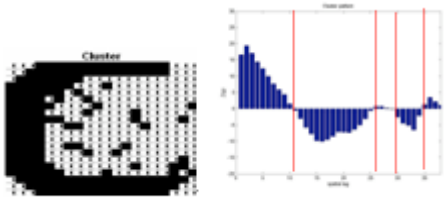
Table 2 shows the accuracy of four procedures for both average and each fold of 4-fold CV datasets. Overall, the proposed method is better than other ones. Especially, the accuracy of DTW outperforms that of Euclidean distance. The experimental results show that multi-lags based DTW is promising alternative for automatic defect classification of wafer map.



(a) Testing wafer map and corresponding correlogram



(b) Best matching wafer map by DTW and corresponding correlogram

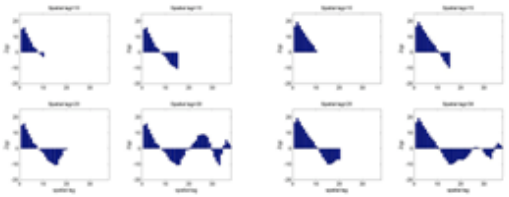


(c) Best matching wafer map by Euclidean distance and corresponding correlogram

Figure 5. Classification results using DTW and Euclidean distance

Figure 5 shows why DTW works to classify diverse defect types. Figure 5 (a)-(c) show the testing wafer map, best matching wafer maps by DTW and Euclidean distance with their corresponding correlograms, respectively. The changes of defect location and size make some horizontal shift of correlograms. DTW accurately classifies the new wafer map into circle pattern whereas Euclidean distance misclassifies it into cluster pattern. In specific, the distance between X_{new} and X_{circle} by DTW is $d^{DTW}_{(X_{new},X_{circle})}$=47.6 while the distance between X_{new} and X_{circle} by DTW is $d^{DTW}_{(X_{new},X_{cluster})}$=257.3, so the new wafer map is classified into circle pattern based on the 1-nearest neighbor classifier. On the other hand, the distance between X_{new} and X_{circle} by Euclidean distance is $d^{ED}_{(X_{new},X_{circle})}$=26.4 while the distance between X_{new} and $X_{cluster}$ by Euclidean distance is $d^{ED}_{(X_{new},X_{cluster})}$=19.1, so the new wafer map is classified into cluster pattern.

For classification of spatial correlogram based on the 1-nearest neighbor classifier (of other classifiers using distance measures), it is important to compute the distance between two correlograms. However, some defect patterns can not be clearly be discriminated using small number of spatial lags. For example, as shown in Figure 6, circle and cluster patterns can not be clearly discriminated using smaller spatial lags while large number of spatial lags (≤30) clearly do.



(a) Circle pattern (b) Cluster pattern

Figure 6. Spatial correlogram of circle and cluster patterns

In order to explore an optimal spatial lags for the proposed classification method, Table 3 shows the classification accuracy of DTW with different spatial lags. As shown in Table 3, larger

spatial lags are used, better classification accuracy is obtained. Therefore, for the classification purpose, full number of spatial lags is suggested for accurate classification.

Table 3. Classification accuracy of DTW with different spatial lags

Testing set	Lags=10	Lags=20	Lags=30	Lags=38
{1},{2}	81.3%	73.8%	98.8%	98.8%
{3},{4}	72.5%	82.5%	90.0%	92.5%
{5},{6}	61.3%	67.5%	76.3%	82.5%
{7},{8}	60.0%	72.5%	77.5%	88.8%
Average	68.8%	74.1%	85.7%	90.6%

To investigate the effectiveness of multi-lags based DTW, the details of classification performance of testing set {3, 4} are shown in the Table 4. As seen in Table 4, multi-lags based DTW misclassify circle pattern into cluster pattern or in opposite because in some cases, the distinction between spatial correlograms of two patterns is not clear due to high random noise. It can accurately classify repetitive and spot pattern regardless to random noise.

Table 4. Detail of DTW performance of testing set {3, 4}

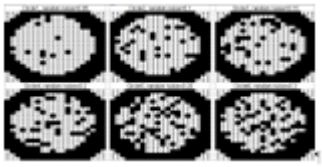
Type	Accuracy	Misclassification description
Circle	18/20 (90%)	Circle → Cluster
Cluster	16/20 (80%)	Cluster → Circle Cluster → Spot
Repetition	20/20 (100%)	None
Spot	20/20 (100%)	None

5.3. Test of robustness to noise, defect location, and defect size

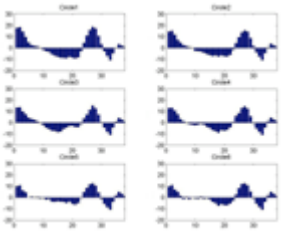
This section compares the multi-lags based DTW with existing algorithm such as binary neural networks in terms of the robustness to random noise, defect location, and defect size.

5.3.1 Robustness to random noise

Figure 7(a) shows same circle patterns with different random noise level ranging 0.05 to 0.3. Figure 7(b) presents correlograms of circle patterns in Figure 7(a). Each correlogram is similar regardless of random noise level.



(a) Wafer maps of circle patterns



(b) Spatial correlograms

Figure 7. Circle patterns with different levels of random noise and their corresponding spatial correlograms

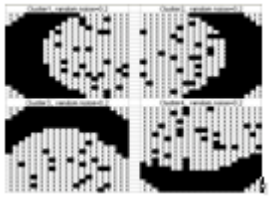
Table 5 summarizes the classification accuracy of circle patterns with different noise level from the experiments in Section 5.1. Our proposed procedure is robust to random noise whereas accuracy of other techniques decreases as noise level becomes large.

Table 5. Classification accuracy of circle pattern with different noise level

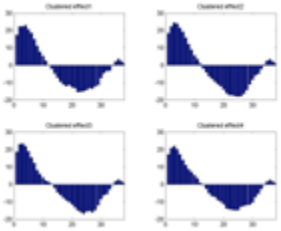
Noise level	B-NNet	M-NNet	DTW
0.05~0.1	100%	100%	100%
0.15~0.2	95%	35%	90%
0.25~0.3	50%	35%	85%

5.3.2 Robustness to defect location

Figure 8(a) shows cluster patterns with different locations for a fixed random noise rate of 0.2 and their corresponding spatial correlograms are shown in Figure 8(b). The correlograms are almost same without regard to defect location.



(a) Wafer maps of cluster patterns



(b) Spatial correlograms

Figure 8. Cluster patterns with different defect locations and their corresponding spatial correlograms

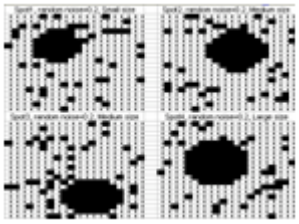
Table 6 summarizes the classification accuracy of cluster patterns with different defect locations from the experiments in Section 5.1. Our proposed procedure is robust to defect locations while other procedures show some classification errors for different defect locations.

Table 6. Classification accuracy of cluster pattern with different defect location

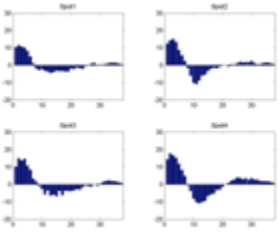
Defect location	B-NNet	M-NNet	DTW
Right	80%	60%	100%
Left	60%	80%	100%
Up	80%	80%	100%
Down	60%	80%	100%

5.3.3 Robustness to defect size

Figure 9(a) shows spot pattern with different defect size for a random noise rate of 0.2 and their corresponding correlograms are shown in Figure 9(b). The shape of correlogram looks somewhat different, but unique characteristics of spot pattern such as three waves are preserved.



(a) Wafer maps of spot patterns



(b) Spatial correlograms

Figure 9. Spot patterns with different defect size and their corresponding spatial correlograms

Table 7 shows the classification accuracy of spot pattern with different defect size. Our proposed procedure shows slightly better robust performance to different defect size compared to neural network-based approaches.

Table 7. Classification accuracy of spot pattern with different defect size

Defect size	B-NNet	M-NNet	DTW
Small	100%	70%	100%
Medium	90%	80%	100%
Large	70%	80%	90%

6. Conclusions

Although an analysis of wafer map helps to better understand ongoing process problems, defect classification can not be easily identified automatically. This paper proposes a new methodology which incorporates spatial correlogram and DTW to detect the anomaly defect patterns and classify them into one of existing spatial defect patterns. The new spatial randomness test procedure based on spatial correlogram of wafer map is used to detect anomaly defect patterns whereas the 1-nearest neighbor classifier using DTW distance with correlogram input is used to classify its corresponding anomaly defect type. Simulation studies show that the proposed methodology is generally more effective in detecting and classifying spatial defect patterns on the wafer map than those methods that use single lag. The experimental results show that our novel methodology is robust to random noise, defect location and defect size. Therefore, this study can be expected to make a contribution to the monitoring and diagnosis of IC manufacturing processes.

As for further study, there is a need to develop more advanced classification techniques of spatial patterns based on spatial correlogram. Also, we may extend our proposed approach to

the wafer bin map that is more informative than the binary wafer map.

Appendix:

A. Proof of Lemma 1

Given n_0 and n_1, the first and the second moments of $c_{00}(g)$ and $c_{11}(g)$, respectively, are obtained as follows (Cliff and Ord 1981, Hansen *et al*. 1997).

$$E(c_{00}(g)) = c(g)n_0^{(2)} / n^{(2)}$$
$$E(c_{11}(g)) = c(g)n_1^{(2)} / n^{(2)}$$
$$E(c_{00}^2(g)) = \sum_{k=1}^{4} b_k(g)n_0^{(k)} / n^{(k)}$$
$$E(c_{11}^2(g)) = \sum_{k=1}^{4} b_k(g)n_1^{(k)} / n^{(k)}$$
$$E(c_{00}(g)c_{11}(g)) = b_4(g)n_0^{(2)}n_1^{(2)} / n^{(4)}$$

where
$c(g) = c_{11}(g) + c_{00}(g) + c_{01}(g)$, $b_1(g) = 0$, $b_2(g) = s_1(g)/4$, $b_3(g) = (s_2(g) - 2s_1(g))/4$, $b_4(g) = (s_0^2(g) - s_2(g) + s_1(g))/4$ and $m^{(k)} = \prod_{i=1}^{k}(m - i + 1)$ for a positive integer m. See Cliff and Ord (1981) to find $s_0(g)$, $s_1(g)$ and $s_2(g)$. Accordingly, we can have the expectation and the variance of $T(g)$ as

$$E(T(g)) = c(g)(\alpha_0 n_0^{(2)} + \alpha_1 n_1^{(2)}) / n^{(2)}$$

$$\begin{aligned} Var(T(g)) = {} & s_1(g)\left[\alpha_0^2\left(\frac{n_0^{(2)}}{n^{(2)}} - 2\frac{n_0^{(3)}}{n^{(3)}} + \frac{n_0^{(4)}}{n^{(4)}}\right) + \alpha_1^2\left(\frac{n_1^{(2)}}{n^{(2)}} - 2\frac{n_1^{(3)}}{n^{(3)}} + \frac{n_1^{(4)}}{n^{(4)}}\right) + 2\alpha_0\alpha_1\frac{n_0^{(2)}n_1^{(2)}}{n^{(4)}}\right] \\ & + s_2(g)\left[\alpha_0^2\left(\frac{n_0^{(3)}}{n^{(3)}} - \frac{n_0^{(4)}}{n^{(4)}}\right) + \alpha_1^2\left(\frac{n_1^{(3)}}{n^{(3)}} - \frac{n_1^{(4)}}{n^{(4)}}\right) - 2\alpha_0\alpha_1\frac{n_0^{(2)}n_1^{(2)}}{n^{(4)}}\right] \\ & + s_0^2(g)\left[\alpha_0^2\left(\frac{n_0^{(4)}}{n^{(4)}} - \left\{\frac{n_0^{(2)}}{n^{(2)}}\right\}^2\right) + \alpha_1^2\left(\frac{n_1^{(4)}}{n^{(4)}} - \left\{\frac{n_1^{(2)}}{n^{(2)}}\right\}^2\right) - 2\alpha_0\alpha_1\left(\frac{n_0^{(2)}n_1^{(2)}}{n^{(4)}} - \frac{n_0^{(2)}n_1^{(2)}}{n^{(2)}n^{(2)}}\right)\right] \end{aligned}$$

respectively. Because $n_0^{(k)} / n^{(k)} \cong q^k$ and $n_1^{(k)} / n^{(k)} \cong p^k$ for a large value of n, the variance of $T(g)$ is approximately derived by

$$\begin{aligned} Var(T(g)) = {} & s_1(g)p^2q^2(\alpha_0 + \alpha_1)^2 \\ & + (s_2(g) - 4s_0^2(g)/n)pq(\alpha_0 q - \alpha_1 p)^2 \end{aligned}$$

Minimizing the above equation subject to $\alpha_0 + \alpha_1 = 1$, we can easily find an optimum solution $(\alpha_0^*, \alpha_1^*) = (p, q)$.

Based on this result, the corresponding expectation and variance are respectively

$$E(T(g)) = c(g)pq$$
$$Var(T(g)) = c(g)p^2q^2.$$

References

1. A. Agresti, *Categorical Data Analysis*, Wiley, 1990.
2. A. D. Cliff and J. K. Ord, *Spatial Processes: Models & Applications*, Pion, 1981.
3. C. A. Ratanamahatana and E. Keogh, "Making time-series classification more accurate using learned constraints," in *Proc. of the 4th SLAM Int. Conf. on data mining,* April 2004a.
4. C. A. Ratanamahatana and E. Keogh, "Everything you know about dynamic time warping is wrong," in *Proc. 10th ACM SIGKDD Int. Conf. on knowledge discovery and data mining,* August 2004b.
5. C. J. Huang, "Clustered defect detection of high quality chips using self-supervised multilayer perceptron," *Expert Sys. App.*, vol. 33, pp. 996-1003, 2007.
6. C. K. Hansen and P. Thyregod, "Use of wafer maps in integrated circuit manufacturing," *Microelectronics Reliability*, vol. 38, pp. 1155-1164, 1998.
7. D. C. Montgomery, *Introduction to Statistical Quality Control*, Wiley, 2005.
8. F. D. Palma, G. D. Nicolao, G. Miraglia, E. Pasquinetti, and F. Piccinini, "Unsupervised spatial pattern classification of electrical-wafer-sorting maps in semiconductor manufacturing," *Pattern Recognit. Lett.*, vol. 26, pp. 1857-1865, 2005.
9 F. L. Chen and S. F. Liu, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication," *IEEE Trans. Semiconduct. Manuf.*, vol. 13, no. 3, pp. 366-373, 2000.
10. G. DeNicolao, E. Pasquinetti, G. Miraglia, and F. Piccinini, "Unsupervised spatial pattern classification of electrical failures in semiconductor manufacturing," *Artificial Neural Nets. Pattern Recognit. Workshop,* pp. 125-131, 2003.
11. H. W. Hsieh and F. L. Chen, "Recognition of defect spatial patterns in semiconductor fabrication," *Int. J. Production Res.*, vol. 42, no. 19, pp. 4153-4172, 2004.

12. J. S. Fenner, M. K. Jeong, and J. C. Lu, "Optimal automatic control multistage production processes," *IEEE Trans. Semiconduct. Manuf.*, vol. 18, no. 1, pp. 94-103, 2005.
13. L. I. Tong, C. H. Wang, and C. L. Huang, "Monitoring defects in IC fabrication using a Hotelling T^2 control chart," *IEEE Trans. Semiconduct. Manuf.*, vol. 18, no. 1, pp. 140-147, 2005.
14. L. Pierre and L. Louis, *Numerical Ecology*, Elsevier, 1998.
15. M. H. Hansen, V. N. Nair, and D. J. Friedman, "Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects," *Technometrics*, vol. 39, no. 3, pp. 241-253, 1997.
16. S. C. Hsu and F. L. Chen, "Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing," *Int. J. Production Econ.*, vol. 107, pp. 88-103, 2007.
17. S. F. Liu, F. L. Chen, and W. B. Lu, "Wafer bin map recognition using a neural network approach," *Int. J. Production Res.*, vol. 40, no. 10, pp. 2207-2223, 2002.
18. S. P. Cunningham and S. McKinnon, "Statistical methods for visual defect metrology," *IEEE Trans. Semiconduct. Manuf.*, vol. 11, no. 1, pp. 48-53, 1998.
19. T. C. Bailey and A. C. Gatrell, *Interactive Spatial Data Analysis*, Prentice Hall, 1995.
20. W. Taam and M. Hamada, "Detecting spatial effects from factorial experiment: an application from IC manufacturing," *Technometrics*, vol. 35, no. 2, pp. 149-160, 1993.