

Semiconductor fault detection and classification for yield enhancement and manufacturing intelligence

Chen-Fu Chien · Chia-Yu Hsu · Pei-Nong Chen

Published online: 29 August 2012
© Springer Science+Business Media, LLC 2012

Abstract Process monitoring and profile analysis are crucial in detecting various abnormal events in semiconductor manufacturing, which consists of highly complex, interrelated, and lengthy wafer fabrication processes for yield enhancement and quality control. To address real requirements, this study aims to develop a framework for semiconductor fault detection and classification (FDC) to monitor and analyze wafer fabrication profile data from a large number of correlated process variables to eliminate the cause of the faults and thus reduce abnormal yield loss. Multi-way principal component analysis and data mining are used to construct the model to detect faults and to derive the rules for fault classification. An empirical study was conducted in a leading semiconductor company in Taiwan to validate the model. Use of the proposed framework can effectively detect abnormal wafers based on a controlled limit and the derived simple rules. The extracted information can be used to aid fault diagnosis and process recovery. The proposed solution has been implemented in the semiconductor company. This has simplified the monitoring process in the FDC system through the fewer key variables. The results demonstrate the practical viability of the proposed approach.

Keywords Fault detection and classification (FDC) · Data mining · Manufacturing intelligence · Yield enhancement · Semiconductor manufacturing

C.-F. Chien · P.-N. Chen
Department of Industrial Engineering and Engineering Management,
National Tsing Hua University, Hsinchu 30013, Taiwan

C.-Y. Hsu (✉)
Department of Information Management, Yuan Ze University,
135 Yuan Tung Road, Chungli 32003, Taiwan
e-mail: cyhsu@saturn.yzu.edu.tw

1 Introduction

Semiconductor manufacturing involves highly complex and lengthy wafer fabrication processes with 300–500 process steps and a large number of interrelated variables. As the feature size and critical dimension of integrated circuits (IC) continuously shrink to accommodate nano-generation, maintaining process specifications and controlling quality is increasingly difficult in semiconductor manufacturing (Chien and Hsu 2011). Thus, process monitoring and profile analysis are critical in detecting various abnormal events in wafer fabrication to enhance yield and control quality. Process monitoring and fault detection involve detecting abnormal processes and equipment including mean and variance shifts in one or more process variables, spikes, and drifts to quickly remove the assignable causes. However, process measurements are often insufficient, incomplete, or unreliable because of data quality or various other causes (Venkatasubramanian et al. 2003).

Therefore, there is a critical need in the semiconductor industry to effectively and efficiently monitor processes in real time and extract useful profile information to support process monitoring and fault detection. The extracted information assists engineers in understanding the process status and also in quickly removing abnormal behaviors and assignable causes. Good process monitoring leads to less downtime, improvement of production quality and reduction of manufacturing costs. With advanced information technology and metrology sensors, massive amounts of data are routinely collected during wafer processes such as data on temperature, pressure, flow rate, and power. The temporal patterns in these equipment parameters or process variables are signals of equipment behavior. Early detection and quick diagnosis of process faults are important to ensure tool effectiveness while process operation is controllable and to reduce yield loss.

With increasing demand for high-quality products and reliable processes, multivariate statistical process control (MSPC) is widely used in the manufacturing industry to ensure that equipment is “statistically controlled” by monitoring two or more related quality characteristics simultaneously (Montgomery 2005). Conventional univariate charts such as the Shewhart control chart, cumulative sum (CUSUM) control chart, and exponentially weighted moving average (EWMA) control chart have been used to monitor the deviation of key variables or parameter performance on the final product. However, univariate control charts have weaknesses. First, an engineer must monitor a large amount of charts with increasing complexity and manufacturing processes. Second, correlation variations between variables are difficult to be detected by univariate control charts. Third, the type I error increases with the number of charts. MSPC is used to monitor correlated variables with few control charts. The Hotelling T^2 control chart is a popular MSPC method to detect the out-of-control signals. Although a T^2 control chart is useful and powerful, it assumes that variables are normally distributed and independent, which does not occur in practice.

Projection methods are alternatives to MSPC, which reduce the dimensionality of process variables. Projection methods include principal component analysis (PCA) and partial least square (PLS) (Skagerberg et al. 1992; MacGregor et al. 1994; Kourti and MacGregor 1996; Ralson et al. 2001). Multi-way PCA (MPCA)

integrates a time dimension into the PCA model. It was developed for fault detection in batch process monitoring (Wold et al. 1987b; Nomikos and MacGregor 1994). Although MSPC methods are used to detect process deviation real-time equipment monitoring is difficult because diagnosis procedures in real settings often rely on many manual actions by human operators.

To address the requirements of real settings, this study aims to develop a manufacturing intelligence approach for semiconductor fault detection and classification (FDC). It uses MPCA and data mining approaches to monitor and diagnose the semiconductor fabrication process. With advanced information technology, data mining approaches have been used to explore large databases automatically or semi-automatically and extract useful rules and patterns to improve decision quality (Chien et al. 2007). Manufacturing intelligence approaches have also been developed to derive decision rules to enhance operation efficiency and effectiveness (Chen and Chien 2011; Chien et al. 2010, 2011; Kuo et al. 2010). MSPC methods and data mining approaches can be used to discover and extract information from historical process data and can assist fault diagnosis and recovery. MPCA is used to unfold three-dimensional (3D) process data and reduce batch process data dimensions by rotating the original axis in the process data. A few principal components (PCs) are extracted to explain the maximal amount of variation. The D statistic and Q statistic control limits from the score space and residual space, respectively, are then constructed to detect the abnormal wafers. The set of process variables relevant to the detected abnormal events are then identified. Finally, decision trees are used to derive fault classification rules. To validate the proposed framework, an empirical study was conducted on the Chemical Vapor Deposition (CVD) process in a leading semiconductor company in Taiwan. The proposed approach and derived rules allowed the results of routine monitoring processes and information extracted from process data to be used to identify critical process variables and remove fault causes.

The study is organized as follows. Section 2 introduces a literature review of the relevant research on fault detection and classification and MSPC approaches. Section 3 presents a description of the proposed approach for semiconductor fault detection and classification. An empirical study in a leading semiconductor company is conducted to validate the proposed approach in Sect. 4. Section 5 offers a conclusion to the paper with a discussion of findings and future research directions.

2 Literature review

The notations and terminologies used in this paper are as follows:

- i Wafer index
- j Process measurement variable index
- k Sample of a time interval index
- n Number of observations
- I Number of wafers

- J Number of process variables
- K Number of total samples recorded by time
- $\underline{\mathbf{X}}$ Three-dimensional historical process data array
- \mathbf{X} Unfolded two-dimensional data matrix
- $\underline{\mathbf{E}}$ Three-dimensional residual array
- \mathbf{E} Unfolded two-dimensional residual data matrix

2.1 Fault detection and classification

For real semiconductor fabrication facilities (fabs), process control and monitoring are necessary to ensure yield, and thus, profitability of large investments. There are four process monitoring procedures: fault detection, fault identification, fault diagnosis, and process recovery (Raich and Cinar 1996). Fault detection determines whether a fault has occurred in the process. Early detection gives engineers more time to perform appropriately to avoid serious equipment abnormality. Fault identification identifies the main effects on observation variables and concentrates on the process variables most relevant to diagnosing abnormalities. Fault diagnosis determines which fault has occurred, that is, the cause of the observed out-of-control status. Fault diagnosis determines the type, location, magnitude, and time of a fault (Isermann 1995). Process recovery removes the cause of the fault to reduce yield loss.

MSPC is widely used for chemical, biotechnical, polymer, and pharmaceutical applications. MSPC methods such as PCA and PLS have been used for many applications. Skagerberg et al. (1992) used PLS to predict polymer properties from measured temperature profiles in a tubular low-density polyethylene reactor and to interpret process behavior. MacGregor et al. (1994) developed a multiblock PLS method to detect and diagnose process faults. Wikströma et al. (1998) applied the multivariate process monitoring charts including a multivariate Shewart control chart, multivariate CUSUM control chart, and multivariate EWMA control chart in an electrolysis process. Ralson et al. (2001) developed a PCA model for process monitoring and fault diagnosis in chemical processes.

Several studies have examined fault detection and diagnosis applications in batch process monitoring. Nomikos and MacGregor (1994) proposed the MPCA method to monitor the batch process in a chemistry process. The MPCA method unfolds a 3D data matrix into a two-dimensional (2D) matrix and then performs PCA to explain variability among batches, process variables, and time. Wise et al. (1999) compared conventional PCA, MPCA, trilinear decomposition, and parallel factor analysis for fault detection in the semiconductor etching process. Yue et al. (2000) introduced batch process monitoring to semiconductor fabrication for plasma etchers by using emission spectra and the MPCA method to analyze multiple scan sensitivity within a wafer for several typical faults. Wise and Gallagher (1996) also used the MPCA and MSPC methods in process monitoring and fault detection for a chemistry process. Spitzlspurger et al. (2005) presented an adaptive Hotelling T^2 control chart for the semiconductor etching process. In addition to statistical methods, the k-nearest neighbor (kNN) method based on distance has also been used

to classify faults in semiconductor manufacturing (He and Wang 2007; Verdier and Ferreira 2011).

2.2 Principal component analysis

PCA is a multivariate method that explains the covariance structure of multivariate data using a few linear combinations of the original variables (Wold et al. 1987a; Wise and Gallagher 1996). PCA decomposes original data matrix \mathbf{X} , including i measurements and j process variables, into uncorrelated t scores matrix \mathbf{T} by an orthogonal loading matrix, \mathbf{P} , and the unexplained variation of \mathbf{X} , that is, residual matrix \mathbf{E} .

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}. \quad (1)$$

Most variations can be explained by the first r PC matrix \mathbf{TP}^T . Given new measurement \mathbf{X} , and then PC score \mathbf{t} , prediction of measurement $\hat{\mathbf{X}}$, and residual matrix $\hat{\mathbf{E}}$ are defined as follows:

$$\mathbf{t} = \mathbf{P}^T \mathbf{X} \quad (2)$$

$$\hat{\mathbf{X}} = \mathbf{PP}^T \mathbf{X} \quad (3)$$

$$\hat{\mathbf{E}} = (\mathbf{I} - \mathbf{PP}^T) \mathbf{X}. \quad (4)$$

After building the PCA model, original data variation can be detected by score space and residual space. The Hotelling T^2 statistic and squared prediction error (SPE) are used to measure the divergence of a new measurement. The Hotelling T^2 statistic is first used to estimate the explained variation of a new measurement by the PCA model in the score space. This is shown in (5).

$$T^2 = \mathbf{t}^T \mathbf{S}^{-1} \mathbf{t}. \quad (5)$$

The matrix \mathbf{S}^{-1} is the diagonal matrix containing the inverse eigenvalues associated with the r eigenvectors (PCs). The T^2 statistic assumes an F distribution. SPE is also used to estimate the unexplained variation of a new measurement in residual space and is defined as

$$SPE = \hat{\mathbf{E}}^T \hat{\mathbf{E}} = \mathbf{X}^T (\mathbf{I} - \mathbf{PP}^T) \mathbf{X} \quad (6)$$

where SPE represents the square error of residual. Therefore, PCA performs process monitoring by constructing a control chart for these two statistics.

3 Proposed approach

With advanced information technology and sensors for data collection, real-time tool data can be recorded for tool or process monitoring in advanced 300 mm fabrication. All identification information and equipment parameters are collected as wafers pass through the process. All process data should be considered multivariable and not mutually independent because they may be correlated with

each other. This study proposes a manufacturing intelligence approach for fault detection, diagnosis, and classification based on MPCA and data mining approaches, as shown in Fig. 1. After problem definition and data preparation, MPCA unfolds the 3D process data and reduces the batch process data dimensions into a few PCs. In fault detection, the D and Q statistics are calculated to detect abnormal events from the score space and residual space, respectively. Different types of process faults can be detected based on the constructed control limits of both statistics. All data points are then clustered using a self-organizing map network. The set of most relevant process variables to the detection of abnormal events are also identified. Decision trees are then used to extract rules from the generated groups to describe the various faults in the process. The generated simple rules are used to predict the wafer class defined by performance.

3.1 Problem definition

Advanced sensors and information technologies enable process data and tool parameters to be recorded for fault detection and diagnosis. Process data can be recorded immediately during operation in a semiconductor fab. Historical data including nominal process information and quantitative measurements are continuously recorded at

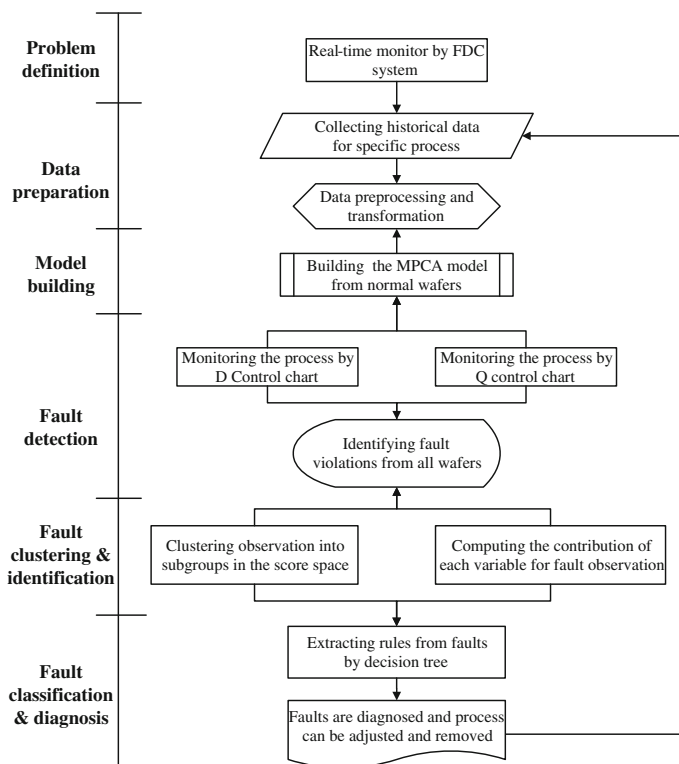


Fig. 1 Research framework for semiconductor FDC

nine sites per wafer. To reduce yield loss during the manufacturing process, tool abnormalities should be detected early during process monitoring. The MPCA unfolds the data and extracts a few PCs for process monitoring by considering the 3D data array in semiconductor manufacturing, wafers, process measurement variables, and samples recorded by time. If an abnormal wafer is detected, the fault type must be identified by classifying several subgroups for quick diagnosis and process recovery. This study also uses data mining to extract valuable process information and manufacturing intelligence to support fault detection and classification.

3.2 Data preparation

The FDC system collects a large amount of historical data including missing, noisy, and redundant data. Therefore, the data must be prepared to improve data quality for effective model construction (Chien et al. 2007). Because the FDC system records historical data every second, this increases the quantity of data and noise, which both influence the effectiveness of the constructed model. Domain knowledge is used to select the key process step in the temporal window. Historical data features are then extracted to detect faults and generate fault classes. Equipment parameters with constant values are removed. Before performing the MPCA method, data should be transformed to zero mean and unit variance. For each process variable, subtract its sample mean and divide it by its standard deviation. The purpose of standardization is to avoid particular variables dominating the model results due to the bias of data scale.

3.3 MPCA model construction

MPCA unfolds the 3D data matrix into a 2D matrix and then performs PCA (Wold et al. 1987b; Nomikos and MacGregor 1994). For batch process monitoring in a semiconductor fab, a set of process variables is measured with time intervals and wafers. All historical data are arranged into 3D array $\underline{\mathbf{X}}$ ($I \times J \times K$), where I is the number of wafers, J is the number of variables, and K is the number of times for each wafer measurement. After unfolding 3D matrix $\underline{\mathbf{X}}$ into 2D matrix \mathbf{X} ($I \times JK$), ordinary PCA is performed on matrix \mathbf{X} . The objective of MPCA is to decompose 3D matrix $\underline{\mathbf{X}}$ into fewer PCs. As shown in Fig. 2, $\underline{\mathbf{X}}$ is decomposed into the score space ($\sum_{r=1}^R \mathbf{t}_r \mathbf{p}_r^T$) with first R PCs and residual space ($\underline{\mathbf{E}}$). \mathbf{t}_r and \mathbf{p}_r represent a score vector and the r th row vector of the loading matrix, respectively.

$$\begin{aligned} \underline{\mathbf{X}} &= \mathbf{TP}^T + \underline{\mathbf{E}} = \sum_{r=1}^R \mathbf{t}_r \mathbf{p}_r^T + \underline{\mathbf{E}} \\ &= \hat{\mathbf{X}} + \underline{\mathbf{E}} \end{aligned} \quad (7)$$

\mathbf{P} is the loading matrix, which includes loading vector \mathbf{p}_r , and \mathbf{T} is the score matrix containing the projection location on the PC subspace. $\underline{\mathbf{E}}$ is the variation that the MPCA model cannot explain.

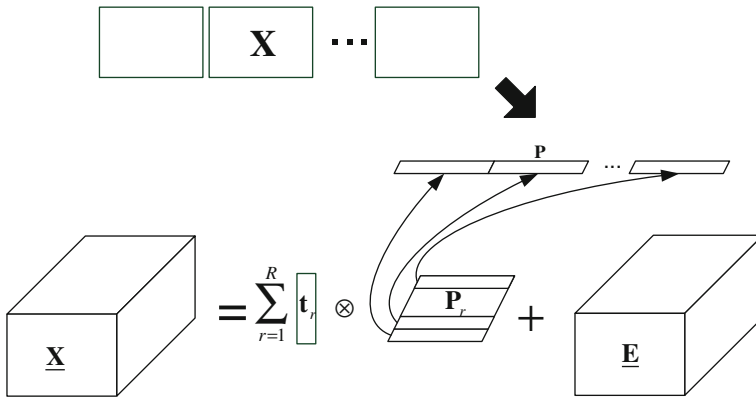


Fig. 2 Decomposed parts of the three-dimensional array

3.4 Fault detection

MPCA decomposes normal wafer variation into model space and residual space by using the D and Q statistics to monitor the multivariate batch process. The D statistic is similar to the Hotelling T^2 , that is, the Mahalanobis distance between the new wafer data and the normal condition in a score space that is formed by fewer PCs. The D statistic monitors systematic variation in the score space. The Q statistic monitors variation that is not explained by the PCs retained by the MPCA model. The D and Q statistics are calculated using the following equations:

$$D_i = \mathbf{t}_i^T \mathbf{S}_R^{-1} \mathbf{t}_i \times \frac{I(I - R)}{R(I^2 - 1)} \times F_{(R, I-R)} \tag{8}$$

$$Q_i = \mathbf{e}_i \mathbf{e}_i^T, \tag{9}$$

where R is the number of PCs retained by MPCA; \mathbf{t}_i is the R score vector corresponding to wafer I ; \mathbf{S}_R is the covariance matrix ($R \times R$) of the t scores calculated by the MPCA model; the diagonal units in the matrix are the variance of first R PCs; and \mathbf{e}_i is residual vector i of residual matrix \mathbf{E} . Therefore, original data variability can be detected in the model space and residual space.

The reference wafers determine the control limits for the D and Q statistics under normal conditions. Based on significance level α determined by the user, the D and Q statistic control limits are determined as follows (Jackson and Mudholkar 1979; Nomikos 1996):

$$D_{\alpha, \text{limit}} = F_{R, I-R} \tag{10}$$

$$Q_{\alpha, \text{limit}} = \theta_1 \left[1 - \frac{\theta_2 h_0 (1 - h_0)}{\theta_1^2} + \frac{z_\alpha (2\theta_2 h_0^2)^{1/2}}{\theta_1} \right]^{1/h_0}, \tag{11}$$

where $h_0 = 1 - 2\theta_1\theta_3/3\theta_2^2$, $\theta_1 = \text{trace}(V)$, $\theta_2 = \text{trace}(V^2)$, $\theta_3 = \text{trace}(V^3)$, and V is the covariance matrix of \mathbf{E} , $V = \mathbf{E}\mathbf{E}'/(\mathbf{I} - 1)$, and z_α is a critical value of standard normal distribution at significance level α .

3.5 Faults clustering and identification

Once the out-of control signal has been detected by the D or Q control charts, the cause of the out-of-control events must be identified. An out-of-control signal indicates abnormal behavior in equipment or processes. Violations in the D and Q control charts are from two types of process faults. The first is because of the systematic variation on t -score, and the other is because of residual space variation that cannot be explained by the MPCA model. Clustering analysis groups the observed wafers to investigate them under normal and abnormal process conditions with possible faults. Clustering analysis segments a diverse group with similar observation characteristics. For fault clustering, self-organizing map (SOM) neural network is used to cluster observations based on PC similarity. The SOM is better than other classical clustering analysis methods because it produces graphic visualizations and parallels the architecture of a neural network. The 2D topological map shows group results based on the relationships among observations.

Fault identification is difficult for process engineers or operators because many process variables and equipment parameters must be monitored. Additionally, abnormal faults could lead to a difference in several process variables. The aim of fault identification is to identify process variables that are most important for judging the out-of-control signal. A contribution plot can be used to show the contribution of the j th variable to the D and Q statistics. To compare the quantity contribution of the process variables, process variables with relatively large contributions should be diagnosed further.

3.6 Fault classification and diagnosis

Fault classification classifies normal and abnormal clusters by considering process variables that are most relevant to the faults. The decision tree method is used to extract rules to explain the relationship between the wafer groups and process variables. Because of the nature of the problem and the data characteristics, a Chi-squared Automatic Interaction Detector (CHAID) is used (Kass 1980). The tree is grown iteratively until all attributes in the decision tree model are not significantly different or the number of instances in the node is less than a determined threshold. Decision trees are better at exploring data for classification and extracting understandable decision rules for process engineers than other classification techniques. All extracted information can assist with fault diagnosis, and the process deviation can be corrected quickly.

4 Empirical study

4.1 Problem definition

To validate the proposed approach, an empirical study was conducted for the CVD Ti/TiN process in a leading semiconductor company in Taiwan. CVD is a chemical process used to develop thin films in ICs. The abnormal event is peeling caused by

an imbalanced link between the film and substrate. In practice, the peeling is related to uniformity of thickness, which is inspected by scanning electron microscope analysis after several process steps. Therefore, equipment sensors record equipment parameters to determine whether the tool condition is normal.

4.2 Data preparation

Total 100 wafers with 21 process variables were collected during CVD process. It was possible to collect 23 records from the main CVD process step for each wafer. There were 68 normal wafers and 32 abnormal wafers. A normal wafer has no peeling faults. Additionally, four variables were excluded because of constant values that did not provide process information. Therefore, 17 process variables were selected to build the monitoring model. Table 1 lists these variables. All data were standardized with zero mean and unit variance for further MPCA modeling.

The trend charts of each process variable are shown in Fig. 3. Identifying the process variable that is the main cause of the peeling fault is difficult. The variation among the 17 process variables is different, and some variables are correlated with each other. The variability among wafers, process variables, and recorded time cannot be explained by conventional PCA.

4.3 MPCA model construction

To characterize equipment behavior under normal conditions, 68 normal wafers were used to construct a reference model for equipment monitoring. First, MPCA unfolded original 3D array $\underline{\mathbf{X}}$ ($68 \times 17 \times 23$) into 2D data set \mathbf{X} (68×391), where the row represents wafer i and the column represents process variable j at time k . MPCA was then used to transform the original data space into score space and residual space. The score space accumulates the variation under first r PCs, and residual space is the variation that the model cannot explain. The maximal number of PCs is 68, and the explained variance percentages for each PC are listed in Fig. 4.

Table 1 Process variable description

Process variable	Description	Process variable	Description
V1	Chamber pressure	V10	RF load position
V2	Gas1 flow	V11	RF tune position
V3	Gas3 flow	V12	RF Vpp
V4	Gas4 flow	V13	RF Vdc
V5	Stage heater temperature	V14	Shower head temperature
V6	Stage heater power_in	V15	Shower head power
V7	Stage heater power_out	V16	Stage heater resistance in
V8	APC angle	V17	Stage heater resistance out
V9	RF power		

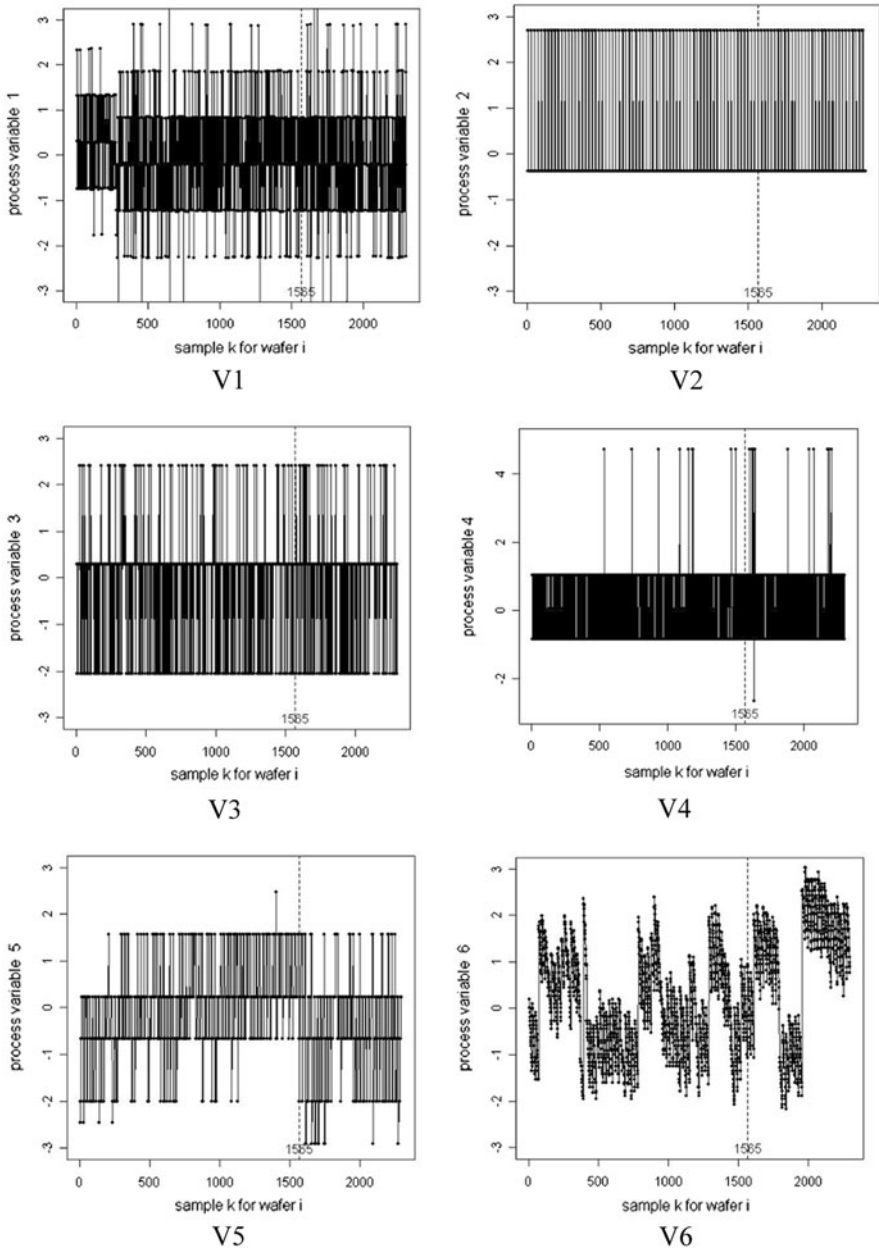


Fig. 3 Trend charts for process variables

The explained variation for each PC decreases as the number of PCs increases. The number of PCs is selected if the proportion of explained variation is larger than 5 %. The first three PCs explain 50.87 % of the process variation.

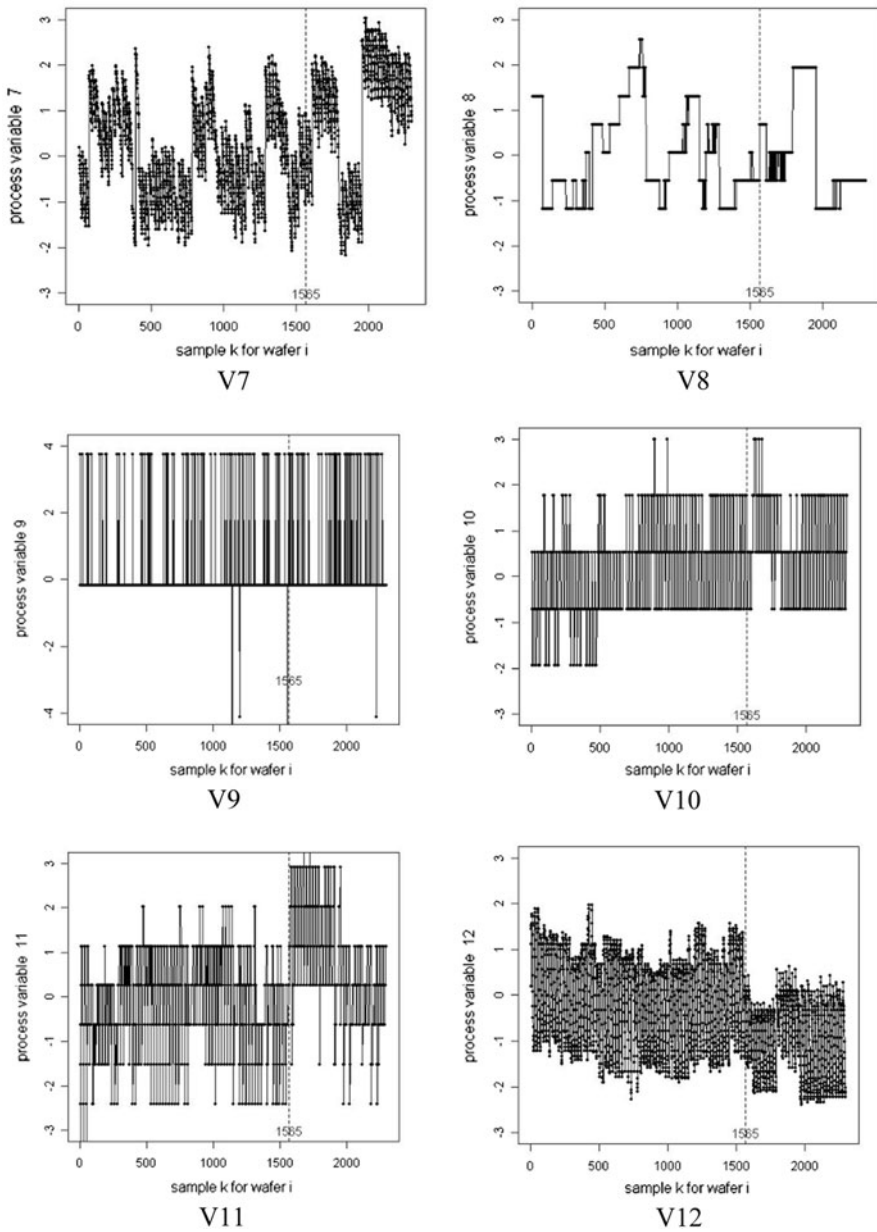


Fig. 3 continued

4.4 Fault detection

To evaluate the effectiveness of peeling fault detection, unfolded data matrix \mathbf{X} (100×391), including abnormal wafers (32×391), was decomposed into score

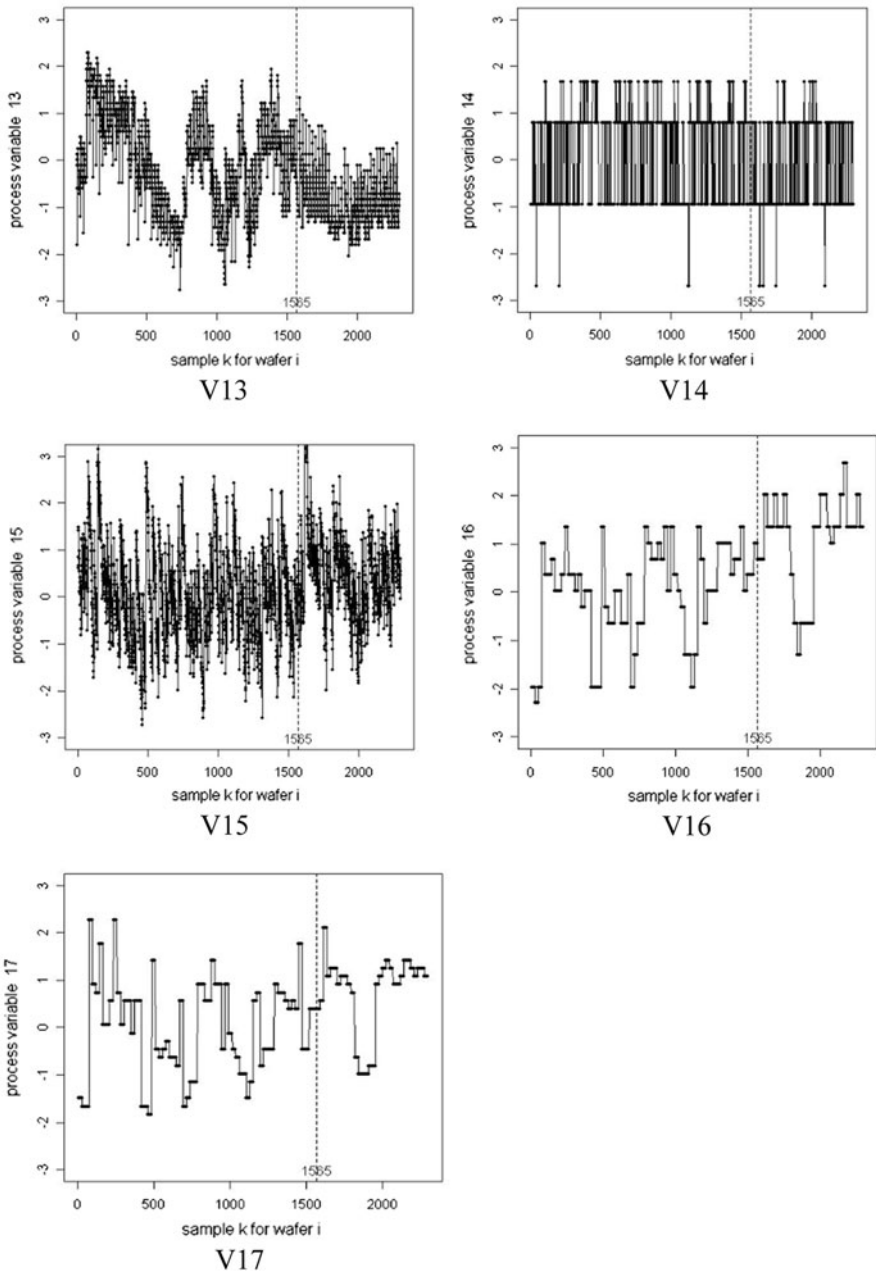


Fig. 3 continued

space and residual space multiplying by loading matrix $\mathbf{L}_{j=3}$ (391×3). The D statistic is the overall wafer performance to compare with normal wafer performance in the MPCA model. The Q statistic represents equipment behavior in the residual space and

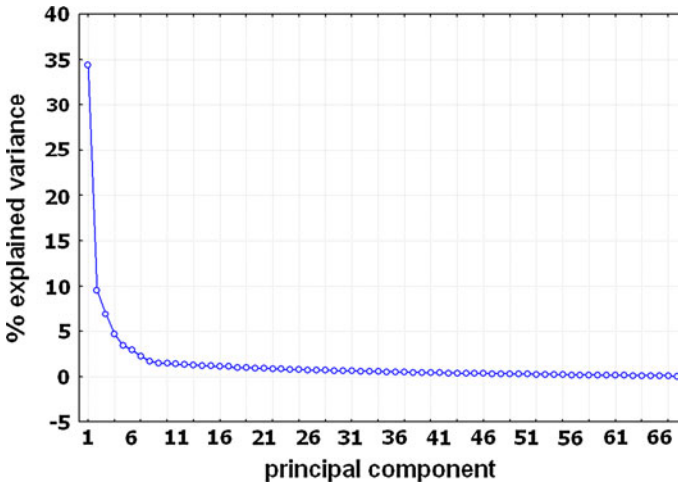


Fig. 4 Explained variation for each principal component

monitors each wafer. The D and Q statistic control charts identify out-of-control signals to monitor processes. They are shown in Figs. 5 and 6. The dashed line is the 95 % control limit, and the solid line is the 99 % control limit. In total, 98 % of the wafers are under the control limit of the D statistic. This shows that most wafers are similar in the score space. The normal and abnormal wafer detection results can be compared by the t-score scatter plot, as shown in Fig. 7, where a circle represents a wafer without the peeling fault and a solid dot represents a wafer with the peeling fault. Normal and abnormal wafers with prior knowledge of peeling faults do not separate significantly in the score space. This implies that the equipment behavior of abnormal wafers cannot be identified in the score space.

In the D statistic control chart, the 64th and 71st wafers are identified as abnormal wafers at the 99 % and 95 % control limits, respectively. Figure 7 shows that the 64th and 71st wafers are different to other wafers, especially in the second and third PCs. The result of the D control chart implies that differentiating between a peeling and a non-peeling wafer is difficult to use only score space. In the Q statistic control chart, several wafers exceed the control limit of the Q statistic. In total, 47 wafers were abnormal beyond the 99 % control limit, and 85 wafers were abnormal beyond the 95 % control limit. The MPCA model explains only 50.87 % of the variation in the process data, which led to the misclassification of 25 % of normal wafers by the Q statistic.

4.5 Fault clustering and identification

Once a deviation is detected in the D or Q statistic control charts, the wafer should be diagnosed to identify the cause of the abnormal event. The fault identification process locates the process variables most responsible for the abnormal event. In the D statistic control chart, the 64th and 71st wafers exceed the out-of-control limit. Figure 8 shows the contribution of each process variable to the D statistic. For the

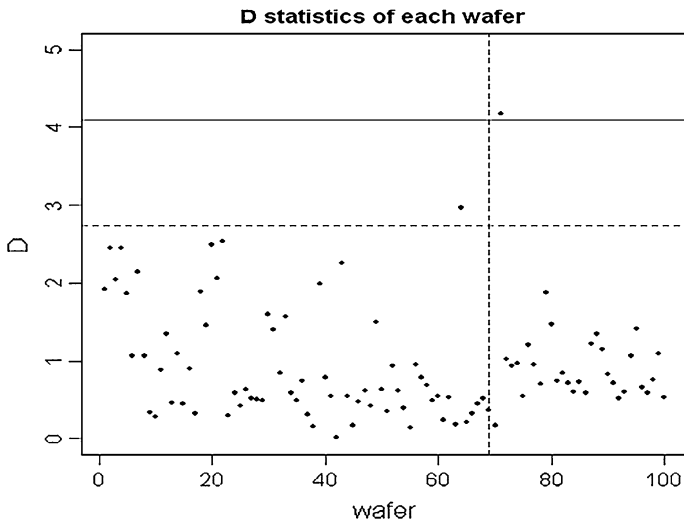


Fig. 5 D statistic control chart

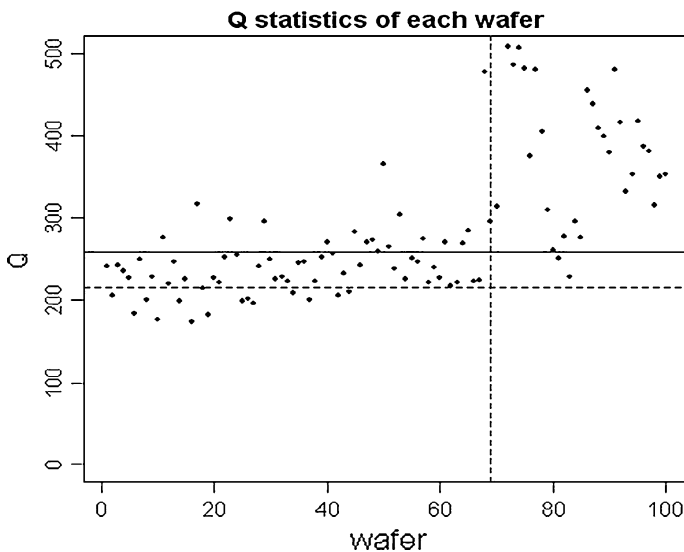


Fig. 6 Q statistic control chart

64th wafer, process variables V6, V7, V15, V16, and V17 contribute more than the other variables. Process variables V10, V16, and V17 contribute the most to the 71st wafer.

The wafers are grouped by SOM into three clusters based on the D and Q statistics. Table 2 lists the basic statistics for each cluster. The percentage of peeling wafers in Clusters (i) and (ii) is lower than 15 %, which is normal. Cluster (iii) is in the peeling wafer class because 91.67 % of wafers have a peeling problem.

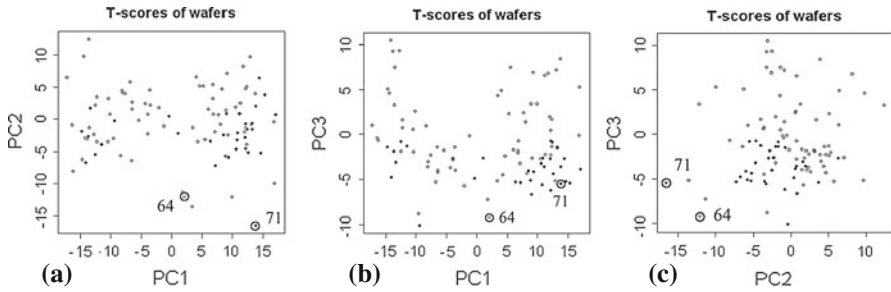


Fig. 7 T score scatter plots, **a** first PC and second PC, **b** first PC and third PC, **c** second PC and third PC

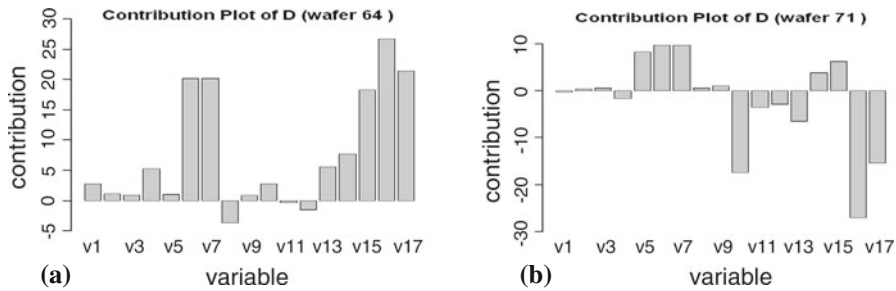


Fig. 8 *D* statistic contribution plot, **a** 64th wafer, **b** 71st wafer

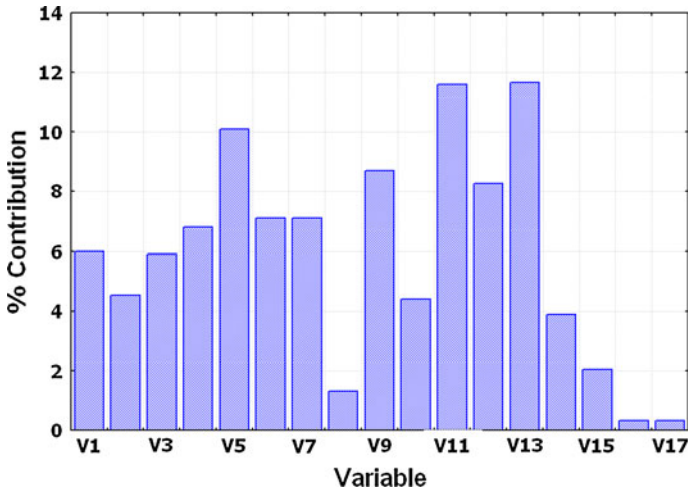
The peeling wafers in Cluster (iii) are selected to identify process variables associated with the abnormality. In the residual space, the *Q* statistic chart is used to monitor and investigate the MPCA model residuals. By averaging the process variable contributions of these selected abnormal events, process variables V13, V11, V5, V9, and V12 are first identified for fault diagnosis, as shown in Fig. 9.

4.6 Fault classification and diagnosis

After building the MPCA model for variation decomposition into score space and residual space, three groups of wafers were clustered using SOM method. A decision tree was then used to extract the relationship between peeling wafers and process variables. The target variable is a cluster result called “1” if the wafer is peeling, and “0” if it is not. Figure 9 shows the five variables identified from the contributing process variables set, V5, V9, V11, V12, and V13, based on the results from Cluster (iii). The average values for each process variable (100×17) were used as independent variables. Figure 10 shows the peeling fault rules extracted by the decision tree. The overall model fit is 97 % for V12 and V11, and three rules were extracted for classifying the peeling fault. The first rule is, “If the value of variable V12 is less than 0.8, then the wafer has the peeling issue.” The second rule is, “If the value of variable V12 is greater than 0.6 and the value of variable V11 is less than 1, then the wafer does not have the peeling issue.” The third rule is, “If the

Table 2 Basic statistics for each cluster

	Cluster (i)	Cluster (ii)	Cluster (iii)
Number of wafers	14	62	24
Peeling wafers	1	9	22
Percentage of peeling wafers in the cluster (%)	7.14	14.51	91.67

**Fig. 9** Average Q contribution plot for Cluster (iii)

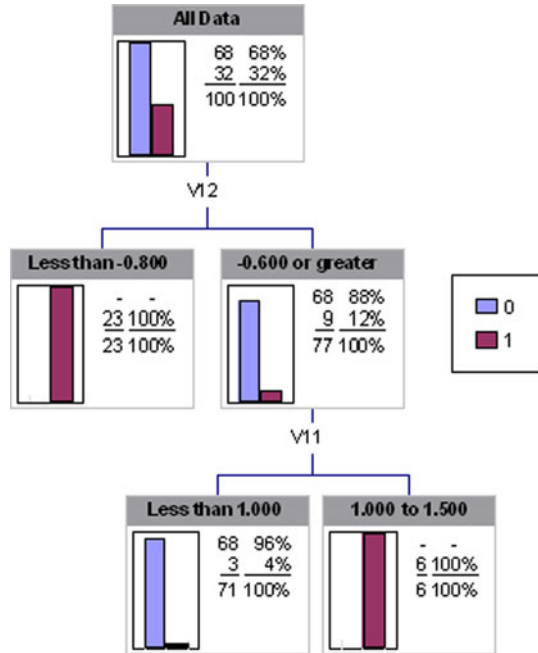
value of variable V12 is greater than 0.6 and the value of variable V11 is between 1 and 1.5, then the wafer has the peeling issue.” Based on the classification results, 29 peeling wafers were detected, and variables “RF Vpp” and “RF tune position” were extracted for a process engineer to diagnose the cause of the peeling issue.

4.7 Results and discussion

The peeling and non-peeling wafers can be identified using three rules. After detecting abnormal wafers, a set of process variables with large contributions to the D or Q statistics were identified by the fault identification procedure. Process variables V5 (Stage Heater Temperature), V9 (RF Power), V11 (RF Tune Position), V12 (RF Vpp), and V13 (RF Vdc) were regarded as the cause of the peeling issue. Variable V12 (RF Vpp) was extracted in the first level of decision tree and classified 23 peeling wafers, implying that it had the largest impact on classifying peeling wafers.

To evaluate the effectiveness and practical viability of the proposed approach, two conventional approaches including MPCA and the kNN method were selected for comparison. The kNN method categorizes an unlabeled wafer as normal or

Fig. 10 Decision tree for classifying peeling wafers



abnormal according to the k -nearest similar wafers in the training data. First, the number of k NNs for each sample is determined in the normal wafers. The sum of the k -smallest squared Euclidean distances from the i th wafer to the j -nearest wafer is then calculated. The number of k NNs was set as five based on domain knowledge and data quality. The fault detection threshold was estimated at the 95 % and 99 % confidence intervals from the 68 normal wafers. Figure 11 shows that the k NN method detected 24 abnormal wafers at the 99 % control limit, and only 1 normal wafer was identified as abnormal.

Table 3 shows a comparison of the three methods by sensitivity and specificity. Sensitivity measures the proportion of peeling wafers that were correctly classified into the peeling wafer group. Specificity measures the proportion of non-peeling wafers that were correctly classified into the non-peeling wafer group. The control limit was set under the 99 % confidence level. MPCA using the D and Q statistic control charts has the large sensitivity (93.75 %) and small specificity (77.94 %), respectively. The k NN method has the small sensitivity (75 %) and the large specificity (98.53 %). The proposed method has the large sensitivity (90.63 %) and the large specificity (100 %). While the results indicate that the proposed method effectively detects abnormal wafers, it is difficult to show significant differences among these methods because of the small number of wafers used in the study. Using a larger data set from a semiconductor company to reproduce this study should produce a more robust comparison and results.

Comparing the fault detection and classification results shows that the accuracy of the MPCA method for classifying the peeling issue is low and many non-peeling wafers exceeded the control limit of Q . Because only 50.87 % of the variation is

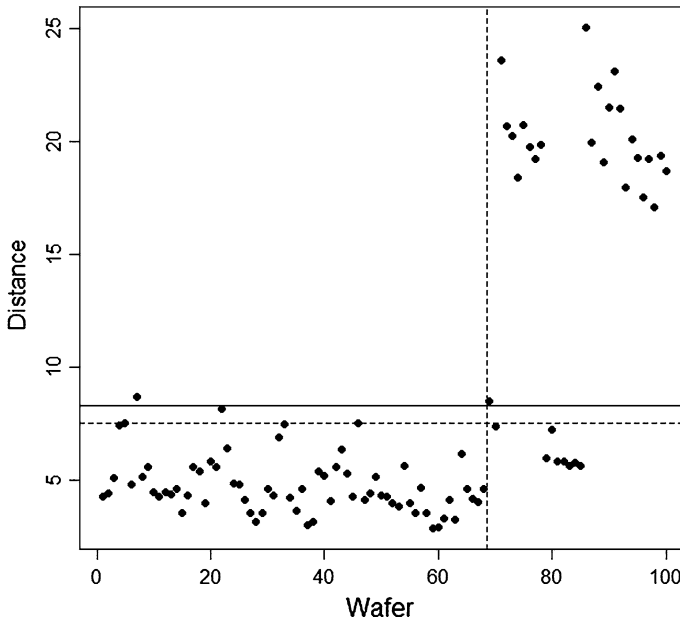


Fig. 11 Fault detection using the kNN method

Table 3 Comparison results for fault detection

	Sensitivity (%)	Specificity (%)
MPCA approach	93.75 (30/32)	77.94 (53/68)
kNN approach	75.00 (24/32)	98.53 (67/68)
Proposed approach	90.63 (29/32)	100.00 (68/68)

explained in the score space, the remaining 49.13 % of the variation remains in the residual space. This results in many wafers that exceed the Q control limit and produces many false alarms. However, most peeling wafers were classified by clustering analysis and the decision tree method. This means that peeling wafers can be detected by the average process variable performance, and the data fluctuation in process time should be smoothed via data preparation for constructing the monitoring model.

5 Conclusion

This study develops an effective approach for semiconductor fault detection and classification. It integrates MSPC methods and data mining approaches for manufacturing intelligence and yield enhancement. The proposed approach detects potential process faults or abnormal events effectively by monitoring fewer key variables than conventional approaches. The results demonstrate the practical

viability of the proposed approach. The MPCA method unfolds 3D historical data and projects the data onto the score space and residual space. Abnormal wafers are then detected based on the D and Q statistic control charts. The process variables most critical to abnormal events are then identified by SOM analysis and contribution plots. Understandable rules are extracted to classify the normal and abnormal wafer classes. With the information extracted from the historical process data, diagnosis of abnormal events can be performed systematically and effectively. The decision to remove the abnormal events and reset the process to normal conditions can be made by engineers or domain experts in a timely and precise manner. Moreover, simple and understandable rules are generated to predict wafer performance and classify wafers into different classes using classification approaches. The proposed approach has been implemented in the FDC system of the company for in situ real-time monitoring for yield enhancement and quality control.

In the semiconductor manufacturing process, some recipes and products result in different behaviors in the same equipment and processes. The effects between change of recipe or product changes and equipment abnormalities are not considered in this study. To detect faults while accounting for different recipes in the same monitoring model, recipe effects must be eliminated before model construction. Further studies can develop an integrated model for fault detection by considering recipe changes and different products. Most existing methods consider the reference model to compare different conditions by using a group of normal measurements. However, equipment conditions are not fixed because of the complex processes and varied products in fab. Further research should develop a rolling scheme for adaptive monitoring and equipment diagnostics and prognostics.

Acknowledgments This research is supported by National Science Council, Taiwan (NSC 99-2221-E-007-047-MY3; NSC 100-2410-H-155-048), and the Ministry of Education (101N2073E1).

References

- Chen L-F, Chien C-F (2011) Manufacturing intelligence for class prediction and rule generation to support human capital decisions for high-tech industries. *Flex Serv Manuf J* 23(3):263–289
- Chien C-F, Hsu C (2011) UNISON analysis to model and reduce step-and-scan overlay errors for semiconductor manufacturing. *J Intell Manuf* 22(3):399–412
- Chien C-F, Wang W, Cheng J (2007) Data mining for yield enhancement in semiconductor manufacturing and an empirical study. *Expert Syst Appl* 33(1):192–198
- Chien C-F, Chen Y-J, Peng J-T (2010) Manufacturing intelligence for semiconductor demand forecast based on technology diffusion and production life cycle. *Int J Prod Econ* 128(2):496–509
- Chien C-F, Hsu C, Hsiao C (2011) Manufacturing intelligence to forecast and reduce semiconductor cycle time. *J Intell Manuf*. doi:10.1007/s10845-011-0572-y
- He QP, Wang J (2007) Fault detection using the k-nearest neighbor rule for semiconductor manufacturing process. *IEEE Trans Semicond Manuf* 20(4):345–354
- Isermann R (1995) Model based fault detection and diagnosis methods. *Proc Am Contr Conf* 3:1605–1609
- Jackson JE, Mudholkar GS (1979) Control procedures for residuals associated with principal component analysis. *Technometrics* 21(3):341–349
- Kass GV (1980) An exploratory technique for investigating large quantities of categorical data. *Appl Stat* 29(2):119–127

- Kourti T, MacGregor JF (1996) Multivariate SPC methods for process and product monitoring. *J Qual Technol* 28(4):409–428
- Kuo C, Chien C-F, Chen C (2010) Manufacturing intelligence to exploit the value of production and tool data to reduce cycle time. *IEEE Trans Autom Sci Eng* 8(1):103–111
- MacGregor JF, Jaccckle C, Kipariisdes C, Kourti M (1994) Process monitoring and diagnosis by multiblock PLS methods. *AIChE J* 40(5):826–838
- Montgomery DC (2005) Introduction to statistical quality control, 5th edn. John Wiley & Sons, Inc., New York
- Nomikos P (1996) Detection and diagnose of abnormal batch operations based on multi-way principal component analysis. *ISA Trans* 35(3):259–266
- Nomikos P, MacGregor JF (1994) Monitoring the processes using multiway principal component analysis. *AIChE J* 40(8):1361–1375
- Raich A, Cinar A (1996) Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. *AIChE J* 42(4):905–1009
- Ralson P, DePuy G, Graham JH (2001) Computer-based monitoring and fault diagnosis: a chemical process case study. *ISA Trans* 40(1):85–98
- Skagerberg B, MacGregor JF, Kipariisides C (1992) Multivariate data analysis applied to low-density polyethylene reactors. *Chemom Intell Lab Syst* 14(1–3):341–356
- Spitzlspurger G, Schmidt C, Ernst G, Strasser H, Speil M (2005) Fault detection for a via etch process using adaptive multivariate methods. *IEEE Trans Semicond Manuf* 18(4):528–533
- Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN (2003) A review of process fault detection and diagnosis part I: quantitative model-based methods. *Comput Chem Eng* 27(3):293–311
- Verdier G, Ferreira A (2011) Adaptive Mahalanobis distance and k-nearest neighbor rule for fault detection in semiconductor manufacturing. *IEEE Trans Semicond Manuf* 24(1):59–68
- Wikströma C, Albano C, Eriksson L, Fridén H, Johansson E, Nordahl A, Rännar S, Sandberg M, Kettaneh-Wold N, Wold S (1998) Multivariate process and quality monitoring applied to an electrolysis process part I. Process supervision with multivariate control charts. *Chemom Intell Lab Syst* 42(1–2):221–231
- Wise BM, Gallagher NB (1996) The process chemometrics approach to process monitoring and fault detection. *J Process Contr* 6(6):329–348
- Wise BM, Gallagher NB, Bulter SW, White DD Jr, Barna GG (1999) A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. *J Chemom* 13(3–4):379–396
- Wold S, Esbensen K, Geladi P (1987a) Principal components analysis. *Chemom Intell Lab Syst* 2(1–3):37–52
- Wold S, Geladi P, Esbensen K, Ohman J (1987b) Multi-way principal components and PLS analysis. *J Chemom* 1(1):41–56
- Yue HH, Qin SJ, Markle RJ, Nauert C, Gatto M (2000) Fault detection of plasma etchers using optical emission spectra. *IEEE Trans Semicond Manuf* 13(3):374–385

Author Biographies

Chen-Fu Chien is a Tsing Hua Distinguished Professor and Vice President and Chief of Staff in National Tsing Hua University (NTHU). He received his BS degree (with Phi Tao Phi Honor) with double majors in Industrial Engineering and Electrical Engineering from NTHU in 1990. He received his MS degree in Industrial Engineering and PhD degree in Operations Research and Decision Sciences from UW-Madison in 1994 and 1996, respectively. He was a Fulbright Scholar in UC Berkeley from 2002 to 2003 and also received the PCMPCL Training in Harvard Business School in 2007. From 2005 to 2008, he had been on-leave as the Deputy Director of Industrial Engineering Division in Taiwan Semiconductor Manufacturing Company (TSMC) that is the world largest semiconductor foundry. His research efforts center on decision analysis, modeling and analysis for semiconductor manufacturing, manufacturing strategy, and data mining. Dr. Chien has received seven invention patents on semiconductor manufacturing and published three books, more than 100 journal papers, and a number of case studies in Harvard Business School. He has been invited to give keynote speech in various conferences, including IEEM, APIEMS, C&IE, IML, KES, and leading universities worldwide. Dr. Chien received the National Quality Award from the Executive Yuan, twice Distinguished Research Awards and Tier 1 Principal Investigator (Top 3%) from

NSC, Distinguished University-Industry Collaborative Research Award from the Ministry of Education, University Industrial Contribution Awards from the Ministry of Economic Affairs, Distinguished University-Industry Collaborative Research Award and Distinguished Young Faculty Research Award from NTHU, Distinguished Young Industrial Engineer Award, Best IE Paper Award, and the IE Award from Chinese Institute of Industrial Engineering (CIIE), Best Engineering Paper Award, and Distinguished Engineering Professor by Chinese Institute of Engineers in Taiwan. He is now an Associate Editor of *IEEE Transactions on Automation Science and Engineering*, Area Editor of *Flexible Services and Manufacturing Journal*, Advisory Board Member of *OR Spectrum*, and on the editorial board of a number of TSSCI/EI journals

Chia-Yu Hsu is an Assistant Professor in Yuan Ze University (YZU). He received his BS degree in Statistics from National Cheng Kung University in 2002. He received his MS and PhD degrees in Industrial Engineering and Engineering Management from National Tsing Hua University in 2004 and 2009, respectively. His research interests include manufacturing intelligence, data mining, yield enhancement, design for manufacturability, and APC/AEC. His research works have appeared in *Computers & Industrial Engineering*, *Flexible Services and Manufacturing Journal*, *Journal of the Chinese Institute of Industrial Engineers*, *Journal of Intelligent Manufacturing*, and *Journal of Quality*. Dr Hsu has received two invention patents on semiconductor manufacturing. He has also served as a Guest Editor for *Journal of Intelligent Manufacturing*.

Pei-Nong Chen received his MS degree in Industrial Engineering and Engineering Management at the National Tsing Hua University in 2005. His research interests include data mining, multivariate data analysis, and yield enhancement.