201st Meeting of the Electrochemical Society, International Symposium on Plasma Processing XIV, Abs. 413, Philadelphia, PA, May 2002.

SIMULTANEOUS FAULT DETECTION AND CLASSIFICATION FOR SEMICONDUCTOR MANUFACTURING TOOLS

Brian E. Goodlin, Duane S. Boning, Herbert H. Sawin

Massachusetts Institute of Technology

Cambridge, MA 02139, USA

Barry M. Wise

Eigenvector Research, Inc.

Manson, WA 98831, USA

Increasingly there is a need for fast, accurate, and sensitive detection of equipment and process faults to maintain high process yields and rapid fault classification (diagnosis) of the cause to minimize tool downtime in semiconductor manufacturing. Current methods treat fault detection and classification as a two-step process. We present a novel method to simultaneously detect and classify faults in a single-step using fault-specific control charts. These control charts are designed to discriminate between specific fault classes and the normal process operation as well as all other fault classes. Using a set of experimental data collected from an industrial plasma etcher, we demonstrate that if the fault-specific charts are constructed using an orthogonal linear discriminant approach, they are effective in simultaneously detecting and classifying a given fault. We also demonstrate that this methodology has improved sensitivity for detection of faults when compared to other commonly used methods of fault detection.

INTRODUCTION

State-of-the-art semiconductor processes are often pushed to the limits of the current technology, resulting in processes that have little or no margin for error. Increasingly there is a need for fast, accurate, and sensitive detection and classification of equipment and process faults to maintain high process yields and high throughput in manufacturing. Current methods treat fault detection and classification (FDC) as a two-step process: 1) first detect the fault, then 2) classify the cause of the fault and apply corrective action. In this paper we propose a method for combining these two steps for simultaneous detection and classification of specific faults.

Detection of process and tool faults in the shortest time possible is critical to minimize scrap wafers and improve product yields for semiconductor manufacturing. Traditionally, there are two ways to detect these faults. One method involves waiting for problems to arise with the end-of-the-line product and searching the various processes and/or equipment that were operating improperly. This method is slow and tedious and can lead to substantial losses in product. Alternatively, one can monitor wafer-state data collected after each manufacturing process step and look for faulty processing. For example, one can monitor etch rates, uniformity, critical dimension, etc. for plasma etching processes, and compare the results with a specified target and control limits using statistical process control (SPC). Since most of these measurements are done off-line and less frequently than every wafer, this can still lead to a large number of scrapped wafers before a fault is detected. Additionally, some faults, e.g. device damage during plasma etching, could pass through these limited number of wafer-state measurements undetected.

More recently there has been a move towards fault detection directly on the manufacturing tool, through monitoring of tool-state data and in-situ process-state sensor data. Many manufacturing tools are beginning to have the ability to collect a large amount of data in real-time, which can be then accessed for tool fault detection. For example, a modern plasma etching tool will collect streaming data for power, pressure, flow rates, valve positions, match network positions, etc. in addition to data from in-situ sensors such as residual gas analyzers (RGAs), optical emission spectroscopy (OES), and RF impedance monitoring. Unfortunately, with such an abundant amount of data available, it is often difficult to extract useful information such as when the tool is no longer operating properly (i.e. detecting a fault).

Over the past 15 to 20 years, a number of analytical techniques have emerged to process large amounts of data and convert them to useful information. Several of these are starting to gain usage for tool fault detection. For example, Wise et al. [1,2] have compared a wide assortment of factor analysis techniques for use in fault detection of plasma etchers. Yue et al. [3] have also examined the use of factor techniques for fault detection. Bunkofske [4] has reported the use of Hotelling's T^2 for fault detection in a manufacturing setting. Ultimately, such techniques attempt to compress the multivariate data into a small number of latent variables or statistical parameters that can be monitored using fault detection control charts. Once the fault is detected, then the fault must be classified for some assignable cause, which ultimately must be corrected before the tool is returned to normal operation.

Classification of the cause of the fault is equally critical to detecting the fault, because rapid classification and corrective action will lead to minimized tool downtime and increased throughput. Traditionally in semiconductor manufacturing, after a fault is detected, the cause is identified in one of two ways. The first way is to use trial and error, inspecting everything that one believes might be associated with the fault. This method is both slow and tedious, but is unfortunately how much of the fault classification in semiconductor manufacturing is still done. The second way involves the exploration of tool, sensor, and process data to gain engineering intuition as to what are the most

probable causes of a fault. One might think of this as a primitive form of pattern recognition. This method is still somewhat slow since the engineer has to mull through a large amount of data, and relies heavily on empirical engineering understanding of the process.

Newer methods have focused on automating the fault classification with automatic pattern recognition. Bunkofske [4] has demonstrated how contribution plots to T^2 represent a specific pattern or fingerprint that can be used to help classify and diagnose faults, by identifying the measured parameter(s) that most likely caused the failure. Ison et al. [5] describe the use of probabilistic models such as tree-based or generalized linear models for automatically classifying faults based on historical data. Love and Simaan [6] describe the use of a knowledge-based approach to fault diagnosis. May and Spanos [7] describe a methodology for automatically detected faults through knowledge-based and probabilistic modeling. In general, these methods try to identify some fault signature which can be compared to previous signatures from faults that have already been classified.

In this work, we take the automatic fault classification idea to the next level, by creating fault-specific control charts for detection of specific classes or types of faults. Once a fault is detected on any given chart, the cause is immediately identified. Further, by looking for a specific kind of fault, we can expect the sensitivity to that fault will be greater than would be expected if one were detecting for any generic fault. As a result, we would expect to detect and classify faults faster and more accurately, resulting in improved process yields and higher throughput. The potential downside to this method is that the false alarm rate may be higher for detecting faults if a large number of fault detection control charts need be created.

EXPERIMENTAL

Metal Etch Process

This project focused on an Aluminum stack etch process performed on the commercially available Lam 9600 plasma etch tool. The goal of this process is to etch the TiN/Al – 0.5% Cu/TiN/oxide stack with an inductively coupled BCl_3/Cl_2 plasma. The key parameters of interest are the line width of the etched Aluminum line (specifically the line width reduction in relation to the incoming resist line width), uniformity across the wafer and the oxide loss.

The standard recipe for the process consists of a series of six steps. The first two are for gas flow and pressure stabilization. Step 3 is a brief plasma ignition step. Step 4 is the main etch of the Al layer termination at the Al endpoint, with Step 5 acting as the over-etch for the underlying TiN and oxide layers. Note that this is a single chemistry etch process, i.e. the process chemistry is identical during steps 3 through 5. Step 6 vents the chamber.

Tool-state Data

The current work utilized machine tool data collected directly on the LAM 9600 etcher, which consisted of 19 variables as shown in Table 1. Data were collected and recorded at one second intervals during the etch for each of these sensors. Since our primary concern in this work was to detect faults occurring from one wafer to the next, we took the average value of each variable during the etch process for each wafer, resulting in a 1x19 array of values for each wafer. We could then look for significant differences between the wafer averages for fault detection. Note that Wise et al. [1,2] have explored methods to compare variable trajectories between different wafers as opposed to just comparing the wafer averages. These methods are not explored in the current work.

1	BCl3 Flow	11	RF Power
2	Cl2 Flow	12	RF Impedance
3	RF Bottom Power	13	TCP Tuner
4	RF Bottom Ref Pwr	14	TCP Phase Error
5	Endpoint A Detector	15	TCP Impedance
6	Helium Pressure	16	TCP Top Power
7	Chamber Pressure	17	TCP Reflected Power
8	RF Tuner	18	TCP Load
9	RF Load	19	Vat Valve Position
10	Phase Error		

Table 1. Tool-state variables used for process monitoring

Induced Fault Experiments

A series of three experiments (experiments 29, 31, and 33) were performed where faults were intentionally induced by changing the TCP power, RF power, pressure, Cl_2 or BCl_3 flow rate, and He chuck pressure on selected wafers. The three experiments consisted of a total of 129 wafers: 107 wafers were normally processed, 21 wafers were processed with induced faults, and one wafer was misprocessed (outlier). The 21 induced faults shown in Table 2 were classified into one of six classes depending on what parameter was varied. The normal process operation is proprietary, but the deviation from the normal process is indicated for the individual fault experiments in the table. The fault parameter together with the deviations will be used as a notation for the faults in the control charts given in the results section. The helium chuck fault magnitude is not known specifically.

Fault Class	Parameter Varied	Number of Faults	Fault Experiments
1	TCP power (watts)	6	TCP +50, +30, +20, +10, -10, -15
2	RF power (watts)	4	RF +12, +10, +8, -12
3	Pressure (mtorr)	4	Pr +3, +2, +1, -2
4	Cl2 flow rate (sccm)	3	Cl2 +5, -5, -10
5	BCl3 flow rate (sccm)	3	BCl3 +10, +5, -5
6	He chuck pressure	1	He chuck*
Total		21	

* Magnitude of He chuck fault is not known

Table 2. Induced Fault Experiments

To make the test more representative of an actual sensor failure, the analysis was done with "reset" values: values for the controlled variable which was intentionally moved off its setpoint were reset to have the same mean as its normal baseline value, i.e. the controlled variable which was changed was reset to look normal in the data file. For example, if the induced fault was a change of the TCP power from 350 to 400 watts, the data file value of the TCP power was reset from a mean of 400 back to 350 by addition of a constant bias. The resulting data looks as if the controller was seeing a biased sensor for TCP power and adjusting accordingly: TCP power would appear normal, but it would not be. The effect of a TCP power offset, however, should be evident (we hope) in the remaining process variables because the apparent relationship between the TCP power and the remaining variables should be different.

The three induced fault experiments were run at widely spaced intervals (in February, March, and April 1996, respectively). Process drift is apparent in the data: each experiment has a significantly different multivariate mean. Issues concerning the process drift in this data are discussed elsewhere [8]. We will assume in this analysis that we are attempting to identify shifts in the data due to the induced faults as deviations away from the normal process drift. As a result, we can pre-treat the data in such a way, so as to eliminate the slow process drift (this is described in the next section).

DATA ANALYSIS

Data pre-processing

Data collected from the etch tool consisted of the average values for each of the 19 tool-state variables for each of the 129 wafers processed, which could be organized into a data matrix of size (129 x 19). Of the 129 wafers, 107 wafers represented normal process operation, 21 wafers were induced faults, and one wafer was an outlier that was removed from the data set. To discriminate against process drift, a high pass frequency filter was employed to remove the low frequency drift in the process. The high pass filter used for this data was the difference between the current point and the last non-fault data point. In practice, one might prefer to take the difference between the current data point and an exponentially-weighted moving average (EWMA) of the preceding data for the filter.

After removing the outlier and taking the difference of the data, the mean and variance of the differenced data were then calculated over the normal process wafer data for each of the machine process variables. SPC control charts were then constructed for each of the machine variables using the mean and variances. For the multivariate algorithms, the entire matrix is then scaled to mean zero and variance one for the normal process wafers, to place the same relative importance on all of the variables.

SPC Charts of Individual Variables

The first method we used for detecting faults on the data was to create statistical process control (SPC) charts for each of the tool-state variables collected. Since the deviations were expected on a wafer to wafer basis, single sample control charts were constructed [9]. The standard deviation and mean values of the variables for the normally processed wafer data were used as an estimate for the control limits. Since there were 19 tool-state variables, if all variables are monitored this would lead to 19 control charts – one for each variable. If each chart is designed with +/- 3-σ limits, then each chart individually would have a type I error (fault is detected when there is none) rate of 0.27%, i.e. every 370 wafers there would be one type I error. Unfortunately, type I errors for multiple charts are cumulative if the variables are independent; the type I error rate for 19 charts is 5%, i.e. every 20 runs there would be one type I error on average. The cumulative type I errors combined with the high maintenance cost associated with the upkeep of these 19 control charts makes this option for fault detection unattractive. Note that one can change the control limits to Bonferroni limits so that the cumulative type I error rate is lower, but only at the expense of decreasing the sensitivity in detecting true faults.

Multivariate T^2 Control Chart / T^2 Contribution Plots

By combining all of the tool-state variables together into one multivariate control chart, one can alleviate the type I error problem, and minimize the maintenance cost associated with a large number of control charts. The most popular way to do this is with a multivariate Hotelling's T^2 control chart. The single sample T^2 statistic is calculated as:

$$T_i^2 = \left(\mathbf{x}_i - \bar{\mathbf{x}}\right)\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (1)$$

where i is the wafer number, $\mathbf{x}_i$ is the (1x19) row vector of tool-state variables, $\bar{\mathbf{x}}$ is the (1x19) row vector of means for the tool-state variables, and $\mathbf{S}^{-1}$ is the inverse of the sample covariance matrix (19x19) for the 19 machine variables. The means and covariance are calculated on the normal process wafers only.

The T^2 statistic represents the multivariate extension of the t-statistic, and calculates the weighted distance of a given process data point from normal process operation. The statistic is squared, so it is always a positive value. The best way to understand the T^2 statistic is to look at what it represents in terms of geometry. If we represent the normal process as a cluster of data points in some n-dimensional space, where n is the number of tool-state variables, then T^2 represents the squared distance from the center of this data

cluster weighted relative to the noise in the normal process, which can be bounded by some hyper-ellipse (see Figure 1). This hyper-ellipse represents an upper control limit (*UCL*) at a given confidence interval for T^2 that can be calculated from an *F*-statistic as follows:

$$UCL = \frac{n(m-1)}{m-n} F_{1-\alpha,n,m-n} \tag{2}$$

where m is the number of wafers used in the training of the mean, and α is the type I error rate ($1-\alpha$ is the confidence interval). Since the only deviations that are important are those away from the mean, no lower control limit is used.

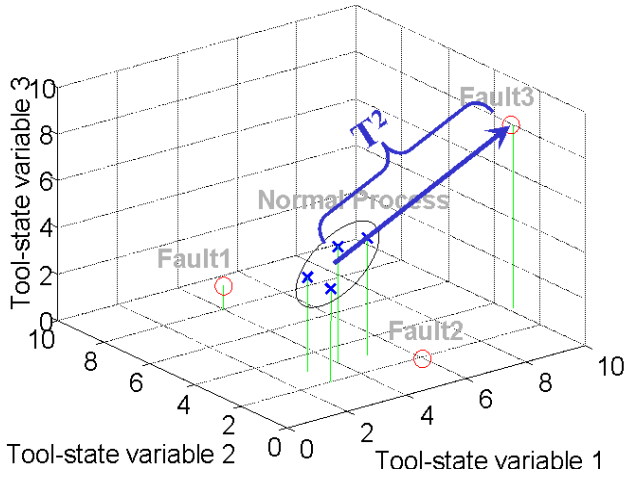

Figure 1. T^2 statistic measures the weighted squared distance from the normal process which tends to cluster in some region in the multivariate tool-state variable space. The normal process data points can be bounded by a hyper-ellipse that represents some confidence interval for detecting faults. Points falling outside of this hyper-ellipse are faults, and will have large values of T^2.

In the present work, we will utilize the entire data set in the determination of the T^2 statistic. In general, using the entire data set for the T^2 calculation can have potential pitfalls, particularly if the covariance matrix is ill-conditioned, i.e. the largest eigenvalue divided by the smallest eigenvalue is large (>100). In these highly collinear cases, it is better to calculate a T^2 statistic on a subset of the original data, e.g. T^2 can be calculated on a few principal component scores using principal component analysis (PCA). For our data set, the covariance matrix has a low condition number of 15, indicating that the entire data set can be used without problem.

Once a fault has been detected, one can calculate the contribution to the T^2 statistic for each variable, which can be used to help classify the cause of the faults. Large values represent a large contribution to T^2, whereas small values represent small contribution to T^2. The sign of each variable represents whether the shift away from the mean for a given fault was in a positive or negative direction for each of the variables.

Fault-specific Control Charts

The basis for our fault-specific control charts is the conjecture that different classes of faults will generally occur in unique directions (paths) away from the normal process operation in our multivariate tool-variable parameter space (see Figure 2). For small deviations around the normal operating process, as the size of a particular kind of fault increases, we would expect the deviation from the normal process to fall on a straight line. We can therefore represent that faults as linear combinations of tool-state variables. Discriminant analysis [10] provides a way to achieve maximum separation between the normal operation and any of these faults - i.e. it provides the best combination of variables, with which one can detect faults most sensitively. It can be shown that the best linear combination of variables is given by the *Fisher linear discriminant*, which can be determined by an appropriate regression of the tool-state data against definable classes [11]. The linear discriminant gives a direction in the multivariate tool-state variable space, onto which the data can be projected, which can be used for detecting a specific class of fault. This is depicted in Figure 2. For each fault that is classified, a new discriminant can be determined and a fault-specific control chart can be created. If the data set is well-suited and the charts are designed properly, a given fault will only occur on one fault-specific control chart, therefore providing simultaneous detection and classification of the fault.

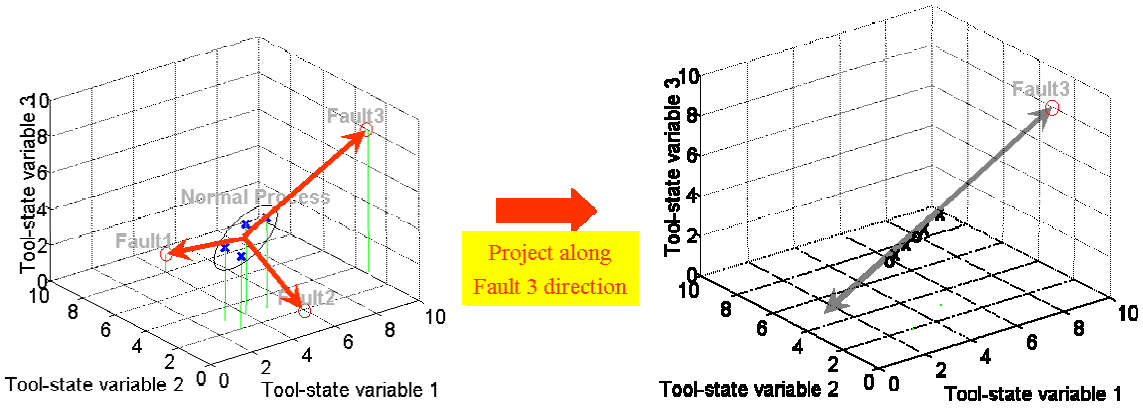

Figure 2. Different classes of faults occur in different directions in the multivariate tool-state space. By projecting the data along a specific direction pointing to a given class of fault, we can create a sensitive detection metric for that specific fault class. The same procedure can be done for each different class of fault to generate a set of fault-specific control charts.

Next we will describe two methods to create fault-specific control charts. In the first method, we make the assumption that the different fault classes occur naturally in different directions in the tool-state variable space, so that each of these fault classes only need to be discriminated against the normal process. In the second method, this is assumption is not made, and the different fault classes are discriminated against one another as well as the normal process. For each method, the input and outputs for a regression must be appropriately specified to achieve the proper discrimination.

Discrimination against normal process only - The first type of fault-specific control charts are designed by regressing each of *k* individual fault classes against the normal process data only. For example, all of the induced pressure fault data were compared with the normal process wafer data, to create a fault detection chart to specifically identify the pressure faults, but not to discriminate against other kinds of faults. The input and output for the regression were determined as follows:

$$\mathbf{X}_k = \begin{bmatrix} \mathbf{X}_{(NP)} \\ \mathbf{X}_{(Fault_k)} \end{bmatrix} \qquad \mathbf{y}_k = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Delta}_k \end{bmatrix} \tag{3}$$

where $\mathbf{X}_k$ is the matrix containing machine variable data from the normal process (NP) wafers and from the fault wafers ($Fault_k$) for a given fault type *k*. $\mathbf{y}_k$ is the output column vector containing zeros corresponding to the normal process wafers and a vector $\mathbf{\Delta}_k$ corresponding to the deviations in the given fault. For example, for the fault corresponding to the change in pressure, there were four fault experiments of +3, -2, +2, and +1 mtorr. Therefore, for pressure the output vector is calculated as:

$$\mathbf{y}_{pressure} = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Delta}_{pressure} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ +3 \\ -2 \\ +2 \\ +1 \end{bmatrix} \tag{4}$$

Using these input matrices and output vectors, a regression is calculated as:

$$\mathbf{b}_k = \mathbf{X}_k^+ \mathbf{y}_k \tag{5}$$

where $\mathbf{X}_k^+$ is the pseudo-inverse of $\mathbf{X}_k$, calculated in this work using the normal equations [12], and $\mathbf{b}_k$ is the vector of regression coefficients for a given fault class *k*. Alternatively, one might choose to use principal components regression (PCR), partial least squares (PLS), or ridge regression (RR) for determining the regression coefficients, particularly for an ill-conditioned data matrix X_k.

The data is then projected onto the regression vectors corresponding to the different classes of faults as follows:

$$\hat{\mathbf{y}}_k = \mathbf{X}_k \mathbf{b}_k \tag{6}$$

where $\hat{\mathbf{y}}_k$ are the multivariate fault parameters for the six different fault types. We can then calculate the mean and standard deviation of these fault parameters and set up SPC charts for these faults in the same manner as we did on the original variables earlier.

Discrimination against other faults and normal process – The second type of fault-specific control charts not only discriminate between each fault and the normal process, but also discriminate between the different faults. In terms of the geometry of the problem, this type of chart tries to orthogonalize the regression vector directions, $\mathbf{b}_k$, so

that only one specific kind of fault can be identified on any given chart. The only downfall is that the sensitivity to a given fault, relative to the normal process noise, is diminished when the vector direction is chosen to discriminate against other faults as well.

Setting up this type of chart is similar to the previous method, except that the data matrix, **X**, and the output vectors **y** are set up differently. In this case, the data matrix is given by:

$$\mathbf{X}_k = \begin{bmatrix} \mathbf{X}_{(NP)} \\ \mathbf{X}_{(Fault_{l \neq k})} \\ \mathbf{X}_{(Fault_k)} \end{bmatrix} \quad ; \quad \mathbf{y}_k = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{\Delta}_k \end{bmatrix} \tag{7}$$

where *k* is the current fault class being discriminated for and *l* are all other fault classes not of the current class. In this case, we are trying to discriminate between the given fault class and both the normal process wafers and all other faults, so that a chart can be created that uniquely identifies a given fault. A regression is performed in a similar manner on the input and output matrix and the data matrix can be projected onto each of the b_k projection directions to form six $\hat{\mathbf{y}}_k$ fault prediction parameters, from which SPC control charts can be constructed and monitored.

The fault-specific control charts are also subject to cumulative type I errors, but since there are hopefully fewer fault classes than tool-state variables, we expect the effect to be diminished. In this case, we have identified six specific classes of faults, which is less then one-third of the original variables. In a more general case, one may have to restrict the fault-specific control charts to some important subset of faults. Alternatively, one might lump several similar classes together to create a more manageable set.

RESULTS

SPC Charts of Individual Variables

For each of the 19 tool-state variables, SPC charts were created and faults were monitored. Two of these charts (for the *TCP load* and *Vat valve* variables) are shown in Figure 3, demonstrating that some of the faults can be captured on any given chart, whereas others are not. Note that the experiments with induced faults are represented on the charts with the corresponding labels given in Table 2 indicating the type and magnitude of the faults, whereas the experiments operating under normal process operation are shown without labels.

To see how sensitive individual variables were to faults and how many type I errors were detected, the faults were tabulated for each of the 19 machine variables and cumulatively for all of the 19 charts combined. The results are shown in the bar graph in Figure 4. One can see that certain variables were more capable of detecting faults than others, and cumulatively 20 out of the 21 induced faults were detected on at least one of these charts. Unfortunately, the number of falsely detected faults (type I errors) is rather

high when using 19 charts; 12 type I errors were detected out of 107 normal process wafers, which is actually higher than the 5% type I errors that we would predict for the 19 charts.

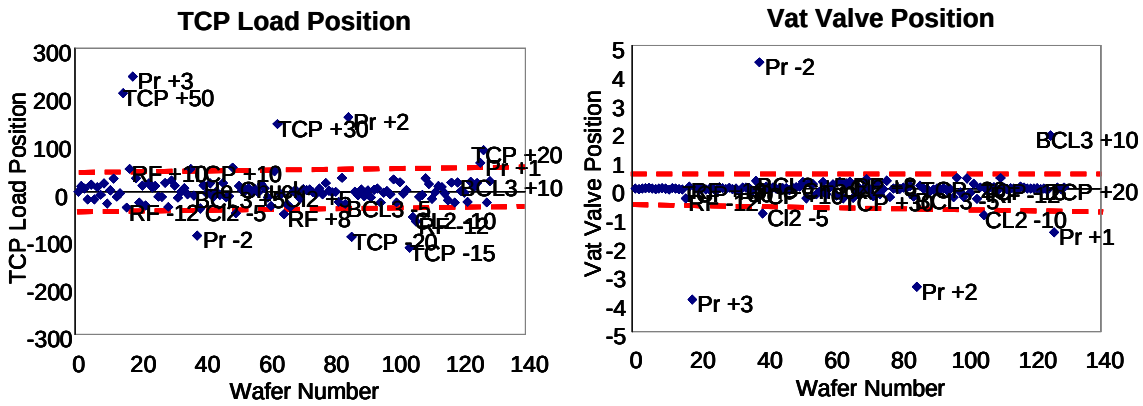

Figure 3. SPC charts were constructed for the 19 machine process variables, two of which are shown here. The dotted lines represent +/- 3 sigma control limits. The normal process wafers are shown as diamonds without any labels; the induced faults are shown with labels to indicate the type and magnitude of the fault. One can see from these plots, that some of the induced faults are captured, whereas others are not in each of these two charts.

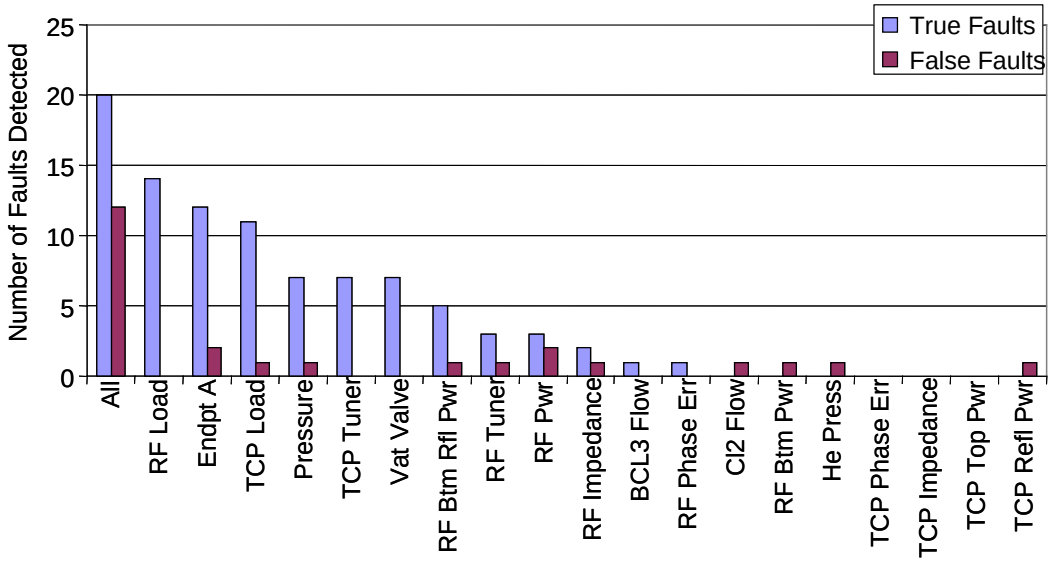

Figure 4. Bar Chart showing the number of true faults detected as well as the number of false faults (type I errors) for each of the individual variables and cumulatively for all of the variables. Cumulatively, 20 of the 21 induced faults were detected with the individual charts, but 12 faults were detected for normal process wafers. This gives a type I error rate of greater than 10%, which is even higher than the 5% rate predicted.

T^2 Control Charts

The T^2 control chart for the data is shown in Figure 5. One can see from this figure that 19 of the 21 faults were detected with the T^2 chart, with a low false alarm rate - only one wafer was falsely detected.

The contribution plots for T^2 can be used to help identify a *signature* or *fingerprint* of a particular fault. For example, the fault corresponding to a +3 mtorr deviation in pressure is shown in Figure 6. One can see that certain variables have large contributions, i.e. large deviations from normal process behavior, and can be associated with a particular fault. By comparing the T^2 contribution plot fingerprint for a given fault against a library of fingerprints from previous faults, this will hopefully enable easier identification of the cause.

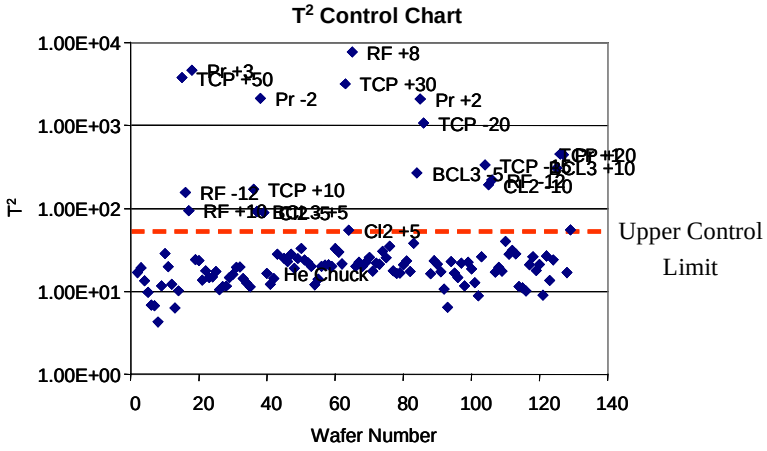

Figure 5. T^2 control chart for data. The dashed line shows the 3-sigma upper control limit for the data. 20 of 21 faults were correctly identified with the T^2 control chart and only one fault was falsely identified.

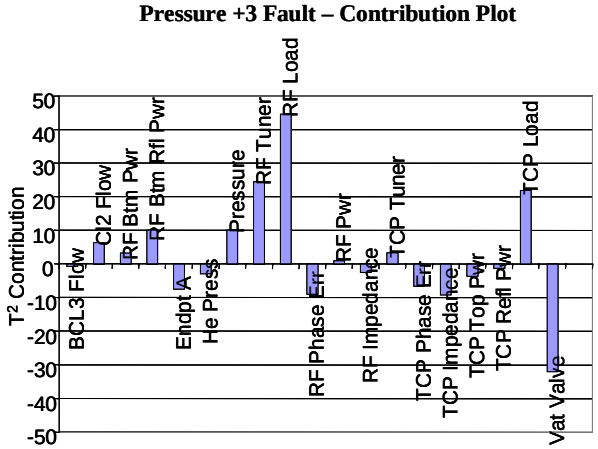

Figure 6. T^2 contribution plot for the fault with a +3 mtorr deviation in pressure reveals that certain machine variables such as RF Load and Tune positions, TCP Load position, and Vat Valve positions have significant shifts from normal process operation. This contribution plot constitutes a fingerprint for the pressure fault, and contribution plots from future faults could be compared with this and other fault fingerprints for identifying possible causes.

Fault-Specific Control Charts

Discrimination against Normal Process Only - Figure 7 shows a plot of two fault-specific control charts for pressure and TCP power faults, that were generated using the method that discriminates against the normal process only. One can see that although the pressure and TCP faults are correctly identified on their respective charts, several other faults are also detected on each chart. Since a given type of fault may appear on multiple charts, the fault can not be uniquely classified without using some additional method. One possible method to aid in classifying a given fault would be to calculate the correlation of the fault direction in the tool variable space with the direction vectors, $\mathbf{b}_k$, for the different fault classes.

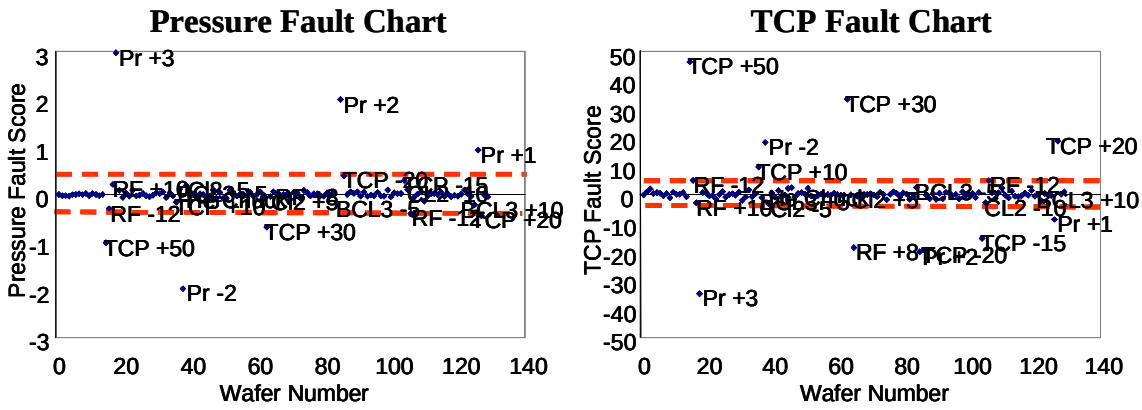

Figure 7. Fault detection control charts for two specific kinds of faults, for charts regressed on individual faults. The dotted line indicates 3-sigma control limits. One can see that for these charts, although the pressure and TCP power faults are picked up by their respective fault detection control charts, each chart also detects several other faults. As a result, these charts do not necessarily give a clear identification of the exact cause of the fault, without the use of some other additional pattern recognition.

The results of all of the charts have been summarized in Figure 8. For each fault detection control chart, the fraction of faults from a given class that were correctly identified are given along with the fraction of faults detected from other classes, and the fraction of faults that were falsely detected from the normal process wafers. These results indicate that for each chart, all of the faults from the class of fault for which the chart was created have been detected. However, in most cases a large number of other fault classes have also been detected. This indicates that the classification of a given fault cannot be uniquely determined, since different fault classes are detected on multiple charts. Nevertheless, we can see that using these charts cumulatively gives better overall detection sensitivity than T^2 or the individual SPC charts, with only one falsely detected fault – the same number detected with T^2 and far less than the 12 faults detected with the individual chart.

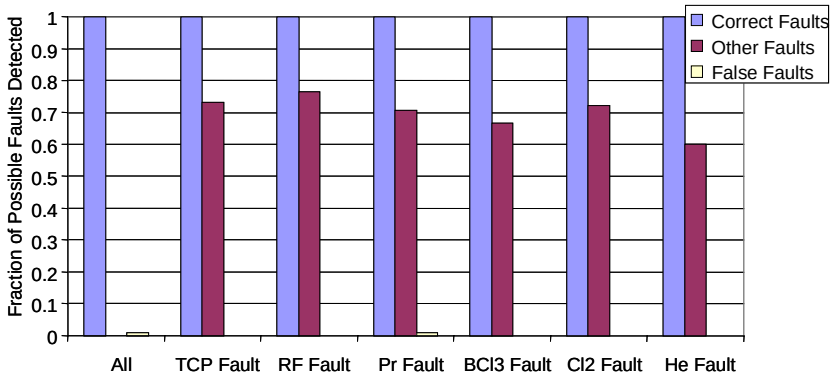

Figure 8. Bar Chart showing the fraction of faults that were detected correctly for that specific fault class, fraction of other induced faults that were detected, and the fraction of falsely detected faults. The cumulative results of all of the charts is also shown, demonstrating that all of the faults were identified correctly with only one false detection out of 107 normal process wafers.

Discrimination against Normal Process and Other Faults - Figure 9 shows the SPC charts for two such fault detection charts revealing that the charts are most sensitive to the specific faults they are trying to detect. The overall performance of these charts is summarized in Figure 10. This figure shows the fraction of faults correctly identified for the particular class of fault represented by each control chart, the fraction of faults from other classes that were detected, and the fraction of normally processed wafers that were falsely detected as faults. One can see that compared to the previous method, this method does a much better job of fault-specific detection, at the expense of a slightly higher false detection rate. In fact, 12 out of 21 of the faults are uniquely detected on one chart, and five more are detected on only two of the charts. This allows for the simultaneous detection and classification of faults that was desired for the majority of the faults.

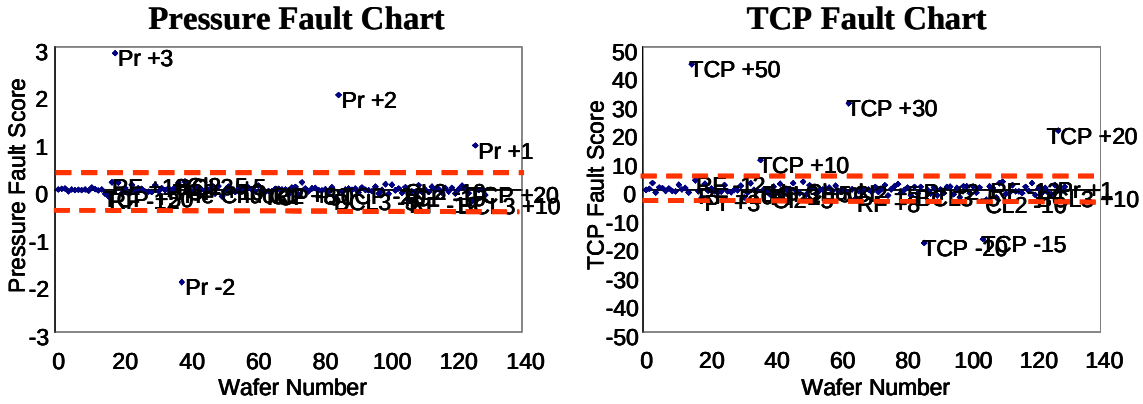

Figure 9. Pressure and TCP fault detection charts are shown here, revealing that the pressure and TCP types of faults can be uniquely detected and classified by these charts.

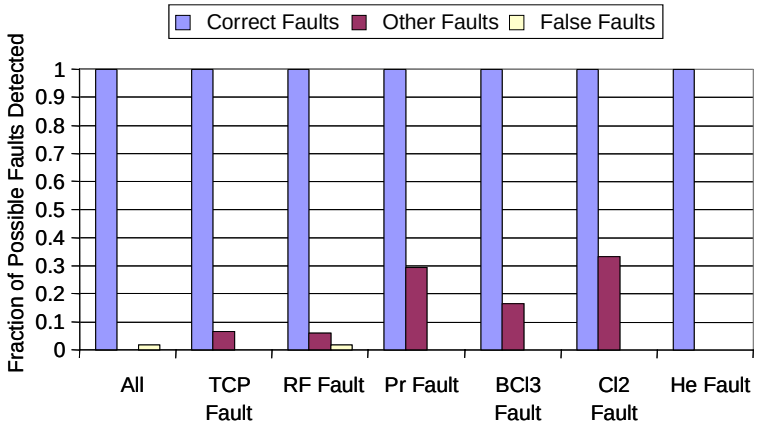

Figure 10. Bar chart showing the fraction of faults correctly identified for each class, the fraction of faults detected from other classes, and the fraction of falsely detected faults (type I errors). Looking cumulatively at all of the charts, one finds that all of the faults are detected properly and only two out of a possible 107 normally processed wafers are incorrectly detected as faults. In this case, where the faults are discriminated against one another, there is a low rate of detection of faults not of the correct class. This allows for simultaneous detection and classification.

CONCLUSIONS

We have proposed and demonstrated a method for simultaneous detection and classification of faults. Two methods have been proposed using discriminant analysis as a basis for generating fault-specific control charts - one method discriminates a given class of fault against the normal process wafers only and the other method discriminates a given class of fault against all other fault classes and the normal process wafers. These methods have been compared with traditional methods for detection of faults on a sample set of data collected on an industrial plasma etching tool. The results indicate that the simultaneous detection and classification methods provide improved performance for detection sensitivity over the traditional methods – all 21 faults were detected when using the simultaneous detection methods, with low type I error rates.

The first method provides the best overall sensitivity for fault detection, but does not do an adequate job of uniquely classifying the fault. This is because it does not provide discrimination between various classes of faults. The second method, however, provides a way to uniquely detect faults on a single chart for most classes of faults, and therefore provides a way to simultaneously detect and classify the fault.

This paper demonstrates the concept of simultaneous fault detection and classification, but does not address key issues concerning the practical implementation of fault-specific control charts in a manufacturing setting. For example, how does one train for the different classes of faults? Is it reasonable to intentionally induce faults or is it better to construct or modify charts as the faults are detected over time? Also, since the fault-specific control charts will generally only capture faults which have previously been characterized, a general-fault control chart such as a T^2 control chart should be

considered for use in conjunction with the fault-specific control charts. Future work should address practical issues such as these that must be taken into consideration for implementation in manufacturing.

ACKNOWLEDGEMENTS

This work was completed under Semiconductor Research Corporation Task 704.001.

REFERENCES

[1] B. M. Wise, N. B. Gallagher, S. W. Butler, D. D. J. White, and G. G. Barna, J. Chemometrics, **13**(3-4), 379 (1999).
[2] B. M. Wise, N. B. Gallagher, and E. B. Martin, J. Chemometrics, **15**(4), 285 (2001).
[3] H. Yue, S. J. Qin, R. J. Markle, C. Nauert, and M. Gatto, IEEE Trans. Semicond. Manufact., **13**(3), 374 (2000).
[4] R. Bunkofske, C. Ambrozic, and M. Sanders, in *AEC/APC Symposium XIII*, p. 31, Banff, Canada (2001).
[5] A. M. Ison, W. Li, and C. J. Spanos, in *Proceedings of the IEEE International Symposium on Semiconductor Manufacturing*, p. B49, San Francisco, CA (1997).
[6] P. L. Love, and M. Simaan, in *Proceedings of the IEEE International Symposium on Intelligent Control*, p. 736, Arlington, VA (1989).
[7] G. S. May, and C. J. Spanos, IEEE Trans. Semicond. Manufact., **6**(1), 28 (1993).
[8] N. B. Gallagher, B. M. Wise, S. W. Butler, D. White, and G. G. Barna, in *IFAC ADCHEM '97*, p. 78, Banff, Canada (1997).
[9] D. C. Montgomery, *Introduction to Statistical Quality Control*, Third ed., John Wiley & Sons, Inc., New York (1997).
[10] P. A. Lachenbruch, *Discriminant Analysis*, Hafner Press, New York (1975).
[11] B. E. Goodlin, Ph.D. Thesis, MIT, Cambridge, MA (2002).
[12] G. Strang, *Introduction to Linear Algebra*, Wellesley-Cambridge Press, Wellesley, MA (1993).