Korean J. Chem. Eng., **25**(1), 13-18 (2008)

Multi-way principal component analysis for the endpoint detection of the metal etch process using the whole optical emission spectra

Kyounghoon Han, Kun Joo Park*, Heeyeop Chae,† and En Sup Yoon**

School of Chemical & Biological Engineering, Seoul National University,
San 56-1 Shillim-dong, Gwanak-gu, Seoul 151-744, Korea
*DMS Co. Ltd, Suwon 445-810, Korea
**Department of Chemical Engineering, Sungkyunkwan University, Suwon 440-746, Korea
(*Received 22 December 2006 • accepted 7 May 2007*)

Abstract−An endpoint detection algorithm based on multi-way principal component analysis (MPCA) is developed for plasma etching processes. Because many endpoint detection techniques use a few manually selected wavelengths, noise renders them ineffective and it is hard to select important wavelengths. Furthermore, process drift and faulty condition should be considered for more robust endpoint detection at the same time. In this paper, MPCA with the whole optical emission spectra is used for effective endpoint detection using a large set of data. And the fault detection was achieved by concept of 'product' and 'mean deviation value' chart with the result of each wafer's endpoint detection. The product was defined by the multiples of OES data with loading vector and mean deviation chart was defined by a chart of the difference between the product value of the target wafer and mean value of previous wafers. Therefore, a robust model for endpoint detection can be developed by excluding faulty wafers. This approach is successfully applied to the metal etch process of TiN/Al-0.5%Cu/TiN/Oxide stack in an inductively coupled BCl_3/Cl_2 plasma. The optical emission signal intensities of the 129 wavelengths were measured and saved in a four-dimensional (wavelengths, time, intensity, and wafers) matrix for the subsequent data processing. With this approach the endpoint signal was improved with the whole emission spectra and the process drift was considered by MPCA after information of faulty wafers was discarded.

Key words: Plasma Etching, Optical Emission Spectrometer, Multi-way Principal Component Analysis, Endpoint Detection

INTRODUCTION

In semiconductor processing, plasma etching is typically employed to define the micro- and nano-scale patterns on a silicon wafer. When the target layer is cleared, it is critical to stop the plasma etching to avoid excessive over-etching, an event called endpoint detection (EPD). Typically, the uniformity of film thickness should be maintained within 5% or so and it is inevitable to over-etch to ensure that all the lines, contacts and vias are completed. However, excessive over-etching may remove the film underneath the target layer, and too much over-etching can cause device failures and subsequent yield reduction. Therefore, it is critical to determine the endpoint without damaging of the underlayer.

The most widely used method for end point detection is to monitor the optical emission trace of reactive species in plasmas by using an optical emission spectrometer (OES) [1]. By measuring optical emission signal intensities at specific wavelengths, one can identify the neutral particles and ions present in the plasma. Most endpoint detection methods using OES focus on identifying a single wavelength corresponding to a chemical species that shows a pronounced transition at the endpoint [2-4]. When the open area is less than 1%, the single wavelength analysis shows its limitation due to weak signal intensity [8].

Biolsi et al. [5] demonstrated an advanced endpoint system for small open-area etching by applying threshold signal processing with single wavelength signal. This single wavelength method cannot avoid the noise problem or time delay associated with filtering. Furthermore selection of appropriate wavelengths requires significant experience of process engineers. So these methods usually reliably work only for the large open area wafers (typically larger than 10%). White et al. [6] proposed T^2 and Q statistics for the endpoint detection of low open-area wafers using PCA in conjunction with T^2 detection and recursive mean update. They improved signal sensitivity, but their model cannot include the drift of the process. To overcome this limitation, recursive mean and covariance updates are needed for real-time adjustment. Yue and co-workers [8] extracted a reliable endpoint signal using the principal component analysis (PCA). In this algorithm, loading vectors are used for the wavelengths selection, and principal component (PC) values are monitored for the EPD. They suggested the sphere criterion method using the loading vectors for the wavelength selection, but this method also does not consider abnormal process conditions. The drift or abnormal process conditions should also be excluded for the exact EPD modeling when the information from the multiple wafers is utilized.

In this paper, the entire optical emission spectra were used for the EPD and fault detection. And the multiple wafer algorithm was also presented for the EPD prediction in real time with normal wafer data.

MULTI-WAY PRINCIPAL COMPONENT ANALYSIS

PCA [8] is a famous tool for data compression and information

†To whom correspondence should be addressed.
E-mail: hchae@skku.edu

extraction. PCA decomposes the data matrix (**X**) as the sum of outer product of vectors $\mathbf{t}_i$ and $\mathbf{p}_i$ plus a residual matrix (**E**):

$$\mathbf{X}=\mathbf{t}_1\mathbf{p}_1^T+\mathbf{t}_2\mathbf{p}_2^T+\ldots+\mathbf{t}_k\mathbf{p}_k^T+\mathbf{E} \quad (1)$$

Here, k must be less than or equal to the smaller dimension of **X**. The $\mathbf{t}_i$ vectors are defined as ***scores***, and contain information on how the samples relate to each other. The $\mathbf{p}_i$ vectors are known as ***loadings*** and contain information on how the variables relate to each other.

In the PCA decomposition, the $\mathbf{p}_i$ vectors are the eigenvectors of the covariance matrix, *i.e.*, for each $\mathbf{p}_i$:

$$\mathrm{Cov}(\mathbf{X})\mathbf{p}_i=\lambda_i\mathbf{p}_i \quad (2)$$

Here, is the λ_i is the eigenvalue associated with the $\mathbf{p}_i$ and Cov(**X**) means covariance matrix of **X**. Note that for **X** and any $\mathbf{t}_i$, $\mathbf{p}_i$ pair:

$$\mathbf{X}\mathbf{p}_i=\mathbf{t}_i \quad (3)$$

The score vector ($\mathbf{t}_i$), is the linear combination of the original **X** variables defined by $\mathbf{p}_i$. For the endpoint detection, the reduced PC variables and score vectors can be used for the prediction of EPD time.

Because this simple PCA method cannot consider the process drift among the wafers, mean and covariance estimation was necessary for the dynamic process condition. In MPCA, the array $\underline{\mathbf{X}}$ is decomposed into the summation of the product of the score vectors (**t**) and loading matrices (**P**), plus a residual array ($\underline{\mathbf{E}}$), which is minimized by least squares.

$$\underline{\mathbf{X}}=\sum_{r=1}^{R}\mathbf{t}_r\otimes\mathrm{P}_r+\underline{\mathbf{E}} \quad (4)$$

Here, $\otimes$ is the outer product operator. This decomposition, which is achieved in accordance with the principles of PCA, optically separates the data into two parts. The systematic part, the sum of $\mathbf{t}_r\otimes\mathrm{P}_r$ expresses the deterministic variation as part (**t**), which is only related to batches. The second part (**P**) is related to variables and their time variation. The noise or residual part ($\underline{\mathbf{E}}$) is a set as small as possible, which is associated with non-deterministic variation in the data.

MPCA is equivalent to performing PCA on a large two-dimensional matrix, which is formed by unfolding the three-way array ($\underline{\mathbf{X}}$). For example, one might unfold $\underline{\mathbf{X}}$ (samples **I**×variables **J**×batch **K**) in such way as to position each of its slices (**I**×**J**) with a batch interval (**K**) as shown in Fig. 1. This particular unfolding allows the variability among batches in $\underline{\mathbf{X}}$ with variables and their time variation. As the same way for principal component prediction, MPCA prediction can also be made with a three dimensional matrix using the scores of $\underline{\mathbf{X}}$.

ALGORITHM FOR EPD WITH MPCA

In this paper, a *single wafer model* was defined when the EPD time of the second wafer was predicted by the PCA loading vector of the first wafer. And a *multiple wafer model* was also defined when the EPD time of the last target wafer was predicted by the MPCA loading vector of the previous several wafers.

1. Single Wafer Model

In a process, the decision for endpoint call is often made with little information. The OES information from the former wafer is typically used to decide the endpoint of the next target wafer with the same process recipe. In the single wafer model, the loading vectors of the first wafer were used as the predictor. And then these loading vectors were multiplied with the real time OES data of the second wafer during the plasma etching process, as shown in Fig. 2.

This multiplying operation between the OES data and loading vectors was defined as ***product*** in this paper.

$$\mathbf{Y}_i=\mathbf{X}'\mathbf{P}_i \quad (5)$$

$\mathbf{Y}_i$ means product value of i^{th} sample time and $\mathbf{X}'$ means the data matrix without normalization. The normalization means centered and divided by its standard deviation for each column in this paper. So there can be the ***product lines*** as the same numbers as the set of loading vectors. Because this product method uses data without scaling or preprocessing, it can be adjusted to a real-time process. Furthermore, these product lines also can be used for fault detection, because the product line itself (actual product line) shows the overall process conditions.

The EPD time can be determined in this algorithm by using these products (estimated product line) as the following procedures. Initially, the entire range of OES signals from the first wafer was captured and normalized. The covariance of this normalized data was obtained and a singular value decomposition (SVD) performed. The loading vectors were obtained from solving the eigenvalue problem of the result of its SVD. Finally, the entire range of OES signals of the second wafer was captured in real time for multiplying with the loading vector of the first wafer. Over 80% of information can be represented only with three products of the first three PCs in most

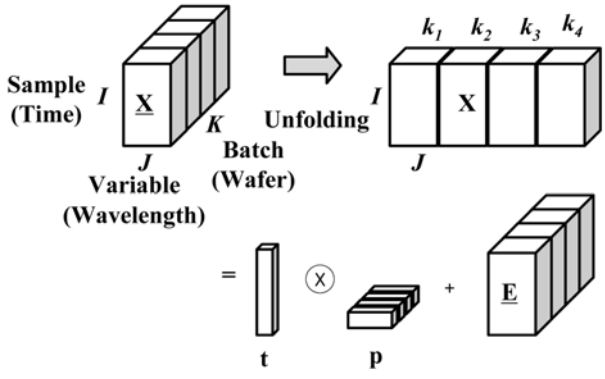

Fig. 1. Algorithm for performing MPCA [10].

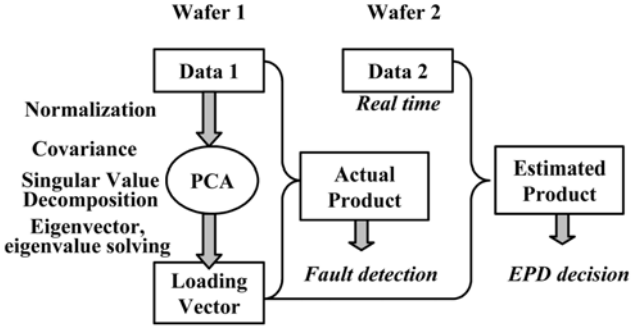

Fig. 2. Algorithm of single wafer model for real time EPD.

cases. The sensitivity can be enhanced further by using the ratios of these three products.

2. Multiple Wafer Model

2-1. Fault Detection Using the Single Wafer Model

If there are some problems or faults in the first wafer, its information cannot be useful for the second wafer's EPD. This abnormal situation can be distinguished by definition by the ***mean deviation value (MDV)***.

$$\mathrm{MDV}=\left|\mathrm{Y}_{n,j}-\frac{\sum_{i=k}^{n-1}\mathrm{Y}_{i,j}}{\mathrm{n}-\mathrm{k}}\right| \tag{6}$$

Here, $Y_{n,j}$ is the product value of n^{th} wafer and j^{th} sample time. Because the product line already was acquired in the single wafer model, these MDV lines can be drawn directly. And the faulty situation can be decided by the counting of MDV numbers which is over a limit value as shown in Fig. 3. The limit can be absolute values or multiples of standard deviation.

2-2. Endpoint Detection Using the Multiple Wafer Model

Now a multiple wafer model was developed with normal wafers after above faulty wafer filtering as shown in Fig. 4. The single wafer model is used for the real time EPD from the first wafer to the $(n-1)^{th}$ wafer with two wafers. However, a database can be compiled with real time PCA information from each wafer. And the variables (wavelengths) are ranked by the loading vectors, which provide for maximum variability, as shown in the following equation [8]:

$$l_i=\sqrt{\sum_{j=1}^{n}l_{ij}^2} \tag{7}$$

Where, l_{ij} denotes the loading vectors for the i^{th} wavelength and j^{th} PC. MPCA prediction was performed with selected high ranking wavelengths for the n^{th} target wafer in the same way as PCA prediction. And this model can be updated by the previous MDV fault decision with n^{th} wafer after its EPD.

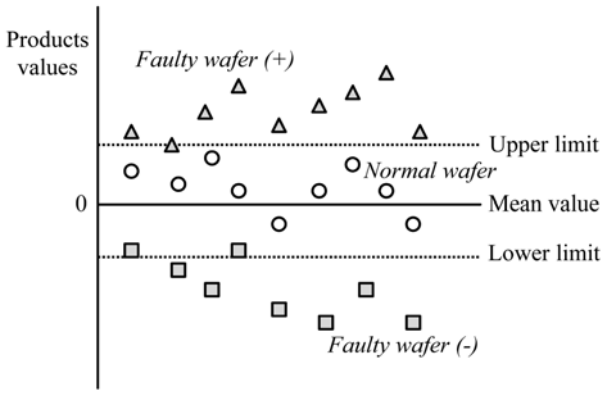

Fig. 3. Mean deviation value (MDV).

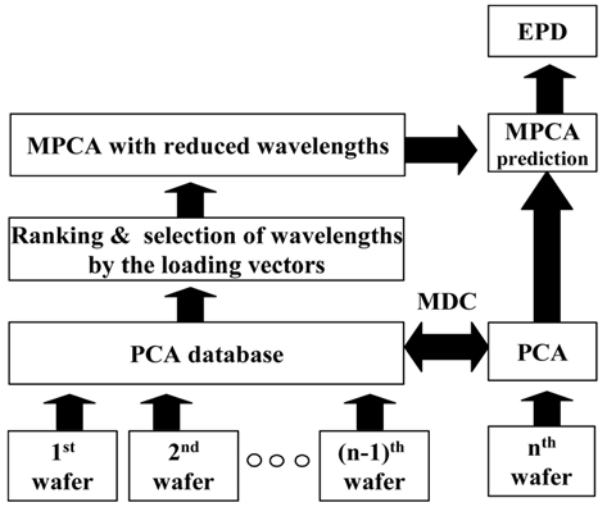

Fig. 4. Algorithm of multiple wafer model for real time EPD.

APPLICATION TO THE END POINT DETECTION OF METAL ETCH

For the case study, the open data source of the monitoring prob-

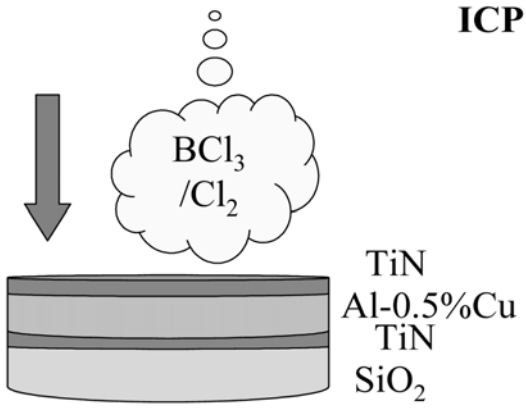

Fig. 5. Wafer composition for the metal etch [11].

Table 1. Wafer set with the intentional faults

Exp	Wafer numbers	Deviation	Detection
29	15	TCP +50	○ (+)
29	16	RF +10	×
29	17	Pr +3	○ (−)
29	35	TCP +10	○ (+)
29	36	BCl_3 +5	○ (−)
29	37	Pr -2	○ (+)
29	38	Cl_2 -5	○ (−)
29	39	He chuck	×
31	62	TCP +30	○ (+)
31	63	Cl_2 +5	○ (+)
31	64	RF +8	×
31	83	BCl_3 -5	○ (+)
31	84	Pr +2	○ (+)
31	85	TCP -20	○ (−)
33	102	TCP -15	○ (−)
33	103	Cl_2 -10	○ (−)
33	104	RF -12	×
33	122	BCl_3 +10	○ (−)
33	123	Pr +1	○ (−)
33	124	TCP +20	○ (+)

lem in the semiconductor processing was used. ([9], Eigenvector Research Inc.) The goal of this process was to etch a TiN/Al-0.5%Cu/TiN/oxide stack to form a metal line employing an inductively coupled BCl_3/Cl_2 plasma, as shown in Fig. 5. Our focus was only on an Al-stack etch process, which was performed on the commercially available Lam 9600® plasma etch tool. The OES data were collected, which consisted of 40 process set points and the measured variables sampled at 1 second intervals. This experiment was done at three different times considering the condition drift, and consisted of 126 wafers with 20 faults totally. Several intentional faulty conditions were applied in three experiments: the changes of the transformer coupled plasma power (TCP), RF bias power (RF),

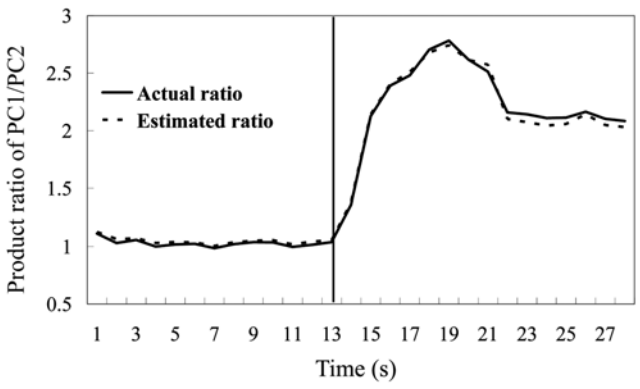

Fig. 6. Endpoint detection with estimated ratio curve of single wafer model.

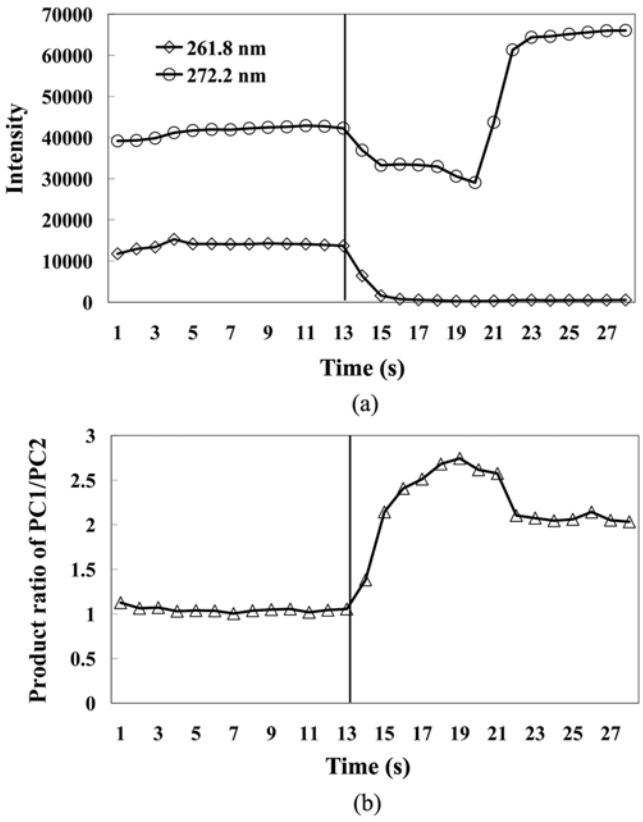

Fig. 7. Comparison of EPD with (a) several single wavelengths and (b) single wafer model.

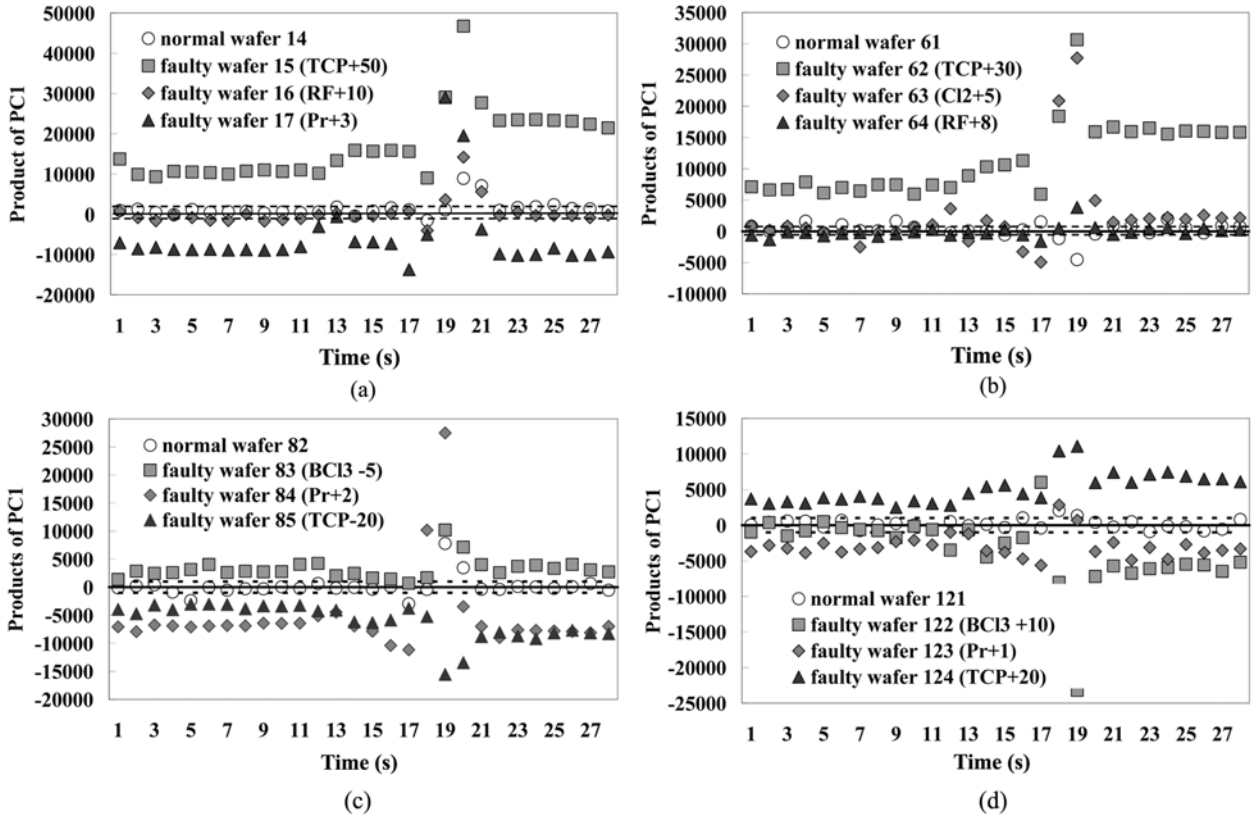

Fig. 8. MDV comparison of normal and faulty wafers (a. 14-17, b. 61-64, c. 82-85, d. 121-124 wafer numbers).

pressure (Pr), Cl_2 or BCl_3 flow rate, and the He chuck pressure, as described in Table 1.

For the single wafer model, the ratio of PC1 product over PC2 product is used for the explicit behaviors (PC1 contains 48.92%, PC2 contains 25.33% of the variance). This % variance is the ratio of each PC value over the total sump of PC values. In Fig. 6, the estimated product ratio (dotted line) shows the product ratio of the first wafer's loading vector as predictor and the second wafer's data. The actual ratio (solid line) shows the real product value by the loading vector and data of the second wafer itself. They show good correlation with our single wafer model estimation. And this estimated product ratio was compared with two important single wavelengths (261.8 nm of AlCl and 272.2 nm of BCl), which are commonly used for the EPD method, as shown in Fig. 7. In this diagram, an EPD of TiN/Al-0.5% layer was decided about 13 s from the simple clear estimated curves, which were no different from the actual curve because the curve significantly changed. It is not needed to consider which wavelength should be collected for using the whole optical emission data. And some filters also are not needed for denoising, which brings time delays.

To verify our fault detection algorithm, MDV with previous 10 wafer's product lines were used for the 11^{th} target wafer as shown in Fig. 8 (*e.g.*, from wafer 4 to wafer 13 for the wafer 14 at (a)). The limit was set at a value of ±1,000 (arbitrary unit) for the decision of the process condition and this value can be changed by the times of standard deviation generally. The MDV shows the normal wafer is almost located within the limit values (about 86% in case of wafer 82 at (c)), but the faulty wafers were not (0% in case of wafer 85 at (c)). Furthermore, the location of the deviation ('+' means located upper than the upper limit, '–' means lower than the lower limit) can explain the trend of the deviation. Because the increasing of (TCP, Cl_2) and decreasing of (Pr, BCl_3) bring more plasma intensity, these conditions correspond with '+' deviation in MDV very well. Changes in the RF power and He chuck pressure were not detected because their change is localized in the chuck area and does not affect much the signal intensity of bulk plasma optical emission spectra. So the information of faulty wafers could be discarded for more robust endpoint detection of the multiple wafer model.

For the verification of our multiple wafer model, the ranking of the wavelengths was conducted by using the two loading vectors (of PC1, PC2) from the previous five normal wafers (*e.g.*, from wafer 1 to 5 for the wafer 6). The 60% highest ranked wavelengths (79 wavelengths from 129) were selected, and then MPCA prediction was performed. Fig. 9 shows the result of the MPCA prediction for the 18^{th} target wafer EPD. Previous five normal wafers (11, 12, 13, 14, and 16) were used whereas two wafers (15, 17) can be discarded by our MCD. The line with circles (○) shows the PCA curve for the entire spectra of the 18^{th} wafer itself (actual product line of 100%). The solid line shows the PCA with 60% reduced variables (actual product line of 60%), and the MPCA predictive dotted line shows a good estimation when compared to this solid line. There was the same time change of these three curves, which confirmed the EPD time to be about 18 sec.

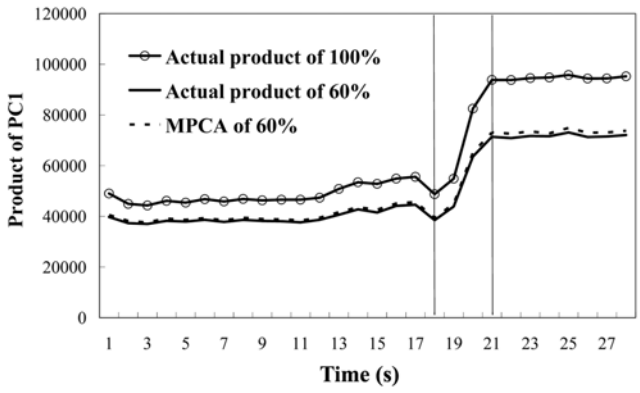

Fig. 9. EPD of 18^{th} wafer from the multiple wafer model.

Table 2. Percent captured by PCA of OES data

Principal component	% Variance	% Variance cumulative
1	48.92	48.92
2	25.33	74.25
3	3.18	77.43
4	2.40	79.83

CONCLUSIONS

In this paper, an endpoint detection algorithm for the plasma etching process was developed based on the MPCA prediction methods. Because the existing single wavelength method shows its limitation for the small open area endpoint detection, we developed the endpoint detection algorithm with the whole wavelengths signals effectively. This algorithm includes a single wafer model and a multiple wafer model. The single wafer model uses real time OES data of the second wafer with the PCA loading vector of the first wafer. And the multiple wafer model uses the piles of the previous wafers for the more robust EPD. The multiple wafer model could be updated with a new wafer's PCA information after its normality is checked by MDV fault detection method. In the case study, this algorithm was applied to the metal etch to verify its effectiveness. So the smooth PCA predictive line could be acquired by the single wafer model. And a more robust MPCA predictive line also could be acquired by the multiple wafer model, because it can include process drift and extraction of a faulty wafer's information. This more robust EPD algorithm can give us more reliable information of the etching systems for process monitoring and run-to-run control.

ACKNOWLEDGMENT

We acknowledge the financial aid for this research provided by the Brain Korea 21 Program, which is supported by the Ministry of Education & Human Resources Development. In addition, we would like to thank the Automation and Systems Research Institute and the Institute of Chemical Processes of Seoul National University.

NOMENCLATURE

$\underline{E}$: residual array
l_{ij} : loading vectors for the i^{th} wavelength and j^{th} PC
l_i : loading vector summation of i^{th} wavelength

p_i : eigenvectors of the covariance matrix of X
Pr : pressure
P_r : loading matrix
RF : radio frequency
t_i : scores
t_r : score vectors
X : data matrix
$\underline{X}$: data array
$\otimes$: outer product operator

REFERENCES

1. A. Grill, *Cold plasma in materials fabrication*, IEEE Press (1994).
2. S. J. Pearton, F. Ren, C. R. Abernathy and C. Constantine, *Materials Science Engineering*, **B23**, 36 (1999).
3. L. Dreeskornfeld, R. Segler, G. Haindl, O. Wehmeyer, S. Rahn, E. Majkova, U. Kleineberg, U. Heinzmann, P. Hudek and I. Kostic, *Microelectronic Engineering*, **54**, 303 (2000).
4. H. S. Kim, Y. J. Sung, D. W. Kim, T. Kim, M. D. Dawson and G. Y. Yeom, *Materials Science & Engineering*, **B82**, 159 (2001).
5. P. Biolsi, D. Morvay, Drachinik and S. Eliinger, *An advanced endpoint solution <1% open area applications; contact and via*, IEEE/SEMI Advanced Semiconductor Manufacturing Conference, 391 (1996).
6. D. White, B. E. Goodlin, A. E. Gower, D. S. Boning, H. Chen, H. H. Sawin and T. J. Dalton, *IEEE Trans. Semiconductor Manufacturing*, **13**(2), 193 (2000).
7. J. Jackson, *J. Qual. Tech.*, 12, 201 (1981).
8. H. H. Yue, S. J. Qin, J. Wiseman and A. Toprac, *Journal of Vacuum Science & Technology*, **A 19**, 66 (2001).
9. http://www.software.eigenvector.com/Data/Etch/
10. B. M. Wise, N. B. Gallagher, S. W. Butler, D. D. White Jr. and G. G. Barna, *Journal of Chemometrics*, **13**, 379 (1999).
11. B. M. Wise, N. B. Gallagher, R. Bro, J. M. Shaver, W. Windig and R. S. Koch, *PLS_Toolbox 3.5 for use with MATLAB™*, Eigenvector Research (2004).