

Principal Component Analysis of Optical Emission Spectroscopy and Mass Spectrometry: Application to Reactive Ion Etch Process Parameter Estimation Using Neural Networks

Reza Shadmehr,¹ David Angell,² Paul B. Chou, Gottlieb S. Oehrlein, and Robert S. Jaffe

IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York 10598

ABSTRACT

We report on a simple technique that characterizes the effect of process parameters (*i.e.*, pressure, RF power, and gas mixture) on the optical emission and mass spectra of CHF_3/O_2 plasma. This technique is sensitive to changes in chamber contamination levels (*e.g.*, formation of Teflon-like thin-film), and appears to be a promising tool for real-time monitoring and control of reactive ion etching. Through principal component analysis, we observe that 99% of the variance in the more than 1100 optical and mass spectra channels are accounted for by the first four principal components of each sensor. Projection of the mass spectrum on its principal components suggests a strong linear relationship with respect to chamber pressure. This representation also shows that the effect of changes in thin-film levels, gas mixture, and RF power on the mass spectrum is complicated, but predictable. To model the nonlinear relationship between these process parameters and the principal component projections, a feedforward, multi-layered neural network is trained and is shown to be able to predict all process parameters from either the mass or the optical spectrum. The projections of the optical emission spectrum on its principal components suggest that optical emission spectroscopy is much more sensitive to changes in RF power than the mass spectrum, as measured by the residual gas analyzer. Model performance can be significantly improved if both the optical and mass spectrum projections are used (so called sensor fusion). Our analysis indicates that accurate estimates of process parameters and chamber conditions can be made with relatively simple neural network models which fuse the principal components of the measured optical emission and mass spectra.

In the reactive ion etching (RIE) process, plasma characteristics depend on many parameters; some of these parameter values are set by the tool operator, *e.g.*, chamber pressure, RF power, and gas flow, while others are internal to the condition of the chamber, *e.g.*, thin-film thickness on the chamber walls, or the amount of material etched. Plasma characteristics can be observed using *in situ* measurements, *e.g.*, via optical emission spectroscopy (OES) or residual gas analysis (RGA). How these measurements can be used to estimate the process parameters is the question that this paper is concerned with.

Currently, an etch process is arrived at through experiments which try to find the particular process parameters that result in the desired etch selectivity and profile. These conditions translate into a recipe which typically defines the set-points for gas flow, chamber pressure, RF power, and etching time. But since a chamber is usually stressed under high throughput, the chamber's internal parameters will vary even though the input parameters (as set by the operator) have not changed. This leads to results that are difficult to reproduce. The situation is further complicated by the fact that an unsuccessful etch may be caused by a mismatch between the process settings and the conditions actually produced by the machine's hardware, which is the case when there is, for example, a malfunctioning mass flow controller.

Research on modeling the relationship between process set-points and the desired etch profile (1-3) has greatly improved our understanding of the basic physical and chemical mechanisms in the RIE process. These models typically try to account for the plasma chemical reactions, the ionization and motion in an electric field, and the surface reaction kinetics (4). For monitoring and control application, however, it is unlikely that such models can be readily applied, since the particular reaction rates and cross sections are generally unknown.

Another approach has been to track the chamber conditions *in situ*. The idea here is to monitor the trajectory of one or two channels of an OES or RGA for the "perfect chamber," and then try to track that trajectory during the manufacturing process. Various artificial intelligence tools have been devised that, through rule-based expert systems, can detect significant deviation from the desired

trajectory (5). However, we know only of the work of Bolker *et al.* (6) where an *in situ* measurement, the RGA, was correlated to the process set-points.

In the present work we use *in situ* measurements in order to identify how the observed state of the plasma depends on each input process parameter (*e.g.* pressure, power, and/or mass flow/composition of gas). If it can be shown that the effect of each process parameter on the observed state of the plasma is uniquely identifiable, then discrepancies between the current and desired state of the plasma can be attributed to a particular parameter—facilitating monitoring and diagnosis. Furthermore, the system should be able to alert the operator to conditions that cannot be overcome by manipulation of the input parameters, such as leaks in the chamber, or significant formation of contaminations (*e.g.*, fluorocarbon, Teflon-like film) inside the chamber (7, 8).

This report describes our effort for identifying the effects of the process parameters on the *in situ* measurements of OES and RGA. Using principal component analysis (PGA), we show that changes in chamber pressure, RF power, gas mixture, and contamination levels result in complex but predictable clustering patterns when the OES and RGA measurements are projected onto their first few principal components. Our results indicate that for the limited set of experiments conducted, nearly all of the information in the OES and RGA measurements can be represented by these principal components. This way of representing the *in situ* measurements allows for very efficient modeling, since the characteristics of the plasma can be represented with only a few variables with almost no loss of information. Furthermore, relationships are much easier to visualize due to the significant reduction in dimensionality of the data set.

We explored both linear and nonlinear estimators for modeling the relationship between process parameters and these principal component projections. For the nonlinear estimator, we use a feed-forward, multi-layered neural network. The results reported here suggest that accurate estimates of process set-points and chamber contamination levels can be made with relatively simple neural network models, using the principal components of the OES and RGA measurements.

Experiments

In the present work, we are concerned with the following problem: A mixture of CHF_3/O_2 is vertically dissociated and ionized at a certain chamber pressure and RF

¹ Present address: Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

² Present address: IBM Advanced Semiconductor Technology Center, East Fishkill, New York 12533-6531.

Table I. Experiment numbers and the corresponding process parameter values.

Exp. no.	Power watts	Mix % O ₂	Press. mTorr	Thin film	Exp. no.	Power watts	Mix % O ₂	Press. mTorr	Thin film
1	200	3	25	9.4	37	200	3	50	20.2
2	300	3	25	9.5	38	300	3	50	20.2
3	350	3	25	9.6	39	350	3	50	20.3
4	200	9	25	9.8	40	200	9	50	20.5
5	300	9	25	9.8	41	300	9	50	20.5
6	350	9	25	9.9	42	350	9	50	20.6
7	200	15	25	10.0	43	200	15	50	20.6
8	300	15	25	10.0	44	300	15	50	20.7
9	350	15	25	10.1	45	350	15	50	20.7
10	200	3	50	10.2	46	200	3	75	20.9
11	300	3	50	10.3	47	300	3	75	21.2
12	350	3	50	10.4	48	350	3	75	21.4
13	200	9	50	10.5	49	200	9	75	21.8
14	300	9	50	10.5	50	300	9	75	22.0
15	350	9	50	10.6	51	350	9	75	22.2
16	200	15	50	10.7	52	200	15	75	22.4
17	300	15	50	10.7	53	300	15	75	22.5
18	350	15	50	10.7	54	350	15	75	22.6
19	200	3	75	11.0	55	200	3	50	30.0
20	300	3	75	11.3	56	300	3	50	30.0
21	350	3	75	11.7	57	350	3	50	30.0
22	200	9	75	12.0	58	200	9	50	30.2
23	300	9	75	12.2	59	300	9	50	30.2
24	350	9	75	12.3	60	350	9	50	30.3
25	200	15	75	12.6	61	200	15	50	30.4
26	300	15	75	12.6	62	300	15	50	30.4
27	350	15	75	12.7	63	350	15	50	30.4
28	200	3	25	19.3	64	200	3	75	30.5
29	300	3	25	19.4	65	300	3	75	30.8
30	350	3	25	19.5	66	350	3	75	31.2
31	200	9	25	19.7	67	200	9	75	31.7
32	300	9	25	19.8	68	300	9	75	31.8
33	350	9	25	19.9	69	350	9	75	32.1
34	200	15	25	19.9	70	200	15	75	32.4
35	300	15	25	20.0	71	300	15	75	32.4
36	350	15	25	20.1	72	350	15	75	32.5

power (no wafer is present in the chamber). From the measured optical emission and mass spectra, the task is to estimate the value of these parameters, as well as the amount of Teflon-like thin-film (7, 8) present on the surface of the chamber's interior.³ Build-up of this film (here after referred to as thin-film) changes the characteristics of the plasma, can significantly reduce the Si etch rate (9-11), and in extreme cases, can flake off and seriously contaminate the wafer.

For the following experiments, the RIE reactor was a 13.56 MHz flexible diode, with a 40 cm diameter lower electrode and a power range from 0 to 1000 W. Chamber pressure was controlled by an automatic throttle valve and measured by an MKS Baratron. The CHF₃/O₂ mix was determined by set-points on the MKS mass/flow controllers with a flow range of 0 to 100 sccm. The optical emission of the plasma was analyzed using a 512 diode array (150 groves/mm) EG+G PARC Model 1460, with a dual position stepper motor adjustment on the diffraction grating that allowed for viewing of emissions from 240 to 900 nm. Downstream mass spectra for masses 1 to 99 were taken on an Inficon Quadrex 200 RGA. Both the OES and RGA were interfaced to an IBM PS/2 Model 80 for efficient data collection and labeling. An Inficon IC6000 Quartz Crystal Microbalance (QCM), with the crystal mounted on the wall, was used as an indicator of the amount of thin-film in the chamber⁴ (12).

Chamber pressure was changed from 25 to 75 mTorr by increments of 25 mTorr. Gas flow was kept constant at 100 sccm, and the ratio of CHF₃/O₂ was changed from

³ Note that because no wafer was present in the chamber, for any given set of process set-points, the plasma's characteristics, and therefore the state of the process, did not vary significantly with respect to time. Although under conditions of high pressure and power one can readily observe thin-film build-up over several minutes, which results in a change in the species present in the optical and mass spectra, these measures remain essentially stationary for a period of a few tens of seconds after the power is applied. We assumed that the relationship between the *in situ* measurements and the process parameters is stationary.

⁴ Thin-film growth is reflected in a lowering of the frequency readout of the QCM. Unfortunately, without knowledge of the film density, this readout cannot be converted to an absolute measure of the thin-film thickness. However, as a relative measure, the QCM provides an accurate estimate of the change in thin-film thickness.

97%/3% to 85%/15% by increments of 6%. RF power was set at either 200, 300, or 350 W for each of the above permutations. To intentionally and quickly deposit thin-film, we set the input parameters at 50 mTorr, 100 sccm CHF₃, and 250 W. The QCM showed rapid build-up of thin-film on the chamber walls, which was accompanied by a drop in V_{DC}. We recorded the change in the QCM's output, and repeated the above set of experiments at medium and high levels of thin-film. In Table I we've listed the experiment numbers and the corresponding process parameter values.

The RGA and OES measurements were stored for each process set-point. We stored the RGA spectrum for masses 1 to 99, and the spectrum from 251.8 to 562.2 nm for the first 510 channels of optical emission (resolution of 0.61 nm), and from 549.2 to 860.9 for the second 510 channels. A total of 72 experiments were performed (see Table I).

Principal Component Analysis

Principal component analysis enables reduction of a data set while retaining most of its variation. The new set of variables are called the principal components derived from linear transformations of the original ones. They are statistically uncorrelated to each other and typically a small fraction of them (the first several) contain the majority of the variation (13). For the problem at hand, changes in the process parameters result in changes in the state of the plasma, which is estimated by the 99 variables of the measured mass spectrum and the 1020 variables of the measured optical emission spectrum. By visual inspection, it is perhaps possible to correlate patterns in parts of these spectra with changes in the process parameters. But because the number of recorded variables is so large, it is very difficult, for example, to see the effect of RF power on the RGA. Furthermore, building a model that estimates four process parameters from some 1119 different variables would be impractical for both efficiency and accuracy considerations.

By employing principal component analysis, we hope to find a small subspace spanned by a set of orthogonal vectors (the first few principal components), where projections of the original data can be effectively studied. If these

few uncorrelated variables reproduce most of the variation in all of the original variables, and if these variables are interpretable (*i.e.*, have a physical meaning), then the PCs give an alternate, and much simpler description of the data than the original variables. Our results demonstrate that this is in fact the case for the RIE process described in the previous section.

Computing the principal components amounts to calculating the eigenvectors in descending order of the covariance matrix of the variables in question. Let $\mathbf{y}_i = [s_i, q_i, p_i, c_i]^T$, represent the process parameter vector for experiment i , where s is the RF power, q is the percent O_2 in the CHF_3/O_2 mixture, p is the chamber pressure, and c is the thin-film contamination level as measured by a quartz crystal microbalance (QCM). Let $\mathbf{m}_i = [m_{i1}, m_{i2}, \dots, m_{i99}]^T$ and $\mathbf{o}_i = [o_{i1}, o_{i2}, \dots, o_{i1020}]^T$, represent the mass and optical spectra for experiment i , respectively. The task is to build a model which, given an observation of the state of the process, \mathbf{m}_i and \mathbf{o}_i , produces an estimate of the process's input and internal parameters, \mathbf{y}_i . Our approach is to use principal component analysis to find $\bar{\mathbf{m}}_i$ and $\bar{\mathbf{o}}_i$ from \mathbf{m}_i and \mathbf{o}_i where the dimensionality of $\bar{\mathbf{m}}_i$ and $\bar{\mathbf{o}}_i$ is much less than \mathbf{m}_i and \mathbf{o}_i , and then build linear and nonlinear models that describe \mathbf{y}_i as a function of $\bar{\mathbf{m}}_i$ and $\bar{\mathbf{o}}_i$.

The matrix $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{72}]^T$ is the set of process set-points visited in the experiments. Matrices $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{72}]^T$ and $\mathbf{O} = [\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_{72}]^T$ are the measured mass and optical emission spectra for the set of process set-points \mathbf{Y} . For example, O_{jk} is the value of the k th OES channel, for the j th process setting, \mathbf{y}_j . The principal components of the matrices \mathbf{M} and \mathbf{O} can be calculated by finding the eigenvectors of their respective covariance matrices; Matrices \mathbf{B} and \mathbf{C} are the sampled covariances of \mathbf{M} and \mathbf{O} , respectively

$$B_{jk} = \frac{1}{l-1} \sum_{i=1}^l (M_{ij} - \bar{M}_j)(M_{ik} - \bar{M}_k) \quad [1]$$

$$C_{jk} = \frac{1}{l-1} \sum_{i=1}^l (O_{ij} - \bar{O}_j)(O_{ik} - \bar{O}_k) \quad [2]$$

where $j, k = 1, 2, \dots, 99$ in Eq. [1], $j, k = 1, 2, \dots, 1020$ in Eq. [2], $l = 72$ (*i.e.*, the number of experiments), and \bar{M}_j and \bar{O}_j are defined as

$$\bar{M}_j = \frac{1}{l} \sum_{i=1}^l M_{ij} \quad [3]$$

$$\bar{O}_j = \frac{1}{l} \sum_{i=1}^l O_{ij} \quad [4]$$

The eigenvectors of \mathbf{B} and \mathbf{C} correspond to the principal components of the mass and optical spectra data, respectively. Let matrices $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_1}]^T$, and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n_2}]^T$, where $\mathbf{u}_i = [u_{i1}, u_{i2}, \dots, u_{i99}]^T$ and $\mathbf{v}_i = [v_{i1}, v_{i2}, \dots, v_{i1020}]^T$ and where \mathbf{u}_i and \mathbf{v}_i are the eigenvectors with the i th largest eigenvalue for the matrices \mathbf{B} and \mathbf{C} , respectively.

Then $\bar{\mathbf{M}}$ and $\bar{\mathbf{O}}$ are the projections of the mass and optical emission spectra onto their respective principal components, where $\bar{\mathbf{M}} = \mathbf{U} \times \mathbf{M}^T$ and $\bar{\mathbf{O}} = \mathbf{V} \times \mathbf{O}^T$, and $\bar{\mathbf{M}} = [\bar{\mathbf{m}}_1, \bar{\mathbf{m}}_2, \dots, \bar{\mathbf{m}}_{72}]$, and $\bar{\mathbf{O}} = [\bar{\mathbf{o}}_1, \bar{\mathbf{o}}_2, \dots, \bar{\mathbf{o}}_{72}]$. \bar{M}_{ij} , for example, is the projection of \mathbf{m}_i onto the j th principal component of \mathbf{M} .

Generally, one is interested in the first few principal components, *i.e.*, those eigenvectors with the largest eigenvalues, since they account for most of the variation in the data. A common practice for deciding on the number of principal components to keep is to compare the eigenvalues of the covariance matrix with the average of the diagonal elements of the covariance matrix (13). If the n th largest eigenvalue is smaller than this average value, then one keeps the first $n-1$ principal components. Alternatively, if the sum of the first n eigenvalues is a large percentage of the sum of the diagonal elements (trace) of the covariance matrix, and the $(n+1)$ th largest eigenvalue adds a negligible amount to this sum, then keep only the first n principal components.

Parameter Estimation

From the measured mass and optical emission spectra \mathbf{m}_i and \mathbf{o}_i , the task is to estimate the process parameters \mathbf{y}_i ($i = 1, 2, \dots, 72$). After the principal component analysis, we have a set of vectors in \mathbf{U} and \mathbf{V} which define the projections $\mathbf{m}_i \rightarrow \bar{\mathbf{m}}_i$, and $\mathbf{o}_i \rightarrow \bar{\mathbf{o}}_i$, respectively. Since the dimensionalities of $\bar{\mathbf{m}}$ and $\bar{\mathbf{o}}$ are small, we found it helpful to plot \mathbf{m}_i and \mathbf{o}_i in the space spanned by the first few principal components, and look for clustering patterns that might provide clues for building a model that relates the process parameters to the measured variables. From these clustering patterns, we observe (see the Results section) a linear relationship between chamber pressure and the mass spectrum data. Similarly, a linear model should also do a fair job in estimating the amount of thin-film in the chamber. Regression analysis is used to construct these linear estimators.

Further inspection of the principal component projections reveals that the relationship between some parameters, power, for example, and the measured variables is not linear. One option for building a nonlinear estimator is to implement a feedforward, multi-layered neural network, and estimate the desired mapping using the error back-propagation training algorithm. The reader is referred to Ref. (14) for a complete treatment on such networks and algorithms. We constructed two neural networks; for experiment i , the input to the first network is $\bar{\mathbf{m}}_i$, and after training, the output is an estimate of \mathbf{y}_i . The input to the second network is $\bar{\mathbf{o}}_i$, and after training, the output is an estimate of \mathbf{y}_i .

An additional advantage that is gained by studying the principal component projections is the ability to see the relative sensitivity of each sensor to changes in process parameters. In the Results section, the projections of the optical emission principal components suggest that OES is much more sensitive to changes in RF power than RGA. Based on this observation, we hypothesize that process parameter estimation could be aided if both the mass and optical emission projections are used simultaneously as inputs to the estimators. Therefore, a third network is built, where, for experiment i , input includes both $\bar{\mathbf{m}}_i$ and $\bar{\mathbf{o}}_i$, and after training, the output is an estimate of \mathbf{y}_i .

The neural networks are initially trained on the entire data set (72 experiments) and their performance (after 100,000 exposures) is compared with the performance of the linear models. For this comparison, the performance of each estimator is measured by finding the root-mean-squared error (RMSE) of each parameter over the entire data set (the same data set that the network was trained on, and the linear model was fitted to).

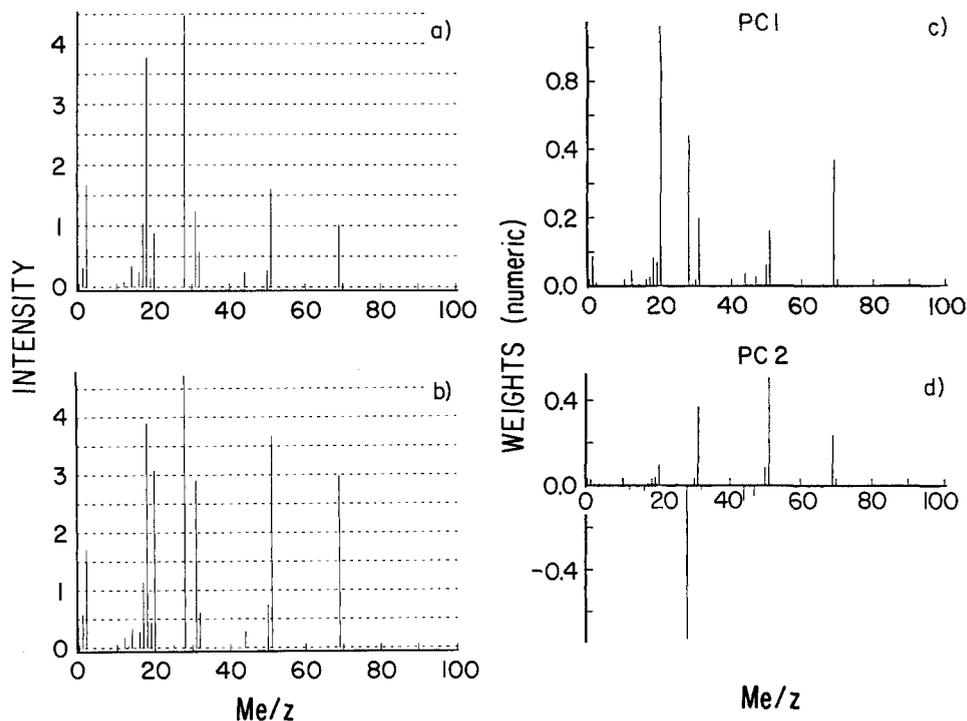
To determine the ability of the neural networks to generalize, we adopt the concept of cross-validation. The neural networks are initialized to weights of small, random values and then trained and tested on nonintersecting data sets. Each training set consists of 65 experiments, and each test set consists of 7 experiments. After exposure to a given training set, the RMSE of the network's estimation is determined for the test set. Eight pairs of such training and test sets are selected, and the degree of generalization of each network is estimated by the average RMSE over these eight test sets.

The data points in a given training set are also fitted to a linear model, and the performance of this linear estimator is gauged by the RMSE for the respective test set. There are eight training and test set pairs, so eight such linear estimators are developed for each parameter and for each sensor. The root-mean-squared errors of the linear estimators are then averaged for each sensor.

Results

Using Eq. [1] and [2], we calculated the covariance matrices \mathbf{B} and \mathbf{C} . The sum of the diagonal elements of \mathbf{B} , for example, is an estimate of the total variance present in the mass spectrum data. This value was 4.689 for \mathbf{B} , and 3.209×10^6 for \mathbf{C} . The vector \mathbf{u}_1 , which is the first principal component of \mathbf{M} , has an eigenvalue of 3.384. This means that projection of the mass spectrum data on \mathbf{u}_1 preserved 72% of the variance in the original data. The eigenvalues

Fig. 1. Mass spectra for two sets of parameter settings on the RIE reactor. (A) Chamber conditions set to 25 mTorr, 3% O₂, 97% CHF₃, 100 sccm, 200 W, and quartz crystal microbalance's (QCM) estimate of thin film at 9.7. (B) Chamber conditions set to 75 mTorr, 3% O₂, 97% CHF₃, 100 sccm, 200 W, and QCM at 11.0. (C) First principal component of the mass spectra. (D) Second principal component of the mass spectra.



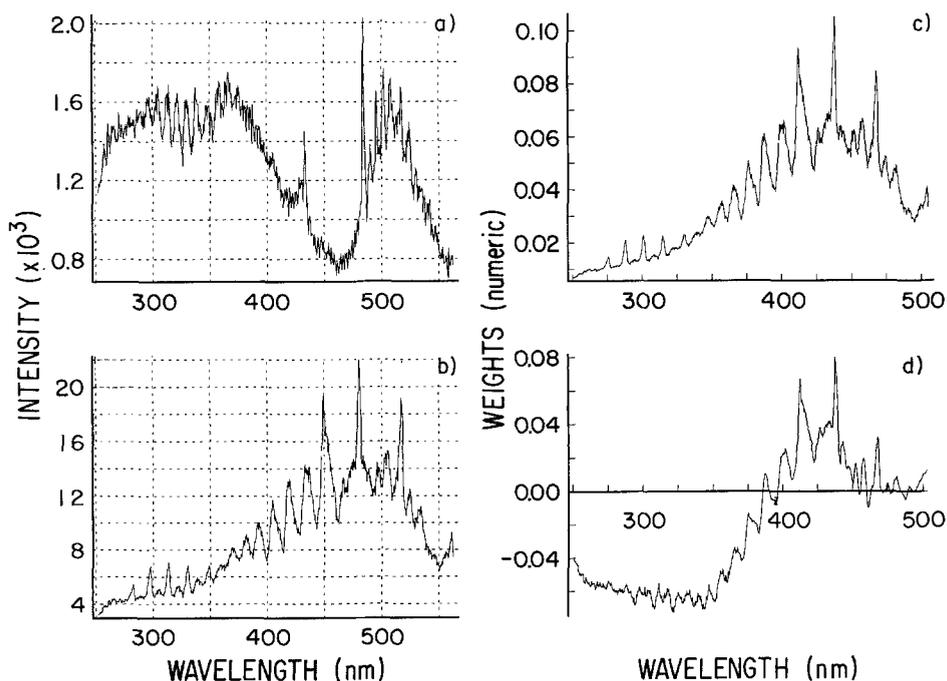
for u_2 , u_3 , and u_4 , are 1.137, 0.149, and 0.010, respectively. These first four principal components of \mathbf{M} together contain 99.8% of the information in \mathbf{M} (the first two PCs alone contain 96.4% of the information). In Fig. 1a and b we have plotted the original raw spectra from two different samples. Figure 1c) and d) show the first two principal components u_1 , and u_2 . In Fig. 1c the variables with the largest weights are those with the greatest variance on the first principal component. Channel 20, HF, 28 (CO and N), and channel 69, which is a fragment from CF₄, varied most for a given change in the external chamber set-points, and this is reflected in their relatively large weights.

The first four principal components of \mathbf{O} represent 99.9% of the information in \mathbf{O} , and the first two PCs alone represent 99.2% of this figure. In Fig. 2a and b we have plotted the raw data from the left spectra and in Fig. 2c and d that of v_1 , v_2 , the first and second principal components.

Although the effect of the process parameters on the RGA and OES may not be evident from Fig. 1 and 2, projection of the measured spectra on their principal com-

ponents allows us to see relationships more clearly: Figure 3 is a projection of \mathbf{M} onto its first and second principal components. Each point is identified with its experiment number from Table I. The lines connect experiments where pressure and thin-film levels are held constant, while power and the gas mixture are changed. The two parallel lines to the far left of the figure are for 25 mTorr, while the three parallel lines in the middle and far right are for 50 and 75 mTorr, respectively. Within each set of parallel lines, we have the condition where an experiment with the same pressure, power, and mixture was repeated under higher thin-film contamination levels. The effect of this contamination then is a translation of the point along the first PC. The first PC is highly sensitive to pressure, and almost has no sensitivity to changes in power or gas mixture. The second PC is sensitive to the gas mixture and power, but not to pressure or thin-film levels. From Fig. 3 we can see that if pressure and gas mixture were known (or could be estimated to a high degree of certainty), then the RF power and thin-film levels can be estimated as well.

Fig. 2. The "left" portion of the optical emission spectra. (A) Chamber conditions set to 25 mTorr, 3% O₂, 97% CHF₃, 100 sccm, 200 W, and quartz crystal microbalance's (QCM) estimate of thin film at 9.7. (B) Chamber conditions set to 75 mTorr, 15% O₂, 85% CHF₃, 100. (C) First principal component of the "left" optical spectra. (D) Second principal component of the "left" optical spectra. sccm, 350 W, and QCM at 12.7.



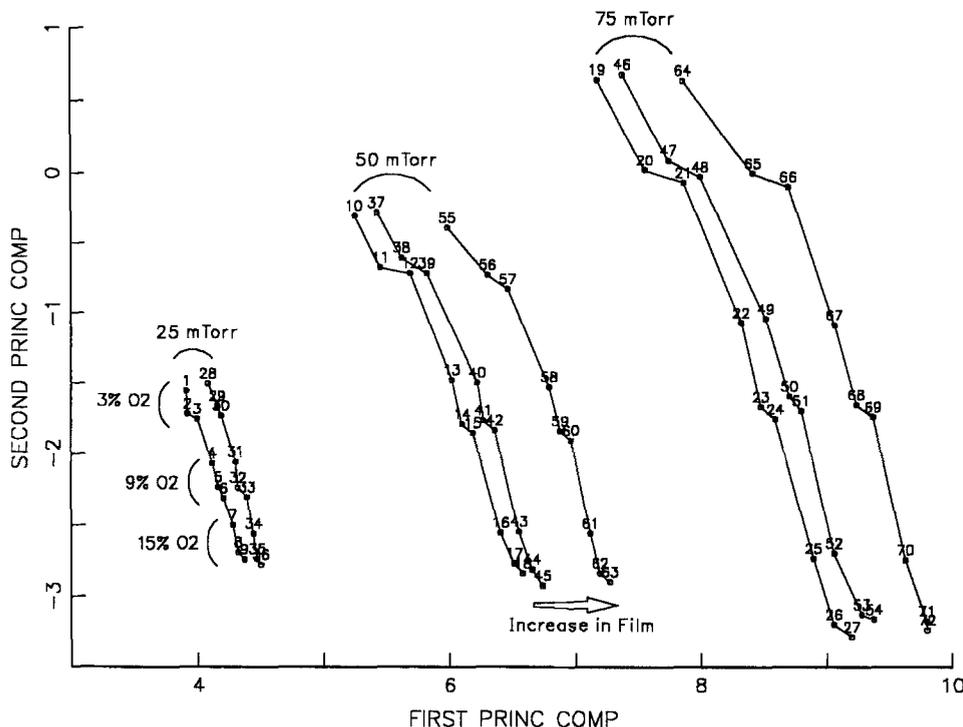


Fig. 3. Projection of the mass spectrum data on its first two principal components. The lines connect experiments which had the same pressure levels. The points are labeled according to their experiment number in Table I.

For estimation of thin-film levels independent of all other parameters, the fourth PC (plotted vs. the first PC in Fig. 4) appears to be the most effective, where we see sensitivity to thin-film levels but not to pressure.

Since changes of equal magnitude in pressure result in more or less equal changes of magnitude in the first PC, we expect that a linear model would do a fair job of estimating chamber pressures from the principal components of *M*. On the other hand, the clustering patterns in the first four PC projections appear to indicate that no PC separated the effects of power from the effects of change in gas mixture. Therefore, we expect that a linear model would not be appropriate for this estimation.

In Fig. 5 we've plotted the projection of *O* onto its first two principal components. The lines again connect experiments where chamber pressure is held constant. The points cluster based on what the chamber pressure is. The distribution of the points within each pressure cluster is

regulated by the RF power, gas mixture, and the thin-film levels. We can see from this figure that the OES is much more sensitive to changes in RF power than the RGA—the more or less straight lines of Fig. 5 have been replaced by jagged lines. There are three sets of jagged lines; the set in the upper left corner consists of two “parallel” jagged lines, the sets in the middle and lower right corner consist of three “parallel” jagged lines each. The effect of an increase in pressure is to translate a point along an axis running from the upper left corner of this figure to the lower right corner. The effect of a change in gas mixture is to translate a point along an axis perpendicular to the before mentioned axis. The effect of an increase in RF power is similar to the increase in pressure, but the translation is of a smaller magnitude.

Linear estimation.—Using regression analysis, we built linear estimators so that, for example, given \bar{m}_i , we can esti-

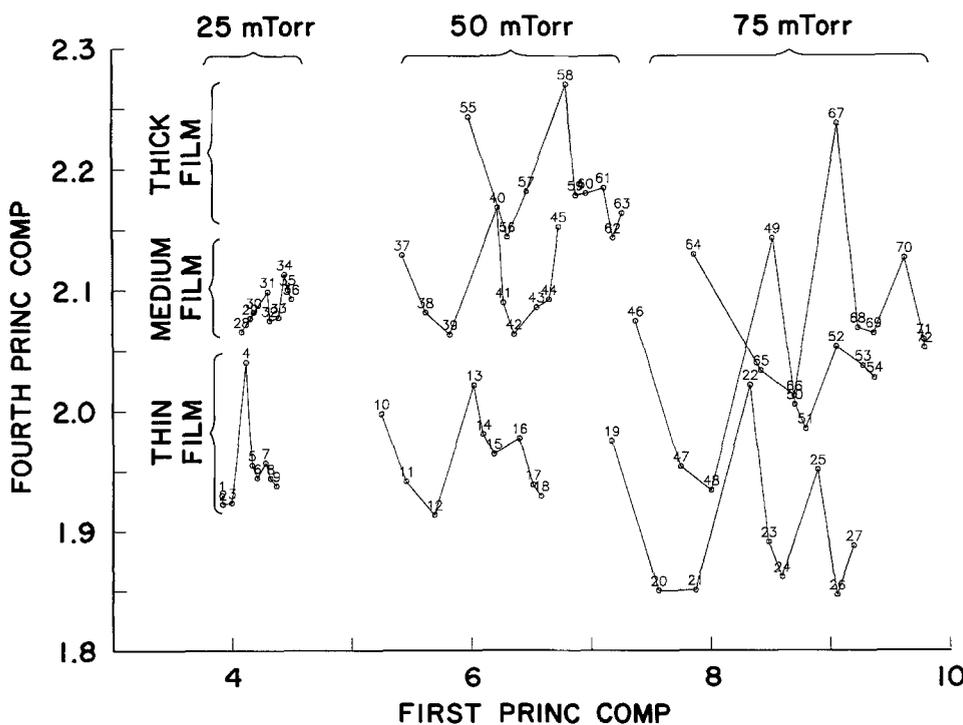
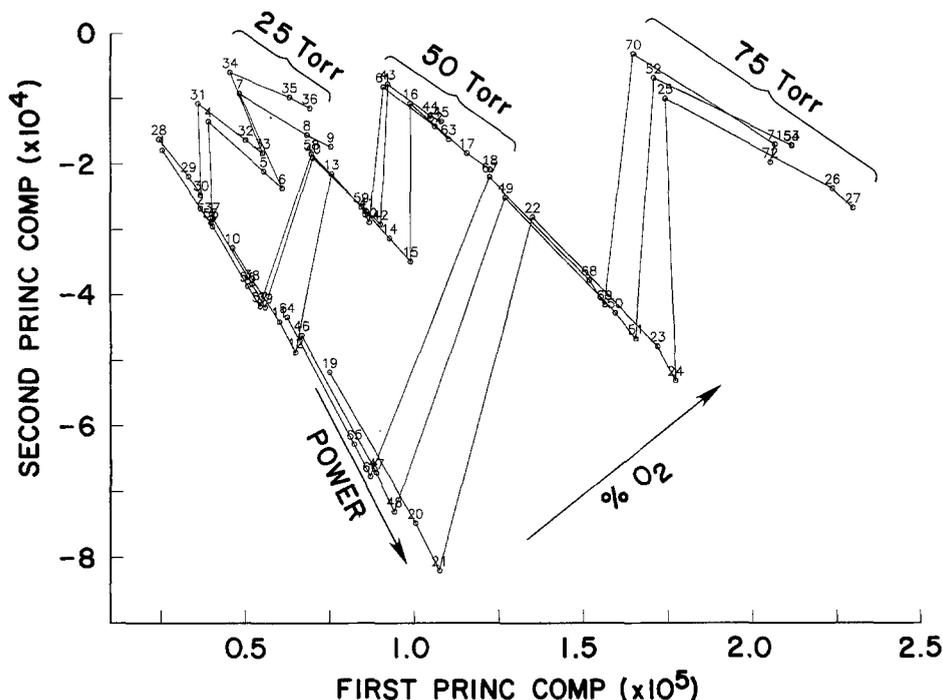


Fig. 4. Projection of the mass spectrum data on its first and fourth principal components. The lines connect experiments which had the same pressure levels. The points are labeled according to their experiment number in Table I.

Fig. 5. Projection of the optical emission spectrum on its first and second principal components. The lines connect experiments which had the same chamber pressure levels. The points are labeled according to their experiment number in Table I.



mate chamber pressure p_i . Three sets of such linear estimators are built (each set consists of four linear estimators, one for each parameter): $\bar{\mathbf{m}}_i \xrightarrow{\text{Lin}} \mathbf{y}_i$, $\bar{\mathbf{o}}_i \xrightarrow{\text{Lin}} \mathbf{y}_i$, and $[\bar{\mathbf{m}}_i, \bar{\mathbf{o}}_i] \xrightarrow{\text{Lin}} \mathbf{y}_i$. Each estimator is developed by using all the data points in the 72 experiments. The performance of each estimator is measured by its RMSE over the same 72 data points. This performance is listed in Table II.

From Table II, it appears that a linear model may be sufficient for estimating chamber pressure. On the other hand, a linear model appears to perform poorly for estimating power from the measured mass spectrum. The $\bar{\mathbf{o}}_i \xrightarrow{\text{Lin}} \mathbf{y}_i$ model does not perform as well as the $\bar{\mathbf{m}}_i \xrightarrow{\text{Lin}} \mathbf{y}_i$ model for estimating pressure, mixture, or thin-film levels, but it performs significantly better in predicting the applied power. If the chamber pressure could be reliably estimated, which appears to be the strength of the $\bar{\mathbf{m}}_i \xrightarrow{\text{Lin}} \mathbf{y}_i$ model, then Fig. 5 suggests that estimation of power from the optical emission data can be improved even further. A "sensor fusion" is beneficial if the resulting model proves more accurate than the model where each of the sensors were used independently. Table II suggests that when both sensors are used, performance of the linear estimators improve for all parameters, but especially so for power.

The output of each of the above mentioned linear models is a weighted sum of the projections of the spectra onto the principal components. The principal components may or may not have any relation to the process parameters, but the weights of the linear models that were developed can be used to produce a set of matched filters, whose outputs will vary linearly as a function of any given process parameter. For example, take a linear model such as the one which mapped $\bar{\mathbf{m}}_i \xrightarrow{\text{Lin}} p_i$, where p is chamber pressure. A filter is produced when the first four PCs of \mathbf{M} are weighted and summed according to the weights of the mentioned linear model. This filter has a special property, however; when chamber pressure is increased from p_1 to p_2 and the measured RGA is fed to the filter, its output

(which is a scalar) will show an increase that is approximately equal to $p_2 - p_1$ (the degree of accuracy depends on how small the RMSE of the original linear model was).

The matched filters for chamber pressure and thin-film levels from mass spectrum data are plotted in Fig. 6. As the thin-film in the chamber builds up, the output of the filter of Fig. 6b will increase proportionally. This is an encouraging result since the QCM is unlikely to be used in the manufacturing environment due to the invasive nature of its data gathering. The filter of Fig. 6b suggests that one may be able to use a transformation of the mass spectrum measurements as a "thin-film monitor."

The matched filters for power and thin-film levels from the optical emission spectrum are plotted in Fig. 7. Since a statistically significant linear relationship exists between the measured optical emission and chamber power and thin-film levels, the spikes of Fig. 7a and b point to species whose magnitude of optical emission vary proportionally as a function of these process parameters. These spikes are approximately at 483 and 656 nm, which are due to CO and H. From Fig. 7a and b we can hypothesize that an increase in chamber power is highly correlated to an increase in CO and H. From Fig. 7b and c we can hypothesize that an increase in thin-film levels is highly correlated to a decrease in H.

Nonlinear estimation using neural networks.—From the results of Table II, it appears that a linear transformation of the mass spectrum does a poor job of estimating the applied RF power and the O₂ mixture. As a point of comparison, we gauge the performance of a nonlinear transformation on the same data set. We choose a feedforward, multi-layered neural network for this estimation task. Three networks are trained using back-propagation (14). The performance of these estimators on the training set, after 100,000 exposures, is listed in Table III. The nonlinear mappings that are provided by the neural networks perform significantly better in estimating power and O₂ mixture, as compared to the $\bar{\mathbf{m}}_i \xrightarrow{\text{Lin}} \mathbf{y}_i$ linear estimator. In comparing $[\bar{\mathbf{m}}_i, \bar{\mathbf{o}}_i] \xrightarrow{\text{Lin}} \mathbf{y}_i$ to $[\bar{\mathbf{m}}_i, \bar{\mathbf{o}}_i] \xrightarrow{\text{NN}} \mathbf{y}_i$, the performance of the neural network is at least twice as accurate in estimating power, gas mixture, and thin-film levels. In fact, if we assume that the actual conditions of the chamber were within 1% of the process set-points, then the root-mean-squared error of the neural network model is within this noise level when both mass and optical emission spectra are used.

Unfortunately, we cannot infer that the neural network has generalized over the space where the data was

Table II. Root-mean-squared errors of the three linear estimators after regression analysis over the entire data set.

Parameter	Range	$\bar{\mathbf{m}}_i \xrightarrow{\text{Lin}} \mathbf{y}_i$	$\bar{\mathbf{o}}_i \xrightarrow{\text{Lin}} \mathbf{y}_i$	$[\bar{\mathbf{m}}_i, \bar{\mathbf{o}}_i] \xrightarrow{\text{Lin}} \mathbf{y}_i$
Power (watts)	200-350	38.6	14.1	9.6
Mixture (% O ₂)	3-15	1.44	1.77	1.22
Pressure (mTorr)	25-75	0.81	2.68	0.65
Thin-film	9-32	1.6	2.4	1.4

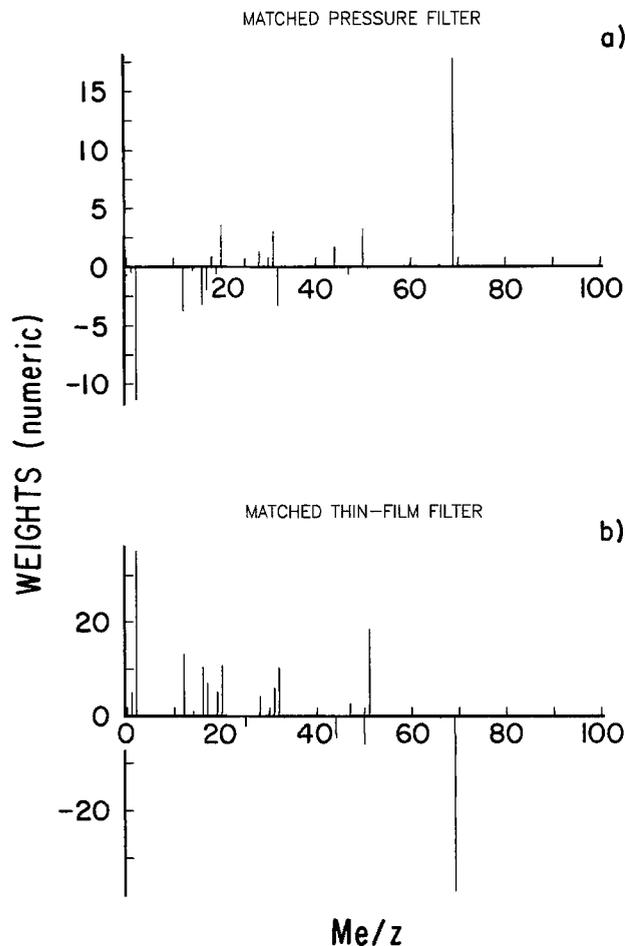


Fig. 6. The matched pressure and thin-film filters for the mass spectrum data. For example, given an increase in chamber pressure of amount Δp , the projection of the measured RGA onto the pressure filter will result in a scalar whose value will increase by an amount equal to Δp .

sampled. This is due to the fact that the number of sampled data points are small compared to the space that has been spanned, and therefore a "test set" has not veri-

Table III. Root-mean-squared errors of the three neural-network-based estimators after training on all the experiments.

Parameter	Range	$\bar{m}_i^{NN} \rightarrow y_i$	$\bar{o}_i^{NN} \rightarrow y_i$	$[\bar{m}_i, \bar{o}_i]^{NN} \rightarrow y_i$
Power (watts)	200-350	22.9	10.6	2.4
Mixture (% O ₂)	3-15	0.70	0.93	0.17
Pressure (mTorr)	25-75	0.85	1.59	0.63
Thin-film	9-32	1.5	1.9	0.6

fied the degree of generalization of either the linear or the neural net mappings. Nevertheless, since the neural network was not able to estimate pressure levels significantly better than the linear models, and since the RMSE of the linear models were relatively small, it appears that the relationship between the measured mass spectrum and chamber pressure is linear. Furthermore, the performances of $\bar{m}_i^{NN} \rightarrow y_i$ and $\bar{o}_i^{NN} \rightarrow y_i$ suggest that estimating chamber power from either mass or optical emission spectra is very difficult, even with a nonlinear model.

To gauge the generalization ability of the estimators, pairs of training and test sets are selected from the original 72 experiments, and the performance of the linear and neural network-based estimators are compared on the test set after exposure to the training set (the training and test sets are mutually exclusive). The average RMSE of the linear estimators and the neural networks for the test sets are listed in Table IV. Performances of the linear estimators are only slightly worse on the test set (Table IV), when compared to the case when models were fitted to the entire data set (Table II). For the case of pressure estimation, performance of the linear model is quite excellent, providing further evidence that chamber pressure and mass spectrum are linearly related. In predicting power and gas mixture of the test set, performance of $\bar{m}_i^{NN} \rightarrow y_i$ is significantly better than $\bar{m}_i^{Lin} \rightarrow y_i$, suggesting that the nonlinear transformation learned by the neural network over the training set is sufficient for a reasonably accurate estimation of the test data. The RMSE of the neural net for estimation from optical emission data is better than the linear estimator for all parameters. However, this result indicates that, except for power, all other parameters can be more accurately estimated using the mass rather than the optical spectrum measurements.

When the optical and mass spectrum data are fused, the RMSE of the neural network is twice as good as the linear model for estimating power, and four times as good for es-

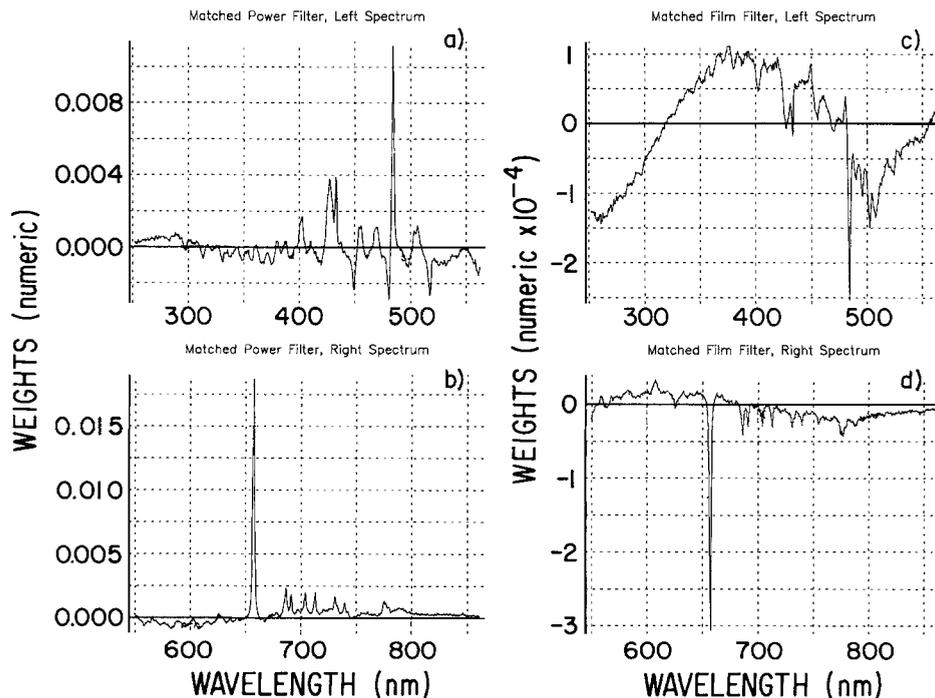


Fig. 7. (a, b) The matched power filter for the optical emission data. The peaks in the power filter are likely to be due to CO, in the left spectrum, and H, in the right spectrum, respectively. (c, d) The matched thin-film filter for the optical emission data.

Table IV. Average root-mean-squared errors of the estimators for the test sets.

Parameter	$\bar{m}_i^{\text{Lin}} \rightarrow y_i$	$\bar{m}_i^{\text{NN}} \rightarrow y_i$	$\bar{o}_i^{\text{Lin}} \rightarrow y_i$	$\bar{o}_i^{\text{NN}} \rightarrow y_i$	$[\bar{m}_i, \bar{o}_i]^{\text{Lin}} \rightarrow y_i$	$[\bar{m}_i, \bar{o}_i]^{\text{NN}} \rightarrow y_i$
Power	40.4	26.9	16.8	14.1	11.2	5.31
Mixture	1.46	0.71	1.91	1.54	1.31	0.30
Pressure	0.85	1.82	3.01	2.30	0.70	0.91
Thin-film	1.82	1.89	2.92	2.44	1.68	1.18

timating the gas mixture. For estimating chamber pressure from the fused data, the RMSE of both the linear and the neural network model are within 2% of the highest pressures tested, suggesting that pressure can be very accurately estimated from the optical and mass spectra of the plasma using a linear estimator, and that the neural network is behaving essentially like a linear estimator for pressure. Results of Table IV also indicate that thin-film levels can be estimated reasonably accurately (to within 4% of the highest values tested) with the neural network when both spectra are available.

Discussion

The ability to estimate process parameters from the optical emission or mass spectrum measurements is essential to real-time monitoring and diagnosis of RIE. For example, if there is a leaky valve, or an inappropriate amount of RF power is being applied, then the estimators should be able to indicate that there is something wrong and point to a specific process parameter as the probable cause. This second ability, *i.e.*, being able to decipher the effect of one process parameter independent of another, is crucial for a controller which attempts to match a predefined "state trajectory" for its process, where in our case, the state of this process is observed by the measured mass and optical emission spectra.

Building this controller requires defining the relationship between the measured variables, *i.e.*, the bulk spectra, and the input parameters of the process, *i.e.*, pressure, power, and gas mixture. But the state of the process also depends on a set of internal parameters, *i.e.*, thin-film contamination levels, whose values cannot be manipulated by the controller. The problem that we are concerned with is how to estimate the process's input and internal parameters from the *in situ* measurements.

Through principal component analysis, we show that nearly all of the information in each of the original spectra can be represented by a set of four new variables. These variables are the projections of the measured spectra onto the first four principal components of each data set. This transformation produces a new data set whose complexity (as measured by its dimensionality) is minimized, while its information content (as measured by its variance) is almost identical to that of the original data set.

The plots of the new data set show a remarkably regular clustering pattern for both the mass and optical emission spectra. The points cluster as a function of pressure, and the pattern within each cluster is regulated by the other process parameters.

The relatively good performance of the linear models on predicting pressure and thin-film levels from the mass spectrum, and power and thin-film levels from the optical emission spectrum, suggests that a matched filter can be developed for each of the above transformations (Fig. 6 and 7). The output of each such filter is a scalar whose change is equal to the change in one of the process parameters. These filters also convey information on how the process parameters relate to chemical and physical phenomenon. For example, from Fig. 7a and b we can forward a conjecture that changes in applied RF power can be estimated by the amount of optical emission produced by CO and H.

Since the performance of the linear models on predicting power and mixture is generally poor, we developed a set of nonlinear estimators by constructing multi-layered neural networks, and trained and tested both estimators on pairs of disjoint "training" and "test" sets. The perform-

ance of the neural network-based estimators is significantly better than that of the linear estimators (see Table IV) for power and gas mixture, for all three sensor modalities tested. However, these results also illustrate that it would be difficult to accurately estimate chamber power from the mass spectrum measurements, even with a nonlinear model, and that for estimating chamber pressure from the mass spectrum data, a nonlinear model is not expected to perform better since the performance of the linear estimator is extremely good. Furthermore, from Table IV we can conclude that estimating chamber pressure, gas mixture, and thin-film levels can be accomplished much more accurately when the mass spectrum rather than optical emission data is used (note that the linear model that used mass spectrum data outperformed both the linear and the neural network models which used the optical emission spectrum). However, since the output of the OES is more sensitive to power but less sensitive to pressure than the RGA, we suggest that use of both of these sensors can result in a model whose performance is better than when each sensor is used independently. Results listed in Table IV appear to confirm this idea; when both sensors are used as inputs to the neural network, thin-film level can be estimated to within 4% of its largest value, while power, mixture, and pressure can be estimated to within 2% of their largest values.

The ability to estimate thin-film contamination levels from the mass and optical emission spectra, as shown in this report, is an important result because it has been shown that growth of this material leads to significant reductions in the Si etch rate (9), and because the current invasive technique as used by the QCM is considered to be unusable in the manufacturing floor.

Manuscript submitted April 22, 1991; revised manuscript received Oct. 10, 1991.

IBM's Research Division assisted in meeting the publication costs of this article.

REFERENCES

1. D. J. Economou, S. K. Park, and G. D. Williams, *This Journal*, **136**, 188 (1989).
2. D. B. Graves, in "Plasma Processing," G. S. Mathad, G. C. Schwartz, and R. A. Gottscho, Editors, PV 87-6, p. 267, The Electrochemical Society Softbound Proceedings Series, Pennington, NJ (1987).
3. P. Schoenborn, R. Patrick, and H. P. Baltes, *This Journal*, **136**, 199 (1989).
4. J. I. Ullacia and J. P. McVittie, *J. Appl. Phys.*, **64**, 1488 (1989).
5. S. B. Dolins, A. Srivastava, and B. E. Flichbaugh, *IEEE Trans Semiconductor Manuf.*, **1**, 23 (1988).
6. B. F. Bolker, T. C. Tison, and T. S. Latos, *J. Vac. Sci. Technol.*, **18**, 328 (1981).
7. M. M. Millard and E. Kay, *This Journal*, **129**, 160 (1982).
8. R. D'Agostino, F. Cramarossa, and S. De Benedictis, *Plasma Chem. Plasma Process.*, **2**, 213 (1982).
9. G. S. Oehrlein and H. L. Williams, *J. Appl. Phys.*, **62**, 662 (1987).
10. M. A. Jaso and G. S. Oehrlein, *J. Vac. Sci. Technol. A*, **6**, 1397 (1988).
11. C. Cardinaud, A. Rhounna, G. Turban, and B. Grolleau, *This Journal*, **135**, 1472 (1988).
12. J. W. Coburn, H. F. Winters, and T. J. Chuang, *J. Appl. Phys.*, **48**, 3532 (1977).
13. I. T. Jolliffe, "Principal Component Analysis," Springer-Verlag, New York (1986).
14. J. L. McClelland, D. E. Rumelhart, and the PDP Research Group, "Parallel Distributed Processing," MIT Press, MA (1986).