APPLICATION OF REDUCED-RANK MULTIVARIATE METHODS TO THE ANALYSIS OF SPATIAL UNIFORMITY OF SILICON WAFER ETCHING

Pratik Misra[1], Michael Nikolaou[1], Andrew D. Bailey III[2]

1. Dept. of Chemical Engineering, University of Houston
Houston, TX 77204-4004
2. Lam Research Corporation
4650 Cushing Parkway Fremont CA 94538

Abstract: We provide a smooth introduction to reduced-rank multivariate analysis, and show how it can be used to monitor images of etched silicon wafers. Results from two industrial case studies are presented and discussed.

Keywords: Plasma etching, principal component analysis, singular value decomposition, image processing.

1 INTRODUCTION

Spatially uniformity is necessary for high yields in a number of crucial processes of the semiconductor manufacturing industry, such as etching or deposition of thin films and chemical-mechanical planarization (CMP). In plasma etching, good spatial uniformity is the result of both appropriate design of etching tools as well as development of successful recipes. For either of these tasks, the designer or operator must be able to assess spatial uniformity characteristics, understand similarities and differences between tools or recipes, and apply criteria for the monitoring of spatial uniformity from tool to tool or run to run. Because uniformity is usually expressed in terms of a single number (e.g., 3σ/[average etch depth]) very different spatial uniformity profiles may result in the same numerical value of uniformity (Figure 1), thus masking important information that could be useful in a number of ways related to tool or recipe performance.

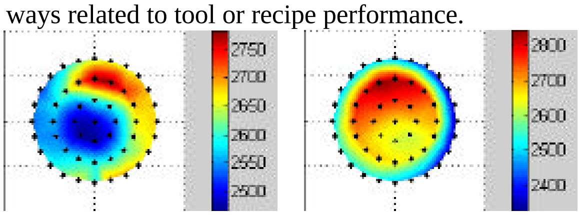

Figure 1 – Etch rate profiles on 300-mm wafer surface, interpolated over 49 measurement points (black dots). Both wafers correspond to virtually the same numerical uniformity value, but exhibit very different etch patterns.

In this presentation we provide a brief tutorial overview of the fundamentals of reduced-rank analysis, a topic that has found widespread use in chemical engineering. We show how it can be applied to the analysis, comparison, monitoring, and control of images corresponding to etch patterns of silicon wafers. Similar rank reduction techniques, especially Karhunen-Loeve (KL) transform, have been used to study spatiotemporal patterns on catalyst surface by Krischer et al.(1993) and in analyis and control of paper machines by Rigopoulos and Arkun (1996).

2 COMPRESSION OF COLLINEAR DATA VIA SVD

2.1 Basic case: Deterministic signals, no noise

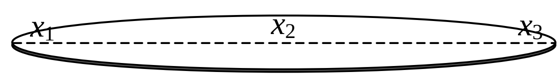

Figure 2 – Etch rate measurement points

An unrealistic but instructional example setting

Suppose that etch rates, x_1, x_2, x_3 are exactly measured at three points (edge/center/edge) along the diameter of a wafer, as shown in Figure 2. We want to know if the etch profiles are similar and etching process consistent.

Noiseless data are collected

Note that, for now, the data are assumed to be exact, i.e. *there is no measurement noise*. A set of data collected is shown in the matrix **X** below, and Figure 3.

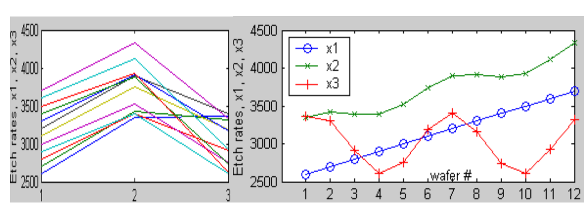

Figure 3 – Hypothetical etch rate profiles for 12 wafers (left) and Hypothetical local etch rates vs. wafer # (right).

$$\mathbf{X} = \overbrace{\begin{bmatrix} 2600 & 3348 & 3361 \\ 2700 & 3423 & 3311 \\ 2800 & 3392 & 2907 \\ 2900 & 3393 & 2609 \\ 3000 & 3527 & 2757 \\ 3100 & 3745 & 3182 \\ 3200 & 3900 & 3400 \\ 3300 & 3919 & 3163 \\ 3400 & 3882 & 2740 \\ 3500 & 3934 & 2614 \\ 3600 & 4118 & 2927 \\ 3700 & 4327 & 3324 \end{bmatrix}}^{x_1 \quad x_2 \quad x_3} \hat{=} [\mathbf{x}_1\, \mathbf{x}_2\, \mathbf{x}_3] \qquad (1)$$

Data collinearity and computation of matrix rank

Are the variables x_1, x_2, x_3 linearly dependent? i.e. is there a nonzero vector $\mathbf{a} \hat{=} [a_1\; a_2\; a_3]^T$ such that

$$a_1 x_1 + a_2 x_2 + a_3 x_3 = 0 \Leftrightarrow \mathbf{x}^T \mathbf{a} = 0 \qquad (2)$$

If so, the data satisfy the relationship (**model equation**)

$$a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + a_3 \mathbf{x}_3 = 0 \Leftrightarrow \mathbf{Xa} = \mathbf{0} \text{ for } \mathbf{a} \neq \mathbf{0} \qquad (3)$$

A numerically robust method to check whether eqn. (3) is valid is the singular value decomposition (SVD). Detailed treatment of SVD can be found in a number of standard texts such as Horn and Johnson (1985). In SVD a matrix of rank r is decomposed as

$$\mathbf{X} = \underbrace{s_1 \mathbf{u}_1}_{\text{"score"1}} \underbrace{\mathbf{v}_1^T}_{\text{"loading"1}} + \cdots + \underbrace{s_r \mathbf{u}_r}_{\text{"score"}r} \underbrace{\mathbf{v}_r^T}_{\text{"loading"}r} \equiv \sum_{i=1}^{r} s_i \mathbf{u}_i \mathbf{v}_i^T \equiv \sum_{i=1}^{r} \mathbf{y}_i \mathbf{v}_i^T \qquad (4)$$

Application of SVD (e.g. in Matlab ®) to the data matrix **X**, eqn. (4) yields that the rank of **X** is 2, and the matrix **X** can be decomposed as

$$\mathbf{X} = \underbrace{19973 \begin{bmatrix} -0.26882 \\ -0.2727 \\ -0.26381 \\ -0.25876 \\ -0.26976 \\ -0.29076 \\ -0.30431 \\ -0.30144 \\ -0.2919 \\ -0.29302 \\ -0.30999 \\ -0.32996 \end{bmatrix}_{12}^{1}}_{\text{"score"1},\, \mathbf{y}_1} \underbrace{\begin{bmatrix} -0.54865 & -0.65112 & -0.52443 \end{bmatrix}_1^3}_{\text{"loading"1},\, \mathbf{v}_1^T}$$

$$+ \underbrace{1233.7 \begin{bmatrix} 0.53687 \\ 0.44752 \\ 0.14112 \\ -0.10007 \\ -0.068095 \\ 0.1339 \\ 0.20911 \\ 0.005124 \\ -0.31245 \\ -0.44865 \\ -0.31508 \\ -0.13062 \end{bmatrix}_{12}^{1}}_{\text{"score"2},\, \mathbf{y}_2} \underbrace{\begin{bmatrix} -0.52217 & -0.22301 & 0.82317 \end{bmatrix}_1^3}_{\text{"loading"2},\, \mathbf{v}_2^T}$$

$$+ \underbrace{\mathbf{0}^{12\times 1}}_{\text{"score"3},\, \mathbf{y}_3} \underbrace{\begin{bmatrix} -0.65293 & 0.72548 & -0.21764 \end{bmatrix}_1^3}_{\text{"loading"3},\, \mathbf{v}_3^T} \qquad (5)$$

The above eqn. (5) implies that each row of the matrix **X** can be written as a linear combination of the row vectors *loading*1 and *loading*2, i.e.

$$\underbrace{[x_1\; x_2\; x_3]}_{\mathbf{x}^T} = \underbrace{y_1}_{\text{"score"1}} \underbrace{\boxed{\mathbf{v}_1^T}^{3}}_{\text{"loading"1}} + \underbrace{y_2}_{\text{"score"2}} \underbrace{\boxed{\mathbf{v}_2^T}^{3}}_{\text{"loading"2}} \qquad (6)$$

Because **V** is orthonormal, eqn. (6) yields the sought eqn. (2), i.e.

$$\mathbf{x}^T \mathbf{v}_3 = 0 . \qquad (7)$$

Loadings can be interpreted as basic shapes that can be used to represent the raw data

Note that the row vectors *loading*1 and *loading*2 in eqn. (6) are **the same** for all rows of data triplets x_1, x_2, x_3; they appear to be related to the system and not to any individual wafer. Therefore, *loading*1 and *loading*2 can be *interpreted* as two basic shapes (Figure 4), whose linear combination (sum weighted by score entries) can produce any of the 12 measured shapes.

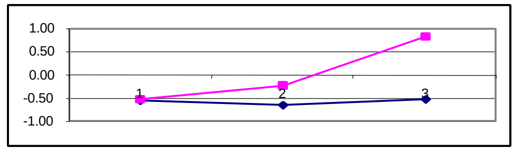

Figure 4 – Loadings, eqn. (5). The two shapes attempt to capture the curvature in the etch rate profile.

Monitoring scores gives a complete picture of the data

It follows from the preceding discussion that one can simply observe the scores (compressed data, values of *principal components* – hence PCA), to capture all information about the original data. In other words, instead of looking at

Figure 3, one can look at Figure 5.

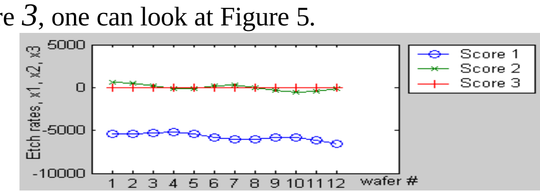

Figure 5 – Scores for the data in Figure 2, according to eqn. (5). Note that Score 3 is identically 0, which is precisely the equation sought in eqn. (2).

$rank(\mathbf{X}) = 2$ implies data points fall on a plane

Figure 6 shows 3-D plots of the data from two different viewpoints. The second viewpoint clearly shows that data fall on a plane.

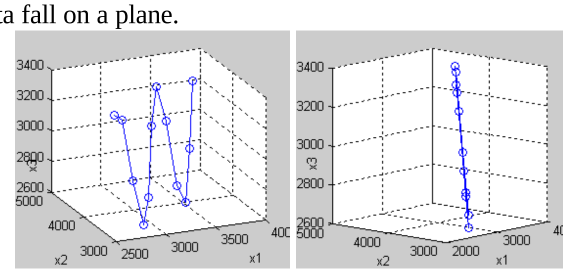

Figure 6 – 2-D world in 3-D data ("collinearity").

Loadings can also be thought of as weights used to relate original data to scores (compressed data)

If the score vectors $\mathbf{y}_1$, $\mathbf{y}_2$ are thought of as corresponding to two new variables, y_1, y_2, then y_1, y_2 are related to x_1, x_2, x_3 as follows: Because the loadings are orthonormal, we can post-multiply eqn (4). by $\mathbf{v}_j$ to get

$$\mathbf{X}^{m\times n}\mathbf{v}_j^{n\times 1} = \underbrace{s_j\mathbf{u}_j^{m\times 1}}_{\text{"score" } j} \hat{=} \mathbf{y}_j \tag{8}$$

or, row by row,

$$y_j = [x_1 \cdots x_n]\mathbf{v}_j \equiv \mathbf{x}^T\mathbf{v}_j = \mathbf{v}_j^T\mathbf{x} \tag{9}$$

or, in vector/matrix form,

$$\mathbf{y} = \mathbf{V}^T\mathbf{x} \Leftrightarrow \mathbf{x} = \mathbf{V}\mathbf{y} \tag{10}$$

(The new variables $\mathbf{y}$ are also called ***principal components***, see section 2.3.)

Thus, for this particular example we get, using eqn.(9), that the two nonzero score variables are

$$y_1 = [x_1\ x_2\ x_3]\begin{bmatrix}-0.54865\\-0.65112\\-0.52443\end{bmatrix},\quad y_2 = [x_1\ x_2\ x_3]\begin{bmatrix}-0.52217\\-0.22301\\0.82317\end{bmatrix} \tag{11}$$

and that the last score variable should be trivially equal to zero, i.e.

$$y_3 = [x_1\ x_2\ x_3]\begin{bmatrix}-0.65293\\0.72548\\-0.21764\end{bmatrix} = 0 \tag{12}$$

which is the same as eqn. (7).

This gives us the *second interpretation of loadings*: They are the vectors of coefficients by which we weight the original variables in linear combinations that produce a new set of variables (the "scores").

The preceding findings about **X** *can be used to monitor the system*

If the system etches subsequent wafers in the same way, it is reasonable to expect that data points (x_1, x_2, x_3) will be produced that are related as before, i.e. by eqn. (2). That means, equivalently, that if one first constructs 2 new variables y_1, y_2 in terms of eqn. (9) then the value of the *residual error* (cf. eqn. (6))

$$\mathbf{e}^T \hat{=} \underbrace{[x_1\ x_2\ x_3]}_{\mathbf{x}^T} - \left(\underbrace{y_1}_{\text{"score"1}} \overbrace{\boxed{\mathbf{v}_1^T}}^{3}{}_{\text{"loading"1}} + \underbrace{y_2}_{\text{"score"2}} \overbrace{\boxed{\mathbf{v}_2^T}}^{3}{}_{\text{"loading"2}} \right) \tag{13}$$

$$= (\mathbf{x} - \mathbf{P}\mathbf{P}^T\mathbf{x})^T$$

for each new data triplet should be equal to zero, or, equivalently,

$$\|\mathbf{e}\|^2 \hat{=} \mathbf{e}^T\mathbf{e} = 0 \Leftrightarrow \mathbf{x}^T(\mathbf{I} - \mathbf{P}\mathbf{P}^T)\mathbf{x} = 0 \tag{14}$$

where the matrix **P** consists of the first r columns of **V**. (The reason for using eqn. (14), instead of simply $\mathbf{e} = \mathbf{0}$, is that it can easily be extended to handle noisy data, as will be shown below).

Consider now the new data shown in Figure 7 .

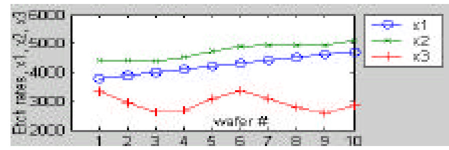

Figure 7 – Data set from 10 new wafers.

Applying the test of eqn. (14) to the data shown above yields the results of Figure 8. It is clear that two data points (#7 and #8) do not fall on the zero line as they should. These points indicate that the behavior of the system that etched these wafers is different from before.

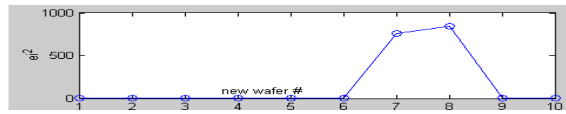

Figure 8 – (Errors)^2 for 10 new data sets, Figure 7.

2.2 Noisy signals

SVD on the noisy counterpart of **X** *reveals similar relationship among* x_1, x_2, x_3.

Table 1 – Noisy data

#	x_1	x_2	x_3
1	2585	3373	3353
2	2874	3586	3374
3	2809	3311	2861
4	2759	3355	2562
5	3175	3602	2763
6	3071	3753	3258
7	3424	3933	3486
8	3368	3974	3263
9	3526	3887	2709
10	3523	4034	2735
11	3546	4209	2910
12	3666	4381	3417

If measurements of x_1, x_2, x_3 are obtained with measurement noise as shown in the data of Table 1, SVD on the data of Table 1 yields singular values of 20219, 1206.5, 226.15 (cf. eqn.(5)). The eigenvalues (singular values squared) are shown in Figure 9. The smallest singular value is two orders of magnitude smaller than the largest one, indicating that it is probably equal to zero. But the second singular value is also an order of magnitude smaller than the largest singular value. Is it really non-zero or zero? How many singular values should be retained? What is the underlying rank of the data? How many singular values of **X** are really nonzero?

Let us call the noiseless data matrix Ξ and

$$\mathbf{X} = \Xi + \mathbf{E} \tag{15}$$

where **E** is a matrix that contains measurements errors. Note that for the data in Table 1

$$rank(\mathbf{X}) = 3 > rank(\Xi) = 2 \tag{16}$$

The singular values of **X**, $\sigma_{\mathbf{X}}$, can be bounded by bounds such as (Horn and Johnson, 1985):

$$|s_i(\mathbf{X}) - s_i(\Xi)| \le \|\mathbf{E}\|_{i2} = s_{\max}(\mathbf{E}) \tag{17}$$

Two simple criteria for detecting the number of essentially nonzero singular values of **X** are

(a) visual inspection of the singular value plot such as in Figure 9, and

(b) fidelity of reconstruction of the original data in **X**

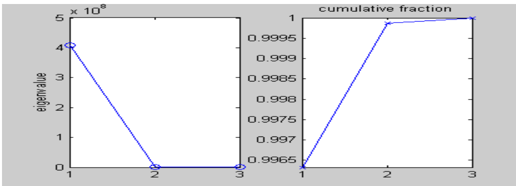

Figure 9 – Squared singular values (eigenvalues) for data in Table 1. (a) individual, (b) cumulative.

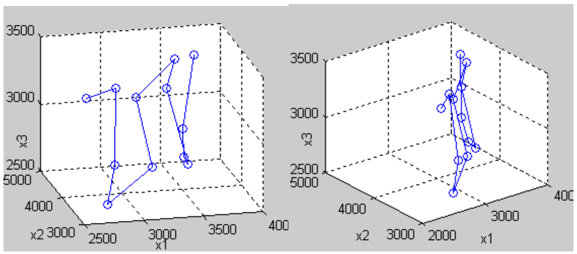

Figure 10 – 2-D world in noisy 3-D data.(cf. Figure 6).

Singular values quantify the goodness of data fit by a matrix of reduced rank

If only a "small" number of principal components is important, what is the best estimate of Ξ (with rank $r<n$) given the data in **X**? Answering this question will allow us to construct scores and loadings, and to monitor the system, in the same way as we did in the noiseless case. The difference is that what should have been ideally zero errors, eqn. (13) should now be "small" (more in the sequel).

To find the best estimate $\hat{\Xi}$ of Ξ given **X** we can minimize the distance between Ξ and **X**, i.e. find

$$\min_{rank(\Xi)=r<n} \|\mathbf{X}-\Xi\| \qquad (18)$$

When the norm in (18) is *induced 2 norm* or *Frobenius* norm, the solution is given by SVD as

$$\hat{\Xi} = \sum_{i=1}^{r} s_i \mathbf{u}_i \mathbf{v}_i^T \qquad (19)$$

Moreover, the optimal difference can be shown to be

$$\min_{rank(\Xi)=r<n} \|\mathbf{X}-\Xi\|_{i2} = \|\mathbf{X}-\hat{\Xi}\|_{i2} = s_{r+1} \qquad (20)$$

and

$$\min_{rank(\Xi)=r<n} \|\mathbf{X}-\Xi\|_F = \|\mathbf{X}-\hat{\Xi}\|_F = \sqrt{\sum_{i=r+1}^{n} s_{r+i}^2} \qquad (21)$$

Note that the singular vectors (loadings) of **X** could be very different from the singular vectors (loadings) of Ξ (Stewart, 1991). Figure 11, shows loadings for **X**. Comparison with Figure 4 shows little difference.

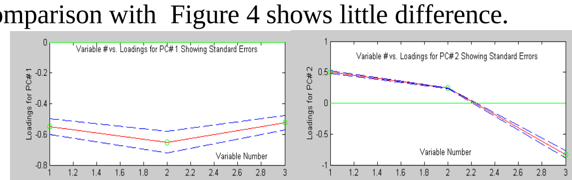

Figure 11 – Loadings(with error bounds) for noisy data of Table 1 (by PLS-toolbox®) (cf. Figure 4).

Process monitoring by looking at residual errors

Once the relationship among x_1, x_2, x_3 has been identified by the counterpart of eqn. (7) with noisy loading $\mathbf{v}_3$, the value of the *residual error* (i.e. counterpart of eqn. (13) for noisy loadings) for each new data point (x_1, x_2, x_3) arriving in the future can be checked. If the relationship among x_1, x_2, x_3 remains the same, then the residual error should be "small". This leads to the counterpart of eqn. (14) for noisy data. Specifically, if the residual error is normally distributed (very often a reasonable assumption) then $\|\mathbf{e}\|^2 = \mathbf{e}^T\mathbf{e}$ follows a chi-square distribution, from which one can construct Q-confidence as (cf. eqn. (14))

$$\mathbf{e}^T\mathbf{e} = \mathbf{x}^T(\mathbf{I}-\mathbf{P}\mathbf{P}^T)\mathbf{x} < d^2 \qquad (22)$$

2.3 Stochastic signals

For multiple random variables principal components *are uncorrelated new variables, a few of which capture most variance*

SVD can provide additional insight if the vector variable **x** is stochastic. The analysis is known as *principal component analysis* (PCA) (Jolliffe, 1986).

Consider the random variable vector $\mathbf{x} \triangleq [x_1 \cdots x_n]^T$, and assume that $E[\mathbf{x}]=\mathbf{0}$ [1] where E denotes expected value. Denote the covariance matrix of **x** by

$$\mathbf{C} = E[\mathbf{x}\mathbf{x}^T] \in \Re^{n\times n} \qquad (23)$$

It can be shown that we can use the modal matrix $\mathbf{A} \triangleq [\mathbf{a}_1 \ \vdots \ \cdots \ \vdots \ \mathbf{a}_n]$ of **C** (i.e. the matrix whose columns are the orthonormal eigenvectors of **C**) to construct a new, zero-mean, vector random variable **y** as

$$\mathbf{y} = \mathbf{A}^T\mathbf{x} \Leftrightarrow \mathbf{x} = \mathbf{A}\mathbf{y} \qquad (24)$$

(**principal components**) that has the following important property

$$\mathrm{var}(y_i) = \max_{\substack{\|a_i\|_2=1 \\ E[y_i y_{j<i}]=0}} \mathrm{var}\left(a_i^T\mathbf{x}\right) = l_i, \qquad (25)$$

That is, each principal component, y_i is a weighted sum of the original variables $x_1, \ldots, x_n$, (eqn. (24)) such that

(a) its variance is maximal and equal to the *i*-th eigenvalue of the original covariance matrix **C** (eqn. (25)), and

(b) y_i is orthogonal to all previous principal components y_{i-j}, $i \geq 2$, $j = 1,\ldots,i-1$ (eqn. (25)).

---

1 If the average of **x** is not zero, a new deviation variable can trivially be defined as $\mathbf{x}-E[\mathbf{x}]$. There is much higher chance that deviation variables (as opposed to original variables) are linearly dependent. Indeed, if the variables **x** satisfy the relationship $\mathbf{f}(\mathbf{x})=\mathbf{0}$, Taylor series expansion around $E[\mathbf{x}]$ yields

$$\mathbf{0} = \mathbf{f}(\mathbf{x}) \approx \underbrace{\mathbf{f}(E[\mathbf{x}])}_{=\mathbf{0}} + \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}=E[\mathbf{x}]} (\mathbf{x}-E[\mathbf{x}]) \triangleq \mathbf{B}\cdot\Delta\mathbf{x}$$

which implies linearly dependent $\Delta\mathbf{x}$.

SVD on covariance estimate produces values of principal components

Because the matrix **C** is unknown, it has to be estimated from data. The best estimate of **C** is

$$\mathbf{C} \approx \frac{1}{m-1}\mathbf{X}^T\mathbf{X} \qquad (26)$$

where **X** is a matrix that contains the data for each random variable in a column. Then, the eigenvalue/eigenvector pairs $(k, \mathbf{w})$ of $\frac{1}{m-1}\mathbf{X}^T\mathbf{X}$ are estimates of the eigenvalue/eigenvector pairs $(l, \mathbf{a})$ of **C**, which implies that

(a) the eigenvectors **w** of $\frac{1}{m-1}\mathbf{X}^T\mathbf{X}$ (hence the estimates of eigenvectors of **C**) are equal to the singular vectors **v** of **X** (eqn.(4)), and

(b) the eigenvalues of $\frac{1}{m-1}\mathbf{X}^T\mathbf{X}$ (hence the estimates of eigenvalues of C) are equal to $(m-1)$ times the squares of the singular values of **X**

Consequently, one can look at the values of

$$\frac{s_i^2}{s_1^2+\cdots+s_r^2} = \frac{s_i^2}{E[\mathbf{x}^T\mathbf{x}]} = \frac{l_i}{l_1+\cdots+l_r} \quad i=1,\cdots,r \qquad (27)$$

to assess what percentage of the total variance of **x**, is captured by each of the principal components. By looking at the first few principal components, one can monitor the system that produces the data

(a) visually, e.g., by plotting PC1 vs. wafer #, PC2 vs. wafer #, etc. or PC1 vs. PC2 vs. PC3.

(b) numerically, by monitoring statistics such as the Hotelling statistic [5].

Principal components are directly related to multivariate SPC

If the zero-mean vector random variable **x** has (non-degenerate) covariance **C**, then one can construct the Hotelling (scalar) random variable

$$\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} = \underbrace{\mathbf{x}^T\mathbf{A}}_{\mathbf{y}^T}\Lambda^{-1}\underbrace{\mathbf{A}^T\mathbf{x}}_{\mathbf{y}} \triangleq \mathbf{y}^T\Lambda^{-1}\mathbf{y} = \sum_{i=1}^{n}\frac{y_i^2}{l_i} \qquad (28)$$

i.e. the Hotelling random variable is the sum of n independent random variables, y_i^2/l_i. If some eigenvalues are zero, then we stop the summation in eqn. (28) at r, the rank of **C**, to ensure $l_i \neq 0$.

3 CASE STUDY 1

Etch profiles (49 measurement points $x_1,\ldots,x_{49}$) from 9 different etching tools were collected, thus creating a 9×49 matrix **X**. Figure 12 indicates that 2 or 3 principal components result in less that 10% or 5% error, respectively. Corresponding scores are shown in Figure 13. Loadings are shown as weights in Figure 14 and as basis surfaces in Figure 15. The quality of reconstruction of the original data by 3 principal components is excellent, in that it captures curvature characteristics, as indicated by the samples shown in Figure 16.

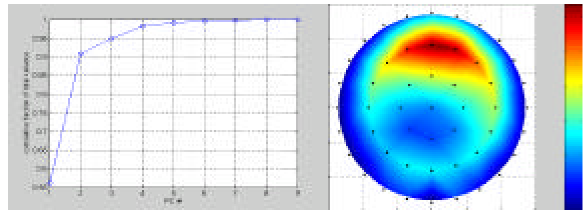

Figure 12 – Cumulative fraction of total variance captured by principal components (left) for variables $x_1,\ldots,x_{49}$ scaled by subtraction of sample averages $\bar{x}_1,\ldots,\bar{x}_{49}$ (right).

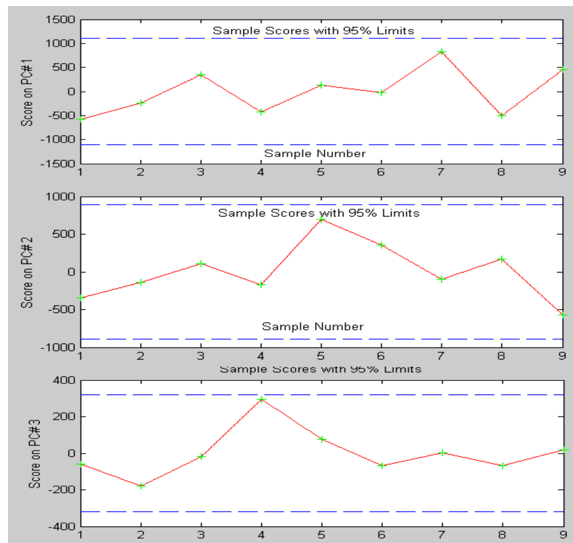

Figure 13 – Scores for the first 3 principal components (cf. Figure 5). (Confidence bounds by PLS-toolbox®)

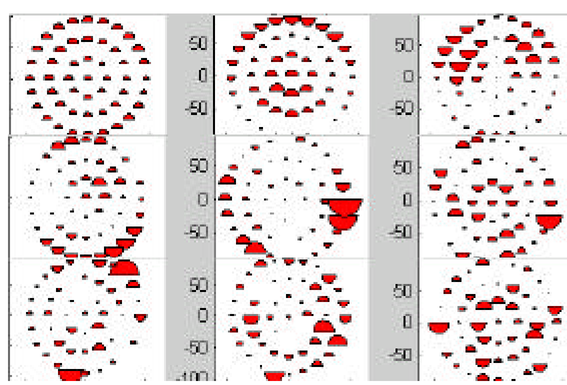

Figure 14 – Loadings as weighting coefficients for all 9 principal components. Semi-disk size and orientation denote magnitude and sign, respectively.

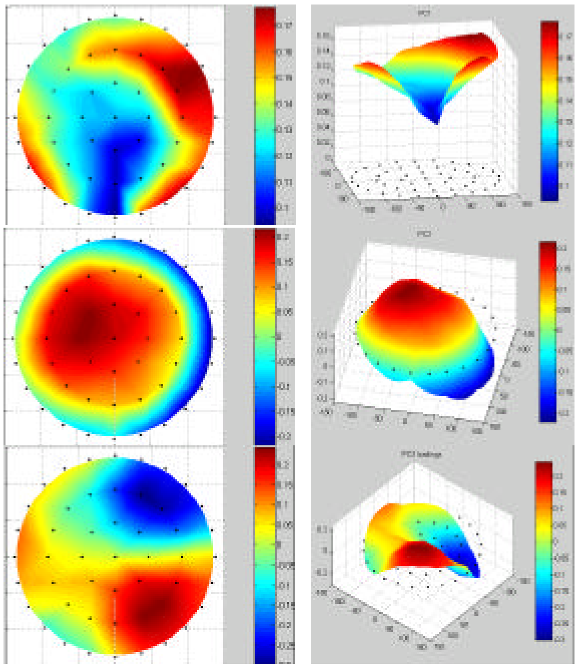

Figure 15 – Top and angle views of loadings as contour surfaces for the first 3 principal components.

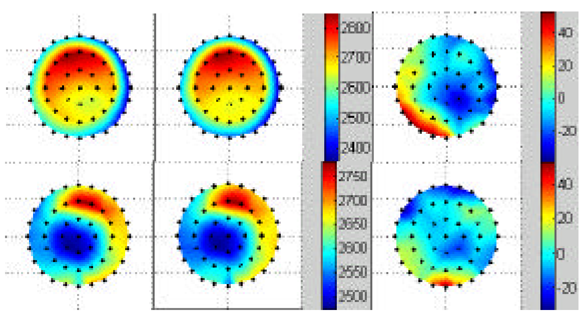

Figure 16 – Original etch profile (column 1), etch profile reconstructed from 3 principal components (column 2) and approximation error (column 3) for two sample wafers (cf. Figure 1)

4 CASE STUDY 2

18 200-mm silicon wafers were etched in an inductively coupled plasma reactor at Lam Research Corporation's facilities in Fremont, CA. Etch rates were measured at 49 points on the wafer, and a 18×49 data matrix **X** was constructed. Three principal components account for 99.94% of variation in data and are considered significant. The three loadings are shown in Figure 17. The scores are shown in Figure 18.

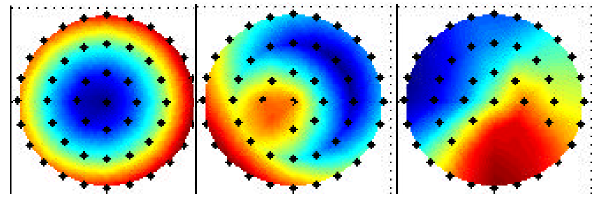

Figure 17 – Loadings of 3 principal components.

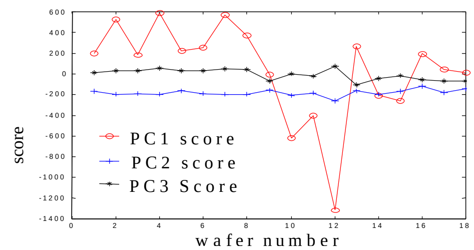

Figure 18 – Scores for PCs for experimental data

It can be observed that PC1 score varies far more than PC2 or PC3 score. There is a strong linear correlation between PC1 score and u_1, u_2 with $R^2 = 0.9686$ and $F = 200.24$. This implies that the first shape can be easily removed from the etch patterns and indeed we can see for wafer 9, PC1 score is almost zero. This information can be used to design better recipes.

5 CONCLUSIONS

Silicon wafer images depicting etch depth uniformity can be analyzed efficiently and effectively using reduced-rank multivariate methods. Two industrial case studies exemplify the basics summarized in this work.

REFERENCES

Horn, RA, and CR Johnson, *Matrix Analysis*, Cambridge University Press (1985).

Dewilde, P., and Ed. F. Deprettere, "Singular Value Decomposition: An Introduction," in *SVD and signal processing: algorithms, applications, and architectures*, edited by Ed. F. Deprettere, North-Holland, 3-41. (1988).

Rigopoulos A. and Y. Arkun, "Principal Component Analysis in Estimation and Control of Paper Machines", *Computers Chem. Engg.*, **20**, S1059-S1064, 1996.

Krischer. K., R. Rico-Martinez, I.G. Kevrikidis, H.H. Rotermund, G Ertl, and J.L. Hudson, "Model Identification Of A Spatiotemporally Varying Catalytic Reaction", *AICHE J.*, 39, 89-98, 1993.

Stewart, G., "Perturbation Theory for the Singular Value Decomposition", in *SVD and Signal Processing, II*, R. J. Vacarro ed., Elsevier, Amsterdam (1991).

Jolliffe, IT, *Principal Component Analysis*, Springer-Verlag (1986).

Acknowledgement – The first two authors gratefully acknowledge partial support for this work from Lam Research Corporation and the National Science Foundation through a GOALI grant.