Matrix Factorisation Techniques for Endpoint Detection in Plasma Etching

E. Ragnoli[a], S. McLoone[a], J. Ringwood[a], N. Macgerailt[b]

[a] Department of Electronic Engineering, NUI Maynooth, Ireland

[b] Intel (Ireland) Ltd

Abstract— **Advanced data mining techniques such as variable selection through matrix factorization have been intensively applied in the last ten years in the area of plasma-etch point detection using optimal emission spectroscopy (OES). OES data sets are enormous, consisting of measurements of over 2000 wavelength recorded at sample rates of $1 - 3$ Hertz, and consequently, these techniques are needed in order to generate compact representations of the relevant process characteristics.**

To date, the main technique employed in this regard has been PCA (Principal Components Analysis), a matrix factorisation technique which generates linear combinations of the original variables that best capture the information in the data (in terms of variance explained).

Recently, an alternative matrix factorisation technique, Non-Negative Matrix Factorisation (NMF) [1], has been gaining increasing attention in the fields of image feature extraction and blind source separation due to its tendency to yield sparse representations of data. The aim of this work is to introduce Non-Negative Matrix Factorisation to the semiconductor research community and to provide a comparison with PCA in order to highlight its properties.

I. INTRODUCTION

Plasma-etch endpoint detection using optical emission spectroscopy (OES) has been a very active area of research in the last ten years. This is largely due to advances in solid-state spectrometer technology and the non-invasive nature of the measurements. The resulting data sets are enormous, consisting of circa 2000 wavelength measurements at a sample rate of 1-3 Hz. A typical spectrum for a 98 time sample etch step of a single wafer from a Hitachi silicon etcher is illustrated in Fig. 4.

In order to make the best use of this data, advanced data mining techniques are needed to obtain compact representations of the relevant process characteristics. The main technique employed to date in this regard has been Principal Components Analysis (PCA). The literature on PCA applications in semiconductor manufacturing research is rather numerous, see for example [2], [3], [4] and [5]. In [3] PCA is applied for spatial characterisation of the wafer state. In [2] and [5] it is exploited for endpoint detection, while in [4] the focus is on fault detection.

PCA is essentially a matrix factorisation technique in which linear combinations of the original variables are used to capture the information in the data in order of significance (in terms of variance explained). Matrix factorisation

decomposes an $n \times m$ (data) matrix $\mathbf{X}$, such that

$$\mathbf{X} \approx \mathbf{WH} \tag{1}$$

where $\mathbf{W}$ and $\mathbf{H}$ are $n \times r$ and $r \times m$ matrices, respectively, that satisfy certain constraints. This can be interpreted as the mapping of the matrix $\mathbf{X}$ onto a new coordinate system, where $\mathbf{H}$ defines the new basis vectors and $\mathbf{W}$ the data coordinates in this new basis. In the case of PCA the new basis are ordered in the directions of the most significant data variation. Other matrix factorisation techniques are, for example, Independent Component Analysis (ICA), that captures the information in the data in terms of statistical independence, Vector Quantisation (VQ), where each column of $\mathbf{H}$ in (1) is constrained to be a unary vector, and the LU Decomposition, where the matrix $\mathbf{X}$ is decomposed as the product of a lower and upper triangular matrix.

Recently, an alternative matrix factorisation technique, Non-Negative Matrix Factorisation (NMF) [1], has been gaining increasing attention in the fields of image feature extraction and blind source separation due to its tendency to yield sparse representations of data.

The purpose of this paper is to introduce NMF to the semiconductor research community. Using PCA as a baseline comparison the main features of NMF will be illustrated. Section 2 provides a brief overview of PCA. Section 3 introduces NMF together with the idea of sparseness and the problem of uniqueness. Section 4 deals with the choice of the number of components in NMF and proposes a PCA guide to finding a solution. Section 5 applies PCA and sparse NMF to a set of plasma-etch OES data and compares the two techniques in terms of sparseness measures and ability to identify a subset of non-zero channels. Section 6 deals with open issues and future directions for research.

II. PRINCIPAL COMPONENT ANALYSIS

Principal Components Analysis (PCA) is a matrix factorisation technique that constrains the columns of $\mathbf{W}$ to be orthonormal and the rows of $\mathbf{H}$ to be orthogonal to each other. Specifically, using standard PCA notation the matrix $\mathbf{X}$ can be expanded as

$$\mathbf{X} = \sum_{i=1}^{r} \mathbf{t}_i \mathbf{p}_i^T = \mathbf{TP}^T \tag{2}$$

where $\mathbf{t}_i$ are $n \times 1$ vectors, $\mathbf{p}_i^T$ is the transpose of the $n \times 1$ vector $\mathbf{p}_i$, $\mathbf{T}$ is a $n \times r$ matrix and $\mathbf{P}^T$ is the transpose of the

$m \times r$ matrix $\mathbf{P}$, $\mathbf{t}_i$ and $\mathbf{p}_i$ are denoted as scores and loadings, respectively. Obviously, the factorisation (2) corresponds to (1), with $\mathbf{W} = \mathbf{T}$ and $\mathbf{H} = \mathbf{P}^T$.

Statistically, the ordered principal components can be interpreted as the directions of largest variance in the data. In general, when there is a significant level of redundancy only a small number of components are needed to capture the information in the data. There are several algorithms which can be used to compute the principal components, for example nonlinear iterative partial least square (NIPALS), singular value decomposition (SVD), the power method (POWER) and eigen-value decomposition (EVD) [6]. NIPALS [7], the one adopted in this work, is an iterative algorithm that computes the eigenvectors of the matrix $\mathbf{X}$, one at the time in order of significance.

III. Non-Negative Matrix Factorisation

The notion of Non-Negative Matrix Factorisation (NMF) was introduced in a recent article in Nature, [1]. NMF is a matrix factorization method for decomposing multivariate data under the constraints of non-negative components. In particular, in the factorization $\mathbf{X} \approx \mathbf{WH}$ negative entries in in the matrix factors $\mathbf{W}$ and $\mathbf{H}$ are not allowed. The non-negativity constraint makes the representation purely additive (allowing no subtractions) which is often more consistent with underling physical processes. Since the scope of this work is to present a comparison between PCA and NMF, the latter will be presented in terms of PCA terminology. Hence, in (1), $\mathbf{W}$ will be referred as scores and $\mathbf{H}$ as loadings.

The task of finding an NMF factorization of $\mathbf{X}$ can be defined as

$$\mathbf{X} = \mathbf{WH} \text{ s.t. } \begin{cases} \mathbf{X} \geq 0 \\ \mathbf{W} \geq 0 \\ \mathbf{H} \geq 0 \end{cases}$$

Moreover, NMF can be expressed as the optimisation problem

$$\min_{\mathbf{W},\mathbf{H}} \|\mathbf{X} - \mathbf{WH}\|^2, \text{ s.t. } \mathbf{W}, \mathbf{H} \geq 0$$

Here $\|\mathbf{X}\|^2 = \sum_{i,j}(x_{ij})^2$, where $\mathbf{X} = (x_{ij})_{m \times n}$. Several iterative optimisation procedures has been proposed for this non-convex problem including multiplicative updates rules [8], projected gradient methods [9] and second order cone programming [10]. In this work an NMF software toolbox [11] that uses a projected gradient algorithm [9] is employed.

Thus, PCA and NMF, both generate matrix factorisations, one with constraints on variance explained, the other with constraints on sign. An important side effect of the non-negativity constraints, not present in PCA factorisation, is that NMF usually produces a sparse representation of the data [9]. The concept of sparseness is introduced and explained in the next section.

To date NMF has been applied to the fields of blind source separation and image feature extraction. An example can be found in [1]. There NMF is applied to a database of facial images and it is shown that it discovers a basis of images that are localised features that intuitively correspond to the notion of the parts of faces.

A. The idea of sparseness

Sparseness is a quantity that measures how much energy in a vector is packed into only a few components. Such a measure is a map from $\mathbb{R}^n$ to $\mathbb{R}$.

The idea comes from [12], where the concept of "sparse coding" refers to a representational scheme where only a few units out of a large population are effectively used to represent typical data vectors. This implies that most units effectively take values close to zero with only a few having values significantly greater than zero.

Several sparseness measures have been proposed and used. For example, in [13] the sparseness of the vector $\mathbf{x} = (x_i)_n$ is defined as the measure

$$s(\mathbf{x}) = \frac{\#(x_i < \epsilon_{tot} x_{max})}{\#(x_i)} \tag{3}$$

where $\#(x_i)$ counts the number of components of $\mathbf{x}$, ϵ_{tot} is a small possible threshold level and x_{max} is the largest element of $\mathbf{x}$. In [14] sparseness is defined as an l_0 norm, such that

$$\|\mathbf{x}\|_0 = \#(x_i \neq 0) \tag{4}$$

It is noted that the l_0 norm can be related to the kurtosis, which consequently can be used as a measure of sparseness. In this work we use the sparseness measure adopted in Hoyer's algorithm [9], based on the relationship between the l_1 and l_2 norm:

$$s(\mathbf{x}) = \frac{\sqrt{n} - \|\mathbf{x}\|_1 / \|\mathbf{x}\|_2}{\sqrt{n} - 1} \tag{5}$$

where $\|\mathbf{x}\|_1 = \sum_i |x_i|$ is the l_1 norm, $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$ is the l_2 norm and n is the dimensionality of $\mathbf{x}$. This function is 1 if and only if $\mathbf{x}$ contains only a single non-zero component, 0 if and only if all the components are equal (up to signs) and in general $0 \leq s(\mathbf{x}) \leq 1$.

In [9] and [15] it is noted that one of the most useful properties of NMF is that it usually produces a sparse representation of the data, making the representation easier to interpret. However, since this is rather a side effect, several researchers have investigated algorithms that attempts to control the sparseness of the representation directly. Such an algorithm is described in [9] and will be referred to as Hoyer's algorithm in this paper.

The task of finding a sparse NMF factorization of $\mathbf{X}$ can be defined as

$$\mathbf{X} = \mathbf{WH} \text{ s.t. } \begin{cases} \mathbf{X}, \mathbf{W}, \mathbf{H} \geq 0 \\ s(\mathbf{w}_i) = \sigma_{\mathbf{W}} \\ s(\mathbf{h}_i) = \sigma_{\mathbf{H}} \end{cases} \tag{6}$$

for $1 \leq i \leq r$, where $s(\mathbf{w}_i)$ and $s(\mathbf{h}_i)$ are the sparseness of the i-th column of $\mathbf{W}$ and the i-th row of $\mathbf{H}$, respectively, and $\sigma_{\mathbf{W}}$ and $\sigma_{\mathbf{H}}$ are the constraints.

B. Uniqueness

A matrix factorization technique generates a unique solution if there exists a unique $\mathbf{W}$ and $\mathbf{H}$ such that (1) is satisfied. Obviously the solutions can only be unique up to scaling
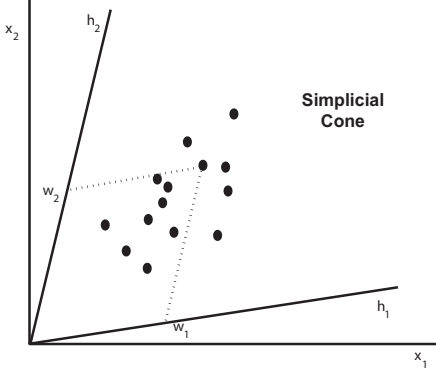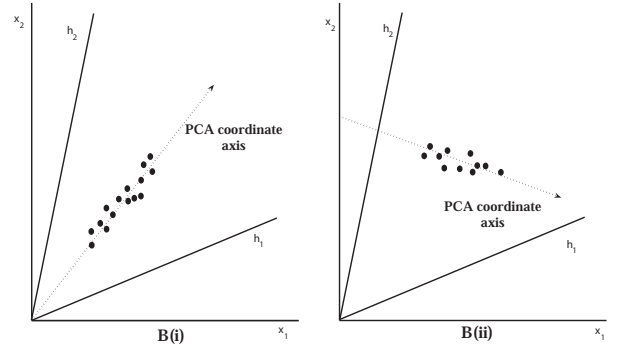
Fig. 1.   Graphical representation of NMF.

Fig. 2.   Graphical representation of NMF dimension estimation when the data can be represented by a single PC component: (i) When the PC axis intersects with the origin $r = r_{PCA} = 1$. (ii) When the PC axis does not intersect with the origin $r = r_{PCA} + 1 = 2$.

and permutations indeterminacies. In fact, if $\mathbf{P}$ denotes a permutation matrix and $\mathbf{L}$ a scaling matrix, then

$$\mathbf{X} = \mathbf{WH} = (\mathbf{WP}^{-1}\mathbf{L}^{-1})(\mathbf{LPH})$$

Hence, when $\mathbf{W}^* = \mathbf{WP}^{-1}\mathbf{L}^{-1}$ and $\mathbf{H}^* = \mathbf{LPH}$ the two factorisations are equivalent.

While the uniqueness of PCA is an established and an obvious result (it can be achieved by ordering of the principal components according to data variance) the same does not apply to NMF.

A first attempt at assessing the uniqueness problem for NMF can be found in [16]. The factorisation (1) can be rewritten as

$$\mathbf{x}_i = \sum_{j=1}^{r} w_{ij}\mathbf{h}_j \qquad (7)$$

where $\mathbf{x}_i$ is the i-th row of $\mathbf{X}$, w_{ij} is the j-th element of the i-th row of $\mathbf{W}$ and $\mathbf{h}_j$ is the j-th row of $\mathbf{H}$. A simplicial cone generated by the family of vectors $\{\mathbf{y}_j\}_{j=1}^r$ is defined as

$$\Lambda = \{x : \ x = \sum_j \alpha_j \mathbf{y}_j, \ \alpha_j \geq 0\}$$

Hence, finding an NMF factorisation for $\mathbf{X}$ is geometrically equivalent to projecting the data $\mathbf{X}$ onto the edges of the simplicial cone generated by $\mathbf{h}_j$. This is illustrated in Fig. 1 in a two dimensional case. Moreover, unless the edges of the simplicial cone are sufficiently near the two positive axis, any simplicial cone can be contained in a bigger simplicial cone. Hence, the authors conclude that, in order to achieve uniqueness, some of the data must be sufficiently near the two positive axis.

An analytic approach to the problem of uniqueness is taken in [17]. There, the non-uniqueness of sparse NMF is investigated. The authors prove that, in the case of the projected gradient algorithms (specifically Hoyer's algorithm), the number of non-unique cases is relatively small compared to the cases in which uniqueness is recovered. In particular, it is shown that the set of non-unique cases is of measure zero when related to the set of unique cases.

IV. DETERMINING THE NUMBER OF NMF COMPONENTS

When applying a matrix factorisation technique, the number of columns and rows of $\mathbf{W}$ and $\mathbf{H}$ (the dimension r in (1)) must be chosen. Since in PCA the principal components are ordered as the directions of largest variance in the data, the choice of r can be determined on the basis of the percentage of the data variation that the factorisation is required to model.

However, in NMF, there is no clear indication which value of r is optimal in order to find the most important components underlying the original dataset. A similar issue applies to the choice of $\sigma_{\mathbf{H}}$ and $\sigma_{\mathbf{W}}$ in (6). In the next section we deal with the choice of r, while the problem of how to find an optimal value of sparseness is not treated in this work.

A. A PCA guide to finding the number of components

PCA can be geometrically interpreted as the reorganisation of the data $\mathbf{X}$ with a new set of coordinates, following the directions of largest variance in the data $\mathbf{X}$. As above, from [16], NMF can be geometrically interpreted as the reorganisation of the data $\mathbf{X}$ with a new set of coordinates, according to the simplicial cone that contains the data $\mathbf{X}$ (although the new coordinates set is no longer orthogonal).

Hence, a methodology for finding the optimal value of r is to map the data into its PCs and then to project these coordinates in a simplicial cone.

(A)   Perform PCA. Decide r_{PCA}, the number of components that describe the percentage of data required. That gives the number of axis in the PCA coordinate system.

(B)   Project the PCA coordinate system in the simplicial cone describing NMF. Two possibilities exist:
(i) If one of the PCA coordinate axis intersects the origin of the simplicial cone then $r = r_{PCA}$.
(ii) If none of the PCA coordinate axis intersects the origin of the simplicial cone then $r = r_{PCA}+1$.
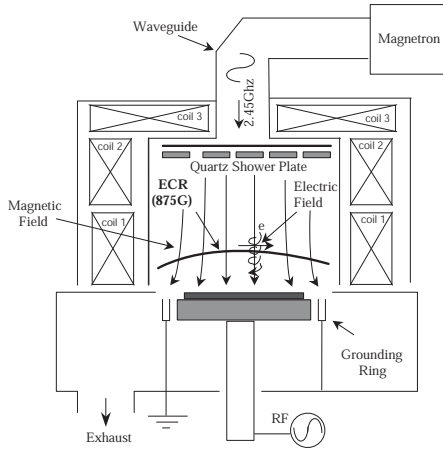
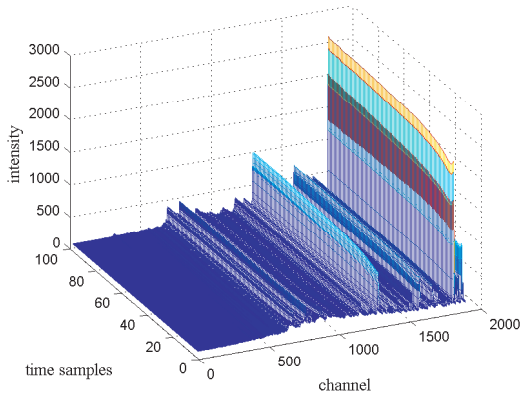Fig. 3. Graphical representation of a Hitachi STR Etch chamber.

Fig. 4. 3D OES Spectrum for a plasma discharge.

This procedure is justified by the fact that any $\mathbb{R}^n$ subsystem of m orthogonal coordinates (the euclidean space $\mathbb{R}^m$), with $n \leq m$, not centred on the origin, can be projected into the r edges of a simplicial cone, with $n \leq r \leq n+1$. This is illustrated, for the two dimensional case, in Fig. 2. Moreover, since NMF with sparseness constraints is applied in this work, the simplicial cone system of coordinates found is unique as in [17].

When dealing with applications, it can be conjectured that the set of cases for which B(i) applies is of measure zero when compared to the set of overall cases and hence statistically irrelevant. Hence, in practice, the number of NMF components can be estimated as $r = r_{PCA} + 1$. When sparseness constraints are included this can be viewed as a lower bound on r.

V. EXPERIMENTAL RESULTS

To demonstrate the properties of NMF, a case study was undertaken involving OES data collected from a Hitachi STR Etch Chamber with a $Ar-O_2-Cl/HBr/C_2F_6$ gas mixture discharge. The etcher, which is illustrated graphically in Fig. 3, is equipped with a magnetron generator with a microwave input of 2.45 GHz. Various mixtures of Ar, Cl_2, SF_6, HBr,
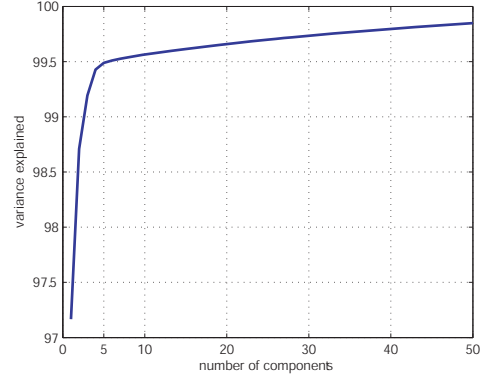

Fig. 5. PCA variance explained as a function of the number of PCs.

CF_4, CHF_3 and/or O_2 gases are used. The gases enter through the gas ring between the quartz window and the quartz shower head. A He backside cooling system is used to control the wafer temperature.

During operation the magnetic field strength is held at 875 Gauss, the RF Power supply (used to generate DC bias independently of plasma ionization) is changed from 0 to 50 W, the magnetron power is switched between 300 and 1500 W, and the chamber pressure is increased from 0.3 to 1.6 Pa. Adjusting RF power at low pressure allows extraction of atoms with a longer free path. The electrons travel far enough in order to gain sufficient energy to strike gas molecules and cause ionization.

Under these conditions OES data was recorded for an experimental run in which a plasma consisting of fixed amounts of Cl_2, HBr and O was used to etch a polysilicon wafer. Measurements were taken using a USB 2000 spectrometer. The resulting 3D spectrum, plotted in Fig. 4, consists of emission intensity measurements for 1890 channels over 90 consecutive time samples.

A. PCA analysis

Initially the OES data was analysed using PCA to assess the level of redundancy and determine its underlying dimensionality. Fig. 5 shows the relationship between the accumulated variance explained by successive principal components (PCs) as a function of the number of PC components (r_{PCA}) and clearly highlights the substantial redundancy in the data. Three PCs can explain 99% of the variation in the data and this increases to 99.5% when five PCs are used. Furthermore, there is a clearly defined 'knee' in the graph at $r_{PCA} = 5$, beyond which each additional PC contributes approximately equally to variance explained, suggesting that the underlying dimensionality of the OES data is 5 and that the remaining components are capturing variation due to noise.

The scores of the first three PCA components are shown in Fig. 6. Note that each one has a clear endpoint signal with PC1 showing the sharpest transition at end-point. Thus, a derivative based threshold on this PC could form the basis of a robust end-point detection algorithm. The one weakness of PCA is highlighted in Fig. 7 which shows the loadings
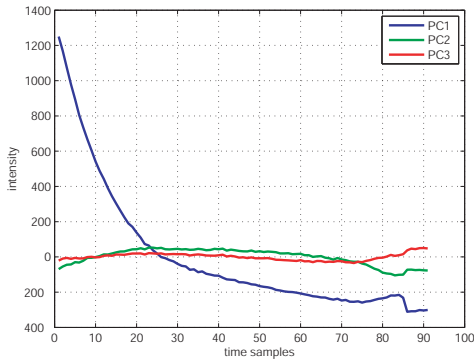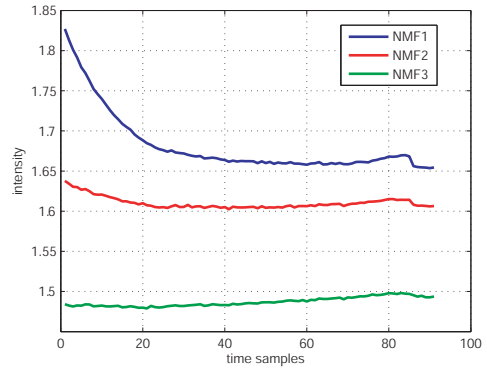
Fig. 6.   1st, 2nd, 3rd PCA component (scores).
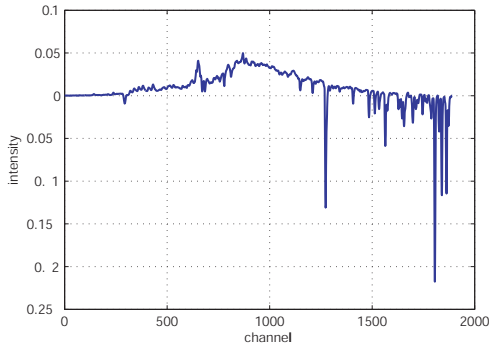

Fig. 8.   3 NMF components $\sigma_{\mathbf{H}} = 0.7$.

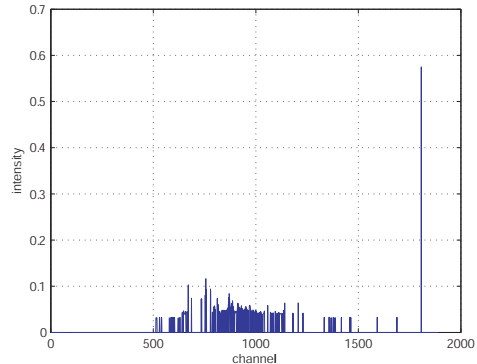Fig. 7.   Loadings of PC1.

Fig. 9.   NMF1 loadings $\sigma_{\mathbf{H}} = 0.7$.

for PC1. As can be seen almost all the channels contribute significantly to the PC direction. This makes it difficult to gain practical insights into the chemistry contributing to process variations. It also means that the full OES spectrum has to be recorded in order to compute the PC. Note, that while Fig. 7 shows that some wavelengths make large contributions to the PC direction, this can not be relied on as indicating significance due to the highly non-uniform spectrometer sensitivity characteristic. Variance scaling can be used to eliminate the effect of non-uniformity, but this in turn leads to the amplification of noise on small magnitude signals. This is a particularly severe problem with OES data as some channels are predominately background noise, hence, in practice variance scaling is not recommended [5].

B. NMF analysis

Since five PCA components can explain 99.5% of the spectral data, it follows from Section IV-A that six NMF components will be sufficient to adequately explain the OES data. To encourage NMF to produce components with a small number contributing channels, thereby providing valuable insight into key factors contributing to process variations, a sparseness constraint is imposed on the NMF loading matrix, $\mathbf{H}$. There is no benefit in imposing a sparseness constraint on the scores matrix, $\mathbf{W}$, hence none is imposed in this work.

Fig. 8 shows the three most significant NMF components obtained when NMF is applied to the OES data with the

sparseness on $\mathbf{H}$ set at $\sigma_H = 0.7$, while Fig. 9 shows the loadings corresponding to the most significant NMF component (NMF1). Since NMF imposes no particular ordering on its components (running the sparse NMF algorithm twice results in the same components but in a different order), these were identified manually by visual inspection of the six computed components.

The NMF components clearly capture the same endpoint feature that appeared in the three PCs obtained with PCA analysis, and are therefore of equivalent value to PCA in this regard. However, in contrast to PCA the NMF loadings are much sparser, with a large number of zero loadings. In the case of NMF1 82.5% of the loadings are zero (Fig. 9), while none of the PC1 loadings are zero.

To gain a better understanding of the relationship between Hoyer's sparseness constraint, and the sparseness of the resulting NMF components the percentage of non-zero channels has been recorded in Table I for each of the six NMF components for σ_H values ranging from 0.1 to 0.9. Since NMF dose not produce a rigorous ordering of the components, they have been ordered in the table according to descending l_2 norm.

Note that increasing σ_H corresponds to a decrease in the percentage of non-zero channels, but that the relationship is not one-to-one or consistent, i.e. the number of non-zero channels varies from component to component for the same

TABLE I

NMF COMPONENTS LOADINGS: PERCENTAGE OF NON ZERO CHANNELS

σ_H	NMF1	NMF2	NMF3	NMF4	NMF5	NMF6
0.1	100	100	100	100	100	100
0.2	100	100	100	79.7	80.8	100
0.3	100	100	94.2	79.7	80.8	100
0.4	91.7	66.9	80.1	91.3	99.5	85.5
0.5	67.3	80.5	50.1	68.7	80.5	55.3
0.6	78.6	44.8	54.6	38.7	30.2	18.4
0.7	13.3	12.1	17.6	12.4	22.2	22.3
0.8	22.3	7.9	13.5	15.4	19.8	21.1
0.9	2.3	8.8	12.2	7.7	7.9	5.1

sparseness level.

To provide a comparison with PCA analysis, Hoyer's sparseness measure and the percentage of non-zero channels have also been computed for each of the 5 PCA loadings. Here the percentage of non-zero channels was found to be 100% for all PCs, while the sparsities were 0.37, 0.42, 0.51, 0.44 and 0.40, respectively. These sparseness values reflect the significant number of low amplitude components in the PC loadings and suggests that the sparseness measure (3) proposed in [13] may be a better reflection of Hoyer's measure, than the strict zero count adopted here.

VI. FUTURE WORK

Development and interpretation of NMF for practical applications such as plasma etch monitoring is ongoing, and several issues remain to be resolved. As noted above, NMF components are not ordered. A descending energy criterion was used in this paper for Table I, but the question of what the best choice is remains open. Ordering in terms of the variance explained, as in PCA, may be better, but this is not straight forward as NMF components are not orthogonal.

Further investigation of the parts based representation is necessary. From Table I it can be seen that NMF chooses a subset of the original channels. Are these significant in terms of the chemistry of the process? Is NMF able to consistently detect the key channels and, hence, is it a valid variable selection technique for OES metrology?

In this work NMF components were computed using Hoyer's algorithm [9]. The sparseness measure adopted (5) has the effect of increasing the number of zero channels. However, it is not a direct measure of this quantity. Can an algorithm that adopts a direct measure of the non-zero channels, like the l_0 norm (4), produce a better parts-based representation?

VII. CONCLUSION

In this paper two matrix factorisation techniques, Principal Component Analysis (PCA) and Non-Negative Matrix Factorisation (NMF) have been compared on a case study involving optical emission spectroscopy data from a plasma-etch process. A methodology for finding the number of components for NMF, based on PCA, has also been introduced.

While PCA is an established technique for applications in the field of semiconductor manufacturing, NMF is relatively new. Both techniques are able to detect critical features in the OES data case study, but they differ substantially in their ability to select a subset of channels relevant to the feature investigated. In fact, for any PCA component 100% of the channels are non-zero and hence contribute to the component. In contrast, NMF is able to select a much smaller percentage of non-zero channels, a percentage that can be controlled by appropriate choice of sparseness constraint. This helps to highlight a small number of channels that are the key contributors to the feature being observed.

REFERENCES

[1] D. D. Lee and H. S. Seung. Learning the parts of objects by Non-Negative Matrix Factorization. *Nature*, 401:788–791, 1999.

[2] White et al. Low Open-Area Endpoint Detection Using a PCA-based T2 Statistics and Q Statistics on Optimal Emission Spectroscopy Measurements. *IEEE Trans. Semiconductor Manufacturing*, 13(2), 2000.

[3] D.A. White, D. Boning, S.W. Butler, and G.G. Barna. Spatial characterisation of wafer state using principal component analysis of optical emission spectra in plasma etch. *IEEE Transactions on Semiconductor Manufacturing*, 10(1):50–61, 1997.

[4] H. H. Yue, S. J. Qin, R. Markle, C. Nauert, and M. Gatto. Fault Detection of Plasma Etchers Using Optical Emission Spectra. *IEEE Transactions on Semiconductor Manufacturing*, 13(3), 2000.

[5] H. H. Yue, S. J. Qin, J. Wiseman, and A. Toprac. Plasma etching endpoint detection using multiple wavelengths for small open-area wafer. *J. Vac. Sci. Technology A*, 1(19):66–75, 2001.

[6] E.R. Malinowski. Factor analysis in chemistry. *John Wiley and Sons*, 1991.

[7] P. Geladi and B.R. Kowalski. Partial Least Squares Regression: a tutorial. *Analytica Chimica Acta*, 1986.

[8] D.D. Lee and H.H. Seung. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing (Proc. NIPS 2000)*, 13, 2001.

[9] P. O. Hoyer. Non-negative Matrix Factorization with Sparsess Constraints. *Journal of Machine Learning Research*, (5):1457–1469, 2004.

[10] M. Heiler and C. Schnorr. Learning Sparse Representations by Non-Negative Matrix Factorization and Sequential Cone Programming. *Journal of Machine Learning Research*, 7, 2006.

[11] P.O. Hoyer. http://www.hiit.fi/node/70.

[12] D.J. Field. What is the goal of sensory coding? *Neural Computation*, (6):559–601, 1994.

[13] K. Stadlthanner, F. Theis, E. W. Lang, A. Tome, and C. Puntonet. Blind Matrix Decomposition Via Genetic Optimisation of Sparseness and Nonnegativity Constraints. *Proceedings of ICANN 2007*, pages 799–808, 2007.

[14] J. Karvanen and A. Cichowski. Measuring Sparseness Of Noisy Signals. *Proceedings of ICA2003*, pages 125–130, 2003.

[15] J. Eggert and E. Korner. Sparse coding and NMF. *Proc. In. Joint Conf. on Neural Networks*, pages 2529–2533, 2004.

[16] D. Donoho and V. Stodden. When does a Non-Negative Matrix Factorisation Give a correct Decomposition into Parts? *Advances in Neural Information Processing Systems*, 16, 2004.

[17] F. Theis. First results on uniqueness of sparse Non-Negative Matrix Factorization. *EUSIPCO*, 2005.