

Target Factor and Neural Network Analyses Applied to Titanium Nitride Composition Recognition by AES

Jill Card,* Anne L. Testoni and Laurie A. Le Tarte

Digital Equipment Corporation, HLO2-3/N08, 77 Reed Road, Hudson, MA 01749, USA

This paper details the development of two pattern recognition approaches—modified target factor analysis (TFA) and artificial neural network analysis—applied to Auger electron spectra for precise discrimination of thin-film TiN_x compound composition. Use of Auger electron spectroscopy for the analysis of titanium nitride films offers advantages over other spectroscopic methods in its ability to analyze either blanket-deposited films or submicron features on patterned wafers and its speed of sample preparation and spectral processing. However, severe overlap of the characteristic Ti LMM and N KLL transitions between 380 and 390 eV prohibit direct stoichiometry measurement. Modified TFA and multilayer perceptron (MLP) neural network approaches are applied to the task of discriminating TiN_x stoichiometries to within 2% of a nominal 1:1 composition. The modified TFA procedure succeeded in accurate prediction of AES spectra TiN_x stoichiometries in five of six sample groupings. The MLP neural network approach accurately predicted stoichiometries for all samples with reduced sample variance over standard analytical methods. This work combined AES and neural network technologies to improve significantly the precision of the stoichiometric composition recognition capability for TiN_x compounds.

INTRODUCTION

This paper details work performed in the development of pattern recognition procedures applied to the precise discrimination of chemical stoichiometries of thin-film titanium nitride (TiN_x) compounds via two methods of mathematical decomposition of Auger electron spectra: multivariate linear target factor and artificial neural network analyses. Auger electron spectroscopy (AES) offers the advantage over other spectroscopic methods such as x-ray photoelectron spectroscopy (XPS) and Rutherford backscattering spectrometry (RBS) in the analysis of titanium nitride films in its ability to analyze either blank-deposited films or submicron features on patterned wafers. In addition, the sample preparation and spectral data processing time of Auger electron spectra can be much less than other spectroscopic methods.¹ The precision of AES is anticipated to be sufficient to enable discrimination of stoichiometries to within 0.5 at.% concentration differences.

The principal difficulty arising in the use of AES for TiN_x stoichiometry discrimination is the appearance of major overlap of the characteristic Ti LMM and N KLL transitions between 380 and 390 eV. Because the individual spectra cannot be measured directly, pattern recognition techniques for discrimination of atomic concentrations of Ti and N are required.

To date, published AES pattern recognition studies^{2,3} have had success in compound concentration discrimination by utilizing targeted factor analysis (TFA) methods developed by Malinowski and Howery.⁴

Stickel, Watson and Diebold's³ findings show Ti and N peak separation and accurate relative concentration

discrimination between TiN_x samples whose atomic concentration span ± 20 at.%. Our study sought to extend their results in yielding accurate discrimination capability for TiN_x stoichiometries that vary from a nominal TiN_x 1:1 composition by 1.5–2.0%, with a between-sample discrimination capability of 0.5%, matching the anticipated precision of the Auger spectra measurement.

This report details the results in model development and employment of statistical and artificial neural network pattern recognition techniques that satisfy the TiN_x concentration problem discrimination requirements. In addition to current results for the TiN_x study, the applicability of these techniques to increased requirements in the precision of discrimination in other areas of spectroscopy is discussed.

BACKGROUND

The structure and composition of TiN_x films are known to affect their electrical resistance and effectiveness as diffusion barriers between aluminum and silicon in ULSI integrated circuits. X-ray photoelectron spectroscopy (XPS) can be used to measure titanium nitride stoichiometries relative to a standard material using the statistical comparison of Ti 2p and N 1s peak area ratios.¹ This method has been shown to discriminate between samples within $\pm 0.5\%$ relative concentration differences in the Ti:N ratio, with relative concentration standard deviations measured in the range 0.4–1.0% when care is taken to produce highly reproducible data. This corresponds to being able to measure differences of ± 0.5 at.% concentration.

While precise, the XPS-based procedure has inherent drawbacks that hamper its use as a time- and labor-

* Author to whom correspondence should be addressed.

efficient composition measurement procedure. The analysis area in XPS for high-sensitivity concentration measurements is $\sim 2 \text{ mm} \times 3 \text{ mm}$ and therefore requires unpatterned films. In addition, the XPS sample preparation, data collection and analysis are time consuming, requiring of the order of 9 h per two-film comparison.¹ Recent advances in XPS hardware and data analysis systems have reduced this time a little but several hours are still required.

Auger electron spectroscopy measurements were expected to be equally sensitive to the precision capabilities of the stoichiometry discrimination of XPS, with the added ability of being applied to small features on patterned wafers. In addition material preparation and data collection times are significantly reduced, owing to the smaller area to be cleaned of surface contamination and the ability to increase easily the excitation species flux to $\sim 1 \text{ h}$ per sample. The actual mathematical computation of the relative Ti:N concentration composition of unknown samples is a matter of minutes, given the solution to the pattern recognition problem addressed in this paper. The challenge in developing accurate pattern recognition algorithms for determination of TiN_x stoichiometry using AES is the severely overlapped characteristic peaks of the Ti LMM and N KLL intensity spectrum profiles (Fig. 1), precluding the use of spectral peak area comparison methods for composition discrimination, as can be done using XPS.

Gaarenstroom² investigated the use of pattern recognition applied to poly(alkyl)methylacrylate compound decomposition using a TFA procedure. Using TFA, he computed the composition of six polyethyl- or polymethylacrylate compounds within polymer samples to within an average estimated error of 9%. Stickle, Watson and Diebold³ successfully applied TFA by utilizing derivative transforms of raw data to the TiN_x stoichiometry discrimination problem. They were able to predict accurately the relative compositions of TiN_x samples that varied sample to sample by $\pm 5\%$. The supervision in their report was done by using assumed known Ti and N concentrations obtained via Rutherford backscattering spectroscopy (RBS).

The TiN_x discrimination problem discussed herein differs from that reported by Stickle, Watson and Diebold in the utility of XPS-determined Ti:N relative concentration values as the assumed "known" target concentration. X-ray photoelectron spectroscopy was

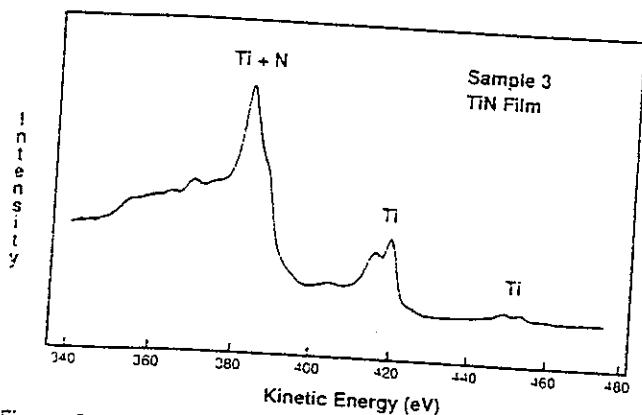


Figure 1. Auger electron spectrum of a TiN_x film. The characteristic N KLL peak appears at $\sim 380 \text{ eV}$. The three major Ti LMM peaks appear at $\sim 385, 420$ and 450 eV .

utilized because of its increased discrimination capability of XPS over RBS in relative concentration determinations⁴ and because it is available in our laboratory. As a consequence, it is intended that the between-sample stoichiometry discrimination capability using Auger methods will be increased by one order of magnitude (to $\pm 0.5\%$ as compared to $\pm 5\%$).

Given the very high precision required for stoichiometry prediction in the $\pm 0.5\%$ atomic concentration discrimination range, it was found that the standard TFA linear approximations were not satisfactory. A modification was introduced that reduced a between-sample shift in intensity measurements due to slight differences in sample placement with respect to the electron energy analyzer. These placement differences resulted in shifts in the absolute magnitude of the collected spectra which is a violation of the assumed constancy required by the linear TFA model. The modified TFA procedure successfully yielded predicted Ti:N atomic concentration ratios that were not significantly different in values and variance from the predictions obtained using XPS of the same samples for the majority of samples analyzed.

The remaining lack-of-fit component using the TFA procedure was assumed to be due to non-linearity of the data system, either because of the physical data collection procedure or because of other unknown chemical factors to which AES is sensitive.

A multilayer perceptron (MLP) was constructed and applied to the Auger TiN_x data as a second approach to developing a prediction model, given the neural network's ability to learn non-linear mapping and generalize pattern learning in the presence of noise. Neural networks have been used extensively in recent years to address chemical decomposition and general spectroscopic analysis problems.⁵⁻¹⁰ Analysis results in this work for training and testing of a back-propagation MLP yielded a predictive model capable of assessing Ti and N atomic concentrations to within $\pm 0.5\%$ atomic concentration for data collected using a single spectrometer.

For purposes of more extensive MLP network validation, additional AES samples were collected, although from the same instrument with a different electron energy analyzer. Data from all samples were used to recompute a TFA model. The modified TFA procedure was not able to estimate accurately the Ti:N ratios owing to a significant between-spectrometer shift effect. The MLP, however, learned the training set well (average sample ratio estimated error = 0.10%) predicted the test set sample ratios to within 1.5% error on the average. These results suggest that the neural network approach to AES stoichiometry decomposition is a feasible avenue of analysis to pursue, given the network's ability to learn complex, non-linear mappings.

PROBLEM DESCRIPTION

We sought to develop a pattern recognition tool that could discriminate TiN_x stoichiometry to within $\pm 0.5\%$ atomic concentration for Ti and N in Auger electron spectra. In keeping with previous AES/ TiN_x analysis results,³ we approached the higher precision model in a

similar fashion by proposing a multivariate linear model, as follows

$$[D]_{o \times v} = [C]_{o \times n} [R]_{n \times v} \quad (1)$$

where: $[D]$ is an observed intensity data matrix of o observation sets, each comprised of $v = 270$ energy measurement points; $[C]$ is a concentration matrix of n columns, each column representing the concentration of each principal feature by o observation sample rows; and $[R]$ is a characteristic feature matrix of n rows, each row corresponding to a v length spectrum of the n th principal feature.

Equation (1) decomposes an observed matrix $[D]$ of AES spectra into $o \times n$ column matrix $[C]$, containing the atomic concentrations of the n principal chemical components of the system, and $[R]$, a corresponding row matrix of AES features, wherein the j th row represents the AES feature of the j th principal chemical component.

We decompose the sample spectra observation matrix $[D]$ such that the rows of $[R]$ correspond to the characteristic AES spectra of titanium and nitrogen, respectively. If such characteristic features are known, the concentration vector elements can be calculated from any 'unknown' AES TiN_x spectra.

We rewrite Eqn (1) in matrix element form as

$$d_{ij} = \sum_k c_{ik} r_{kj} \quad (2)$$

where d_{ij} is the j th energy interval intensity measure of the i th sample, c_{ik} is the k th chemical component concentration of the i th sample, and r_{kj} is the j th energy interval standard intensity value of the k th chemical component.

The representation Eqn (2) of d as a linear combination of standard chemical feature vectors is assumed in the TFA model.

Alternatively, a more general model can be constructed that is not constrained by the linear relationship of Eqn (2). An MLP was developed for the AES/ TiN_x discrimination problem, having a general

model design as follows:

$$x_{k,p} = g\left(\sum_n x_{j,n} w_{k,np}\right) \quad (3)$$

where

$$x_{j,n} = f\left(\sum_m x_{i,m} w_{j,mn}\right) \quad (4)$$

and: $x_{k,p}$ is the p th node value of the k (output) layer, corresponding to the x_k nodes, which are the predicted Ti or N atomic concentration values; $x_{i,m}$ is the equivalent m th node value of the i (input) layer, corresponding to the m th AES sample spectrum element, $x_{j,n}$ is the equivalent n th node value of the j (hidden) layer; $w_{k,np}$ is the weight coefficient between the n th and p th nodes of the j (hidden) and k (output) layers, respectively (subscript k identifies the destination layer); $w_{j,mn}$ is the equivalent weight between the i (input) and j (hidden) layers; and $f(z)$ and $g(z)$ are functional transforms for the hidden and output layers, respectively.

The incorporation of a second (hidden) layer in an MLP construction (Fig. 2) enables solution of the non-linear prediction problem.¹¹ This ability to perform non-linear mapping and the MLP's ability to generalize in the presence of non-random nuisance parameters made it a logical analytical tool choice following exhaustion of modification to the TFA based on understood physical principles affecting TiN_x stoichiometry.

EXPERIMENTAL

Samples

Six titanium nitride films of varying Ti:N stoichiometry were prepared by reactive sputter deposition (Ti target, N_2/Ar gas mixture) onto Si $\langle 100 \rangle$ substrates in an MRC Eclipse system; for this paper, the films are referred to as samples 1-6. Nitrogen partial pressure

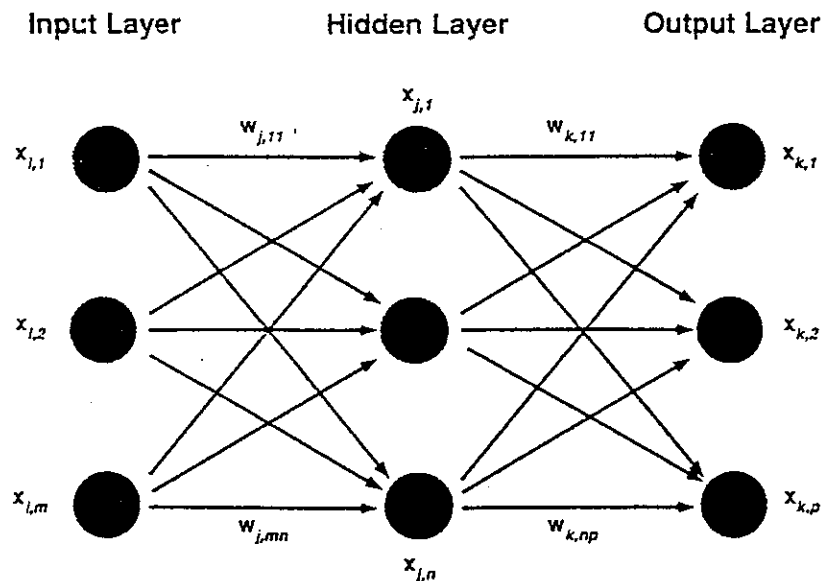


Figure 2. Diagram of a multilayer perceptron (MLP) neural network showing input, hidden and output layers.

Table 1. Summary of data collection by sample stoichiometry

Sample no.	No. of repetitions	ratio from XPS	Ti/N atomic concentration*	Ti:N
1	10	1.057	0.514:0.486	
2	10	1.065	0.516:0.484	
3	10	1.595	0.615:0.385	
4	10	0.973	0.493:0.507	
5	10	1.000	0.500:0.500	
(ref.)		(assigned)		
6	5	1.002	0.501:0.499	
After analyzer change				
7	5	0.981	0.495:0.505	
8	5	0.930	0.482:0.518	
9	5	0.914	0.477:0.523	
10	5	0.951	0.487:0.513	

* Calculated from Ti/N ratio by assuming that only Ti and N are present within film.

and substrate bias were varied in order to produce films with a range of TiN_x composition. One film (sample 5) was designated as 'standard' and assigned a Ti:N stoichiometry of 1.00. At a later date, four additional TiN_x films were prepared, also by reactive sputter deposition, using conditions that would produce films showing nitrogen enrichment (Ti:N < 1.00). These films are referred to as samples 7-10.

Individual pieces, 25 mm \times 25 mm in size, were cleaved from the center of each wafer for analysis. The pieces were glued onto stainless-steel (316L) sample holders using colloidal silver paint and allowed to dry for a minimum of 2 h prior to analysis.

Instrumentation

All data were collected using a VG Scientific Microlab Mk III electron spectroscopy system (~1987), equipped as follows:

Electron energy analyzer: hemispherical sector (150°), 150 mm radius, with three parallel detectors (version 1) or five parallel detectors (version 2) and a focusing transfer lens.

Electron source: LaB_6 .

Electron optical column: dual electromagnetic lenses.

X-ray source: dual anode (Mg, Al); no monochromator.

Ion sputtering gun: noble gas, differentially pumped; electron impact source; single electrostatic lens, rasterable.

Vacuum system pressure: $< 1 \times 10^{-9}$ mbar during data collection; $\sim 8 \times 10^{-7}$ mbar during sputter cleaning.

Between the time when data were collected from samples 1-6 and from samples 7-10, the electron energy analyzer was replaced. The only difference between analyzer configurations was an increase in the number of parallel detectors from three (version 1) to five (version 2). This analyzer replacement should affect the absolute magnitude of individual spectra, owing to the increased electron detection. Other changes in spectra due to differences in transmission function should be minor.

Reference composition—XPS measurements

Quantitative comparison of the 'unknown' film compositions, with respect to the 'standard' film, was made using XPS. Details of the procedure are presented elsewhere;¹ a summary of the procedure is given below:

- (1) remove surface contamination and native oxide using ion sputtering (no evidence of O 1s or C 1s peaks in spectra);
- (2) record Ti 2p and N 1s regions 10 times, as data pairs;
- (3) calculate areas of Ti 2p and N 1s photoelectron peaks after Shirley background subtraction from each spectrum in the data pair;
- (4) calculate Ti 2p/N 1s ratios for each data pair;
- (5) normalize the Ti 2p/N 1s ratio to the assigned 'standard' film;
- (6) perform Student's *t*-test to determine whether a significant difference in composition exists between samples.

The XPS data collection parameters are as follows:

Electron energy analyzer: mode, CAE; pass energy = 10 eV; acceptance angle = 22°; emission angle = 30°.

X-ray source: type, Mg K α ; power = 300 W; incident angle = 30°.

Ion sputtering gun: gas, Ar; 5×10^{-5} mbar in ionization chamber; energy = 3.0 keV; current = 250 nA raster = 10 mm \times 10 mm, 10 frames per second.

The results of this analysis, yielding nominal, relative Ti:N composition, are given in Table 1. While possibly not representative of the absolute composition of the individual TiN_x films, it is believed that these values are an accurate indication of their relative composition.

Composition recognition via AES

Auger electron spectroscopy was performed on the same samples as used for the XPS measurements. Data were collected immediately after completion of the XPS experiments so that the sample surface was still free of contamination and/or oxidation; cleanliness was verified by collecting wide-energy-range (20-2000 eV) Auger electron spectra from each sample surface. The following AES data collection conditions were employed to produce spectra for this work:

Electron energy analyzer: mode, CRR = 10; acceptance angle = 22°; nominal emission angle = 30°.

Incident electron beam: energy = 10 keV; current = 25.0 nA; angle = 30°; no raster during data collection; spot size ~ 2000 Å.

Spectral range: start energy = 340.00 eV; end energy = 475.00 eV; increment = 0.50 eV; dwell time increment = 100 ms; spectra averaging = 10 scans.

Spectra were collected at different positions on each sample to enable estimation of within-sample variance. The number of repetitions is indicated in Table 1. A total of 55 spectra were collected from samples 1-6 and 20 from samples 7-10, yielding a total of 75 spectra for the data analysis.

DATA PREPROCESSING

Data preprocessing was designed to achieve data compression and noise reduction, as well as to highlight data features in such a way as to facilitate recognition by an analysis tool. Initial investigation of composition prediction of TiN_x Auger electron spectra utilized data preprocessing in the form of the fast Fourier transform (FFT) of each 270 element spectrum sample. The FFT serves both to reduce the input vector size from 270 intensity readings per spectrum to 20–35 Fourier coefficients with adequate stoichiometric information for pattern discrimination, while concurrently reducing noise.

In addition, for the linear model description of the TiN_x prediction problem, the concentration matrix [C] can be shown to be invariant under the FFT.

The FFT is defined as

$$f_i = \frac{1}{v} \sum_{j=0}^{v-1} d_j \exp(-iv_i j) \tag{5}$$

and

$$d_j = \sum_{i=0}^{v-1} f_i \exp(-iw_i j) \tag{6}$$

where f_i is the i th Fourier coefficient, v is the total number of original variable observations per data set, d_j is the j th original data set variable observation, $i = \sqrt{-1}$, w_j is the frequency ($2\pi f/v$) and j is the sequence number of variable observations within each spectrum.

Bloomfield¹² shows superposability for the FFT

$$d_j = y_j + z_j \Rightarrow f_{d,i} = f_{y,i} + f_{z,i} \tag{7}$$

Assigning $y_j = c_1 r_{1j}$ and $z_j = c_2 r_{2j}$

$$d_j = c_1 r_{1j} + c_2 r_{2j} \Rightarrow f_{d,i} = f_{c_1 r_{1i}} + f_{c_2 r_{2i}} \tag{8}$$

We see the FFT equality in terms of an input spectrum vector d of variable length j in the TiN_x model in Eqn (1). Furthermore

$$\begin{aligned} y_j^* &= a y_j \Rightarrow f_{y^*,i} \\ &= \frac{1}{v} \sum_{j=0}^{v-1} a y_j \exp(-iw_i j) \\ &= a f_{y,i} \end{aligned} \tag{9}$$

It follows that

$$\begin{aligned} d_j &= c_1 r_{1j} + c_2 r_{2j} \Rightarrow f_{d,i} \\ &= c_1 f_{r_{1i}} + c_2 f_{r_{2i}} \end{aligned} \tag{10}$$

In matrix notation

$$[F_D] = [C][F_R] \tag{11}$$

where $[F_D]$ are the FFT coefficients of $[D]$ and $[F_R]$ are the FFT coefficients of $[R]$. This shows the concentration matrix $[C]$ to be invariant under the FFT.

No initial spectrum variable normalization (i.e. variance stabilization or mean adjustment) was performed for any of the analyses presented in this work. While variance stabilization is common for data in which different variable measurements come from widely differing parent populations, it is not necessarily desired for chemometric application where loss of infor-

mation of the relative lengths of the component axes may in fact damp features.³ For the AES spectra analyzed here, the samples were expected to vary in intensity primarily in the Ti- and N-sensitive energy range (i.e. 380–390 eV) with variation observed in other ranges due principally to random error. Stabilization of variances in this case where required precision approaches the level of noise would effectively mask the desired stoichiometric patterns.

TARGET FACTOR ANALYSIS

Method

In TFA we seek to decompose the observation matrix $[D]$ into a concentration matrix $[C]_{ox2}$ and a characteristic feature spectra matrix $[R]_{2xv}$, which corresponds to the concentrations and characteristic spectra of Ti and N for a given sample.

To accomplish this, the covariance matrix $[Z] = [D]^T[D]$ is constructed. The notation $[D]$ represents the oxv raw or transposed data matrix to be decomposed. The matrix $[Z]$ can be diagonalized by identifying an orthogonal matrix, $[Q]$, such that

$$[Q]^{-1}[Z][Q] = [\lambda_i d_{ij}] = [\lambda] \tag{12}$$

where the λ_i are the set of eigenvalues of $[Q]$ and d_{ij} the Kronecker Delta function. The columns of $[Q]$ represent the eigenvectors of $[Z]$. Malinowski⁴ derives the following relationships between $[Q]$ and the matrices $[R]$ and $[C]$

$$[Q]^T = [R] \tag{13}$$

$$[D][Q] = [C] \tag{14}$$

The eigenvectors Q_i of $[Z]$ are orthogonal and n (the number of eigenvectors) equals the rank of $[Z]$ (which is generally the minimum of $v =$ number of variables or $o =$ number of observations, owing to the presence of experimental error). For illustration, we assume $v < o$ throughout the remainder of this section.

It is required to show that the system of equations is adequately represented by only $n' = 2$ dimensions, corresponding to Ti and N components of the spectra.

The procedures used to determine n' , the dimensions of the system that explain variance in the data points not attributed to experimental error, center on use of eigenvalues. For $[Z]$ representing a covariance matrix, the corresponding eigenvector Q_i represents the relative magnitude of error explained by Q_i to the total variance $v; (\lambda_i/v) \times 100 =$ per cent of total variance explained by Q_i . Consequently, the larger λ , the more important its corresponding eigenvector as a dimension in explaining system variance.

Dimension reduction is achieved by only including 'significant' eigenvectors in the reconstruction of the data matrix $[D]$. Assuming the Q_i vectors are ordered 1, 2, ..., v by decreasing magnitudes of their corresponding λ_i values, decision rules for inclusion of the n' largest eigenvectors are employed and the reduced dimension $[D^{++}]$ matrix is computed

$$[D^{++}]_{oxv} = [C^{++}]_{oxn'} [R^{++}]_{n'xv} \tag{15}$$

This dimension $n' < v$ is acceptable when $[\Delta D] = [D] - [D^{++}]$ represents error equivalent to an estimate of experimental error.

Principal component analysis (PCA) is utilized to perform the orthogonal decomposition of $[D]$ and determination of the eigenvector matrix $[Q]$. We refer the reader to Ref. 14 for a comprehensive description of the PCA procedure. For our application, a simple Scree chart¹⁵ of percentage variance explained by eigenvalue order was sufficient to determine the dimension beyond which the plot leveled to a percentage variance attributed to experimental error alone. The order of the eigenvalue prior to the leveling out of the variance represents the n' value.

The decomposition of $[D^{++}]$ into orthogonal principal component matrices $[C^{++}]$ and $[R^{++}]$ does not ensure that the columns of $[C^{++}]$ represent the Ti and N atomic concentrations, respectfully, for the o sample rows; nor does it ensure that the rows of $[R^{++}]$ correspond to the characteristic features of Ti and N, respectively. In fact, given the high degree of overlap at ~ 387 eV, Ti and N components indicate a strong correlation rather than orthogonality of dimensions. Consequently, an oblique rotation of the $[R^{++}]$ and $[C^{++}]$ matrices is required that will match that of the natural system.

Malinowski details a least-squares approach to identifying a transformation matrix $[T]$

$$[D^{++}] = ([C^{++}][T])([T]^{-1}[R^{++}]) \quad (16)$$

$$[D^{++}] = ([C_{\text{real}}])([R_{\text{real}}]) \quad (17)$$

where $[C_{\text{real}}]$ is the column matrix of Ti and N atomic concentrations and $[R_{\text{real}}]$ is the row matrix of Ti and N characteristic AES features.

A matrix $[C_{\text{seed}}]$, comprised of the known Ti and N atomic concentrations for the samples, is used in the model training. For all analysis the $[C_{\text{seed}}]$ matrix of known sample composition coefficients is obtained from analyses of XPS spectra obtained for each of the training samples.

Analysis results

An initial FFT of the 55 sample spectra indicates that 20–30 FFT coefficients are sufficient to recreate the original spectra accurately. This is shown in Fig. 3, where the rate of change of absolute magnitude of FFT coefficients drops noticeably after 20. The following TFA set of results was obtained with reduced data consisting of the first 30 FFT coefficients per sample. All analysis utilized the 55 observation spectra described in Table 1.

The PCA analysis was performed with no variable variance stabilization or mean adjustment, i.e. the data covariance matrix was used for matrix decomposition. The PCA percentage variance explained by eigenvector results (Table 2) indicates that the data system of equations is accurately defined by the first one or two principal components, with $> 99.9\%$ variance explained.

The TFA results (Table 3) indicate low average error values of the TFA-predicted Ti/N ratios compared to the given XPS seed ratios, but sufficient lack of fit exists to warrant a re-evaluation of the basic linear model assumption and/or source of estimation error.

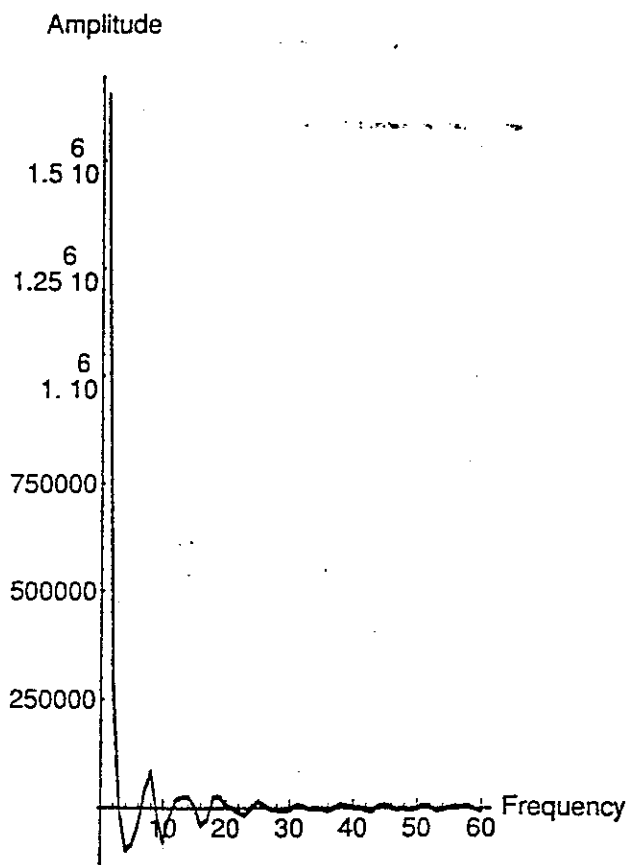


Figure 3. Representation of the first 60 fast fourier transform (FFT) coefficients of the original AES spectrum of 270 points by energy.

Examination of the original data spectra reveals a between-sample intensity-scale 'shift' that cannot be attributed to random error or to the difference in Ti:N concentration. For example, Fig. 4 displays 19 Auger

Table 2. Principal component analysis—no variance stabilization transform

Principal component no.	Eigenvalue	Cumulative % variance explained
1	3.012×10^{12}	99.995
2	1.190×10^6	99.999
3	15034714	100.000
4	7277168	100.000

Table 3. Target factor analysis results for predicting Ti/N composition ratio: no variance stabilization transform

Sample no.	XPS seed value	TFA/AES predicted value	% Error
1	1.057	1.087	+2.84
2	1.065	1.083	+1.69
3	1.595	1.570	-1.57
4	0.973	0.967	-0.62
5	1.000	0.975	-2.47
6	1.002	1.015	+1.30

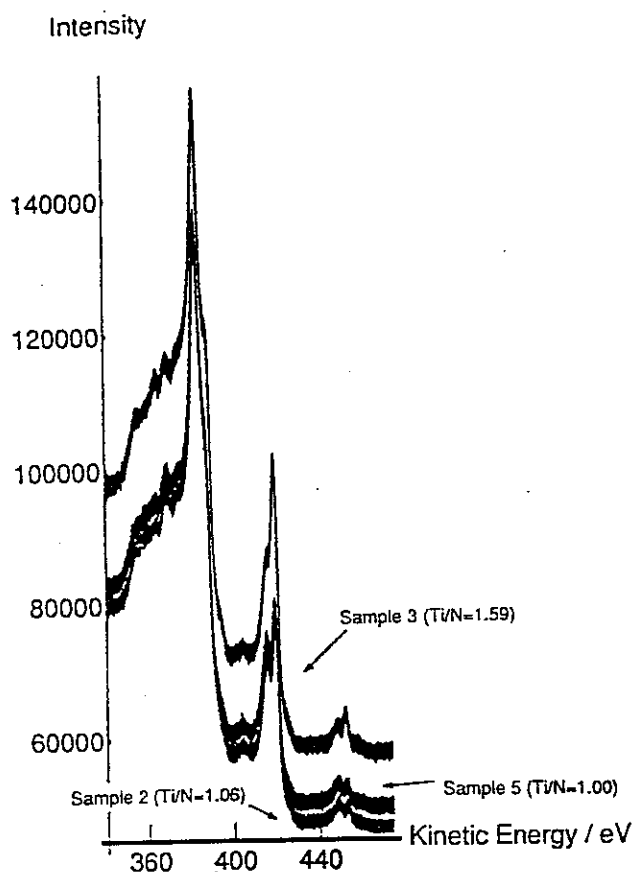


Figure 4. Raw data AES spectra of three samples with the between-sample 'shift' effect far exceeding more subtle peak shape changes due to film stoichiometry. The shifting effect is uncorrelated with Ti/N component magnitude and assumed to be a function of the data collection procedure.

electron spectra obtained from three samples (samples 2, 3 and 5). The XPS-generated Ti/N ratios are effectively no different between samples, yet a visual shift exists in the spectra. This nuisance variation was hypothesized to be due to the difference in exact positioning of a sample at the focal point of the electron energy analyzer input lens, causing multiplicative shifting of intensity values between samples. This only occurs when samples are changed, hence the good precision for data collected for a single sample run. The relative intensity value differences within any sample spectra are preserved between samples of like Ti:N concentration ratios under this model.

We capture the concept of a scaling effect between samples

$$d_{ij} = \sum_{l=1}^n K_l (c_{il} r_{lj}) \quad (18)$$

where d_{ij} is the j th observation of the i th sample spectrum, $i = 1, \dots, o$, $j = 1, \dots, v$, n is the number of significant dimensions and K_l is the scalar specific on the i th sample.

The basic assumption here is that K_l does not affect the relative intensities recorded from one electron energy level to another, thereby preserving the concept of standard Auger transition features in like materials. Only the intensity slope from energy point to successive point is impacted and such an effect (k_j) is constant throughout any given spectrum.

Given this 'scale' model as defined in Eqn (18), the PCA decomposed matrix, denoted $[C^*]$ for two significant dimensions or factors, is now represented by

$$[C^*] = \begin{bmatrix} K_1 c_{11} & K_1 c_{12} \\ K_2 c_{21} & K_2 c_{22} \\ \vdots & \vdots \\ K_o c_{o1} & K_o c_{o2} \end{bmatrix} \quad (19)$$

where c_{il} is the concentration of the l th principle component of the i th sample.

It is no longer appropriate to target transform $[C^*]$ to the XPS-seeded matrix, where $c_{i1} + c_{i2} = 1$ for all i . The relation that holds instead is the relative concentration of the first column element of $[C^*]$ to the second column element

$$\frac{C_{seedi1}}{C_{seedi2}} \sim \frac{K_i c_{i1}}{K_i c_{i2}} = \frac{c_{i1}^*}{c_{i2}^*} \quad (20)$$

Consequently, the sum-of-squares expression involved in the target procedure minimization is of the form

$$\sum_{i=1}^o [\text{Ratio}(XPS)_i - \text{Ratio}(C^*T)_i]^2 \quad (21)$$

where

$$\text{Ratio}(XPS)_i = \frac{C_{seedi1}}{C_{seedi2}}$$

$$\text{Ratio}(C^*T)_i = \frac{K_i (c_{i1} t_{11} + c_{i2} t_{12})}{K_i (c_{i1} t_{21} + c_{i2} t_{22})}$$

Minimization of expression (21) with respect to the four elements of (T) no longer has a closed form solution. A random initial value search procedure employing a steepest descent algorithm to locate a global minimum was performed. The resulting TFA percentage deviations of the predicted atomic concentration ratios from XPS-seed ratios are given in Table 4 along with an associated two-sided Student's t -hypothesis-test for equality. At $\alpha = 0.05$ (i.e. to 95% confidence), the predicted ratio is not significantly different from the XPS-seed ratio for all but one sample (sample 5), whose P value is 0.008.

Table 4. Target factor analysis results for predicting Ti/N composition ratios: scalar shift modification applied

Sample no.	XPS seed value	TFA/AES predicted value	% Error	P value for two-sided t -test H_0^* ($Ti/N_{XPS} = Ti/N_{AES}$)
1	1.057	1.064	+0.66	0.07
2	1.065	1.061	-0.38	0.57
3	1.595	1.593	-0.13	0.84
4	0.973	0.981	+0.82	0.52
5	1.000	0.987	-1.30	0.008
6	1.002	1.013	+1.10	0.07

* Probability, calculated under the null distribution, of obtaining a test result as extreme as that observed in the sample. A high P value (e.g. $P > 0.05$) supports failure to reject H_0 .

Discussion

The TFA method yields the marginal predictive capability of the Ti/N atomic concentration ratio from FFT AES spectra when a scaling modification is employed. In five of six samples, the AES-based predicted Ti/N ratios were not statistically different than XPS-measured ratios for $\alpha = 0.05$. The borderline fit for the remaining sample using this modified TFA model does not appear to be due to additional factors larger than two that are unaccounted for, given the >99.9% cumulative variance explained by the first two principal components. It is more likely that the basic model assumption of linearity is not sufficient within the very precise Ti/N relative concentration window of interest. Knowledge of variation in sample positioning during data collection suggested the scaling modification that resulted in an average 59% reduction in absolute deviation between seed and predicted values. However, additional physical causes for system non-linearity are unknown and no readily apparent transforms are suggested from residual analysis to improve the fit.

NEURAL NETWORK ANALYSIS

Method

In the face of a potentially non-linear model and a stringent concentration precision requirement of $\pm 0.5\%$ atomic concentration, an artificial neural network approach was implemented. Among the reported strengths of this approach are a network's ability to converge to correct prediction or classification estimates in the presence of highly non-linear relationships between input and output information and effectively filter noise or non-essential nuisance effects present in the data sample. The apparent shifting of Auger electron spectra between samples (Fig. 4), which was attributed to the data collection methods, could be filtered as non-essential by a correctly designed neural network.

The MLP used in this study is a two-layer feed-forward neural network (Fig. 2) that takes as input an information vector and utilizes a network of weights and hidden nodes to compute, in forward fashion, an output prediction vector. Back propagation refers to the method by which the feed-forward network is 'trained' to map input to output vectors and is widely attributed to Rumelhart,¹⁶ with a similar algorithm having been introduced by Packer.¹⁷ During training, the network computes a prediction vector. The output prediction vector is compared to the 'known' output and an error metric is computed. A steepest descent gradient is computed, which indicates how the weight structures should change in order to minimize the total error. This descent gradient information is 'back propagated' to the weighting structures at both the input and hidden layers and the weights modified accordingly.

The MLP is an extension of the single-layer perceptron first introduced by Rosenblatt in 1958.²⁰ The single-layer discrete network can be shown to converge upon a correct training solution when the pattern classes are linearly separable. The MLP can, on the

other hand, implement arbitrary complex (non-linear) input/output mappings for separating classes.

For the feed-forward MLP utilized to predict the Ti/N relative atomic concentration ratio, two layers were constructed consisting of a 20 node input vector of FFT coefficients from the spectra, a 10 node hidden layer and a single-node output value of the Ti:N composition ratio. Each layer is fully interconnected to the successive layer via a weight structure. The value of a hidden node $x_{j,n}$ is the hyperbolic tangent transform of the inner product of the input vector X_i and connecting weight vector W_j

$$x_{j,n} = f\left(\sum_{m=1}^{20} x_{i,m} w_{j,mn}\right) \quad (22)$$

for

$$f(z) = \frac{e^{+z} + e^{-z}}{e^{+z} - e^{-z}} \quad (23)$$

Similarly, the output prediction ratio is the hyperbolic tangent of the inner product of X_j and W

$$x_{k,1} = f\left(\sum_{n=1}^{40} x_{j,n} w_{k,n1}\right) \quad (24)$$

During training of this supervised network, the global error between the network computed Ti/N ratio and XPS-seeded ratio is computed for a given input vector. In the case of a single output node, the global error E is simply half of the squared difference between known and predicted ratios. Back-propagation of error is achieved by computing the gradient of the total error with respect to the weights $w_{k,n1}$ and adjusting the weight proportionally in the direction of the steepest error descent. The error gradient with respect to the $w_{k,n1}$ values is accomplished by application of the chain rule as follows

$$\frac{\partial E}{\partial w_{k,n1}} = \left(\frac{\partial E}{\partial x_{k,1}}\right) \left(\frac{\partial x_{k,1}}{\partial net_{k,1}}\right) \left(\frac{\partial net_{k,1}}{\partial w_{k,n1}}\right) \quad (25)$$

for

$$net_{k,1} = \sum_{n=1}^{40} (x_{j,n} w_{k,n1}) \quad (26)$$

and all other terms defined as in Eqn (3). By similar application of the chain rule, the weights $w_{j,mn}$ are updated in the direction of steepest global error descent using the generalized delta rule. The specific rule used to train the network in this study is a modification of the generalized delta rule called the Extended Delta Bar Delta (EDBD) algorithm.¹⁹ The EDBD algorithm utilizes heuristic rules by which a learning coefficient (a coefficient that dictates the size of the step taken along the path of steepest descent) and momentum term (a weight change that is proportional to the previous weight change) are adjusted to accomplish faster convergence. For more comprehensive discussion of MLP, back-propagation algorithms or EDBD algorithms, the reader is directed to Zurada,²⁰ Schalkoff²¹ or Minai and Williams.¹⁹ Following back-propagation of error and weight update, the network is then presented with another input vector and the entire process repeats until

the network converges to within the required precision of the global error.

Results

The first 30 FFT coefficients per spectrum for 45 of the 55 sample spectra described in Table 4 were used to train the MLP back-propagation network. The remaining 10 spectra were used to test the network. The network was trained for a total of 27 000 input presentations. The plot of presentation cycle number versus root-mean-square error of the predicted vs. supervised output ratio is shown in Fig. 5.

Table 5 lists the Ti/N atomic concentration ratios predicted by the network along with percentage error from associated XPS-seed ratio values for the train and test sets. Results of two-sided *t*-tests comparing the network-predicted sample means to the XPS-analyzed sample mean yield no significant difference between means for any sample types, with both train and test group *P* values ≥ 0.50 . The large *P* values strongly support the failure to reject a null hypothesis of Ti/N mean equality.

The sample variance of the within-sample group MLP predicted ratios for all six sample groups was not significantly larger than the XPS ratio sample variances. The null hypothesis H_0 [$\sigma^2_{Ti/N}(\text{AES}) \leq \sigma^2_{Ti/N}(\text{XPS})$] could not be rejected, with all sample test *P* values > 0.65 .

Owing to the relatively small number of different sample materials used in the above analyses, 20 additional spectra collected using the same spectrometer after the electron energy analyzer had been changed from one with three parallel detectors to one with five parallel detectors (see Experimental section). The 20 spectra consisted of five repetitions, each of four samples; the XPS-estimated Ti/N ratios are shown in Table 1 for the samples, referred to as samples 7-10. Figure 6 shows clearly the relatively large intensity shift effect evident between spectra collected with the two different analyzers, reflecting the increased sensitivity of the system with five parallel detectors. An equivalent scalar transform is applied within subsequent TFA and MLP analysis to minimize the impact of the non-linear effect on the AES spectra due to the analyzers.

Table 5. Neural network Ti/N composition ratio predicted values

Sample no.	XPS seed value	Ti/N composition ratio		<i>P</i> value for <i>t</i> -test H_0 * ($Ti/N_{XPS} = Ti/N_{AES}$)
		NN/AES predicted value	% Error	
<i>Training set</i>				
2	1.065	1.065	0.000	>0.90
3	1.595	1.595	0.000	>0.90
4	0.973	0.975	0.206	0.74
5	1.000	1.000	0.000	>0.90
6	1.002	1.000	0.200	0.71
<i>Test set</i>				
1	1.057	1.059	0.203	0.67

* A large *P* value supports failure to reject H_0 .

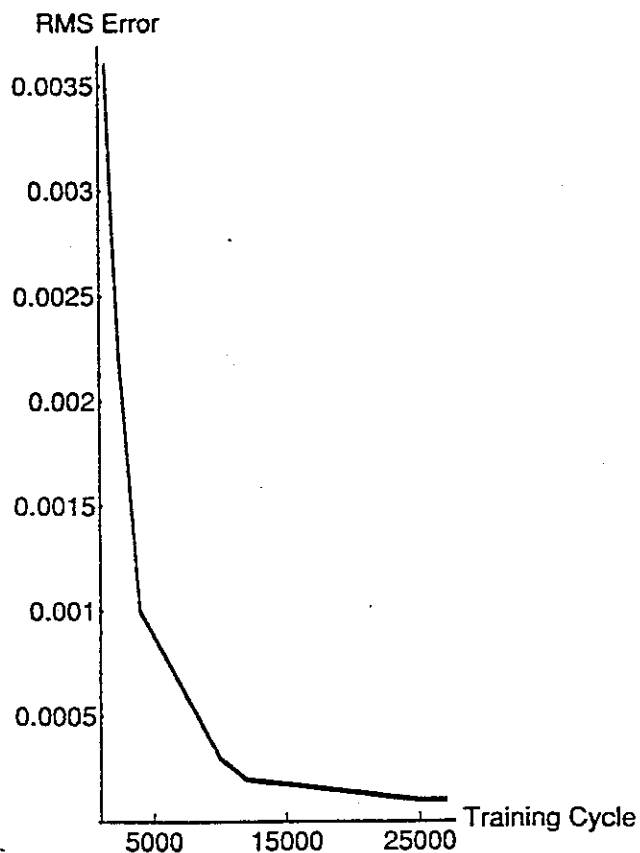


Figure 5. Plot of network training cycle versus root-mean-square (RMS) error of the predicted versus supervised output Ti:N ratio value for network trained using samples 2-6.

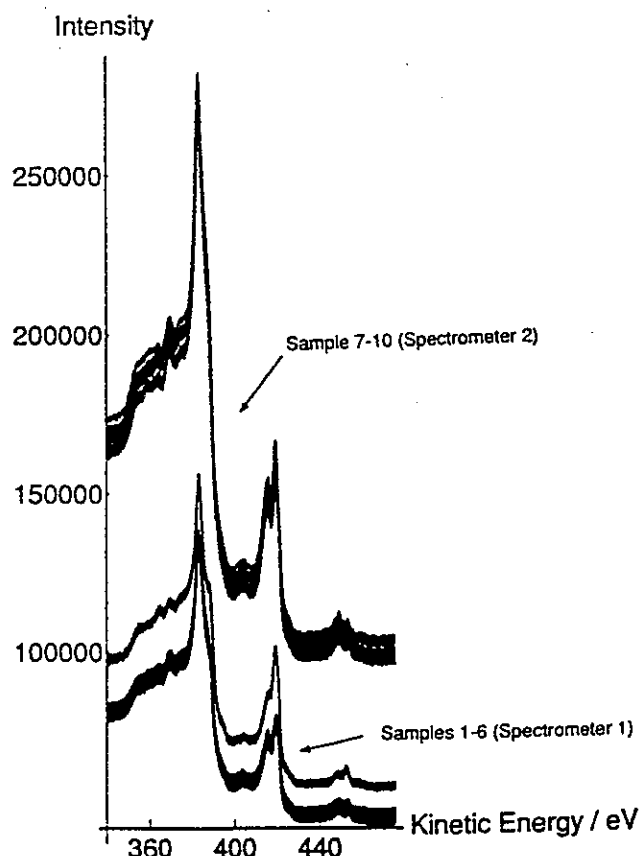


Figure 6. Raw data AES spectra of 10 samples from two different electron energy analyzer configurations on the same instrument. A significant intensity shift effect exists. Scalar transform partially removes the effect.

Table 6. Target factor analysis results for predicting Ti/N composition ratios from data from two different energy analyzer configurations

Sample no.	XPS seed value	Ti/N composition ratio		
		TFA/AES predicted value	% Error	P value for t -test H_0^* ($Ti/N_{XPS} = Ti/N_{AES}$)
1	1.057	1.005	4.92	<0.001
2	1.065	1.004	5.73	<0.001
3	1.595	1.566	1.82	0.377
4	0.973	0.987	1.44	0.006
5	1.000	0.988	1.20	0.005
6	1.002	0.994	0.80	0.134
7	0.981	1.007	2.65	<0.001
8	0.930	0.994	6.88	<0.001
9	0.914	0.991	8.42	<0.001
10	0.957	1.078	12.64	<0.001

* A large P value supports failure to reject H_0 .

The modified TFA analysis was repeated using all 10 samples. Results of the two-sided t -tests comparing TFA Ti/N ratio means to XPS sample means are displayed in Table 6. The percentage error between the TFA means and targeted XPS seed values is significantly larger than that measured for the original, single analyzer data, with eight of the 10 samples exhibiting significantly different mean value differences. This indicated that the scalar transform adjustment for the shift effect was not sufficient to eliminate nuisance sources of non-linearity between spectra. Table 7 includes the estimated mean variances for all samples obtained from the TFA analysis. With the exception of sample 3, all other mean variances are within the range exhibited by the XPS sample analyses.

A second MLP neural network was trained and tested with the combined 10 sample data set, totaling 75 spectra. A scaling transform was applied to the raw data spectra to help mitigate the between-spectrometer sampling shift. The transform was computed as follows

$$x_{ij} = d_{i(j+1)}/d_{ij} \quad (27)$$

Table 7. Comparison of sample variances of predicted Ti/N ratios between TFA and neural network (NN) analyses

Sample no.	XPS seed value	Ti/N composition ratio estimated standard deviation (P value)*	
		TFA/AES value	NN/AES value
1	0.047	0.006 (>0.50)	0.096 (0.04)
2	0.099	0.006 (>0.50)	0.035 (>0.50)
3	0.140	0.456 (<0.01)	0.035 (>0.50)
4	0.061	0.012 (>0.05)	0.025 (>0.50)
5	0.051	0.007 (>0.50)	0.047 (>0.50)
6	0.051	0.003 (>0.50)	0.101 (0.05)
7	0.051	0.005 (>0.50)	0.019 (>0.50)
8	0.050	0.010 (>0.50)	0.028 (>0.50)
9	0.043	0.008 (>0.50)	0.037 (>0.50)
10	0.058	0.065 (0.42)	0.023 (>0.50)

* P values listed are F test results for the null hypothesis $H_0: \sigma_{AES}^2 \leq \sigma_{XPS}^2$. P values < 0.05 indicate a failure to match or diminish the Ti/N estimate by AES over XPS.

Table 8. Neural network training and test set results of Ti/N composition ratio predictions

Sample no.	XPS seed value	Ti/N composition ratio		
		Auger predicted value	% Error	P value for t -test H_0^* ($Ti/N_{XPS} = Ti/N_{AES}$)
Training set				
2	1.065	1.064	0.09	>0.90
3	1.595	1.595	0.00	>0.90
4	0.973	0.974	0.10	>0.90
5	1.000	0.999	0.10	0.88
7	0.981	0.981	0.00	0.85
9	0.914	0.918	0.44	0.42
10	0.957	0.957	0.00	0.88
Test set				
1	1.057	1.024	3.12	<0.01
6	1.002	1.005	0.30	0.72
8	0.930	0.941	1.18	0.09

* A large P value supports failure to reject H_0 .

where d_{ij} is the j th observation of the i th sample spectrum $j = 1, \dots, 270$ and $i, \dots, 75$.

The network consisted of an input layer of 30 Fourier coefficients and a hidden layer of 10 nodes, pruned to four. Hidden node removal (pruning) occurred when the network training performance did not increase after 8000 presentations of the training set and the network performance did not degrade by more than 25% when the candidate node was disabled. This procedure of network performance stabilization followed by node pruning continued until the remaining four hidden nodes could not be disabled without network performance falling below 75% accuracy. Pruning attempts to minimize network complexity and is based on the concept that a simpler network that does well on a training set will generalize better and consequently perform better on a test set. The final output layer consisted of a single Ti/N ratio node, as before. The network was trained for a total of 99 000 input vector presentations.

The network was trained using the EDBD algorithm, as before, with a hyperbolic tangential transfer function used for both the hidden and output layers.

Table 8 lists the results of the test and training sample Ti/N ratio estimates. Only sample 1 of the test set results in rejection of the null hypothesis of no difference between XPS and AES mean estimates, with an error of 3.12%. The remaining nine samples (seven training and two test samples) show no significant differences. The mean variance estimates for this MLP model are found in Table 7, with all sample P values ≥ 0.04 .

Discussion

Both of the MLP neural networks trained for this study accurately predicted Ti and N atomic concentrations for Auger spectra when XPS-analyzed spectra were used as the seed or known concentrations. In addition, the variance of the concentration estimates was generally smaller or not greater than that of the reported XPS-based values. We note that the neural network

solution accommodated non-linear shift effects introduced both within a single electron energy analyzer data set as well as between analyzer samplings. The marginal estimation of one test sample ratio by the neural network would likely improve with availability of additional training samples. The modified TFA procedure produced marginal results when faced with between-sample shifts on a single analyzer, but failed to estimate accurately the Ti/N concentrations for samples taken between two similar analyzers. Given these results, the neural network models appear preferable over the TFA procedure in being able to accommodate suspected non-linear nuisance effects and yield consistent and accurate Ti/N concentration predictions within $\pm 0.5\%$ atomic concentration precision.

CONCLUSION

This work sought to realize the logistic and analysis benefits of AES methods over other spectroscopy methods by developing a data analysis method that matched the stoichiometric discrimination precision

capabilities of XPS. A modified TFA approach was developed and yielded, at best, marginally comparable results to standard XPS analysis of TiN_x films. However, an MLP analysis design applied to the AES spectra was successful in both matching the discrimination capability of XPS analysis methods and yielding estimated Ti/N ratio variance estimates that were not larger or significantly smaller than those produced by the XPS analysis methods for the majority of samples analyzed. The improved performance of the MLP over the TFA pattern recognition systems is attributed to the neural network's ability to learn complex, non-linear mappings.

From these results, we conclude that neural network technology holds great promise for achieving precision pattern recognition in other chemical spectroscopy applications. Our work continues in the areas of resolution of other spectral interference/overlap problems, such as chemical state separation in XPS data, development of rapid and routine analyses packages for complex spectral data sets and performing routine spectral analysis and sample classification, such as particle and defect identification using Auger electron and electron microprobe methods.

REFERENCES

1. A. L. Testoni, submitted to *Surf. Interface Anal.*
2. S. Gaarenstroom, *J. Vac. Sci. Technol.* **16**, 600 (1979).
3. D. G. Watson, W. F. Stickle and A. C. Diebold, *Thin Solid Films* **193/194**, 305 (1990).
4. E. R. Malinowski and D. G. Howery, *Factor Analysis in Chemistry*. Krieger Publishing, FL (1989).
5. A. L. Sumner, S. K. Rogers, G. L. Tarr, M. Kabrisky and D. Norman, *Proc. SPIE* **1294**, 138 (1990).
6. J. U. Thomsen and B. Meyer, *J. Magn. Reson.* **84**, 212 (1989).
7. B. R., Upadhyaya, G. Muthai and E. Eryurek, *Proc. 29th IEEE Conference on Decision and Control*, Vol. 6, p. 3671. IEEE, New York (1990).
8. J. M. Minor, *Proc. ISA/89 International Conference and Exhibition*, Vol. 3, p. 1155 (1989).
9. R. Sobral, S. Canu, C. Bernard and H. Axelrad, *10th Int. Workshop: Expert Systems and Their Applications*, p. 53. EC2, Nanterre, France (1990).
10. J. R. Long, V. G. Gregorior and P. J. Gremperline, *Anal. Chem.* **62**, 1791 (1990).
11. R. P. Lippmann, *IEEE ASSP Mag.* **4**, 4 (1987).
12. P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*. Wiley, New York (1976).
13. R. W. Rozett and E. M. Peterson, *Anal. Chem.* **47**, 1301 (1975).
14. I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag, New York (1986).
15. R. B. Cattell, *Multivar. Behav. Res.* **1**, 245 (1966).
16. D. E. Rumelhart and J. L. McClelland (eds), *Processing: Explorations in the Microstructure of Cognition*, Vol. 1. MIT Press, Cambridge, MA (1986).
17. D. B. Packer, *Report TR-47*. MIT Center for Computational Research in Economics and Management Science, Cambridge, MA (1985).
18. F. Rosenblatt, *Psychol. Rev.* **65**, 386 (1958).
19. A. A. Minia and R. D. Williams, *International Joint Conference on Neural Networks*, Vol. 1, p. 676. IEEE TAB Neural Network Committee, New York (1990).
20. J. M. Zurada, *Introduction to Artificial Neural Systems*, pp. 163-250. West Publishing Co., New York (1992).
21. Schalkoff, *Pattern Recognition*, pp. 236-263. Wiley, New York (1992).