Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

Development of a multiple-element XPS spectrum decomposition method based on multiple peaks

Ryo Murakami[a*], Hiromi Tanaka[a], Hiroshi Shinotsuka[b] ,
Kenji Nagata[b], Hideki Yoshikawa[b]

[a]*Department of Electrical and Computer Engineering,*
National Institute of Technology, Yonago College, Tottori 683-8502, Japan
[b]*Research and Services Division of Materials Data and Integrated System,*
National Institute for Materials Science, Tsukuba 305-0047, Japan

Keywords
X-ray photoelectron spectroscopy, Automatic spectrum analysis, Active Shirley, Bayesian information criterion, Multi-element spectra.

Abstract

The increasing sophistication of measurement methods such as operand analysis and three-dimensional analysis has led to a need for the automatic analysis of a large number of spectra. A measured spectrum is presently compared with a reference spectrum of a known single-phase compound sample, often to estimate the component ratio of a studied compound. However, it is difficult to automate all processes and there is the problem in that the analyzer's arbitrariness is included. The present study analyzed the spectrum of a multiphase compound using a reference spectrum and developed a method of fully estimating the component ratio of a compound in an automatic manner. To use a reference spectrum comprising multiple peaks as a basis function, a peak separation method based on the Bayesian information criterion was adopted. In particular, it was clarified that the estimation accuracy was improved by analyzing multiple-element spectra simultaneously rather than by analyzing only a single-element spectrum.

1. Introduction

X-ray photoelectron spectroscopy (XPS) is a measuring method capable of analyzing element species and compound species on a substance surface. It is used for material development and quality control in industry. Studies have investigated the visualization of the chemical composition distribution of materials by depth direction analysis and in-plane direction analysis and the evaluation of process reactions by operand analysis. The analysis data comprise a large number of spectra, and there is thus a need to develop a method that automates the analysis of a spectrum in the estimation of the component ratio of a compound.

In the analysis of a measured spectrum of an unknown sample, the spectrum is compared with a reference spectrum of a known sample using data analysis software. As a reference spectrum of a known sample, measurements obtained by different devices and literature data are often used. Reference spectra with different energy resolutions relating to the measuring device and different peak shift amounts relating to the surface charge of the sample may be mixed, and it is thus effective for the data analysis software to perform peak separation of the reference spectra using the pseudo-Voigt function [1] to correct the shape difference. This processing can be interpreted as sparse modeling of the reference spectrum, because the measured spectrum comprising many points is represented by a few parameters of the mathematical function.

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

Algorithms that automatically search for optimal solutions of mathematical functions describing reference spectra have been proposed [2–6]. However, their effective use has not yet been established, automation from the measured spectra of unknown samples to the estimation of component ratios of compounds is difficult, and the analyst's arbitrariness is included.

To solve the above problems, we developed a method of expressing the reference spectrum of a single-phase compound using a mathematical function and automatically analyzing the spectrum of a multi-phase compound considering fluctuation relating to the device measuring the peak parameter of the reference spectrum. To use the reference spectrum as a basis function comprising multiple peaks (represented by mathematical functions), we use the peak separation method based on the Bayesian information criterion (BIC) developed by Shinotsuka et al. to derive the basis function [2]. This method estimates the number of peaks appropriately using the BIC and searches for the optimum solution from many candidate solutions using the initial values of a large number of different peak parameters (i.e., intensity, position, and width). In this manner, the reference spectrum of a single-phase compound was automatically peak separated such that it could be used as a basis function of the analysis. However, if only the spectra of single elements are analyzed, the solutions vary widely owing to the superposition and similarity of the reference spectra obtained when the XPS spectra of multiphase compounds are analyzed using the reference spectra of multiple candidate substances. It was therefore clarified that the dispersion of the solution can be suppressed by simultaneously analyzing the multiple element spectrum.

2. Method

Generally, XPS spectra are measured for each element in multiple energy bands. The present study aims at simultaneously analyzing the XPS spectra measured for each element l and estimating the component ratios of multiphase compounds.

2.1. Model

2.1.1. Analysis Model of XPS Spectra Based on Reference Spectra

The analysis of the measured spectrum $D = \{x_i, y_i\}_{i=1}^{N}$ of specific energy bands using reference spectra is first described. Here, x is the binding energy, y is the intensity of the measured spectrum, and N is the number of data points. The spectra observed in XPS are expressed as the sum of the signal spectra and background as follows.

$$f(x) = S(x) + B(x). \tag{1}$$

Here, $f(x)$ is the observed spectrum, $S(x)$ is the signal spectrum, and $B(x)$ is the background. The signal spectrum

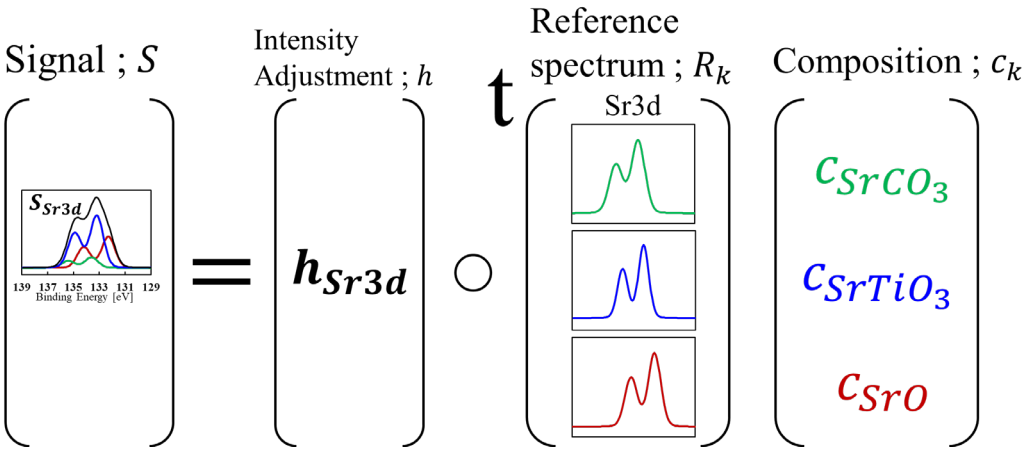

Figure 1: Matrix representation of the independent analysis model using reference spectra for single-element spectra.

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

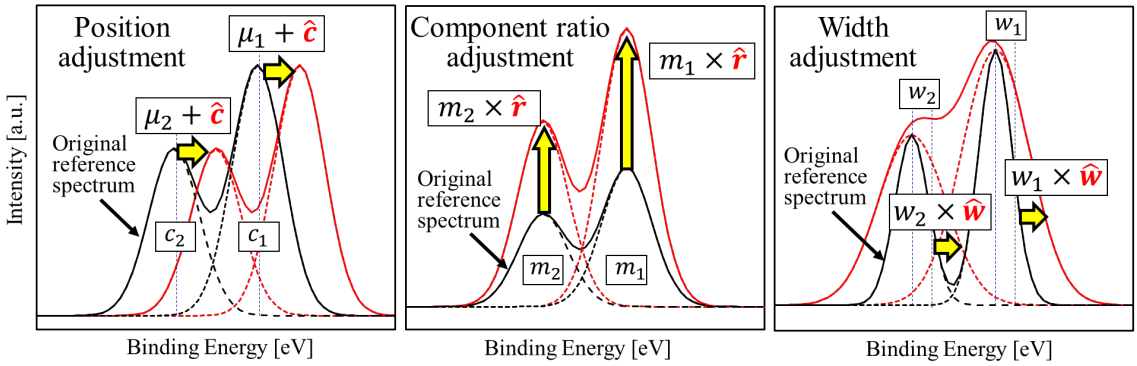

Figure 2: Overview of adjustments of the reference spectrum: (left) adjustment of position μ, (middle) adjustment of component ratio c, (right) adjustment of width w.

$S(x)$ comprises a linear sum of the reference spectra. $S(x)$ is therefore defined as follows.

$$S(x) = h\sum_{k=1}^{K} c_k R_k(x). \tag{2}$$

Here, $R_k(x)$ is the reference spectrum of compound k normalized by area considering the atomic fraction, K is the number of assumed compound species, c_k is the area ratio of the reference spectrum, and h is the intensity of the fitting function. In other words, c_k is the component ratio of the compound in the measured substance while h is the intensity of the signal spectrum depending on the intensity of the X-rays and measuring time. Figure 1 shows that the signal spectrum $S(x)$ can be decomposed into the reference spectra $R_k(x)$, which express the compound species, the component ratios of the compounds c_k, and the signal spectral intensity h. The circle symbol in Figure 1 represents the element-by-element matrix product and is known as the Hadamard product. Meanwhile, t in Figure 1 represents the transpose of the matrix. The reference spectrum $R_k(x)$ is defined by the following equations.

$$R_k(x; \hat{\mu}_k, \hat{w}_k) = \sum_{j=1}^{M_k} m_{kj}\, V\left(x; \hat{\mu}_k + \mu_{kj},\ \hat{w}_k w_{kj}, r_{kj}\right), \tag{3}$$

$$V(x; \mu,\ w, r) = (1-r)\frac{1}{(\ln 2)\sqrt{\pi} w}\exp\left\{-(\ln 2)\left(\frac{x-\mu}{w}\right)^2\right\} + r\frac{1}{\pi}\frac{w}{(x-\mu)^2 + w^2}. \tag{4}$$

Here, M_k is the number of peaks constituting the reference spectrum k, $V(x; c,\ w, r)$ is the pseudo-Voigt function normalized by the area, m_{kj} is the area ratio of peak j constituting the reference spectrum k, μ_{kj} is the position of peak j in the reference spectrum k, w_{kj} is the width of peak j in the reference spectrum k, and r_{kj} is the Lorentz–Gaussian ratio of peak j in reference spectrum k. $\{m_{kj}, c_{kj}, w_{kj}, r_{kj}\}$ are given by the separation peaks of the reference spectra of known single-phase compounds and literature data and are fixed values when we analyze the spectrum. Figure 2 is a conceptual diagram of the parameters $\{c_k, \hat{\mu}_k, \hat{w}_k\}$ used in adjusting the reference spectrum. c_k is the component ratio of the compound (reference spectrum k), $\hat{\mu}_k$ is a parameter used to adjust the peak position of reference spectrum k, and $\hat{w}_k$ is a parameter used to adjust the width of reference spectrum k.

In this paper, the background estimation adopts the commonly used Shirley method [7, 8]. The background $B(x)$ is defined by the following equation.

$$B(x; a, b) = \frac{\int_0^x S(u)du}{\int_0^\infty S(u)du}(a-b) + b. \tag{5}$$

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

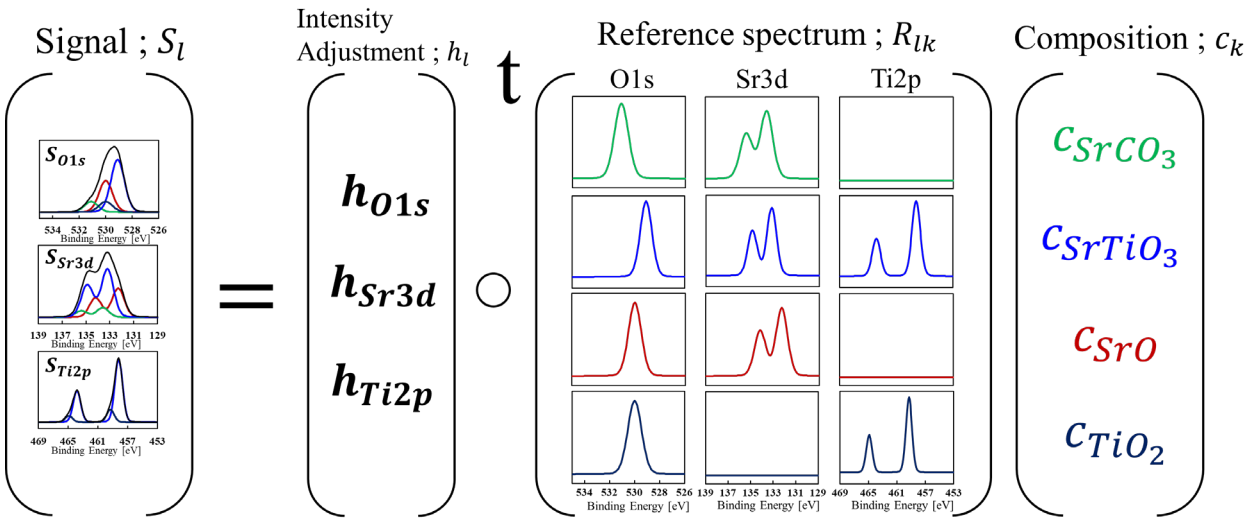

Figure 3: Matrix representation of the simultaneous analysis model using reference spectra for multiple-element spectra.

Here, a is the endpoint intensity on the high-bound energy side while b is the endpoint intensity on the low-bound energy side. In this paper, the parameter to be estimated is $\boldsymbol{\theta} = \{h, \{c_k, \hat{\mu}_k, \hat{w}_k\}_{k=1}^{K}, a, b\}$. The parameter $\boldsymbol{\theta}$ is determined by minimizing the error function $E(\boldsymbol{\theta})$ defined by the following equation.

$$E(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - f(x_i;\boldsymbol{\theta})\right)^2. \tag{6}$$

In this study, the search ranges of parameters were set as a half-width at half maximum $w = [0.3, 0.7]$, a position adjustment $\hat{\mu} = [-0.3, 0.3]$, and ratio of compound $c = [0, 1]$.

2.1.2. Simultaneous Analysis of Multiple-element XPS Spectra Measured in Multiple Energy Bands

XPS spectra are generally measured for each element l in multiple energy bands. Simultaneous analysis of multiple XPS spectra measured for each element l is described. Assuming that the distribution of the component ratio of the compound is uniform in the depth direction of the measurement sample, the component ratio of the compound is the same in the multiple-element XPS spectrum measured for the same sample. In other words, the component ratio c_k of the compound is a common parameter in the measured XPS spectrum for each element l. Therefore, as shown in Figure 3, the analysis model is formulated as matrix decomposition in which the reference spectrum is used as the basis matrix and the component ratio of the compound is used as the coefficient vector. The signal spectrum $S_l(x_l)$ in the energy band of element l is formulated identically as Equation (3), with the subscript l applied to variables and functions other than the component ratio c_k of the compound independent of the energy band, as in the following equations.

$$S_l(x_l) = h_l \sum_{k=1}^{K} c_k R_{lk}(x_l;\hat{\mu}_{lk},\hat{w}_{lk}), \tag{7}$$

$$R_{lk}(x_l;\hat{\mu}_{lk},\hat{w}_{lk}) = \sum_{j=1}^{M_{lk}} m_{lkj}\, V\left(x_l;\hat{\mu}_{lk}+\mu_{lkj},\ \hat{w}_{lk}w_{lkj}, r_{lkj}\right). \tag{8}$$

Here, the reference spectrum $R_{lk}(x_l;\hat{\mu}_{lk},\hat{w}_{lk})$ is zero if there is no reference spectrum $R_{lk}(x_l;\hat{\mu}_{lk},\hat{w}_{lk})$ in the energy band of element l.

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

2.2. Algorithm

2.2.1. Algorithm for Searching the Values of Initial Peak Parameters

We optimized the Levenberg–Marquardt method [9, 10] to carry out analysis in practical time. However, the Levenberg–Marquardt method is a gradient method, and the initial dependency is therefore strong and the method gives a local solution. Therefore, parameters $\{c_k, \hat{\mu}_k, \hat{w}_k\}$ for adjusting the reference spectrum were searched for globally to obtain suitable initial values (i.e., search ranges were a half-width at half maximum $w = [0.3, 0.7]$, a position adjustment of $\hat{\mu} = [-0.3, 0.3]$, and a ratio of compound $c = [0, 1]$). This search used a genetic algorithm [11–14] to perform 5000 trials of 30 points. The settings of the genetic algorithm were a mutant probability of 10% and crossover probability of 90%. In this search, considering the shift of the energy axis due to the charge-up effect, the energy position was corrected by ensuring that the energy position of the maximum intensity of the fitting function coincided with the energy position of the maximum intensity of the measured spectrum before calculating the error function. In the search of the initial value by the genetic algorithm, the endpoint intensities $\{a,\ b\}$ of the background were fixed at the endpoint intensities of the measured spectrum, and the maximum value h of the fitting function was also fixed at the maximum value of the measured spectrum. At this stage, the parameters $\{h, a, b\}$ are strongly affected by the spectrum noise. Afterward, to greatly improve the parameters, we apply the Levenberg–Marquardt method to optimize only $\{h, a, b\}$, which are fixed in the genetic algorithm. We use these parameters as initial values in the next optimization stage.

2.2.2. Automatic Identification of Compound Species in Actual Spectra

Actual analysis requires the selection of a combination of reference spectra because the compound species of the measured spectrum are often unknown. This processing is synonymous with model selection in the field of statistics. We analyzed the spectrum for each combination of candidate compounds and selected optimal solutions according to the BIC [15]. Adoption of the BIC allows the selection of a solution that has low error and is the combination of a few compounds. The BIC is defined by the following equation.

$$\mathrm{BIC} = -2\log\hat{L} + \lambda\log N. \tag{9}$$

Here, $\log\hat{L}$ is the maximum logarithmic likelihood, λ is the number of parameters, and N is the number of data points. We assume that the noise in the XPS spectrum follows a Gaussian distribution of variance σ^2. The probability distribution p_i and log-likelihood $\log L$ are given by the following equations.

$$p_i = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y_i - f_i)^2}{2\sigma^2}\right\}, \tag{10}$$

$$\log L = \sum_{i=n}^{N}\log p_i. \tag{11}$$

The maximum log-likelihood $\log\hat{L}$ is therefore given by the following equations.

$$\log\hat{L} = -\frac{N}{2}\{\log(2\pi\hat{\sigma}^2) + 1\}, \tag{12}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - f_i)^2\,. \tag{13}$$

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

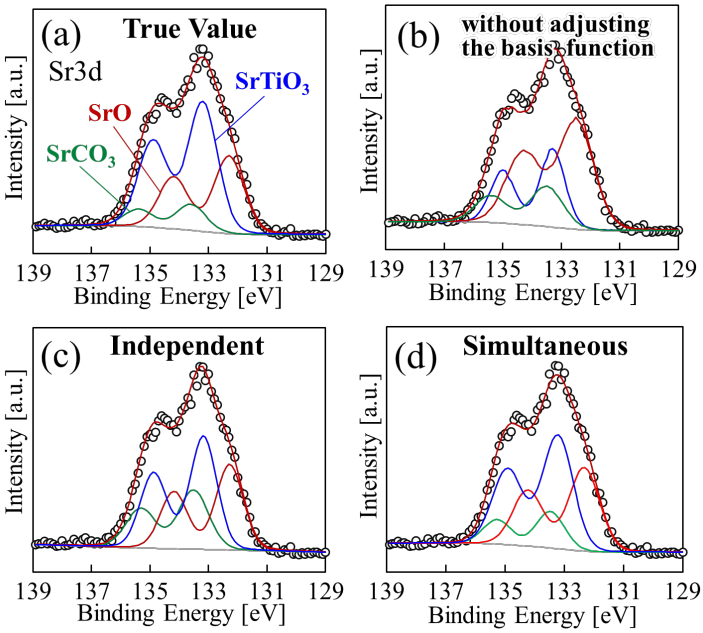

Figure 4: Analysis results of the Sr3d spectrum (artificial data) obtained using different methods: (a) true separated peaks, (b) analysis results obtained by expressing spectra to be analyzed as the linear summation of measured spectra of single-phase known samples (i.e., without adjusting the basis function), (c) analysis results obtained by independently analyzing the Sr3d spectrum using the proposed method, (d) analysis results obtained by simultaneously analyzing the spectra of three elements using the proposed method. Only the Sr3d spectrum is shown.

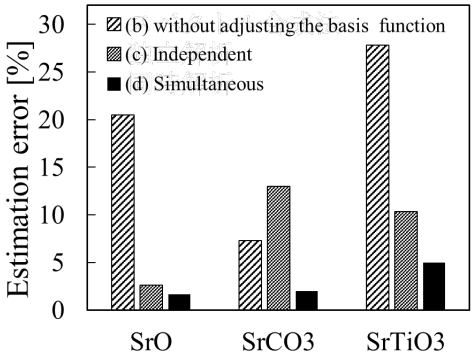

Figure 5: Error in the estimation of the component ratio of compounds in Fig. 3.1.1. (b) Without adjusting the basis function: the method of expressing spectra to be analyzed as the linear summation of measured spectra of single-phase known samples. (c) Independent: Independent analysis in the Sr3d spectrum. (d) Simultaneous: Simultaneous analysis of spectra of three elements (O1s, Sr3d, Ti2p).

3. Results

3.1. Simultaneous Analysis of Multiple-element XPS Spectra Using Reference Spectra

Referring to peak positions μ_{kj} of [SrO, $SrCO_3$, $SrTiO_3$, TiO_2] taken from the literature [16–21], we artificially produced spectra of three elements (O1s, Sr3d, Ti2p) using pseudo-Voigt functions, with a half-width at half maximum w_{kj} of 0.6 eV and a Lorentz–Gaussian ratio r_{kj} of 0.1 (i.e., the component of the Lorentz function is 10%) while considering the atomic fraction. Here, assuming the component ratio of the compound in the depth direction of the sample is uniform, we

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

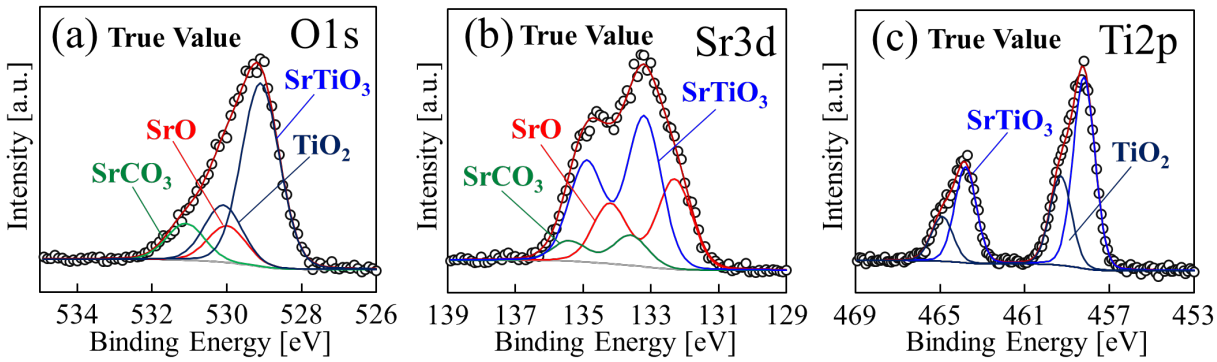

Figure 6: Artificial data of multiphase compounds ($SrCO_3$: 11.6%, SrO: 14.0%, $SrTiO_3$: 58.1%, TiO_2: 16.3%) generated from literature data: (a) O1s spectrum, (b) Sr3d spectrum, (c) Ti2p spectrum.

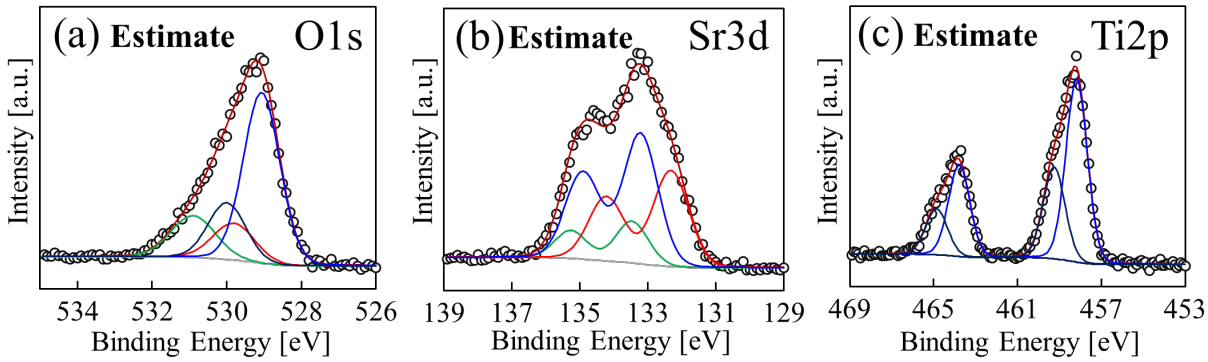

Figure 7: Analysis results obtained by simultaneously analyzing spectra of three elements using the proposed method: (a) O1s spectrum, (b) Sr3d spectrum, (c) Ti2p spectrum.

generated artificial data in which the component ratio of [SrO, $SrCO_3$, $SrTiO_3$, TiO_2] is approximately [11.6%, 14.0%, 58.1%, 16.3%]. We added Poisson noise to the artificial data generated at a level of about 3%.

Figure 4 shows (a) the true separated peaks of the Sr3d spectrum, (b) analysis results obtained by expressing the analysis spectrum by a linear sum of measured spectra of known samples (i.e., without adjusting the basis function), (c) analysis results obtained by independently analyzing the Sr3d spectrum using the proposed method, and (d) analysis results obtained by simultaneously analyzing the spectra of three elements (O1s, Sr3d, Ti2p) using the proposed method. Figure 5 shows the estimation errors of the component ratios of the compounds when using the different methods. The reference spectra used in the proposed method and the method without adjusting the basis function varied by ±0.1–0.2 eV in the width parameter, suggesting that that reference spectra were measured by devices with different energy resolution. Figure 4(a) shows that the analysis result obtained without adjusting the basis function differs from the true peak configuration, with an estimated maximum error of about 28%. Meanwhile, Figure 4(b) shows that the analysis result obtained using the proposed method in the independent analysis of the single-element XPS spectra was about 13% of the estimated maximum error of the component ratio. The analysis result obtained using the proposed method was more similar to the true separated peaks than the method without adjusting the basis function because the change in the reference spectrum due to the difference in device was corrected. Figure 4(d) shows that the result obtained using the proposed method in the simultaneous analysis of multiple-element XPS spectra was closest to the true separated peaks, with the estimated

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

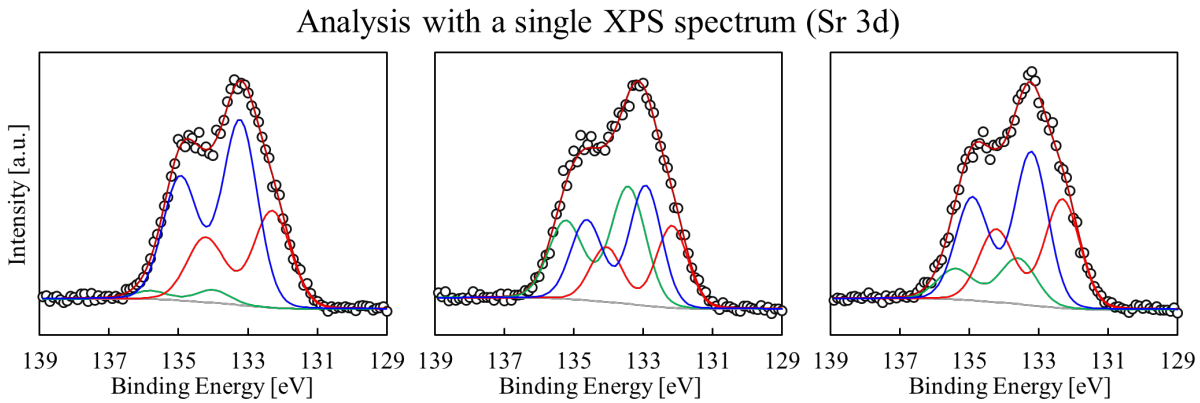

Figure 8: Three examples of results of independent analysis using the proposed method for Sr3d XPS spectra with the same noise level but different seeds of random-number generation used in noise generation.

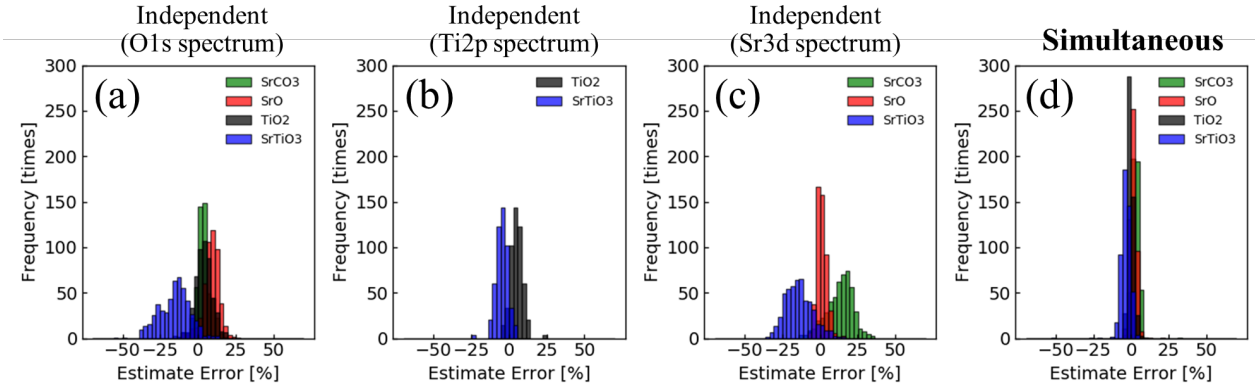

Figure 9: Variation in the estimation error of the component ratio of the compound in analyzing data with different seeds of random-number generation used in noise generation: (a) independent analysis of only the O1s spectrum, (b) independent analysis of only the Ti2p spectrum, (c) independent analysis of only the Sr3d spectrum, (d) simultaneous analysis if spectra of the three elements (O1s, Sr3d, Ti2p).

maximum error of the component ratio being about 5%. This confirms that the estimation accuracy of the component ratio of multiphase compounds was improved by the simultaneous analysis of multiple element spectra. Figure 6 shows the artificial data generated and the true separated peaks while Figure 7 shows the results obtained using the proposed method in the simultaneous analysis of multiple-element XPS spectra. Figures 6 and 7 reveal that the analysis result obtained using the proposed method in the simultaneous analysis of multiple-element XPS spectra was close to the true constituents. The time required for analysis was about 1.1 minutes (PC: MacBook Air, CPU: 1.6 GHz Intel Core i5, main memory: 8.0 GB).

Figure 8 shows three examples of the results obtained in an independent analysis of Sr3d spectra with identical noise levels but different seeds of random-number generation used in noise generation. It was seen that the trends of the three fitting results are almost the same but the analysis results vary because the reference spectra were largely superimposed. Therefore, with adoption of the proposed method, the dispersion of analytical results was compared for the case in which only XPS spectra of single-element species were analyzed independently and for the case in which multiple-element XPS

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

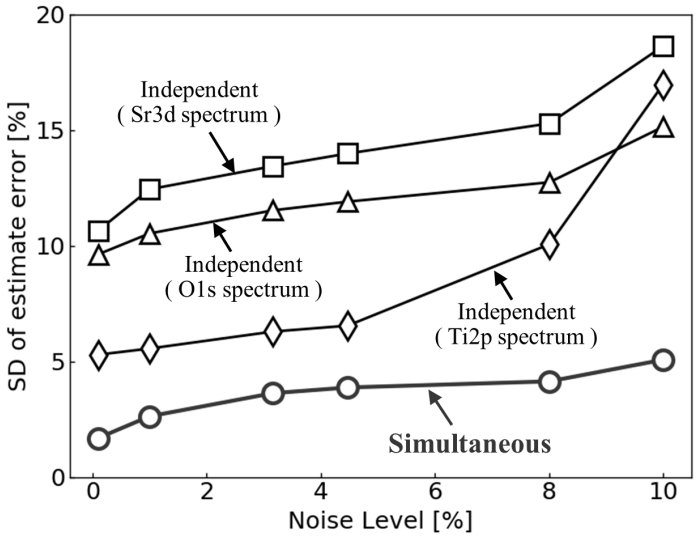

Figure 10: Variation in the estimation error of the component ratio of compound species with the noise level.

spectra were analyzed simultaneously. Figure 9 shows a histogram of the estimation error of the component ratio when 500 artificial data were generated by changing only the seed of random-number generation used in noise generation, and each of the data was analyzed using the proposed method. Figure 9(a)–(c) show that the estimation error greatly varied when the single-element XPS spectra was analyzed independently. Meanwhile, Figure 9(d) shows that the dispersion of the solution of the estimation error was small and the median value was around 0 when multiple-element XPS spectra were analyzed simultaneously. Simultaneous analysis of multiple-element XPS spectra reduced the standard deviation of the estimation error by about 70% on average compared with the case of performing independent analysis.

To verify robustness, the proposed method was applied to data with various noise levels and the dispersion of the estimation accuracy was confirmed. Five-hundred artificial data for which the seed of random-number generation used in noise generation was changed were generated at each noise level, and the data were analyzed using each proposed method. Figure 10 shows the standard deviation of the estimation error with respect to the noise level. It was confirmed that the variation in the standard deviation of the estimation error with the variation in noise level was small when multiple-element XPS spectra were analyzed simultaneously. It was therefore clarified that analysis robust against the noise level can be realized by simultaneously analyzing multiple-element XPS spectra.

3.2. Simultaneous Analysis of Multiple-element XPS Spectra with the Identification of Compound Species

In Section 3.1, we performed an analysis assuming that combinations of reference spectra [SrO, $SrCO_3$, $SrTiO_3$, TiO_2] are known. However, the combination of truly included reference spectra is often unknown at an actual analysis site. We therefore analyzed artificial data (spectra of [SrO, $SrCO_3$, $SrTiO_3$, TiO_2]) for different combinations of reference spectra of the candidate compounds, and the combinations of the reference spectra were automatically selected using the BIC. The candidate compounds were 10 major compound species [22–28] composed of three elements (O, Ti, Sr). Figure 11 shows the combinations of compound species with smaller BIC among the $2^{10} - 1 = 1023$ candidate combinations. It was seen that the true combination [SrO, $SrCO_3$, $SrTiO_3$, TiO_2] was selected using the BIC.

It is considered that the selection of the combination of reference spectra varies depending on the noise level and the

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

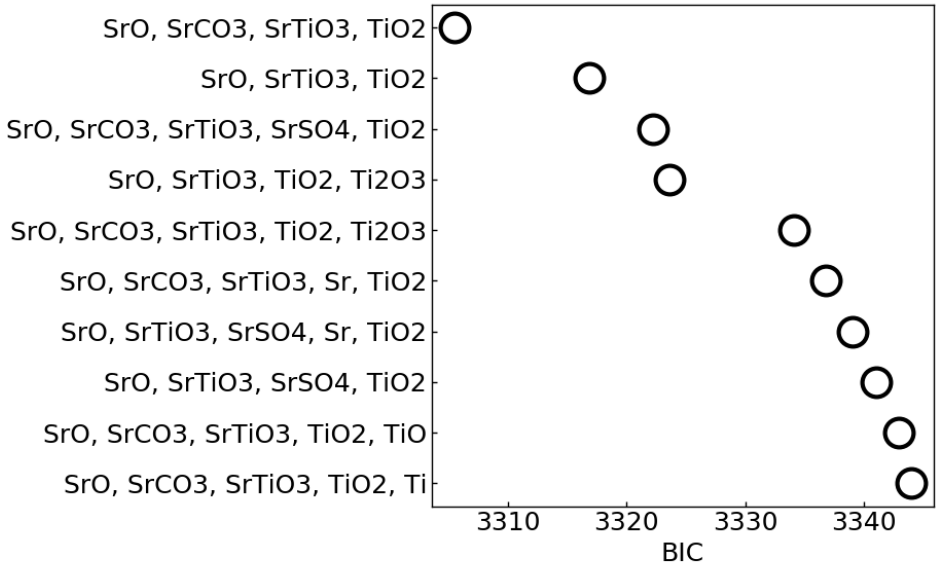

Figure 11: Results of selecting a combination of reference spectra using BIC. the true combination is [SrO, $SrCO_3$, $SrTiO_3$, TiO_2].

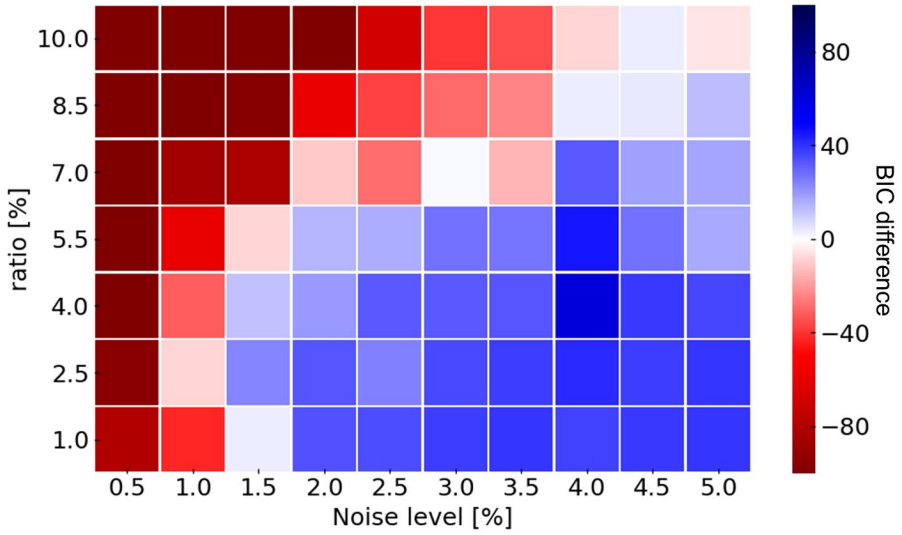

Figure 12: Heat map of the difference in the BIC when analyzing [SrO, $SrCO_3$, $SrTiO_3$, TiO_2] and [SrO, $SrTiO_3$, TiO_2] as reference spectra. The X-axis gives the noise level of spectra while the Y-axis gives the $SrCO_3$ component ratio of the spectrum.

component ratio of the compound. The detection limit of the automatic selection of the combination of reference spectra was therefore verified. The component ratio of the $SrCO_3$ (i.e., the component represented by green in Figure 6) and the noise level were varied to generate artificial data, and the reference spectrum was automatically selected using the proposed method. The [SrO, $SrCO_3$, $SrTiO_3$, TiO_2] as the true combination or [SrO, $SrTiO_3$, TiO_2] not including $SrCO_3$ were selected using the BIC, and both combinations of reference spectra are discussed here. Figure 12 shows a heat map of the difference in the BIC between the two models ([SrO, $SrCO_3$, $SrTiO_3$, TiO_2] and [SrO, $SrTiO_3$, TiO_2]) for the noise level and the component ratio of $SrCO_3$ (i.e., the component represented by green in Figure 6). The red color in Figure 12 shows the area

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

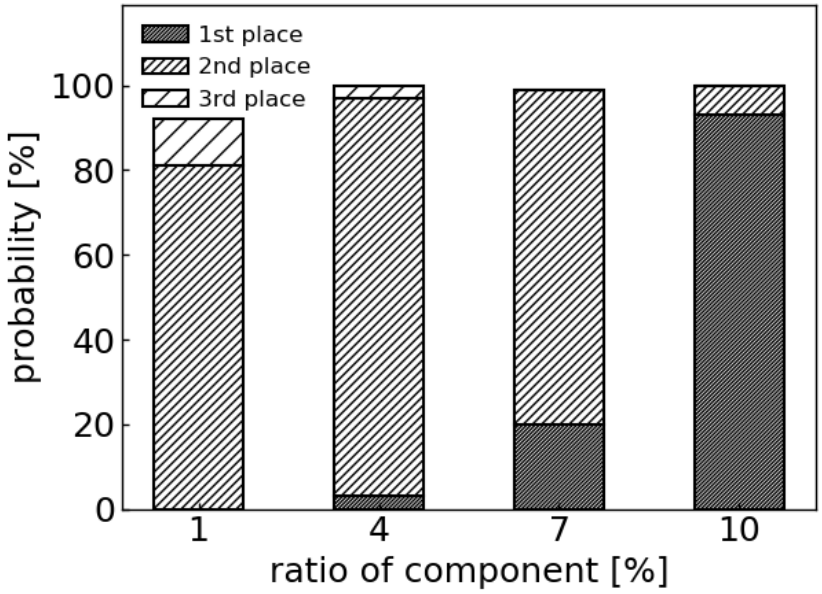

Figure 13: Probability of selecting the true reference spectrum combination from first to third place using the BIC. The X-axis gives the $SrCO_3$ component ratio of the spectrum.

in which the true reference spectrum combination [SrO, $SrCO_3$, $SrTiO_3$, TiO_2] was selected. Meanwhile, the blue color shows the area in which [SrO, $SrTiO_3$, TiO_2] was selected. It was clarified that a few components can be detected by reducing the noise level. In addition, by reducing the noise level, the red color became darker, indicating that micro-components are detected more reliably by reducing the noise level.

In the automatic selection of the reference spectrum using the BIC, the true reference spectrum may not be selected depending on the intensity and noise level of the components. It was practically desirable that the combination of true reference spectra was a promising candidate in the evaluation using the BIC, and we thus verified whether the true reference spectrum is obtained as a promising candidate using the BIC. Concretely, 100 artificial data were generated by changing the seed of random-number generation used in noise generation, and the combination of reference spectra was selected using the proposed method for each of the artificial data. This validation was carried out for several artificial data in which the component ratios of the $SrCO_3$ differed. Figure 13 shows the probability that the true reference spectrum combinations are selected from the first to third ranks for the ratio of components in $SrCO_3$. It was confirmed that the true combination of reference spectra was selected in first, second, or third place with more than 99% probability when the component ratio of $SrCO_3$ was 4% to 10%. It was also confirmed that the combination of true reference spectra was selected in either second or third place with 90% probability when the component ratio of $SrCO_3$ was 1%.

4. Conclusion

The development of the automatic analysis of a large number of spectra in XPS is required with the expectation of the industrial application of the visualization of the chemical composition distribution of a substance through three-dimensional analysis and the evaluation of process reactions by operand analysis. In spectral analysis, the measured spectrum is often compared with the reference spectrum of a known single-phase compound sample manually. The present study developed

a method that automatically analyzes the XPS spectrum of a multiphase compound using a reference spectrum comprising multiple peaks as a basis function. In particular, when analyzing only the single-element spectrum, it was confirmed that the analysis result was not uniquely determined because the reference spectra have superposition and similarity, and was clarified that accuracy is improved by analyzing multiple-element spectra simultaneously.

In addition, the developed method using the BIC allows us to automatically estimate a proper combination of compounds among many candidate combinations of reference spectra. The method is applicable to the automated analysis of spectra from unknown samples. We also clarified that the identification of minute compound components is possible by reducing the noise level of the measured spectrum.

Acknowledgment

We thank Glenn Pennycook, MSc, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

References

[1] R. Hesse, P. Streubel, R. Szargan, Product or sum: comparative tests of Voigt, and product or sum of Gaussian and Lorentzian functions in the fitting of synthetic Voigt-based X-ray photoelectron spectra, Surf. Interface Anal. 39 (2007) 381–391.

[2] H. Shinotsuka, et al., Automated information compression of XPS spectrum using information criteria, J. Electron Spectrosc. Relat. Phenom. 239 (2020) 146903.
https://doi.org/10.1016/j.elspec.2019.146903

[3] A. Herrera-Gomez, M. Bravo-Sanchez, O. Ceballos-Sanchez, M.O. Vazquez-Lepe, Practical methods for background subtraction in photoemission spectra, Surf. Interface Anal. 46 (2016) 897–905.

[4] R. Matsumoto, et al., J. Electron Spectrosc. Relat. Phenom. 207 (2016) 55–59.

[5] R. Murakami, et al., e-Journal of Surface and Nanotechnology, 17 (2019) 61–68.

[6] K. Nagata, et al., Bayesian spectral deconvolution with the exchange Monte Carlo method. Neural

[7] D.A. Shirley, High-resolution X-ray photoemission spectrum of the valence bands of gold, Phys. Rev. B 5 (1972) 4709.

[8] A. Proctor, P.M.A. Sherwood, Data analysis techniques in X-ray photoelectron spectroscopy, Anal. Chem. 54 (1982) 13–19.

Networks 28 (2012) 82–89.

[9] K. Levenberg, A method for the solution of certain non-linear problems in least squares, Q. Appl. Math. 2 (1944) 164–168.

[10] D.W. Marquardt, An algorithm for least squares estimation of nonlinear parameters, J. Soc. Ind. Appl. Math. 11 (1963) 431–441.

[11] J.H. Holland, Adaptation in Natural and Artificial Systems, The University of Michigan Press, Ann Arbor, MI, 1975.

[12] D.E. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning, Addison-Wesley, 1989.

[13] V.S. Gordon, D. Whitley, Serial and parallel genetic algorithms as function optimizers, Proceedings of the Fifth International Conference on Genetic Algorithms, 177–183 (1993).

[14] L. Davis, Handbook of Genetic Algorithms, Van Nostrand Reinhold, New York, 1990.

[15] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

[16] M.E. Pilleux, C.R. Grahmann, V.M. Fuenzalida, Hydrothermal Strontium Titanate Films on Titanium: An XPS and AES Depth‐Profiling Study, J. Am. Ceram. Soc. 77(1994) 1601-1604.
https://doi.org/10.1111/j.1151-2916.1994.tb09763.x

[17] M.I. Sosulnikov, Y.A. Teterin, A.N. Doklady, X-ray photoelectron studies of Ca, Sr and Ba and their oxides and carbonates, XPS Study of Calcium, Strontium, Barium and Their Oxides, DAN of USSR317, 2 (1991) 418-421.

[18] N. Aas, T.J. Pringle, M. Bowker, J. Chem. Adsorption and decomposition of methanol on TiO2, SrTiO3 and SrO, Soc. Faraday Trans. 90 (1994) 1015-1022.
DOI : https://doi.org/10.1039/FT9949001015

[19] I.T. Chashechnikova, V.M. Vorotyntsev, V.V. Borovik, G.I. Golodets, I.V. Plyuto, A.P. Shpak, Strong metal-carrier interaction in cobalt- and nickeltitanium dioxide co-hydrogenation catalysts, *Theor. Eksp. Khim*. 28, 216 (1992) 174-176.
DOI : https://doi.org/10.1007/978-3-319-44439-0_2

[20] R.P. Vasquez, X-ray photoelectron spectroscopy study of Sr and Ba compounds, *J. Electron Spectrosc. Relat. Phenom.* 56, 3 (1991) 217-240.
DOI : https://doi.org/10.1016/0368-2048(91)85005-E

[21] F. Lange, H. Schmelz, H. Knozinger, An X-ray photoelectron spectroscopy study of oxides of arsenic supported on TiO2, J. Electron Spectrosc. Relat. Phenom. 57, 3 (1991) 307-315.
DOI : https://doi.org/10.1016/0368-2048(91)80017-O

[22] A.A. Galuska, J.C. Uht, N. Marquez, Reactive and nonreactive ion mixing of Ti films on carbon substrates, *J. Vac. Sci. Technol.* A 6 (1988) 110.
DOI : https://doi.org/10.1116/1.574992

[23] D. Simon, C. Perrin, J. Bardolle, ESCA study of Nb and Ti oxides. Applications to the determination of the nature of the superficial films formed during the oxidation of Nb-Ti and Nb-Ti alloys, *J. Microsc. Spectrosc. Electron.* 1 (1976) 175-186.

[24] D. Gonbeau, C. Guimon, G. Pfister-Guillouzo, A. Levasseur, G. Meunier, R. Dormoy, XPS study of thin films of titanium oxysulfides, *Surf. Sci.* 254 (1991) 81-89.
DOI : https://doi.org/10.1016/0039-6028(91)90640-E

[25] S. Badrinarayanan, S. Sinha, A.B. Mandale, XPS studies of nitrogen ion implanted zirconium and titanium, *J. Electron Spectrosc. Relat. Phemon.* 49 (1989) 303-309.
DOI : https://doi.org/10.1016/0368-2048(89)85018-2

[26] M.V. Kuznetsov, J.F. Zhuravlev, V.A. Zhilyaev, V.A. Gubanov, XPS analysis of adsorption of oxygen molecules on the surface of Ti and TiNx films in vacuum, *J. Electron Spectrosc. Relat. Phenom.* 58, 3 (1992) 169-176.
DOI : https://doi.org/10.1016/0368-2048(92)80016-2

[27] F. Werfel, O. Brummer, Corundum Structure Oxides Studied by XPS, *Phys. Scripta.* 28 (1983) 92.

[28] A.B. Christie, J. Lee, I. Sutherland, J.M. Walls, An XPS study of ion-induced compositional changes with group II and group IV compounds, Appl. Surf. Sci. 15 (1983) 224-237.
DOI : https://doi.org/10.1016/0378-5963(83)90018-1

Preprint:
The final version was published as: Journal of Electron Spectroscopy and Related Phenomena. 245 (2020) 147003.
https://doi.org/10.1016/j.elspec.2020.147003

Highlights

- *Automatic analysis method developed using a sparsely modeled reference spectrum.*
- *Method applied to the XPS spectrum of a multiphase compound.*
- *Estimation accuracy clarified through simultaneous analysis of multi-element spectra.*
- *Method established for effective use of the sparse modeling of reference spectrum.*
- *Spectrum analysis method developed with the automatic selection of compound species.*