Efficient Removal of Noise-derived Components for Automatic XPS Spectral Decomposition Using Hierarchical Clustering

Ryo Murakami, Kazuki Nakamura, Hiromi Tanaka, Hiroshi Shinotsuka, Hideki Yoshikawa

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

Corresponding author: natsu8199916@gmail.com

Received: 19 March, 2020; Accepted: 19 May, 2020; Published: 28 May, 2020

In this paper, we aim to automatically provide a solution to peak separation in an X-ray photoelectron spectroscopy (XPS) spectrum with non-negligible statistical noise that is inevitably accepted in multi-dimensional (e.g., 2-dimensional/3-dimensional XPS profiles) XPS measurement. To achieve this, in our previous study [H. Shinotsuka et al., J. Electron Spectros. Relat. Phenomena 239, 146903 (2020)], we automatically selected optimal solutions using the Bayesian information criterion (BIC) for measured XPS spectra. This was successfully performed for many varieties of XPS spectra. However, the optimal solution rarely included a small and sharp peak that was likely to be caused by statistical noise. In this study, we investigate a practical method to eliminate the infrequent solution with a noise-derived peak. This method uses hierarchical clustering with peak parameters (i.e., width and area) as a preprocessing step before selecting the solutions using the BIC.

Keywords: X-ray photoelectron spectroscopy; Automatic analysis; Noise removal; Bayesian information criterion; Clustering

I. INTRODUCTION

X-ray photoelectron spectroscopy (XPS) is an analytical method that is capable of identifying the chemical bonding states of surfaces and interfaces [1]. It is widely used in materials development and quality control in industry. XPS spectral decomposition is equivalent to solving nonlinear least-squares problems. Its solution is not uniquely determined because it depends on statistical noise, the freedom of peak parameters, and their initial values at the time of peak separation. Therefore, the problem is that the reproducibility of the solution is not guaranteed and analysts obtain different results.

To overcome this problem, Shinotsuka et al. introduced the process of searching for optimal solutions from a large number of candidates derived from different initial peak fitting parameters [2]. The initial peak parameters depend on smoothing the XPS spectrum that has statistical noise. The introduced process changed the number of repetitions of the smoothing processes applied to the raw spectrum to obtain patterns of initial peak parameters (the number of peaks of the initial value decreased as the number of smoothing repetitions increased). After that, the introduced process obtained a large number of solutions using the gradient method for each pattern of initial peak parameters, and selected optimal solutions using the information criterion (Bayesian information criterion: BIC [3–5] or Akaike information criterion: AIC [6]) from a large number of solutions depending on the initial peak parameters. The introduced process succeeded in greatly reducing the analysts’ dependence on the solution, even in the analysis of an XPS spectrum with statistical noise and a complicated spectral structure. Additionally, Shinotsuka et al. recommended the BIC because it obtained results more similar to the result analyzed manually by XPS.
experts than the AIC. Additionally, they used the active
Shirley method [7−11] to automatically separate photoelec-
tron peaks and background to complete the analysis without
depending on analysts.

However, even if the BIC was used, a sharp and small
peak similar to a noise-derived peak, such as a peak near
125 eV in Figure 1(b), was rarely selected as part of the op-
timal solution. The sharp and small peak was removed when
we took a long time to make the smoothing processes very
fine and select the optimal solution using the BIC.

We enhanced the practicality of the methods by setting an
upper limit on the number of repetitions of the smoothing
processes so that the analysis could be completed within
about 1 min using a typical PC (CPU: Intel(R) Core(TM)
i5-7200U CPU @ 2.50 GHz and main memory: 8.0 GB).
For example, the number of repetitions of the smoothing
processes was 155 and the analysis time was 0.8 min. Figure
1(a) shows a scatter plot of the solutions obtained by making
the smoothing processes finer than previously reported. In
this study, the fine smoothing and previous coarse smooth-
ing mean that the total numbers of searched solutions were
2880 and 155, respectively. In Figure 1(a), each axis repre-
sents the minimum full-width at half-maximum (FWHM)
for all peaks, minimum area-intensity for all peaks, and the
BIC of the fitting result. Taking note of the area surrounded
by the red dashed box, it can be seen that there are several
solutions that are as small as the BIC and have a different
FWHM [see ▽ and ○ in Figure 1(a)]. Panels (b) and (c)
in Figure 1 show solutions to peak separation corresponding
to ○ and ▽ in Figure 1(a), respectively. The visual pat-
terns of the peak separation are different, but the BIC has a
tiny difference of 0.02%. By increasing the number of
searches to 2880 and densifying the searching steps (by al-
lowing 12 min of analysis time), as shown in Figure 1(c), we
obtained an optimal solution without sharp peaks that likely
to originate from noise. The small difference in the BICs
between the two solutions shown in Figure 1(a) suggests
that there are several mathematically promising solutions.
The details are described in Appendix A.

In this study, we investigate a new practical method that
provides the optimal solution without a noise-derived peak
in a short-time (< 1 min). Specifically, while maintaining the
number of searches (i.e., the number of repetitions of the
smoothing processes) at the same limited value (e.g., 155 as
in the previous report), we investigate the method using
clustering [12−16] to eliminate the solutions that include
noise-derived peaks.

II. METHOD

The proposed method aims to effectively select optimal
solutions that do not include a sharp and small peak that
originates from statistical noise within a practical calculation
time for automatic peak separation. In this study, an im-
portant point is to objectively filter solutions and, thus, to
reduce analyst arbitrariness. This method filters the can-
didate solutions using the clustering result before selecting the
optimal solutions using the BIC.

A noise-derived peak has two features: a peak with a
smaller FWHM than the FWHM of the main peaks and a
peak with a smaller area intensity than the area intensity of
the main peaks. Therefore, we focused on the FWHM and
the area intensity of the separated peaks and performed
clustering of the solutions using the minimum FWHM
($W_{\text{min}}$) and minimum area intensity ($A_{\text{min}}$) for all peaks for
each solution. These solutions were obtained by varying the

Figure 1: Solution pattern of Al 2s XPS spectral decomposition:
(a) Scatter plot of solutions of spectral decomposition (total num-
ber of searched solutions was 2880) as spectral parameters of the
minimum FWHM, minimum area intensity, and BIC. (b) Solution
pattern, which corresponds to ○, with the smallest BIC in the
groups with a small FWHM. (c) Solution pattern, which corre-
sponds to ▽, with the smallest BIC in the groups with a large
FWHM.
number of repetitions of the smoothing processes.

The XPS peak cannot have an FWHM less than the energy resolution $\Delta E$ (convolution of the X-ray energy width and the energy resolution of the spectrometer) of the XPS equipment. The peak that has an FWHM below the threshold $\sigma_{th} = \Delta E$ is clearly assessed to be noise-derived. When clustering is performed, the cluster that has solutions with an FWHM below the threshold can be assessed as a noise-derived cluster (more precisely, it is difficult to reasonably separate the cluster into the noise-derived part and non-noise-derived part). The proposed method eliminates all parts of data in the noise-derived cluster from the solution candidates. The solutions that belong to other clusters are assessed as non-noise-derived data, and the method selects the optimal solution from non-noise-derived data using the BIC. The formula for the BIC is

$$BIC = -2 \log L + m \log N,$$

where $L$ is the maximum likelihood, $m$ is the number of parameters, and $N$ is the number of data points. The BIC is an index to select the optimal solutions that have both a low error [the first term of Eq. (1)] and a simple model with a small number of peaks [the second term of Eq. (1)].

Figure 2 shows the flow chart of the proposed method to obtain the optimal solutions without noise-derived peaks.

Considering the energy resolution of XPS equipment, the threshold $\sigma_{th}$ was set to 0.4 eV. The hierarchical clustering (the Ward method [17, 18]) was used because it was expected that the number of peaks would vary between each cluster depending on the noise level. As a preprocessing step for clustering, we normalized $W_{\text{min}}$ and $A_{\text{min}}$ to have a mean of 0 and variance of 1, respectively.

### III. RESULTS AND DISCUSSION

We applied the proposed method to an Al 2s XPS spectrum shown in Figure 1. Figure 3 shows the results of applying hierarchical clustering with $W_{\text{min}}$ and $A_{\text{min}}$ as parameters to a large number of solutions obtained by suppressing the number of practical searches (i.e., the number of smoothing operations) to 155 solutions. Figure A2 in Appendix B shows a dendrogram of this clustering. The threshold in the dendrogram used to determine the number of clusters was 20% of the maximum distance. In this paper, “distance” means the Euclidean distance between the data in parameter space. The threshold was determined empirically when we applied the proposed method to 47 XPS spectra with different noise levels obtained from high-$T_c$ superconductors of Bi$_2$Sr$_2$Ca$_{n-1}$Cu$_n$O$_{(n=1−3)}$ [19–22].

In the proposed method, solutions that belonged to clusters (red clusters in Figure 3) that included peaks with an FWHM below $\sigma_{th}$ were eliminated from the solution candidates as solutions including noise-derived peaks; that is, the proposed method selected optimal solution using the BIC from clusters other than the red cluster in Figure 3.

Figure 4(a) shows the solution with the smallest BIC that belonged to the excluded clusters (the red clusters in Figure 3), and a sharp and small peak can be observed around 125 eV, as mentioned in Figure 1(b). Figure 4(b) shows the solution.
tion with the smallest BIC within clusters other than the red cluster in Figure 3 and does not include any sharp and small peaks. Additionally, a peak structure that could be clearly distinguished from noise was not confirmed in the residual spectra in Figure 4. Therefore, a sharp and small peak at around 125 eV in Figure 4(a) was considered to be a component derived from noise.

As shown in Figure 4, the same five peaks in the solution to peak separation mean that the number of peak parameters m in Eq. (1) was the same in both figures. The number of data points N was a common value. Therefore, the second term in Eq. (1) was the same for Figure 4, and the selection of a solution changed with a slight difference in the first term in Eq. (1). Depending on the signal-to-noise level of the measured spectrum, a single sharp and small peak, which likely originated from noise, was rarely included in the optimal solution, as shown in Figure 4(a). Note that any multiple peaks that likely originated from noise were not included in the optimal solutions. This results from the BIC suppressing the total number of peaks [the second term in Eq. (1)]. This means that the BIC is effective for XPS spectral decomposition but rarely accepts a noise-derived peak because the BIC does not evaluate the peak shape (e.g., the area and the width). By contrast, the proposed method selected the solution shown in Figure 4(b) after eliminating the solution candidates with a noise-derived peak by clustering with Wmin and Amin before selecting the solution using the BIC.

As the simplest method for eliminating noise-derived peaks, it is also conceivable to eliminate only solutions including peaks with FWHMs smaller than the energy resolution of the XPS equipment. However, this method is not sufficient for eliminating noise-derived peaks if the FWHM of the noise-derived peak is larger than the energy resolution. The proposed method simply eliminates the solution with confusing noise-derived peaks that are not easily distinguished from non-noise-derived peaks.

Table 1 shows the number of searched solutions, computation time, and FWHMs of the peaks included in the optimal solutions (the smallest BIC solutions) for the previous method using the BIC and for the proposed method using clustering before the BIC. When we finely changed the increment of the number of repetitions of the smoothing processes and searched for 2880 solutions, a solution including sharp and small peaks derived from noise was not selected. However, the analysis took about 12 min. By contrast, the proposed method allowed us to select the solution without the noise-derived peak shown in Figure 4(b) as the optimal solution, and the time required for analysis was about 1 min.

To demonstrate the versatility of the proposed method, we applied it to the Cu 2p3/2 XPS spectrum for high-Tc superconductors of Bi2Sr2Ca_{n-1}Cu_{n}O (n = 1−3) with worse signal-to-noise level and more complex structures than the Al 2s XPS spectrum shown Figure 1. Figure 5(a) shows the
results of hierarchical clustering using $W_{\text{min}}$ and $A_{\text{min}}$ as parameters. The dendrogram is shown in Figure A3 in Appendix B. Figure 5(b, c) shows the solutions in which the BICs have minimum values in (b) a red cluster and (c) a blue cluster, respectively. The solutions including peaks with an FWHM less than $\sigma_{\text{th}}$ [the red cluster in Figure 5(a)] were eliminated by the proposed method. As shown in Figure 5(b), in the solution belonging to the excluded cluster, there was clearly a noise-derived peak similar to the blue peak near 946 eV.

In addition to Cu 2p$_{3/2}$, the proposed method was applied to other XPS spectra (O 1s, C 1s, Bi 4f, Sr 3d, and Ca 2p) of 50 high-$T_c$ superconductors of Bi$_2$Sr$_2$Ca$_{n-1}$Cu$_n$O ($n = 1$−$3$) with various noise levels from about 2% to 10%. The results confirmed that the same trend was obtained.

**IV. CONCLUSIONS**

In this study, we aimed to automatically provide a solution to peak separation in the XPS spectrum. In the previous study, the optimal solution was automatically selected using the BIC from a large number of solutions with the initial values of different peak parameters (i.e., the height, the

![Figure 5](image_url)

*Figure 5:* Solution pattern of Cu 2p$_{3/2}$ XPS spectral decomposition: (a) Result of clustering many solution patterns of spectral decomposition using the minimum FWHM and the minimum area intensity as parameters. (b) Spectral decomposition solution with the smallest BIC in the red cluster and its residual spectrum. (c) Spectral decomposition solution with the smallest BIC in the blue cluster and its residual spectrum.

![Figure A1](image_url)

*Figure A1:* Top three optimal solutions in Al 2s XPS spectral decomposition. The first (a), the second (b), and the third (c) ranks were evaluated using their BICs.
width, and the position), and greatly reduced the analysts’
dependence on the solution. Unfortunately, it was rarely
confirmed that the optimal solution included a sharp and
small peak derived from statistical noise.

In this study, we clarified that noise-derived components
can be effectively removed by introducing a clustering pro-
cess with peak parameters (i.e., the width and the area). This
method filtered solutions using peak parameter clustering
and subsequent BIC selection, and thus enabled us to pro-
vide a solution without noise-derived components at high
speed (< 1 min).

Appendix A

The results show that the best and most time-consuming solution
in Figure 1(c) for peak separation obtained by making the search
step finer than previously reported differed from the practically
best solution in Figure 4(b) for peak separation obtained using the
proposed method. However, both solutions had almost the same
BIC, the difference of which was 0.3%, and therefore, the differ-
ence is considered to be within the fluctuation of solutions caused
by statistical noise. Figure A1 shows the redrawn Al 2s XPS de-
composition solutions mentioned above, where the ranking of their
BICs is as follows: The best rank is Figure 1(c); the second-best
rank is Figure 1(b), that is, Figure 4(c); and the third-best rank is
Figure 4(b). The difference between these BICs was within 0.3%,
which was caused by statistical noise and, therefore, it is difficult
to assess which solution is really the best. This allows us to accept
multiple best solutions depending on statistical noise from the
viewpoint of the BIC. Note that the proposed method can automatical-
ically eliminate solutions with noise-derived peaks using clustering
with $W_{\text{min}}$ and $A_{\text{min}}$ as parameters, even if plural solutions should be
accepted.

Appendix B

Figures A2 and A3 show the dendrogram in the clustering results
shown in Figures 3 and 5(a), respectively. The dendrogram shows
the process of the hierarchical clustering of the data which are dis-
tributed in the parameter space of the minimum FWHM and the
minimum area intensity. The hierarchical clustering is a method of
fusing nearby data to progressively create larger clusters. The Ward
method [16−18] was used in this study for the hierarchical cluster-
ing to fuse data so that the residual sum of squares in all the clus-
ters is smaller and clusters with relative dense data is obtained. The
distance between each data and a gravity center of cluster was cal-
culated as the Euclidean distance. The Ward method is better for
the automatic XPS analysis because the method is little affected by
outliers and its result is robust.

A dendrogram is a binary tree in which terminal nodes are data,
and non-terminal nodes are the fusing clusters. The vertical axis of
the dendrogram represents the Euclidean distance between the data
in the parameter space. The threshold (a horizontal black line) was
set to 20% of the maximum distance to fuse the final two clusters.
In conclusion, the number of clusters was four in Figure A2 and
three in Figure A3, and the result leads the clear separation of the
clusters in Figures 3 and 5(a).

References

[1] S. Hofmann, Auger- and X-Ray Photoelectron Spectroscopy in
Materials Science (Springer, Berlin, Heidelberg, 2013) Chap. 7.
[2] H. Shinotsuka, H. Yoshikawa, R. Murakami, K. Nakamura, H.
Tanaka, and K. Yoshihara, J. Electron Spectros. Relat. Phenomena
239, 146903 (2020).
[3] G. Schwarz, Ann. Stat. 6, 461 (1978).
[4] C. M. Bishop, Pattern Recognition and Machine Learning
(Springer-Verlag, New York, 2006) Chap. 5.
[5] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Sta-
tistical Learning: Data Mining, Inference, and Prediction
(Springer, New York, 2009) Chap. 7.
[6] H. Akaike, IEEE Trans. Automat. Contr. 19, 716 (1974).
[7] A. Herrera-Gomez, M. Bravo-Sanchez, O. Ceballos-Sanchez,
and M. O. Vazquez-Lepe, Surf. Interface Anal. 46, 897 (2014).
[8] R. Matsumoto, Y. Nishizawa, N. Kataoka, H. Tanaka, H. Yo-
shikawa, S. Tanuma, and K. Yoshihara, J. Electron Spectros. Relat.
Phenomena 207, 55 (2016).
[9] R. Matsumoto, Y. Nishizawa, N. Kataoka, H. Tanaka, H. Yoshikawa, S. Tanuma, and K. Yoshihara, J. Surf. Anal. 22, 155 (2016) (in Japanese).
[10] Y. Nishizawa, R. Matsumoto, N. Kataoka, H. Tanaka, H. Yoshikawa, S. Tanuma, and K. Yoshihara, J. Surf. Anal. 24, 36 (2017) (in Japanese).
[11] R. Murakami, H. Kageyama, K. Nakamura, H. Tanaka, H. Shinotsuka, H. Yoshikawa, and K. Yoshihara, e-J. Surf. Sci. Nanotechnol. 17, 61 (2019).
[12] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, Cluster Analysis, 5th ed. (John Wiley & Sons, 2011) Chap. 4.
[13] A. K. Jain and R. C. Dubes, Algorithms for Clustering Data (Prentice Hall, New Jersey, 1988) Chap. 3.
[14] R. Xu and D. Wunsch, IEEE Trans. Neural Netw. 16, 645 (2005).
[15] S. Guha, R. Rastogi, and K. Shim, ACM SIGMOD Rec. 27, 73 (1998).
[16] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer, New York, 2009) Chap. 4.
[17] J. H. Ward Jr., J. Am. Stat. Assoc. 58, 236 (1963).
[18] D. Wishart, Biometrics 25, 165 (1969).
[19] Y. Yamauchi, S. Kishida, M. Kimura, H. Yoshikawa, S. Fukushima, and H. Tokutaka, Physica C: Supercond. 357-360, 410 (2001).
[20] H. Tanaka, S. Kishida, H. Yoshikawa, A. Kimura, A. Tanaka, and S. Fukushima, Physica C: Supercond. 392-396, 153 (2003).
[21] Z.-X. Shen, P. A. P. Lindberg, I. Lindau, W. E. Spicer, C. B. Eom, and T. H. Geballe, Phys. Rev. B 38, 7152(R) (1988).
[22] F. U. Hillebrecht, J. Frixedas, L. Ley, H. J. Trodahl, J. Zaanen, W. Braun, M. Mast, H. Petersen, M. Schaible, L. C. Bourne, P. Pinsukanjana, and A. Zettl, Phys. Rev. B 39, 236 (1989).

All articles published on e-J. Surf. Sci. Nanotechnol. are licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0). You are free to copy and redistribute articles in any medium or format and also free to remix, transform, and build upon articles for any purpose (including a commercial use) as long as you give appropriate credit to the original source and provide a link to the Creative Commons (CC) license. If you modify the material, you must indicate changes in a proper way.

Published by The Japan Society of Vacuum and Surface Science