



Sample structure prediction from measured XPS data using Bayesian estimation and SESSA simulator

Hiroshi Shinotsuka^{a,*}, Kenji Nagata^a, Malinda Siriwardana^a, Hideki Yoshikawa^a, Hayaru Shouno^b, Masato Okada^c

^a Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba, Japan

^b Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu, Japan

^c Graduate School of Frontier Science, The University of Tokyo, Kashiwa, Japan

ARTICLE INFO

Keywords:

X-ray photoelectron spectroscopy
Bayesian estimation
Exchange Monte Carlo method
SESSA

ABSTRACT

We have developed a framework for solving the inverse problem of X-ray photoelectron spectroscopy (XPS) by incorporating an XPS simulator, Simulation of Electron Spectra for Surface Analysis (SESSA), into Bayesian estimation to obtain an overall picture of the distribution of plausible sample structures from the measured XPS data. The Bayesian estimation framework automated the very tedious task of adjusting the sample structure parameters manually in the simulator. As an example, we performed virtual experiments of angle-resolved XPS on a four-layered sample, and we estimated the sample structures based on the XPS intensity data obtained from experiments. We succeeded in not only obtaining an optimal solution, but also visualizing the distribution of the solution through the Bayesian posterior probability distribution.

1. Introduction

X-ray photoelectron spectroscopy (XPS) is a non-destructive surface analysis technique that identifies the elements contained in a solid sample and their chemical states by measuring the energy spectrum of photoelectrons emitted by irradiating an X-ray on the sample [1]. The XPS spectrum is reproducibly obtained if the sample structure (e.g., the number of layers, ratio of the elements in each layer, and thickness of each layer) and geometric arrangement of the XPS system (e.g., the X-ray source, photoelectron analyzer, and sample arrangement) are known. Simulation of Electron Spectra for Surface Analysis (SESSA) is a well-known simulator to facilitate quantitative interpretation of the spectra of XPS and Auger electron spectroscopy [2]. In principle, it is possible to solve an inverse problem using the simulator to estimate a sample structure from measurement data, but there is no automatic way to solve the inverse problem. Because users have to run the simulator and manually compare the results with measured data, there is a need to automate the process.

The aim of this study is to incorporate the SESSA simulator into a Bayesian estimation framework [3] to automatically obtain an overall picture of the plausible parameter distribution of the sample structure from the XPS measurement data.

In 1992, Cumpson et al. pioneered the application of Bayesian estimation in surface composition analysis [4]. They considered samples consisting of uniform single-phase compounds and analytically evaluated the impact on the compositional error, as determined by the sensitivity coefficient method, when given observed XPS peak intensities and associated error values (assuming Gaussian noise).

In Bayesian estimation, the data and parameters are described in terms of probabilities. In XPS, the number of photoelectrons is measured by pulse counting, and noise is observed based on Poisson processes. Bayesian estimation naturally takes these noise effects into account and finds the distribution of the parameters [5,6]. The presence of noise in the measurements means that the estimated solution also has some variance, that is, the estimated solution is more accurate for smaller noise, and vice versa.

To estimate the composition and thickness of a multilayer structure, it is necessary to solve an inverse problem with high-dimensional parameters. The parameter search by the commonly used least-squares method can easily become trapped in a local solution, making it impossible to determine whether the obtained solution is a global solution or not. The replica-exchange Monte Carlo method [7,8] is a type of Markov-chain Monte Carlo method [9–11] and a powerful tool for solving such complex high-dimensional parameter searches. Thus, we

* Corresponding author.

E-mail address: Shinotsuka.Hiroshi@nims.go.jp (H. Shinotsuka).

<https://doi.org/10.1016/j.elspec.2023.147370>

Received 2 March 2023; Received in revised form 30 May 2023; Accepted 5 July 2023

Available online 6 July 2023

0368-2048/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

developed a new algorithm using the SESSA simulator in the replica-exchange Monte Carlo method that automatically performs parameter searches for solving the inverse problem of XPS under the framework of Bayesian estimation. This method automated the very tedious task of adjusting the sample structure parameters manually in the simulator. The Bayesian estimation allows not only the optimal solution for various parameters to be obtained, but also the estimation accuracy of those parameters to be calculated [12]. Although SESSA is computationally expensive for a global-solution search using a large number of SESSA results in the replica-exchange Monte Carlo method, we overcame this problem by combining the accurate solution obtained by SESSA with an approximate solution obtained by the straight-line path approximation (SLA).

As an example, we applied this method to angle-resolved XPS (ARXPS), a technique that combines measurements at several emission angles and is used to estimate the compositions and thicknesses of surface film structures. ARXPS is a suitable method for non-destructive tomographic analysis of samples with internal structures, which is important for materials engineering. As a concrete example of ARXPS, we performed a virtual experiment of a four-layered model sample of CON/HfON/SiON/Si to obtain the photoelectron intensities from the Hf4f, Si2p, C1s, N1s, and O1s orbitals at two emission angles (0° and 54.7°). The two-emission-angle measurement is the simplest ARXPS that roughly estimates whether each layer is shallow or deep, and the result is practically useful for determining whether more precise multiangle measurements are necessary. Therefore, it is useful to know how much information about the composition and thickness of each layer can be obtained from inverse-problem analysis of the simplest ARXPS measurement.

After performing inverse-problem analysis of two-emission-angle ARXPS for the composition and thickness of each layer as unknown parameters, we succeeded in estimating the sample structure, including the true model. The estimated parameter was obtained as a distribution. For compositions and film thicknesses near the sample surface, where the number of counts is large, we obtained narrow distributions with high estimation accuracy. Conversely, for compositions and film thicknesses in the internal structure of the sample, where the number of counts is smaller, the distribution of estimated parameters became wider, although some appropriate values were included. The proposed method is effective for inverse estimation of the sample structure from ARXPS measurements, which enables to quantify how accurate the sample structure can be obtained depending on the statistical noise of a XPS spectrum.

2. Inverse-problem analysis system

A schematic diagram of the inverse-problem analysis system proposed in this study is shown in Fig. 1. The system estimates the elemental compositions and structures of the layers in a sample from experimental ARXPS data. From the ARXPS measurements, the peak area intensities of

Table 1

Example of an intensity data set obtained by ARXPS.

	Intensity at $\alpha = 0^\circ$	Intensity at $\alpha = 54.7^\circ$
Hf4f	100,586	94,540
Si2p(substrate)	6586	1689
Si2p(compound)	2131	864
C1s	14,211	25,570
N1s	2932	887
O1s	74,280	72,269

the dominant core levels (e.g., O1s and C1s) of the elements in the sample are obtained depending on several emission angles. To estimate the internal structure of the sample from these data, a spectral generation method such as SESSA is used to compare the calculated and measured data. Bayesian estimation, a statistical framework, is used for the comparison. Using Bayesian estimation, not only can the optimal values of various parameters be calculated, but their estimation accuracy can also be calculated via the Bayesian posterior distribution.

The framework of the proposed method is as follows. In XPS, the core-level spectra of the elements of interest are measured, and the peak area intensities after removal of the background spectra are calculated. Furthermore, in ARXPS, the peak area intensities are measured for several photoelectron emission angles to obtain a table, an example of which is Table 1. The experimental XPS intensity at photoelectron emission angle α and core level of interest i is defined as $I_{\alpha,i}^{\text{Exp}}$.

In Table 1, we assumed that the dominant core levels are Hf4f, Si2p, C1s, N1s, and O1s, and that Si2p for the Si substrate and Si compounds (silicon oxide and silicon oxynitride) can be distinguished by the difference in the chemical shift. The core levels of interest are denoted $i = \{\text{Hf4f, Si2p (substrate), Si2p (compound), C1s, N1s, O1s}\}$. The emission angles are denoted $\alpha = \{0^\circ, 54.7^\circ\}$.

We attempted to estimate a sample structure that reproduces the measured data in Table 1. We made the following three assumptions: the sample has a multilayer structure with a steep interface; the number of layers is predetermined, including the substrate; and the candidate elements constituting each layer can be predicted. The number of layers is denoted L , and the thickness of the l -th layer from the surface is denoted T_l . K_l is the number of candidate elements in the layer, k denotes an individual element, and x_{lk} is the content of element k in layer l . The parameter set to be estimated (θ) is as follows:

$$\theta = \left\{ T_l, \{x_{lk}\}_{k=1, \dots, K_l} \right\}_{l=1, \dots, L}. \quad (1)$$

The following constraint is imposed on the element contents:

$$\sum_{k=1}^{K_l} x_{lk} = 1, (x_{lk} \geq 0) (\text{for } l = 1, \dots, L). \quad (2)$$

Hereafter, we assume that the sample has a four-layer structure, including the substrate. The top layer contains C, O, and N, assuming contamination compounds owing to atmospheric exposure; the second

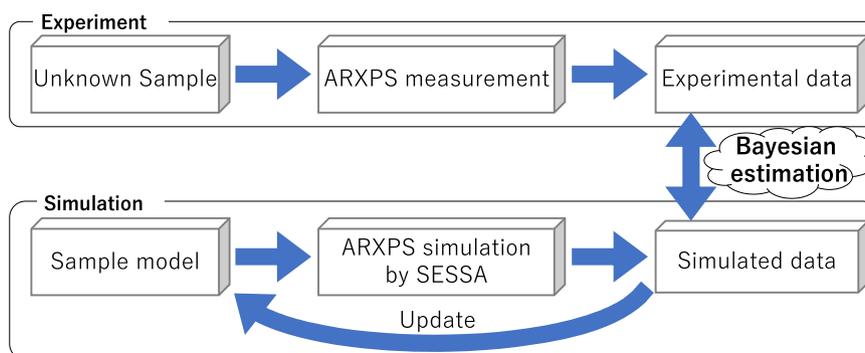


Fig. 1. Schematic of the proposed inverse-problem system.

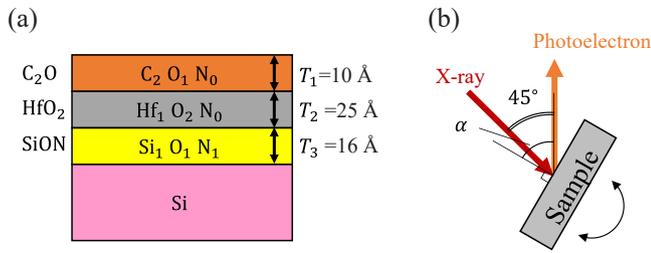


Fig. 2. (a) Predetermined sample configuration. (b) Geometrical configuration. The angle between the incident X-ray and the photoelectron detector is fixed at 45°. The photoelectron emission angle α is changed by tilting the sample.

layer contains Hf, O, and N, assuming Hf cationic compounds; the third layer contains Si, O, and N, assuming Si cationic compounds; and the fourth layer is the pure Si substrate. The sample model is a generalized hafnium oxide thin film on a Si substrate, and it is expressed as $C_{x_{11}}O_{x_{12}}N_{x_{13}}(T_1 \text{ \AA})/Hf_{x_{21}}O_{x_{22}}N_{x_{23}}(T_2 \text{ \AA})/Si_{x_{31}}O_{x_{32}}N_{x_{33}}(T_3 \text{ \AA})/Si$. The parameter set to be estimated (θ) is then

$$\theta = \{T_1, T_2, T_3, x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}, x_{31}, x_{32}, x_{33}\}. \quad (3)$$

The thickness of the substrate layer is $T_4 = \infty$. The aim is to find a parameter set of the sample structure that reproduces the ARXPS data given in Table 1. The two-emission-angle ARXPS spectra of the hafnium oxide thin film were presented in the ISO 13424 document “XPS - reporting of results of thin film analysis” as a typical example of thin film analysis [13].

For a sample structure, the photoelectron intensity can be calculated and is denoted $I_{\alpha,i}^{\text{Calc}}(\theta)$. Once the measured and calculated intensity data $I_{\alpha,i}^{\text{Exp}}$ and $I_{\alpha,i}^{\text{Calc}}(\theta)$ are obtained, their similarity can be calculated using a probabilistic model. Assuming that the measured intensities are counting data, their statistical noise follows a Poisson distribution. When the expected value of the Poisson distribution is $\lambda = I_{\alpha,i}^{\text{Calc}}(\theta)$, the probability of observing $y = I_{\alpha,i}^{\text{Exp}}$ as the experimental value is

$$p_\lambda(y) = \frac{\lambda^y e^{-\lambda}}{y!}. \quad (4)$$

The probabilities for all α, i are multiplied, and the cost function $E(\theta)$ is defined as the sign-reversal of its logarithm:

$$E(\theta) = \sum_i \sum_\alpha \left(I_{\alpha,i}^{\text{Calc}}(\theta) - I_{\alpha,i}^{\text{Exp}} \log I_{\alpha,i}^{\text{Calc}}(\theta) + \log \Gamma(I_{\alpha,i}^{\text{Exp}} + 1) \right), \quad (5)$$

where $\Gamma(x)$ is the gamma function. The reproducibility of the measured data is higher for smaller $E(\theta)$. The parameter set that minimizes $E(\theta)$ is then searched for.

Bayesian estimation allows not only the optimal value of each parameter to be estimated, but also the Bayesian posterior distribution to be visualized and its estimation accuracy to be calculated. The framework of Bayesian estimation is described in detail in Appendix A.1. The posterior distribution of the parameter θ for a given data set D ($p(\theta|D)$) can be expressed as follows using the cost function $E(\theta)$:

$$p(\theta|D) \propto \exp(-E(\theta))p(\theta), \quad (6)$$

Table 2
Three ARXPS data sets for analysis.

	Data set 1 100 min		Data set 2 10 min		Data set 3 1 min	
	$\alpha = 0^\circ$	$\alpha = 54.7^\circ$	$\alpha = 0^\circ$	$\alpha = 54.7^\circ$	$\alpha = 0^\circ$	$\alpha = 54.7^\circ$
Hf4f	100,586	94,540	10,185	9476	1059	945
Si2p(substrate)	6586	1689	645	170	59	15
Si2p(compound)	2131	864	216	86	22	9
C1s	14,211	25,570	1441	2658	144	252
N1s	2932	887	282	96	23	7
O1s	74,280	72,269	7480	7144	781	695

where $p(\theta)$ is the prior distribution of the parameter of interest, which must be determined in advance. Because it is difficult to visualize this posterior distribution analytically, a replica-exchange Monte Carlo method, a type of Markov-chain Monte Carlo method, is used. Exchange Monte Carlo methods can find global solutions more efficiently than ordinary optimization methods. The framework is described in Appendix A.2.

3. Simulation

To verify the effectiveness of the proposed method, we performed a virtual experiment and considered the results as real data for estimating the sample structure.

3.1. Sample structure model and data generation

We performed a virtual experiment using SESSA with a predetermined sample structure. The sample was assumed to have a four-layered structure consisting of C_2O (10 Å)/ HfO_2 (25 Å)/ $SiON$ (16 Å)/ Si , as shown in Fig. 2(a). Here, C_2O is a provisional notation of the surface contamination on the sample, ignoring hydrogen because hydrogen cannot be observed by XPS. We assumed that a mixture of C, H, and O was deposited as the surface contamination layer. Following the notation of Eq. (3), the true values of the parameter set $\hat{\theta}$ are $\hat{\theta} = \{ \hat{T}_1 = 10, \hat{T}_2 = 25, \hat{T}_3 = 16, \hat{x}_{11} = \frac{2}{3}, \hat{x}_{12} = \frac{1}{3}, \hat{x}_{13} = 0, \hat{x}_{21} = \frac{1}{3}, \hat{x}_{22} = \frac{2}{3}, \hat{x}_{23} = 0, \hat{x}_{31} = \frac{1}{3}, \hat{x}_{32} = \frac{1}{3}, \hat{x}_{33} = \frac{1}{3} \}$. The geometrical configuration is shown in Fig. 2(b). The angle between the incident X-ray and the photoelectron detector was fixed at 45°, assuming the geometry of the actual experimental apparatus. By tilting the sample plane, we measured the photoelectron peak intensities at photoelectron emission angles of $\alpha = 0^\circ$ and 54.7° . The aperture angle of the photoelectron detector was set to $\pm 20^\circ$.

Under these conditions, a data set of peak area intensities can be generated using SESSA. However, the peak area intensities generated by SESSA are not count data. Therefore, we scaled the simulated intensity data to count data and then applied Poisson noise. In particular, to investigate the relationship between the measurement noise and the estimation accuracy, we prepared three different data sets with measurement times of 100, 10, and 1 min (Table 2). The measurement times were arbitrarily chosen for convenience only to indicate that the peak intensity ratio of data sets 1–3 was 100:10:1. The number of counts is larger and the relative noise is lower for longer measurement time. Here, we generated data such that the maximum counts (peak area intensity of Hf4f at $\alpha = 0$) for data sets 1–3 were approximately 100,000, 10,000, and 1000, respectively. We investigated how the confidence intervals of the parameters of the sample structure changed with the measurement time.

3.2. Experimental conditions for the estimation method

In the framework of Bayesian estimation, it is necessary to set up a prior distribution. For our sample structure model, all of the layers consist of three elements, the composition of each layer is always greater than or equal to 0, and the sum is 1. Therefore, the prior distributions for

the compositions $\{x_{11}, x_{12}, x_{13}\}$, $\{x_{21}, x_{22}, x_{23}\}$, and $\{x_{31}, x_{32}, x_{33}\}$ are all assumed to be uniform Dirichlet distributions of order three. The prior distributions for the thicknesses of the three overlayers were assumed to be uniform in the range from 0 to 50 Å for T_1 , and in the range from 0 to 100 Å for T_2 and T_3 .

For the exchange Monte Carlo setup, the number of replicas was set to 24 and the inverse temperature adjustment parameter was set to $\gamma = 2.0$. The first 2000 steps of the Monte Carlo simulation were discarded as burn-in, and the posterior probability distributions of the parameters were obtained by sampling the subsequent steps. In exchange Monte Carlo calculations, we have reduced the computational cost using the SLA (which is less accurate but very fast) instead of SESSA in regions where coarse estimation is sufficient, i.e., the burn-in period or high-temperature regions. Appendix B shows more details on how to reduce computation time.

An essential issue in implementing this method is how to incorporate the SESSA simulator into the exchange Monte Carlo method, which is explained in Appendix C.

3.3. Calculation results

We applied the proposed method to each of the three data sets obtained with different measurement times, obtaining more than about 10,000 sampling data in each case. The convergence of the Markov-chain Monte Carlo calculation was verified by the autocorrelation function, which is explained in Appendix E. The optimal parameter sets (labelled by the maximum a posteriori (MAP) estimator) obtained for each data set, together with the true parameter $\hat{\theta}$ assumed in the virtual measurement, are given in Table 3. The MAP estimators obtained from data set 1 approximately agreed with the true values. The composition of the first layer was $C_{69.3}N_{30.7}O_{0.1}$, compared with the true composition of $C_{66.7}O_{33.3}N_0$. The composition of the second layer was $Hf_{31.1}O_{68.7}N_{0.2}$, compared with the true composition of $Hf_{33.3}O_{66.7}N_0$. The composition of the third layer was $Si_{47.9}O_{7.8}N_{44.3}$, compared with the true composition of $Si_{33.3}O_{33.3}N_{33.3}$. The thicknesses of the layers ($\{T_1, T_2, \text{ and } T_3\}$) were $\{9.7, 28.1, 12.8\}$ Å compared with the true thicknesses of $\{10.0, 25.0, 16.0\}$ Å. For data set 2, it was more difficult for the MAP estimators to estimate the true values, and it was even more difficult for the MAP estimators to estimate the true values for data set 3. For example, in data set 3, the composition of the first layer was $C_{69.0}O_{31.0}N_{0.0}$, the composition of the second layer was $Hf_{29.5}O_{70.1}N_{0.3}$, the composition of the third layer was $Si_{54.5}O_{9.3}N_{36.2}$, and the layer thicknesses were $\{9.6, 31.6, 13.1\}$ Å. Because the measurement time increased from data set 3 to data set 1, the estimated parameters became closer to the true values. Although data set 3 had a very small electron count of at most 1000, it correctly estimated that there was almost no nitrogen in the first and second layers.

Bayesian posterior probability distributions were obtained from the sampling data generated by the Monte Carlo simulation, which allows us

to discuss the accuracy of the estimation of each parameter. The 5% and 95% percentile values for each parameter, indicating the ranges of the estimated parameters, are given in Table 3. For data set 1, the ranges of the C, O, and N contents in the first layer were [67.0%, 70.5%], [29.4%, 32.9%], and [0.0%, 0.2%], respectively. The ranges of the Hf, O, and N contents in the second layer were [30.2%, 32.9%], [66.9%, 69.6%], and [0.0%, 0.4%], respectively. The ranges of the Si, O, and N contents in the third layer were [36.9%, 50.5%], [1.2%, 27.6%], and [34.8%, 48.9%], respectively. The ranges of the thicknesses T_1 , T_2 , and T_3 were [9.3 Å, 10.1 Å], [26.0 Å, 29.0 Å], and [12.3 Å, 15.0 Å], respectively. This indicates that the estimation accuracy of the parameters was higher for the layers close to the surface and lower for the layers buried inside the sample. For data sets 2 and 3, the estimated ranges of the parameters widened and the estimation accuracy decreased compared with those for data set 1.

The posterior distributions of all 12 parameters for data set 1 are shown in Fig. 3. The method proposed by Cumpson et al. [4] aimed to estimate the chemical composition of single-phase compounds using Bayesian estimation. In contrast, our method simultaneously estimates both the film thickness and chemical composition of a multilayered sample. The three-component triangular diagrams of the compositions of the first CON layer, second HfON layer, and third SiON layer are shown in Fig. 3(a)–(c), respectively. The posterior distribution for the thickness of each layer is shown in Fig. 3(d). The points indicated by open circles in Fig. 3(a)–(c) and the dotted lines in Fig. 3(d) correspond to the true values of the sample structure parameters. The triangular diagram in Fig. 3(a) shows the posterior distribution of the compositions of the first CON layer, with the top vertex representing 100% carbon, the bottom left vertex representing 100% oxygen, and the bottom right vertex representing 100% nitrogen. In the triangular diagram, red areas indicate more samples, blue areas indicate fewer samples, and white areas contain no samples. The posterior distributions of the compositions of the first CON and second HfON layers were very sharp (Fig. 3(a) and (b)). The distribution of the compositions of the third SiON layer was a linearly spreading distribution (Fig. 3(c)), indicating that the composition ratio of silicon to nitrogen was constant. The true value indicated by the open circle was included in this distribution. The posterior distributions obtained from data sets 2 and 3 are shown in Figs. 4 and 5, respectively. For data sets 2 and 3, the posterior distribution clearly became wider and parameter estimation became more difficult than for data set 1. For example, the shape of the posterior distribution of the composition of the SiON layer changed from a linearly spreading distribution to a planar spreading distribution.

Compared with data set 1, parameter estimation for data set 3 was much more difficult. Although all of the distributions included the true values, the estimation accuracies were not good owing to the wide posterior distributions. This is consistent with the expectation that the relatively large noise (electron counts below approximately 1000) would make determination of the sample structure parameters difficult.

Table 3

Optimal values and statistics of each parameter for the three data sets. True are the true values given in advance, MAP are the MAP estimators, and 5% and 95% are the 5% and 95% percentile values, respectively.

		Data set 1 (Fig. 3)			Data set 2 (Fig. 4)			Data set 3 (Fig. 5)			
		True	5%	MAP	95%	5%	MAP	95%	5%	MAP	95%
Composition of the 1st layer (%)	C	66.7	67.0	69.3	70.5	65.2	69.2	74.8	58.8	69.0	86.2
	O	33.3	29.4	30.7	32.9	24.9	30.6	34.6	13.3	31.0	40.7
	N	0	0.0	0.1	0.2	0.0	0.3	0.6	0.0	0.0	1.4
Composition of the 2nd layer (%)	Hf	33.3	30.2	31.1	32.9	28.8	30.7	35.6	25.9	29.5	42.5
	O	66.7	66.9	68.7	69.6	63.8	69.1	70.7	56.6	70.1	73.3
	N	0	0.0	0.2	0.4	0.0	0.1	1.2	0.1	0.3	2.7
Composition of the 3rd layer (%)	Si	33.3	36.9	47.9	50.5	28.2	51.3	55.0	18.6	54.5	76.8
	O	33.3	1.2	7.8	27.6	1.7	2.1	48.8	1.9	9.3	71.9
	N	33.3	34.8	44.3	48.9	22.6	46.6	46.2	4.8	36.2	35.8
T_1 (Å)		10.0	9.3	9.7	10.1	8.8	9.8	10.9	7.0	9.6	12.5
T_2 (Å)		25.0	26.0	28.1	29.0	23.2	28.4	30.1	19.2	31.6	36.5
T_3 (Å)		16.0	12.3	12.8	15.0	11.5	12.2	17.9	9.5	13.1	25.9

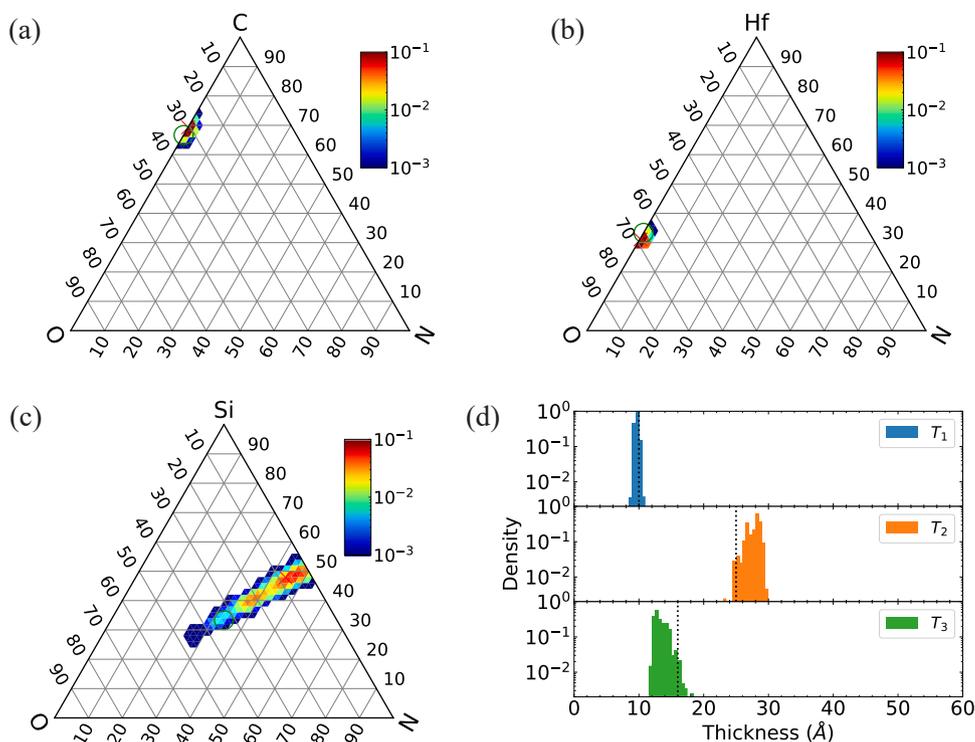


Fig. 3. Posterior distributions of the sample structure parameters obtained for data set 1 (measurement time 100 min). (a), (b), and (c) Compositional fractions of the first CON layer, second HfON layer, and third SiON layer, respectively. The posterior distributions are represented on triangular diagrams. The open circles in the figures indicate the true values, and the X markers indicate the MAP values. (d) Posterior distributions of film thicknesses T_1 , T_2 , and T_3 . The dotted lines indicate the true values.

4. Discussion

As shown in Section 3, although data set 1 had sufficient electron counts, it was difficult to identify the sample structure parameters for the deep third layer. It seems to be possible to reproduce the same intensity data by adjusting the composition and thickness of each layer. The model sample assumed in this study contained oxygen atoms in all

of the layers. It is expected that the total oxygen intensity can be kept constant by appropriately distributing the amount of oxygen to each layer. Because the intensity of photoelectrons emitted from the deeper layers of the sample decreases exponentially, the uncertainty of the oxygen content in the third layer is considered to be larger than that in the first and second layers. The other anion, nitrogen, was only contained in the third layer of the sample. A model in which nitrogen is also

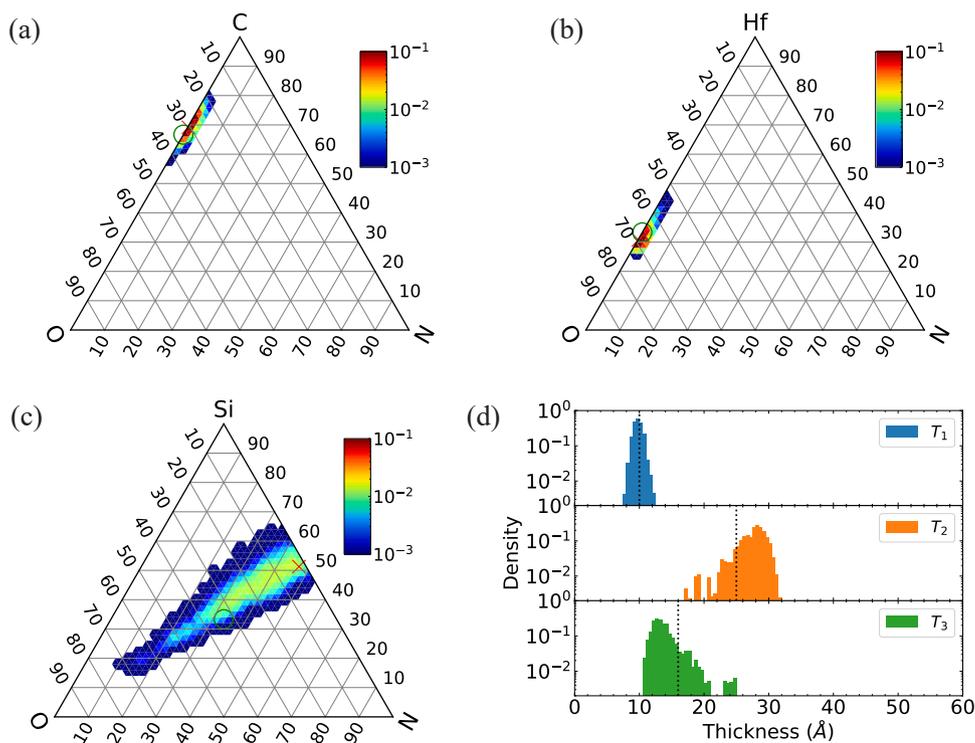


Fig. 4. Posterior distributions of the sample structure parameters obtained for data set 2 (measurement time 10 min). (a), (b), and (c) Compositional fractions of the first CON layer, second HfON layer, and third SiON layer, respectively. The posterior distributions are represented on triangular diagrams. The open circles in the figures indicate the true values, and the X markers indicate the MAP values. (d) Posterior distributions of film thicknesses T_1 , T_2 , and T_3 . The dotted lines indicate the true values.

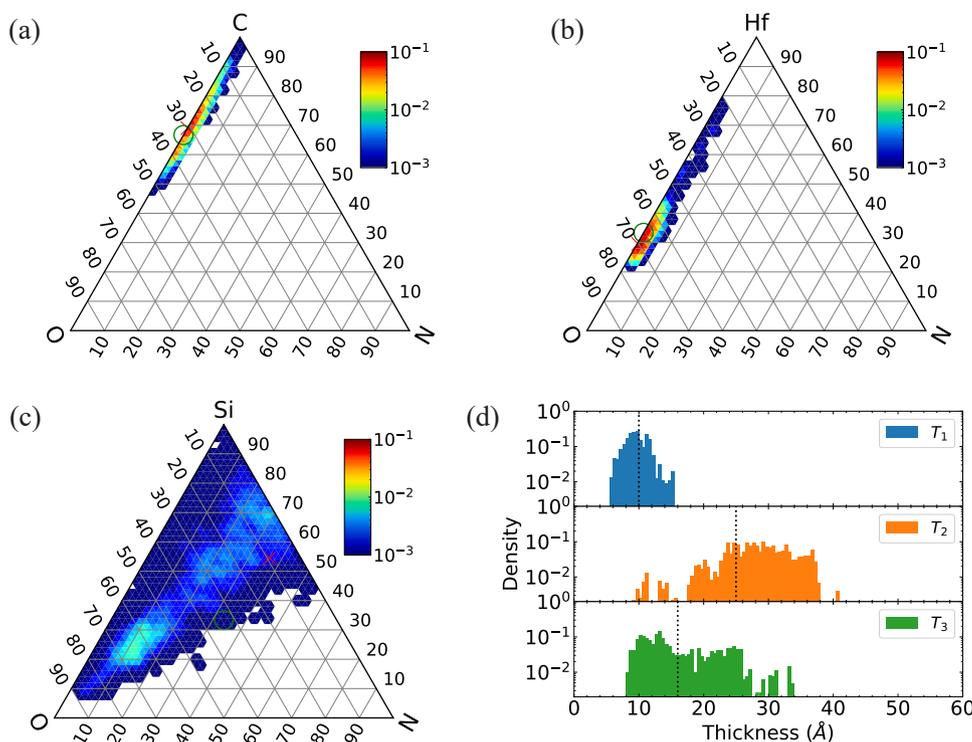


Fig. 5. Posterior distributions of the sample structure parameters obtained for data set 3 (measurement time 1 min). (a), (b), and (c) Compositional fractions of the first CON layer, second HfON layer, and third SiON layer, respectively. The posterior distributions are represented on triangular diagrams. The open circles in the figures indicate the true values, and the X markers indicate the MAP values. (d) Posterior distributions of film thicknesses T_1 , T_2 , and T_3 . The dotted lines indicate the true values.

present in the first or second layer cannot achieve the relationship between the nitrogen intensities at the two emission angles, as for data set 1, so such a model would have been rejected.

Next, we focus on the total thickness of the overlayers, $T_{\text{total}} = T_1 + T_2 + T_3$. Histograms of T_{total} for data sets 1–3 are shown in Fig. 6(a)–(c), respectively. Compared with Figs. 3(d), 4(d), and 5(d), the spread of the posterior distribution of T_{total} was smaller than those of the individual distributions of T_1 , T_2 , and T_3 . For example, the distribution of T_{total} for data set 1 was very sharp ($51.1 \pm 0.8 \text{ \AA}$). We believe this is because the peak area intensity of Si2p can be distinguished from the substrate silicon and compound silicon, because these peaks have a relatively large difference in the chemical shifts ($\sim 2 \text{ eV}$). If the inelastic mean free paths (IMFPs) at all of the overlayers are approximately the same, it is possible to determine the total film thickness that reproduces the Si2p intensities from the substrate at the two emission angles, which explains the sharp posterior distribution of T_{total} . It is interesting to note that this observation, which was not expected a priori, was obtained from the sampling results.

We have reported the results of precise analysis using SESSA, which was performed after a coarse primary search based on the SLA. For comparison, we only performed a coarse search based on the SLA to analyze data set 1 by the exchange Monte Carlo method. The posterior

distribution of each parameter obtained from the sampling results is shown in Fig. 7. Compared with the results of precise analysis using SESSA after a coarse primary search based on the SLA (Fig. 3), the compositions of the first and second layers and the thickness of the first layer had sharper distributions that only slightly differed from the true values. Conversely, the composition of the third layer and the thicknesses of the second and third layers had sharp distributions that significantly differed from the true values. The main reason for this is that the elastic scattering effect on the XPS spectral intensities is stronger in deeper layers from the surface, and thus the difference in the precision of the physical model between SESSA and the SLA is significant in terms of the accuracy and uncertainty of the results.

Let us assume that all the film thicknesses have been determined in advance by a method other than XPS. In this case, with the thickness parameters T_1 , T_2 , and T_3 fixed, only the composition of each layer is searched for, and, as a result, the posterior distributions of the compositions are expected to be more sharply distributed near their true values. Conversely, let us assume that the information about the sample preparation process is available and that the compositions are fixed. In this case, only the thickness parameter is searched for with the composition parameters fixed, and, as a result, the posterior distribution of the thickness of each layer will be close to the true value.

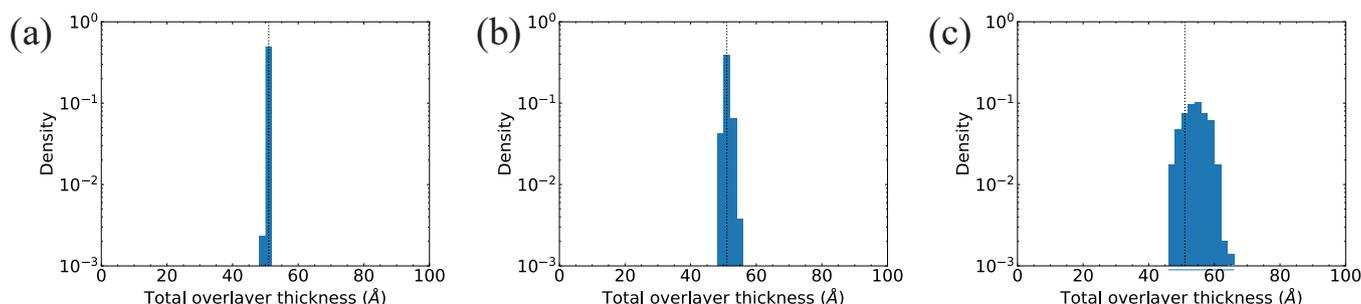


Fig. 6. Histograms of the total over-layer thickness $T_{\text{total}} = T_1 + T_2 + T_3$ for data sets (a) 1, (b) 2, and (c) 3.

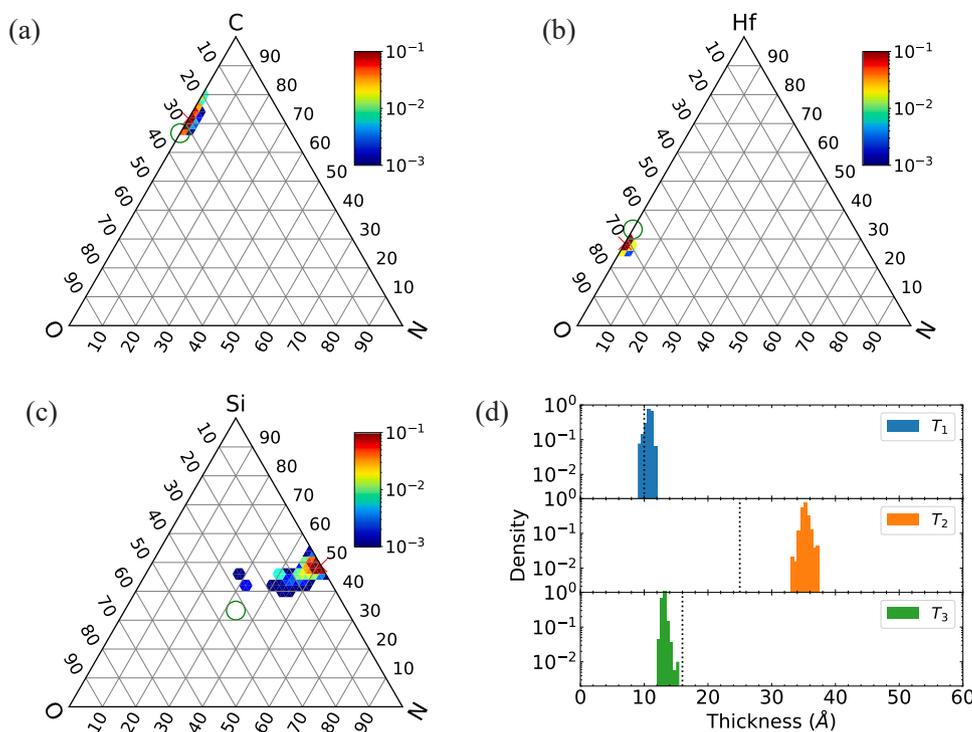


Fig. 7. Posterior distributions of the sample structure parameters obtained for data set 1 (measurement time 100 min) using only the SLA. (a), (b), and (c) Compositional fractions of the first CON layer, second HfON layer, and third SiON layer, respectively. The posterior distributions are represented on triangular diagrams. The open circles in the figures indicate the true values, and the X markers indicate the MAP values. (d) Posterior distributions of film thicknesses T_1 , T_2 , and T_3 . The dotted lines indicate the true values.

In Section 3.1, we made restrictive assumptions regarding the sample structure. Specifically, we assumed that the sample consists of four layers and limited the elemental species in each layer. The validity of these assumptions can be assessed through model selection using Bayesian estimation. For instance, we can prepare models with different numbers of layers, layer orders, and elemental species. It is possible to estimate from the measurement data which model is appropriate among them. These considerations should also be explored as future research tasks.

In Bayesian estimation, the estimated parameters are more accurate for more data. For example, by increasing the variation of the emission angles in ARXPS, it is possible to further refine the accuracy of the estimated parameters. It should be noted that further refinement of the accuracy of the parameters requires precise calculation by a simulator, such as SESSA.

5. Conclusions

We have developed a framework for solving the XPS inverse problem by incorporating the SESSA simulator into Bayesian estimation to obtain an overall picture of the distribution of plausible sample structure parameters from the measured XPS data. In this way, the procedure for running the simulator, which originally required human trial and error, was fully automated. In addition, it is now possible to not only obtain the optimal values of the sample structure parameters to be estimated, but also to evaluate the accuracy of the parameters from the posterior distribution in Bayesian estimation. The drawback of SESSA, which is its high computational cost, was overcome by combining the replica-exchange Monte Carlo method with the SLA.

As a concrete example, a virtual sample with a four-layered structure of CON/HfON/SiON/Si was designed, and the simplest ARXPS intensity data were generated by superimposing statistical noise on the SESSA data that simulate the photoelectron intensities from the Hf4f, Si2p, C1s, N1s, and O1s orbitals at two emission angles (0° and 54.7°). Using the above ARXPS intensity, we performed an inverse-problem solution in the framework to obtain the composition and thickness of each layer of the pre-designed sample, and we succeeded in estimating the sample

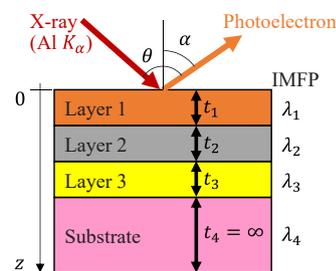


Fig. A.1. Definition of the sample structure and parameters for calculating the photoemission intensity in the SLA.

structure, including the true model. The estimation accuracy was higher for higher intensity of the spectrum. For example, for a data set with a maximum count of 1000, the estimated film thickness was $10.3 \pm 9.1 \text{ \AA}$, compared with the true value of 10 \AA for the thickness of the top layer. Conversely, for a data set with a maximum count of 100,000, the estimated thickness was $10.1 \pm 1.0 \text{ \AA}$, which is very good accuracy. Moreover, even for the same measurement intensity, the estimation accuracy was higher for the thickness and composition parameters of a layer closer to the surface, where the electron counts are higher.

The advanced features and many adjustable parameters of SESSA make it difficult for users to manually search for the sample structure parameters that match the experimental data. The proposed method incorporates SESSA into Bayesian estimation to find the appropriate sample structures among many candidates, paving the way for future use of SESSA.

From a practical perspective, the proposed method requires pre-processing of the spectrum by separating the peaks and calculating the signal area as count data, assuming that the noise in the count data follows a Poisson distribution. However, as is well known, the signal area depends on the choice of background, and it is difficult to separate the signal-derived noise from the background-derived noise. In an actual raw XPS spectrum, Poisson noise is added to the sum of background and signal at each point along the x-axis. Ideally, Bayesian estimation should

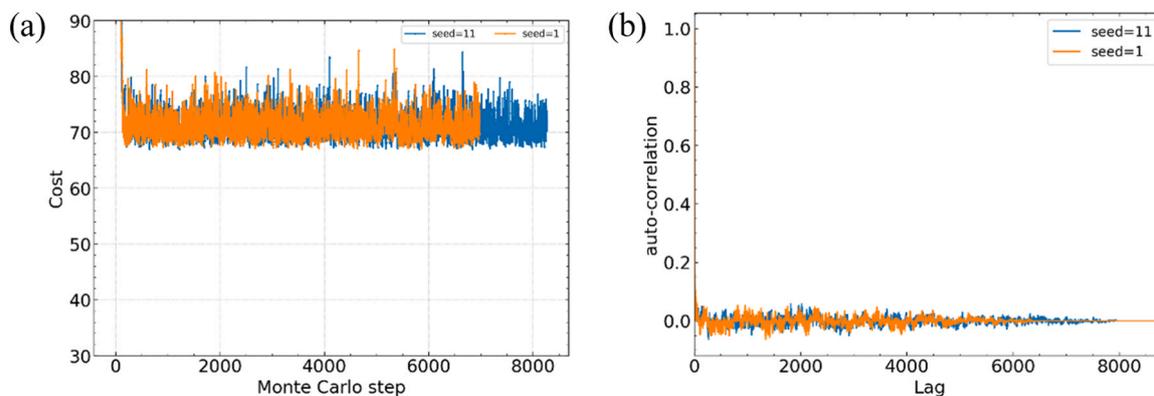


Fig. E.1. (a) History of the cost function for data set 1. (b) Autocorrelation function of the cost function after 1000 steps.

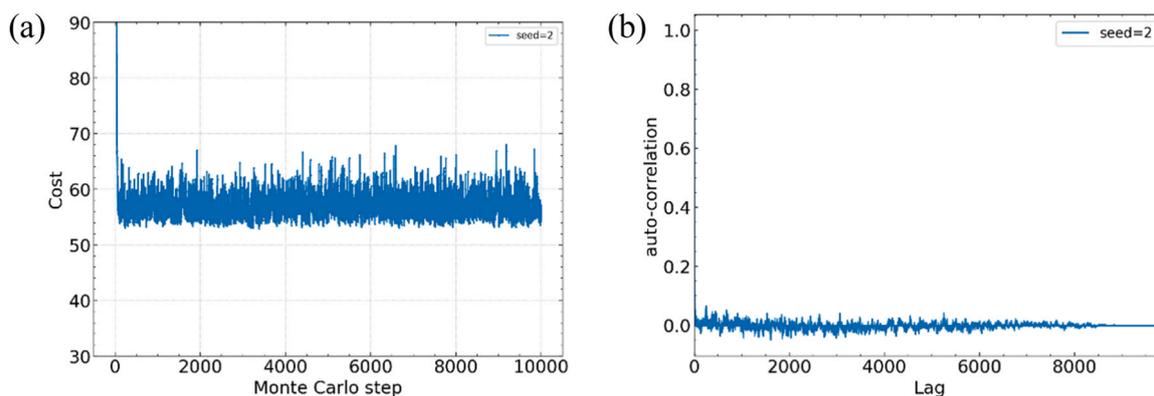


Fig. E.2. (a) History of the cost function for data set 2. (b) Autocorrelation function of the cost function after 1000 steps.

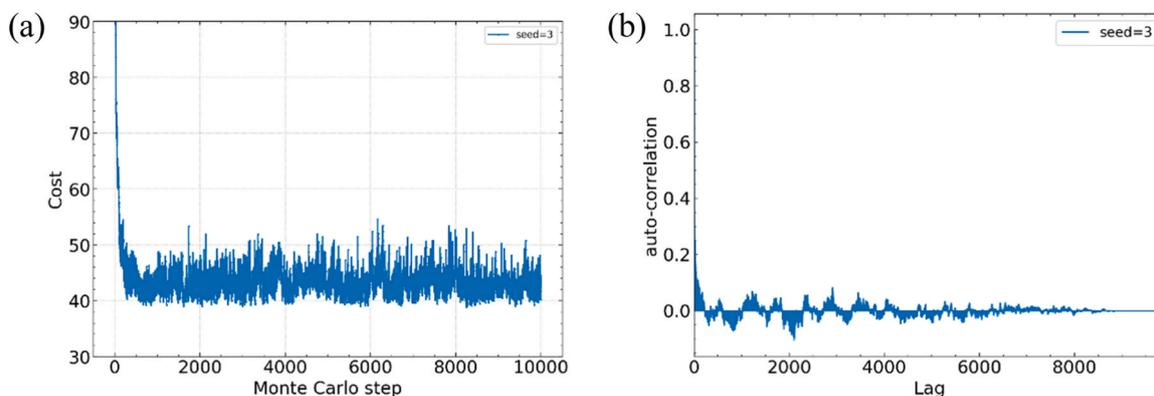


Fig. E.3. (a) History of the cost function for data set 3. (b) Autocorrelation function of the cost function after 1000 steps.

be performed on the spectrum data themselves, including the total counts of the signal and background. SESSA can simulate spectra with backgrounds due to inelastic scattering, and is therefore suitable for this purpose. It should be considered as future work to investigate this approach further.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgements

We would like to thank Dr. Shigeo Tanuma, who provided invaluable advice and guidance for this research. This work was supported by JST CREST under grant number JPMJCR1761 and by a JSPS KAKENHI Grant-in-Aid for Scientific Research (C) under grant number 19K12154.

Appendix A. Bayesian estimation and exchange Monte Carlo method

In this section, we describe a framework of the Bayesian estimation and the exchange Monte Carlo method used in our proposed method.

Bayesian estimation

Bayesian estimation is a framework in which the process of generating data in a probabilistic model is formulated and an estimation is made by tracing back the causal relationship using the Bayesian theorem. For more information on Bayesian statistics, please refer to Ref. [3]. We first define a probabilistic model. Suppose that the measured photoelectron peak intensity data set is $D = \{y_i\}_{i=1}^n$, the sample model parameter is θ , and the calculated intensity data set is $\{f_i(\theta)\}_{i=1}^n$. Given a calculated intensity value of f , the probability of obtaining the measured value y can be calculated by the Poisson distribution function $f^y e^{-f}/y!$. That is, the conditional probability of obtaining the measured data y_i when the parameter θ is given can be calculated via

$$p(y_i|\theta) = \frac{f_i(\theta)^{y_i} e^{-f_i(\theta)}}{y_i!}.$$

Assuming that the data $\{y_i\}_{i=1}^n$ are independent and identically distributed, the cost function $E(\theta)$ can be used to calculate the probability of the data set D via

$$p(D|\theta) = \prod_{i=1}^n p(y_i|\theta) \propto \exp(-E(\theta)),$$

$$E(\theta) = \sum_{i=1}^n (f_i(\theta) - y_i \log f_i(\theta) + \log \Gamma(y_i + 1)).$$

In the Bayesian estimation, we regard not only the data D , but also the parameter set θ as random variables. That is, we construct a model of the simultaneous probability distribution $p(D, \theta)$. In the Bayesian estimation, we consider the simultaneous probability distribution $p(D, \theta)$:

$$p(D, \theta) = p(D|\theta)p(\theta) \propto \exp(-E(\theta))p(\theta).$$

Here, $p(\theta)$ is a prior distribution, and it should be set in advance. According to the Bayesian theorem, we estimate the parameter θ using the posterior distribution $p(\theta|D)$:

$$p(\theta|D) = \frac{p(D, \theta)}{p(D)} = \frac{p(D, \theta)}{\int p(D, \theta) d\theta} = \frac{\exp(-E(\theta))p(\theta)}{\int \exp(-E(\theta))p(\theta) d\theta} = \frac{1}{Z} \exp(-E(\theta))p(\theta),$$

where

$$Z = \int \exp(-E(\theta))p(\theta) d\theta.$$

This technique for obtaining the parameter θ that maximizes the posterior distribution $p(\theta|D)$ is a MAP estimation, and the parameter θ at this time is called the MAP estimator. Bayesian estimation has the advantage that both the optimum value of the parameter and the confidence interval can be evaluated by obtaining the width of the posterior distribution $p(\theta|D)$.

Exchange Monte Carlo method

How to evaluate the posterior probability $p(\theta|D)$ is a problem when calculating the spectral decomposition through Bayesian estimation. In the case of spectral decomposition, the posterior distribution for the parameter θ is difficult to handle because the distribution does not become a known distribution such as a Gaussian distribution. For example, it is difficult to determine the MAP estimators because of the presence of local solutions. Furthermore, when the confidence interval of the parameter θ needs to be determined, it is necessary to determine the shape of the distribution itself, which is much more difficult.

The exchange Monte Carlo method [7,8], a type of Markov-chain Monte Carlo method, can evaluate the posterior distribution $p(\theta|D)$. We prepare a plurality of distributions

$$p_{\beta}(\theta|D) = \frac{1}{Z_{\beta}} \exp(-E(\theta))p(\theta),$$

$$Z_{\beta} = \int \exp(-\beta E(\theta))p(\theta) d\theta,$$

in which the inverse temperature β is introduced into the posterior distribution, and simultaneously perform sampling in parallel. That is, the target distribution becomes the simultaneous distribution

$$p(\theta_1, \dots, \theta_M) = \prod_{m=1}^M p_{\beta_m}(\theta_m|D).$$

Here, the inverse temperature $\{\beta_m\}_{m=1}^M$ is $0 = \beta_1 < \beta_2 < \dots < \beta_M = 1$, and M is the number of prepared inverse temperatures. θ_m are the fitting parameters for inverse temperature β_m . For $\beta_M = 1$, $p_{\beta_M}(\theta_M|D)$ matches the posterior distribution $p(\theta|D)$.

The algorithm of the EMC method comprises two updates.

Metropolis sampling for each temperature. For each inverse temperature β_m , θ_m from the distribution of $p_{\beta_m}(\theta|D)$ is sampled using the Metropolis algorithm.

State exchange between adjacent temperatures. The states between adjacent temperatures are exchanged and $\{\theta_m, \theta_{m+1}\} \rightarrow \{\theta_{m+1}, \theta_m\}$. Whether to exchange is determined on the basis of the probability

$$p(\theta_m \leftrightarrow \theta_{m+1}) = \min(1, v),$$

$$v = \frac{p_{\beta_m}(\theta_{m+1}|D)p_{\beta_{m+1}}(\theta_m|D)}{p_{\beta_m}(\theta_m|D)p_{\beta_{m+1}}(\theta_{m+1}|D)} = \exp((\beta_{m+1} - \beta_m)(E(\theta_{m+1}) - E(\theta_m))).$$

As a result, the parameter θ_M obtained from the M -th inverse temperature β_M can be regarded as sampling from the posterior distribution $p(\theta|D)$. Thus, by repeating the above procedure sufficiently many times and recording θ_M , a sample sequence from the posterior distribution $p(\theta|D)$ is obtained, and the MAP estimators and confidence interval can be determined.

Using the exchange Monte Carlo method, the heating and annealing effects can be introduced by stochastically exchanging samples between adjacent temperatures during sampling of each distribution. Thereby, the algorithm can escape from local solutions and efficiently search for the global solution. This is an important advantage for accurate estimation of MAP estimators and confidence intervals.

Appendix B. Techniques for reducing the calculation time

To reduce the computational cost, we partially use the SLA instead of SESSA in the exchange Monte Carlo method. The major bottleneck of our technique is that SESSA requires approximately 1.6 min of computation time per sample structure. Therefore, when coarse estimation is acceptable, it is sufficient to use the SLA instead of SESSA. SLA calculations are approximately a million times faster than SESSA calculations, although they are less accurate. The details of the SLA calculation are described in [Appendix D](#).

The first 2000 Monte Carlo steps are performed with the SLA because they only require coarse estimation. We also prepare 24 replicas in the exchange Monte Carlo method, of which the 12 inverse temperatures on the high-temperature side are computed using the SLA, because coarse estimation is sufficient. The SESSA calculations are only performed after 2000 steps, and they are only performed for the 12 inverse temperatures on the low-temperature side. The exchange Monte Carlo method is suitable for parallel computing. The computation time can be reduced by dividing the 24 replicas into 12 sets of two replicas and using 12-thread parallel computation. These innovations reduce the computational cost to one-sixth of the original computational cost and allow the computation to be performed in a reasonable time.

The specifications of the computer were as follows: the operating system was Ubuntu 20.04LTS, the processor was an AMD EPYC 7451 (2.3 GHz) processor, and the RAM was 128 GB. The time required to calculate 1000 steps was 14.8 days for data set 1, 2.9 days for data set 2, and 1.5 days for data set 3. We increased the number of samplings by running the exchange Monte Carlo method simultaneously with 10 different initial random values.

Appendix C. How to incorporate the SESSA simulator into a C program

An essential issue in implementing this method is how to incorporate the SESSA simulator into the parameter search algorithm. We developed the method by combining the Linux version of the SESSA simulator, a C program, and virtual frame buffer (XVFB) technology [14].

The graphical user interface (GUI) is generally used by SESSA users. A user-friendly GUI allows users to visually set up the sample structure and instrument geometry and to visualize the simulation results. Experienced users can give dedicated commands to the command line interface of SESSA to perform various operations equivalent to GUI operations. A user can also prepare a SESSA script file containing a series of commands, give it to the SESSA command line interface, and run it. This allows a user to set a large number of parameters at once without time-consuming GUI operations.

For a more-advanced approach, SESSA can also be called from external applications. The SESSA manual gives the procedure for using SESSA from a C program [2]: start SESSA in pipe mode, assign it a file pointer, and execute arbitrary processes by relying on SESSA-specific commands through the pipe. Although it requires advanced programming skills, it is highly compatible with fully automated parameter searches. The user needs to create the connecting parts, such as converting the tuning parameters to SESSA commands and reading the SESSA output file in the external application, but this is not very difficult.

When invoking SESSA in this way, it is also important to suppress the display of the SESSA GUI. Even if SESSA is started in pipe mode, the GUI window of SESSA opens on the real screen of the computer. This GUI image processing is unnecessary for simulating XPS, and it significantly slows down the computational speed. This is where the Linux technology XVFB plays an important role. The XVFB can suppress the display of the GUI on the real screen, dramatically improving the performance.

Note that in rare cases, SESSA may hang up for unknown reasons. It is also necessary to design the timeout, force SESSA to terminate the process, and rerun it after a certain period of time has elapsed.

Appendix D. Photoelectron intensity for a multilayer sample within the SLA

The SLA is a method to calculate the photoelectron intensity under the approximation that photoelectrons generated in a solid sample only travel in a straight path. Elastic scattering effects are ignored, such as photoelectrons changing their direction of motion in the solid. Because of this very simple approximation, the intensity of photoemission from multilayer samples can be analytically calculated. Therefore, it is very fast, although its quantitative accuracy is questionable.

In the SLA, the detected photoelectron intensity is expressed as

$$I = I_0 F_A F_B F_C,$$

where I_0 is the X-ray intensity, which is regarded as a constant value here. F_A is a geometric factor, and $F_A \propto 1/\cos\alpha$ when the X-ray spot area and detector acquisition solid angle are constant, where α is the photoelectron emission angle. F_B is a factor related to photoexcitation, and it is expressed as

$$F_B = \sigma_{nl}(h\nu) W(\beta_{nl}(h\nu), \theta),$$

where σ_{nl} is the photoionization cross section, which depends on the atomic species, atomic orbital n, l , and energy of the X-ray $h\nu$. Under the dipole approximation, $W(\beta_{nl}(h\nu), \theta)$ is expressed as

$$W(\beta_{nl}(h\nu), \theta) = \frac{1}{4\pi} \left[1 + \frac{\beta_{nl}(h\nu)}{2} \left(\frac{3}{2} \sin^2\theta - 1 \right) \right],$$

using the anisotropy factor $\beta_{nl}(h\nu)$ and the angle between the X-ray and detector θ .

F_C is the photoelectron intensity including electron attenuation in the solid. This term depends on the sample structure and the approximation of electron transport in the solid. Hereafter, the calculations are performed assuming a four-layered structure, as shown in Fig. A.1. In the SLA, the escape depth distribution function $\phi(z)$ can be analytically obtained as follows:

$$\phi(z) = \begin{cases} \exp\left(-\frac{z}{\lambda_1\mu}\right) & (\text{for } 0 \leq z \leq t_1) \\ \exp\left(-\frac{t_1}{\lambda_1\mu} - \frac{z-t_1}{\lambda_2\mu}\right) & (\text{for } t_1 \leq z \leq t_1+t_2) \\ \exp\left(-\frac{t_1}{\lambda_1\mu} - \frac{t_2}{\lambda_2\mu} - \frac{z-(t_1+t_2)}{\lambda_3\mu}\right) & (\text{for } t_1+t_2 \leq z \leq t_1+t_2+t_3) \\ \exp\left(-\frac{t_1}{\lambda_1\mu} - \frac{t_2}{\lambda_2\mu} - \frac{t_3}{\lambda_3\mu} - \frac{z-(t_1+t_2+t_3)}{\lambda_4\mu}\right) & (\text{for } z \geq t_1+t_2+t_3) \end{cases}$$

where z is the depth from the surface.

The depth distribution function was defined such that $\phi(z=0) = 1$. The thicknesses of the first, second, and third layers from the surface were set to t_1, t_2 , and t_3 , respectively, and the thickness of the fourth layer, the substrate layer, was set to $t_4 = \infty$. The IMFPs of the four layers were set to $\lambda_1, \lambda_2, \lambda_3$, and λ_4 . Note that the IMFPs depend on the material and photoelectron kinetic energy. We also defined $\mu \equiv \cos\alpha$, where α is the photoelectron emission angle. The atomic number densities of the atomic species corresponding to the atomic orbitals of interest are now denoted n_1, n_2, n_3 , and n_4 , in order from the top layer to the bottom layer. The F_C term can then be calculated as follows [1]:

$$\begin{aligned} F_C &= \int_0^{\infty} n(z)\phi(z)dz \\ &= n_1 \int_0^{t_1} \phi(z)dz + n_2 \int_{t_1}^{t_1+t_2} \phi(z)dz + n_3 \int_{t_1+t_2}^{t_1+t_2+t_3} \phi(z)dz + n_4 \int_{t_1+t_2+t_3}^{\infty} \phi(z)dz \\ &= n_1 \left\{ 1 - \exp\left(-\frac{t_1}{\lambda_1\mu}\right) \right\} \lambda_1\mu \\ &\quad + n_2 \exp\left(-\frac{t_1}{\lambda_1\mu}\right) \left\{ 1 - \exp\left(-\frac{t_2}{\lambda_2\mu}\right) \right\} \lambda_2\mu \\ &\quad + n_3 \exp\left(-\frac{t_1}{\lambda_1\mu}\right) \exp\left(-\frac{t_2}{\lambda_2\mu}\right) \left\{ 1 - \exp\left(-\frac{t_3}{\lambda_3\mu}\right) \right\} \lambda_3\mu \\ &\quad + n_4 \exp\left(-\frac{t_1}{\lambda_1\mu}\right) \exp\left(-\frac{t_2}{\lambda_2\mu}\right) \exp\left(-\frac{t_3}{\lambda_3\mu}\right) \lambda_4\mu. \end{aligned}$$

The atomic number density is zero for layers in which the element of interest is not present. The photoionization cross sections, asymmetry parameters, IMFPs, and atomic number densities required for the calculations can be obtained from SESSA.

Appendix E. Convergence verification of the Markov-chain Monte Carlo method using the autocorrelation function

In this section, we confirm that our Monte Carlo sampling was properly executed by examining the autocorrelation function. Because our method is based on the Markov-chain Monte Carlo (MCMC) approach, the sampling at short time intervals becomes dependent on past values. If such correlation is small, it indicates the convergence of the MCMC method. The autocorrelation function can be used to confirm this convergence.

Figure E.1 (a) represents the history of the cost function for data set 1. Because of its high computational cost, we prepared two random seeds to increase the number of samples. The first 1000 steps were discarded as burn-in, and the remaining steps were obtained as samples. Figure E.1 (b) shows the autocorrelation function calculated for the obtained cost function. The horizontal axis represents the lag, and the vertical axis represents the

autocorrelation values. From the plot, we can observe that the autocorrelation values are mostly below 0.1 within 100 steps, indicating little correlation between samples that are further apart. This suggests that the sampling was performed with sufficient convergence. Figures E.2 and E.3 present similar autocorrelation function analyses for data sets 2 and 3, respectively. These figures also show that sampling was performed with sufficient convergence.

References

- [1] J. Matthew, in: D. Briggs, J.T. Grant (Eds.), *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, IMPublications, Chichester, UK and SurfaceSpectra, Manchester, UK, 2003.
- [2] W.S.M. Werner, W. Smekal, C.J. Powell, *Simulation of Electron Spectra for Surface Analysis (SESSA) - 2.2.0*, National Institute of Standards and Technology, Gaithersburg, MD, 2021.
- [3] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006. ISBN-10: 0387310738. ISBN-13: 978-0387310732.
- [4] P.J. Cumpson, M.P. Seah, Random uncertainties in AES and XPS: II: a quantification using either relative or absolute measurements, *Surf. Interface Anal.* 18 (1992) 361–367, <https://doi.org/10.1002/sia.740180509>.
- [5] K. Nagata, R. Muraoka, Y. Mototake, T. Sasaki, M. Okada, Bayesian spectral deconvolution based on Poisson distribution: Bayesian measurement and virtual measurement analytics (VMA), *J. Phys. Soc. Jpn.* 88 (2019) 1–7, <https://doi.org/10.7566/JPSJ.88.044003>.
- [6] A. Machida, K. Nagata, R. Murakami, H. Shinotsuka, H. Shouno, H. Yoshikawa, M. Okada, Bayesian estimation for XPS spectral analysis at multiple core levels, *Sci. Technol. Adv. Mater.: Methods* 1 (2021) 123–133, <https://doi.org/10.1080/27660400.2021.1943172>.
- [7] K. Nagata, S. Sugita, M. Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Netw.* 28 (2012) 82–89, <https://doi.org/10.1016/j.neunet.2011.12.001>.
- [8] K. Hukushima, K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, *J. Phys. Soc. Jpn.* 65 (1996) 1604–1608, <https://doi.org/10.1143/JPSJ.65.1604>.
- [9] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, New York, 1996. ISBN-10: 0412055511. ISBN-13: 978-0412055515.
- [10] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, *Bayesian Data Analysis, Third ed.*, Chapman & Hall/CRC, New York, 2013. ISBN-10: 1439840954. ISBN-13: 978-1439840955.
- [11] *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition*, Chapman & Hall/CRC, New York, 2006. ISBN-10: 1584885874. ISBN-13: 978-1584885870.
- [12] H. Shinotsuka, K. Nagata, H. Yoshikawa, Y. Mototake, H. Shouno, M. Okada, Development of spectral decomposition based on Bayesian information criterion with estimation of confidence interval, *Sci. Technol. Adv. Mater.* 21 (2020) 402–419, <https://doi.org/10.1080/14686996.2020.1773210>.
- [13] ISO 13424:2013 Surface chemical analysis – X-ray photoelectron spectroscopy – Reporting of results of thin-film analysis, 2013.
- [14] XVFB. <https://www.x.org/releases/X11R7.7/doc/man/man1/Xvfb.1.xhtml> (Accessed on Dec. 16, 2022).