Taylor & Francis
Taylor & Francis Group

Science and Technology of Advanced Materials: Methods

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tstm20

# Maximum a posteriori estimation for high-throughput peak fitting in X-ray photoelectron spectroscopy

**Tarojiro Matsumura, Naoka Nagamura, Shotaro Akaho, Kenji Nagata & Yasunobu Ando**

**To cite this article:** Tarojiro Matsumura, Naoka Nagamura, Shotaro Akaho, Kenji Nagata & Yasunobu Ando (2024) Maximum a posteriori estimation for high-throughput peak fitting in X-ray photoelectron spectroscopy, Science and Technology of Advanced Materials: Methods, 4:1, 2373046, DOI: 10.1080/27660400.2024.2373046

**To link to this article:** https://doi.org/10.1080/27660400.2024.2373046

SCIENCE AND TECHNOLOGY OF ADVANCED MATERIALS: METHODS
2024, VOL. 4, NO. 1, 2373046
https://doi.org/10.1080/27660400.2024.2373046



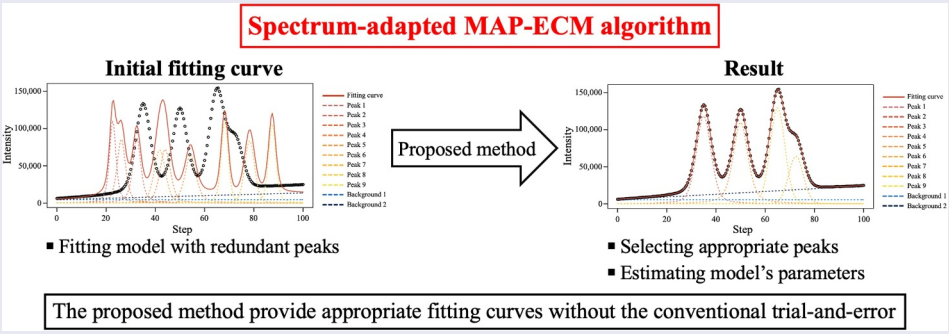
# Maximum a posteriori estimation for high-throughput peak fitting in X-ray photoelectron spectroscopy

Tarojiro Matsumura [a], Naoka Nagamura [b,c,d], Shotaro Akaho [e], Kenji Nagata [b] and Yasunobu Ando [a]

[a]National Institute of Advanced Industrial Science and Technology (AIST), Research Center for Computational Design of Advanced Functional Materials (CD-FMat), Tsukuba, Japan; [b]National Institute for Materials Science (NIMS), Center for Basic Research on Materials, Tsukuba, Japan; [c]PRESTO, Japan Science and Technology Agency, Saitama, Japan; [d]Faculty of Advanced Engineering, Tokyo University of Science, Tokyo, Japan; [e]National Institute of Advanced Industrial Science and Technology (AIST), Human Informatics and Interaction Research Institute (HIIRI), Tsukuba, Japan

**ABSTRACT**

We introduce a peak fitting method to estimate the model parameters and the number of peaks without using the conventional trial-and-error approach. The proposed method automatically removes excess peaks using maximum a posteriori estimation. The computation is performed efficiently by the spectrum-adapted expectation–conditional maximisation algorithm with deterministic annealing. We apply the proposed method to synthetic and experimental data from a tunnel field-effect transistor. The proposed method identified two peak components in the experimental data from a $MoS_2$ sheet, which are interpreted to be the Mo $3d_{3/2}$ and Mo $3d_{5/2}$ peaks. No peaks were detected on the $p^+$-$WSe_2$ sheet and hexagonal boron nitride (*h*-BN) within the measured binding energy range.




**IMPACT STATEMENT**

The spectrum-adapted MAP – ECM algorithm estimates the model's parameters and selects the peaks in a single calculation without the conventional trial-and-error. The method is a valuable tool for high-throughput analysis.

**ARTICLE HISTORY**
Received 27 February 2024
Revised 10 June 2024
Accepted 23 June 2024

**KEYWORDS**
Peak fitting; X-ray photoelectron spectroscopy; maximum a posteriori estimation; expectation–conditional maximisation algorithm; high-throughput analysis

**JEL CLASSIFICATION**
Methodology, apparatus, and experimental design; Materials, data analysis, and utilisation

## 1. Introduction

Recent spectroscopic techniques have enabled the investigation of broad areas of samples at high resolution. Developing an efficient method for analysing the large volume of spectral data generated by this approach is a crucial issue in materials science [1–5]. The peak fitting of spectral data requires manual trial-and-error to determine the number of peaks, subtract the background, and estimate the model parameters. Because this trial-and-error approach is a major bottleneck, only a part of the spectral data can be analysed despite the large volume of data available.

An effective method would improve on the conventional trial-and-error approach [6–12]. Matsumura et al. [6] proposed the spectrum-adapted expectation maximisation (EM) algorithm as a high-throughput peak fitting method. The spectrum-adapted expectation conditional maximisation (ECM) algorithm is an improved fitting model [7] that has been applied to Mo 3*d* and S 2*s* core-level spectra measured in monolayer $MoS_2$–Nb-doped $MoS_2$ lateral homojunctions [13]. By extending the model, peak fitting and background subtraction can be conducted using joint optimisation [8], which has been demonstrated using a large volume of spectral data collected from an SnS monolayer sheet.

A limitation of the spectrum-adapted EM algorithm is that the number of peaks cannot be selected

**CONTACT** Yasunobu Ando ando.y.am@m.titech.ac.jp National Institute of Advanced Industrial Science and Technology (AIST), Research Center for Computational Design of Advanced Functional Materials (CD-FMat), Tsukuba central 2, 1-1-1, Umezono, Ibaraki, Tsukuba 305-8568, Japan
Supplemental data for this article can be accessed online at https://doi.org/10.1080/27660400.2024.2373046

automatically and must be fixed as a hyperparameter [6–8]. The number of peaks can be estimated using criteria [9,12,14]. For example, Nagata et al. [14] calculated the marginal likelihood using a replica exchange Markov chain Monte Carlo method and determined the appropriate number of peaks in the analysis of synthetic spectral data and reflectance spectral data from olivine. In contrast, because the spectrum-adapted EM algorithm is not based on Bayesian inference, it cannot be used to obtain the marginal likelihood. Moreover, calculating the marginal likelihood is computationally expensive and not compatible with high-throughput analysis. Recently, Shinotsuka et al. [9] and Murakami et al. [12] used the Bayesian information criterion (BIC) [15], an asymptotic approximation of the marginal likelihood, to determine the number of peaks. The BIC is easier to calculate than the marginal likelihood; however, the accuracy of peak fitting using BIC requires verification.

An alternative approach without criteria is sparse modelling. Sparse modelling optimises the model by removing unnecessary variables. The model assumes that the number of relevant parameters is less than the total number of parameters prepared in advance and allows variable selection by requiring that the number of variables is small and the model accurately corresponds to the data. The Least Absolute Shrinkage and Selection Operator (LASSO) [16] is a typical sparse modelling technique used in materials science (e.g. [17,18]).

We introduce the spectrum-adapted maximum a posteriori (MAP) – ECM algorithm for high-throughput peak fitting. The proposed method estimates the parameters of the fitting model and selects the number of peaks based on a single sparse modelling calculation by applying MAP estimation. MAP estimation is an approach for point estimation by maximizing a posterior distribution that consists of a prior distribution and a likelihood function. The parameters are assumed to be random variables, and the prior distribution incorporates constraints on the values of the parameters using probability distributions. Excess peaks, which do not contribute to the explanation of the spectral data, are removed by reducing their mixing coefficients to approach 0 during the calculation [19,20]. As the proposed method calculates a model with many peak components, the estimated values of the parameters may be trapped in a local optimum. To avoid this, we use a deterministic annealing technique [21–23].

We apply the spectrum-adapted MAP – ECM algorithm to synthetic and experimental data. The experimental data are 7396 spectra collected from a tunnel field-effect transistor (TFET) by X-ray photoemission spectroscopy (XPS), and we produced a map of Mo $3d_{5/2}$ to represent the spatial variation in peak binding energy.

## 2. MAP estimation

### 2.1. Likelihood function

We define a likelihood function $L$ that shows the goodness-of-fit between the model and the data. The spectral data, collected by X-ray photoelectron spectroscopy (XPS), consist of the measured energy step $\boldsymbol{x} = x_n$ and corresponding intensity $\boldsymbol{w} = w_n$. As the intensity is the count of photoelectrons, the error can be approximated by a Poisson distribution [6–8]. Given that a Poisson distribution takes into account the error distribution of the spectral data, it is preferable to the conventional non-linear least square approach that uses the residual sum of squares between the model and the spectra, which assumes the error to be Gaussian (see also supporting information). We emphasise that the use of the Poisson distribution is a notable feature of our method.

The probability distribution of the intensity $w_n$ corresponding to each energy step $x_n$ using the Poisson distribution can be expressed as follows:

$$p(w_n|x_n,\boldsymbol{\theta}) = Poisson(w_n|\xi_n) = \frac{\xi_n^{w_n}\exp(-\xi_n)}{w_n!}, \quad (1)$$

where the mean parameter $\xi_n > 0$ is a function of model parameter $\boldsymbol{\theta}$, and here it is modeled by the sum of probability density $\mathbf{P}(x_n|\theta_k), k = 1,\ldots,M$ as

$$\xi_n = \sum_{k=1}^{M} \mathbf{P}(x_n|\theta_k). \quad (2)$$

Therefore the likelihood function $L$ can be written as

$$L = \prod_{n=1}^{N} p(w_n|x_n,\boldsymbol{\theta})$$

$$= \prod_{n=1}^{N}\left\{\frac{\xi_n^{w_n}\exp(-\xi_n)}{w_n!}\right\}. \quad (3)$$

Taking the logarithm of Equation (3), we get

$$\ln(L) = \sum_{n=1}^{N}\left\{w_n \ln\left\{\sum_{k=1}^{M}\mathbf{P}(x_n|\theta_k)\right\} - \ln(w_n!) - \sum_{k=1}^{M}\mathbf{P}(x_n|\theta_k)\right\}. \quad (4)$$

The second term on the right side of Equation (4) is a constant. When the normalization constraint is applied to a peak component, the third term in Equation (4) becomes constant,

$$\sum_{n=1}^{N}\sum_{k=1}^{M}\mathbf{P}(x_n|\theta_k) = 1. \quad (5)$$

Hence, the second and third terms of Equation (4) are not dependent on the parameter set $\boldsymbol{\theta}$. Let $E(\boldsymbol{x}|\boldsymbol{\theta})$ be the function that extracts the term in Equation (4) that affects the optimisation of $\boldsymbol{\theta}$, as follows:

$$E(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{n=1}^{N}\left\{\boldsymbol{w}_n \ln\left\{\sum_{k=1}^{M}\mathbf{P}(\boldsymbol{x}_n|\boldsymbol{\theta}_k)\right\}\right\}. \quad (6)$$

Equation (6) has the same form as the weighted log-likelihood function in the spectrum-adapted EM algorithm [6] and spectrum-adapted ECM algorithm [7].

This study uses a pseudo-Voigt mixture model with a linear background model $\mathbf{P}_{\mathbf{PVMMB}}$ as a fitting model. $\mathbf{P}_{\mathbf{PVMMB}}$ is defined as a mixture model consistintg of $K$ peaks $\mathbf{P}_{\mathbf{PV}}(x_n|\mu,\sigma,\eta)$ of the pseudo-Voigt model and two background components $\mathbf{P}_{\text{uni}}(x_n;x_1,x_N)$ and $\mathbf{P}_{\text{tri}}(x_n;x_1,x_N)$. Here, $\mu$ is the mean; $\sigma$ is the standard deviation ($0<\sigma$); and $\eta$ is the mixing parameter of the Lorentz and Gauss distributions ($0\le\eta\le1$). Thus, our fitting model consists of $K+2$ components (i.e. $M=K+2$). Details of the fitting model are given in Appendix A.

### 2.2. Prior and posterior probability distribution

A posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{w})$ is required for MAP estimation. The posterior distribution has the following relationship with the likelihood function $p(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{\theta})$, which is the same as $L$ (Equation 4), and prior distribution $r(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{w}) \propto p(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{\theta})r(\boldsymbol{\theta}). \quad (7)$$

For ease of calculation, we take the logarithm of Equation (7) as follows:

$$\ln(p(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{w})) = \ln(p(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{\theta})) + \ln(r(\boldsymbol{\theta})) + \text{const.} \quad (8)$$

In this study, $\mathbf{P}_{\mathbf{PVMMB}}(x_n|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\eta},\boldsymbol{\lambda})$ is used as the fitting model. Here, $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_K)$ is the mixture ratio for each component ($0\le\lambda_k\le1$ and $\sum_{k=1}^{K}\lambda_k=1$) (see also Appendix A.1). We set the prior distribution for the parameters of $k$-th component as follows:

$$r(\mu_k) = \frac{\text{Gauss}(\mu_k;\chi_k,\psi_k)}{\text{CDF}_{\text{Gauss}}(x_{\max};\chi_k,\psi_k) - \text{CDF}_{\text{Gauss}}(x_{\min};\chi_k,\psi_k)}, \quad (9)$$

$$r(\sigma_k) = \text{Gamma}(\sigma_k;\tau_k,\upsilon_k), \quad (10)$$

$$r(\eta_k) = \text{Beta}(\eta_k;\alpha_k,\beta_k), \quad (11)$$

and

$$r(\lambda_k) = \text{Dir}(\lambda_k|\boldsymbol{\phi}_k) = \text{C}(\boldsymbol{\phi}_k)\prod_{k=1}^{K+2}\lambda_k^{\phi_k--1}, \quad (12)$$

where $\text{CDF}_{\text{Gauss}}$, Gamma, Beta and Dir are the cumulative Gaussian, Gamma, Beta, and Dirichlet distributions, respectively; $\chi$ is the mean ($x_{min}\le\chi\le x_{max}$); $\psi$ is the standard deviation ($0<\psi$); $\upsilon$ is the scale parameter ($0<\upsilon$); $\alpha$, $\beta$, and $\tau$ are the shape parameters ($0<\alpha$, $0<\beta$, and $0<\tau$); $\boldsymbol{\phi}$ is the concentration parameter ($0<\boldsymbol{\phi}$); and $C(\boldsymbol{\phi}_k)$ is the normalizing constant in the Dirichlet distribution. As the values of $\mu$ are constrained to the range of the measurement steps ($x_{min}\le\mu\le x_{max}$), we apply a truncated Gaussian distribution (Equation 9). Since the applied prior distribution of $\sigma$ and $\eta$ (Equations 10 and 11) satisfies the defined domain (i.e. $0<\sigma$ and $0\le\eta\le1$), it is not necessary to use truncated distributions. The Dirichlet distribution is appropriate for $\lambda$ (Equation 12), as it satisfies the defined domain ($0\le\lambda\le1$ and $\sum_{k=1}^{K+2}\lambda_k=1$).

The proposed method extracts the appropriate peaks using the Dirichlet distribution as the prior distribution of $\lambda$. We recommend setting the concentration parameter $\boldsymbol{\phi}_k$ to $\le1$ to effectively remove excess peaks. When $\boldsymbol{\phi}_k=1$, the Dirichlet distribution becomes uniform. The rule for updating $\lambda$, which is the same as used by Corduneanu [19], allows for the removal of excess components. When $\boldsymbol{\phi}_k<1$, the Dirichlet distribution has high probabilities at the edges, resulting in some $\lambda_k$ values being close to 0, effectively removing them from the model. In contrast, when $\boldsymbol{\phi}_k>1$, the Dirichlet distribution has high probabilities near the centre, leading to the convergence of $\lambda_k$ values away from 0, and the peaks cannot be selected by setting the $\lambda_k$ values of the excess peaks to close to 0. The parameters of the other prior distributions also affect the estimation of the corresponding parameters. In this study, we set the prior distribution with a broad distribution, thereby reducing the effect on the estimation to $\mu$, $\sigma$ and $\eta$.

The original objective function to maximize is the log posterior (Equation 8). The first term on the right-hand side of Equation 8 (i.e. Equation 4) is maximized when Equation 6 is maximized so that it can be replaced by Equation 6. The second term ($\ln(r(\boldsymbol{\theta}))$) is obtained from Equations 9-12. Put together, the modified objective function can be defined as $E_{MAP}(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{w})$ given in Equation 13.

$$\begin{aligned} &E_{MAP}(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{w}) \\ &= \sum_{n=1}^{N}\{w_n \ln\{\sum_{k=1}^{K}\lambda_k\mathbf{P}_{\mathbf{PV}}(x_n|\theta_k) \\ &\quad + \lambda_{K+1}\mathbf{P}_{\text{uni}}(x_n;x_1,x_N) + \lambda_{K+2}\mathbf{P}_{\text{tri}}(x_n;x_1,x_N)\}\} \\ &\quad + \sum_{k=1}^{K}\{\ln\{\text{truncatedGauss}(\mu_k;\chi_k,\psi_k) \\ &\quad + \ln\{\text{Gamma}(\sigma_k;\tau_k,\upsilon_k) \\ &\quad + \ln\{\text{Beta}(\eta_k;\alpha_k,\beta_k) + \ln\{\text{Dir}(\lambda_k|\boldsymbol{\phi}_k)\}, \end{aligned} \quad (13)$$

where $\boldsymbol{\theta}$ is a vector of the parameters. The parameters for peak fitting are obtained by maximizing Equation (13). As the parameters of the prior distributions are fixed before the calculation, the prior distributions can be interpreted as the regularisation term of the likelihood function in the optimisation of the model.

## 3. Peak fitting

### 3.1. Deterministic annealing

Deterministic annealing is a method for solving an optimisation problem that avoids parameters from becoming trapped in local optima [21–23]. A temperature parameter ($T$) is introduced to the likelihood function or posterior distribution, and the optimisation is conducted by changing the value of $T$ from $T_1$ ($1 < T_1$) to $T_H = 1$. Deterministic annealing has been used to estimate the maximum likelihood of mixture models [24,25] and neural networks [26–28]. Katahira et al. [29,30] applied deterministic annealing to variational Bayesian inference. Figure 1 provides a schematic illustration of peak fitting using deterministic annealing. Peak fitting is performed for each value of $w_n/T_h$, and the values of the parameters at $T_H = 1$ are adopted as the solution. In the supporting information, we have demonstrated the sensitivity of hyperparameters to peak fitting through numerical experiments. Tuning the hyperparameters is crucial for reducing computational costs and efficiency analyzing large amounts of spectral data.

### 3.2. Spectrum-adapted ECM algorithm

We use the spectrum-adapted ECM algorithm to optimise the parameters. The spectrum-adapted ECM algorithm conducts calculations by iteration between E- and CM-step, and has the advantages of numerical stability and low computation cost.

First, the initial value of the temperature parameter $T_1$ and the parameters ($\boldsymbol{\mu}_0$, $\boldsymbol{\sigma}_0$, $\boldsymbol{\eta}_0$, and $\boldsymbol{\lambda}_0$) are set. $E_{MAP}(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{w})$ is calculated using these values and Equation (13).

In E-step, the following responsibilities ($\gamma_{z_{nk}}$, $\gamma_{z_{nK+1}}$, and $\gamma_{z_{nK+2}}$) are calculated using the current parameters ($\boldsymbol{\mu}^{old}$, $\boldsymbol{\sigma}^{old}$, $\boldsymbol{\eta}^{old}$, $\boldsymbol{\lambda}^{old}$), with the initial values ($\boldsymbol{\mu}_0$, $\boldsymbol{\sigma}_0$, $\boldsymbol{\eta}_0$, $\boldsymbol{\lambda}_0$) being used in the first iteration:

$$\gamma_{z_{nk}} = \frac{\lambda_k^{old}\mathbf{P}_{\mathbf{PV}}(x_n|\mu_k^{old}, \sigma_k^{old}, \eta_k^{old})}{\sum_{k=1}^{K}\lambda_k\mathbf{P}_{\mathbf{PV}}(x_n|\mu_k^{old}, \sigma_k^{old}, \eta_k^{old}) + \boldsymbol{\lambda}_{K+1}^{old}\mathbf{P}_{\mathrm{uni}}(x_n; x_1, x_N) + \boldsymbol{\lambda}_{K+2}^{old}\mathbf{P}_{\mathrm{tri}}(x_n; x_1, x_N)}, \tag{14}$$

$$\gamma_{z_{nK+1}} = \frac{\lambda_{K+1}^{old}\mathbf{P}_{\mathrm{uni}}(x_n; x_1, x_N)}{\sum_{k=1}^{K}\lambda_k^{old}\mathbf{P}_{\mathbf{PV}}(x_n|\mu_k^{old}, \sigma_k^{old}, \eta_k^{old}) + \lambda_{K+1}^{old}\mathbf{P}_{\mathrm{uni}}(x_n; x_1, x_N) + \lambda_{K+2}^{old}\mathbf{P}_{\mathrm{tri}}(x_n; x_1, x_N)}, \tag{15}$$

and

$$\gamma_{z_{nK+2}} = \frac{\lambda_{K+2}^{old}\mathbf{P}_{\mathrm{tri}}(x_n; x_1, x_N)}{\sum_{k=1}^{K}\lambda_k^{old}\mathbf{P}_{\mathbf{PV}}(x_n|\mu_k^{old}, \sigma_k^{old}, \eta_k^{old}) + \lambda_{K+1}^{old}\mathbf{P}_{\mathrm{uni}}(x_n; x_1, x_N) + \lambda_{K+2}^{old}\mathbf{P}_{\mathrm{tri}}(x_n; x_1, x_N)}. \tag{16}$$

Equations (14)–(16) represent the peak, baseline, and monotonically increasing background trend, respectively. $K$ is the number of peaks and a positive integer. As our fitting model consists of $K$ peaks and two background components, the total number of components is $K + 2$. $\gamma_{z_{nk}}$ is the posterior probability of a latent variable $z_{nk}$ for a given $x_n$ [20]. In this study, $x_n$ is assumed to be generated from one of the components of the fitting model (see Appendix A). $z_{nk}$ represents the component generated: $z_{nk}$ is equal to 1 when $x_n$ is generated from the $k$-th component; otherwise, $z_{nk}$ is equal to 0. For theoretical details, see Bishop [20] and McLachlan [31].

In CM-step for $\lambda$, $\lambda$ is updated to $\lambda^{new}$ by solving $\frac{dE_{MAP}}{d\lambda} = 0$. The following equations calculate $\lambda_k^{new}$, $\lambda_{K+1}^{new}$ and $\lambda_{K+2}^{new}$, using $\gamma_{z_{nk}}$, $\gamma_{z_{nK+1}}$ and $\gamma_{z_{nK+2}}$:

$$\lambda_k^{new} = \frac{\phi_k - 1 + \sum_{n=1}^{N} w_n\gamma_{z_{nk}}}{M(\phi_k - 1) + \sum_{k=1}^{K+2}\sum_{n=1}^{N} w_n\gamma_{z_{nk}}}, \tag{17}$$

$$\lambda_{K+1}^{new} = \frac{\phi_{K+1} - 1 + \sum_{n=1}^{N} w_n\gamma_{z_{nK+2}}}{M(\phi_{K+1} - 1) + \sum_{k=1}^{K+2}\sum_{n=1}^{N} w_n\gamma_{z_{nk}}}, \tag{18}$$

and

$$\lambda_{K+2}^{new} = \frac{\phi_{K+2} - 1 + \sum_{n=1}^{N} w_n\gamma_{z_{nK+2}}}{M(\phi_{K+2} - 1) + \sum_{k=1}^{K+2}\sum_{n=1}^{N} w_n\gamma_{z_{nk}}}, \tag{19}$$

where $M$ is the total number of components including the peaks and background ($K + 2$ in this study). In CM-step for the other parameters, $\mu_k$ is updated from $\mu_k^{old}$ to $\mu_k^{new}$ by maximisation of the following function:

$$Q(\boldsymbol{x}, \boldsymbol{w}, T|\boldsymbol{\theta}) = \sum_{n=1}^{N}\sum_{k=1}^{K}\{\frac{w_n}{T}\gamma_{z_{nk}}(\ln\lambda_k + \ln\mathbf{P}_{\mathbf{PV}}(x_n|\mu_k, \sigma_k, \eta_k))\}$$

$$+ \sum_{k=1}^{K}\{\ln\{\mathrm{Gauss}(\mu_k; \chi_k, \psi_k\} + \ln\{\mathrm{Gamma}(\sigma_k; \tau_k, \upsilon_k\}$$

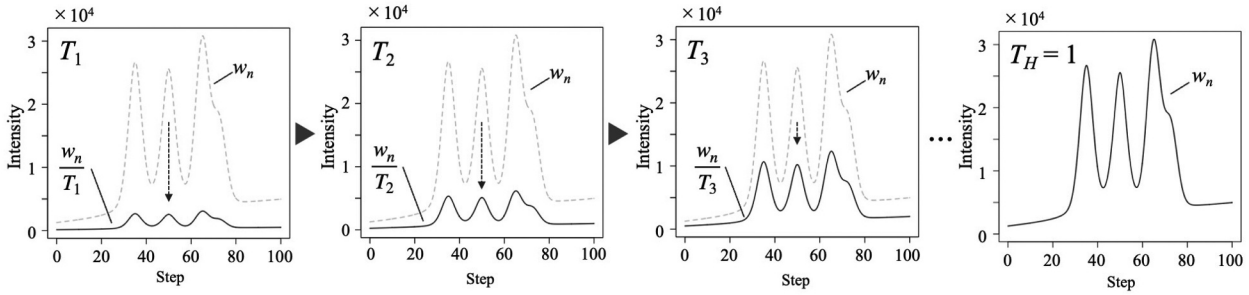



**Figure 1.** Peak fitting with deterministic annealing. Peak fitting is performed at each value of $w_n/T_h$. By decreasing $T_h$ from $T_1$ to $T_H = 1$, peak fitting is eventually performed on the original intensities ($w_n$).

$$+ \ln\{\mathrm{Beta}(\eta_k; \alpha_k, \beta_k\} + \ln\{\mathrm{Dir}(\lambda_k|\boldsymbol{\phi}_k\}\}, \qquad (20)$$

where $\lambda$, $\sigma$, and $\eta$ are fixed to be $\lambda^{new}$, $\sigma^{old}$, and $\eta^{old}$, respectively, with the initial values used in the first iteration. The maximisation is conducted using Brent's method [32]. In CM-step for $\sigma$, $\sigma_k$ is updated from $\sigma_k^{old}$ to $\sigma_k^{new}$ by maximizing Equation (20), and $\lambda$, $\mu$, and $\eta$ are fixed to be $\lambda^{new}$, $\mu^{new}$, and $\eta^{old}$, respectively. In CM-step for $\eta$, $\eta_k$ is updated from $\eta_k^{old}$ to $\eta_k^{new}$ by maximizing Equation (20), and $\lambda$, $\mu$, and $\sigma$ are fixed to be $\lambda^{new}$, $\mu^{new}$, and $\sigma^{new}$, respectively.

The iteration between E-step and CM-step is stopped when the convergence criteria is satisfied; otherwise, the algorithm returns to E-step. When the iteration converges at $T = 1$, the calculation is complete and the current parameters are adopted as the solution. Otherwise $T \neq 1$, the current temperature parameter $T_h$ is updated to the next temperature parameter $T_{h+1}$ and the current parameters are used as the initial values when returning to E-step.

## 4. Analysis of synthetic data

### 4.1. Generation of synthetic data

We applied the spectrum-adapted MAP-ECM algorithm to synthetic spectral data analysis. We define the synthetic spectral data **D** as follows:

$$\mathbf{D} = \begin{pmatrix} x_1^* & \cdots & x_n^* & \cdots & x_N^* \\ w_1^* & \cdots & w_n^* & \cdots & w_N^* \end{pmatrix}, \qquad (21)$$

where $x_n^*$ is the measurement energy step and $w_n^*$ is the intensity of the spectrum at the corresponding energy. We generated a synthetic pseudo-Voigt mixture model for MAP estimation. As the intensity in XPS analysis is the number of photoelectrons measured at each measurement step, we assume that $w_n^*$ are randomly generated from a Poisson distribution ($\mathrm{Poisson}(X|\xi^*)$), where $X$ is a discrete random variable integer starting from 0 (0, 1, 2, …) and $\xi^*$ is the mean of the Poisson distribution. We defined $\xi^*$ corresponding to $w_n^*$ as follows:

$$\mathrm{Poisson}(w_n^*|\xi^*) = \frac{\xi^{*w_n^*}\exp(-\xi^*)}{w_n^*!}. \qquad (22)$$

We use a pseudo-Voigt mixture model with four peak components and a linear background, $\xi^*$, which is expressed as follows:

$$\xi^* = \{\sum_{k=1}^{4}\{\lambda_k^* \mathbf{P}_{\mathbf{PV}}(x_n^*|\mu_k^*, \sigma_k^*, \eta_k^*)\} + \mathbf{B}\} \times 10^d, \qquad (23)$$

where $d$ is a constant ($d \geq 0$) related to the measurement time. The background component (**B**) is expressed as follows:

$$\mathbf{B} = \lambda_{k+1}^*\{\frac{ax_n^* + b}{\sum_{n=1}^{N}(ax_n^* + b)}\}, \qquad (24)$$

where $a$ and $b$ represent the slope and intercept of the background, respectively, and $x_n$ was selected from between 0 and 100 in steps of 0.5, so that the total number of steps was 201. The parameters of the peak components were as follows: $\mu_k^* = \{35, 50, 65, 72.5\}$, $\sigma_k^* = \{3, 3, 3, 3\}$, $\eta_k^* = \{0.4, 0.6, 0.1, 0.1\}$, and $\lambda_k^* = \{0.2, 0.2, 0.2, 0.2, 0.1\}$. We set $a = 500$, $b = 15000$, and $\lambda_{k+1}^* = 0.3$ for the background component. The synthetic spectral data were generated with $d$ set to 7 and 4. The synthetic data when $d = 7$ represent high-resolution spectral data with a large number of counts, and the synthetic data when $d = 4$ represent noisy, low-intensity spectral data.

### 4.2. Initial conditions of synthetic data analysis

We applied the proposed method to synthetic data. In the calculation, the initial number of peaks was set to $K = 9$ so that the fitting model included redundant peaks. The initial values of $\mu$, $\sigma$, and $\eta$ were generated randomly from each prior distribution, and the initial $\lambda$ values were 1/11, as the total number of components $M$ is 11. The parameters of the prior distribution are listed in Table 1.

The temperature parameter set $T_h$ was prepared as follows:

**Table 1.** Parameters of the prior distribution for the synthetic data analysis.

| Parameter | Value |
|---|---|
| $\chi_1$ | 20.0 |
| $\chi_2$ | 27.5 |
| $\chi_3$ | 35.0 |
| $\chi_4$ | 42.5 |
| $\chi_5$ | 50.0 |
| $\chi_6$ | 57.5 |
| $\chi_7$ | 65.0 |
| $\chi_8$ | 72.5 |
| $\chi_9$ | 80.0 |
| $\psi_1$ | 10 |
| $\psi_2$ | 10 |
| $\psi_3$ | 10 |
| $\psi_4$ | 10 |
| $\psi_5$ | 10 |
| $\psi_6$ | 10 |
| $\psi_7$ | 10 |
| $\psi_8$ | 10 |
| $\psi_9$ | 10 |
| $\tau_1$ | 10 |
| $\tau_2$ | 10 |
| $\tau_3$ | 10 |
| $\tau_4$ | 10 |
| $\tau_5$ | 10 |
| $\tau_6$ | 10 |
| $\tau_7$ | 10 |
| $\tau_8$ | 10 |
| $\tau_9$ | 10 |
| $\upsilon_1$ | 5 |
| $\upsilon_2$ | 5 |
| $\upsilon_3$ | 5 |
| $\upsilon_4$ | 5 |
| $\upsilon_5$ | 5 |
| $\upsilon_6$ | 5 |
| $\upsilon_7$ | 5 |
| $\upsilon_8$ | 5 |
| $\upsilon_9$ | 5 |
| $\alpha_1$ | 7 |
| $\alpha_2$ | 7 |
| $\alpha_3$ | 7 |
| $\alpha_4$ | 7 |
| $\alpha_5$ | 7 |
| $\alpha_6$ | 7 |
| $\alpha_7$ | 7 |
| $\alpha_8$ | 7 |
| $\alpha_9$ | 7 |
| $\beta_1$ | 7 |
| $\beta_2$ | 7 |
| $\beta_3$ | 7 |
| $\beta_4$ | 7 |
| $\beta_5$ | 7 |
| $\beta_6$ | 7 |
| $\beta_7$ | 7 |
| $\beta_8$ | 7 |
| $\beta_9$ | 7 |
| $\phi_1$ | 0.01 |
| $\phi_2$ | 0.01 |
| $\phi_3$ | 0.01 |
| $\phi_4$ | 0.01 |
| $\phi_5$ | 0.01 |
| $\phi_6$ | 0.01 |
| $\phi_7$ | 0.01 |
| $\phi_8$ | 0.01 |
| $\phi_9$ | 0.01 |
| $\phi_{10}$ | 0.01 |
| $\phi_{11}$ | 0.01 |

$$T_h = \frac{\sum_{n=1}^{N} w_n}{10^{t_h}}, \tag{25}$$

where

$$t_h = t_0 + \frac{(h-1)\{\log_{10}(\sum_{n=1}^{N} w_n) - t_0\}}{H-1}, \tag{26}$$

$$t_0 = 3, \tag{27}$$

and $H$ is the total number of temperature parameters. We set $H = 5$ ($h = \{1, 2, 3, 4, 5\}$). These values were heuristically selected. A theoretical analysis to find the appropriate values is an ongoing problem.

The iterative calculation is considered to have converged when the difference between the values of Equation (20) before and after the CM-steps is $< 10^{-8}$. The calculation is terminated after 50,000 iterations. A solution is obtained when the iterative calculations are complete at $T = 1$. The calculation was conducted using code written in R (http://cran.r-project.org/), an open-source programming language and software environment for statistical analysis. The computer running the calculations had an Intel Core m7 processor, dual cores running at 1.3 GHz, and 8 GB memory.

### 4.3. Results of the synthetic data analysis

The estimated fitting curve shows a good fit to the synthetic spectral data and contains the correct number of peaks ($K = 4$), including a minor peak ($\mu = 72.5$) near the major peak ($\mu = 65$). Figure 2 shows the fitting curves using the initial and estimated parameters. The proposed method extracts a four-component pseudo-Voigt mixture model from the initial nine-component model.

Figure 3 shows the variation in each parameter with each iteration. Each parameter converged within $< 400$ seconds. Five mixture ratios ($\lambda_1, \lambda_3, \lambda_4, \lambda_6, \lambda_9$) converged to 0, and the other six mixture ratios were correctly assigned to four peak components and two background components. This results show that the proposed method extracts appropriate peaks by assigning a value of 0 to the mixture ratios $\lambda$ of the excess peaks. In addition, the parameters of the removed peaks converged to the mode of the prior distribution (Table 2).

Figure 4 shows a comparison of three related methods: (1) the proposed method, (2) the proposed method without deterministic annealing, and (3) the spectrum-adapted ECM algorithm [8]. The calculation times of the three methods were 285.1 seconds (5816 iterations), 3251.7 seconds (50,000 iterations), and 2268.0 seconds (50,000 iterations), respectively. The calculations were terminated after 50,000 iterations. A large calculation cost was required to satisfy the convergence criteria using the methods that did not use deterministic annealing. This suggests that the proposed method can more efficiently produce a suitable fitting curve. The proposed method performed appropriate peak fitting by removing excess peaks; however, the proposed method without deterministic annealing produced a curve with two components for the left and centre peaks. Similarly, the curve produced

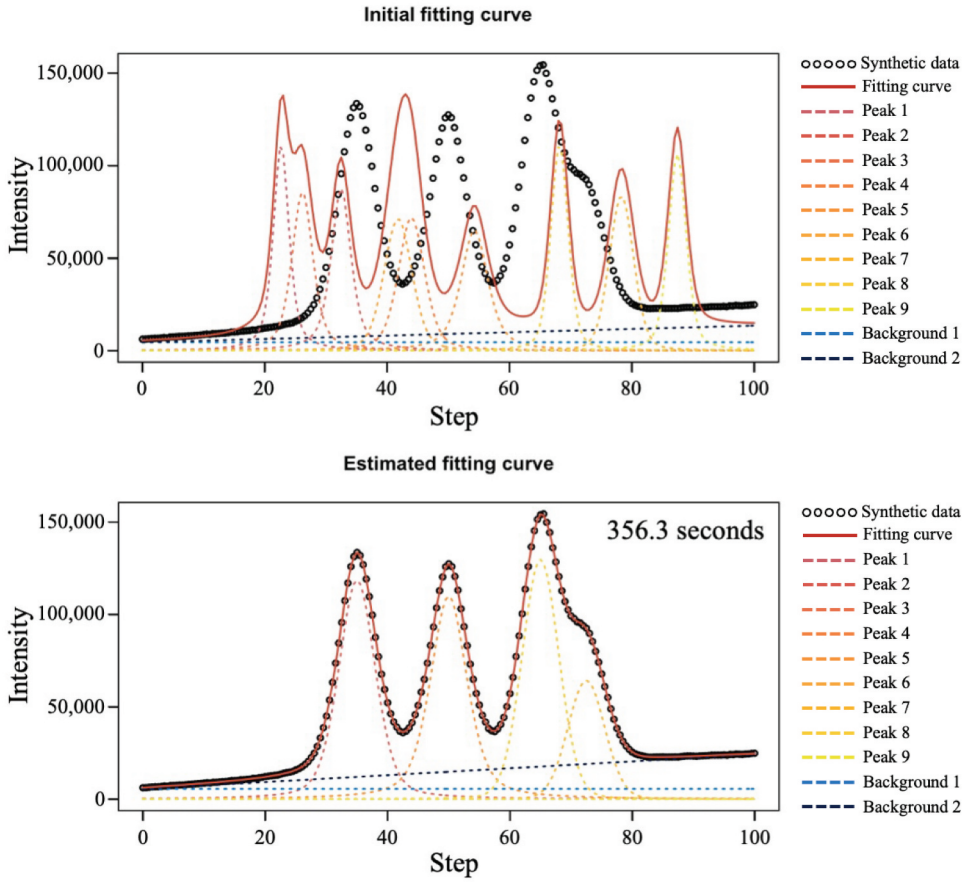



**Figure 2.** (Upper panel) Initial fitting curves and (lower panel) estimated fitting curves. The proposed method successfully produced a fitting curve with an appropriate number of peaks ($K = 4$).

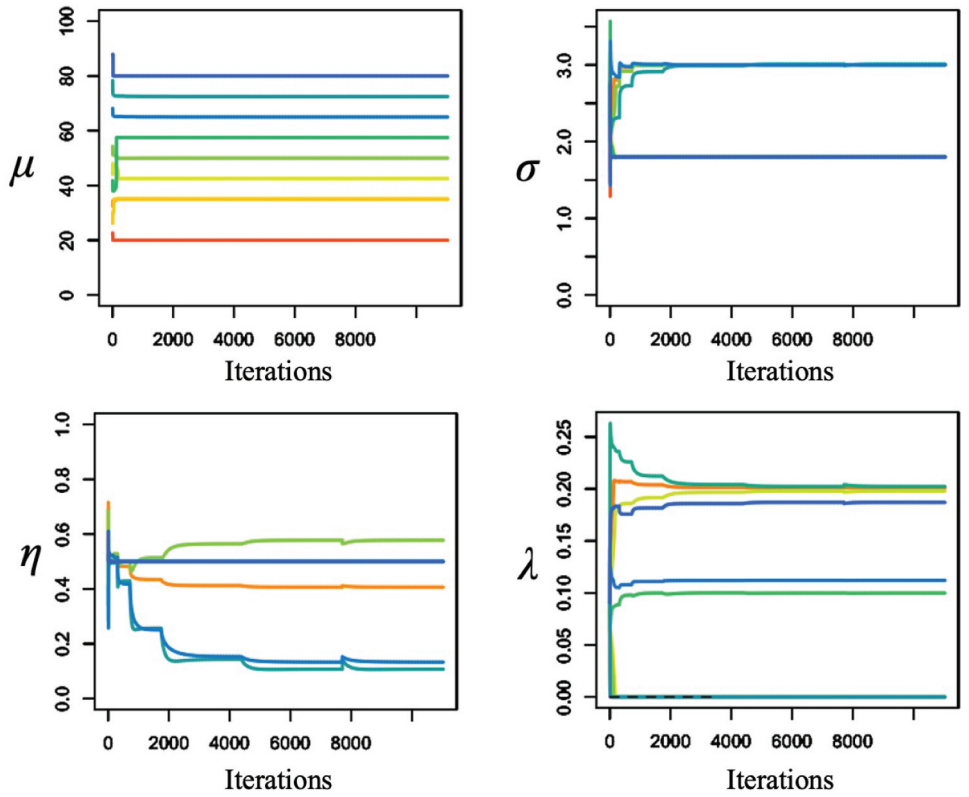



**Figure 3.** Variation in each parameter during the iterative calculations. The mixture ratio ($\lambda$) of the redundant peak components is reduced to 0.

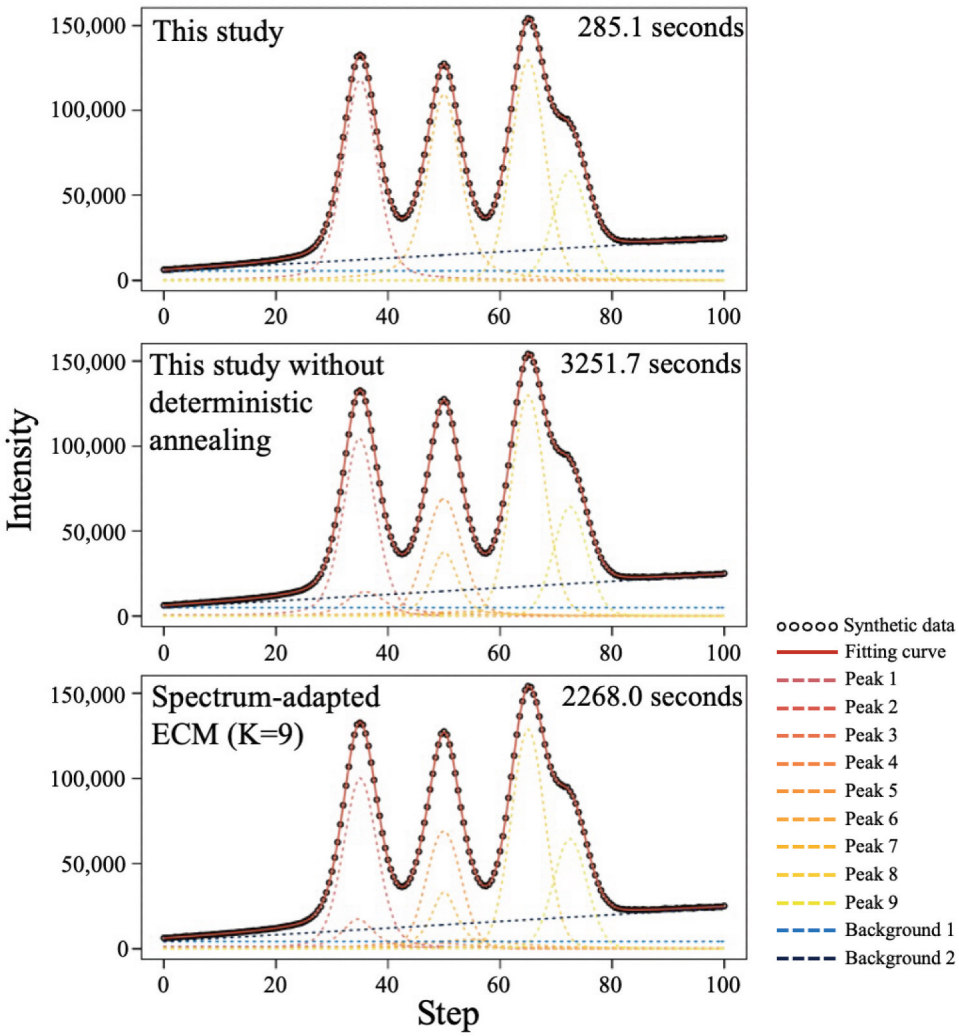



**Figure 4.** Fitting curves obtained using (upper panel) the proposed method, (middle panel) the proposed method without deterministic annealing, and (lower panel) the spectrum-adapted ECM algorithm. The proposed method determined the appropriate number of peaks and performed more efficiently than the other methods.

by the spectrum-adapted ECM algorithm includes excess peaks; therefore, individual peaks cannot be decomposed adequately.

Figure 5 shows that the proposed method extracted a four-component pseudo-Voigt mixture model from an initial nine-component pseudo-Voigt mixture model even from spectral data with noisy peaks due to low signal intensity. The obtained fitting curve comprised three major peaks and one minor peak, correctly estimating the true peak shape. The estimated parameters were similar to those from higher-intensity spectral data (Table 2). In addition, the calculation time was shorter than that required to analyse high-intensity spectral data; therefore, the proposed method can perform peak fitting robustly and efficiently even with low-intensity spectral data. This calculation behavior is due to the sensitivity of the objective function to changes in parameters depending on the intensity of the spectra. This study observes that when the intensity of the spectra is high, more iterations are required to satisfy the convergence criteria. This is because the objective function becomes sensitive to changes in parameters on the peak fitting of high-intensity spectra. Even slight changes in the parameters cause significant fluctuations in the objective function, necessitating more computational time to satisfy the convergence criteria compared to when analyzing noisy spectra with lower intensity.

## 5. Analysis of experimental data

### 5.1. Experimental data acquisition

Experimental data were acquired by XPS mapping the surface of a TFET, as undertaken by Nakamura et al. [33]. The TFET data were acquired using a 3D nanoESCA core-level photoelectron microspectrometer [34,35] at the BL07LSU soft X-ray beamline at the SPring-8 synchrotron radiation facility. The photon energy of the synchrotron radiation used for the analysis was 1000 eV. 3D nanoESCA can scan a sample with high lateral spatial resolution (~ 70 nm) and record photoelectron spectra for quantitative analysis of the

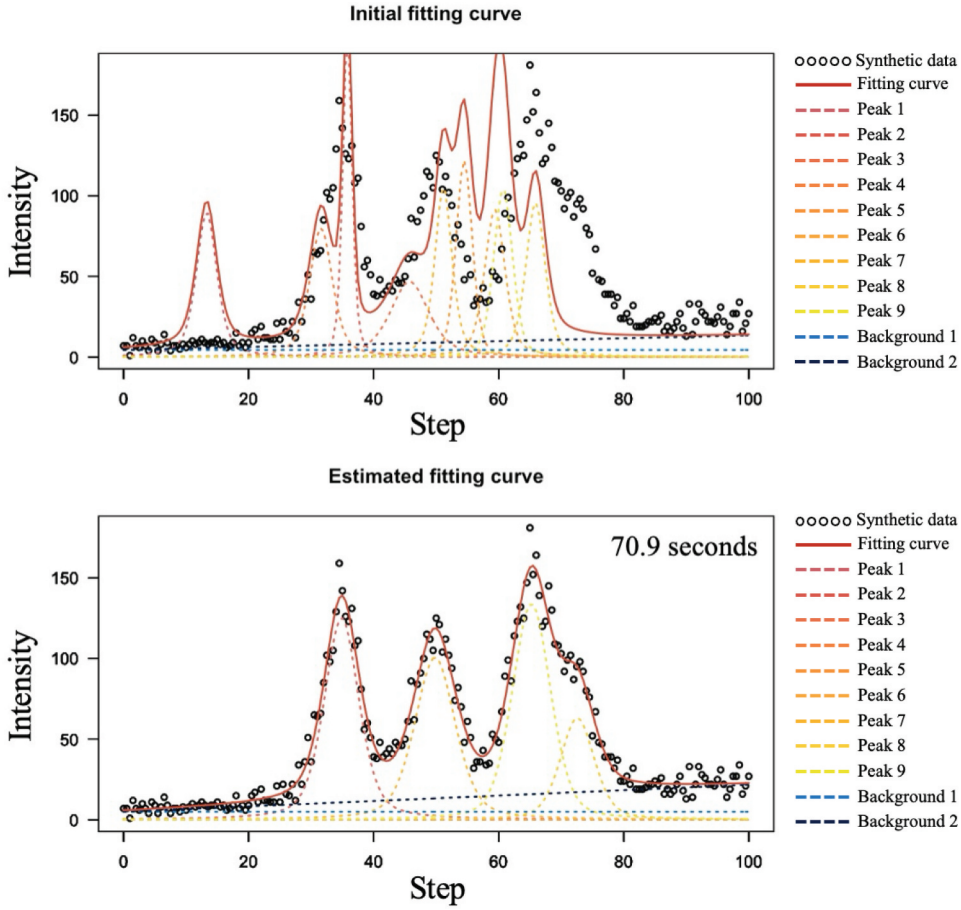



**Figure 5.** (Upper panel) Initial fitting curves and (lower panel) estimated fitting curves for noisy spectral data. The proposed method produced a fitting curve with no excess peaks in the analysis of noisy spectral data.

chemical state. The TFET data consist of 7396 spectra from 86 × 86 measurement points. Figure 6 shows a map of the total intensity of the spectra. More details about the TFET are given by Nakamura et al. [33].

### 5.2. Initial conditions

The number of peaks $K$ in the fitting model was set to nine, yielding a fitting model with redundant peak components. We set the initial values of $\mu$ over equal intervals within the range of binding energy of the spectra: $\{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9\} = \{222.000, 224.875, 227.750, 230.625, 233.500, 236.375, 239.250, 242.125, 245.000\}$. The initial $\sigma$ and $\eta$ values were generated randomly from each prior distribution. Each initial $\lambda$ is 1/11 because the total number of components including the background components $M$ is 11. The prior distribution parameters were set heuristically and are listed in Table 3. These prior distributions provided minor constraints on the possible values of the parameters and allowed the convergence of each to a reasonable value. To use deterministic annealing, a set of temperature parameters $T_h$ was prepared using the same procedure as with the synthetic data (Equations 21 and 22), with $t_0 = 2.25$ and $H = 5$ ($h = \{1, 2, 3, 4, 5\}$).

### 5.3. Results of the experimental data analysis

Figure 7 shows typical fitting curves obtained using the proposed method. Seven peak components were removed from the fitting model as the mixture ratio converged to 0. The other two peak components remained in the fitting model, and the background on the low energy side was linear (Figure 7a–d). These results show that the spectral data from the $MoS_2$ sheet were successfully decomposed into a Mo $3d_{3/2}$ peak, $3d_{5/2}$ peak, and background components. Although the S 2*s* peak should occur near 227 eV, it was not detected due to the noise level of the spectra. If the signal-to-noise ratio of the measurements can be improved, this peak should be detected. In contrast, no Mo $3d_{3/2}$ or $3d_{5/2}$ peaks were detected in the spectral data on the $p^+$-$WSe_2$ sheet (Figure 7e) and hexagonal boron nitride (*h*-BN; Figure 7f). These fitting curves consist predominantly of the background component. As the spectral data from the $p^+$-$WSe_2$ sheet and *h*-BN have no clear peaks within the measured range of binding energies, fitting curves consisting predominantly of background components are realistic. These fitting curves (Figure 7a–f) were obtained using the same initial conditions. Thus, appropriate fitting curves are obtained without tuning the initial

**Table 2.** Summary of the estimated values of the parameters for the synthetic data analysis.

| Parameter | High-intensity spectra | Noisy spectra |
|---|---|---|
| $\mu_1$ | 20.00 | 20.00 |
| $\mu_2$ | 35.00 | 34.88 |
| $\mu_3$ | 35.00 | 35.00 |
| $\mu_4$ | 42.50 | 42.50 |
| $\mu_5$ | 49.99 | 50.00 |
| $\mu_6$ | 57.50 | 57.50 |
| $\mu_7$ | 72.49 | 49.82 |
| $\mu_8$ | 65.00 | 72.59 |
| $\mu_9$ | 80.00 | 65.27 |
| $\sigma_1$ | 1.80 | 1.80 |
| $\sigma_2$ | 3.00 | 2.79 |
| $\sigma_3$ | 1.80 | 1.80 |
| $\sigma_4$ | 1.80 | 1.80 |
| $\sigma_5$ | 3.00 | 1.80 |
| $\sigma_6$ | 1.80 | 1.80 |
| $\sigma_7$ | 3.01 | 3.25 |
| $\sigma_8$ | 3.00 | 2.74 |
| $\sigma_9$ | 1.80 | 3.08 |
| $\eta_1$ | 0.50 | 0.50 |
| $\eta_2$ | 0.41 | 0.53 |
| $\eta_3$ | 0.50 | 0.50 |
| $\eta_4$ | 0.50 | 0.50 |
| $\eta_5$ | 0.58 | 0.50 |
| $\eta_6$ | 0.50 | 0.50 |
| $\eta_7$ | 0.11 | 0.49 |
| $\eta_8$ | 0.13 | 0.47 |
| $\eta_9$ | 0.50 | 0.38 |
| $\lambda_1$ | 0.00 | 0.00 |
| $\lambda_2$ | 0.20 | 0.21 |
| $\lambda_3$ | 0.00 | 0.00 |
| $\lambda_4$ | 0.00 | 0.00 |
| $\lambda_5$ | 0.20 | 0.00 |
| $\lambda_6$ | 0.00 | 0.00 |
| $\lambda_7$ | 0.10 | 0.19 |
| $\lambda_8$ | 0.20 | 0.10 |
| $\lambda_9$ | 0.00 | 0.23 |
| $\lambda_{10}$ | 0.11 | 0.10 |
| $\lambda_{11}$ | 0.19 | 0.17 |

conditions individually, even for spectral data collected from composite materials.

The total time to analyse the 7396 spectra was 271.4 hours, and the median calculation time for individual spectra was 92.44 seconds. Thus, most spectra were analysed within 2 minutes (Figure 7a–e), although 206 spectra took >600 seconds to fit (Figure 7f). Subtracting the background and determining the number of peaks is conducted efficiently using a single calculation using the ECM algorithm. Thus, the proposed method fits the peaks sufficiently quickly for practical use.

We collected the peaks where $229.4 < \mu < 230.0$, $0.10 < \sigma < 0.45$, and $\lambda > 0.06$, and assumed that these peaks correspond to the Mo $3d_{5/2}$ peak. Figure 8 provides a map of the binding energy of the Mo $3d_{5/2}$ peak, which shows that it varies spatially between 229.4 and 230.0 eV (Figure 8). The Mo $3d_{5/2}$ peak on the thick-$MoS_2$ sheet was 0.1–0.2 eV lower than that on the three-layered $MoS_2$. In the region where the thick $MoS_2$ sheet overlaps with the $p^+$-$WSe_2$ sheet, the binding energy of the Mo $3d_{5/2}$ peak was 0.2–0.3 eV lower than in the other areas. These shifts in binding energy with $MoS_2$ thickness were due to differences in charge doping from the $p^+$-$WSe_2$. In addition, some areas have binding energies that are 0.3–0.4 eV lower than those in the surrounding region. These regions of low binding energy are bubbles that were introduced during sample preparation.

## 6. Discussion and implications

This paper has introduced a novel method for the complex task of selecting the optimal number of peaks in peak fitting analyses. We present this new technique and demonstrate its effectiveness using both synthetically generated and real spectral datasets. Our findings confirm that the method accurately identifies the appropriate number of peaks, proving its practical value in real-world experimental settings. Additionally, the versatility of the proposed method allows it to be applied across a broad range of spectral

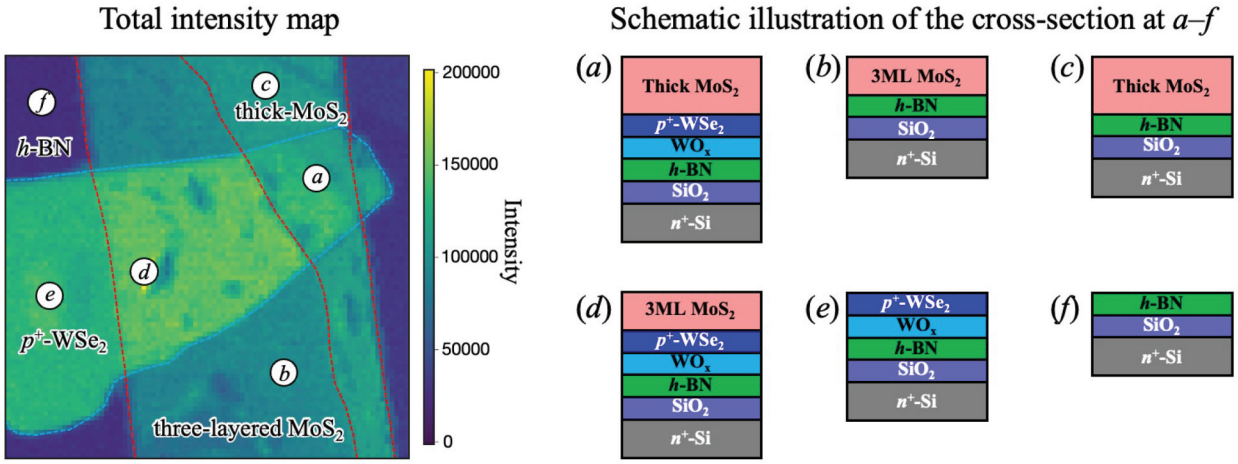



**Figure 6.** (Left) Total spectral intensity map and (right) schematic cross-sections corresponding to positions *a*–*f* on the map. *a*–*f* indicate the locations where the spectral data shown in Figure 7 were collected. The regions on the thick (*a*, *c*) and three-layered (*b*, *d*) $MoS_2$ sheet and $p^+$-$WSe_2$ sheet (*e*) yield higher intensities than the region of *h*-BN (*f*). The red dashed line indicates the boundary between the thick and three-layered $MoS_2$ sheet and the blue dashed line indicates the edge of the $p^+$-$WSe_2$ sheet.

**Table 3.** Parameters of the prior distribution for the experimental data analysis.

| Parameter | Value |
|---|---|
| $\chi_1$ | 222.000 |
| $\chi_2$ | 224.875 |
| $\chi_3$ | 227.750 |
| $\chi_4$ | 230.625 |
| $\chi_5$ | 233.500 |
| $\chi_6$ | 236.375 |
| $\chi_7$ | 239.250 |
| $\chi_8$ | 242.125 |
| $\chi_9$ | 245.000 |
| $\psi_1$ | 5 |
| $\psi_2$ | 5 |
| $\psi_3$ | 5 |
| $\psi_4$ | 5 |
| $\psi_5$ | 5 |
| $\psi_6$ | 5 |
| $\psi_7$ | 5 |
| $\psi_8$ | 5 |
| $\psi_9$ | 5 |
| $\tau_1$ | 10 |
| $\tau_2$ | 10 |
| $\tau_3$ | 10 |
| $\tau_4$ | 10 |
| $\tau_5$ | 10 |
| $\tau_6$ | 10 |
| $\tau_7$ | 10 |
| $\tau_8$ | 10 |
| $\tau_9$ | 10 |
| $\upsilon_1$ | 0.1 |
| $\upsilon_2$ | 0.1 |
| $\upsilon_3$ | 0.1 |
| $\upsilon_4$ | 0.1 |
| $\upsilon_5$ | 0.1 |
| $\upsilon_6$ | 0.1 |
| $\upsilon_7$ | 0.1 |
| $\upsilon_8$ | 0.1 |
| $\upsilon_9$ | 0.1 |
| $\alpha_1$ | 7 |
| $\alpha_2$ | 7 |
| $\alpha_3$ | 7 |
| $\alpha_4$ | 7 |
| $\alpha_5$ | 7 |
| $\alpha_6$ | 7 |
| $\alpha_7$ | 7 |
| $\alpha_8$ | 7 |
| $\alpha_9$ | 7 |
| $\beta_1$ | 7 |
| $\beta_2$ | 7 |
| $\beta_3$ | 7 |
| $\beta_4$ | 7 |
| $\beta_5$ | 7 |
| $\beta_6$ | 7 |
| $\beta_7$ | 7 |
| $\beta_8$ | 7 |
| $\beta_9$ | 7 |
| $\phi_1$ | 0.01 |
| $\phi_2$ | 0.01 |
| $\phi_3$ | 0.01 |
| $\phi_4$ | 0.01 |
| $\phi_5$ | 0.01 |
| $\phi_6$ | 0.01 |
| $\phi_7$ | 0.01 |
| $\phi_8$ | 0.01 |
| $\phi_9$ | 0.01 |
| $\phi_{10}$ | 0.01 |
| $\phi_{11}$ | 0.01 |

types, making it ideal for addressing various analytical challenges in material sciences. The proposed method's broad applicability and practical analysis capabilities make it an invaluable tool for experimental researchers.

The proposed method selects peak components without the conventional trial-and-error approach. The proposed method removes excess peaks using MAP estimation. The computation is performed efficiently using the spectrum-adapted ECM algorithm with deterministic annealing.

Because the proposed method does not use a criterion for evaluating the number of peaks, it is not necessary to conduct the calculation for each number of peaks independently. This reduces the computational cost of repeated analysis for individual spectra and contributes significantly to the efficiency of the analysis, especially in high-throughput analysis where individual analyses must be processed quickly. In contrast, a weakness of the proposed method is that it requires many iterations to converge, as a fitting model with multiple peaks is optimised. This weakness can be improved by deterministic annealing, as shown by our analysis of synthetic data.

Deterministic annealing is essential for performing the proposed method effectively. With deterministic annealing, it is easier to optimise a fitting model with multiple peak components rapidly. Peak fitting can be performed faster and more accurately than when using the conventional spectrum-adapted EM algorithm by setting the temperature parameters appropriately. Hyperparameter tuning has the potential to accelerate the calculation process, increasing the computational speed several to tens of times (Figure S2). Although achieving accurate fitting curves is also possible with ample time, it is impractical for high-throughput analysis without deterministic annealing technique. In this study, hyperparameters are set heuristically; precise adjustment could enhance method efficiency. Systematic exploration of optimal settings is essential for future research. Although the choice of temperature parameters is heuristic, the synthetic and experimental data analyses show promising results. Establishing a tuning procedure for the temperature parameters is essential to further develop a high-throughput peak fitting method.

MAP estimation enables us to employ domain knowledge, including the theoretical peak positions determined by previous research, into the prior distribution. Even when faced with spectral data containing complicated peak shape, it is possible to detect peaks using an appropriate prior distribution. In addition, a broad prior distribution does not affect peak fitting and is almost equivalent to peak fitting using the spectrum-adapted ECM algorithm based on maximum likelihood estimation [7,8].

The conventional workflow of peak fitting involves multiple steps: (1) background subtraction; (2) determining the number of peaks: (3) setting initial conditions for the fitting algorithm; (4) optimizing the fitting model. These tasks often require repetitive effort to achieve an accurate fitting curve. The process, which is manually intensive and not standardized, involves handling each spectrum individually.

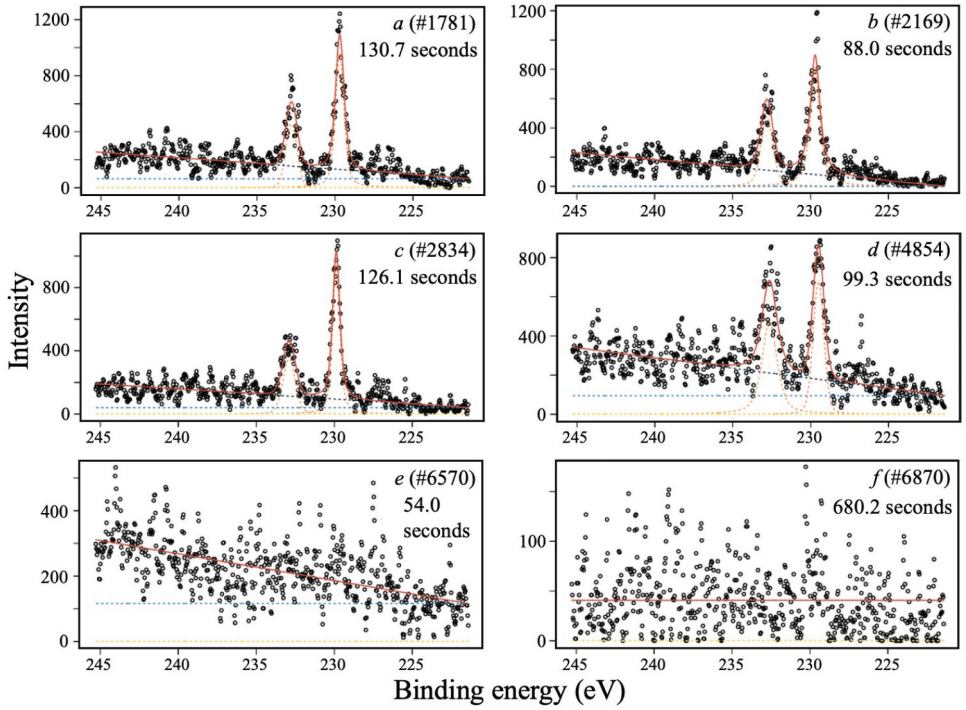



**Figure 7.** Typical fitting curves for the experimental data. *a*–*f* correspond to the locations shown on Figure 6. Clear Mo $3d_{3/2}$ and Mo $3d_{5/2}$ peaks are seen in *a*–*d*. The proposed method successfully selected only these two peaks. In contrast, spectra *e* and *f* have no clear peaks, and a fitting model consisting only of the background was produced.

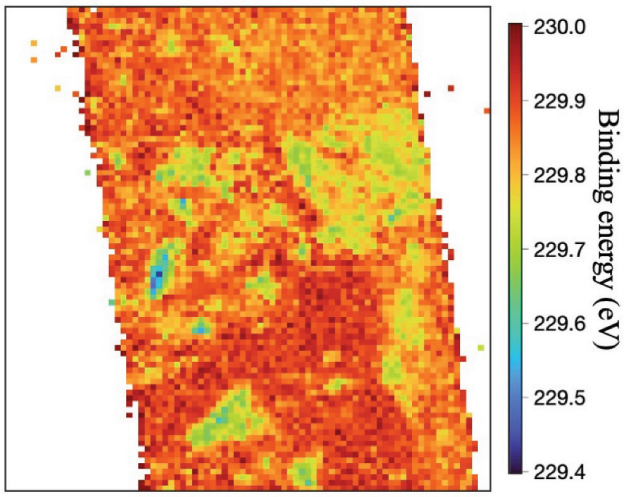



**Figure 8.** Map of the binding energy of the Mo $3d_{5/2}$ peak. As there are no Mo peaks in the spectra collected on the $p^{+}$-$WSe_2$ sheet and *h*-BN, peaks detected in these regions are artifacts of the noisy spectral data.

Analysts, relying heavily on their experience, invest considerable time in finding better solution. Given the many analyzed spectra, this conventional approach is costly and impractical.

The proposed method directly integrates background subtraction and the selection of peak numbers into the peak-fitting calculations and simplifies the workflow by focusing on setting on the initial conditions. Consequently, it reduces analysts' workload and dependency, significantly enhancing the efficiency of the peak fitting. Moreover, the proposed method's versatility allows its application to various types of spectra, such as Raman and XAF, not just the XPS spectra demonstrated in this study. The proposed method contributes to obtaining low-dimensional features, including the peak position and full width at half maximum, extracted from the considerable volumes of spectral data produced.

We are committed to further enhancing our method. We plan to develop optimization techniques for hyperparameters and refine the algorithm to improve computational cost. We also aim to develop graphical user interfaces to enhance usability. Moreover, we recognize that the choice of initial

values affects the computational cost (e.g. Figure 2 bottom and Figure 4 top), and a few initial value sets can take exceptionally long calculation times. Although this is not the focus of this research, it will be an important issue when advancing the proposed method. These advancements are poised to significantly evolve the traditional peak fitting workflow, the potential impact of our research in the material science.

## 7. Conclusions

We introduced the spectrum-adapted MAP – ECM algorithm for high-throughput peak fitting. The method estimates the model's parameters and selects the peaks in a single calculation. It produced a curve that fits synthetic spectral data well, with the correct number of peaks. Furthermore, a similar result was obtained for noisy synthetic spectral data. The method can provide appropriate fitting curves without the conventional trial-and-error approach, even for experimental spectral data from a tunnel field-effect transistor. The method is therefore a valuable tool for high-throughput analysis.

## Acknowledgements

This paper is based on results from a project (P16010) commissioned by the New Energy and Industrial Technology Development Organization (NEDO), the Japan Science and Technology Agency (JST) PRESTO (Precursory Research for Embryonic Science and Technology) (JPMJPR20T7,JPMJPR17NB), JST CREST (Core Research for Evolutional Science and Technology) (JPMJCR1761, JPMJCR21O1, and JPMJCR21O4), and the Research Program for CORE laboratories (2016002) of the Dynamic Alliance for Open Innovation Bridging Human, Environment and Materials of the Network Joint Research Center for Materials and Devices. This work was also supported by JSPS KAKENHI (JP21387393, 21H01638, 21H04696), MEXT KAKENHI (JP20291983) and JST-Mirai Program (JPMJMI21G2) grants. This work was partially supported by the Innovative Science and Technology Initiative for Security of ATLA, Japan (JPJ004596). We thank Prof. Kosuke Nagashio of the University of Tokyo, Prof. Hirokazu Fukidome of Tohoku University, and Dr. Akada of Tsukuba University for acquiring the presented data. The spectral data were obtained with the support of the University of Tokyo outstation beamline at SPring-8 (2018B7580, 2019A7451, 2021A7422, 2018B7575, and 2021B7435).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The work was supported by the Core Research for Evolutional Science and Technology [JPMJCR1761]; JST-Mirai Program [JPMJMI21G2]; Japan Society for the Promotion of Science [21H01638]; New Energy and Industrial Technology Development Organization [P16010]; Research Program for CORE lab [2016002]; Acquisition, Technology & Logistics Agency [JPJ004596]; Precursory Research for Embryonic Science and Technology [JPMJPR17NB].

## ORCID

Tarojiro Matsumura http://orcid.org/0000-0001-9950-509X
Naoka Nagamura http://orcid.org/0000-0002-7697-8983
Shotaro Akaho http://orcid.org/0000-0002-4623-2718
Kenji Nagata http://orcid.org/0000-0001-9894-4461
Yasunobu Ando http://orcid.org/0000-0003-3702-034X

## References

[1] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science. 2015;349 (6245):255–260. doi: 10.1126/science.aaa8415
[2] Mueller T, Kusne AG, Ramprasad R. Machine learning in materials science: recent progress and emerging applications. Rev Comput Chem. 2016;29:186–273.
[3] Ramprasad R, Batra R, Pilania G, et al. Machine learning in materials informatics: recent applications and prospects. Npj Comput Mater. 2017;3(1):1–13. doi: 10.1038/s41524-017-0056-5
[4] Saal JE, Oliynyk AO, Meredig B. Machine learning in materials discovery: confirmed predictions and their underlying approaches. Annu Rev Mater Res. 2020;50 (1):49–69. doi: 10.1146/annurev-matsci-090319-010954
[5] Hong S, Liow CH, Yuk JM, et al. Reducing time to discovery: materials and molecular modeling, imaging, informatics, and integration. ACS Nano. 2021;15(3):3971–3995. doi: 10.1021/acsnano.1c00211
[6] Matsumura T, Nagamura N, Akaho S, et al. Spectrum adapted expectation–maximization algorithm for high–throughput peak shift analysis. Sci Technol Adv Mater. 2019;20(1):733–745. doi: 10.1080/14686996.2019.1620123
[7] Matsumura T, Nagamura N, Akaho S, et al. Spectrum adapted expectation–conditional maximization algorithm for extending high–throughput peak separation method in XPS analysis. Sci Technol Adv Mater: Methods. 2021;1(1):45–55. doi: 10.1080/27660400.2021.1899449
[8] Matsumura T, Nagamura N, Akaho S, et al. High-throughput xps spectrum modeling with autonomous background subtraction for 3d5/2 peak mapping of sns. Sci Technol Adv Mater: Methods. 2023;3 (1):2159753. doi: 10.1080/27660400.2022.2159753
[9] Shinotsuka H, Nagata K, Yoshikawa H, et al. Development of spectral decomposition based on Bayesian information criterion with estimation of confidence interval. Sci Technol Adv Mater. 2020;21 (1):402–419. doi: 10.1080/14686996.2020.1773210
[10] Ueno T, Ishibashi H, Hino H, et al. Automated stopping criterion for spectral measurements with active learning. Npj Comput Mater. 2021;7(1):1–9. doi: 10.1038/s41524-021-00606-5
[11] Murakami R, Kageyama H, Nakamura K, et al. Background estimation in X-ray photoelectron

spectroscopy data using an active Shirley method with automated selection of the analytical range. E-J Surf Sci Nanotechnol. 2019;17:61–68. doi: 10.1380/ejssnt.2019.61

[12] Murakami R, Yoshikawa H, Nagata K, et al. Automatic estimation of unknown chemical components in a mixed material by xps analysis using a genetic algorithm. Sci Technol Adv Mater: Methods. 2022;2(1):91–105. doi: 10.1080/27660400.2022.2061878

[13] Okada M, Nagamura N, Matsumura T, et al. Growth of mos2–nb-doped mos2 lateral homojunctions: a monolayer p–n diode by substitutional doping. APL Mater. 2021;9(12):121115. doi: 10.1063/5.0070333

[14] Nagata K, Sugita S, Okada M. Bayesian spectral deconvolution with the exchange Monte Carlo method. Neural Networks. 2012;28:82–89. doi: 10.1016/j.neunet.2011.12.001

[15] Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6(2):461–464. doi: 10.1214/aos/1176344136

[16] Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B: Stat Methodol. 1996;58(1):267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

[17] Yada A, Nagata K, Ando Y, et al. Machine learning approach for prediction of reaction yield with simulated catalyst parameters. Chem Lett. 2018;47(3):284–287. doi: 10.1246/cl.171130

[18] Yada A, Matsumura T, Ando Y, et al. Ensemble learning approach with lasso for predicting catalytic reaction rates. Synlett. 2021;32(18):1843–1848. doi: 10.1055/a-1304-4878

[19] Corduneanu A, Bishop CM. Variational bayesian model selection for mixture distributions. In: Jaakkola T, Richardson T, editors. Artificial intelligence and statistics. Vol. 2001. Waltham (MA): Morgan Kaufmann; 2001. p. 27–34.

[20] Bishop CM. Pattern recognition and machine learning. New York (NY): Springer; 2006.

[21] Rose K, Gurewitz E, Fox G. A deterministic annealing approach to clustering. Pattern Recognition Letters. 1990;11(9):589–594. doi: 10.1016/0167-8655(90)90010-Y

[22] Rose K, Gurewitz E, Fox GC. Vector quantization by deterministic annealing. IEEE Trans Inf Theory. 1992;38(4):1249–1257. doi: 10.1109/18.144705

[23] Rose K. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. Proc IEEE. 1998;86(11):2210–2239. doi: 10.1109/5.726788

[24] Yuille AL, Kosowsky J. Statistical physics algorithms that converge. Neural Comput. 1994;6(3):341–356. doi: 10.1162/neco.1994.6.3.341

[25] Ueda N, Nakano R. Deterministic annealing variant of the em algorithm. Adv Neural Inf Process Syst. 1994;7:545–552.

[26] Wang J. A deterministic annealing neural network for convex programming. Neural Networks. 1994;7(4):629–641. doi: 10.1016/0893-6080(94)90041-8

[27] Ueda N, Nakano R. Deterministic annealing em algorithm. Neural Networks. 1998;11(2):271–282. doi: 10.1016/s0893-6080(97)00133-0

[28] Wu Z, Karimi HR, Dang C. An approximation algorithm for graph partitioning via deterministic annealing neural network. Neural Networks. 2019;117:191–200. doi: 10.1016/j.neunet.2019.05.010

[29] Katahira K, Watanabe K, Okada M. Deterministic annealing in variational Bayesian algorithm. IEICE Tech Rep. 2007;106(589):177–182.

[30] Katahira K, Watanabe K, Okada M. Deterministic annealing variant of variational Bayes method. In: Journal of Physics: Conference Series; Kyoto, Japan. IOP Publishing; 2008; Vol. 95. p. 012015.

[31] McLachlan GJ, Krishnan T. The EM algorithm and extensions. 2nd ed. New York (NY): Wiley; 2008.

[32] Brent RP. Algorithms for minimization without derivatives. Englewood Cliffs, NJ: Prentice-Hall; 1973.

[33] Nakamura K, Nagamura N, Ueno K, et al. All 2d heterostructure tunnel field-effect transistors: impact of band alignment and heterointerface quality. ACS Appl Mater Interfaces. 2020;12(46):51598–51606. doi: 10.1021/acsami.0c13233

[34] Horiba K, Nakamura Y, Nagamura N, et al. Scanning photoelectron microscope for nanoscale three-dimensional spatial-resolved electron spectroscopy for chemical analysis. Rev Sci Instrum. 2011;82(11):113701. doi: 10.1063/1.3657156

[35] Nagamura N, Fukidome H, Nagashio K, et al. Influence of interface dipole layers on the performance of graphene field effect transistors. Carbon. 2019;152:680–687. doi: 10.1016/j.carbon.2019.06.038

## Appendix A. Derivation of the fitting model

### A1. *Pseudo-Voigt mixture model*

We use a pseudo-Voigt mixture model ($\mathbf{P_{PVMM}}$) to represent the peak shape. The pseudo-Voigt distribution is a general peak-fitting model for spectral data and is easy to calculate. $\mathbf{P_{PVMM}}$ is defined as a linear superposition of $K$ normalised pseudo-Voigt distributions $\mathbf{P_{PV}}(x_n|\mu,\sigma,\eta)$ as follows:

$$\mathbf{P_{PVMM}}(x_n|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\eta},\boldsymbol{\lambda}) = \sum_{k=1}^{K} \lambda_k \mathbf{P_{PV}}(x_n|\mu_k,\sigma_k,\eta_k), \quad \text{(A1)}$$

where $\lambda$ is the mixture ratio for each component ($0 \leq \lambda \leq 1$ and $\sum_{k=1}^{K} \lambda_k = 1$) and $\mathbf{P_{PV}}(x_n|\mu,\sigma,\eta)$ is expressed as follows:

$$\mathbf{P_{PV}}(x_n|\mu,\sigma,\eta) = \frac{\text{pseudoVoigt}(x_n|\mu,\sigma,\eta)}{\sum_{n=1}^{N} \text{pseudoVoigt}(x_n|\mu,\sigma,\eta)}, \quad \text{(A2)}$$

where

$$\text{pseudoVoigt}(x_n|\mu,\sigma,\eta) = \eta \mathbf{L}(x_n|\mu,\sqrt{2\ln 2}\sigma) + (1-\eta)\mathbf{N}(x_n|\mu,\sigma), \quad \text{(A3)}$$

**L** and **N** are Lorentz and Gauss distributions, respectively; $\mu$ is the mean; $\sigma$ is the standard deviation ($0 < \sigma$); and $\eta$ is the mixing parameter of the Lorentz and Gauss distributions ($0 \leq \eta \leq 1$). As the distribution has a broad tail and finite measurement energy steps $x_n$, each distribution is expressed as a normalised distribution over the range of $x_n$.

### A2. *Linear background model*

We introduce a linear background model that is used for background subtraction. This model assumes that a straight line can approximate the background shape. We define the background model $\mathbf{B}(x_n;x_1,x_N)$ on $N$ measurement points $x_n$. We assume that the background intensity monotonically increases from $x_1$ to $x_N$ and that a continuous function can approximate the background component. Based on these assumptions, we define the linear background model using triangle $\mathbf{P}_{\text{tri}}(x_n;x_1,x_N)$ and uniform $\mathbf{P}_{\text{uni}}(x_n;x_1,x_N)$ distributions as follows:

$$\mathbf{P_B}(x_n;x_1,x_N) = \lambda_{K+1}\mathbf{P}_{\text{uni}}(x_n;x_1,x_N) + \lambda_{K+2}\mathbf{P}_{\text{tri}}(x_n;x_1,x_N), \quad \text{(A4)}$$

where

$$\mathbf{P}_{\text{uni}}(x_n;x_1,x_N) = \frac{1}{x_N - x_1} \quad \text{(A5)}$$

and

$$\mathbf{P}_{\text{tri}}(x_n;x_1,x_N) = \frac{2(x_n - x_1)}{(x_N - x_1)^2}. \quad \text{(A6)}$$

Equation (A4) is determined uniquely by the end points of the measurement step ($x_1$, $x_N$) and the mixing ratio of the background component ($\lambda_{K+1}$, $\lambda_{K+2}$).

### A3. *Fitting model*

The fitting model is a pseudo-Voigt mixture model ($\mathbf{P_{PVMM}}$) with a linear background model ($\mathbf{P_{PVMMB}}$), which can be expressed as follows:

$$\mathbf{P_{PVMMB}}(x_n|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\eta},\boldsymbol{\lambda}) = \sum_{k=1}^{K} \lambda_k \mathbf{P_{PV}}(x_n|\mu_k,\sigma_k,\eta_k) + \lambda_{K+1}\mathbf{P}_{\text{uni}}(x_n;x_1,x_N) + \lambda_{K+2}\mathbf{P}_{\text{tri}}(x_n;x_1,x_N). \quad \text{(A7)}$$

The first term on the right side of Equation (A7) represents the peak components, the second term represents the baseline of the background, and the third term represents the trend of the background. In addition, as the total number of mixture ratios $\lambda$ is $K+2$, $\lambda$ sums to 1 (i.e. $0 \leq \lambda \leq 1$, $\sum_{k=1}^{K+2} \lambda_k = 1$). Optimizing Equation (A7) for spectral data means that peak fitting and background subtraction are performed simultaneously.