

Full Length Article

Bayesian estimation analysis of X-ray photoelectron spectra: Application to Si 2p spectrum analysis of oxidized silicon surfaces

Hiroshi Shinotsuka^a, Kenji Nagata^b, Hideki Yoshikawa^{a,*}, Shuichi Ogawa^c, Akitaka Yoshigoe^d

^a Research Network and Facility Services Division, National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan

^b Center for Basic Research on Materials, National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan

^c College of Industrial Technology, Nihon University, Narashino, Chiba 275-8575, Japan

^d Materials Sciences Research Center, Japan Atomic Energy Agency, Sayo, Hyogo 679-5148, Japan

ARTICLE INFO

Keywords:

Bayesian estimation
X-ray photoelectron spectroscopy
Statistical analysis
Silicon surface oxidation

ABSTRACT

X-ray Photoelectron Spectroscopy (XPS) is known as a powerful experimental technique that provides information on the chemical states of material surfaces, and curve-fitting analyses based on methods such as least-squares are widely used for spectrum analysis. However, a major issue with these analyses is that the number of spectral components and other analytical conditions often include qualitative or arbitrary settings. In response to this, we applied Bayesian estimation to the spectral data of the oxidized state of silicon (Si) surfaces, which we previously reported. Bayesian estimation discussed in this paper is based on stochastic modelling without incorporating physical properties such as Si oxidation. By applying this to the time-dependent oxidation spectra of Si surfaces, we reconfirmed that latent information on number of peaks, peak positions, intensity changes and spin-orbit splitting conditions can be extracted through Bayesian posterior probabilities.

1. Introduction

The photoelectric effect, which occurs when light is irradiated onto a material, is a physical phenomenon that makes it possible to capture the electrons present in the material. X-ray photoelectron spectroscopy (XPS), which measures the kinetic energy of the generated electrons (photoelectrons), provides direct information about the binding energy of the electrons present in a material. In particular, when light (X-rays) of sufficiently high energy is used, it can capture the element-specific binding energies of electrons bound to the inner-shell orbitals of atoms, making elemental analysis possible. Furthermore, the binding energy of the inner-shell electrons of an element of interest varies depending on the chemical environment surrounding the element. This difference in binding energy is called the chemical shift, which enabled Siegbahn to develop electron spectroscopy for chemical analysis (ESCA) through the implementation of high-resolution XPS [1]. XPS allows the chemical analysis of solids, liquids, and gases, and has been applied to not only fundamental science but also engineering.

Among the wide range of XPS applications, the analysis of the oxidation state (oxidation number) of materials is of considerable importance. Oxides on the surface of semiconductor materials such as gate insulators of metal-oxide-semiconductor transistors, interlayer

dielectrics, or protective films between devices can be analyzed by XPS. Himpsel *et al.* used high-energy-resolution synchrotron radiation as an excitation source to separate the chemical shift components corresponding to the distinct oxidation number in the silicon (Si) 2p photoemission spectra of silicon oxides (SiO_x) on Si single-crystal surfaces [2]. They clarified in detail the differences in the oxidation states at the interface between the oxide and substrate depending on the crystallographic orientation of the Si substrate. Since this pioneering report, SiO_x films and interfaces have been actively studied by high-energy-resolution photoemission spectroscopy, and it has been recognized as a powerful tool for investigating the chemical state of surface oxides.

The third-generation high-brilliance synchrotron radiation facilities that began operating in the late 1990s have increased the sensitivity and energy resolution of XPS. This has enabled us to obtain detailed information about the chemical state and formation process of surface oxides; for example, the charge of oxides, atomic arrangement differences on Si single-crystal surfaces, the structure of oxide-substrate interfaces, and kinetics of oxide formation. In such precise analysis by XPS using synchrotron radiation, the resulting spectrum of, for example, an SiO_x film on an Si(001) surface, is complicated, as shown in Fig. 1. The Si 2p inner-shell orbital line splits into 2p_{3/2} and 2p_{1/2} signals because of spin-orbit coupling. Each spectrum consists of several peaks with different

* Corresponding author.

E-mail address: YOSHIKAWA.Hideki@nims.go.jp (H. Yoshikawa).

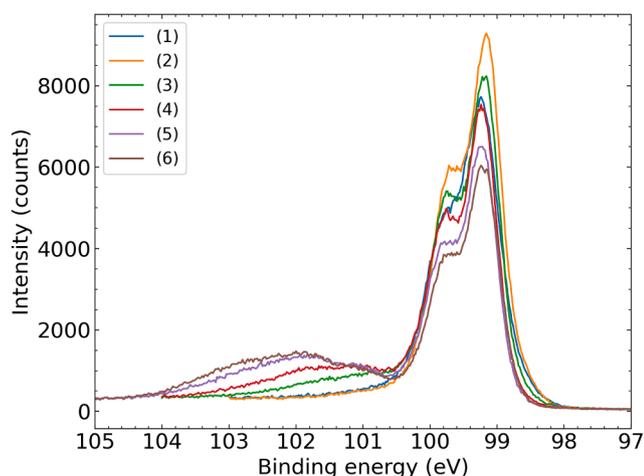


Fig. 1. Representative silicon 2p spectra measured during thermal oxidation of a clean silicon surface.

chemical shifts depending on the electronic state of Si. Such spectra may include signals from bulk Si; Si^{S} and Si^{SS} from surface dimers on clean surfaces; Si^+ , Si^{2+} , Si^{3+} , and Si^{4+} from SiO_x ; and Si^{α} and Si^{β} from Si with interfacial strain [3–5].

In peak separation analysis of such complex spectra, there is a widely known classical approach to remove the $2p_{1/2}$ component that appears because of spin–orbit coupling and analyze only the $2p_{3/2}$ component [5]. First, the Proctor–Sherwood algorithm is used to remove the Shirley-type background from the spectrum, and then Fourier transformation is used to extract the $2p_{3/2}$ component. However, this spectral processing by Fourier transformation has two important intrinsic problems. First, one needs to find the optimum spin–orbit splitting width and branching ratio for each spectrum to remove the $2p_{1/2}$ component. The determined splitting width is 0.608 eV, which is a good approximation to use [6]. The branching ratio then needs to be determined from the theoretical value of 0.5 by taking into account slight variations. However, this is difficult because of the noise in the spectrum. The shape of $2p_{3/2}$ spectra extracted by Fourier transformation is sensitive to the setting of the branching ratio, which hinders the detection of small peaks. The second problem is the loss of information in the noise of the observed spectra; because XPS signals are measured as count data, the noise is known to follow a Poisson distribution. This noise generation process provides useful information for data analysis, but the relationship between intensity and noise is lost during the Fourier transformation process in the analysis of spectra over a finite energy range. For correct analysis, it is desirable to analyze spectra while preserving the Poisson character of the noise without background removal or Fourier transformation.

Even if the $2p_{3/2}$ component can be extracted, there is another issue when further peak separation is performed using the extracted $2p_{3/2}$ component. Although it is known that multiple peaks exist, it is difficult to objectively determine their number, which leads to subjectivity. In particular, it is difficult to discriminate small peaks at the base of large peaks in bulk Si, and the existence of such peaks has long been discussed. In the above separation process of the $2p_{1/2}$ and $2p_{3/2}$ signals of Si 2p spectra via Fourier transformation and curve-fitting analysis by the least-squares method, the assumptions regarding the analysis environment, such as the computer, the existence of constituent components, and the shape of the spectrum, may depend on the knowledge and/or subjectivity of the analyst. Clarifying the details of photoelectron spectra, such as the number of components and the minimum value of signal intensity, which are essential for data processing, is not only necessary for capturing the true picture of the chemical state of materials, but also an important issue for efficient data processing.

Along with recent advances in computer hardware, such as machine

learning, there has been remarkable progress in data science, including algorithms and mathematical statistical analysis. In the field of chemical analysis, the mathematical analysis of spectra using Bayesian estimation has advanced considerably, and its applications are expanding to data acquisition and analysis in a wide range of techniques, including nuclear magnetic resonance and infrared spectroscopies, X-ray diffraction, and electron-beam diffraction. The application of Bayesian estimation to XPS is also being pioneered [7].

In this study, we reinvestigate the Si 2p photoelectron spectra of thermally oxidized Si(001) surfaces reported by Ogawa *et al.* [5] by taking into account the spin–orbit coupling using Bayesian estimation. We discuss the accuracy of the estimation of fitting parameters based on Bayesian posterior distributions as well as the model selection of the number of peaks. In the Si 2p spectral analysis by Ogawa *et al.*, [5] the curve-fitting analysis was performed assuming the existence of various chemical states of Si atoms on a clean Si surface and at the interface between the oxide layer and Si substrate, as is often done in general peak separation. In contrast, our method in this paper does not use such chemical-state assumptions and performs spectral analysis without any prior information about the positions of Si peaks apart from the prominent bulk Si peak. Through the above analysis, we showed that latent information on number of peaks, peak positions, intensity changes and spin–orbit splitting conditions can be extracted through Bayesian posterior probabilities.

2. Theory

2.1. Fitting model

Fig. 2 shows a schematic diagram of the model function used for the spectrum fitting. The model function is expressed as the sum of the spectrum, which is divided into $2p_{3/2}$ and $2p_{1/2}$ components by spin–orbit splitting, and the background. The model function $f(x; \theta)$ is defined as:

$$f(x; \theta) = s(x; \theta) + b(x; \theta), \quad (1)$$

$$s(x; \theta) = s_{3/2}(x; \theta) + s_{1/2}(x; \theta), \quad (2)$$

$$s_{3/2}(x; \theta) = \sum_{k=1}^K A_k \tilde{V}(x - \mu_k; \sigma_k, \gamma_k), \quad (3)$$

$$s_{1/2}(x; \theta) = R \cdot s_{3/2}(x - \Delta E; \theta), \quad (4)$$

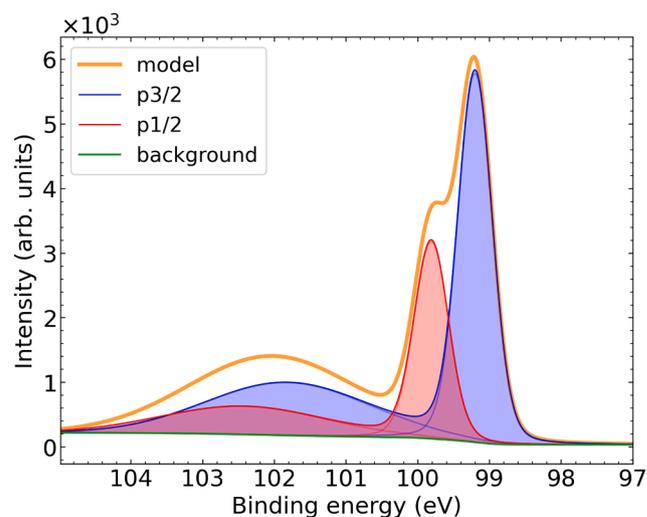


Fig. 2. Schematic of a model function consisting of two basis functions, considering spin–orbit splitting.

$$b(x; \theta) = (b_H - b_L) \frac{\int_{x_S}^x s(x'; \theta) dx'}{\int_{x_S}^{x_E} s(x'; \theta) dx'} + b_L, \quad (5)$$

where $s(x; \theta)$ is the signal function, $b(x; \theta)$ is the background function, x is the binding energy, and θ is the fitting parameter set. In the case of p orbitals, the signal function $s(x; \theta)$ is divided into the contribution from $2p_{3/2}$ ($s_{3/2}(x; \theta)$) and the contribution from $2p_{1/2}$ ($s_{1/2}(x; \theta)$) because of the spin-orbit splitting. Eq. (4) expresses the relation between $s_{1/2}(x; \theta)$ and $s_{3/2}(x; \theta)$ by introducing the spin-orbit splitting width ($\Delta\epsilon$) and branching ratio (R). The basis function representing a peak is assumed to be the convolutional Voigt function normalized to unity height ($\tilde{V}(x; \sigma, \gamma)$). $s_{3/2}(x; \theta)$ consists of the sum of K basis functions, where a single basis function $\tilde{V}(x; \sigma, \gamma)$ is defined as:

$$\tilde{V}(x; \sigma, \gamma) = V(x; \sigma, \gamma) / V(0; \sigma, \gamma), \quad (6)$$

$$V(x; \sigma, \gamma) = \int_{-\infty}^{+\infty} G(x'; \sigma) L(x - x'; \gamma) dx', \quad (7)$$

$$G(x; \sigma) = (\sigma\sqrt{2})^{-1} \exp(-x^2/2\sigma^2), \quad (8)$$

$$L(x; \gamma) = \gamma(\pi(x^2 + \gamma^2))^{-1}, \quad (9)$$

where $V(x; \sigma, \gamma)$ is the convolutional Voigt function, $G(x; \sigma)$ is the Gaussian distribution function with standard deviation σ , and $L(x; \gamma)$ is the Lorentz function with half-width γ . The parameter set θ to be optimized is $\theta = \{ \{ \mu_k, A_k, \sigma_k, \gamma_k \}_{k=1}^K, b_H, b_L, \Delta\epsilon, R \}$, where μ_k, A_k, σ_k , and γ_k are the peak position, height, standard deviation of the Gaussian distribution function, and half-width of the Lorentz function of the k -th peak, respectively. b_H and b_L are the intensities of the background functions at high- and low-binding-energy sides, respectively. The background function $b(x; \theta)$ is the widely used Shirley background [8].

2.2. Bayesian estimation

Ogawa et al. [5] performed a three-step numerical procedure for fitting spectra: background removal, decomposition of spin-orbit splitting by Fourier transformation, and peak fitting. This method has some problems: the effect of noise cannot be evaluated and the results can vary depending on the values of spin-orbit splitting and branching ratio used. In contrast, the method proposed in this study uses Bayesian estimation to handle spin-orbit splitting and background in the fitting process, while considering the effect of noise when estimating the parameters of the model.

Bayesian estimation is a framework in which the process of data generation is formulated in a probabilistic model, and estimation is performed by tracing back causal relationships using Bayes' theorem [7,9]. In ordinary spectral peak fitting, the results depend on initial values and tend to fall into local solutions. In contrast, combining Bayesian estimation with the exchange Monte Carlo method allows not only global peak fitting of spectra, but also confidence intervals for the fitting parameters to be obtained from Bayesian posterior probability distributions. Furthermore, the number of peaks in a spectrum can be estimated by comparing the Bayesian free energies obtained from fitting models with different numbers of peaks [10]. Further details are given in Appendix A and B.

3. Results and Discussion

3.1. Experimental data

Fig. 1 shows a series of Si 2p spectra of a clean Si surface during thermal oxidation [5]. The 2p-level photoelectron spectra split into $2p_{1/2}$ and $2p_{3/2}$ peaks because of the interaction of spin and orbital angular

momentum. There is a large doublet peak in the 99–100 eV range that originates from bulk Si and a feature in the 100–104 eV range that originates from SiO_x . Spectrum (1) in Fig. 1 is of the clean Si surface and lacks oxide-derived peaks. As oxidation progresses, the oxide-derived peaks gradually become larger (spectra (2)–(6)).

3.2. Bayesian estimation calculation settings

In Bayesian estimation, it is necessary to set a prior distribution for each parameter to be estimated. The prior distribution is related to the search range of the parameters and affects the results of parameter estimation and model selection. In the case of a typical spectral analysis, the same probability distribution is set for all peaks and a sufficiently wide search range is used [7]. Because the data set for this analysis traced the oxidation process of an Si surface, bulk Si-derived peaks were expected to appear as the main spectral component. Therefore, we set the prior distributions corresponding to the bulk peaks separately from the prior distributions of other peaks (see Appendix C in detail).

When designing the prior distribution, several parameters were extracted from the measured spectral data, as shown in Fig. 3. First, the peak top intensity (y_{top}) and minimum intensity (y_{min}) of each spectrum in the entire data set were extracted. The approximate intensity of the maximum peak is $y_{\text{max}} = y_{\text{top}} - y_{\text{min}}$. These parameters were used to design the prior distribution of peak intensity parameters. In addition, as a design related to the peak position, the maximum and minimum values of the binding-energy of the spectral data were extracted as x_{min} and x_{max} , respectively. The peak top position of the spectral data was extracted as x_{top} . As a prior distribution design for background parameters, the average intensities of the left- and rightmost ten points in the spectral data were extracted as b_H and b_L , respectively. Detailed prior distribution settings and conditions for the exchange Monte Carlo method are provided in Appendix C. The conventional process of decomposition of spin-orbit splitting by Fourier transformation is described in Appendix D.

3.3. Bayesian estimation calculation results

We will first present the results of the analysis of spectra (1) and (6) from Fig. 1. Fig. 4 shows results of the analysis of spectrum (1) in Fig. 1. Fig. 4(a) illustrates a model selection based on Bayesian estimation. The probability that the peak number $K = 4$ is 99.1 %, $K = 5$ is 0.9 %, and other probabilities are less than 0.1 %, indicating that $K = 4$ is plausible. Fig. 4(b) shows a fitting result that minimizes cost function $E(\theta, K)$, defined in Appendix A, with respect to parameter θ when $K = 4$. The thick blue and thick red lines are the Si $2p_{3/2}$ and $2p_{1/2}$ spectra, respectively. The blue areas indicate the individual Voigt functions that

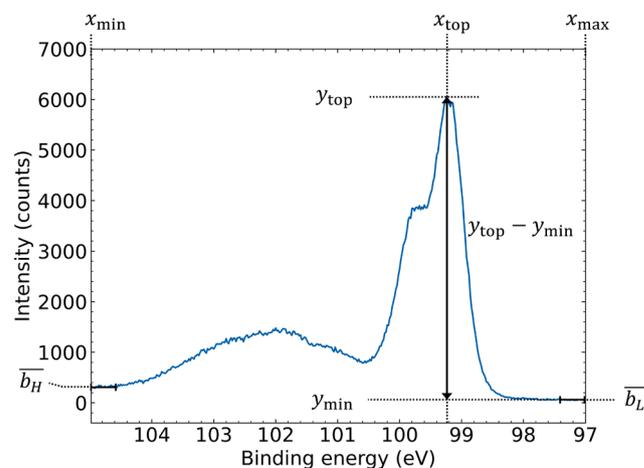


Fig. 3. Schematic diagram of setting parameters.

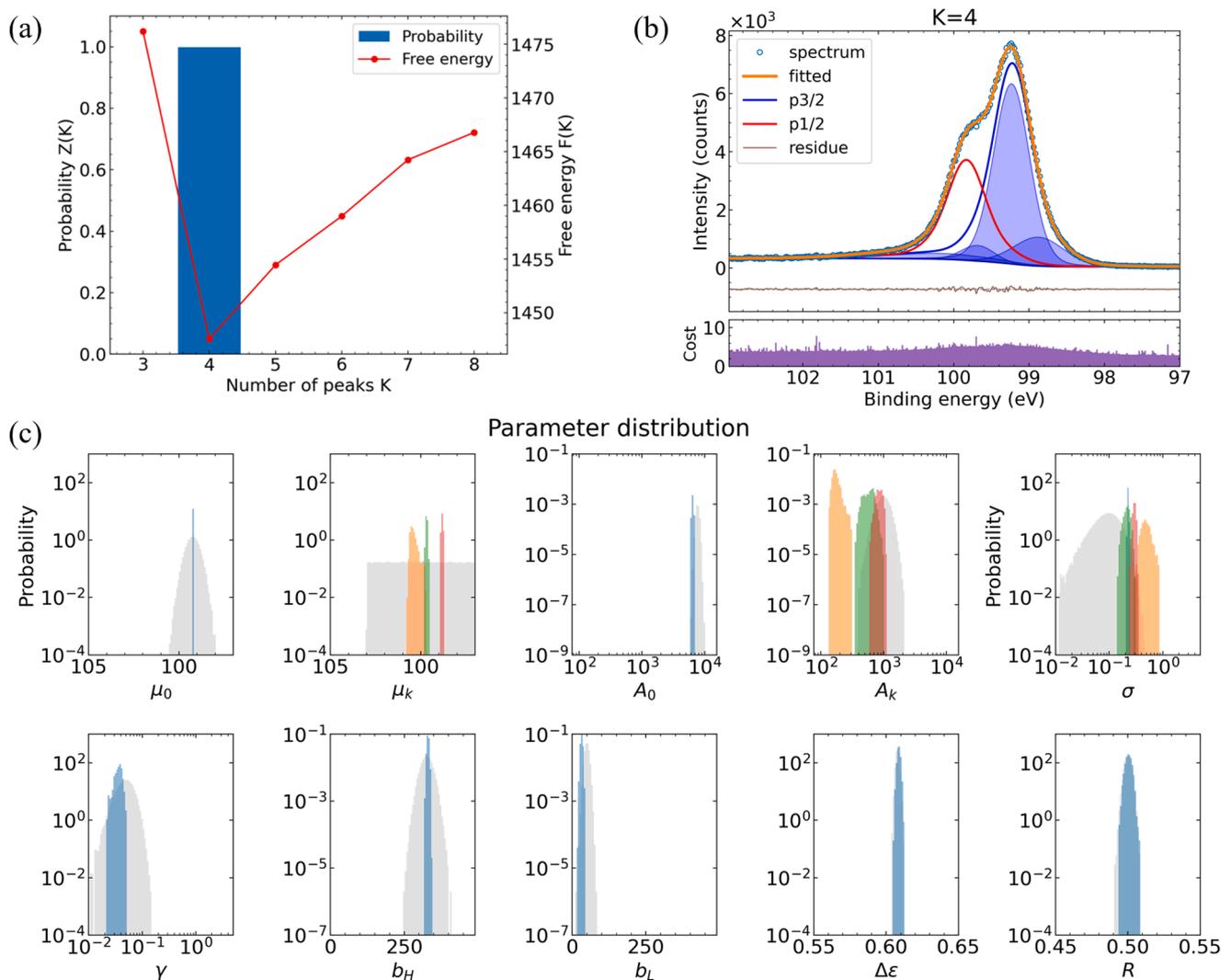


Fig. 4. Analysis results for spectrum (1) in Fig. 1. (a) Result of model selection by Bayesian estimation. The dots and bars represent the free energy $F(K)$ and the probabilities $Z(K)$, respectively, as functions of the number of peaks K . (b) Fitting result when $K = 4$. (c) Posterior probability distributions for all parameters when $K = 4$. The light gray distribution in each graph is the prior distribution.

make up Si $2p_{3/2}$. There is one prominent bulk Si peak, which is sandwiched between two smaller peaks. Fig. 4(c) shows the posterior probability distribution of each parameter when $K = 4$. The gray areas in this figure represent the prior distributions. The posterior distributions of peak height A_k , Gaussian width σ_k , and Lorentzian width γ_k are represented on a logarithmic scale along the horizontal axis. This choice reflects that these parameters only take positive values and can vary significantly between peaks. These posterior distributions are normalized to integrate to 1. The posterior distribution of the peak position μ_0 of bulk Si is much sharper than the prior distribution, meaning that the peak position is clearly determined. The prior distribution of the peak position μ_k for the subpeaks other than that of bulk Si is given by a uniform distribution, whereas the posterior distribution is unimodal, indicating that the peak position is automatically narrowed down. The posterior distributions of the other parameters are also unimodal, indicating that the fitting model is well refined.

Fig. 5 shows results of the analysis of spectrum (6) in Fig. 1. Fig. 5(a) depicts the model selection based on Bayesian estimation. The probability of $K = 7$ is 97.9 %, $K = 6$ is 1.7 %, $K = 5$ is 0.3 %, and the other probabilities are less than 0.1 %, indicating that $K = 7$ is plausible. The fitting result that minimizes cost function $E(\theta, K)$ for $K = 7$ is presented in Fig. 5(b). Fig. 5(c) shows the posterior probability distribution of each parameter for $K = 7$. Despite the complex shape of the spectrum, the

peak positions are sharply localized. The posterior distributions of the other parameters are also unimodal, indicating that a specific model has been derived. The analysis results for spectra (2) to (5) in Fig. 1 are presented in Fig. 6.

So far, we have shown the fitting of six spectra selected as snapshots of the oxidation process. Actually, Ogawa *et al.* [5] collected a series of 70 Si $2p$ spectra during the thermal oxidation of Si. We analyzed all 70 spectra using our method. Fig. 7 shows the posterior probability distribution of the peak positions obtained from this analysis in chronological order of oxidation. Bayesian model averaging [11] was performed for each spectrum. Posterior distributions for the peak positions were obtained for each of the three to eight candidate peak numbers. As a result of the Bayesian estimation, probability for the number of peaks were obtained. The probabilities were used to calculate a weighted average of the posterior distribution of peak positions. The nine red dotted lines in Fig. 7 show the Si $2p$ peak positions listed in Table 1, which is referred by Ogawa *et al.* [5]. The peaks originating from bulk Si are prominent throughout the oxidation process. The spectra contain four peaks during in the initial stage of oxidation and then the number of peaks increases as oxidation proceeds until finally they contain seven peaks.

In the early stage of oxidation, the S peak (derived from the dimer on the Si(001) 2×1 surface) appears to gradually shift to the Si^{IV} position.

In previous peak separation analysis, S and Si^{IV} were treated as

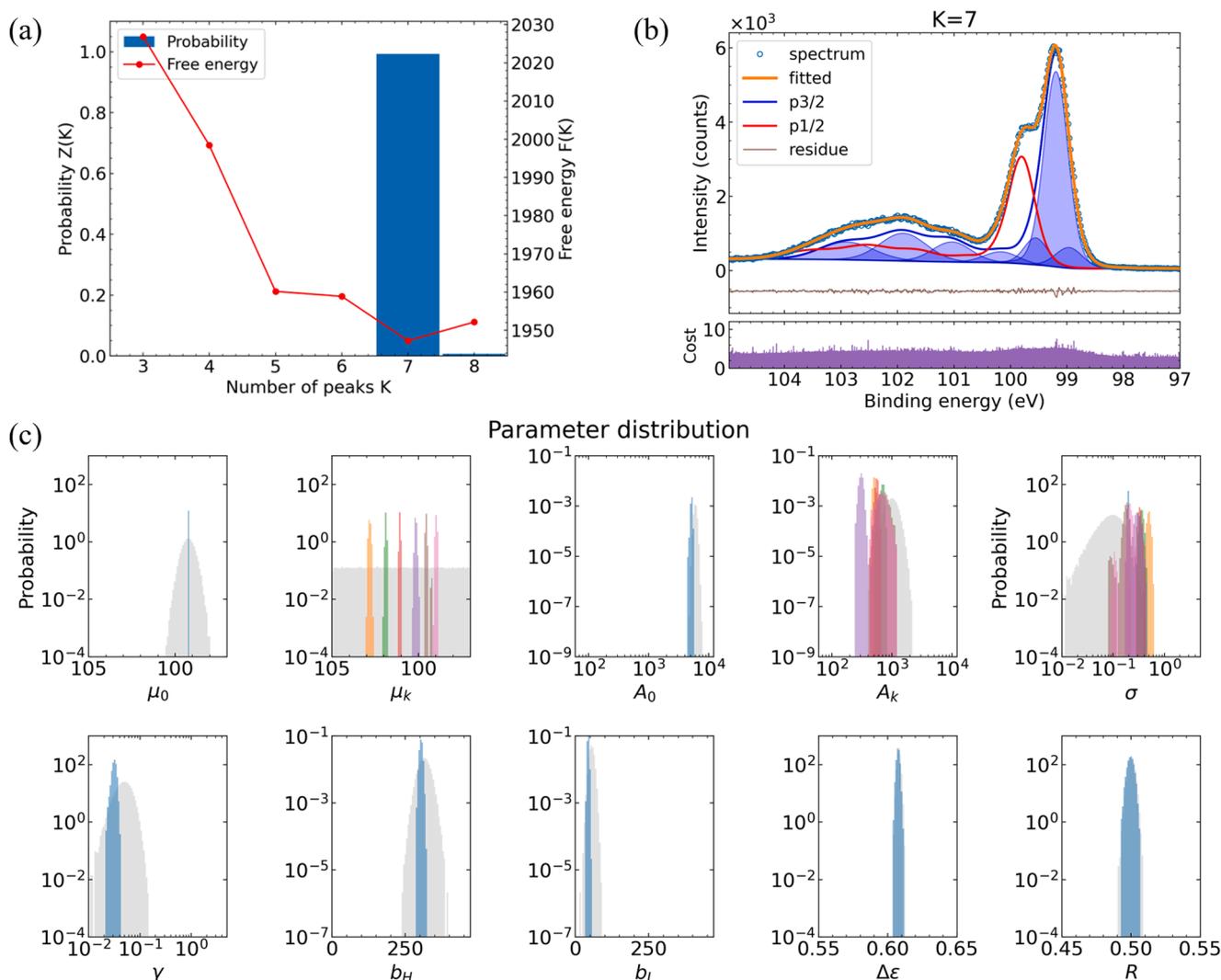


Fig. 5. Analysis results for spectrum (6) in Fig. 1. (a) Result of model selection by Bayesian estimation. (b) Fitting result when the number of peaks $K = 7$. (c) Posterior probability distributions for all parameters when $K = 7$. The description of each figure is the same as that of Fig. 4.

different peaks [5]. This is because the S peak shifts from the Si^β position due to charge transfer driven by asymmetric dimer formation [12], whereas Si^α shifts as a result of the charge density change between Si-Si bonds caused by strain [13]. Because the Si-Si dimer disappears when the Si(001) surface oxidation is completed, the S peak is considered to have zero intensity at this point. In contrast, as the oxidation progresses, oxidation-induced strain is generated at the oxide/Si interface, and Si^α and Si^β peaks appear because of this strain. In the conventional least-squares method, Si^α and Si^β signals appear near the S and S' peaks, so the positions of the strain-induced Si^α and Si^β peaks need to be fixed. If the least-squares fitting was attempted with the peak positions as variables, the Si^α and Si^β peaks would shift toward the S and S' peaks and overlap with them. However, the Si^α and Si^β peak positions vary with the magnitude of strain [13]. Therefore, it is not desirable to fix the peak positions of Si^α and Si^β even when fitting by the least-squares method. In addition, the shape of both peaks reflects the strain distribution at the oxide/Si interface and does not necessarily have the shape of the Gaussian or Voigt functions. Until now, it has not been possible to track the change in the magnitude of the Si^α and Si^β shifts with oxidation because of the limitations of data analysis. Therefore, the main focus was obtaining the changes in the Si^α and Si^β peak intensities during oxidation; the changes in strain caused by oxidation were considered only qualitatively [14].

In this paper, using Bayesian estimation, we have verified that there

are seven components reported in the curve-fitting analysis of Si 2p XPS spectra of oxidized Si surfaces. This result provides a quantitative evidence for the curve-fitting analysis from a mathematical point of view and will be important information for the previous XPS analysis of the Si oxidation model and the Si substrate-oxide film interface. As mentioned above, the results of this analysis are consistent with the physical interpretation of the Si 2p XPS spectra of the oxidized surface in our previous paper [5], but it should be noted that they do not prove the uniqueness of the physical interpretation in the paper. As a future research, it is expected that applying this Bayesian estimation to higher quality XPS data will enable deeper discussions on more complex physical phenomena that were not addressed in previous data analyses, such as how strain in Si-Si bonds affects the dynamic response of the electronic system (final state effects) [15]. It should be noted that when discussing such delicate spectral decomposition, the selection of background models [16] other than the Shirley method used in this study should also be taken into consideration.

4. Conclusions

We reinvestigated Si 2p spectra during thermal oxidation through Bayesian estimation considering spin-orbit splitting. Then, the model selection of the number of peaks was performed, and the estimation accuracy of each fitting parameter was obtained from the corresponding

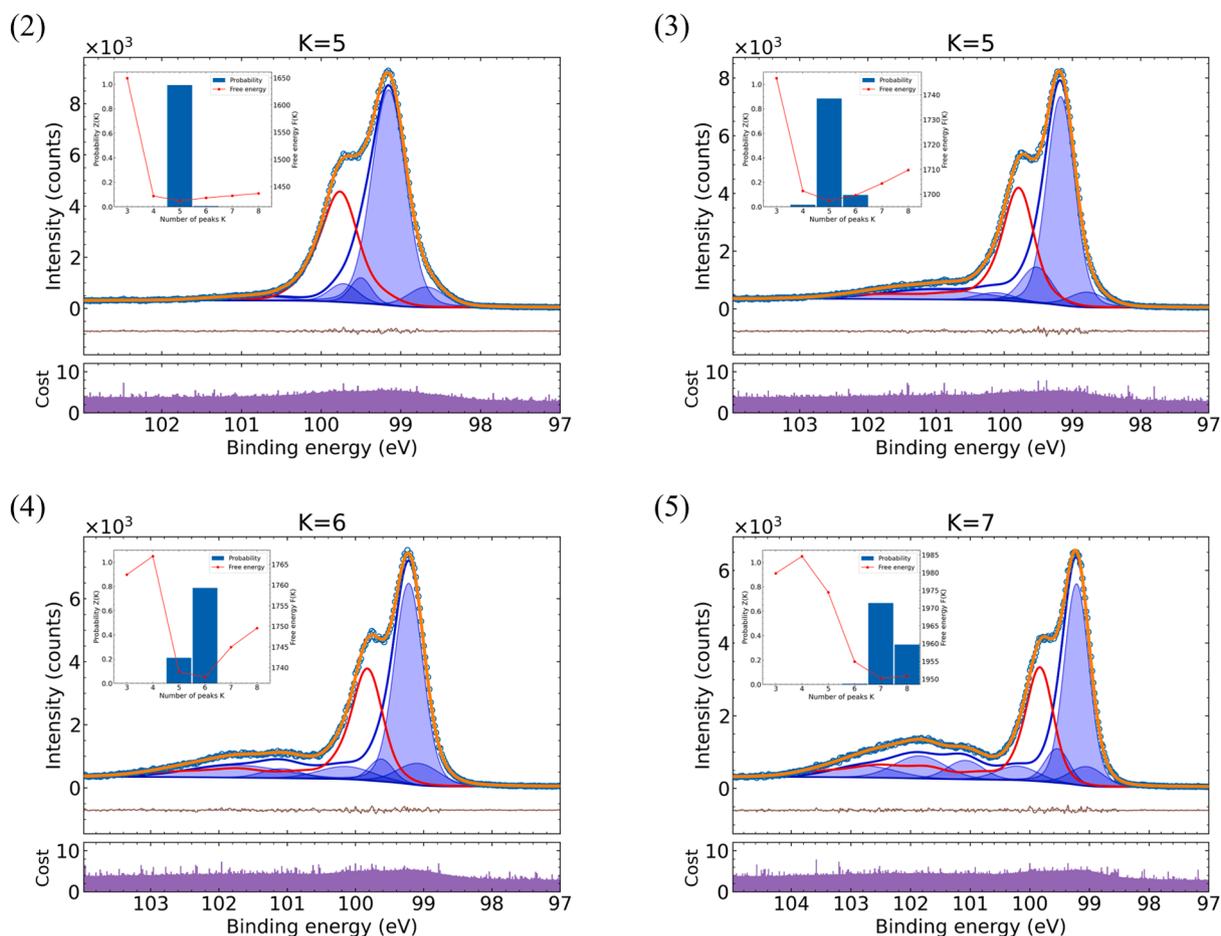


Fig. 6. Fitting results for spectra (2)–(5) in Fig. 1. The description of the fitting results is the same as in Fig. 4(b). The inset of each figure is the result of model selection by Bayesian estimation; the description is the same as that of Fig. 4(a).

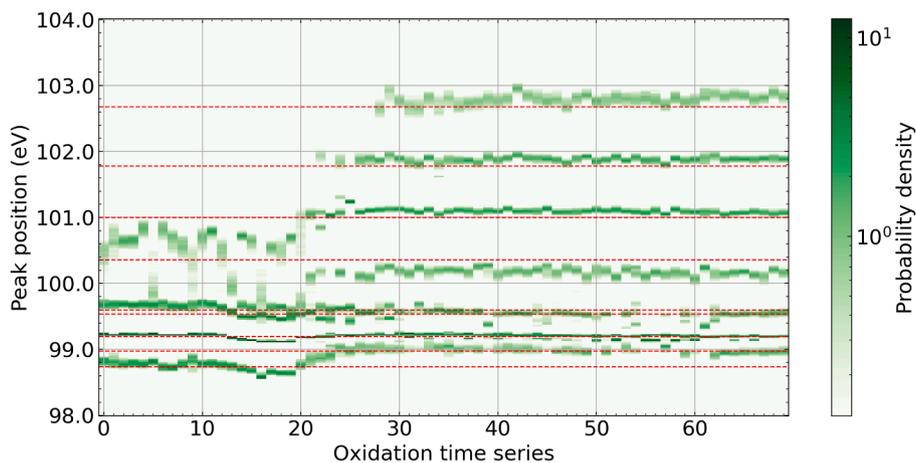


Fig. 7. Posterior probability distribution of peak positions with Bayesian model averaging, in chronological order of oxidation. The red dotted lines indicate representative peak positions assigned in the literature [5].

Table 1
Silicon Peak Assignments and Chemical Shifts [5].

Peak	Si ⁴⁺	Si ³⁺	Si ²⁺	Si ⁺	SS	Si ^β	Si ^β	Si ^α	S
Chemical shift (eV)	3.48	2.59	1.81	1.15	0.41	0.34	0	−0.22	−0.46

Bayesian posterior distribution. This method allows direct fitting to observed spectra without any prior processing such as background removal or extraction of the $2p_{3/2}$ component. As such, it allows the analysis to take into account a stochastic interpretation of the noise inherent in the spectrum.

As oxidation progressed, Si^+ , Si^{2+} , Si^{3+} , and Si^{4+} peaks appeared sequentially, and Si^α and Si^β peaks derived from interfacial strain were detected instead of S and SS peaks. The shift of Si^α , Si^β , S, and SS peaks caused by the change of oxidation-induced strain at the oxide/Si interface with the progress of oxidation was clearly observed. When the Si surface oxidation is completed, the spectra can be deconvoluted into seven peaks, which is consistent with a previous study.

CRediT authorship contribution statement

Hiroshi Shinotsuka: Writing – original draft, Validation, Software.

Kenji Nagata: Writing – review & editing, Formal analysis. **Hideki Yoshikawa:** Supervision. **Shuichi Ogawa:** Writing – review & editing, Resources, Investigation. **Akitaka Yoshigoe:** Writing – review & editing, Resources, Investigation.

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.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP20K05338, JP26420289 and JP23K04578.

Appendix A. 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 [7,9,10]. We first define a probabilistic model. Because XPS measures the number of photoelectrons by pulse counting, its statistical noise follows a Poisson process.

Given a dataset $D = \{x_i, y_i\}_{i=1}^n$ of narrow photoelectron spectra, the probability that the measured count number is y_i when the intensity calculated from the model function at x_i is $f_i = f(x_i; \theta)$ can be expressed using the Poisson distribution function as $p(y_i|x_i, \theta) = \frac{f_i^{y_i} e^{-f_i}}{y_i!}$, where θ is a parameter set of the model function. Assuming that data $\{y_i\}_{i=1}^n$ are independent and identically distributed, using the cost function $E(\theta, K)$, we obtain the conditional probability $p(D|\theta, K)$ of data set D as:

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

$$E(\theta, K) = \sum_{i=1}^n (f(x_i; \theta) - y_i \log f(x_i; \theta) + \log \Gamma(y_i + 1)). \quad (A.2)$$

A smaller value of the cost function $E(\theta, K)$ indicates a better fit between the model and spectral data.

In the Bayesian estimation, we regard not only the data D but also the parameter set θ and the number of peaks K as random variables; i.e., we construct a model of the simultaneous probability distribution $p(D, \theta, K)$:

$$p(D, \theta, K) = p(D|\theta, K)p(\theta|K)p(K) \propto \exp(-E(\theta, K))p(\theta|K)p(K). \quad (A.3)$$

Here, $p(K)$ and $p(\theta|K)$ are prior distributions, and they should be set in advance. According to Bayes' theorem, the simultaneous probability distribution can also be expressed as $p(D, \theta, K) = p(\theta|D, K)p(D, K)$. Here, $p(\theta|D, K)$ is the posterior probability distribution of the fitting parameters given the spectral data and the number of peaks, and it can be calculated as follows:

$$\begin{aligned} p(\theta|D, K) &= \frac{p(D, \theta, K)}{p(D, K)} = \frac{p(D, \theta, K)}{\int p(D, \theta, K) d\theta} \\ &= \frac{\exp(-E(\theta, K))p(\theta|K)p(K)}{p(K) \int \exp(-E(\theta, K))p(\theta|K) d\theta} \\ &= \frac{1}{Z(K)} \exp(-E(\theta, K))p(\theta|K) \end{aligned} \quad (A.4)$$

where

$$Z(K) = \int \exp(-E(\theta, K))p(\theta|K) d\theta. \quad (A.5)$$

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

We used the posterior probability $p(K|D)$ to estimate the number of peaks K by adopting the procedure of probability marginalization:

$$\begin{aligned}
p(K|D) &= \frac{p(D, K)}{p(D)} = \frac{\int p(D, \theta, K) d\theta}{\sum_{k=1}^K \int p(D, \theta, K) d\theta} \\
&= \frac{\int \exp(-E(\theta)) p(\theta|K) p(K) d\theta}{\sum_{k=1}^K \int \exp(-E(\theta)) p(\theta|k) p(k) d\theta} \\
&= \frac{p(K) Z(K)}{\sum_{k=1}^K p(k) Z(k)}
\end{aligned} \tag{A.6}$$

where $\sum_{k=1}^K$ represents the sum of all peaks, labelled with k . The denominator in Eq. (A.6) does not depend on K because it is the sum of all peaks. We thus obtain:

$$p(K|D) \propto p(K) Z(K). \tag{A.7}$$

Therefore, to obtain the posterior probability $p(K|D)$ with respect to the number of peaks, we need to calculate $Z(K)$. The Bayesian free energy is defined as:

$$\begin{aligned}
F(K) &= -\log Z(K) \\
&= -\log \int \exp(-E(\theta, K)) p(\theta|K) d\theta.
\end{aligned} \tag{A.8}$$

If $p(K)$ is uniformly distributed, the maximization of the posterior probability $p(K|D)$ is equivalent to the minimization of the Bayesian free energy $F(K)$. The number of peaks K can therefore be estimated by minimizing the Bayesian free energy.

Appendix B. Exchange Monte Carlo method

The replica exchange Monte Carlo (EMC) method, one of the Markov chain Monte Carlo methods, was used to evaluate the confidence intervals of the posterior probability distribution $p(\theta|D, K)$ and Bayesian free energy $F(K)$ [7,9,10,17]. We prepared multiple distributions:

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

$$Z_{\beta}(K) = \int \exp(-\beta E(\theta, K)) p(\theta|K) d\theta, \tag{B.2}$$

in which the inverse temperature β was introduced into the posterior distribution, and simultaneously performed sampling in parallel. That is, the target distribution became the simultaneous distribution:

$$p(\theta_1, \dots, \theta_L) = \prod_{i=1}^L p_{\beta_i}(\theta_i|D). \tag{B.3}$$

Here, the inverse temperature $\{\beta_i\}_{i=1}^L$ is $0 = \beta_1 < \beta_2 < \dots < \beta_L = 1$ and L is the number of prepared temperatures. θ_i is the fitting parameter for inverse temperature β_i . When $\beta_L = 1$, $p_{\beta_L}(\theta_L|D)$ matches the posterior distribution $p(\theta|D, K)$.

The algorithm of the EMC method has two updates.

Metropolis sampling for each temperature. For each inverse temperature β_i , we sampled θ_i from the distribution of $p_{\beta_i}(\theta|D)$ using the Metropolis algorithm.

State exchange between adjacent temperatures. We exchanged states between adjacent temperatures and set $\{\theta_i, \theta_{i+1}\} \rightarrow \{\theta_{i+1}, \theta_i\}$. We determined whether to exchange on the basis of the probability:

$$p(\theta_i \leftrightarrow \theta_{i+1}) = \min(1, \nu), \tag{B.4}$$

$$\begin{aligned}
\nu &= \frac{p_{\beta_i}(\theta_{i+1}|D) p_{\beta_{i+1}}(\theta_i|D)}{p_{\beta_i}(\theta_i|D) p_{\beta_{i+1}}(\theta_{i+1}|D)} \\
&= \exp\{(\beta_{i+1} - \beta_i)(E(\theta_{i+1}, K) - E(\theta_i, K))\}.
\end{aligned} \tag{B.5}$$

As a result, the parameter θ_L obtained from the L -th inverse temperature β_L can be regarded as sampling from the posterior distribution $p(\theta|D, K)$. Thus, by repeating the above procedure a sufficient number of times and recording θ_L , we obtained a sample sequence from the posterior distribution $p(\theta|D, K)$, which allowed us to determine the maximum a posteriori (MAP) estimators and confidence interval.

There are two advantages of the EMC method. The first is that heating and annealing effects can be introduced by stochastically exchanging samples between adjacent temperatures during the sampling of each distribution. We thereby escape local solutions and efficiently search for global solutions. This is an important advantage for the accurate estimation of MAP estimators and confidence intervals. The second is that the free energy $F(K)$ can also be calculated using the results of the EMC method. We introduced the inverse temperature β into the free energy:

$$f(\beta) = -\log \int \exp(-\beta E(\theta, K)) p(\theta|K) d\theta. \tag{B.6}$$

At the high-temperature limit $\beta = 0$, it holds that $f(\beta) = 0$. The desired free energy $F(K) = f(\beta = 1)$ is expressed as:

$$F(K) = f(\beta = 1) = \int_0^1 d\beta \frac{\partial f}{\partial \beta} \quad (\text{B.7})$$

$$\begin{aligned} \frac{\partial f}{\partial \beta} &= \frac{\int E(\theta, K) \exp(-\beta E(\theta, K)) p(\theta|K) d\theta}{\int \exp(-\beta E(\theta, K)) p(\theta|K) d\theta} \\ &= \int E(\theta, K) p_\beta(\theta|D) d\theta \equiv \langle E(\theta, K) \rangle_{p_\beta(\theta|D)}. \end{aligned} \quad (\text{B.8})$$

$F(K)$ is therefore an integral of the expected value $\langle E(\theta, K) \rangle_{p_\beta(\theta|D)}$ under the probability distribution $p_\beta(\theta|D)$ with respect to the inverse temperature β in the range from 0 to 1. In addition, because the results of the EMC method provide samples at each inverse temperature β_i , the expected value $\langle E(\theta, K) \rangle_{p_{\beta_i}(\theta|D)}$ can be calculated. Therefore, by using the piecewise quadrature method, we can calculate the free energy and perform model selection [7,17].

Appendix C. Settings used in Bayesian estimation analysis

Here we describe the prior distribution of each parameter used in the Bayesian estimation process. It is easy to identify the peak derived from bulk Si because of its position and strong intensity relative to those of other peaks. Therefore, we set a different prior distribution for the peak parameter of peak number 1, which is derived from bulk Si, than for the other peaks. The prior distributions for different parameters are summarized in Table C.1. Details of other parameters (x_{top} , x_{min} , x_{max} , h_{max} , \bar{b}_H , and \bar{b}_L) can be found in the main text. The distribution types denoted as Gauss, Uniform, and Gamma refer to Gaussian, uniform, and gamma distributions, respectively.

Fig. 3 in the main text shows a schematic diagram of a spectrum illustrating setting parameters. We supposed that the peak top position and height of the target spectral data were x_{top} and y_{top} , respectively, and the minimum value of the spectral data was y_{min} . x_{min} and x_{max} were assigned as the minimum and maximum values of the binding energy of the spectral data, respectively. The prior distribution of the peak position μ_1 of bulk Si was assumed to be Gaussian with a mode of x_{top} and standard deviation of 0.1 eV. Assuming that the locations of the peaks other than that of bulk Si were completely unknown, the prior distributions of their peak positions $\{\mu_k\}_{k=2}^K$ were considered to be uniform in the range of $(x_{\text{min}}, x_{\text{max}})$. The prior distribution of the bulk Si peak height A_1 was determined according to the height of the target spectrum and was a gamma distribution with mode $h_{\text{max}} = y_{\text{top}} - y_{\text{min}}$ and standard deviation of $0.05h_{\text{max}}$. The prior distributions of peak heights $\{A_k\}_{k=2}^K$ for the peaks other than that of bulk Si were gamma distributions with a mode of 1000.0 and standard deviation of 250.0. These values were determined based on the subpeak heights of a series of spectral data. The prior distributions of the Gaussian and Lorentzian widths of the peaks, σ_k and γ_k , respectively, were empirically chosen to be gamma distributions of 0.300 ± 0.010 eV and 0.050 ± 0.016 eV, respectively. The prior distribution of the spin-orbit splitting of Si 2p, $\Delta\varepsilon$, was assigned as a gamma distribution with a mode of 0.608 eV and standard deviation of 0.001 eV by referring to the literature [6]. Because it is obvious that the branching ratio R of the spin-orbit splitting of the p orbitals is approximately 0.5, the prior distribution of R was assumed to be a gamma distribution with a mode of 0.50 and standard deviation of 0.01. The Lorentzian widths $\{\gamma_k\}_{k=1}^K$ were assumed to be the same for all peaks; this is because all peaks originated from Si 2p and the lifetime of holes associated with photoionization can be considered to be identical. The prior distributions of the endpoint intensities b_H and b_L on the high- and low-binding-energy sides of the background were Gaussian distributions of $\bar{b}_H \pm \sqrt{\bar{b}_H}$ and $\bar{b}_L \pm \sqrt{\bar{b}_L}$, respectively, where \bar{b}_H and \bar{b}_L are the average of the intensities of ten spectral data points from each endpoint.

Table C1
Prior Distribution Settings for Various Parameters.

Parameter	Distribution type	Distribution properties	
μ_1 (eV)	Gauss	mode = x_{top}	std = 0.100
μ_k ($k \neq 1$) (eV)	Uniform	min = x_{min}	max = x_{max}
A_1	Gamma	mode = h_{max}	std = $0.05h_{\text{max}}$
A_k ($k \neq 1$)	Gamma	mode = 1000	std = 250
σ_k (eV)	Gamma	mode = 0.300	std = 0.010
γ_k (eV)	Gamma	mode = 0.050	std = 0.016
$\Delta\varepsilon$ (eV)	Gamma	mode = 0.608	std = 0.001
R	Gamma	mode = 0.500	std = 0.005
b_H	Gauss	mode = \bar{b}_H	std = $\sqrt{\bar{b}_H}$
b_L	Gauss	mode = \bar{b}_L	std = $\sqrt{\bar{b}_L}$

The gamma distribution is commonly expressed in terms of shape and scale parameters. However, in our case, we represented the gamma distribution using its mode and standard deviation to associate it with meanings of the spectrum. The shape parameter η_G and scale parameter θ_G of the gamma distribution can be easily calculated when the mode m_G and standard deviation σ_G of the gamma distribution are given as follows:

$$\theta_G = 0.5 \left(-m_G + \sqrt{m_G^2 + (2\sigma_G)^2} \right), \quad (\text{C.1})$$

$$\eta_G = 1 + \frac{m_G}{\theta_G}. \quad (\text{C.2})$$

As a configuration for the calculation of the exchange Monte Carlo (MC) method, the number of inverse temperatures was set to $L = 96$, and each inverse temperature was assigned as follows:

$$\beta_l = \begin{cases} 0 & (\text{for } l = 1) \\ 1.2^{l-L} & (\text{otherwise}) \end{cases}. \quad (\text{C.3})$$

In the MC simulation, 10,000 MC steps were used for the burn-in and a subsequent 10,000 MC steps for the sampling.

Appendix D. Deconvolution of spin-orbit splitting by Fourier transformation

Here we describe a procedure for extracting Si 2p_{3/2} from an Si 2p spectrum with spin-orbit splitting using Fourier transformation, and point out the problems involved with this process. The spectrum $y(x)$ of the 2p orbital after removing the background is the sum of the 2p_{3/2} component $y_{3/2}(x)$ and 2p_{1/2} component $y_{1/2}(x)$. Assuming that the spin-orbit splitting width is $\Delta\varepsilon$ and the branching ratio is R , then we have:

$$\begin{aligned} y(x) &= y_{1/2}(x) + y_{3/2}(x) \\ &= R \cdot y_{3/2}(x + \Delta\varepsilon) + y_{3/2}(x). \end{aligned} \quad (\text{D.1})$$

Fourier transformation and inverse transformation give:

$$\mathcal{F}[y](\omega) = (1 + R e^{i\omega\Delta\varepsilon}) \mathcal{F}[y_{3/2}](\omega), \quad (\text{D.2})$$

$$\therefore y_{3/2}(x) = \mathcal{F}^{-1} \left[\frac{\mathcal{F}[y](\omega)}{1 + R e^{i\omega\Delta\varepsilon}} \right] (x). \quad (\text{D.3})$$

Although the formula is simple, it is difficult to manually search for an appropriate combination of splitting width and branching ratio $\{\Delta\varepsilon, R\}$. We used the negative area of $y_{3/2}(x)$ relative to the positive area as a possible criterion and searched for a model in which this criterion became small. The target spectral data were Si 2p spectra of thermally oxidized Si films, which are available on COMPRO software [18]. Figure D.1(a) shows a grid search of $\Delta\varepsilon$ and R in the ranges of $[-1.0, 0.0]$ eV and $[0.0, 1.0]$, respectively, where the red region indicates that the criterion value is small. In the case of Si 2p, it is known that good separation can be obtained in the vicinity of $\{\Delta\varepsilon, R\} = \{0.6, 0.5\}$, and it was confirmed that such a good region does indeed exist. However, the shape of the resulting $y_{3/2}(x)$ is sensitive to parameters. As an example, Figure D.1(b) shows $y_{3/2}(x)$ calculated by fixing $\Delta\varepsilon = -0.608$ eV and changing R from 0.4 to 0.6. A small change in the value of R causes a large change in the shape of $y_{3/2}(x)$, especially in the low intensity region from 100 to 102 eV. This is especially critical for the initial stage of Si oxidation, which is the target of this study. Changing the value of $\{\Delta\varepsilon, R\}$ may affect the presence or absence of small subpeaks, and the Fourier transformation may produce artifactual peaks. Thus, the combination of $\{\Delta\varepsilon, R\}$ produces numerous candidate solutions, but the choice of the optimal value is left to the final judgment of the analyst, which introduces subjectivity.

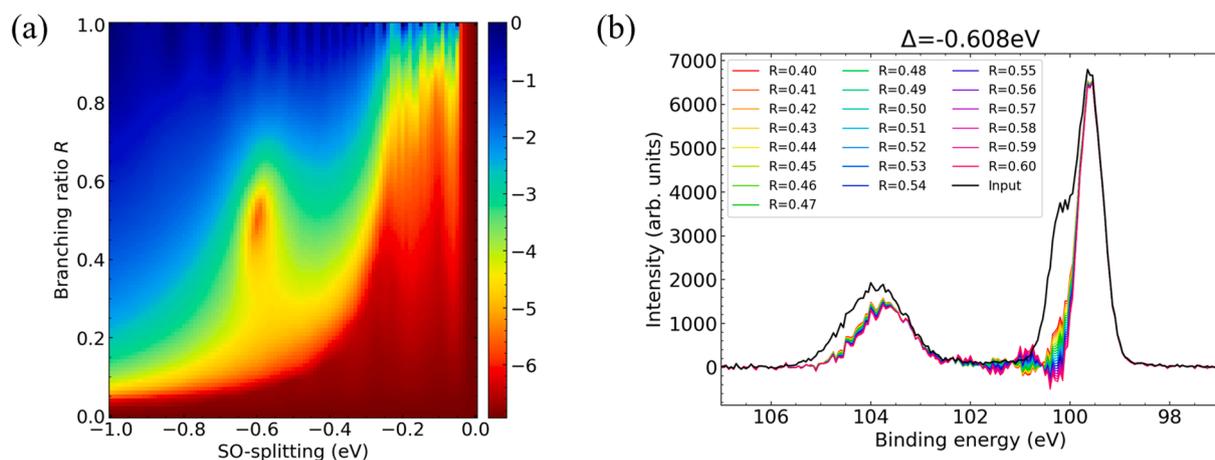


Fig. D1. Example of deconvolution of an Si 2p spectrum of a thermally oxidized Si film by Fourier transformation. (a) Heatmap of the determined values when the spin-orbit splitting width and branching ratio R are varied. (b) Si 2p_{3/2} spectra for $\Delta\varepsilon = -0.608$ eV and $R = 0.4$ to 0.6 .

Data availability

Data will be made available on request.

References

- [1] K. Siegbahn, C. Nordling, A. Fahlman, R. Nordberg, K. Hamrin, J. Hedman, G. Johansson, T. Bergmark, S.-E. Karlsson, I. Lindgren, B. Lindberg, ESCA, atomic, molecular and solid state structure studied by means of electron spectroscopy, *Nov. Act. Reg. Soc. Sci. Upsaliensis Ser. IV* 20 (1967) 1.
- [2] F.J. Himpsel, F.R. McFeely, A. Taleb-Ibrahimi, J.A. Yarmoff, G. Hollinger, Microscopic structure of the SiO₂/Si interface, *Phys. Rev. B* 38 (1988) 6084–6096, <https://doi.org/10.1103/PhysRevB.38.6084>.
- [3] J.H. Oh, H.W. Yeom, Y. Hagimoto, K. Ono, M. Oshima, N. Hirashita, M. Nywa, A. Toriumi, A. Kakizaki, Chemical structure of the ultrathin SiO₂/Si(100) interface:

An angle-resolved Si 2p photoemission study, *Phys. Rev. B* 63 (2001) 1–6, <https://doi.org/10.1103/PhysRevB.63.205310>.

- [4] S. Dreiner, M. Schürmann, M. Krause, U. Berges, C. Westphal, Determination of the source of two extra components in Si 2p photoelectron spectra of the SiO₂/Si(100) interface, *J. Electr. Spectr. Rel. Phenom.* 144–147 (2005) 405–408, <https://doi.org/10.1016/j.elspec.2005.01.120>.
- [5] S. Ogawa, J. Tang, A. Yoshigoe, S. Ishidzuka, Y. Teraoka, Y. Takakuwa, Relation between oxidation rate and oxidation-induced strain at SiO₂/Si(001) interfaces during thermal oxidation, *Jpn. J. Appl. Phys.* 52 (2013) 110128, <https://doi.org/10.7567/JJAP.52.110128>.
- [6] J. Stober, B. Eisenhut, G. Rangelov, T. Fauster, Initial stages of the thermal nitridation of the Si(100) surface with NH₃ and NO: a surface sensitive study of Si 2p core-level shifts, *Surf. Sci.* 321 (1994) 111–126, [https://doi.org/10.1016/0039-6028\(94\)90032-9](https://doi.org/10.1016/0039-6028(94)90032-9).
- [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] D.A. Shirley, High-resolution x-ray photoemission spectrum of the valence bands of gold, *Phys. Rev. B* 5 (1972) 4709–4714, <https://doi.org/10.1103/PhysRevB.5.4709>.
- [9] 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>.
- [10] 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. Tech. Adv. Mat.* 21 (2020) 402–419, <https://doi.org/10.1080/14686996.2020.1773210>.
- [11] A.E. Raftery, D. Madigan, J.A. Hoeting, Bayesian model averaging for linear regression models, *J. Am. Stat. Assoc.* 92 (1997) 179–191, <https://doi.org/10.1080/01621459.1997.10473615>.
- [12] E. Landemark, C.J. Karlsson, Y.C. Chao, R.I.G. Uhrberg, Core-level spectroscopy of the clean Si(001) surface: charge transfer within asymmetric dimers of the 2×1 and $c(4 \times 2)$ reconstructions, *Phys. Rev. Lett.* 69 (1992) 1588, <https://doi.org/10.1103/PhysRevLett.69.1588>.
- [13] O.V. Yazyev, A. Pasquarello, Origin of fine structure in Si 2p photoelectron spectra at silicon surfaces and interfaces, *Phys. Rev. Lett.* 96 (2006) 157601, <https://doi.org/10.1103/PhysRevLett.96.157601>.
- [14] Y. Tsuda, A. Yoshigoe, S. Ogawa, T. Sakamoto, Y. Yamamoto, Y. Takakuwa, Roles of excess minority carrier recombination and chemisorbed O₂ species at SiO₂/Si interfaces in Si dry oxidation: comparison between p-Si(001) and n-Si(001) surfaces, *J. Chem. Phys.* 157 (2022) 234705, <https://doi.org/10.1063/5.0109558>.
- [15] E. Pehlke, M. Scheffler, Evidence for site-sensitive screening of core holes at the Si and Ge (001) surface, *Phys. Rev. Lett.* 71 (1993) 2338–2341, <https://doi.org/10.1103/PhysRevLett.71.2338>.
- [16] A. Herrera-Gomez, M.O. Vazquez-Lepe, P.G. Mani-Gonzalez, P. Pianetta, F. S. Aguirre-Tostado, O. Ceballos-Sanchez, The ST component in the Si 2p photoemission spectrum from H-terminated and oxidized Si (001) surfaces, *J. Vac. Sci. Technol. A* 41 (2023), <https://doi.org/10.1116/6.0002690>.
- [17] Y. Ogata, A monte carlo method for an objective bayesian procedure, *Ann. Inst. Stat. Math.* 42 (1990) 403–433, <https://doi.org/10.1007/BF00049299>.
- [18] K. Yoshihara, The introduction of common data processing system version 12, *J. Surf. Anal.* 23 (2017) 138–148, <https://doi.org/10.1384/jsa.23.138>.