



# Multivariate analysis for scanning tunneling spectroscopy data

Junsuke Yamanishi<sup>a</sup>, Shigeru Iwase<sup>b</sup>, Nobuyuki Ishida<sup>c</sup>, Daisuke Fujita<sup>c</sup>

<sup>a</sup> Department of Applied Physics, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

<sup>b</sup> Graduate School of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan

<sup>c</sup> National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan

## ARTICLE INFO

### Article history:

Received 16 August 2017

Accepted 14 September 2017

Available online 18 September 2017

### Keywords:

Scanning tunneling microscopy

Multivariate analysis

Principal component analysis

## ABSTRACT

We applied principal component analysis (PCA) to two-dimensional tunneling spectroscopy (2DTS) data obtained on a Si(111)-(7 × 7) surface to explore the effectiveness of multivariate analysis for interpreting 2DTS data. We demonstrated that several components that originated mainly from specific atoms at the Si(111)-(7 × 7) surface can be extracted by PCA. Furthermore, we showed that hidden components in the tunneling spectra can be decomposed (peak separation), which is difficult to achieve with normal 2DTS analysis without the support of theoretical calculations. Our analysis showed that multivariate analysis can be an additional powerful way to analyze 2DTS data and extract hidden information from a large amount of spectroscopic data.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

Scanning tunneling spectroscopy (STS) is one of the most important operation modes of scanning tunneling microscopy (STM), and can evaluate the local density of states (LDOS) [1–4] and the elementary excitations [5–7] at surfaces with high spatial and energy resolution. Normally, tunneling differential conductance ( $dI/dV$ ) is measured as a function of bias voltage ( $V$ ) in STS experiments since  $dI/dV$  is proportional to the LDOS of the surface [4,8,9]. The observed features in the spectra are related to the states at energy  $E$  with the simple relationship  $E - E_F = eV$ , where  $E_F$  is the Fermi level in the sample and  $V$  is the tip-sample voltage [9].

In addition to local tunneling spectroscopy, two-dimensional tunneling spectroscopy (2DTS) (also known as current imaging tunneling spectroscopy) is also used to visualize the spatial distribution of LDOS at energy  $E$  [9,10]. In a 2DTS measurement, STS is performed at each pixel of the STM image. The typical image size is about  $128 \times 128$  points, thus providing a large set of tunneling spectra (more than 10,000). Although the spectral features in each curve are related to the LDOS distribution in energy, it is not straightforward to determine their origin due to the lack of standard spectra in contrast to other spectroscopic methods, such as X-ray photo electron spectroscopy (XPS) and Auger electron spectroscopy (AES) [11–13]. Therefore, generally, density-functional theory (DFT) calculation is required to interpret the spectral features.

Multivariate analysis has been used in a wide range of fields as a method to extract significant information from a large amount of data [14]. One of the most frequently used multivariate analyses is principal component analysis (PCA), which finds combinations of variables (called principal components, PCs) that describe important trends in large amounts of data. PCA is often used to decrease the number of dimensions of large amounts of data and makes it easier to interpret them. In fact, PCA has so far been applied to large amounts of spectroscopic data from XPS, AES and time-of-flight secondary ion mass spectrometry for peak identifications, peak decompositions, estimation of the number of linearly independent chemical species in a collection of spectra, and reduction of noise in spectra [15–20].

In this study, we analyzed 2DTS data (obtained on a Si(111)-(7 × 7) surface) using PCA to explore the effectiveness, potential, and limitations of the multivariate analysis for interpreting 2DTS data (which is in the form of a large set of tunneling spectra). Components classified by PCA can be assumed to originate dominantly from the rest atoms, center adatoms in the faulted half (FH), center adatoms in the unfaulted half (UFH), and dimer atoms. It is difficult to discern some of them in the raw spectral data due to the superposition of those components with other larger components. Thus, it is difficult to extract those components from 2DTS data using normal analysis methods without the support of theoretical calculation. Our analysis showed that multivariate analysis can be an additional powerful tool for analyzing 2DTS data to extract hidden information.

E-mail addresses: [yamanishi@ap.eng.osaka-u.ac.jp](mailto:yamanishi@ap.eng.osaka-u.ac.jp), [de.brogli.e.wavelength@gmail.com](mailto:de.brogli.e.wavelength@gmail.com) (J. Yamanishi).

## 2. Methods

### 2.1. STS

STS measurements were performed at 78 K under ultrahigh vacuum (UHV) with a pressure of less than  $1 \times 10^{-8}$  Pa. Commercial PtIr tips were used. Tip cleaning was performed *in situ* using electron bombardment. The Si(111)-(7 × 7) surface was prepared by heating a piece of n-type Si(111) wafer (P doped, <0.01 Ωcm) at 1470 K for 5 s under UHV. In the 2DTS measurement, we obtained the tunneling current spectrum (*I*-*V* curve) at each pixel of the STM image with a grid size of 128 × 128. The number of spectral points was 256 in a bias range from -1.3 V to 1.3 V. The obtained spectra were numerically differentiated using a Savitzky-Golay differentiation filter. After that, the *dI/dV*-*V* spectra were normalized by the averaged tunneling conductance (*I/V*) to remove the tunneling probability component [21].

### 2.2. PCA

The data set of 128 × 128 normalized *dI/dV* spectra from the 2DTS measurement can be viewed as a matrix of the normalized *dI/dV* values where the columns correspond to the variables (bias voltage) and rows correspond to the samples (spectrum at each pixel). PCA was performed on this matrix, *D*(*m*, *n*), where *m* is the number of samples and *n* is the number of variables. To calculate the PCs of *D*, we solve an eigenvalue problem of the corresponding covariance matrix [14]. First, matrix *D* is transformed to matrix *X* through mean-centering and division by  $\sqrt{m-1}$  as follows.

$$X = \frac{1}{\sqrt{m-1}}(d_1 - \langle d \rangle)(d_2 - \langle d \rangle) \dots (d_m - \langle d \rangle), \quad (1)$$

where  $d_i$  ( $i = 1, 2, \dots, m$ ) is a row vector corresponding to the *i*th normalized *dI/dV* spectrum and  $\langle d \rangle$  is the row vector corresponding to the averaged normalized *dI/dV* spectrum calculated from *m* (128 × 128) spectral data. Then, the covariance matrix of *D* can be calculated as  $X^T X$ . To solve the eigenvalue problem of the covariance matrix, singular value decomposition (SVD) was used. SVD gives the eigenvalues and eigenvectors of  $XX^T$  and  $X^T X$ . According to SVD,

$$X = U\Lambda^{1/2}V^T, \quad (2)$$

where  $U(m, l)$  and  $V(n, l)$  ( $l$  is a rank of matrix *X*) are matrices consisting of the eigenvectors of  $XX^T$  and  $X^T X$ , respectively.  $\Lambda^{1/2}(l, l)$  is a diagonal matrix whose diagonal elements are the eigenvalues,  $\lambda_j$  ( $j = 1, 2, \dots, l$ ), where  $\lambda_1 > \lambda_2 > \dots > \lambda_l$ . The eigenvectors of  $X^T X$  correspond to the PCs,  $P_j$  ( $j = 1, 2, \dots, l$ ), of *D*. The PCs are orthogonal to each other, and the norm is equal to 1, that is,

$$P_i \cdot P_j = \delta_{ij}, \quad (3)$$

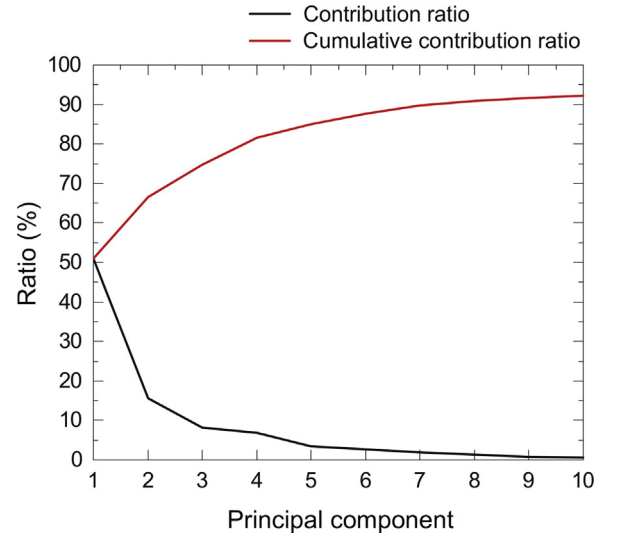
where  $\delta_{ij}$  is the Kronecker delta. The percentage of original variance retained by each PC (contribution ratio) can be calculated using the following equation:

$$\text{Contribution ratio of PC}_j = 100 \times \frac{\lambda_j}{\sum_{k=1}^l \lambda_k} \quad (j = 1, 2, \dots, l). \quad (4)$$

The original normalized *dI/dV* spectra can be expressed by the sum of the averaged normalized *dI/dV* spectrum and the linear combination of the PCs, as follows:

$$d_i = \langle d \rangle + \sum_{j=1}^l c_j P_j \quad (i = 1, 2, \dots, m), \quad (5)$$

where  $c_j$  is the coefficient of the *j*th PC, which reflects the amount of correlation between  $d_i$  and  $P_j$ . Since the PCs are orthogonal to each



**Fig. 1.** Contribution ratio and cumulative contribution ratio of principal components from PC1 to PC10.

other,  $c_j$  can be calculated by taking the inner product between  $d_i - \langle d \rangle$  and  $P_j$ . Thus, the score matrix  $S(m, l)$  can be calculated using the following equation,

$$S = XV = U\Lambda^{1/2}, \quad (6)$$

where each row of  $S(m, l)$  contains the projections of the corresponding spectrum on the different PCs. Also, score plot images corresponding to  $P_j$  can be obtained by arranging the values in each column of  $S(m, l)$  into two-dimensional arrays (128 × 128) with the same size as that of the STM image.

### 2.3. Computational details

DFT simulations were performed using localized atomic basis sets implemented in the GPAW code [22–24]. Optimized atomic structures have been obtained within the Perdew–Burke–Ernzerhof (PBE) generalized-gradient approximation [25]. Partial density of states (PDOS) calculations were conducted using the GLLB-sc functional [26]. The electron-core interactions were described by the projector-augmented wave method [27]. We used double-zeta plus polarized basis sets and 8 × 8 k-points in the two-dimensional Brillouin zone.

## 3. Results and discussion

We performed PCA on a data set of the normalized *dI/dV* spectra from a 2DTS measurement that contains 128 × 128 spectra (see the Section 2 for details). Fig. 1 shows the contribution ratio and the cumulative contribution ratio of the calculated PCs. The contribution ratio gradually decreases as the PC number increases. After PC5, the values were less than 5% and decrease nearly linearly. Thus, we roughly estimated the number of dominant principal components to be four. Since the PCs with a low contribution ratio have little information included in the spectra data, we discuss only about PC1–PC4 in detail below.

The loading plots of PC1–PC4 (PC spectrum) are shown in Fig. 2. As discussed with regards to Eq. (5) in the Methods section, the normalized *dI/dV* spectra can be expressed as a sum of the averaged spectrum ( $\langle d \rangle$ ) and the linear combination of calculated PCs ( $\sum c_j P_j$ ). This means that the spectral shape of each PC spectrum exhibits the difference from the averaged spectrum ( $\langle d \rangle$ ). Thus, the loading values can be negative. The PC1 spectrum shows a large negative peak at -0.89 V and a small one at 0.45 V. The PC2 spec-

Download English Version:

<https://daneshyari.com/en/article/5347173>

Download Persian Version:

<https://daneshyari.com/article/5347173>

[Daneshyari.com](https://daneshyari.com)