doi:10.1017/S1431927619001363

Distinguishability of Structures via Principal Component Analysis: Application to 4D STEM

Mark P. Oxley[1,2], Sergei V. Kalinin[1,2] and Rama K. Vasudevan[1,2*]

[1.] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge TN, USA.
[2.] Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge TN, USA.
* Corresponding author: vasudevanrk@ornl.gov

Work on atomically precise placement of Bi dopants within a Silicon lattice has attracted attention due to potential uses in quantum information technologies, as opposed to the originally studied Phosphorous dopants [1]. Therefore, precise control first requires accurate knowledge of the dopant position in all three dimensions, i.e. (x,y,z). Recently, with the advent of 4D scanning transmission electron microscopy (4D STEM), substantially more information on the chemistry and structure of a cross-sectioned sample is (in principle) available [2]. Yet, one of the main challenges with the proliferation of 4D STEM is the inversion of the diffraction patterns to yield the precise structural information required.

Here, we present methods based on principal component analysis to distinguish between substitutional Bi dopants placed at different depths within a Si lattice in the [110] zone axis orientation. We perform convergent beam electron diffraction (CBED) simulations of Bi dopants in Si placed at various depths, in a 100 Å thick specimen for 200 kV incident electrons and a 30 mrad. probe forming aperture. To account for temporal incoherence CBEDs are simulated in a defocus range of 40 to 140 Å resulting in incoherent CBEDs in a 0-100 Å range.

We wish to answer the questions (1) how can we distinguish between the two different depths, and (2) which probe position and focal length is optimal for maximizing our distinguishability? The probe positions simulated are indicated with green crosses in Fig. 1(a). Since the CBED patterns should be dependent on the depth of the Bi within the Si column, but it is unclear exactly how so, we may perform a simple statistical analysis. Principal component analysis (PCA) is a widely deployed statistical method that performs linear decomposition with constraints of orthogonal components and ordering based on accounted variance, with the first component accounting for the most variance, the second component accounting for the second most, and so on. Before applying PCA, we first flatten the datasets into 2D via flattening of the spatial pixels, i.e., we reshape the (x,y,f) dataset (where x,y are spatial positions and f is the focal length) to an (xy,f) dataset. The flattened dataset for one probe position and one Bi dopant depth is appended to the flattened dataset for the same probe position for another Bi dopant depth, so the resulting dataset is still a 2D matrix.

On this dataset, we perform PCA, with the results shown in Fig. 1(b) for the position marked X_2 in Fig. 1(a). The decomposition results in both eigenvectors (effectively, images), which are plotted on the left, as well as how they vary with focal length (eigenvalues) on the right. We plot the results for the two distinct depths separately on the same plot so that differences in these eigenvalues can be easily observed. The key point here is that both datasets are represented by the same eigenvectors; only the eigenvalues differ, i.e. the relative spectral weights. For the first five PCA components, all focal points appear to show similar eigenvalues regardless of the depth of the Bi dopant. The distinguishing component appears to be the sixth, where there are substantial differences at a focal length of between 50-90 Å. Since the sixth component represents less than 5% of the variance of the dataset, this suggests that distinguishability will

https://doi.org/10.1017/S1431927619001363 Published online by Cambridge University Press

rely on fine structure features. Moreover, the component (or more accurately, the explained variance) at which the components diverge for the two depths can be used as an information-theoretic measure of the distinguishability.

We performed the same test on the CBED simulations in position X_4, which is further away from the doped column. CBEDs include information that is both highly local but can also be sensitive to non-local structural details. The PCA analysis for this case is shown in Fig. 1(c), and in this case the PCA component at which the two depths of the Bi dopant can be distinguished is component 7. This component accounts for less than 3% of the total variance, indicating that, at least in terms of linear separability, it is more difficult to do so when the CBEDs are captured at position X_4 than X_2.

In addition to providing a simple measure of distinguishability, the method presented here can be useful for both understanding the differences in CBED patterns as a function of small structural or chemical changes, and can likely be extended to nonlinear unmixing, e.g. via kernel methods [3].

References:

[1] G Morely et al., Nat. Mater. **9** (2010), p. 725.

[2] P Midgley and J Thomas, Angewandte Chem. **53** (2014), p. 8614.

[3] This work was supported by the U.S. Department of Energy, Office of Science, Materials Sciences and Engineering Division (MPO, SVK, RKV). Research was conducted at the Center for Nanophase Materials Sciences, which is a US DOE Office of Science User Facility.

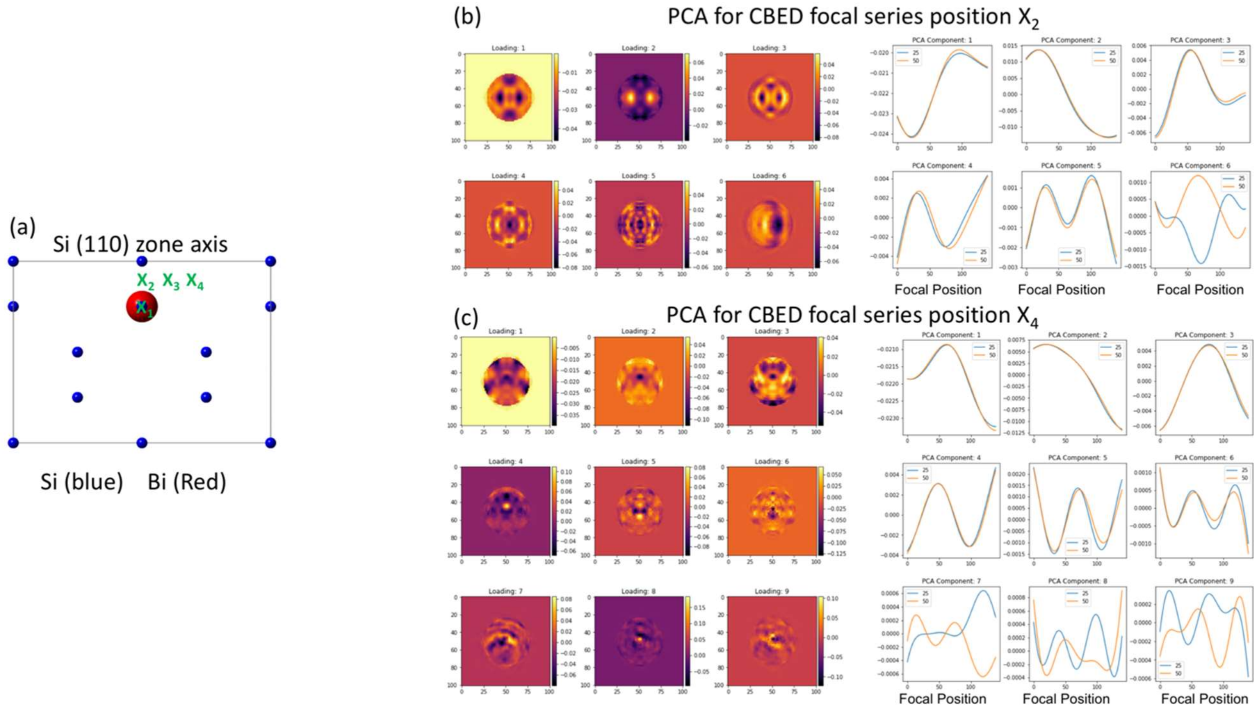

Figure 1. Principal Component Analysis as a tool for measuring distinguishability of CBED patterns. Here, simulations were performed for different probe positions and focal lengths for two distinct Bi dopant depths within a Si column. The schematic of the structure is shown in (a), note the size of Bi is greatly exaggerated. (b) PCA results for CBED focal series taken at position X_2. (c) PCA results for CBED focal series taken at position X_4.

https://doi.org/10.1017/S1431927619001363 Published online by Cambridge University Press