

Navigating the Nanoworld: Automatic Feature Recognition

Stephen J. Pennycook^{1,2*}, Jiadong Dan^{3,4,5}, Xiaoxu Zhao⁶, Shoucong Ning⁴, Wu Zhou², Qian He⁴, N. Duane Loh^{5,7,8}

¹Dept. of Materials Science and Engineering, University of Tennessee, Knoxville, TN, USA.

²School of Physical Sciences and CAS Key Laboratory of Vacuum Sciences, University of Chinese Academy of Sciences, Beijing, China.

³Graduate School for Integrative Sciences and Engineering, National University of Singapore.

⁴Department of Materials Science and Engineering, National University of Singapore.

⁵NUS Centre for Bioimaging Sciences, National University of Singapore.

⁶School of Materials Science and Engineering, Nanyang Technological University, Singapore.

⁷Department of Physics, National University of Singapore.

⁸Department of Biological Sciences, National University of Singapore.

* Corresponding author: stephen.pennycook@cantab.net

Our ability to navigate materials and classify both atomic structure and defects is key to developing new materials with improved properties, such as piezoelectrics [1], thermoelectrics [2] and 2D materials [3]. So far this process has been mostly human based, relying on our finding the critical defects and deciding which are important ourselves. However, machine learning is beginning to enable automatic classification of defects and structures without human supervision or the need for training datasets. Such capability could lead to improved analysis of structural evolution during in-situ experiments, or even to a self-driving microscope in the future.

We have developed a promising unsupervised approach to defect classification based on Zernike polynomial analysis of image patches, which is able to automatically find and classify atomic defects in a variety of materials. It is highly robust to noise, and separates different features more successfully than other image analysis methods such as t-SNE or UMAP.

Fig. 1a shows the first 15 terms of the Zernike polynomial basis set, into which image patches are decomposed. The Zernike polynomials are especially sensitive to the distribution of neighboring atoms, which provides their effectiveness in classifying atomic-scale defects. Fig. 1b,c shows how the image is first split into patches entered on symmetry points, such as atom columns, and then decomposed into Zernike components, Fig. 1d. This can be done either with or without sensitivity to orientation (rotational invariance on or off). The 16,384 pixels of the each original image patch have been replaced by 66 Zernike elements. Further dimension reduction is by principal component analysis, Fig. 2a,b, followed by a second force relaxed clustering approach using k-nearest neighbors, Fig 2b,c. This stage greatly improves cluster separation.

Fig. 3 shows this process applied to an image of monolayer WS₂ which identifies several different point defect types, Fe substituting for W, Te substituting for S, S next to Fe and Te next to W, denoted #, and single S vacancies [4]. The patches identified as the same defect can then be averaged to remove noise, Fig. 3c, and compared to simulations to confirm the identity of the point defect, Fig. 3d.

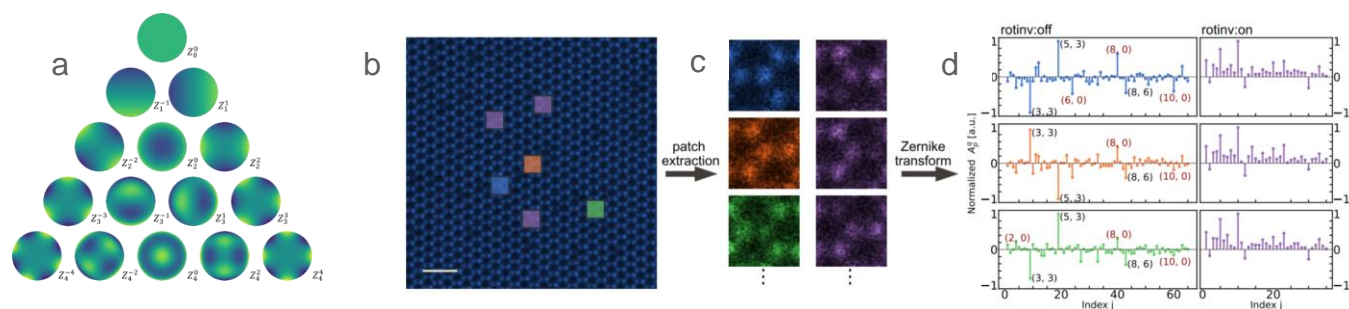


Figure 1. (a) The first 15 Zernike polynomials, (b) Z-contrast image of monolayer MoSe₂, (c) patch extraction and (d) expansion into Zernike coefficients.

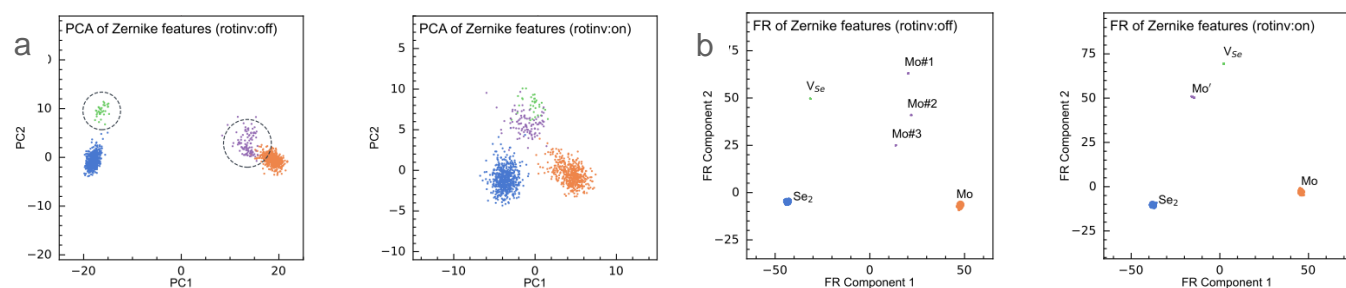


Figure 2. (a) Principle component analysis of the Zernike components followed by (b) a force relaxed clustering approach.

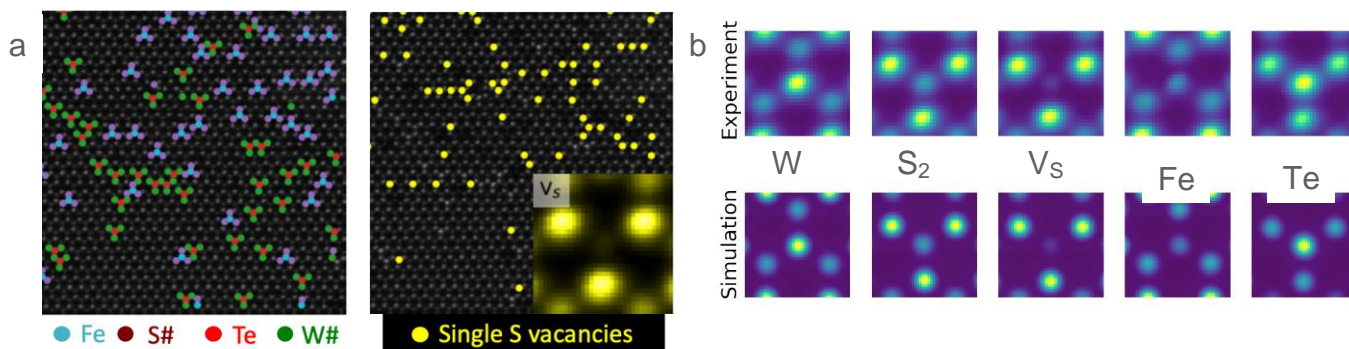


Figure 3. (a) Classification into different point defect types. (b) Averaged images of different point defects compared to simulations.

References:

- [1] H. Liu, et al., *Science* **369**, 292 (2020).
- [2] W. He, et al., *Science*, 365, 1418–1424 (2019).
- [3] X. Zhao, et al., *Nature*, **581**, 171 (2020).
- [4] J. Dan, et al., *Science Advances*, accepted (2022).
- [5] S. J. P. acknowledges funding from Singapore Ministry of Education Tier 1 grant R-284-000-172-114, Tier 2 grant R-284-000-175-112. N.D.L acknowledges funding support from the Singapore National Research Foundation (grant number NRF-CRP16-2015-05) and a NUS Early Career award (A-0004744-00-00).