

Phase determination from atomically resolved images: physics-constrained deep data analysis through an unmixing approach

R. K. Vasudevan^{1,2}, M. Ziatdinov^{1,2}, S. Jesse^{1,2} and S. V. Kalinin^{1,2}

¹ Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge TN 37831 USA

² Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge TN 37831 USA

It is now commonplace to obtain atomically-resolved, real-space images on materials through both scanning tunneling microscopy as well as scanning transmission electron microscopy, and such data is invaluable to determining a host of interactions of order parameters with defects, observe surface reconstructions and nanoscale phase separations, determine the atomic structure of interfaces such as grain boundaries, dislocation cores and domain walls, etc. However, analysis techniques that seek to automate the phase determination and identification remain absent: in fact, it is already becoming impossible to manually identify and analyze all collected images from a single sitting on a microscope, which can yield several ~Gb worth of image data. As such, the need for such techniques is paramount, and here we present our approach towards tackling this problem

We present a method based on the synergy between sliding window fast Fourier transforms (FFT) [1] with endmember extraction based on the N-FINDR algorithm. By sliding a window of fixed size over the image and capturing the FFT at each site, we generate a 2D-stack of FFT spectra, and recast the problem as an unmixing problem similar to the determination of structure factors from powder diffraction. We apply techniques that have been used for unmixing hyperspectral imaging data, specifically, the N-FINDR algorithm.

As an example, in Fig. 1 we show a topographic scanning tunneling microscopy image (1024px x 1024px) of a mixed-terminated surface of $\text{La}_{5/8}\text{Ca}_{3/8}\text{MnO}_3$ (LCMO). The sliding FFT method was used to generate a stack of 2D FFTs with a window size of 64px, and a step size of 8px. The 2D stack was then unmixed using the N-FINDR algorithm, which is described in detail elsewhere [2]. Briefly, it is part of a family of geometric endmember determination algorithms, and seeks to determine the endmembers by computing and maximizing the volume of the simplex formed by a random selection of points within the dataset. The algorithm is run iteratively until the maximum volume is achieved. It should be noted that the method requires the existence of pure endmembers, i.e., physically the window size must be smaller than the characteristic size of the domains present (but larger than unit cell size). The results of this analysis for the LCMO image are shown in Fig. 1(b), with the number of endmembers set to 2. The overlay of abundance map 1 with the topography image, shown in (c), confirms the utility of this approach. We can also compute a metric termed ‘overlap’ between abundance maps (values of 1 are ideal, meaning no overlap, shown in Fig. 2(a)), and determine its variation over the window size and number of components, to help determine the ideal numbers for both parameters. Additionally, the errors in the abundance maps, calculated for each endmember pixel by subtracting the sum of the spectral weights for all endmembers from its own spectral weight can be computed and are shown in Fig. 2(b). It is apparent that larger number of endmembers for this example only led to poorer values for the overlap, as well as for the error maps. Overall, these results suggest that the method described above is highly useful for automatically determining the types of phases present, and can form a framework towards realization of automatic structure-property databases or materials ‘genomic libraries’.

References:

- [1] RK Vasudevan, A Tselev, AP Baddorf, SV Kalinin and S Jesse, *Appl. Phys. Lett.* 106 (2015), p. 091601.
- [2] ME Winter, SPIE's International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics (1999), p. 266.
- [3] This research was sponsored by the Division of Materials Sciences and Engineering, BES, DOE (RKV, MZ, SVK). Research was conducted at the Center for Nanophase Materials Sciences, which also provided support (SJ) and is a DOE Office of Science User Facility.

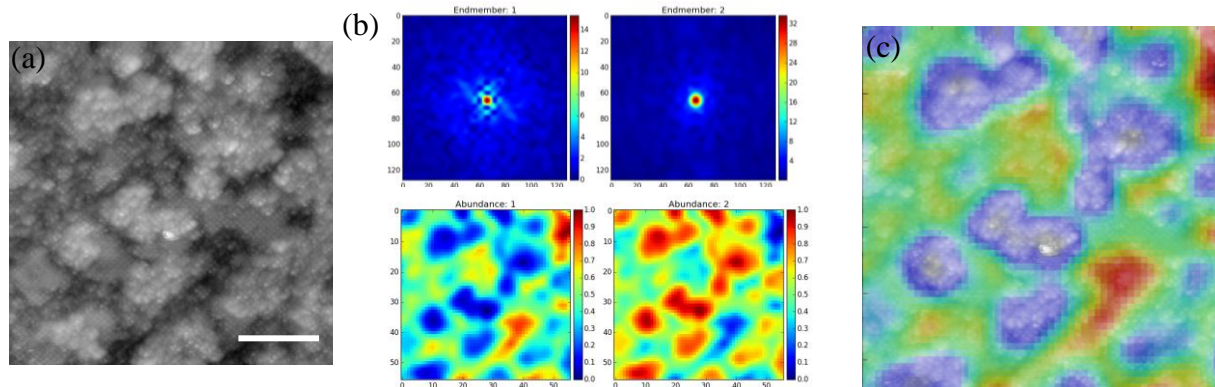


Figure 1. (a) STM topography image of mixed-terminated surface of $\text{La}_{5/8}\text{Ca}_{3/8}\text{MnO}_3$ film grown on (001) SrTiO_3 (scale bar, 10nm). (b) N-FINDR analysis of sliding FFT data, with window size 64 pixels, window step of 8pixels. The endmembers are shown above, and the abundance maps are shown below. (c) Overlay of abundance map 1 on the topography image, showing clearly the isolation of the two terminations by the Sliding FFT-NFINDR method.

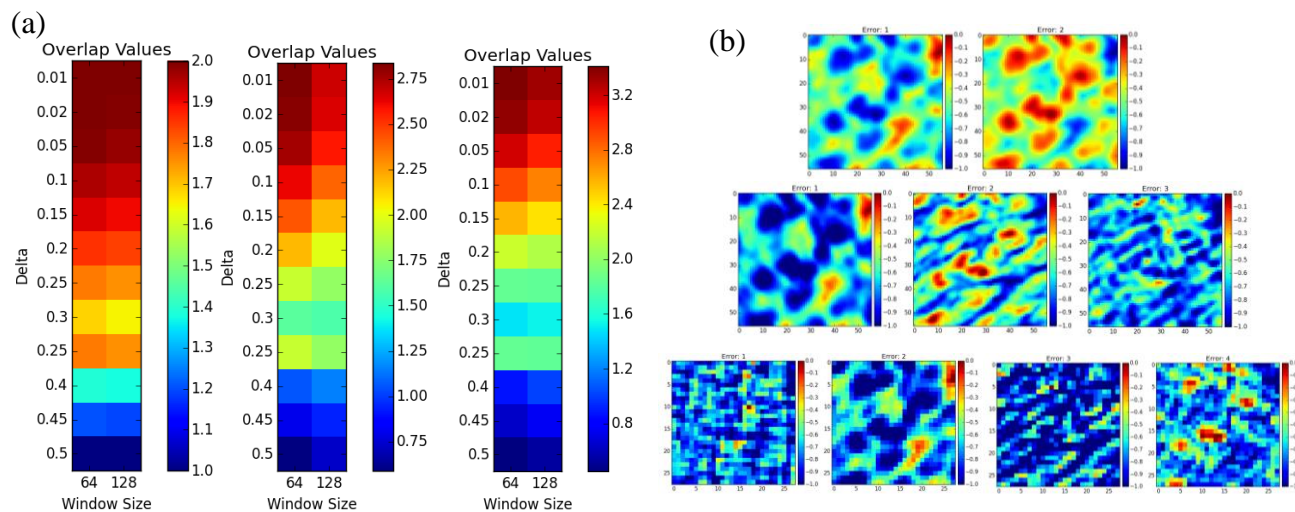


Figure 2. (a) Overlap values for 2, 3 and 4 endmembers, respectively for different threshold values Δ . (b) Error maps for 2, 3 and 4 endmember choices, respectively. Error values should be 0 or -1 in the ideal case; values in between suggest poor separation. The separation becomes worse with increasing choice of number of endmembers, suggesting 2 endmembers is the optimum for this image.