JOURNAL OF SYNCHROTRON RADIATION

ISSN 1600-5775

Received 14 May 2020
Accepted 13 October 2020

Edited by Y. Amemiya, University of Tokyo, Japan

Keywords: hard X-ray nanodiffraction; cluster analysis; thin films; complex oxides.

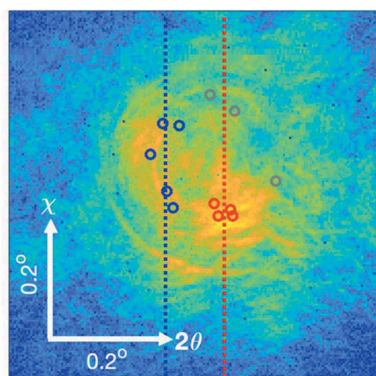
Supporting information: this article has supporting information at journals.iucr.org/s

Understanding nanoscale structural distortions in $Pb(Zr_{0.2}Ti_{0.8})O_3$ by utilizing X-ray nanodiffraction and clustering algorithm analysis

Joyce Christiansen-Salameh,[a] Morris Yang,[a] Geoffrey Rippy,[a] Jianheng Li,[a] Zhonghou Cai,[b,c] Martin Holt,[b] Guillaume Agnus,[d] Thomas Maroutian,[d] Philippe Lecoeur,[d] Sylvia Matzen[d] and Roopali Kukreja[a]*

[a]Department of Materials Science and Engineering, University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA, [b]Center for Nanoscale Materials, Argonne National Laboratory, Lemont, IL 60439, USA, [c]Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, USA, and [d]Center for Nanoscience and Nanotechnology (C2N), CNRS, Université Paris-Saclay, 91120 Palaiseau, France. *Correspondence e-mail: rkukreja@ucdavis.edu

Hard X-ray nanodiffraction provides a unique nondestructive technique to quantify local strain and structural inhomogeneities at nanometer length scales. However, sample mosaicity and phase separation can result in a complex diffraction pattern that can make it challenging to quantify nanoscale structural distortions. In this work, a *k*-means clustering algorithm was utilized to identify local maxima of intensity by partitioning diffraction data in a three-dimensional feature space of detector coordinates and intensity. This technique has been applied to X-ray nanodiffraction measurements of a patterned ferroelectric $PbZr_{0.2}Ti_{0.8}O_3$ sample. The analysis reveals the presence of two phases in the sample with different lattice parameters. A highly heterogeneous distribution of lattice parameters with a variation of 0.02 Å was also observed within one ferroelectric domain. This approach provides a nanoscale survey of subtle structural distortions as well as phase separation in ferroelectric domains in a patterned sample.



© 2021 International Union of Crystallography

1. Introduction

In the last few decades, hard X-ray microscopy has emerged as a novel method to characterize nanoscale heterogeneities in functional oxides (Holt *et al.*, 2013; Logan *et al.*, 2006; Hruszkewycz *et al.*, 2013, 2014; Winarski *et al.*, 2012). Specifically, hard X-ray microscopy techniques have shown certain advantages over complementary electron microscopy techniques for functional oxides, as no alterations of local boundary conditions occur during sample preparation or measurement due to comparatively weak sample–beam interaction. Hard X-ray nanodiffraction, which combines the benefits of diffraction and imaging by using a zone plate to focus the X-ray beam to a spot size of 25 nm, has been utilized to investigate strain fields and structural heterogeneities at the nanoscale in a variety of functional oxides systems (Jo *et al.*, 2011; Rippy *et al.*, 2019; Pagliero *et al.*, 2014; Zhu *et al.*, 2016). However, the majority of these studies have been restricted to epitaxial oxide thin films with negligible mosaicity, as current analysis methods for generating local strain maps rely on mapping the variation of a single Bragg peak over a specific sample region (Jo *et al.*, 2011; Rippy *et al.*, 2019). Sample mosaicity and heterogeneities present in a sample can result in a complex diffraction pattern overlaid with the zone plate diffraction pattern and can comprise of more than one local intensity maxima as the X-ray beam diffracts from multiple strain states across several grains along the sample depth. In

order to separate these multiple effects, and identify the role of local strain on microscopic functional properties, robust analysis methods for complex diffraction patterns are critically needed.

In this article, we present a *k*-means clustering algorithm based analysis of complex diffraction patterns which extends the capabilities of X-ray nanodiffraction to highly mosaic and heterogeneous samples. Cluster analysis is based on self-consistently identifying and classifying common elements within a dataset (diffraction patterns in this case), and has been widely used for image segmentation and quantitative image analysis (Zhao & Karypis, 2005; Nugent & Meila, 2010; Hu *et al.*, 2015; Fraix-Burnet *et al.*, 2012; Sánchez Almeida *et al.*, 2012, 2010). *k*-means is a clustering algorithm whereby the centroids of a dataset are found by iteratively alternating between two steps which include assigning data points to clusters based on the present centroids, and then choosing the next iteration of centroids based on the present assignment of data points to the cluster. The *k*-means algorithm is used in medical imaging, life sciences, astronomy and astrophysics to segment superpixels, or sets of pixels with similar attributes, from an image background (MacQueen, 1967; Jain & Dubes, 1998; Simpson *et al.*, 2012; Balazs *et al.*, 1996; Ordovás-Pascual & Sánchez Almeida, 2014; Haralick & Robert, 1985; Juang & Wu, 2010; Zheng *et al.*, 2018). We utilized the *k*-means algorithm to identify local maxima of intensity by partitioning the diffraction data within a region of interest (ROI) in a 3D feature space of detector coordinates and intensity. This enabled us to ascertain centroids present in complex nanodiffraction patterns, which were then used to obtain local lattice parameters and strain.

This method is applied here to X-ray nanodiffraction measurements of a patterned $Pb(Zr_{0.20}Ti_{0.80})O_3$ (PZT) sample. PZT is an archetypical piezoelectric and ferroelectric material whereby the polarization state is coupled with lattice strain and can be controlled by application of electric field. Ferroelectric materials such as PZT show a thermodynamically stable and electrically switchable polarization (Scott, 2006, 2007; Dawber *et al.*, 2005), which is of interest in applications such as sensor and actuator devices for microelectromechanical systems (MEMS) (Setter *et al.*, 2006; Trolier-McKinstry & Muralt, 2004) and nonvolatile memories (Hoffman *et al.*, 2010). However, at the nanoscale the structural distortions can be heterogeneous, and local strain fields often couple strongly with functional properties (Sulpizio *et al.*, 2014). While piezoresponse force microscopy (PFM) has been utilized to manipulate and probe ferroelectric polarity in these systems, it provides relatively little information about the atomic structure of the ferroelectric domains, especially as the information obtained is limited to the surface of the sample. The structural interpretation of PFM data is further complicated by the complex interaction between the probe tip and surface of ferroelectric material (Jo *et al.*, 2011). Transmission electron microscopy (TEM) has also been used to map out atomic structure and ferroelectric polarization; however, it is restricted to extremely thin samples and sample preparation can be challenging and destructive (Gao *et al.*, 2014; Mundy *et al.*, 2017; Agar *et al.*, 2016). In order to map out the local strain variation non-destructively, we performed X-ray nanodiffraction studies on patterned PZT samples. Complex diffraction patterns obtained from PZT samples were then analyzed using clustering algorithms to quantify local strain in ferroelectric domains. Our analysis shows the presence of two distinct phases with different lattice parameters distributed across the film thickness. One of the phases shows a clear distinction in intensity and lattice parameter consistent with up and down ferroelectric domains. The second phase shows a relatively homogeneous distribution of intensity, which could be attributed to the structural response to unscreened charges at the surface or the interface.

2. Methods

X-ray nanodiffraction experiments were conducted at the Hard X-ray Nanoprobe (26-ID-C) beamline of the Advanced Photon Source and Center for Nanoscale Materials at the Argonne National Laboratory. Additional details about the beamline and nanodiffraction technique can be found in the literature (Holt *et al.*, 2013; Hruszkewycz *et al.*, 2014, 2015; Winarski *et al.*, 2012). The monochromatic incident X-ray beam (9 keV) was focused to a 25 nm spot size with a Fresnel zone plate, and an order sorting aperture and center stop blocked all but the first-order beam. The focusing led to an overall effective beam divergence of 0.24°. This first-order diffracted X-ray beam was utilized to perform X-ray nanodiffraction studies in reflective geometry. A two-dimensional pixel area detector (PAD) with a 55 µm pixel size placed 860 cm from the sample was used to collect the diffraction pattern. The detector was aligned to a 2θ value of 38.4° to access the (002) Bragg peak.

A patterned PZT thin film sample grown using pulsed laser deposition on an $La_{0.7}Sr_{0.3}MnO_3$ (LSMO)-buffered $SrTiO_3$ (STO) substrate was utilized. LSMO and PZT layers were successively grown by pulsed laser deposition (KrF excimer laser, 248 nm wavelength) on a single-crystal, TiO_2-terminated STO (001) substrate held at a temperature of 630°C. The laser was operated at a pulse rate of 4 Hz, with a target-to-substrate distance of 5 cm and a fluence of 2.5 J cm^{-2}. The PZT ceramic target was sintered with a stoichiometric Zr/Ti ratio and 12% excess PbO. The 30 nm-thick LSMO bottom electrode was first deposited under an oxygen pressure of 120 mTorr, then replaced by nitrous oxide (N_2O) at the same pressure to grow a 100 nm-thick PZT layer. After deposition, the sample was cooled to room temperature under a static oxygen pressure of 300 Torr. A rocking curve [around the (002) PZT diffraction Bragg peak] is presented in the supporting information to show the averaged mosaicity of the as-grown PZT film. Electron beam lithography was then performed to pattern 2 µm by 2 µm islands of PZT. The ferroelectric polarization of the PZT was measured and manipulated by PFM. The as-grown polarization pointed upwards (away from the bottom LSMO electrode). Fig. 1(*a*) presents a PFM image of one PZT island showing a two-domain state with an up and a down domain written by applying −5 V and +5 V, respectively, to the

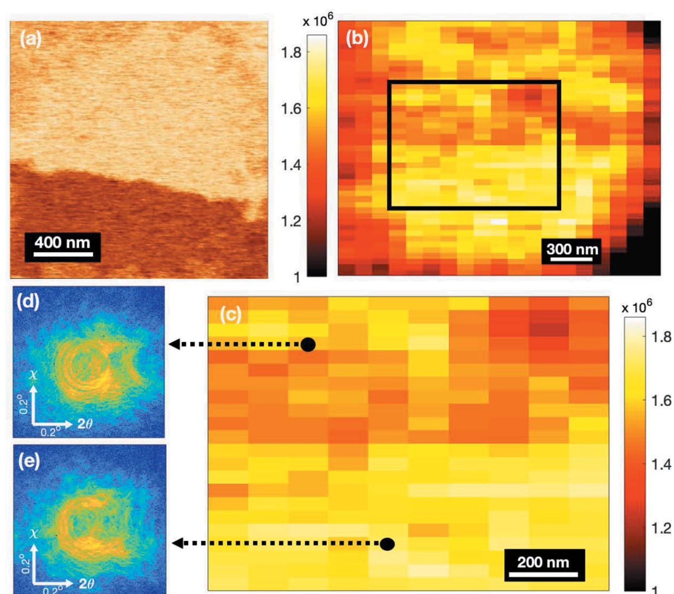


Figure 1
(*a*) PFM phase image of one PZT 2 μm × 2 μm island showing the up (light brown) and down (dark brown) domains, (*b*) nanodiffraction map over the same region as (*a*) using integrated intensity of the diffraction pattern obtained near the (002) Bragg peak. A clear difference in intensity is observed between the top and bottom of the image due to the up and down domains. (*c*) Nanodiffraction map of a selected region in the box which was used for further analysis, (*d*) and (*e*) detector images showing the diffraction patterns for the up and down domains, respectively.

tip. The top of the image corresponds to the up polarization domain (light brown) and the bottom of the image corresponds to the down polarization domain (dark brown).

3. X-ray nanodiffraction measurements of a micro-patterned PZT thin film

Nanodiffraction maps of a PZT sample were collected by raster scanning the X-ray beam and acquiring a diffraction pattern for each sample position. Fig. 1(*b*) shows a nanodiffraction map of the PZT sample, where the integrated intensity of the (002) PZT Bragg peak is plotted as a function of sample position. The intensity difference between the top and the bottom of the image is caused by oppositely polarized domains, which results in a 30% variation of the diffracted intensity due to lack of centrosymmetry of the tetragonal perovskite unit cell (Jo *et al.*, 2011; Do *et al.*, 2004; Grigoriev *et al.*, 2006). This intensity variation corresponds to the up and down domain seen in the PFM image [Fig. 1(*a*)] for the same sample region. However, while the PFM image shows only slight nonuniformity in overall up and down written domains, the nanodiffraction map reveals a significant variation in the diffracted intensity, highlighting potential inhomogeneities within a single domain.

Fig. 1(*c*) shows a zoomed-in version of the nanodiffraction map used for further data analysis. Figs. 1(*d*) and 1(*e*) show detector images of the diffraction pattern for different areas on the sample. The total intensity of such a diffraction image was used to plot the nanodiffraction maps shown in Figs. 1(*b*) and 1(*c*). Here we note that the concentric ring pattern or the annulus observed in the diffraction pattern arises from the Fresnel zone plate and center stop which are used to focus the X-rays and block all but the first-order beam. The spread of intensities across the annulus in the diffraction pattern is caused by the presence of several mosaic blocks in the sample. The contribution from each mosaic block is difficult to distinguish using simple descriptive statistics (*e.g.* the center of mass of the intensity) which have been previously used for nanodiffraction studies of epitaxial films (Jo *et al.*, 2011; Rippy *et al.*, 2019) to characterize the local variation of strain. Although the total intensity of the diffraction pattern can be used to differentiate between the two domains, the complex diffraction pattern makes it challenging to accurately quantify the lattice parameter of the up and down domains as well as its local variation within a single domain. Section S1 of the supporting information provides such an analysis where a typical centroid calculation method for nanodiffraction datasets for epitaxial samples (Holt *et al.*, 2013; Jo *et al.*, 2011; Rippy *et al.*, 2019) is applied to the PZT sample. However, it can be seen from Fig. S2(*b*) of the supporting information that the lattice parameters for the up and the down domains show no clear difference, highlighting that such a method can easily miss the nuances of real-world mosaic samples. Failure of the centroid calculation method for a mosaic sample can be attributed to: averaging of the peak position if multiple peaks are present or completely missing the intensity maxima within the ROI typically used for nanodiffraction analysis, or dependency on the partial overlap of the chosen ROI with the annular diffraction pattern.

4. *k*-means cluster analysis

The *k*-means algorithm (MacQueen, 1967; Jain & Dubes, 1998; Simpson *et al.*, 2012; Balazs *et al.*, 1996; Ordovás-Pascual & Sánchez Almeida, 2014; Haralick & Robert, 1985; Juang & Wu, 2010; Zheng *et al.*, 2018) was utilized to negate these effects and partition the diffraction data within each ROI in 3D feature space of detector coordinates (horizontal axis – 2θ, vertical axis – χ) and intensity in order to isolate high-intensity regions that are uniquely distributed in each diffraction pattern. Multiple ROIs were selected to cover the diffraction pattern in different ways so that the new approach is independent of the ROI selected (see Fig. S3).

Fig. 2 outlines the process of applying the *k*-means algorithm (available in *MATLAB*) (MathWorks, 2019; Lloyd, 1982) to discern a local maximum within one such ROI. Fig. 2(*a*) shows the ROI along with the three initial guesses for centroids which were used to encourage clustering of high- and low-intensity data points. Fig. 2(*b*) shows the resulting three clusters obtained upon convergence of the algorithm based on iteratively minimizing the Euclidean distance between cluster points and cluster centroids in the 3D feature space of the detector 2θ, χ and intensity. Here we stress that the separation of high-intensity data points into cluster 3 (red

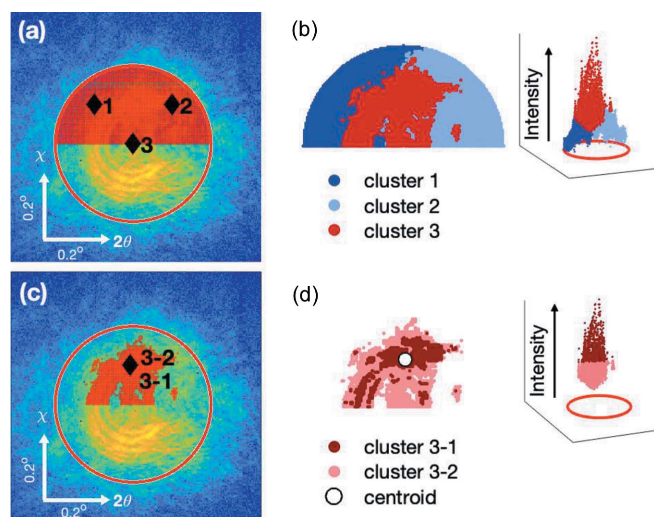


Figure 2
Cluster algorithm for identifying a local maximum within an ROI: (*a*) initial guesses for clusters within the ROI; (*b*) resulting clusters and segmentation of the diffraction data; (*c*) further refinement of the highest intensity cluster obtained in (*b*) using the cluster algorithm, the initial guesses are also shown; (*d*) resulting clusters and further segmentation of the data. Centroid of cluster 3-1 is the local maxima of the ROI.

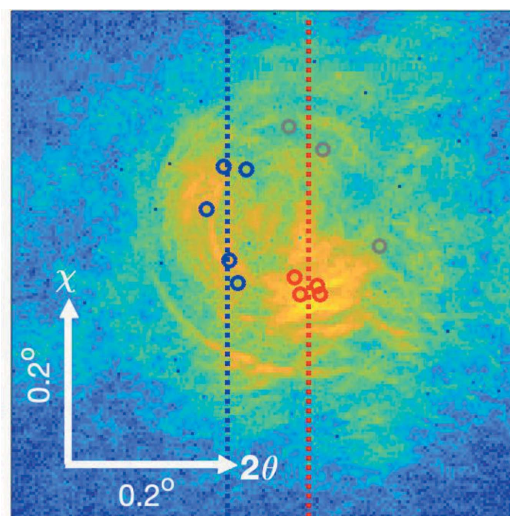


Figure 3
Local maxima (or centroids) obtained for twelve different ROIs. Centroids below 40% intensity threshold are discarded (gray circles). Centroids above this threshold can be grouped in two groups (blue and red) using *k*-means. The *x* location of these two groups is shown using red and blue dashed lines.

distribution) is not equivalent to setting an intensity threshold for the diffraction pattern, as the cluster is determined by both spatial (detector coordinates) and intensity metrics. This allows us to locate the centroid for clusters taking into account both the distance between the pixels and their intensity. In order to further refine the high-intensity dataset, cluster 3 is further partitioned into two additional clusters using the same algorithm. Fig. 2(*c*) shows the initial guess for centroids, again to encourage clustering of high- and low-intensity data points. Fig. 2(*d*) shows the resulting two clusters (3-1 and 3-2), where the centroid of cluster 3-2 marks the local maximum of the ROI. This two-stage implementation of *k*-means first separates a superpixel (or a cluster) of general high-intensity signal from the background of the diffraction pattern, and then in the second step, calculates a centroid using the high-intensity points that maintain a spatial relationship within this superpixel. This two-step method enables identification of the centroid of the highest intensity cluster which would be associated with the Bragg peak following diffraction from the PZT sample.

This method is then applied for all the ROIs to generate a set of twelve centroids (for twelve ROIs, see Fig. S3) denoting local maxima of a diffraction pattern. It can be seen that the centroids are not randomly distributed over the diffraction pattern, but cluster around two higher intensity regions. In order to obtain the final centroid position of these two higher intensity regions, the twelve centroids were grouped based on the horizontal distance (along 2θ) using the *k*-means algorithm. Only the horizontal distance along the 2θ direction was used, as it is directly related to *d*-spacing via Bragg's law, thus allowing calculation of the lattice parameter. Two final centroid positions (the *x* value only, corresponding to a specific 2θ) are obtained corresponding to the local maxima of two higher intensity regions, and are shown in Fig. 3 by red and blue dashed lines. The centroids for ROIs below a certain intensity threshold were neglected in this calculation, as they do not contain local maxima (*i.e.* the gray dots in the Fig. 3).

The analysis described in Figs. 2 and 3 was applied for every diffraction pattern obtained for each sample position shown in Fig. 1(*c*). In most of the diffraction patterns (>80%) the presence of two final centroids was observed, highlighting the importance of this approach. Fig. 4(*a*) plots a histogram of both centroids for every diffraction pattern obtained for each sample position. The *x* axis represents the position of the centroid (along 2θ) in the pixel value away from the center of the annulus (which is taken as reference and marked as 0 pixel), and the *y* axis represents the number of centroids. The histogram clearly shows a bimodal distribution, labeled 1 and 2, indicating that the centroid positions (*i.e.* intensity maxima) were primarily localized near two regions on opposite sides of the annulus. The bimodal distribution for the intensity maxima clearly point towards two different values for the *d*-spacing and the lattice parameter indicating the presence of two phases. Here we stress that two values for the lattice parameters were observed for almost all of the measured sample positions (>80%), implying that two phases are present across the entire sample. These two phases have slightly different lattice parameters resulting in two different 2θ values within a diffraction pattern identified by the *k*-means clustering algorithm.

For these two phases, diffracted intensity was separately plotted in Figs. 4(*b*) and 4(*c*) using the distributions 1 and 2, respectively. The intensity map for distribution 1 aligns with the total intensity map shown in Fig. 1(*c*) and polarization shown by the PFM image in Fig. 1(*a*), indicating that this is

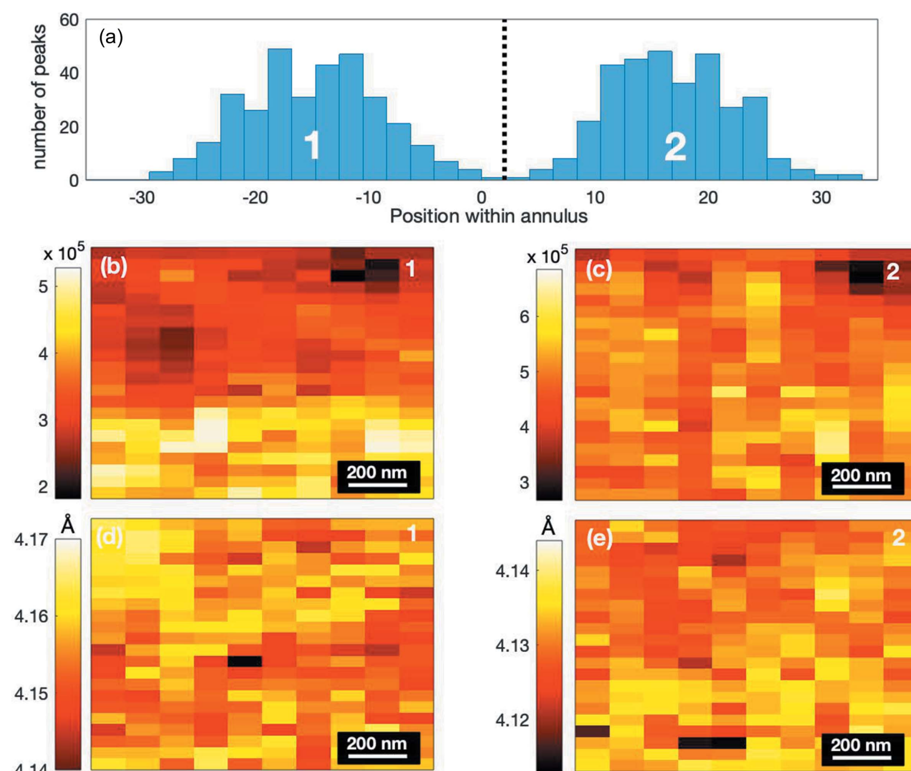


Figure 4
(*a*) Histogram of the local maxima location for all sample positions showing a bimodal distribution, (*b*) and (*c*) integrated intensity maps for distributions 1 and 2, respectively, and (*d*) and (*e*) lattice parameter maps for distributions 1 and 2, respectively.

indeed the ferroelectric phase with two oppositely polarized domains. These domains produce Bragg reflections at the same angle but with opposite sign of crystallographic indices, *i.e.* (002) and ($00\bar{2}$), and with different intensities (Jo *et al.*, 2011). It is interesting to note that no such demarcation between the two domains is observed for the intensity map for distribution 2. Figs. 4(*d*) and 4(*e*) show the lattice parameter map for the two phases calculated using the obtained centroid values for distributions 1 and 2, respectively. For distribution 1, the lattice parameters for the two oppositely polarized domains are 4.157 ± 0.004 Å and 4.153 ± 0.004 Å. The lattice parameter variation observed within each domain is approximately 0.02 Å. For the phase associated with distribution 1, we thus observe higher diffracted intensity and a lower lattice parameter for the polarization down domain compared with the polarization up domain, which is consistent with the literature (Jo *et al.*, 2011). Surprisingly, a small variation in the lattice parameter is observed for the phase associated with distribution 2, in spite of no clear demarcation in the intensity. The lattice parameters in this case are 4.132 ± 0.004 Å and 4.129 ± 0.004 Å and the lattice parameter variation observed within each domain is approximately 0.02 Å. It is interesting to note that for phase 1 the top (bottom) part of the image has a higher (lower) lattice parameter, whereas for phase 2 this contrast is reversed and the top (bottom) part of the image has a lower (higher) lattice parameter. In addition, the lattice parameter distributions in both phases show a variation of 0.02 Å within one domain, highlighting the non-homogeneous nature of these domains.

5. Discussion

The *k*-means clustering based analysis algorithm enables us to identify two phases present in a patterned PZT sample. This identification of two phases in a mosaic and complex sample cannot be achieved using traditional centroid calculation methods. Two phases in the patterned sample correspond to slightly different lattice parameters: 4.155 Å (phase 1) and 4.131 Å (phase 2). For phase 1, a clear demarcation of intensity and lattice parameter distribution is observed between the up and the down domains consistent with the polarized ferroelectric domains observed in PFM [Fig. 1(*a*)]. On the other hand, for phase 2, the intensity variation is relatively homogenous across the top and the bottom of the image, which could imply that either this second phase is not ferroelectric in nature or does not switch between two domains. Surprisingly, lattice parameter variation across the top and the bottom of the image is still observed. It is also interesting to note that the lattice parameter contrast is flipped between the two phases, *i.e.* for phase 1 the up domain has a higher *c* lattice parameter while phase 2 shows a lower *c* lattice parameter.

The variation in the lattice parameter across the film is more likely to occur near the PZT surface or PZT/LSMO interface in comparison with bulk, due to different electro-mechanical boundary conditions that can modify ferroelectric properties. It is therefore possible that the strain state closer to the film surface or near PZT/LSMO interface is not of purely ferroelectric origin and/or was not switched by PFM resulting in the observed behavior of intensity and lattice parameter distribution. As X-rays penetrate the entire film thickness, it is challenging to discern whether this phase 2 is formed at the surface or the bottom interface. Another explanation for the second phase could be a slightly different composition of PZT resulting in different tetragonality and ferroelectric behavior. In order to identify the exact origin of this second phase, additional characterization techniques such as cross-sectional high-resolution TEM or atomic resolution spectroscopy are required.

The identification and isolation of two different phases with slightly different lattice parameters demonstrates the importance for developing and using the *k*-means clustering method to investigate complex samples which present mosacitiy, phase separation, and nanoscale heterogeneities. Although

the traditional centroid calculation works well for ideal and epitaxial systems, our results show that the clustering algorithm-based approach can be utilized to study complex and heterogeneous samples. Furthermore, our X-ray nanodiffraction measurements also highlight that both strain and intensity distribution show highly heterogeneous behavior, which was not observed in PFM images. Micrometer-scale non-uniformity of strain states within ferroelectric domains has been previously reported in a multiferroic composite that included a PMN–$PbTiO_3$ layer (Lo Conte *et al.*, 2018). In PZT, we observed nanoscale non-uniformity of strain states within both up and down domains with a lattice parameter variation of 0.02 Å within one domain. These measurements demonstrate high sensitivity of X-ray nanodiffraction technique, which is able to measure existence of local strain or mosaicity due to resulting large variations in diffracted intensity.

6. Conclusions

In this article, we have utilized *k*-means cluster analysis for nanodiffraction measurements to investigate nanoscale morphology including phase separation and mosaicity in a patterned PZT sample. The use of a clustering algorithm overcomes the limitations of previous methods and allows us to identify multiple distributions resulting in the presence of two phases with different lattice parameters. While one phase shows a clear difference in intensity and lattice parameter distribution consistent with PFM images indicating ferroelectric domains, the second phase shows a relatively homogenous intensity distribution, and could be present near the surface or the PZT/LSMO interface due to electro-mechanical boundary conditions at surface and interfaces. This method provides a unique nondestructive approach for quantifying subtle structural distortions in ferroelectric domains. Local nanoscale structural effects probed by this method can have a profound impact on the manipulation of domains in ferroelectric based devices. This approach can be also extended to other complex oxides, and can be utilized to understand the effects of local strain and phase separation on strongly inter-coupled electronic, optical and structural properties at nanometer-length scales.

Funding information

The work at UC Davis is supported by National Science Foundation (DMR-1902652). This research used resources from the Advanced Photon Source and the Center for Nanoscale Materials, both US Department of Energy (DOE) Office of Science User Facilities operated for the DOE Office of Science by Argonne National Laboratory (contract No. DE-AC02-06CH11357). Research at the Center of Nanoscience and Nanotechnology was supported by a grant from the DIM OXYMORE (Ile-de-France) and by the RENATECH network.

References

Agar, J. C., Damodaran, A. R., Okatan, M. B., Kacher, J., Gammer, C., Vasudevan, R. K., Pandya, S., Dedon, L. R., Mangalam, R. V. K., Velarde, G. A., Jesse, S., Balke, N., Minor, A. M., Kalinin, S. V. & Martin, L. W. (2016). *Nat. Mater.* **15**, 549–556.

Balazs, L. G., Garibjanyan, A. T., Mirzoyan, L. V., Hambaryan, V. V., Kun, M., Fronto, A. & Keleman, J. (1996). *Astron. Astrophys.* **311**, 145–151.

Dawber, M., Rabe, K. M. & Scott, J. F. (2005). *Rev. Mod. Phys.* **77**, 1083–1130.

Do, D.-H., Evans, P. G., Isaacs, E. D., Kim, D. M., Eom, C. B. & Dufresne, E. M. (2004). *Nat. Mater.* **3**, 365–369.

Fraix-Burnet, D., Chattopadhyay, T., Chattopadhyay, A. K., Davoust, E. & Thuillard, M. (2012). *Astron. Astrophys.* **545**, A80.

Gao, P., Britson, J., Nelson, C. T., Jokisaari, J. R., Duan, C., Trassin, M., Baek, S., Guo, H., Li, L., Wang, Y., Chu, Y., Minor, A. M., Eom, C., Ramesh, R., Chen, L. & Pan, X. (2014). *Nat. Commun.* **5**, 3801.

Grigoriev, A., Do, D.-H., Kim, D. M., Eom, C.-B., Adams, B., Dufresne, E. M. & Evans, P. G. (2006). *Phys. Rev. Lett.* **96**, 187601.

Haralick, M. & Robert, L. (1985). *Comput. Vis. Graph. Image Process.* **29**, 100–132.

Hoffman, J., Pan, X., Reiner, J. W., Walker, F. J., Han, J. P., Ahn, C. H. & Ma, T. P. (2010). *Adv. Mater.* **22**, 2957–2961.

Holt, M., Harder, R., Winarski, R. & Rose, V. (2013). *Annu. Rev. Mater. Res.* **43**, 183–211.

Hruszkewycz, S. O., Highland, M. J., Holt, M. V., Kim, D., Folkman, C. M., Thompson, C., Tripathi, A., Stephenson, G. B., Hong, S. & Fuoss, P. H. (2013). *Phys. Rev. Lett.* **110**, 177601.

Hruszkewycz, S. O., Holt, M. V., Allain, M., Chamard, V., Polvino, S. M., Murray, C. E. & Fuoss, P. H. (2014). *Philos. Trans. Soc. R.* **372**, 20130118.

Hruszkewycz, S. O., Holt, M. V., Allain, M., Chamard, V., Polvino, S. M., Murray, C. E. & Fuoss, P. H. (2015). *Opt. Lett.* **40**, 3241–3244.

Hu, C. W., Kornblau, S. M., Slater, J. H. & Qutub, A. A. (2015). *Sci. Rep.* **5**, 12894.

Jain, A. K. & Dubes, R. C. (1998). *Algorithms for Clustering Data.* Englewood Cliffs, NJ: Prentice Hall.

Jo, J. Y., Chen, P., Sichel, R. J., Baek, S., Smith, R. T., Balke, N., Kalinin, S. V., Holt, M. V., Maser, J., Evans-Lutterodt, K., Eom, C. & Evans, P. G. (2011). *Nano Lett.* **11**, 3080–3084.

Juang, L. & Wu, M. (2010). *Measurement*, **43**, 941–949.

Lloyd, S. (1982). *IEEE Trans. Inf. Theory*, **28**, 129–137.

Lo Conte, R., Xiao, Z., Chen, C., Stan, C. V., Gorchon, J., El-Ghazaly, A., Nowakowski, M. E., Sohn, H., Pattabi, A., Scholl, A., Tamura, N., Sepulveda, A., Carman, G. P., Candler, R. N. & Bokor, J. (2018). *Nano Lett.* **18**, 1952–1961.

Logan, J., Harder, R., Li, L., Haskel, D., Chen, P., Winarski, R., Fuesz, P., Schlagel, D., Vine, D., Benson, C. & McNulty, I. (2006). *J. Synchotron Rad.* **23**, 1210–1215.

MacQueen, J. (1967). *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, Statistics, pp. 281–297. Berkeley: University of California Press.

MathWorks (2019). *k*-means. Available at https://www.mathworks.com/help/stats/k-means.html.

Mundy, J. A., Schaab, J., Kumagai, Y., Cano, A., Stengel, M., Krug, I. P., Gottlob, D. M., Doğanay, H., Holtz, M. E., Held, R., Yan, Z., Bourret, E., Schneider, C. M., Schlom, D. G., Muller, D. A., Ramesh, R., Spaldin, N. A. & Meier, D. (2017). *Nat. Mater.* **16**, 622–627.

Nugent, R. & Meila, M. (2010). *An Overview of Clustering Applied to Molecular Biology* In *Statistical Methods in Molecular Biology*, Vol. 620. Totowa, NJ: Humana Press.

Ordovás-Pascual, I. & Sánchez Almeida, J. (2014). *Astron. Astrophys.* **565**, A53.

Pagliero, A., Mino, L., Borfecchia, E., Truccato, M., Agostino, A., Pascale, L., Enrico, E., Leo, N. D., Lamberti, C. & Martínez-Criado, G. (2014). *Nano Lett.* **14**, 1583–1589.
Rippy, G., Trinh, L., Kane, A. M., Ionin, A. L., Lee, M. S., Chopdekar, R. V., Christiansen-Salameh, J. M., Gilbert, D. A., Grutter, A. J., Murray, P. D., Holt, M. V., Cai, Z., Liu, K., Takamura, Y. & Kukreja, R. (2019). *Phys. Rev. Mater.* **3**, 082001.
Sánchez Almeida, J., Aguerri, J. A. L., Muñoz-Tuñón, C. & de Vicente, A. (2010). *Astrophys. J.* **714**, 487–504.
Sánchez Almeida, J., Terlevich, R., Terlevich, E., Cid Fernandes, R. & Morales-Luis, A. B. (2012). *Astrophys. J.* **756**, 163.
Scott, J. F. (2007). *Science*, **315**, 949–959.
Scott, J. F. (2006). *J. Phys. Condens. Matter*, **18**, 361–386.
Setter, N., Damjanovic, D., Eng, L., Fox, G., Gevorgian, S., Hong, S., Kingon, A., Kohlstedt, H., Park, N. Y., Stephenson, G. B., Stolitchnov, I., Taganstev, A. K., Taylor, D. V., Yamada, T. & Streiffer, S. (2006). *J. Appl. Phys.* **100**, 051606.
Simpson, J. D., Cottrell, P. L. & Worley, C. C. (2012). *Mon. Not. R. Astron. Soc.* **427**, 1153–1167.
Sulpizio, J. A., Ilani, S., Irvin, P. & Levy, J. (2014). *Annu. Rev. Mater. Res.* **44**, 117–149.
Trolier-McKinstry, S. & Muralt, P. (2004). *J. Electroceram.* **12**, 7–17.
Winarski, R. P., Holt, M. V., Rose, V., Fuesz, P., Carbaugh, D., Benson, C., Shu, D., Kline, D., Stephenson, G. B., McNulty, I. & Maser, J. (2012). *J. Synchrotron Rad.* **19**, 1056–1060.
Zhao, Y. & Karypis, G. (2005). *Mol. Biotechnol.* **31**, 55–80.
Zheng, X., Lei, Q., Yao, R., Gong, Y. & Yin, Q. (2018). *Image Video Proc.* **68**, 7578.
Zhu, Y., Cai, Z., Chen, P., Zhang, Q., Highland, M. J., Jung, I. W., Walko, D. A., Dufresne, E. M., Jeong, J., Samant, M. G., Parkin ,S. S. P., Freeland, J. W., Evans, P. G. & Wen, H. (2016). *Sci. Rep.* **6**, 21999.