

## ARTICLE OPEN

## Sequential piezoresponse force microscopy and the ‘small-data’ problem

Harsh Trivedi<sup>1</sup>, Vladimir V. Shvartsman<sup>1</sup>, Marco S. A. Medeiros<sup>2</sup>, Robert C. Pullar<sup>2,3</sup> and Doru C. Lupascu<sup>1</sup>

The term big-data in the context of materials science not only stands for the volume, but also for the heterogeneous nature of the characterization data-sets. This is a common problem in combinatorial searches in materials science, as well as chemistry. However, these data-sets may well be ‘small’ in terms of limited step-size of the measurement variables. Due to this limitation, application of higher-order statistics is not effective, and the choice of a suitable unsupervised learning method is restricted to those utilizing lower-order statistics. As an interesting case study, we present here variable magnetic-field Piezoresponse Force Microscopy (PFM) study of composite multiferroics, where due to experimental limitations the magnetic field dependence of piezoresponse is registered with a coarse step-size. An efficient extraction of this dependence, which corresponds to the local magnetoelectric effect, forms the central problem of this work. We evaluate the performance of Principal Component Analysis (PCA) as a simple unsupervised learning technique, by pre-labeling possible patterns in the data using Density Based Clustering (DBSCAN). Based on this combinational analysis, we highlight how PCA using non-central second-moment can be useful in such cases for extracting information about the local material response and the corresponding spatial distribution.

npj Computational Materials (2018)4:28; doi:10.1038/s41524-018-0084-9

## INTRODUCTION

In an age where advanced material characterization techniques offer access to a wide range of information, the problem essentially lies in extracting a sensible meaning out of the large amount of information collected. However, with an ever-increasing data generation tendency in scientific instruments, the necessary data organization and analysis techniques have also come a long way.<sup>1–7</sup> This is also valid for high-throughput combinatorial searches in materials science, as well as chemistry, in which a large number of parallel or complementary properties may be measured with changes in composition or processing variables. Such data is used to identify previously unknown trends or materials-property connections, but a problem exists in efficient analysis and visualization of the data.<sup>8</sup>

As far as Scanning Probe Microscopy (SPM) is concerned, the recent developments in functional imaging-modes have been paralleled with a growing interest in studying various local phenomena. A few examples of such phenomena include electrical conduction at ferroelectric/ferroelastic domain walls,<sup>9–12</sup> polarization dynamics in ferroelectrics,<sup>13–18</sup> temperature/time/voltage dependent study of ergodicity (time dependence) of polarization in relaxor-ferroelectrics,<sup>14,19</sup> and local magnetoelectricity.<sup>20–28</sup> Probing such local phenomena essentially requires measuring a response  $I_{ij}(x_i, P_j)$  on an  $X \times Y$  grid, as a function of a spectral parameter  $P_j$  ( $j = 1, \dots, M$ ); where  $x_i$  is the spatial co-ordinate index ( $i = 1, \dots, N$ ;  $N = X \times Y$ ). The format of such spectroscopic acquisition could be selected in two different ways: (i) a point-by-point acquisition of the response as a function of  $P_j$ , or (ii) a sequence of scans at different values of  $P_j$ . The choice between either of the formats largely depends upon the time step necessary to stabilize the spectral parameter, taking into

consideration the scanner drift. In most cases, voltage is the spectral parameter, which could be swiftly varied, facilitating a drift-free point-by-point acquisition. However, when the spectral variable of interest is temperature, time, or magnetic field, the acquisition time between each step of the spectral variable is larger. In such cases, it is more suitable to opt for the sequential scan format, where a better track of the drift could be kept via simultaneously acquired topography images.

One implication of such spectroscopic techniques includes the local study of magnetoelectric (ME) effect in multiferroic composites. Such composites are a realistic alternative for achieving the foreseen device applications at room-temperature.<sup>29–31</sup> The ME effect in such composites takes place via stress mediation, where the field induced strain in one phase leads to a stress on the adjacent phase, which ultimately varies the order parameter (polarization/magnetization) of the other phase.<sup>32–36</sup> Interestingly, the performance of such composites, gauged by the ME coupling coefficient  $\alpha$ , is found to be highly sensitive to microstructure.<sup>37–40</sup> On the other hand, modeling predicts complex domain evolution in composites.<sup>41–43</sup> PFM based investigation (direct ME effect) is expected to reveal a strong modulation of the local piezoresponse (i.e., polarization) as a function of the externally applied magnetic field<sup>20–28</sup> giving an indirect insight into the spatial distribution of the local magneto-electro-mechanical interactions.

In the present work, two particulate composites of BaTiO<sub>3</sub> (piezoelectric phase) and two different ferrites (magnetostrictive phase) were studied. As a consequence of the applied magnetic field  $H$ , the local polarization is altered as:

$$\Delta P_i = \alpha_{ij} \Delta H_j \quad (1)$$

<sup>1</sup>Institute for Materials Science and Center for Nanointegration Duis order ferroic switching induced by sburg-Essen (CENIDE), University of Duisburg-Essen, Essen, Germany;

<sup>2</sup>CICECO-Aveiro Institute of Materials, University of Aveiro, Aveiro, Portugal and <sup>3</sup>Department of Materials and Ceramic Engineering, University of Aveiro, Aveiro, Portugal

Correspondence: Harsh Trivedi (harsh.trivedi@uni-due.de) or Vladimir V. Shvartsman (vladimir.shvartsman@uni-due.de)

Received: 24 July 2017 Revised: 7 May 2018 Accepted: 9 May 2018

Published online: 21 June 2018

with  $\alpha_{ij}$  is the ME coefficient. This change in polarization should lead to a change of the local piezoelectric coefficient owing to its relation to polarization:<sup>44</sup>

$$d_{kij} = Q_{ijml}(P_l \epsilon_{mk} + P_m \epsilon_{lk}) \quad (2)$$

where  $\epsilon_{mk}$  is the dielectric permittivity tensor, and  $Q_{ijml}$  is the electrostriction coefficient tensor. Substituting equation 1 in 2, a magnetoelectrically induced change in  $d_{ijk}$  can be expressed as:

$$\Delta d_{kij} = Q_{ijml}(\alpha_{ln} \Delta H_n \epsilon_{mk} + \alpha_{mn} \Delta H_n \epsilon_{lk}) \quad (3)$$

In the variable-field PFM experiment one effectively measures  $\Delta d_{eff}(x, y) = f(x, y, H_0, E_0, \Delta H)$ ,<sup>32</sup> and hence an exact observation of equation 3 is not trivial. Despite this limitation, PFM can reveal the intensity of the ME effect at a local scale, which could be useful information to relate the ME coupling to the material properties, as well as microstructure.

An experimental validation of equation 3 could be achieved by measuring local piezoresponse as a function of magnetic field, in other words variable magnetic-field PFM. This was realized in the present work by opting for the sequential spectroscopy approach, where PFM images were acquired in sequence at different in-plane magnetic fields. The problem with this approach is that the large time of acquisition required per image restricts the number of magnetic field steps. The total acquisition time can be expressed as the sum of individual scan times, and the ramp time for achieving a stable field value:

$$T_{total} = M \cdot T_{scan} + \frac{|H_{end} - H_{start}|}{r} \quad (4)$$

$T_{scan}$  –Time per scan,  $M$  –No. of magnetic field steps,  $r$  –Ramp rate,  $H_{start}$  –Starting magnetic field,  $H_{end}$  –End magnetic field

Equation 4 highlights the inevitable limitation in the choice of a large  $M$  value, which could lead to an increased total acquisition time with all the other parameters being fixed. For the present case, where we consider  $H_{start} = H_{end} = 6000$  Oe, and  $r = 600$  Oe/min, an optimal acquisition time without any break in the sequence comes out to be around 3-4 h by choosing  $M = 20$ . An acquisition time larger than this is considered impractical for manual acquisition, and too large for automated acquisition due to the risk of large thermal drifts. We opted for automated acquisition, and had to discard certain images from the sequence due to drift and/or noise being too high, leading to an  $M$  value lower than 20.

The problem at hand here constitutes an efficient extraction of the actual material response from  $I_{ij}$ , without any *a priori* knowledge about the functional form of the response. In addition to the material response, the acquired image  $I_{ij}$  is expected to contain noise, as well as certain additional non-stochastic variations which have no relation to the concerned phenomena. Such variation in  $I_{ij}$  can occur due to (i) varying PFM tip-sample contact conditions,<sup>45-47</sup> and/or (ii) the presence of relative drifts between each image (columns of  $I_{ij}$ ). The linear drift between the images is easily removed as a part of the preprocessing step (see supplementary material); however, removal of the non-linear drift is not always trivial. Since we are interested in extracting the ME-induced variation of piezoresponse, it is useful to consider the problem format similar to that of time-series predictions or blind source-separation tasks,<sup>48</sup> where the different non-stochastic variations, including the material response, are expected to be stochastically mixed. The corresponding latent-variable model can be written as:

$$I_{ij} = a_{ik} R_{kj} + n \quad (5)$$

where  $I_{ij}$  ( $M \times N$ ) is the data-matrix,  $n$  represents white noise,  $R_{kj}$  encompasses the actual material response, as well as the non-stochastic components, and  $a_{ik}$  is the weight matrix representing weights of each component at each spatial location. At this point, we have two physical constraints for  $R_{kj}$  and  $a_{ik}$ : (i) the material

response in  $R_{kj}$ , which is indirectly related to magnetostriction, should be symmetric in magnetic-field, and (ii) it should be localized within the piezoactive (ferroelectric) phase.

The value chosen for  $N$  can be sufficiently large, but the choice of  $M$  is limited. This effectively limits the application of higher order statistics (e.g., Independent Component Analysis; ICA) due to insufficient sampling available for characterizing the marginal densities of the data-vector components (rows of  $I_{ij}$ ).<sup>49</sup> At the same time, if the physical phenomena (ME effect) is weaker, then the data could be further obscured. In such circumstances, it is recommended to be restricted to low-order statistics, and Principal Component Analysis (PCA) is a simpler alternative. In the present case, we prefer usage of non-central moments for PCA, based on the intuition that the material response is expected to lie farther away from the origin in the feature-space, contributing to the maximum univariate moment which the PCA algorithm searches for.<sup>50</sup> However, it is known that unlike ICA, PCA is limited due to the condition of orthogonal bases, and cannot correctly fit data with non-orthogonal density distribution.<sup>51</sup> In other words, PCA is prone to show artifacts in cases where there is more than one independent component present in the data. To monitor this problem, we opt for an approach where we initially label the data using Density Based Clustering (DBSCAN) and visualize the labeled data by means of non-linear projection in 3D. The visualization of labeled data allows us to assess the outcomes of PCA.

## RESULTS AND DISCUSSION

### Data-structure visualization

With  $I_{ij}$  as the data-matrix, we end up with an  $M > 3$  dimensional data-space. A 3D projection of this multi-dimensional data is necessary to characterize the data-structure, which will also serve as a starting point for the analysis that will follow. We have utilized t-Stochastic Neighbor Embedding (t-SNE)<sup>52</sup> for the 3D projection of the  $M$ -dimensional data (Fig. 1).

t-SNE is known to restore the local structure of the high dimensional data into its low dimensional representation. In other words, the visually observable clusters in the 3D projections should correspond to the individual non-stochastic components present in the data. Here, the data-structure of the two different material systems manifest different traits. The data corresponding to the BaTiO<sub>3</sub>-BaFe<sub>12</sub>O<sub>19</sub> system (BTO-BaM) hardly manifests any clusters, suggesting a weaker presence of the individual non-stochastic components. However, the BaTiO<sub>3</sub>-CoFe<sub>2</sub>O<sub>4</sub> system (BTO-CFO) data show a clear distinction between different groups of data-points. Next, we apply DBSCAN to rigorously identify these clusters, irrespective of the visual obscurity in the projected data.

### Density based clustering

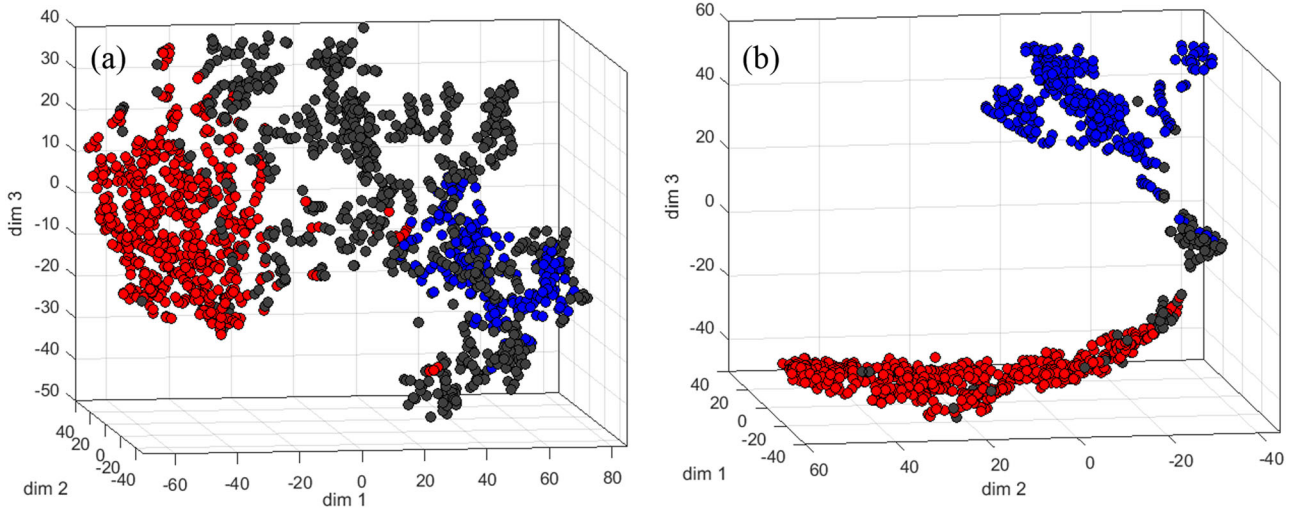
Identical responses tend to cluster in the data-space, and a suitable parameter which can be used to define a cluster is the density based connectivity of the corresponding data-points.<sup>53</sup> DBSCAN is a unique clustering algorithm which looks for clusters based on their local density, and is hence more suitable for the current data-sets. The input criteria required for DBSCAN, namely the neighborhood radius  $\epsilon$  (Euclidian), and the minimum number of neighborhood points  $p$ , were determined from the data-sets. First, the value of  $p$  was heuristically fixed for each data-set based on the total number of data-points, followed by the determination of the average  $p$ -nearest neighbor distance for the entire data-set, which was set to be  $\epsilon$ . Based on these criteria, three clusters were identified in each data-set, as shown in Fig. 1.

Based on visual comparisons of the PFM images (Fig. 2a,b) and the cluster domain-maps (Fig. 2c,d) of both systems, it becomes evident that cluster-2 corresponds to the ferroelectric phase, and should hence represent the material response. On the other hand,

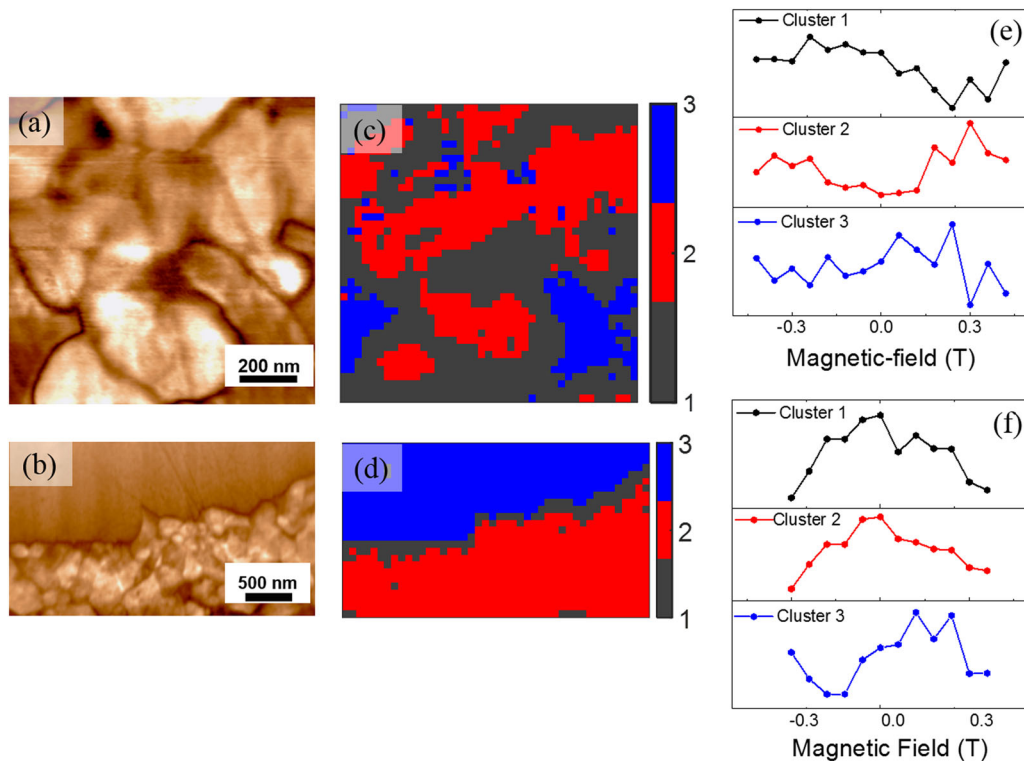
cluster-1 in both cases localizes itself in an area dispersed around the interface, and corresponds to the outliers of DBSCAN. The data-points corresponding to the magnetostrictive phase in both systems have been classified as cluster-3. Figure 2e,f shows the mean-vectors corresponding to each cluster, highlighting interesting nuances. In both cases, the cluster-3 mean-vector does not carry any physical significance as it corresponds to the piezo-inactive magnetic phase. Whereas the cluster-2 mean-vector should ideally reveal an average trend in the magnetic-field dependence of the piezoresponse. In case of BTO-CFO, a reasonable trend in the piezoresponse is visible, which roughly fulfills the physical constraint of symmetry w.r.t. magnetic-field.

However, in the case of BTO-BaM although the corresponding mean-vector does show a symmetry w.r.t. magnetic-field, the relative variation of the piezoresponse is subtle. These nuances in the mean-vector, combined with the features of the data-structure, suggest that the observed local ME effect in BTO-BaM as is weaker as compared to that in BTO-CFO.

Upon labeling the data (Fig. 1), it becomes evident that the non-stochastic components in  $R_{ij}$  have been classified into isolated clusters of high density in the feature-space (cluster-2, cluster-3). The fact that the outliers in both cases are localized around the interface suggests that they correspond to a drift-based variation in piezoresponse. As a consequence of the non-linear drift residual



**Fig. 1** Labeled 3D representation of data corresponding to **a** BTO-BaM, and **b** BTO-CFO systems, based on the clusters identified by DBSCAN (gray: cluster-1, red: cluster-2, blue: cluster-3)



**Fig. 2** Spatial locations of the identified clusters (domain maps) for system-1 **c**, and system-2 **d**. **e**, **f** show the mean-vectors of clusters corresponding to system-1 and system-2, respectively. Corresponding PFM images for BTO-BaM **a**, and BTO-CFO **b** are shown for reference



in the images, the interface data-points undergo a larger variation due to a larger difference in piezoresponse present on either side of the interface. Based on this intuition BTO-BaM possesses a relatively larger non-linear drift, leading to larger quantities of data-points being affected. The interface points in BTO-CFO are seemingly not so strongly affected by the non-linear drifts, and hence correspond to a nominal change from the material response (cluster-2). This conclusion is further corroborated by the fact that in BTO-CFO the cluster-1 more closely corresponds to the cluster-2 (Fig. 2f).

#### Principal component analysis

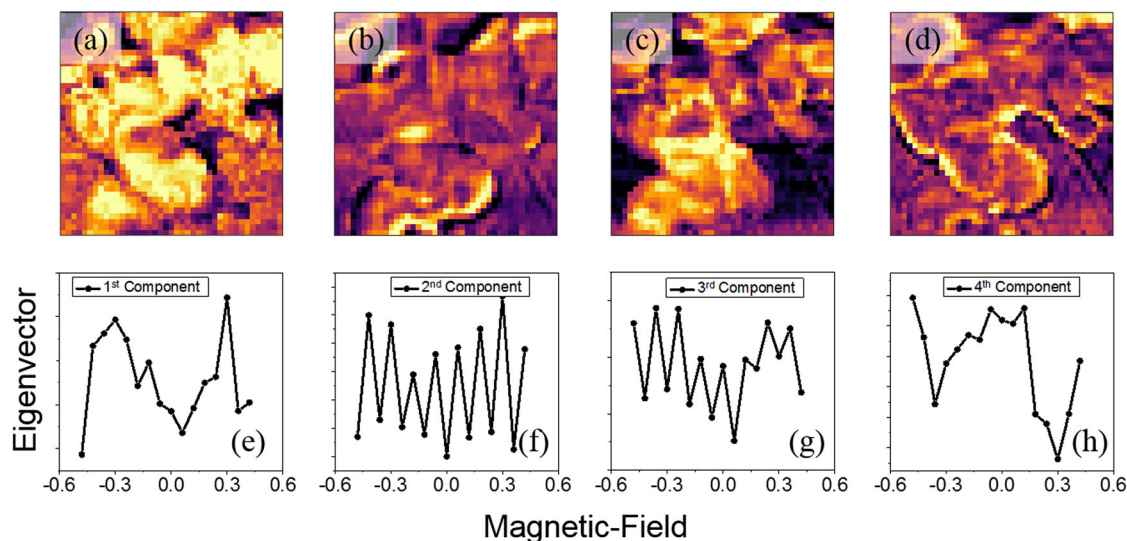
Figure 3 shows results corresponding to the Singular Value Decomposition (SVD) of the BTO-BaM data-matrix. The 1<sup>st</sup> component in PCA corresponds to the highest univariate second-moment (equivalent to variance for centered data) direction, followed by components with a decreasing order of the moment. Simultaneously, the score maps represent the spatial score of the corresponding patterns. We observe that the 1<sup>st</sup> component, which has a higher score in the ferroelectric phase, fulfills the physical requirement of horizontal (i.e., w.r.t. the magnetic field) symmetry (Fig. 3a,e), whereas the subsequent 2<sup>nd</sup> and 3<sup>rd</sup> components do not seem to have any obvious physical

correlation (Fig. 3b,c,f-g). Interestingly, the 4<sup>th</sup> component seems to be more relevant (Fig. 3d,h), but apparently the corresponding moment is too small (Fig. 5a) to consider it as being significant.

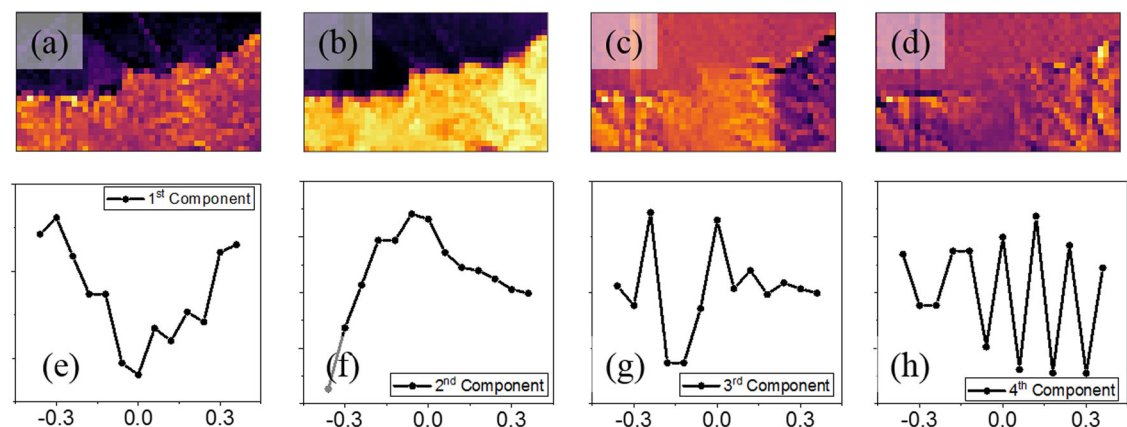
Upon SVD of the BTO-CFO data-matrix (Fig. 4), the 1<sup>st</sup> component indeed shows the expected field dependence (Fig. 4a,e), but at the same time the 2<sup>nd</sup> component also shows a significant related phenomenon in the data (Fig. 4f), which is localized within the ferroelectric phase (Fig. 4b). Subsequently, the 3<sup>rd</sup> and 4<sup>th</sup> components possess negligible moments (Fig. 4c,d,g, h), and are hence ignored from further analysis.

#### DISCUSSION

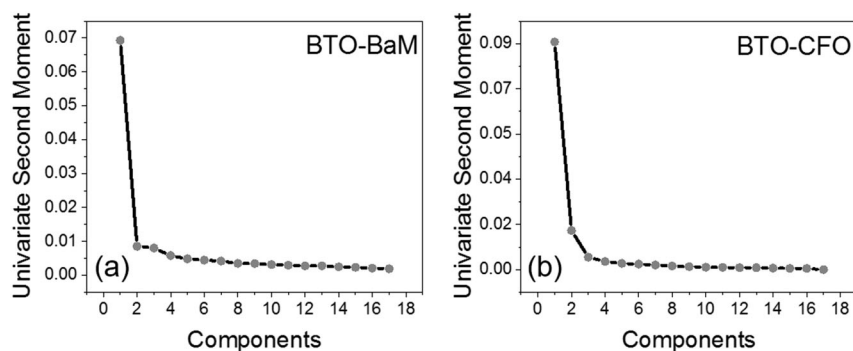
In order to understand the different results obtained by PCA of the two material systems, we consider the basic properties of PCA, which is minimization of the bivariate second-moments and maximization of the univariate second-moments. In the case of a centered data-set, the univariate moments could be considered as the average of the distance of the data-points from their mean projected on the component axis. However, in the present case, where the data is not centered, this distance is measured from the origin, and hence the farthest lying data-points will contribute the maximum univariate moment. In reality, the piezoresponse value corresponding to the magnetostrictive phase is much smaller than



**Fig. 3** The first four principal components e–h, and the corresponding score maps a–d, decomposed from the data-matrix of system-1 (The bright and dark contrast corresponds to high and low score value, respectively)



**Fig. 4** The first four principal components e–h, and the corresponding score maps a–d, decomposed from the data-matrix of system-2 (The bright and dark contrast corresponds to high and low score value, respectively)



**Fig. 5** Moment vs. components curves corresponding to BTO-BaM **a**, and BTO-CFO **b**, showing the decreasing order of the non-central univariate moment among the components

that of the ferroelectric phase and, hence, the corresponding data-points are located closer to the origin. As a result, the first principal component will be predominantly decided by the data-points corresponding to the ferroelectric phase. In order to fulfill its second requirement of minimization of the bivariate second-moments, PCA looks for an axis with a minimum average of squared orthogonal distances from the data points. In other words, it looks for a regression straight-line passing through the data along the direction of the maximal univariate-moments, which also intercepts the origin. Intuitively, the data-points corresponding to the magnetostrictive phase, which has relatively high stochastic spread, have relatively low potential to affect the minimization of the bivariate moments compared to the data-points of the ferroelectric phase.

Based on the theoretical arguments mentioned above, the interesting outcomes of PCA can be analyzed. In the case of BTO-BaM, the blue cluster lies closer to the origin, whereas the red cluster lies farthest from the origin. Hence, the first component essentially represents the red cluster, as evident from the spatial correlation, as well as its qualitative resemblance to the corresponding mean-vector (Fig. 2e). Next, the drift-induced spread in the data-points pertinent to the interface decides the next highest direction of moment, namely the 2<sup>nd</sup> component. Evidently, the score of the 2<sup>nd</sup> component is also higher near the interface (Fig. 3b). However, in the case of BTO-CFO a different analysis is required. In order to understand the difference, we refer to the mapped data-structure for BTO-CFO (Fig. 1b). It can be seen that the red cluster in BTO-CFO has a large and uniform spread in the feature-space. This signifies that the field dependence of piezoresponse undergoes continuous deviation within the ferroelectric phase, with the largest deviation occurring at the interface (outliers in DBSCAN). In other words, there is a continuous spread in the data-points lying farther away from origin, which leads to the 1st and the 2nd component both dividing themselves optimally between the two important events in the data, namely the pattern in piezoresponse, as well as the spread of that pattern. It is interesting to note how efficiently PCA takes into account the main event, as well as the spread in both cases, despite the different nature of the spread.

Apparently, PCA seems to demonstrate a best-suited extraction of the physical phenomena in the concerned data-structure. Up to this point, the observed magnetic field dependence of piezoresponse does not seem to have a non-physical origin, since along with the fact that it localizes itself within the ferroelectric phase, we have also cross-checked the simultaneously acquired topography response using PCA (see supplementary material). Also, we conclude that the two systems show different types of magnetic-field dependency: on one hand, BTO-BaM shows a unique but

weak dependence, whereas on the other hand BTO-CFO shows a stronger dependency on magnetic field which apparently undergoes nominal deviation within the ferroelectric phase. This deviation could be associated with the varying degrees of ME coupling due to a varying texture of BaTiO<sub>3</sub>.

In conclusion, by carrying out combinational analysis using PCA and DBSCAN, we were able to extract physical meaning out of the variable magnetic-field PFM data-sets. At the same time, using valid experimental arguments, we highlighted the inherent ‘small-data’ problem of the sequential imaging data-sets, which needs to be tackled owing to the general interests in such type of experiments. The 3D data representation obtained using t-SNE suggested a varied distribution of the data-points, highlighting the sparse nature of the data-vectors. The data-points were efficiently clustered using DBSCAN. The labeled data-structure of both systems clearly manifested data-points corresponding to the ferroelectric phase, the magnetic phase, and the interface. A key difference between the data-structures of both the systems was in terms of their secondary spread. Despite the individual complexities, both data-sets were optimally analyzed by PCA. This leads to the conclusion that utilizing second order non-central moments should provide an optimal feature extraction for such data-sets. As a general physical conclusion, we notice that the locally observed ME effect is weaker in the hexaferrite based composite as compared to that in the spinel-ferrite based composite. This difference could be associated with the fact that in hexaferrite the number of possible easy axes per crystallite, along which the magnetostriction is highest, is limited to 2, whereas in case of spinel-ferrites there are 4 such possible axes, leading to a more uniform stress distribution at the interface.

## METHODS

The studied systems are particulate composites prepared by means of standard ceramic processing route.<sup>23</sup> The prepared ceramic disks were ground down to under 0.5 mm thickness, in order to be able to apply high magnetic field in a confined region. The studied surfaces on the disks were then finely polished down to grit size of 0.25  $\mu\text{m}$ .

The measurements were carried out using a commercial scanning probe microscope (MFP-3D, Asylum Research). The PFM amplitude images were collected in a dual amplitude resonance tracking mode (DART-PFM™), such that the operational frequency stays close to the contact resonance frequency. Fresh uncoated doped silicon cantilevers (SEIHR, Nanosensors) with a spring constant  $k = 15 \text{ N/m}$  and a free air resonance frequency of  $f = 130\text{--}250 \text{ kHz}$  were used. The corresponding contact resonance frequency was about 600 kHz. The choice of the cantilevers was optimized after several trial and error steps, in order to provide an optimal PFM contrast. A commercial variable magnetic field module (VFM2, Asylum Research) was used to apply in-plane magnetic field up to 3 kOe. The scan resolution was  $512 \times 512$  pixels.

The t-SNE 3D representation was obtained using an open source MATLAB tool distributed by Laurens van der Maaten. The DBSCAN was implemented using the open source Python library distributed by scikit-learn.org.<sup>54</sup>

#### Code availability

The codes generated during the study are available from the corresponding author (H. Trivedi) upon reasonable request. The codes are not publicly available due to them containing information that could compromise research participants consent.

#### Data availability

The datasets generated and analyzed during the study are made available as supplementary information files.

### ACKNOWLEDGEMENTS

This work was supported by the European Commission within FP7 Marie Curie Initial Training Network “Nanomotion” (grant agreement n° 290158). Support through Deutsche Forschungsgemeinschaft via Forschergruppe 1509 “Ferroic Functional Materials” (LU-729/12) is acknowledged. This work was developed within the scope of the project CICECO-Aveiro Institute of Materials, POCI-01-0145-FEDER-007679 (FCT Ref. UID /CTM /50011/2013), financed by national funds through the FCT/MEC and when appropriate co-financed by FEDER under the PT2020 Partnership Agreement. R. C.P. thank the FCT for funding under grant IF/00681/2015. We would like to acknowledge Matthias Labusch and Jörg Schröder (Institute of Mechanics, University of Duisburg-Essen) for their help with the constitutive formulation.

### AUTHOR CONTRIBUTIONS

H.T. carried out the measurement and the analysis presented, and V.V.S. supervised the project, R.C.P. and M.A.M. synthesized and characterized the hexaferrite-based composite system discussed in the paper, H.T. wrote the paper, and V.V.S., D.C.L., and R.C.P. contributed to the revision, V.V.S., D.C.L., and R.C.P. contributed to the discussion of results and the theoretical basis.

### ADDITIONAL INFORMATION

**Supplementary information** accompanies the paper on the *npj Computational Materials* website (<https://doi.org/10.1038/s41524-018-0084-9>).

**Competing interests:** The authors declare no competing interests.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### REFERENCES

- Laanait, N., Zhang, Z. & Schleputz, C. M. Imaging nanoscale lattice variations by machine learning of x-ray diffraction. *Nanotechnology* **27**, 374002–374011 (2016).
- Shinzawa, H., Awa, K., Kanematsu, W. & Ozaki, Y. Multivariate data analysis for Raman spectroscopic imaging. *J. Raman Spectrosc.* **40**, 1720–1725 (2009).
- Kalinin, S. V. et al. Big, deep, and smart data in scanning probe microscopy. *ACS NANO* **10**, 9068 (2016).
- Ceguerra, A. V. et al. The rise of computational techniques in atom probe microscopy. *Curr. Opin. Solid State Mater. Sci.* **17**, 224–235 (2013).
- Hu, Y. et al. A nonrelational data warehouse for the analysis of field and laboratory data from multiple. *IEEE J. Photovolt.* **7**, 230–236 (2017).
- AbuOmar, O. et al. A polymer nanocomposites case study. *Adv. Eng. Inform.* **27**, 615–624 (2013).
- Gaponenko, I. et al. Computer vision distortion correction of scanning probe microscopy images. *Sci. Rep.* **7**, <https://doi.org/10.1038/s41598-017-00765> (2017).
- Pullar, R. C. Combinatorial materials science, and a perspective on challenges in data acquisition, analysis and presentation. In *Information Science for Materials Discovery and Design*. (eds Lookman, T., Alexander, F. & Rajan, K.) Ch. 13, 241–270 (Springer, Heidelberg, 2016).
- Balke, N. et al. Enhanced electric conductivity at ferroelectric vortex cores in BiFeO<sub>3</sub>. *Nat. Phys.* **8**, 81–88 (2011).
- Meier, D. et al. Anisotropic conductance at improper ferroelectric domain walls. *Nat. Mater.* **11**, 284–288 (2012).
- Jiang, J. et al. Temporary formation of highly conducting domain walls for non-destructive read-out of ferroelectric domain-wall resistance switching memories. *Nat. Mater.* **17**, 49–56 (2018).
- Seidel, J. et al. Conduction at domain walls in oxide multiferroics. *Nat. Mater.* **8**, 229–234 (2009).
- Balke, N. et al. Deterministic control of ferroelastic switching in multiferroic materials. *Nat. Nanotechnol.* **4**, 868–875 (2009).
- Dittmer, R. et al. Nanoscale Insight Into Lead-Free BNT-BT- x KNN. *Adv. Funct. Mater.* **22**, 4208–4215 (2012).
- Alsubaie, A., Sharma, P., Liu, G., Nagarajan, V. & Seidel, J. Mechanical stress-induced switching kinetics of ferroelectric thin films at the nanoscale. *Nanotechnology* **28**, 75709–75716 (2017).
- Rodriguez, B. J. et al. Spatially resolved mapping of ferroelectric switching behavior in self-assembled: strain, size, and interface effects. *Nanotechnology* **18**, 405701–405708 (2007).
- Abplanalp, M., Fousek, J., & Günter, P. Higher order ferroic switching induced by scanning force microscopy. *Phys. Rev. Lett.* **86**, 5799–5802 (2001).
- Agar, J. C. et al. Highly mobile ferroelastic domain walls in compositionally graded ferroelectric thin films. *Nat. Mater.* **15**, 549–556 (2016).
- Gobeljic, D., Dittmer, R., Rödel, J., Shvartsman, V. V., & Lupascu, D. C. Macroscopic and nanoscopic polarization relaxation kinetics in lead-free relaxors Bi<sub>1/2</sub>Na<sub>1/2</sub>TiO<sub>3</sub>-Bi<sub>1/2</sub>K<sub>1/2</sub>TiO<sub>3</sub>-BiZn<sub>1/2</sub>Ti<sub>1/2</sub>O<sub>3</sub>. *J. Am. Ceram. Soc.* **97**, 3904–3912 (2014).
- Zheng, T. et al. Local probing of magnetoelectric properties of PVDF/Fe<sub>3</sub>O<sub>4</sub> electrospun nanofibers by piezoresponse force microscopy. *Nanotechnology* **28**, 065707 (2017).
- Pan, D. F. Local magnetoelectric effect in la-doped BiFeO<sub>3</sub> multiferroic thin films revealed by magnetic-field-assisted scanning probe microscopy. *Nanoscale Res. Lett.* **11**, 318 (2016).
- Yan, F., Chen, G., Lu, L., Finkel, P., Spanier, J. E. Local probing of magnetoelectric coupling and magnetoelastic control of switching in BiFeO<sub>3</sub>-CoFe<sub>2</sub>O<sub>4</sub> thin-film nanocomposit. *Appl. Phys. Lett.* **103**, 42906/1–4 (2013).
- Trivedi, H. et al. Local manifestations of a static magnetoelectric effect in nanostructured. *Nanoscale* **7**, 4489–4496 (2015).
- Khodaei, M., Eshghinejad, A., Seyyed Ebrahimi, S. A. & Baik, S. Nanoscale magnetoelectric coupling study in (111)-oriented PZT-Co ferrite multiferroic. *Sens. Actuators A* **242**, 92–98 (2016).
- Vasudevan, R. K., Jesse, S., Kim, Y., Kumar, A. & Kalinin, S. V. Spectroscopic imaging in piezoresponse force microscopy: New opportunities for studying polarization dynamics in ferroelectrics and multiferroics. *MRS Commun.* **2**, 61–73 (2012).
- Miao, H., Zhou, X., Dong, S., Luo, H. & Li, F. Magnetic-field-induced ferroelectric polarization reversal in magnetoelectric. *Nanoscale* **6**, 8515–8520 (2014).
- Xie, S. H. et al. Magnetoelectric coupling of multilayered Pb(Zr<sub>0.52</sub>Ti<sub>0.48</sub>)O<sub>3</sub>-CoFe<sub>2</sub>O<sub>4</sub> film by piezoresponse force microscopy under magnetic field. *J. Appl. Phys.* **112**, 074110–074115 (2012).
- Xie, S., Ma, F., Liu, Y. & Li, J. Multiferroic CoFe<sub>2</sub>O<sub>4</sub>-Pb(Zr<sub>0.52</sub>Ti<sub>0.48</sub>)O<sub>3</sub> core-shell nanofibers and their magnetoelectric coupling. *Nanoscale* **3**, 3152–3158 (2011).
- Jahns, R. et al. Giant magnetoelectric effect in thin-film composites. *J. Am. Ceram. Soc.* **96**, 1673–1681 (2013).
- Zhao, T. et al. Electrical control of antiferromagnetic domains in multiferroic BiFeO<sub>3</sub> films at room. *Nat. Mater.* **5**, 823–829 (2006).
- Bibes M., & Barthélémy, A. Towards a magnetoelectric memory. *Nat. Mater.* **7**, 425–426 (2008).
- Lupascu, D. C. et al. Measuring the magnetoelectric effect across scales. *GAMM-Mitt.* **38**, 25–74 (2015).
- Srinivasan, G. et al. Magnetoelectric bilayer and multilayer structures of magnetostrictive and piezoelectric oxides. *Phys. Rev. B* **64**, 214408 (2001) 1–6.
- Nan, C. W., Liu, G., Lin, Y. & Chen, H. Magnetic-field-induced electric polarization in multiferroic nanostructures. *Phys. Rev. Lett.* **94**, 197203 (2005) 1–4.
- Bichurin, M. I., Kornev, I. A., Petrov, V. M. & Lisnevskaya, I. V. Investigation of magnetoelectric interaction in composite. *Ferroelectrics* **204**, 289–297 (1997).
- Boomgaard, J. VanDen, Van Run, A. M. J. G. & van Suchtelen, J. Piezoelectric-piezomagnetic composites with magnetoelectric effect. *Ferroelectrics* **14**, 727–728 (1976).
- Schmitz-Antoniak, C. et al. Electric in-plane polarization in multiferroic CoFe<sub>2</sub>O<sub>4</sub>/BaTiO<sub>3</sub> nanocomposite tuned by magnetic fields. *Nat. Commun.* **4**, 2051 (2013).
- Duong, G. V., Groessinger, R. & Sato Turtelli, R. Driving mechanism for magnetoelectric effect in CoFe<sub>2</sub>O<sub>4</sub>-BaTiO<sub>3</sub> multiferroic composite. *J. Magn. Mater.* **310**, 1157–1159 (2007).
- Etier, M. et al. Magnetoelectric coupling on multiferroic cobalt ferrite-barium titanate ceramic composites with different connectivity schemes. *Acta Materialia* **90**, 1–9 (2015).
- Corral-Flores, V., Bueno-Baques, D., Carrillo-Flores, D. & Matutes-Aquino, J. A. Enhanced magnetoelectric effect in core-shell particulate composites. *J. Appl. Phys.* **99**, 08J503 (2006).

41. van Lich et al. Colossal magnetoelectric effect in 3-1 multiferroic nanocomposites originating from. *Appl. Phys. Lett.* **107**, 232904 (2015).
42. Ma, F. D., Jin, Y. M., Wang, Y. U., Kampe, S. L. & Dong, S. Phase field modeling and simulation of particulate magnetoelectric composites: Effect of connectivity, conductivity, poling, and bias field. *Acta Mater.* **70**, 45–55 (2014).
43. Ma, F. D., Jin, Y. M., Wang, Y. U., Kampe, S. L. & Dong, S. Effect of magnetic domain structure on longitudinal and transverse magnetoelectric. *Appl. Phys. Lett.* **104**, 112903 (2014).
44. Lupascu, D. C. & Rodel, J. Fatigue In bulk lead zirconate titanate actuator materials. *Adv. Eng. Mater.* **7**, 882–898 (2005).
45. Giannakopoulos, A. E. & Suresh, S. Theory of indentation of piezoelectric materials. *Acta Mater.* **47**, 2153–2164 (1999).
46. Mirman, B. & Kalinin, S. V. Resonance frequency analysis for surface-coupled atomic force microscopy cantilever in ambient. *Appl. Phys. Lett.* **92**, 83102 (2008).
47. Kalinin, S. & Bonnell, D. Imaging mechanism of piezoresponse force microscopy of ferroelectric surfaces. *Phys. Rev. B* **65**, 125408 (2002).
48. Hvyärinen, A., Karhunen, J., Oja, E. *Independent Component Analysis*, 5–7 (Wiley-Interscience, 2001).
49. Hvyärinen, A., Karhunen, J., Oja, E. *Independent Component Analysis*, 152–153 (Wiley-Interscience, 2001).
50. Honeine, P. An eigenanalysis of data centering in machine learning. Preprint at arxiv:1407.2904v1 (2014).
51. Lewicki, M. S. & Sejnowski, T. J. Learning overcomplete representations. *Neural Comput.* **12**, 337–365 (2000).
52. van der Maaten, L. & Hinton, G. Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
53. Ester, M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226–231, (AAAI Press, Palo Alto, California, USA, 1996).
54. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **9**, 2825–28230 (2011).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018