


ARTICLE OPEN

Reconstructing phase diagrams from local measurements via Gaussian processes: mapping the temperature-composition space to confidence

Dhiren K. Pradhan^{1,6}, Shalini Kumari^{1,7}, Evgheni Strelcov², Dillip K. Pradhan³, Ram S. Katiyar¹, Sergei V. Kalinin^{4,5}, Nouamane Laanait^{4,5} and Rama K. Vasudevan^{4,5} 

We show the ability to map the phase diagram of a relaxor-ferroelectric system as a function of temperature and composition through local hysteresis curve acquisition, with the voltage spectroscopy data being used as a proxy for the (unknown) microscopic state or thermodynamic parameters of materials. Given the discrete nature of the measurement points, we use Gaussian processes to reconstruct hysteresis loops in temperature and voltage space, and compare the results with the raw data and bulk dielectric spectroscopy measurements. The results indicate that the surface transition temperature is similar for all but one composition with respect to the bulk. Through clustering algorithms, we recreate the main features of the bulk diagram, and provide statistical confidence estimates for the reconstructed phase transition temperatures. We validate the method by using Gaussian processes to predict hysteresis loops for a given temperature for a composition unseen by the algorithm, and compare with measurements. These techniques can be used to map phase diagrams from functional materials in an automated fashion, and provide a method for uncertainty quantification and model selection.

npj Computational Materials (2018)4:23; doi:10.1038/s41524-018-0078-7

INTRODUCTION

Phase diagrams serve as descriptors of material behavior and material properties, and their forms are intrinsically linked to the underpinning physics driving the system. Traditionally, phase diagrams are determined from property measurements across material compositions, when the anomalies in specific heat, lattice parameter, polarization, or other macroscopic parameters are interpreted as signatures of phase transitions, allowing the phase boundaries to be drawn.¹ Recently, progressively large attention has been focused on extracting this information from theoretical or combinatorial data.^{2–7} The groups involved in synthesis of materials via combinatorial methods have recently turned to using machine-learning-based approaches to characterize the structural phases present, applying clustering methods to large volumes of X-ray diffraction data. These techniques have been explored by several groups worldwide in an effort to accelerate experimental materials design and discovery.^{8–13} They rely on either unsupervised or semi-supervised learning, to group (cluster) diffraction patterns based on some given similarity metric, enabling the distinguishing of phase boundaries across the composition spread.¹⁴ However, the same techniques are seldom applied to local measurements:¹⁵ for instance, while studies of piezoresponse force microscopy (PFM) of piezo/ferroelectric samples as a function of temperature are routine, studies of functional response (such as hysteresis loops) as a function of both temperature and

composition are rare. Partly, the difficulty is related to the appropriate methods of analysis of large multidimensional data sets across temperature and voltage space in a consistent fashion.

Fundamentally, there is appreciable difficulty in accurately determining the phase transition temperature from local measurements, given that they (usually) do not provide access to the order parameter directly and often are not quantitative. Among the scanning probe microscopy methods, piezoresponse force microscopy offers an example of technique which intrinsically quantitative due to specific voltage-dependent contact mechanics.^{16–19} The PFM response measures the remnant effective piezocoefficient, which is proportional to polarization order parameter in proper ferroelectrics. However, it can also arise from a number of other mechanisms that can greatly complicate the interpretation of the measured hysteresis loops.²⁰ Furthermore, in relaxor ferroelectrics,^{21–24} the nature of order parameter is significantly more complex,^{25–27} as therefore is the nature of PFM signal formation.

However, the ability to reconstruct the phase diagram from local spectroscopic PFM measurements could be useful, because such experiments are often simpler to perform (machines and software are commercially available), and can be carried out on extremely small sample sizes (<~20 nm). More importantly, these measurements provide access to the phase transition at the surface, which can differ significantly from the bulk due to

¹Department of Physics and Institute of Functional Nanomaterials University of Puerto Rico San Juan PR 00936, USA; ²Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, MD 207425, USA; ³Department of Physics & Astronomy, National Institute of Technology, Rourkela 769008, India; ⁴Center for Nanophase Materials Sciences Oak Ridge National Laboratory Oak Ridge TN 37831 USA and ⁵Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Correspondence: Rama K. Vasudevan (vasudevanrk@ornl.gov)

⁶Present address: Geophysical Laboratory, Carnegie Institution for Science, Washington, DC 20015, USA

⁷Present address: Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA

Received: 14 August 2017 Revised: 14 March 2018 Accepted: 29 March 2018

Published online: 25 April 2018

symmetry breaking, chemical segregation, adsorbates, etc.²⁸ and thus enables the study of how the phase transition is affected by these factors.

As a secondary difficulty in constructing phase diagrams from experimental data, the uncertainty in these measurements is often hard to compute. Since measurement points in the composition-temperature space are discrete, demarcating the phase boundary requires interpolation, typically using functional forms such as high-order polynomials, or piecewise fitting such as cubic splines. However, these approaches have the disadvantage in that they impose a form of the function on the data, and cannot provide robust uncertainty quantification (the uncertainty depends on the choice of fitting model). Gaussian process (GP) methods are a class of machine learning, and have been utilized to tackle this problem, and provide an elegant (if computationally expensive) method that allows for regressions of data without the need to postulate an underlying functional form. The GP regression is non-parametric, naturally extends to higher-dimensional spaces, and provides robust uncertainty quantification.²⁹

In this letter, we show the use of these methods in exploring the phase diagram of a relaxor-ferroelectric system, $(1-x)$ Pb($\text{Fe}_{0.5}\text{Nb}_{0.5}$) $\text{O}_3-x\text{Ni}_{0.65}\text{Zn}_{0.35}\text{Fe}_2\text{O}_4$ ($x = 0, 0.10, 0.20,$ and 0.30). The four distinct compositions are investigated using band-excitation piezoresponse spectroscopy (BEPS),^{30,31} yielding hysteresis loops as a function of temperature for all four samples below and above the Curie temperature. Basic clustering methods allow coarse determination of the phase diagram, which is compared to the bulk phase transition temperatures as deduced from dielectric spectroscopy measurements. Gaussian process regression is then used to model the response as a function of temperature (T) and voltage (V), providing uncertainty bounds in the temperature space, and allow for the prediction of the response at arbitrary points in the 2D $T-V$ space, resulting in a full reconstruction of the phase diagram. Clustering methods are carried out on the GP-predicted loops and compared to the results from the raw hysteresis loops, indicating similar trends. The results point to a reduction in the surface phase transition temperature with respect to the bulk for the $x = 0.2$ composition, with little difference seen for all other compositions investigated. Further, we use the method to predict hysteresis loops for a composition unseen by the algorithm, and compare it with experimental results, confirming the validity of the approach. The method shown here is

general for determining phase diagrams from local response measurements, allows for uncertainty in the data to be accurately determined, and represents an important step toward automated high-throughput characterization via local spectroscopic measurements in scanning probe microscopy.

RESULTS

The $(1-x)$ Pb($\text{Fe}_{0.5}\text{Nb}_{0.5}$) $\text{O}_3-x\text{Ni}_{0.65}\text{Zn}_{0.35}\text{Fe}_2\text{O}_4$ ($x = 0, 0.10, 0.20,$ and 0.30) films investigated in this study will be abbreviated as PFN, PN1, PN2, and PN3, respectively (see “Methods” for sample preparation details). We abbreviate Pb($\text{Fe}_{0.5}\text{Nb}_{0.5}$) O_3 as PFN and $\text{Ni}_{0.65}\text{Zn}_{0.35}\text{Fe}_2\text{O}_4$ as NZFO. This system is a perovskite/spinel composite that shows promise in magnetoelectric applications. The Pb($\text{Fe}_{0.5}\text{Nb}_{0.5}$) $\text{O}_3-x\text{Ni}_{0.65}\text{Zn}_{0.35}\text{Fe}_2\text{O}_4$ composite structures have been chosen for the present study to observe enhanced ferroelectric and magnetic properties with strong magnetoelectric coupling above room temperature. PFN is a well-known multi-ferroic material with a ferroelectric Curie temperature between 379 and 385 K having high dielectric constant, low loss tangent, and high piezoelectric constant; it also shows magnetic ordering below RT ($T_N \sim 150-200$ K).³²⁻³⁵ Recently, the existence of ferroelectric, ferromagnetic, and ferroelastic properties in PFN with interesting ferroelectric and magnetic properties have been reported.³⁶ On the other hand, Nickel zinc ferrites are soft magnetic materials with high saturation magnetization, low coercivity, high resistivity, reasonable magnetostriction, low dielectric losses, high dielectric constant, and high magnetic Curie temperatures. $\text{Ni}_{0.65}\text{Zn}_{0.35}\text{Fe}_2\text{O}_4$ has been chosen for the present work as this composition exhibits the combination of highest saturation magnetization and high magnetostriction in the entire Ni-Zn series with a magnetic Curie temperature of ~ 710 K.³⁷ As such, exploration of the phase diagram of this system in terms of ferroelectric properties is important in material optimization, and provides the basis for our work.

To examine the phase purity and crystalline quality of PFN, PN1, PN2, and PN3 thin films of 70 nm thickness grown on LaNiO_3 (LNO)-buffered (LaAlO_3)_{0.3}($\text{Sr}_2\text{AlTaO}_6$)_{0.7} (LSAT) substrate, X-ray diffraction (XRD) measurements were carried out at room temperature and the XRD patterns for PFN sample are shown in Fig. 1a. The $\theta-2\theta$ large-angle XRD patterns ($20-80^\circ$) showed only the diffraction peaks from the substrate and (00 l) pseudocubic

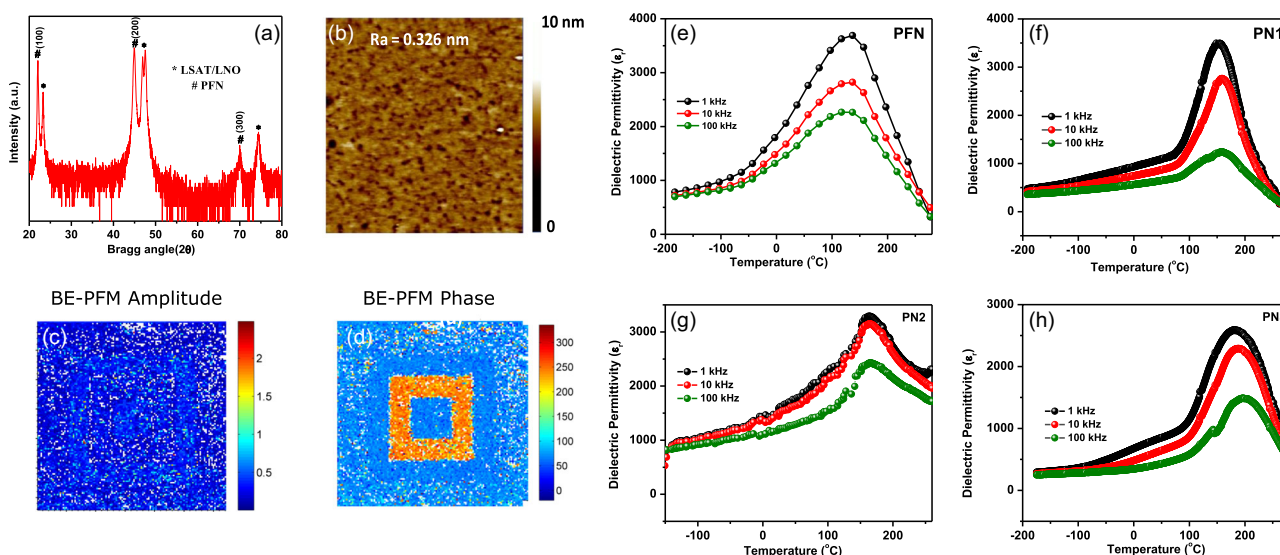


Fig. 1 Surface and bulk characterization of PFN composition series. **a** XRD 2θ scan of PFN thin film, **b** ($5 \mu\text{m} \times 5 \mu\text{m}$) AFM topography image showing low surface roughness of the film. **c**, **d** ($5 \mu\text{m} \times 5 \mu\text{m}$) Band-excitation piezoresponse force microscopy of the PFN sample after poling with ± 8 V applied to the AFM tip, with amplitude image shown in **c** and associated phase **d**. **e-h** Dielectric spectroscopy measurements as a function of temperature and frequency for the four distinct compositions studied

reflections from PFN and NZFO confirms that these films are highly *c*-axis oriented in nature, and that the secondary phases are absent. The XRD peaks with symbols *, #, and & correspond to the peaks of LSAT/LNO (overlapped), PFN, and NZFO, respectively. The superposition of both the crystallographic phases of PFN and NZFO are observed in the XRD patterns, confirming the formation of the composites. The intensity of the peaks corresponding to NZFO phase continues increasing with the increase in ferrite concentration in the composites. No intermediate phases, i.e., extra reflection peaks that would be indicative of secondary phases and peaks from lead-deficient pyrochlore phases were detected, indicating the high purity and high quality of the thin films (see supplementary for XRD scans of all samples). The atomic force microscopy topography image for the PFN sample is shown in Fig. 1b, along with the vertical PFM amplitude and phase image in Fig. 1c, d. The results indicate that the film is extremely flat, and monodomain in the as-grown state, but can be switched. All composite thin films also show similar behaviors.

To probe the ferroelectric phase transitions and to explore the effect of concentration of the magnetic material (NZFO) on the nature of phase transition in composite thin films, the dielectric permittivity of PFN, PN1, PN2, and PN3 as a function of temperature have been measured in a wide range of frequencies and is shown in Fig. 1e–h, respectively. It is observed that for all compositions the dielectric constant increases with increasing temperature, attains its maximum value at a certain temperature and then decreases. The temperature at which maximum value of dielectric constant is observed corresponds to the ferroelectric to paraelectric phase transition temperature (T_C). Single crystal and ceramic samples of PFN exhibit diffuse type of ferroelectric phase transition having T_C ranging from 379 to 385 K.³⁸ From the dielectric spectra, the values of T_C were determined (see “Methods”) for PFN, PN1, PN2, and PN3, which were found to be 400, 430, 440, and 460 K, respectively. The significantly larger values of the transition temperature as compared to the bulk may be attributed to the compressive epitaxial strain,³⁹ which is expected to be on the order of ~4.6%. Additionally, the transition temperature increases with increase of the magnetic material (NZFO) concentration. This behavior, i.e., the increase of ferroelectric T_C with increase of magnetic content has been theoretically predicted for 1–3 type multiferroic composite thin films.⁴⁰ All the measured thin films exhibit a diffuse (broadening) type of phase transition behavior.

The spectra can be fit to a modified Curie Weiss law to describe the diffuseness of the phase transition⁴¹ and is given by the relation

$$\frac{1}{\epsilon} - \frac{1}{\epsilon_m} = \frac{(T - T_m)^\gamma}{C_1} \quad (1)$$

where γ corresponds to the degree of relaxation and its value varies from $1 < \gamma < 2$. For the limiting value of $\gamma = 1$, the above equation reduced to the Curie Weiss law which is applicable to normal ferroelectrics. But in the case of $\gamma = 2$, the quadratic dependence is applicable to relaxor-ferroelectric materials.^{41,42} It should be noted that when γ varies from 1 to 2, this parameter is a measure of the degree of relaxation. Simple models that assume microscopic compositional fluctuations, where the micro-regions have similar dielectric properties but different Curie temperatures, can reproduce the $\gamma = 2$ case, but in real situations this is actually more complex. From fitting of the dielectric spectra, the value of γ for PFN, PN1, PN2, and PN3 are found to be 1.98 ± 0.05 , 1.95 ± 0.01 , 1.80 ± 0.01 , and 1.97 ± 0.02 , respectively (reported uncertainties are 1 s.d.). Thus, the PN2 sample appears to stand out from the other three samples in that the diffuseness of the phase transition is actually lowered. Elaboration of the reasons behind this difference is provided in the “Discussion” section.

Having established the main features of the bulk phase diagram of the system, we next turned to local measurements of the

hysteresis loops using BEPS. Briefly, this involves applying a DC train of pulses comprising a triangular amplitude envelope, with BE packets applied after each DC pulse to determine the strain state (piezoresponse) of the system. Here, we applied BE packets both during (on-field) and after (off-field) the DC pulse application, but from hereon we will focus on the off-field piezoresponse only. Measurements were performed on a 5×5 spectroscopic grid across a [$1 \mu\text{m} \times 1 \mu\text{m}$] region of each sample, at different temperatures. At each point in the spectroscopic grid, two hysteresis loops were acquired, resulting in 50 hysteresis loops at a given temperature for each sample investigated. The average of the 50 hysteresis loops, plotted as a function of temperature for each of the four compositions studied, are shown in Fig. 2. The data indicate that all four samples show a phase transition, although the collapse of the hysteresis loops (and onset of the paraelectric phase) is more rapid with temperature for the PFN than for the other three compositions.

Clustering

To map the phase diagram based on the shape of the hysteresis loops, we turned to the *k*-means clustering algorithm. This is a common method used to cluster data in high dimensional spaces, and aims to segment data into *k* sets (clusters), such that the within-cluster sum of squares is minimized.⁴³ We performed the *k*-means clustering method (as implemented in the open source scikit-learn python package⁴⁴) on the BEPS loop data, with $k = 3$ (clusters). It should be noted that each composition was investigated at similar (but slightly different) temperatures, but the data were grouped together before clustering (this is explored later). The results are plotted in Fig. 3a, b, with the labels in Fig. 3a and the cluster centers plotted in Fig. 3b. Given that there are 50 hysteresis loops at each temperature and composition, the *k*-means clustering algorithm assigns each hysteresis loop a label as belonging to one of the three clusters. In the RGB labels diagram, the color of each pixel in the phase diagram is indicative of the weights, i.e., entirely red pixels signify that all 50 loops are classified as belonging to cluster 0, entirely green are classified as cluster 1, and entirely blue are cluster 2. The average response within each cluster (“Cluster Centers”, Fig. 3b) indicates that cluster 1 is effectively the paraelectric phase, while the red cluster signifies the ferroelectric phase. Meanwhile, a cluster with much larger response is present for the PFN, due to the superior piezoelectric characteristics of this composition and is colored in blue. Remarkably, the simple clustering algorithm is capable of automatically demarcating the phase diagram from the given data set.

Gaussian processes modeling

As noted above, each sample was investigated at slightly different temperatures, which is visualized in Fig. 4a. To adequately compare each measurement across samples, one needs to resample the measured hysteresis loops in the temperature–voltage space. This resampling can be performed using a number of methods, ranging from functional fits to multidimensional interpolation techniques. To surmount difficulties associated with choosing the appropriate form of a functional fit or the unknown statistical confidence of interpolated data, we utilized Gaussian processes modeling (GP). The latter is a multidimensional and non-parametric Bayesian modeling technique used in various statistical learning tasks such as regression and classification. In regression, one aims to find an optimized function *f* that maps the inputs **x** (i.e., vectors of temperature and DC bias) to output **y** (hysteresis loops), given some observations **Y** of the latter at some temperature and bias data points.

Briefly, in Gaussian processes modeling, the input vectors are projected into a higher-dimensional space via a mapping

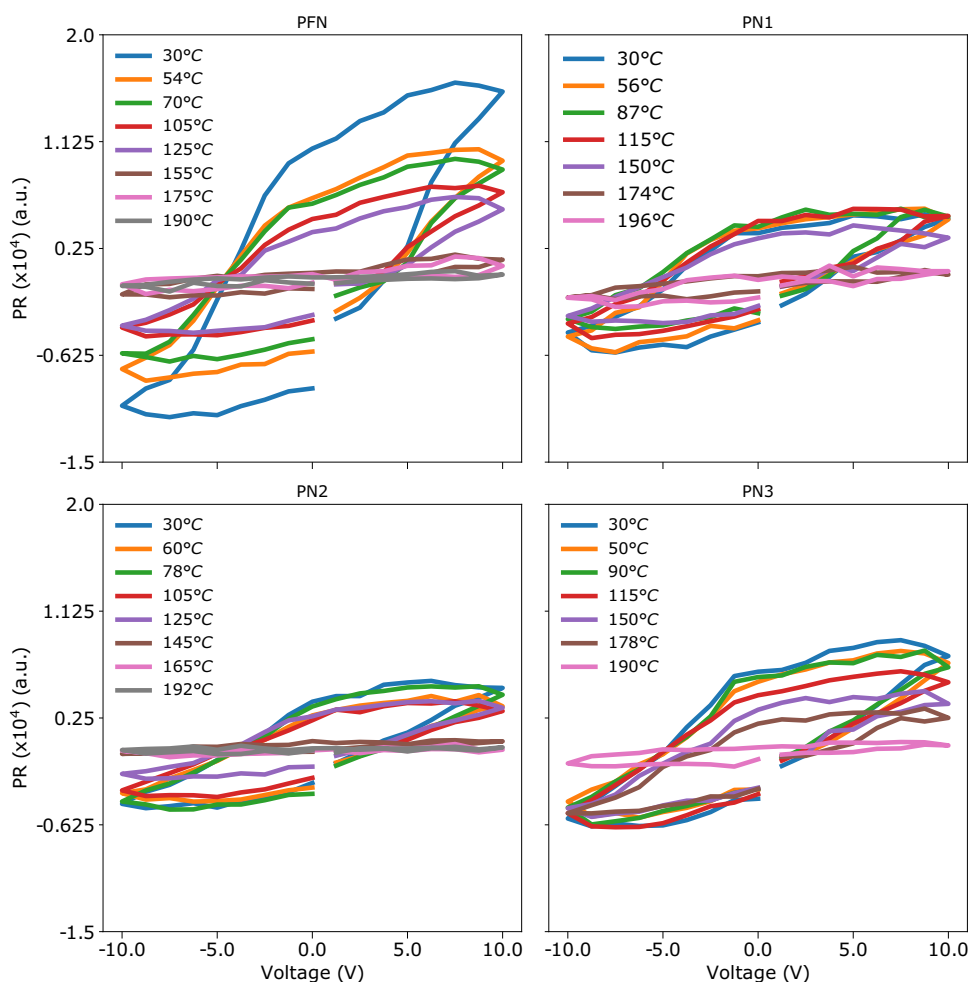


Fig. 2 Local characterization of samples with band-excitation piezoresponse spectroscopy. Shown are average (median) of 50 hysteresis loops captured at each temperature, for each composition

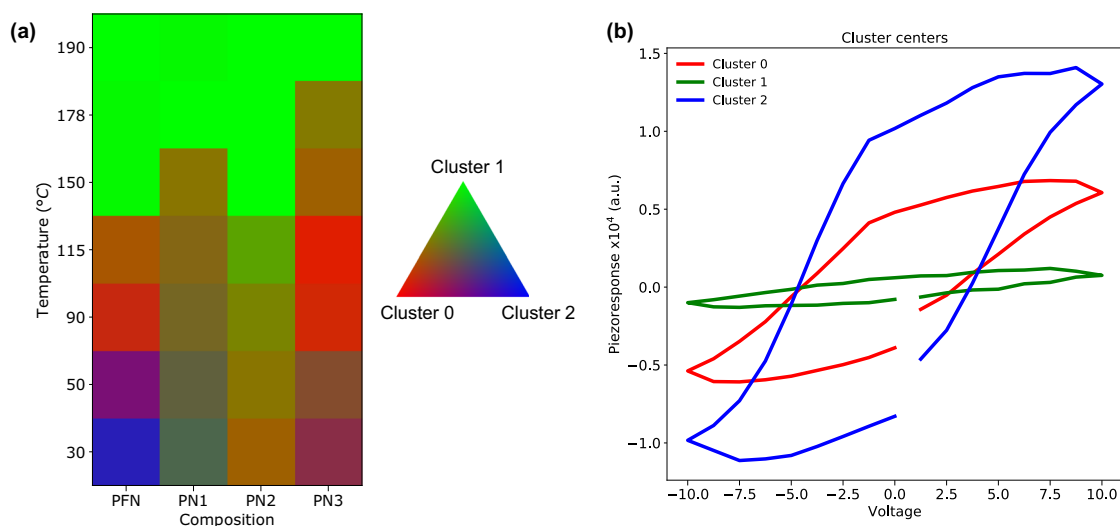


Fig. 3 *k*-means clustering on the raw hysteresis loop data. **a** Label matrix and **b** cluster centers for *k*-means applied to the hysteresis loop data in Fig. 2. The green cluster appears to represent the paraelectric phase, and the red and blue are ferroelectric phases. The color of the squares in **a** correspond to the number of loops within each acquisition composition and temperature that reside in a particular cluster; red colors signify predominant membership in the 0th cluster, green indicates membership in the first cluster, and blue indicates membership in the second cluster, as shown in the RGB triangle

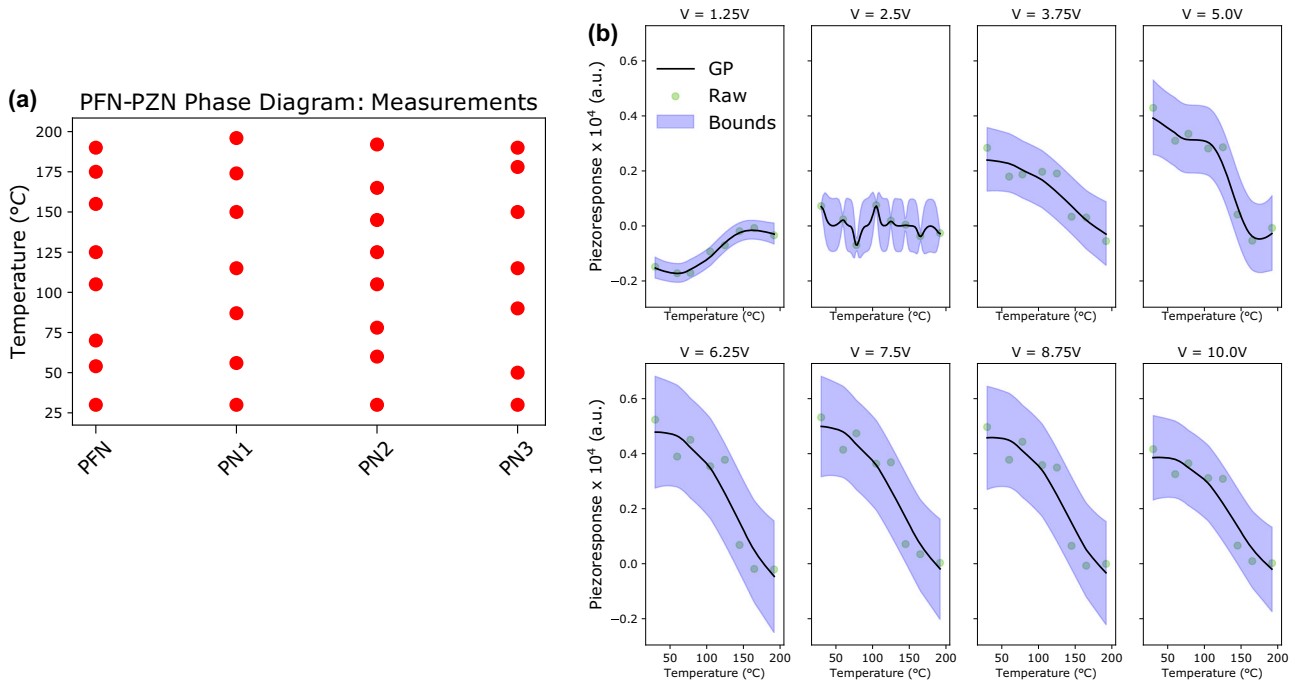


Fig. 4 Interpolation in temperature. **a** Measurements taken on the four samples at different temperatures. **b** Gaussian process modeling of hysteresis loops in temperature for the PN2 sample. We plot the response at a particular voltage as a function of temperature (circles, “Raw”), and GP modeling allows to predict the response in the temperature space with confidence (black line, “GP” with 2 s.d. intervals, “Bounds”)

$\phi_1 = \exp\left(-\frac{1^2}{2\ell} |x - x'|^2\right)$. This allows to model both linear as well as nonlinear functions, f and is now given by

$$f(\mathbf{x}) = \phi(\mathbf{x}^T \mathbf{w}), \text{ and } \mathbf{w} \sim \mathcal{N}(\mathbf{m}, \Sigma), \quad (2)$$

where \mathbf{w} are (unknown) parameters that define f . The parameters of the model are assumed to follow a (prior) multivariate normal distribution \mathcal{N} with mean values \mathbf{m} , and covariance matrix Σ . The covariance matrix defined by $\Sigma = \sigma^2 K(\mathbf{x}, \mathbf{x}')$ is fully determined once the kernel K is chosen, while the variance σ is estimated by maximum likelihood using the observations \mathbf{Y} (incidentally, the choice of K also fully specifies the mapping ϕ). The most popular kernel is the Gaussian correlation function or radial basis function (RBF), given by

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\ell^2} |\mathbf{x} - \mathbf{x}'|^2\right), \quad (3)$$

where ℓ is a hyper-parameter of the GP model. To take into account the noise inherent in each measurement, we used a RBF kernel with a white kernel, i.e.,

$$K = \alpha\phi_1 + \beta\phi_2$$

where $\phi_1 = \exp\left(-\frac{1^2}{2\ell} |x - x'|^2\right)$ is the RBF Kernel in (3), and ϕ_2 is a noise kernel given by

$$\phi_2 = \begin{cases} \epsilon, & x = x' \\ 0, & \text{else} \end{cases} \quad (4)$$

where ϵ is a noise level. We choose $\alpha = 0.1$ and $\beta = 0.5$, with bounds on the noise level ϵ ($[1 \times 10^{-10}, 1 \times 10^{-2}]$).

Once an optimal value of the hyper-parameter is found (by maximizing the log-likelihood of the model), one arrives at an optimal function f , which can be used to predict hysteresis loops at new temperature and bias points. The predictions of hysteresis loops using GP regression (and the RBF kernel) along the temperature axis for the PN2 case is seen in Fig. 4b, for different individual voltages, along with a confidence interval (2 s.d.) filled

in blue. This uncertainty quantification is an advantageous feature of GP regression, and can be useful for applying or developing models to describe the nature of the phase transition.

The GP regression is also fully generalizable across higher dimensions, although the computational cost is higher given that it is not a sparse method. It would appear that the uncertainties represented in Fig. 4b are rather large, and are a result of the sparsity of data points. In a sense, this is the price to be paid for unknown functional forms; while other methods (such as polynomial or spline interpolation) could work, the uncertainty estimates would not be accurate (given that this uncertainty depends on the model choice, which is largely arbitrary, especially for ferroelectric hysteresis loops⁴⁵). However, the ability to generalize and produce predictions with quantified uncertainty is of critical importance, and not available through the other standard means.

Shown in Fig. 5 are the GP regressions on the 2D surfaces of the average (i.e., spatial average) piezoresponse as a function of voltage and temperature for all four compositions. In Fig. 5a–d, the raw data and the GP-predicted loops as a function of temperature are plotted for each composition; in general, the GP-predicted loops closely align with the raw data for each composition. The full GP-predicted 2D surface plots are shown in Fig. 6a–d for the four compositions, and can be regenerated for arbitrary values using the model. The distinct behaviors and closure of the hysteresis loops with increasing temperature are readily observed in this representation. Given the wrapped nature of the 3D plots, we also provide the surface as a function of voltage step, effectively unfolding the hysteresis loops along the voltage axis (see supplementary material).

With the GP-predicted data sets for each composition, we were then able to perform K -means on the full predicted data set containing the loops at different compositions and temperatures. In other words, we took the data in Fig. 6 from the Gaussian process prediction and concatenated them into a single 2D matrix, and subsequently performed k -means clustering. The cluster centers and the label map is shown in Fig. 7a, b, and again clearly demarcates the line between the paraelectric and ferroelectric

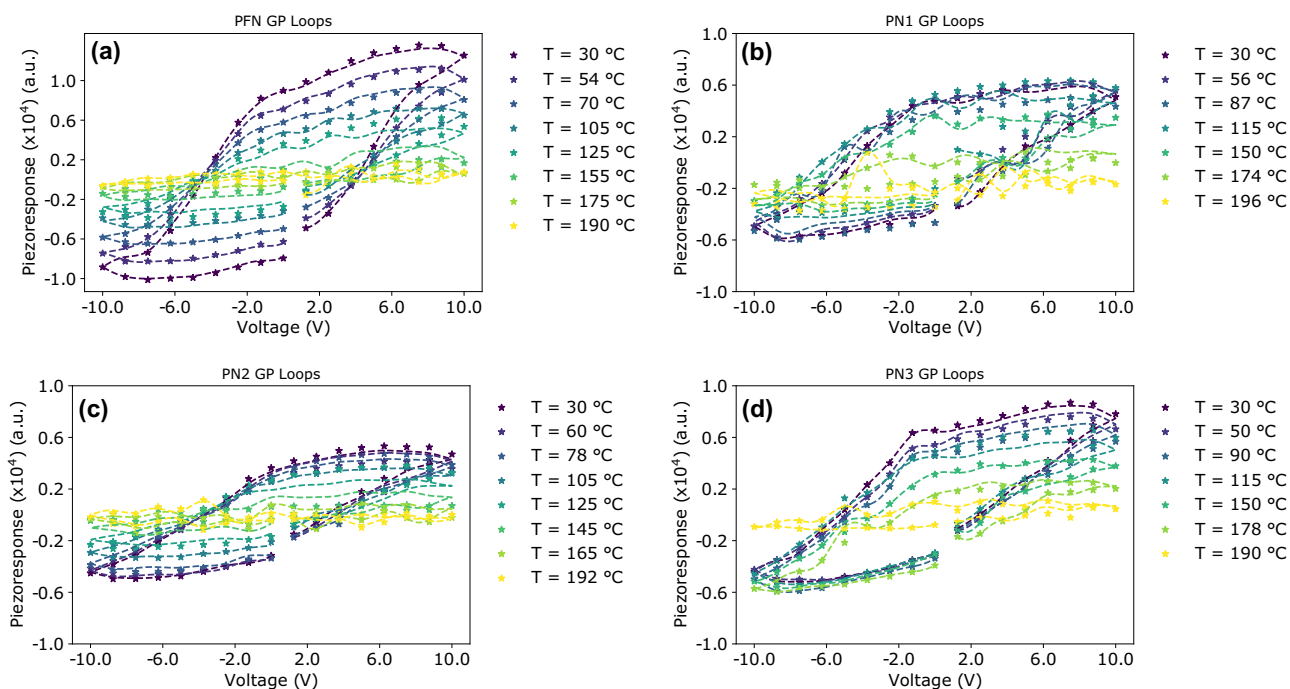


Fig. 5 Gaussian process modeling in 2D. **a–d** Predicted hysteresis loops (dashed lines) with raw data (star markers) as a function of temperature, for the four compositions

phases. Note that here we performed this for $k=4$ clusters (as opposed to 3 earlier), because we found it more accurately depicted the changed loop behavior of the PN2 sample (nonetheless, trends are similar for $k=3$ clusters, shown in supplementary, with the main difference being slightly lowered T_c values, though with a more drastic reduction for the PN2 case). Based on the label matrix, we can determine the phase transition temperatures for the surface, and compare it with the phase transition temperature of the bulk, as shown in Fig. 7c. The surface phase transition temperatures are marked as crosses, while the bulk phase diagram is plotted as colored circles (the latter are taken from the dielectric spectroscopy data in Fig. 1). As with Figs. 3a and 7a, the most interesting feature of this phase diagram is the behavior of PN2, which appears markedly different from the other compositions.

Using GP for prediction in composition space

Perhaps of most importance is the ability to predict functionality as a function of composition,^{9,11,46–48} given that this can potentially reduce the time spent on exploring the composition space (which is generally more time consuming than performing measurements at different conditions, for instance). We attempt to do this on our existing system, for the PN2 composition, given hysteresis loops at a single temperature for the PFN, PN1, and PN3 compositions. The first step is to perform the GP in the composition space given the input data (three hysteresis loops, one for each composition). The results of the GP are shown in Fig. 8a, with the mean plotted as a black line and the variance estimates as filled blue regions. We can then use the trained model to predict the hysteresis loop for the missing PN2 composition, which is shown in Fig. 8b. Note that the response is plotted against the voltage step (i.e., time), not voltage, to more easily visualize the upper and lower confidence bounds (see supplementary for plot as a function of voltage). The actual measurements of the PN2 composition are shown as filled circles with associated standard deviation (arising from spatial variation in the measured hysteresis loops).

In all cases, the actual measured values are within 1 s.d. of the prediction bound, suggesting good agreement and validation of the GP model.

DISCUSSION

The phase transition temperatures as measured by PFM are slightly different from the bulk, especially for the PN2 sample. This is likely due to a combination of the fact that PFM is not measuring the polarization directly (piezoresponse is strain), whereas macroscopic dielectric spectra measure the film polarization as well as interfacial polarization; furthermore, PFM is inherently surface sensitive, with associated issues relating to change of the surface crystal structure, chemical segregation, or symmetry breaking, which can all affect the nature and temperature of the transition in relaxor-ferroelectric systems.²⁸ Indeed, we found that the PN2 transition appears more diffuse in the local measurements, as evidenced by the more gradual transition from green to red in Fig. 3a, whereas it shows a less diffuse transition than the other compositions in the dielectric spectroscopy. There can be several reasons for the discrepancy in the phase transition temperature of the PN2 sample with respect to the others; in addition to the basic substrate strain-related effects, given this is a composite, one can expect effects from the strain impact of one phase on another, as well as possible interdiffusion of cations across the interfaces in this system leading to more spatial variability. It is also possible that there is an interplay between the ferroelectric and magnetic order parameters of the system that can affect the T_c in a nonlinear fashion in the composition space, especially as the parent compound PFN is known to exhibit weak ferromagnetism up to 400 K.⁴⁹ Additional surface-sensitive local PFM investigations can target the influence of surface-based modifications of the sample, e.g., local tip pressure, to uncover the main driving mechanisms behind the locally diffuse phase transition in PN2. The results confirm the utility of the clustering approach in demarcating the important

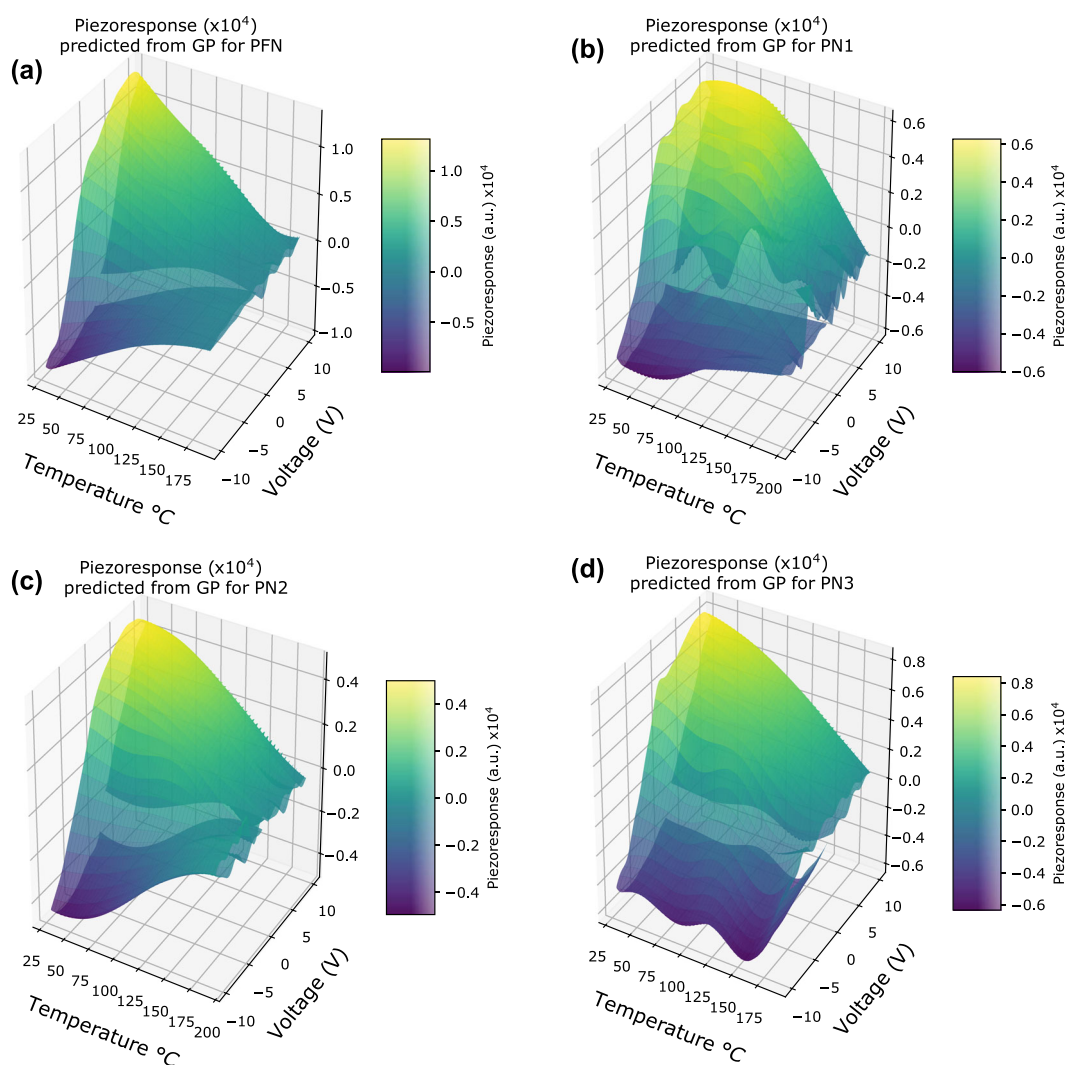


Fig. 6 Gaussian process modeling. **a–d** Full predicted piezoresponse values as a function of temperature and voltage, for the four different compositions studied plotted in **a–d**, respectively

features of the phase diagram, as well as illustrate the differences between the surface and bulk measurements in this system.

An interesting future direction in the application of Gaussian processes in the reconstruction of phase diagrams from local measurements is the prediction of the onset of diffuse transitions in relaxor-ferroelectric systems as a function of composition. Traditionally, an accurate determination of the dependence of piezoelectric and dielectric properties in relaxors on the dependence of ionic substitution requires the synthesis of a large number of materials. As we demonstrated, however, Gaussian processes machine-learning techniques can accurately reconstruct phase diagrams in the presence of a finite and small subset of temperature and voltage measurements. Therefore, one may explore the feasibility of accurately predicting the ionic substitution ratio at which relaxor-ferroelectric behavior takes place from a small and finite number of material compositions via an application of GP machine learning.

The local and global ferroelectric phase transitions of a relaxor-ferroelectric system were probed using dielectric and band-excitation PFM measurements. Through acquisition of hysteresis loops as a function of temperature for four distinct compositions, we compiled a large data set of the functional response of the system. Basic unsupervised clustering algorithms applied to the

data set facilitated the reconstruction of the temperature-composition phase diagram. Given the discrete and non-uniform nature of the measured parameter space, as well as inherent nonlinear dependence of the hysteresis loops on thermodynamic potentials, we employed Gaussian process regression to reconstruct the data in both the 1D case (temperature) and the 2D case (temperature and voltage), accompanied with confidence quantification without imposing functional forms. Clustering was performed on the GP-reconstructed data set to reproduce the phase diagram, with results compared to bulk dielectric spectroscopy and allowing comparison between surface and bulk transition temperatures. These techniques can be used for automated determination of phase diagrams with uncertainty quantification, across a wide range of imaging modalities.

METHODS

Pulsed laser deposition

Phase pure ceramic targets of $\text{Pb}(\text{Fe}_{0.5}\text{Nb}_{0.5})\text{O}_3$ and $(1-x)\text{Pb}(\text{Fe}_{0.5}\text{Nb}_{0.5})\text{O}_3-x\text{Ni}_{0.65}\text{Zn}_{0.35}\text{Fe}_2\text{O}_4$ ($x = 0.10, 0.20, \text{ and } 0.30$) composites were synthesized by high-temperature solid-state reaction technique. The detailed synthesis conditions and structural properties of the above-mentioned ceramic targets are reported elsewhere.²¹ PFN, PN1, PN2, and PN3 thin films were grown on LaNiO_3 (LNO) bottom electrode deposited on

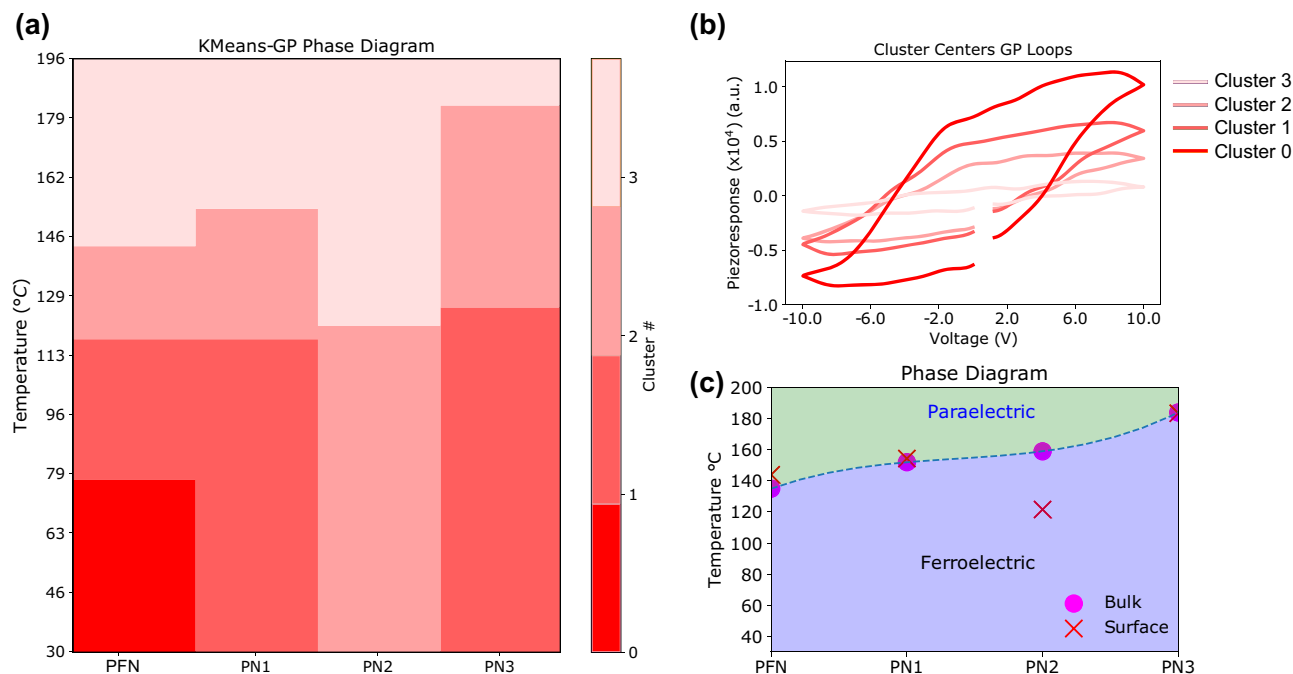


Fig. 7 Comparing the phase diagrams. **a**, **b** Label Matrix (**a**) and Cluster centers (**b**) after performing K -means on the full GP-predicted data set (from Fig. 5), for $k = 4$ clusters. This allows one to calculate the transition temperature from the PFM measurements, which are plotted on top of the bulk phase diagram in **c**. The bulk phase transition temperatures are taken from the dielectric spectroscopy in Fig. 1. The transition temperatures are mostly similar, except for the anomalous PN2 composition

commercially available $(\text{LaAlO}_3)_{0.3}(\text{Sr}_2\text{AlTaO}_6)_{0.7}$ (LSAT) (100) substrates using optimized pulsed laser deposition (PLD) with an excimer laser (KrF, 248 nm). Initially, LNO bottom electrode of 50 nm was deposited on LSAT substrate at 700 °C in an oxygen pressure of 200 mTorr, followed by annealing in same oxygen ambient of 300 Torr for 30 min at same temperature and then slowly cooled down to room temperature. PFN, PN1, PN2, and PN3 thin films were deposited at a fixed temperature of 600 °C under oxygen ambient of 20 mTorr. The laser pulse repetition rate was 2 Hz with a laser energy density ~ 1.5 J/cm². Later, the PFN and other composite thin films were annealed in an oxygen atmosphere of 300 Torr for 30 min at 700 °C and later cooled down to room temperature slowly. The thickness of all above-mentioned thin films were kept constant ~ 70 nm for the comparison of physical functionalities.

X-ray diffraction

The phase purity and crystallinity of PFN, PN1, PN2, and PN3 thin films were examined using high-resolution X-ray diffraction (HRXRD) using CuK_α radiation with wavelength of $\lambda = 1.5405$ Å operated at a scan rate of 1°/min over the angular range (2θ) of 20–80 at room temperature. The thicknesses of all the thin films were measured utilizing XP-200 profilometer along with Filmetrics F-20.

Dielectric characterization

Pt top electrodes of area $\sim 10^{-4}$ cm² and thickness ~ 40 nm were grown by dc sputtering utilizing a metal shadow square mask for electrical characterization. Temperature-dependent dielectric parameters, i.e., capacitance, loss tangent, phase angles, and impedance were measured from 100 to 550 K in a wide frequency range using an impedance analyzer HP4294A with MMR Technologies K-20 programmable temperature controller with fixed ac voltage amplitude of 0.1 V. To determine the ferroelectric-paraelectric phase transition temperature, we plotted the derivative of inverse of dielectric permittivity ($1/\epsilon$) against temperature for the frequency 10 kHz. When the derivative tends to zero, it is known as the phase transition temperature, and the phase transition temperature of PFN, PN1, PN2, and PN3 were found to be 400, 430, 440, and 460 K, respectively.

Band-excitation piezoresponse spectroscopy

The PFM spectroscopy measurements were carried out by moving the tip across a pre-defined coordinate (x, y) grid on the surface of the samples and perturbing it with a DC bias waveform, and then measuring the strain (piezoresponse) as a function of frequency via band-excitation (BE) technique. BE technique includes the generation of a band of AC frequencies near the contact resonance of the cantilever and measure the electromechanical response over time with subsequent Fourier transformation back to the frequency domain to generate the frequency-dependent response. This measurement is repeated for many time steps after every DC bias pulse; thus, at each (x, y) position, local spectroscopic information is obtained as a function of perturbation voltage (V) and excitation frequency (f), characterizing the system's mechanical response (vertical deflection) $R = R(x, y, V)$. The response R at each (x, y, V) step is fitted to a simple harmonic oscillator (SHO) model yielding the amplitude (A), phase (ϕ), quality factor (Q), and resonant frequency (ω) corresponding with the response. The thin film was glued onto a sample plate with silver epoxy. The PFM experiments were carried out using moderately stiff Budget Sensors ElectriMulti75-G cantilevers ($k \sim 1$ N/m) and a free resonance frequency (in air) of ~ 75 kHz, on a multimode (Veeco) AFM equipped with a Nanonis controller. National Instruments DAQ cards were utilized for signal generation and acquisition for the band-excitation measurements, which were performed using scripts written in Labview v11 and python 3. All analyses were also performed utilizing python 3, including scikit-learn⁴⁴ for Gaussian Process implementation and Pycroscopy (<https://pycroscopy.github.io/pycroscopy/>) for BEPS data analysis. Temperature-dependent PFM spectroscopic measurements were carried out in a wide temperature range of 300–600 K with a high precision temperature controller. In order to avoid the moisture effect, all the thin films were heated up to 50 °C for 30 min prior to the measurements.

Data availability

The data and Jupyter notebook that details the analysis are included in the supplementary files accompanying this manuscript.

DISCLAIMER

This manuscript has been authored by UT-Battelle, LLC, under contract no. DE-AC0500OR22725 with the U.S. Department of

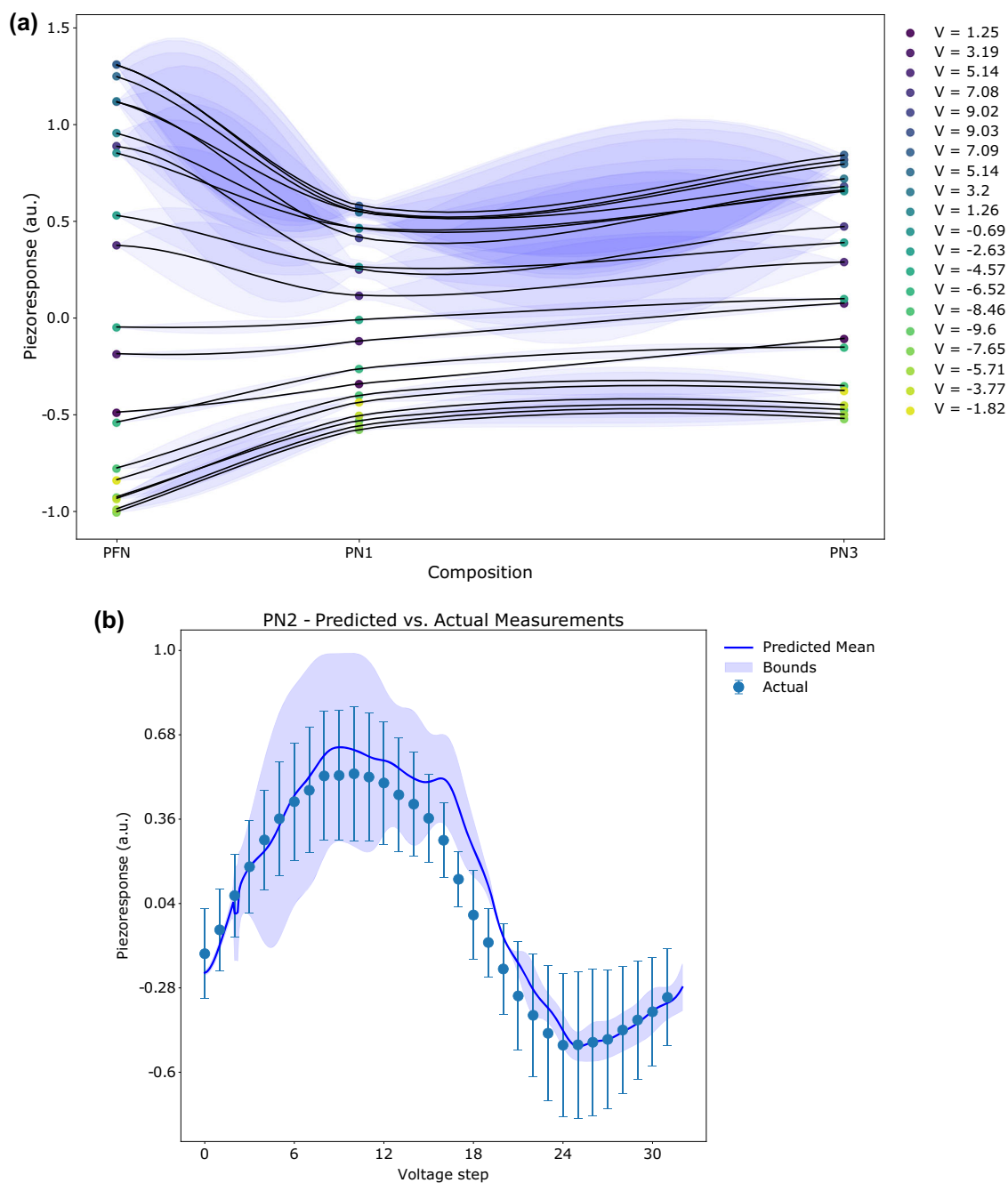


Fig. 8 GP prediction across compositions. **a** GP performed in composition space, for hysteresis loops captured at 30 °C. The measurements are plotted as filled circles, with the color corresponding to the voltage at which the piezoresponse measurement was acquired. Solid black lines indicate the mean from the GP regression, and blue regions indicate the variance estimates (1 s.d.). **b** Comparison of predicted response with the measured response. The predicted mean for this composition is plotted as a solid blue line, with measurements (and 1 s.d.) of this composition plotted as filled circles. The shaded blue regions indicate the variance estimates from GP. The measurements fall within 1 s.d. of the blue bounded regions, confirming the validity of the approach

Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for the United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doi-public-access-plan>).

ACKNOWLEDGEMENTS

The PFM and Gaussian process portion of this research was supported by the U.S. Department of Energy, Office of Science, Materials Sciences and Engineering Division (R.K.V. and S.V.K.). The synthesis and characterization of samples work was supported by DoD-AFOSR (Grant #FA9550-16-1-0295). D.K.P. and S.K. acknowledge IFN (NSF Grant No. EPS-01002410) for fellowship. The scanning probe microscopy studies were conducted at the Center for Nanophase Materials Sciences, which is a US DOE Office of Science User Facility. N.L. acknowledges support from the Eugene P. Wigner Fellowship program at Oak Ridge National Lab. D.K.P. and R.S.K. acknowledge CNMS facilities through CNMS user Proposal ID: CNMS2014-095. E.S. acknowledges support

under the Cooperative Research Agreement between the University of Maryland and the National Institute of Standards and Technology Center for Nanoscale Science and Technology, Award 70NANB10H193, through the University of Maryland.

AUTHOR CONTRIBUTIONS

D.K.P. and S.K. synthesized and characterized the samples. Piezoresponse force microscopy measurements were performed at ORNL by the advice of R.K.V., E.S., N.L., and S.V.K. The data were analyzed by D.K.P., E.S., D.K.P., N.L., and R.K.V. R.S.K. and S.V.K. supervised this research work. N.L. developed the Gaussian process modeling, analyzed data, and co-wrote the paper. R.K.V. conceived of the idea, analyzed data, and co-wrote the paper. All authors assisted in writing and commented on the manuscript.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Computational Materials* website (<https://doi.org/10.1038/s41524-018-0078-7>).

Competing interests: The authors declare no competing financial interests.

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

REFERENCES

- Hillert, M. *Phase Equilibria, Phase Diagrams and Phase Transformations: Their Thermodynamic Basis* (Cambridge University Press, Cambridge, 2007).
- Diéguez, O. et al. Ab initio study of the phase diagram of epitaxial BaTiO₃. *Phys. Rev. B* **69**, 212101 (2004).
- Tavares, J., da Gama, M. T., Teixeira, P., Weis, J. & Nijmeijer, M. Phase diagram and critical behavior of the ferromagnetic Heisenberg fluid from density-functional theory. *Phys. Rev. E* **52**, 1915 (1995).
- Li, L., Muckerman, J. T., Hybertsen, M. S. & Allen, P. B. Phase diagram, structure, and electronic properties of (Ga 1 – x Zn x)(N 1 – x O x) solid solutions from DFT-based simulations. *Phys. Rev. B* **83**, 134202 (2011).
- Zhao, J.-C. Combinatorial approaches as effective tools in the study of phase diagrams and composition–structure–property relationships. *Prog. Mater. Sci.* **51**, 557–631 (2006).
- Daniels, J. E., Jo, W., Rödel, J., Honkimäki, V. & Jones, J. L. Electric-field-induced phase-change behavior in (Bi 0.5 Na 0.5) TiO 3–BaTiO 3–(K 0.5 Na 0.5) NbO 3: a combinatorial investigation. *Acta Mater.* **58**, 2103–2111 (2010).
- Potyrailo, R. et al. Combinatorial and high-throughput screening of materials libraries: review of state of the art. *ACS Comb. Sci.* **13**, 579–633 (2011).
- Bunn, J. K. et al. Generalized machine learning technique for automatic phase attribution in time variant high-throughput experimental studies. *J. Mater. Res.* **30**, 879–889 (2015).
- Hattrick-Simpers, J. R., Gregoire, J. M. & Kusne, A. G. Perspective: composition–structure–property mapping in high-throughput experiments: turning data into knowledge. *APL Mater.* **4**, 053211 (2016).
- Stein, H. S., Jiao, S. & Ludwig, A. Expediting combinatorial data set analysis by combining human and algorithmic analysis. *ACS Comb. Sci.* **19**, 1–8 (2017).
- Kusne, A. G. et al. On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets. *Sci. Rep.* **4**, 6367 (2014).
- LeBras, R. et al. in *The 2nd International Workshop on Constraint Reasoning and Optimization for Computational Sustainability* (CROCS at CPAIOR-10, Bologna, Italy, 2010).
- Green, M. et al. Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies. *Appl. Phys. Rev.* **4**, 011105 (2017).
- Iwasaki, Y., Kusne, A. G. & Takeuchi, I. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. *NPJ Comput. Mater.* **3**, 4 (2017).
- Laanait, N., Zhang, Z. & Schlepütz, C. M. Imaging nanoscale lattice variations by machine learning of x-ray diffraction microscopy data. *Nanotechnology* **27**, 374002 (2016).
- Karapetian, E., Kachanov, M. & Kalinin, S. V. Nanoelectromechanics of piezoelectric indentation and applications to scanning probe microscopies of ferroelectric materials. *Philos. Mag.* **85**, 1017–1051 (2005).
- Shin, J. et al. Simultaneous elastic and electromechanical imaging by scanning probe microscopy: Theory and applications to ferroelectric and biological materials. *J. Vac. Sci. Technol. B* **23**, 2102–2108 (2005).
- Kalinin, S. V., Karapetian, E. & Kachanov, M. Nanoelectromechanics of piezoresponse force microscopy. *Phys. Rev. B* **70**, 184101 (2004).
- Kalinin, S. V. & Gruverman, A. *Scanning Probe Microscopy: Electrical and Electro-mechanical Phenomena at the Nanoscale* (Springer, New York, 2006).
- Vasudevan, R. K., Balke, N., Maksymovych, P., Jesse, S. & Kalinin, S. V. Ferroelectric or non-ferroelectric: Why so many materials exhibit “ferroelectricity” on the nanoscale. *Appl. Phys. Rev.* **4**, 021302 (2017).
- Shvartsman, V. et al. Polar nanodomains and local ferroelectric phenomena in relaxor lead lanthanum zirconate titanate ceramics. *Appl. Phys. Lett.* **86**, 202907 (2005).
- Rodriguez, B. J., Jesse, S., Bokov, A. A., Ye, Z.-G. & Kalinin, S. Mapping bias-induced phase stability and random fields in relaxor ferroelectrics. *Appl. Phys. Lett.* **95**, 092904 (2009).
- Shvartsman, V. V., Kleemann, W., Lukasiewicz, T. & Dec, J. Nanopolar structure in Sr x Ba 1 – x Nb 2 O 6 single crystals tuned by Sr/ Ba ratio and investigated by piezoelectric force microscopy. *Phys. Rev. B* **77**, 054105 (2008).
- Shvartsman, V. & Kholkin, A. Domain structure of 0.8 Pb (Mg 1/3 Nb 2/3) O 3–0.2 PbTiO 3 studied by piezoresponse force microscopy. *Phys. Rev. B* **69**, 014102 (2004).
- Glinchuk, M. D. & Stephanovich, V. A. Dynamic properties of relaxor ferroelectrics. *J. Appl. Phys.* **85**, 1722–1726 (1999).
- Westphal, V., Kleemann, W. & Glinchuk, M. D. Diffuse phase-transitions and random-field-induced domain states of the relaxor ferroelectric PBMG1/3NB2/3O3. *Phys. Rev. Lett.* **68**, 847–850 (1992).
- Vugmeister, B. E. & Rabitz, H. Coexistence of the critical slowing down and glassy freezing in relaxor ferroelectrics. *Phys. Rev. B* **61**, 14448–14453 (2000).
- Kholkin, A. et al. Surface domain structures and mesoscopic phase transition in relaxor ferroelectrics. *Adv. Funct. Mater.* **21**, 1977–1987 (2011).
- Rasmussen, C. E. *Evaluation of Gaussian Processes and Other Methods for Non-linear Regression* (University of Toronto, Toronto, 1999).
- 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).
- Jesse, S., Maksymovych, P. & Kalinin, S. V. Rapid multidimensional data acquisition in scanning probe microscopy applied to local polarization dynamics and voltage dependent contact mechanics. *Appl. Phys. Lett.* **93**, 112903 (2008).
- Correa, M., Kumar, A., Priya, S., Katiyar, R. & Scott, J. Phonon anomalies and phonon-spin coupling in oriented PbFe 0.5 Nb 0.5 O 3 thin films. *Phys. Rev. B* **83**, 014302 (2011).
- Kumar, A., Katiyar, R., Rinaldi, C., Lushnikov, S. G. & Shaplygina, T. A. Glasslike state in PbFe 1/2 Nb 1/2 O 3 single crystal. *Appl. Phys. Lett.* **93**, 232902 (2008).
- Correa, M., Kumar, A., Katiyar, R. & Rinaldi, C. Observation of magnetoelectric coupling in glassy epitaxial PbFe 0.5 Nb 0.5 O 3 thin films. *Appl. Phys. Lett.* **93**, 192907 (2008).
- Peng, W. et al. Spin-lattice coupling in multiferroic Pb (Fe 1/2 Nb 1/2) O 3 thin films. *Appl. Phys. Lett.* **94**, 012509 (2009).
- Carpenter, M. A. et al. Elastic and magnetoelastic relaxation behaviour of multi-ferroic (ferromagnetic + ferroelectric + ferroelastic) Pb (Fe0.5Nb0.5) O3 perovskite. *J. Phys. Cond. Mat.* **27**, 285901 (2015).
- Pradhan, D. K. et al. Studies on dielectric, optical, magnetic, magnetic domain structure, and resistance switching characteristics of highly c-axis oriented NZFO thin films. *J. Appl. Phys.* **122**, 033902 (2017).
- Pradhan, D. K., Barik, S. K., Sahoo, S., Puli, V. S. & Katiyar, R. Investigations on electrical and magnetic properties of multiferroic [(1 – x) Pb (Fe0.5Nb0.5) O3 – x Ni0.65Zn0.35Fe2O4] composites. *J. Appl. Phys.* **113**, 144104 (2013).
- Schlom, D. G. et al. Elastic strain engineering of ferroic oxides. *Mrs Bull.* **39**, 118–130 (2014).
- Lu, X., Wang, B., Zheng, Y. & Ryba, E. Critical phase transition temperatures of 1–3 type multiferroic composite thin films. *J. Phys. D Appl. Phys.* **40**, 1614 (2007).
- Smolenskii, G. & Agranovskaya, A. Dielectric polarization of a number of complex compounds. *Sov. Phys. Solid State* **1**, 1429–1437 (1960).
- Uchino, K. & Nomura, S. Critical exponents of the dielectric constants in diffused-phase-transition crystals. *Ferroelectrics* **44**, 55–61 (1982).
- Hartigan, J. A. & Wong, M. A. Algorithm AS 136: a k-means clustering algorithm. *Journal of the Royal Statistical Society. Ser. C* **28**, 100–108 (1979).
- Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- Jesse, S., Baddorf, A. P. & Kalinin, S. V. Switching spectroscopy piezoresponse force microscopy of ferroelectric materials. *Appl. Phys. Lett.* **88**, 062908 (2006).
- Suram, S. K., Haber, J. A., Jin, J. & Gregoire, J. M. Generating information-rich high-throughput experimental materials genomes using functional clustering via multitree genetic programming and information theory. *ACS Comb. Sci.* **17**, 224–233 (2015).
- Ding, S. et al. Combinatorial development of bulk metallic glasses. *Nat. Mater.* **13**, 494–500 (2014).
- Thienhaus, S., Naujoks, D., Pftzing-Micklich, J., König, D. & Ludwig, A. Rapid identification of areas of interest in thin film materials libraries by combining

- electrical, optical, X-ray diffraction, and mechanical high-throughput measurements: a case study for the system Ni–Al. *ACS Comb. Sci.* **16**, 686–694 (2014).
49. Majumder, S. et al. Dielectric and magnetic properties of sol-gel-derived lead iron niobate ceramics. *J. Appl. Phys.* **99**, 024108 (2006).



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