Information Recovery via Matrix Completion for Piezoresponse Force Microscopy Data

Henry S. Yuchi[1] **Kerisha N. Williams**[2] **Gardy K. Ligonde**[2] **Matthew W. Repasky**[1]
Yao Xie[1] **Nazanin Bassiri-Gharb**[2,3]
[1]H. Milton Stewart School of Industrial & Systems Engineering
[2]School of Material Science and Engineering
[3]George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA 30318
{shaowu.yuchi,kerishaw8991,kligonde,mwrepasky}@gatech.edu
yao.xie@isye.gatech.edu
nazanin.bassirigharb@me.gatech.edu

Abstract

Piezoresponse force microscopy (PFM) is a scanning microscopy technique that is used to evaluate the nanoscale strain response to an electric voltage applied to the surface of a ferroelectric material. PFM is a powerful tool for imaging, manipulation, and studying the nanoscale functional response of ferroelectric materials, which has been extensively used as a first-pass test for ferroelectricity in novel materials with unknown functional properties. However, low signal-to-noise ratio observations arising from the loss of electromechanical signal during polarization switching often result in unreliable information extraction at these observations, hampering our understanding of the material characteristics. To address this challenge, we propose an information recovery framework utilizing subspace-based matrix completion to achieve improved characterization from PFM data. It enables us to efficiently recover and extract reliable information from the data, assisting the modeling efforts for PFM and providing insights for characterization and experimentation practices.

1 Introduction

Piezoresponse force microscopy (PFM) is a scanning probe microscopy (SPM) technique used to study the nanoscale electromechanical response of ferroelectric materials [7]. It is a powerful tool for high-resolution imaging, manipulation, and spectroscopic probing of polarization dynamics [12]. PFM has been applied in a range of applications on the nanoscale, including human bones [9], biological systems [13], and complex polymer materials [16]. Furthermore, PFM has been utilized in the characterization of various piezoelectric materials, where further material design and discovery can be leveraged by its capacity to evaluate material properties and derive insights [5, 22]. Machine learning approaches have been brought into studying the material properties and evolution for PFM [14]. PFM can indeed be used to test for ferroelectricity and assist the design of new materials [21], acting as an efficient high-throughput material characterization platform. Thus it is a key step in material discovery.

The resonant-frequency modes (R-PFM) of PFM capture the signal across a band of frequencies. They are particularly useful because they minimize both direct (topographic) and indirect (cantilever) crosstalk with the measured ferroelectric and piezoelectric properties [11]. To do so, we typically fit the measured signal from microscopy, denoted by $f(x)$, a function of probing frequency (denoted

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

by x) with a simple harmonic oscillator (SHO) model [18]. This converts detected signals into tip-surface interactions and material properties. The analysis is often based on this model for frequency, amplitude, phase, and quality factor responses [17]. In this work, we utilize the model described in Eqn. 1, which produces model parameters from the complex data directly:

$$f(x) = \frac{Ae^{i\theta}\omega^2}{x^2 + \frac{i\omega x}{Q} - \omega^2}, \tag{1}$$

where A refers to amplitude, θ refers to phase, ω refers to the contact resonance frequency, and Q refers to the quality factor. The four parameters correspond to surface displacement magnitude, polarization orientation at the material surface, viscoelastic, and energy dissipation properties of the material, respectively.

Although PFM is effective in characterizing ferroelectric materials and interpreting their physical properties, it is not without challenges. Fitting data points with low signal-to-noise (SNR) ratio is often challenging, leading to poor model parameter estimates. Furthermore, these low SNR data points usually correspond to data points where the material's inherent polarization is expected to vanish, for example at the coercive field during polarization switching or in proximity to domain walls. In other words, at data points that are crucial for us to understand the material properties, PFM measurements are "unreliable".

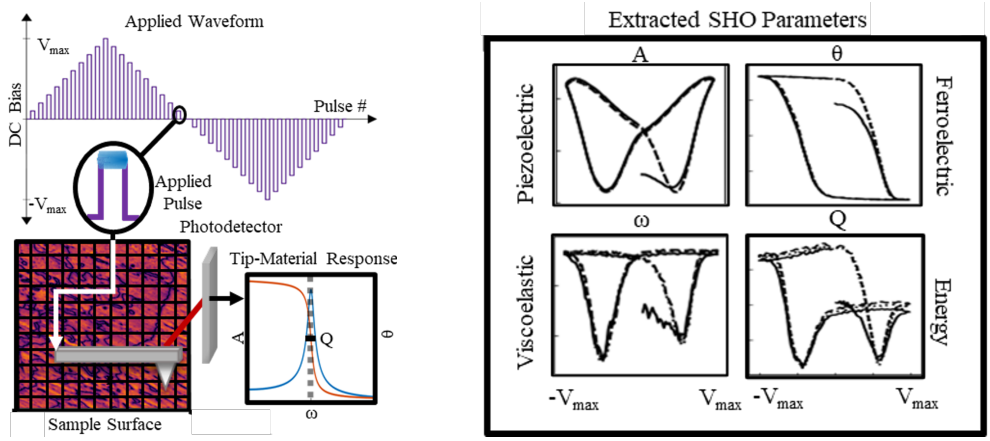
There have been a few related works approaching the noise issue. The authors of [6] use the k-means algorithm to find local outliers and anomalies in switching spectroscopy PFM data, identifying potential systematic issues. Principal component analysis (PCA) is applied to decompose and interpret SPM data [10], as the principal directions and weights carry physical meanings. The work closest to ours, [20], directly attempts to identify and eliminate the poor SHO model fits in PFM data due to significant noise, where PCA is applied to remove the noise. However, it does not address recovering information from low-SNR data points. Simply denoising the data cannot recover the true signals accurately, and the problem with low-SNR data remains. Additionally, the pattern of the location of the low-SNR cases is not considered either.

To tackle the challenges brought by low SNR data points in PFM analysis, we propose a new information recovery approach leveraging subspace matrix completion methods. By interpolating the estimated dominant subspaces across data points, we can reconstruct these low-SNR observations.

2 PFM Problem Setting

Information related to functional properties is extracted from multiple levels of R-PFM data. For each data point, we obtain the tip-material response to an applied voltage pulse across the probed frequency band in the waveform as shown in Fig. 1a. At each grid location, a series of incrementally changing voltage pulses from the waveform is applied to the sample surface. The electromechanical strain recorded as the result of each voltage pulse captures the local hysteresis response, from which SHO parameters can be extracted at this location as shown in Fig. 1b. The grid usually takes several microns in each dimension of the sample surface. The data we use is collected from the average response across three switching cycles (an applied waveform of three triangular periods and 475 individual applied pulses) over a 50 x 50 grid. Here, we focus on the switching response collected from a *single* grid point of a solid solution of lead magnesium niobate and lead titanate relaxor-ferroelectric single crystal. The response was excited at each pulse with an AC bias of 2V, across a frequency band of 50kHz in width, centered around 250kHz using an Olympus AC240TM-R3 (k = 2N/m) cantilever.

We utilize the SHO model to fit the complex frequency response within the range of probing frequency (240-260kHz) at this grid point via non-linear least squares regression. For the complex response data at each frequency, we subtract a linear background from the complex frequency response to remove the instrumental noise [15]. As Eqn. (1) fits the complex data directly, we split them into real and imaginary parts for analysis. We obtain the SHO model fits for each impulse bias, some of which are plotted in Fig. 2. It is observed that in most cases, the SHO model can capture the trend in the frequency response. However, there are some cases where the SNR is low such that the SHO model can no longer produce a fit. This causes a failure in the fitting of model parameters. Moreover, these cases are usually when polarization switching takes place, which bears significance in material properties. Therefore, the loss of information here seriously undermines the efficacy of PFM. To improve PFM data analysis, we need to find a way to recover the missing information for these low SNR cases.



(a) PFM polarization switching measurements on a grid with applied voltage waveform.

(b) Functional SHO parameters extracted from data on a single grid point (ideal scenario).

Figure 1: PFM setup with applied waveform and the extracted parameters using the SHO model.

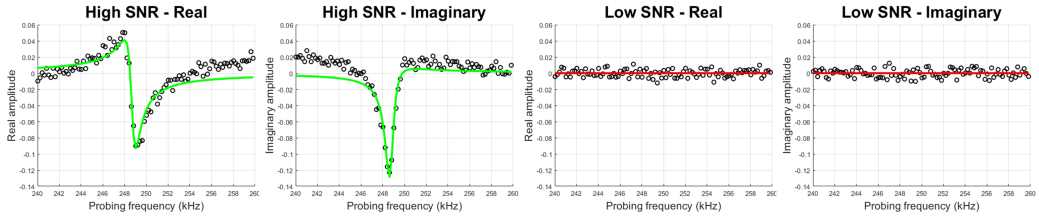


Figure 2: SHO fits on amplitude response. The two plots on the left are fits for a high SNR case, with real and imaginary component magnitudes. The two plots on the right are fits for a low SNR case.

3 Information Recovery

In this section, we will detail the information recovery approach to tackle the low SNR cases in PFM data. We will split the procedure into three major steps: (i) identify missing entries in the data matrix, (ii) estimate the subspace via BayeSMG, and (iii) recover missing information with subspace weight interpolation.

(i) Identify: For both the real and imaginary components of the frequency response, we can concatenate the series of real and imaginary response across pulses together into a data matrix. The matrix contains m_1 rows corresponding to the number of pulses in the waveform and m_2 columns corresponding to the number of probed frequencies. Since SHO fits fail at the response of some pulses as found in Section 2 due to the noise, we identify these low-SNR data points based on the error associated with Q values fit by the SHO model. Poor fits to the SHO have previously been detected using Q values [20]. We label the response from these identified pulses as missing entries in the data matrix.

(ii) Estimate: Denote this data matrix with noisy measurements by $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$. Denote the index set of the observed entries (where observations are determined to be reliable) by $\Omega \subseteq [m_1] \times [m_2]$. Denote the underlying accurate SHO fits for all pulses by the matrix $\mathbf{X}$. Since the work by [20] has identified that there are only a few dominant eigenvectors in the PFM data, we take advantage of it to make the assumption that when stacked together, the true matrix $\mathbf{X}$ is of low rank. We assume it to be of rank $R := \text{rank}(\mathbf{X}) \ll \min(m_1, m_2)$. Then the noisy observations become:

$$\mathbf{Y}_{i,j} = \mathbf{X}_{i,j} + \epsilon_{i,j}, \quad (i,j) \in \Omega. \tag{2}$$

That is, the observation at index (i,j) is denoted $Y_{i,j}$, which is corrupted by a noise term $\epsilon_{i,j}$. We assume the noise term $\epsilon_{i,j} \sim \mathcal{N}(0, \eta^2)$, which are independent and follow a zero-mean Gaussian distribution with variance η^2. Our aim is then to estimate the row space of the true matrix $\mathbf{X}$ given the partial and noisy observations $\mathbf{Y}_\Omega$, leveraging the fact that the true matrix is of low rank.

To recover missing entries in a matrix, completion is a natural option. There have been numerous works in this area [2, 1, 3], but many look at estimating missing entries only. In our problem, each row is the frequency response from a voltage pulse in our waveform which bears a periodic pattern. This makes it more appealing for us to estimate the subspace of the matrix $\mathbf{X}$ from partial and noisy observations, as the missing pattern of the entries is not uniform. Additionally, polarization switching is expected to be accompanied by a change in the SHO parameters. Therefore, it is pertinent to utilize a method where subspaces are explicitly estimated. The BayeSMG model fits our objective well [19]. It utilizes the singular matrix-variate Gaussian (SMG) distribution to model the matrix [8], where a random matrix $\mathbf{X} \stackrel{d}{=} \mathcal{P}_{\mathcal{U}}\mathbf{D}\mathcal{P}_{\mathcal{V}}$ for some projection matrices $\mathcal{P}_{\mathcal{U}} = \mathbf{U}\mathbf{U}^T$ and $\mathcal{P}_{\mathcal{V}} = \mathbf{V}\mathbf{V}^T$, where orthogonal subspace matrices $\mathbf{U}$ and $\mathbf{V}$ span the column space and the row space of rank R, respectively. BayeSMG then estimates the subspace matrices by conducting a Gibbs sampler on the posterior distributions of the model parameters, including the subspace matrices.

(iii) Recover: Since the frequency response by the pulse is stacked by row, we elect to interpolate on the row space $\mathbf{V}$ to recover the missing information. The data on each row can be represented as:

$$\mathbf{X}[i,:] = \sum_{k=1}^{R} \alpha_k \hat{\mathbf{V}}[k,:], \tag{3}$$

for $i = 1, 2, \ldots, m_1$, and α_k is the weight for the kth dimension. For the rows with good SHO fits, we can calculate $\{\alpha_k\}_{k=1}^{R}$ directly since the row space is already estimated. For low SNR rows where SHO fails, we use some interpolation methods for each of the linear weights α_k between rows to estimate the weights $\hat{\alpha}_{k'}$ where entries are missing, then use them to reconstruct the particular rows of data. We use either piecewise linear interpolation or cubic spline interpolation to find the weight parameters. By doing this, we can successfully recover the information which is lost to low-SNR PFM measurements. This method enables us to better fit the SHO models and further our understanding of the polarization switching of the material.

4 Results & discussions

Applying the information recovery framework to our PFM experiment, we obtain the SHO parameters across all pulses and use them to draw the hysteresis response plots. Out of the 949 pulses, we find approx. 7.2% of them are low SNR and need to be recovered. This is carried out by utilizing the SHO fitting error, denoted by σ_Q, which is directly associated with the Q value. We identify the unreliable data points using the following rule:

$$\frac{\sigma_Q}{Q} > h, \tag{4}$$

where h is the threshold set to 50% here. Most of the identified low SNR cases are clustered around polarization switching, which is where key information regarding material properties is located. We verify the data matrix constructed using the PFM measurements are indeed of low rank, containing 10 significant singular values. Therefore, we set $R = 10$ for the matrix completion process. To evaluate the performance of our proposed approach, we compare them to two other methods. One is a simple linear interpolation across each of the four SHO parameters to fill in those which fall under the low SNR category as judged by the Q factor. That means values recovered by interpolation in the A curves are unrelated to those recovered in the θ, ω & Q curves. The second method is a third-order spline interpolation across the individual SHO parameters instead. For our proposed method, we use linear and spline interpolations on the subspace weights α_k respectively. In total, there are two baseline methods and two from the proposed framework with different subspace interpolation options.

We plot the extracted SHO parameters and the calculated piezoresponse values from our BayeSMG and the two interpolation methods in Fig. 3. We focus on the polarization-switching part of the plots, which usually contains the most useful information.

From the plots, we observe our proposed recovery method predicts reasonable A values in proximity to switching bias (at low amplitude) without artificial plateaus as the comparison methods do since this plateauing behavior attributes to the original SHO parameters deemed unreliable. Similar 'notch' features have previously been associated with ferroelastic switching events, which are deemed problematic [4]. For the reconstructions from our proposed methods, there are clear minima found for both triangle curves, suggesting missing information at switching has been recovered. The corresponding piezoresponse (PR) curves also show improved and more continuous patterns at

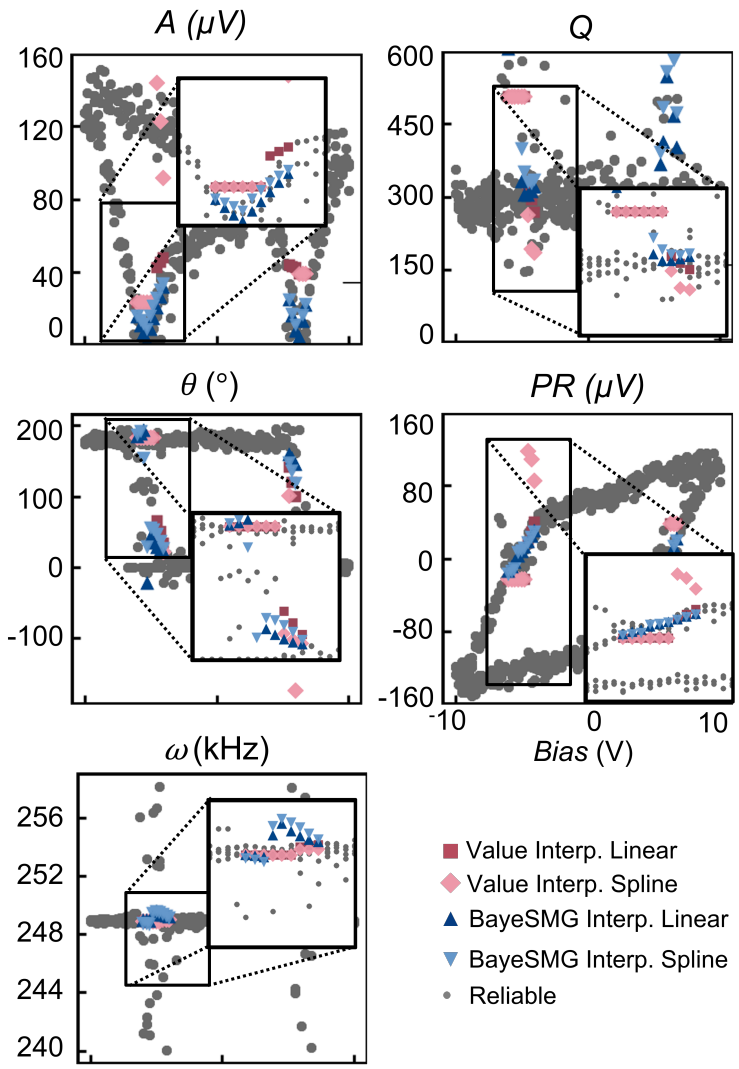


Figure 3: Extracted A, θ, ω, Q, and calculated piezoresponse (PR) values recovered by BayeSMG using linear interpolation (dark blue triangles), BayeSMG using spline interpolation (light blue triangles), linear interpolation (red squares), and third-order spline interpolation (golden diamonds) respectively. Gray markers denote reliable values of the SHO fitting.

switching. This is in contrast to the plateaus and then steep jumps in the PR curves by the linear and spline interpolation methods, which are less reasonable. For both A and PR value curves, we have observed the proposed methods are able to recover missing values better than the comparison ones. This can be because there are consecutive missing values to fit, simply carrying out interpolation on individual parameters cannot product good quality estimates. Interpolation on the subspace weights, on the other hand, is able to alleviate the issue. Additionally, the ω curve shows polarization switching is accompanied by slight hardening. Such behavior is not identified from the results of the other two methods, and may entail some unknown factors which call for further investigation.

To conclude, the proposed information recovery approach utilizing the BayeSMG model enables us to consistently detect and recover unreliable information from PFM data. It is a reliable methodology to extract SHO model parameters from low SNR points. The simultaneous extraction of SHO model parameters facilitated by BayeSMG in our approach suggests a potential route for robust and consistent extraction of piezoresponse values from low SNR data points. Currently, we have only demonstrated qualitative results and improvements found from experiments, but there is clear potential in further developing it into an efficient automated characterization tool for ferroelectric materials.

References

[1] E. Candes and B. Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012.

[2] E. J. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

[3] Y. Chen, J. Fan, C. Ma, and Y. Yan. Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences*, 116(46):22931–22937, 2019.

[4] A. R. Damodaran, S. Pandya, J. C. Agar, Y. Cao, R. K. Vasudevan, R. Xu, S. Saremi, Q. Li, J. Kim, M. R. McCarter, et al. Three-state ferroelastic switching and large electromechanical responses in $PbTiO_3$ thin films. *Advanced Materials*, 29(37):1702069, 2017.

[5] C. R. Foschini, M. A. Ramirez, S. R. Simoes, J. A. Varela, E. Longo, and A. Z. Simoes. Piezoresponse force microscopy characterization of rare-earth doped $BiFeO_3$ thin films grown by the soft chemical method. *Ceramics International*, 39(3):2185–2195, 2013.

[6] L. A. Griffin, I. Gaponenko, and N. Bassiri-Gharb. Better, faster, and less biased machine learning: Electromechanical switching in ferroelectric thin films. *Advanced Materials*, 32(38):2002425, 2020.

[7] A. Gruverman and S. V. Kalinin. Piezoresponse force microscopy and recent advances in nanoscale studies of ferroelectrics. *Journal of Materials Science*, 41(1):107–116, 2006.

[8] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. CRC Press, 1999.

[9] C. Halperin, S. Mutchnik, A. Agronin, M. Molotskii, P. Urenski, M. Salai, and G. Rosenman. Piezoelectric effect in human bones studied in nanometer scale. *Nano Letters*, 4(7):1253–1256, 2004.

[10] S. Jesse and S. V. Kalinin. Principal component and spatial correlation analysis of spectroscopic-imaging data in scanning probe microscopy. *Nanotechnology*, 20(8):085714, 2009.

[11] S. Jesse and S. V. Kalinin. Band excitation in scanning probe microscopy: sines of change. *Journal of Physics D: Applied Physics*, 44(46):464006, 2011.

[12] S. Jesse, H. N. Lee, and S. V. Kalinin. Quantitative mapping of switching behavior in piezoresponse force microscopy. *Review of Scientific Instruments*, 77(7):073702, 2006.

[13] S. V. Kalinin, B. J. Rodriguez, S. Jesse, T. Thundat, and A. Gruverman. Electromechanical imaging of biological systems with sub-10 nm resolution. *Applied Physics Letters*, 87(5):053901, 2005.

[14] Y. Liu, B. Yu, Z. Liu, D. Beck, and K. Zeng. High-speed piezoresponse force microscopy and machine learning approaches for dynamic domain growth in ferroelectric materials. *ACS Applied Materials & Interfaces*, 12(8):9944–9952, 2020.

[15] S. M. Neumayer, S. Saremi, L. W. Martin, L. Collins, A. Tselev, S. Jesse, S. V. Kalinin, and N. Balke. Piezoresponse amplitude and phase quantified for electromechanical characterization. *Journal of Applied Physics*, 128(17):171105, 2020.

[16] P. Sharma, D. Wu, S. Poddar, T. J. Reece, S. Ducharme, and A. Gruverman. Orientational imaging in polar polymers by piezoresponse force microscopy. *Journal of Applied Physics*, 110(5):052010, 2011.

[17] R. K. Vasudevan, K. P. Kelley, E. Eliseev, S. Jesse, H. Funakubo, A. Morozovska, and S. V. Kalinin. Bayesian inference in band excitation scanning probe microscopy for optimal dynamic model selection in imaging. *Journal of Applied Physics*, 128(5):054105, 2020.

[18] S. Yang, L. Mazet, M. B. Okatan, S. Jesse, G. Niu, T. Schroeder, S. Schamm-Chardon, C. Dubourdieu, A. P. Baddorf, and S. V. Kalinin. Decoupling indirect topographic cross-talk in band excitation piezoresponse force microscopy imaging and spectroscopy. *Applied Physics Letters*, 108(25):252902, 2016.

[19] H. S. Yuchi, S. Mak, and Y. Xie. Bayesian uncertainty quantification for low-rank matrix completion. *Bayesian Analysis*, 1(1):1–28, 2022.

[20] F. Zhang, K. N. Williams, D. Edwards, A. B. Naden, Y. Yao, S. M. Neumayer, A. Kumar, B. J. Rodriguez, and N. Bassiri-Gharb. Maximizing information: a machine learning approach for analysis of complex nanoscale electromechanical behavior in defect-rich pzt films. *Small Methods*, 5(12):2100552, 2021.

[21] H. Zhang, X. Chen, Y. Tang, W. Liao, F. Di, X. Mu, H. Peng, and R. Xiong. PFM (piezoresponse force microscopy)-aided design for molecular ferroelectrics. *Chemical Society Reviews*, 2021.

[22] M. Zhang, H. Thong, Y. Lu, W. Sun, J. Li, and K. Wang. (K, Na) NbO 3-based lead-free piezoelectric materials: An encounter with scanning probe microscopy. *Journal of the Korean Ceramic Society*, 54(4):261–271, 2017.