PAPER • OPEN ACCESS

Decoding the shift-invariant data: applications for band-excitation scanning probe microscopy[*]

To cite this article: Yongtao Liu *et al* 2021 *Mach. Learn.: Sci. Technol.* **2** 045028

View the article online for updates and enhancements.

MACHINE LEARNING
Science and Technology

PAPER

Decoding the shift-invariant data: applications for band-excitation scanning probe microscopy*

Yongtao Liu[1] , Rama K Vasudevan[1] , Kyle K Kelley[1], Dohyung Kim[2], Yogesh Sharma[3], Mahshid Ahmadi[2], Sergei V Kalinin[1,**] and Maxim Ziatdinov[1,4,**]

1   Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
2   Department of Materials Science and Engineering, Joint Institute for Advanced Materials, University of Tennessee, Knoxville, TN 37996, United States of America
3   Center for Integrated Nanotechnologies (CINT), Los Alamos National Laboratory, Los Alamos, NM 87545, United States of America
4   Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
**   Authors to whom any correspondence should be addressed.

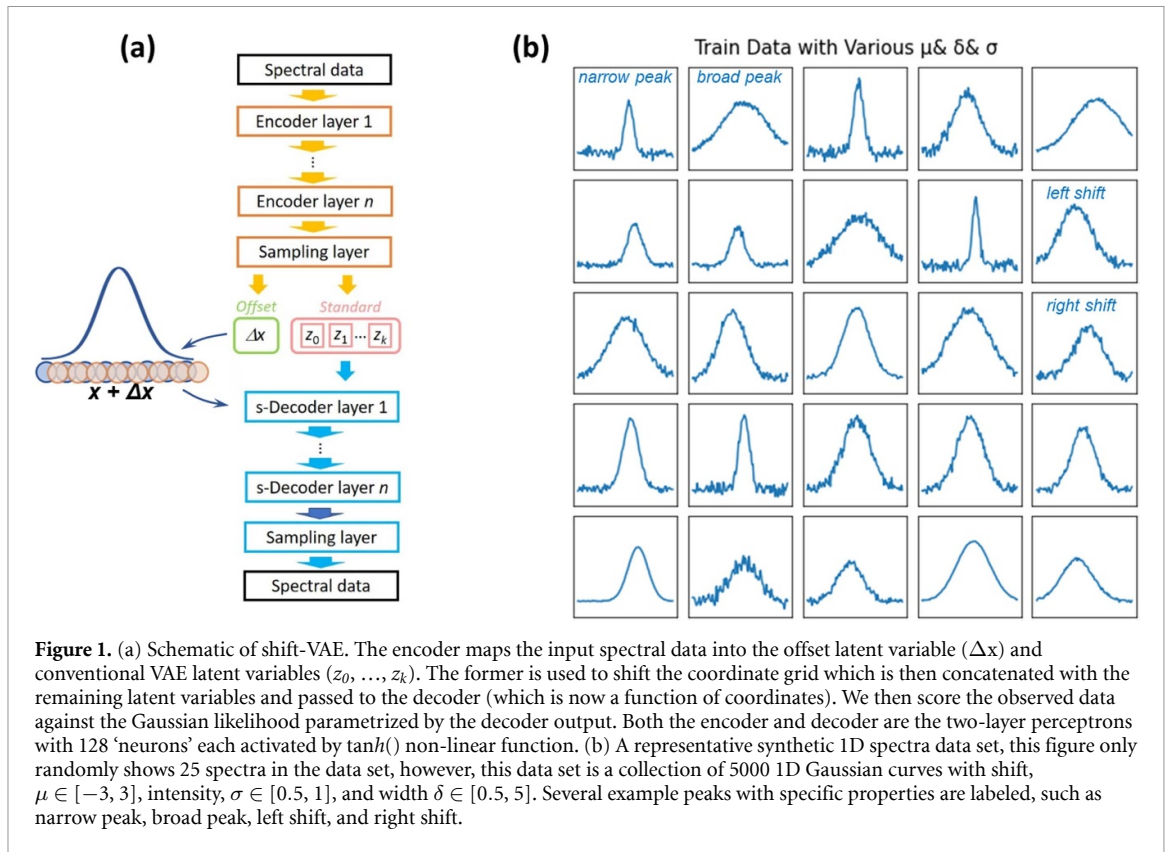**E-mail:** sergei2@ornl.gov and ziatdinovma@ornl.gov

Abstract

A shift-invariant variational autoencoder (shift-VAE) is developed as an unsupervised method for the analysis of spectral data in the presence of shifts along the parameter axis, disentangling the physically-relevant shifts from other latent variables. Using synthetic data sets, we show that the shift-VAE latent variables closely match the ground truth parameters. The shift VAE is extended towards the analysis of band-excitation piezoresponse force microscopy data, disentangling the resonance frequency shifts from the peak shape parameters in a model-free unsupervised manner. The extensions of this approach towards denoising of data and model-free dimensionality reduction in imaging and spectroscopic data are further demonstrated. This approach is universal and can also be extended to analysis of x-ray diffraction, photoluminescence, Raman spectra, and other data sets.

In materials science, a common characterization of materials properties is the presence of sharply localized peaks that can shift across the horizontal axis. Case in point, for x-ray diffraction this is due to changes in solid solution composition and strain, etc; for photoluminescence this is due to changes in bandgap and defect, etc; for band excitation piezoresponse force microscopy (BE-PFM) this is due to variation of contact resonance frequency. This behavior makes the applications of the linear multivariate methods such as principal component analysis (PCA) and related techniques [1] impractical, since decomposition gives rise to a large number of components. In BE, this was recognized since the first application of PCA in 2009[2]. This behavior persists for more complex analysis methods, including those based on the manifold learning, conventional and variational autoencoders (VAEs) [3].

In this work, we propose an approach naturally accounting for the shift of peaks, the application of this approach for BE-PFM datasets is demonstrated. Prior to discuss the analysis approach, we first introduce BE-PFM as follow. Following the invention of atomic force microscopy (AFM) in 1986[4], scanning probe microscopy (SPM) has emerged as an extremely powerful tool for probing and modifying nanoscale systems. Multiple variants of SPM methods were developed for probing electric [5, 6], mechanical [7, 8],

---

Figure 1. (a) Schematic of shift-VAE. The encoder maps the input spectral data into the offset latent variable (Δx) and conventional VAE latent variables ($z_0, \ldots, z_k$). The former is used to shift the coordinate grid which is then concatenated with the remaining latent variables and passed to the decoder (which is now a function of coordinates). We then score the observed data against the Gaussian likelihood parametrized by the decoder output. Both the encoder and decoder are the two-layer perceptrons with 128 'neurons' each activated by $\tanh()$ non-linear function. (b) A representative synthetic 1D spectra data set, this figure only randomly shows 25 spectra in the data set, however, this data set is a collection of 5000 1D Gaussian curves with shift, $\mu \in [-3, 3]$, intensity, $\sigma \in [0.5, 1]$, and width $\delta \in [0.5, 5]$. Several example peaks with specific properties are labeled, such as narrow peak, broad peak, left shift, and right shift.

electromechanical [9–11], and magnetic phenomena [12], as well as probing electronic [13, 14] and ionic transport [15]. Despite the broad gamut of measured signals and imaging conditions, the basic detection principles of these techniques remained almost invariant for two decades, and were based either on detection of static signal in e.g. contact AFM, amplitude detection via lock-in amplifier, or frequency detection via phase locked loop. In all these cases, the dynamic response of the cantilever is essentially reduced to one or several parameters linked to the functionality and visualized as spatially resolved maps.

This paradigm changed with the development of band excitation (BE) method [16], which utilized parallel detection of response in multiple frequencies. BE has enabled quantitative measurements avoiding the frequency-dependent cross-talk in a broad variety of SPM methods including piezoresponse force microscopy (PFM) [17], Kelvin probe force microscopy [18, 19], and magnetic force microscopy [20, 21]. It has also enabled new SPM modalities including electrochemical strain microscopy [22] and electrochemical force microscopy [23]. However, the central element in BE, i.e. conversion from the measured amplitude/phase—frequency dependence to local dynamic characteristics has remained unchanged since its inception [16]. Namely, the response curve is fitted by a simple harmonic oscillator (SHO) model, and the derived response amplitude, resonance frequency, and quality factor are visualized as a function of spatial coordinates or control parameters such as voltage and time in complex spectroscopies.

Recently, deep neural networks were proposed as a way to improve the initial guesses to BE analyses [24]. However, this approach still postulates the SHO functional form of response, ignoring more complex mechanical and nonlinear responses. A Bayesian inversion approach was proposed to separate linear and non-linear responses [25]. However, this approach requires a prior set of models and is extremely computationally intensive. Hence, of interest is the development of unsupervised machine learning methods capable of analysis of BE data, and extendable to other similar data sets such as those emerging in x-ray scattering, mass-spectrometry, or optical and Raman spectroscopies.

Here, we propose and implement a novel manifold learning method based on a shift-invariant VAE (figure 1(a)). This approach allows naturally accounting for the shift of peaks along the stimulus axis. We show that under certain conditions, the latent variables derived from the unsupervised learning are linear functions of the ground truth peak shift and other ground truth parameters of the peaks, corresponding to the full unsupervised disentanglement of physically-relevant variables. Here, the application for band-excitation PFM is developed. However, extensions for other methods such as x-ray scattering, photoluminescence, and Raman spectra are straightforward.
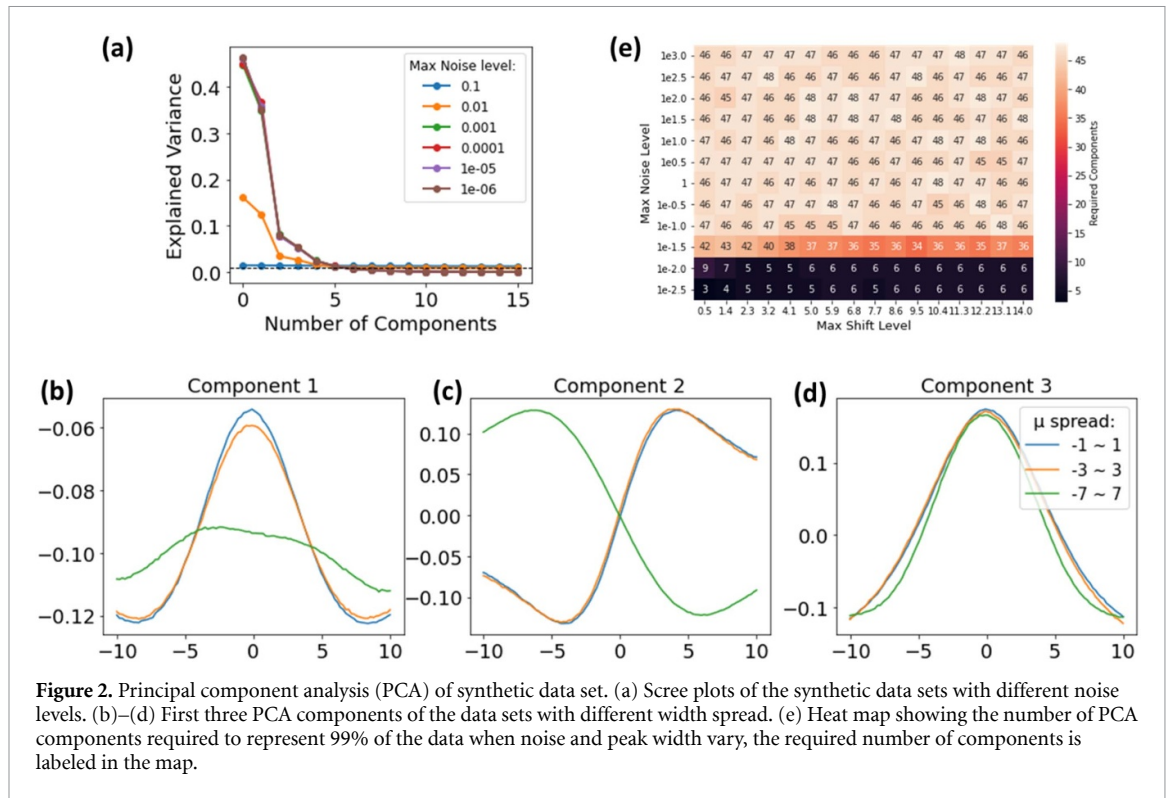
Figure 2. Principal component analysis (PCA) of synthetic data set. (a) Scree plots of the synthetic data sets with different noise levels. (b)–(d) First three PCA components of the data sets with different width spread. (e) Heat map showing the number of PCA components required to represent 99% of the data when noise and peak width vary, the required number of components is labeled in the map.

To illustrate the origins of this problem, we illustrate analysis on synthetic data set. Here, the data set is formed as a collection of $N = 5000$ Gaussian curves defined on $x \in [-10, 10]$ by function:

$$y = \sigma \times e^{\frac{-(x-\mu)^2}{2 \times \delta^2}} + \sigma_{\text{noise}}$$

with the shifts uniformly distributed on $\mu \in [\mu_{\text{min}}, \mu_{\text{max}}]$, amplitudes on $\sigma \in [\sigma_{\text{min}}, \sigma_{\text{max}}]$, and width on $\delta \in [\delta_{\text{min}}, \delta_{\text{max}}]$. In addition, the white noise of amplitude σ_0, or amplitude uniformly distributed on $\sigma_{\text{noise}} \in [0, \sigma_{\text{max - noise}}]$ interval, can be added. The data can be optionally normalized. Figure 1(b) shows a representative synthetic data set with $\mu \in [-3, 3]$, $\sigma \in [0.5, 1]$, and $\delta \in [0.5, 5]$. More example of synthetic data sets are shown in supplementary materials (figure S1 available online at stacks.iop.org/MLST/2/045028/mmedia) and the provided notebook allows adjusting the synthetic data set parameters and subsequent analytics.

The PCA scree plots of the synthetic data is shown in figure 2(a) for different noise levels $\sigma_{\text{noise}} \in [0, \sigma_{\text{max-noise}}]$, $\sigma_{\text{max-noise}} = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$, or 10^{-6}. Here, the curve illustrates characteristic shape with rapid decay of the values followed by the saturation in the white noise tail. The inflection point separating the two regimes defines a number of significant components required to fully represent the data. The characteristic PCA components are shown in figures 2(b)–(d) and illustrate the average and factor of variability in the degree of importance. Note that these components, generally speaking, do not have a well-defined physical sense. Finally, a heat map in figure 2(e) shows the number of components required to represent the data as a function of the maximum noise level and maximum shift μ_0. Here, the number of components is estimated as components needed to represent 99% of information in data set. Note that for large distributions or low noise levels the number of components can be very significant. These behaviors were observed experimentally, with often highly non-trivial distribution of spatial information between components. Shown in figure S2 is an analysis of the correlation between PCA components and Gaussian curves' properties (i.e. μ, σ, δ), where none of the components shows a correlation with curves properties, clearly illustrating limitations of the PCA or other linear analysis methods for such data sets.

To alleviate this problem, we introduce a shift-VAE technique. The key idea behind a regular VAE is that complicated real-world observations can be explained by a small number of disentangled latent variables capturing the ground truth factors of variation. The VAE consists of a decoder (generative model) that reconstructs observations from a latent code and an encoder (inference model) that approximates the true posterior probability via the amortized variational inference. Here, it is important to note that Locatello *et al* demonstrated [26] (theoretically) that unsupervised learning of disentangled latent representations is

fundamentally impossible without inductive biases. The latter typically involves modification of the loss function and/or of the architecture of the encoder and decoder neural networks [26, 27]. Here, we argue that for applications of VAE (and other deep generative models) in domain sciences, the necessary inductive bias(es) can come from prior domain knowledge such as understanding the role of measurement (instrumental factors) and the information available from theory about fundamental length scales and symmetries present in a system. This concept is illustrated by the shift-VAE technique for analyzing 1D spectral data in the presence of arbitrary shifts in peak position.

In shift-VAE (figure 1(a)), we designate one of the latent variables to absorb the information about the relative position of spectral features (we refer to it as 'offset' latent variable), whereas the rest of the latent variables capture other (than position) factors of variation. Specifically, we start by creating a 1D x-coordinate grid whose length is equal to the number of points in the spectra. Our encoder maps the input spectra into the offset latent variable and several (usually two) standard VAE latent variables. We assume that shifts in the position are normally distributed and sample our offset latent vector from a Gaussian distribution, although the usage of other distributions is also possible. The sampled values are used to shift the 1D coordinate grid by Δx. The shifted grid is then concatenated with the standard latent vector z and passed to the VAE's generator network expressed as a function of spectral coordinates to enforce consistency in geometric features between the shifted spectra. Practically, we multiply the offset latent vector by a coefficient k ($0 < k \leqslant 1$) whose value reflects our prior belief about a degree of 'disorder' in the system. The loss (negative evidence lower bound, ELBO) is computed according to
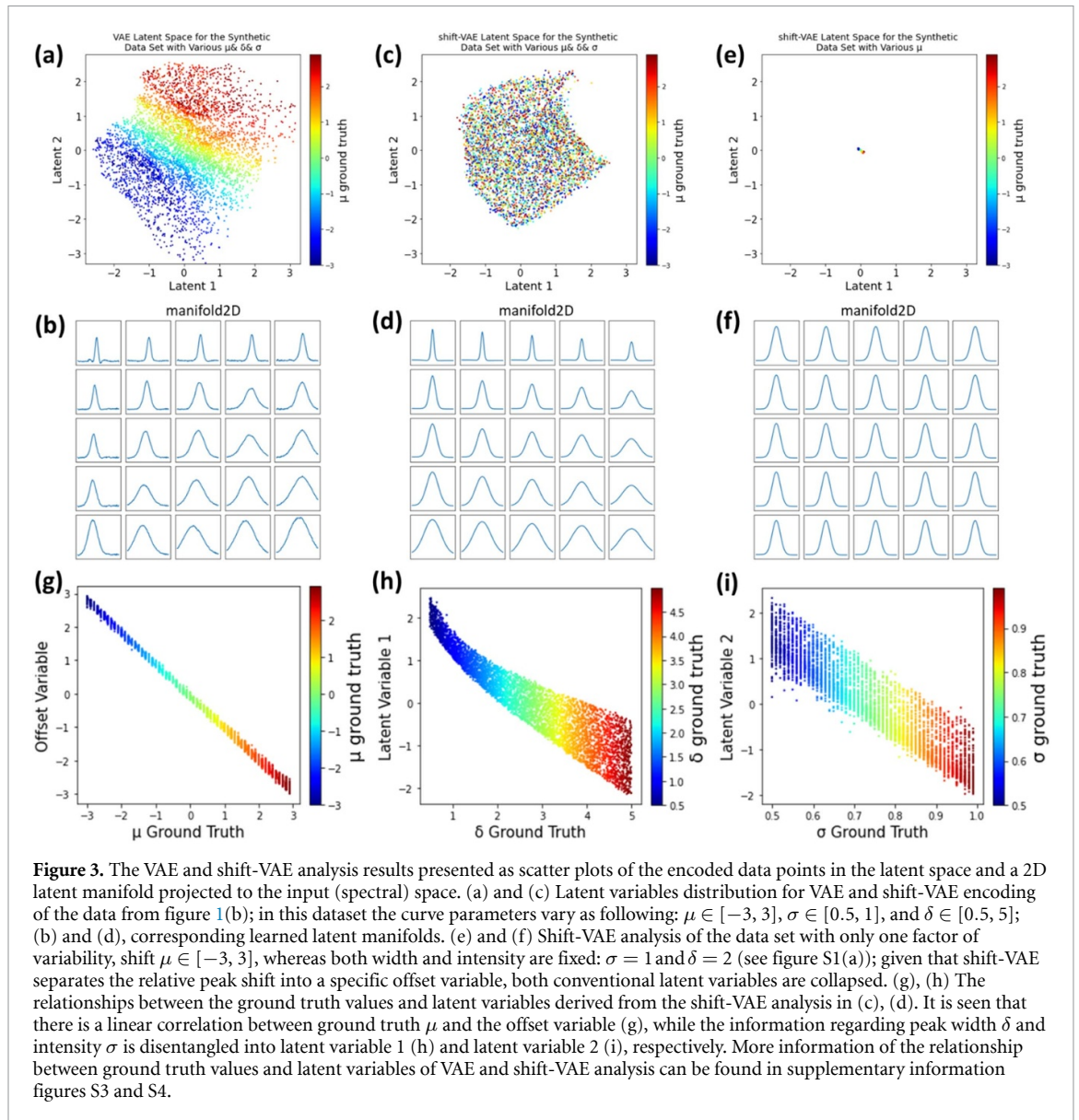
$$\mathcal{L}(y) = RE + \beta_1(t) D_{KL}(q(z|y) \| p(z)) + \beta_2(t) D_{KL}(q(\Delta x|y) \| p(\Delta x)),$$

where RE is a reconstruction error, D_{KL} is a Kullback–Leibler divergence term, and β are (optional) 'time'-dependent regularization coefficients (here 'time' is expressed through a training iteration number).

Shown in figure 3 are the results of the VAE and shift-VAE analysis. Note that, compared to shift-VAE, VAE technique does not have an offset latent variable, so the relative position of spectra is captured in conventional latent variables. Here, the curves in the original data set (figure 1(b)) are encoded into two latent variables, and resultant distribution is plotted on the 2D plane. Shown in figure 3(a) is the latent space distribution for the VAE encoding. The distribution is reminiscent of the ground-truth distributions of widths and positions. The color scale corresponds to the ground truth value of peak shift, μ. Notable is that the labels corresponding to the position are changing (mostly) from top to bottom of the image. The labels corresponding to the width are changing (mostly) from left to right (not shown). We further reconstruct the curves (figure 3(b)) from the uniform square grid of points in the latent space and observe that indeed the width changes from top to bottom and position changes from left to right. Hence, our VAE has (mostly) disentangled the representations of the data.

We also explored the correlation between the VAE's latent variables and the ground truth and found that there is indeed a certain degree of disentanglement between the two. In other words, the latent representations of the data disentangled by VAE show the factors of variability in the original synthetic data set. However, while there is a relationship between the ground truth values and latent variables, they are not equal and have large deviations (figure S4). Furthermore, the offsets are encoded in arbitrary units, which is not of practical use. Most importantly, the VAE disentanglement works well for narrow distributions, for which the peaks are fully confined within the data interval. For broad distributions, the shape of the latent manifold changes, and the correlation between the ground truth and the disentangled representation breaks down. In other words, the VAE behavior becomes controlled by the cut-off at the edges of the interval, which can be seen from the supplementary video 1 of the VAE latent space evolution when δ of training data set gradually changes from [0.5, 1] to [0.5, 50].

Next, we analyzed the shift-VAE results. Shown in figures 3(c) and (d) is the shift-VAE analysis of the same data set (see figure 1(b)). In this case, the relative shift is separated as a special offset latent variable, and the remaining variability is encoded as two conventional latent variables. In this case the offset variable solely represents peak shift (μ) and two conventional latent variables encode the other parameters. Indeed, shown in figure 3(c) is the distribution of the two latent variables encoded by shift-VAE, with the color corresponding to the ground truth shift. The latter is randomly distributed, clearly indicating that the shift variability was separated from the other latent variables. This is also illustrated for the learned latent manifolds projected to the input (spectral) space (figure 3(d)), where we only observe peak width changes, and the peak position is the same for all the curves. In figures 3(g)–(i) we explore the relationships between the encoded latent variables of the shift-VAE and the ground truth values. We observe that the offset variable shows a clear linear relationship with the ground truth μ (figure 3(g)). Here, of particular importance is that the absolute values of the offset variable and ground truth μ are equal. Moreover, the offset variable is independent of ground truth width (δ) and intensity (σ), as shown in figures S4(c) and (e), respectively. This

Figure 3. The VAE and shift-VAE analysis results presented as scatter plots of the encoded data points in the latent space and a 2D latent manifold projected to the input (spectral) space. (a) and (c) Latent variables distribution for VAE and shift-VAE encoding of the data from figure 1(b); in this dataset the curve parameters vary as following: $\mu \in [-3, 3]$, $\sigma \in [0.5, 1]$, and $\delta \in [0.5, 5]$; (b) and (d), corresponding learned latent manifolds. (e) and (f) Shift-VAE analysis of the data set with only one factor of variability, shift $\mu \in [-3, 3]$, whereas both width and intensity are fixed: $\sigma = 1$ and $\delta = 2$ (see figure S1(a)); given that shift-VAE separates the relative peak shift into a specific offset variable, both conventional latent variables are collapsed. (g), (h) The relationships between the ground truth values and latent variables derived from the shift-VAE analysis in (c), (d). It is seen that there is a linear correlation between ground truth μ and the offset variable (g), while the information regarding peak width δ and intensity σ is disentangled into latent variable 1 (h) and latent variable 2 (i), respectively. More information of the relationship between ground truth values and latent variables of VAE and shift-VAE analysis can be found in supplementary information figures S3 and S4.

indicates that the shift-VAE disentanglement works very well for encoding peak shift. In addition, variabilities of width (δ) and intensity (σ) are also mostly encoded into separated latent variables by shift-VAE, as shown in figures 3(h) and (i), respectively. In addition, shift-VAE also performs considerably well for broad peaks, where even if partial peaks are cut off at edges of the interval, as shown in figure S5.

Note that for data sets with fewer factors of variability, the latent space will be partially or even completely collapsed. For example, shown in figures 3(e) and (f) are latent space distribution and the learned latent manifolds of a shift-VAE analysis on the data set in figure S1a with only one factor of variability, shift μ, for which both latent variables are collapsed. This is because the only existing variability factor (μ) is captured by the offset variable and hence conventional latent variables are variability-free in this case. In contrast, the latent space of the VAE analysis (shown in figure S6(a)) on this data set is only partially collapsed, with the second latent variable being related to the ground truth shift. Classically in VAEs, the collapse of latent space is perceived as a problem calling for the adjustment of the 'loss' function. However, in our case, we know that the ground truth data set has only limited factor of variability. Hence, the dimensionality of our latent space hints at the true physical dimensionality of the data. This is further confirmed by analyses on the data set with two factors of variability, as illustrated in detail in supplementary figure S6.

To obtain further insight into this behavior, we explore the structure of the latent space of VAE and shift-VAE in terms of function behavior. Here we fit the curves reconstructed from the latent manifold by the known ground truth functional form (i.e. Gaussian) and show the maps of fit parameters σ, μ, δ in figures 4(a)–(c) (shift-VAE) and figure S7(a) (VAE). In figures 4(a)–(c), we observe that μ maps of shift-VAE are almost uniform and all values are very close to zero. This is because that the shift-VAE separates the peak
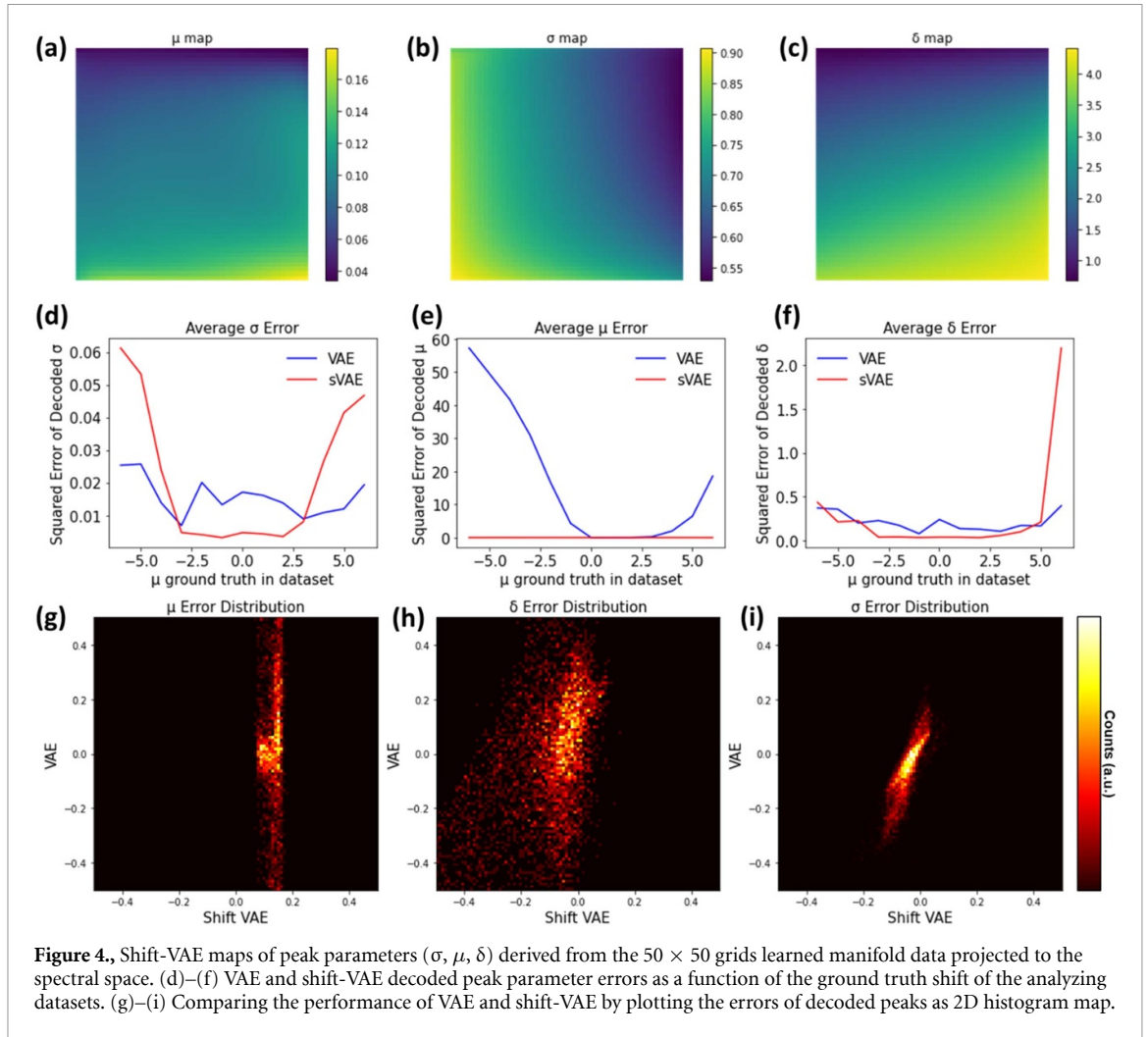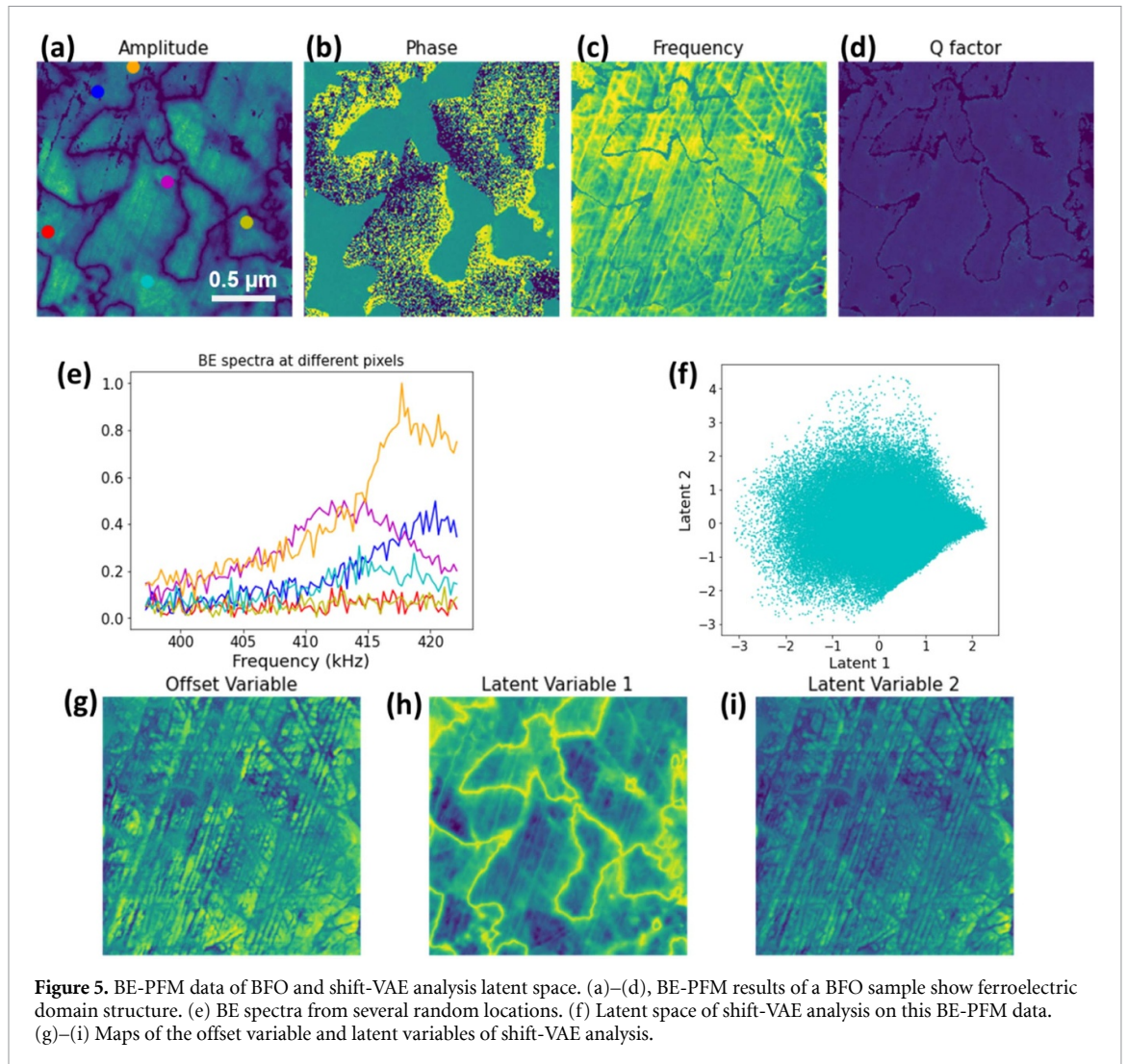
Figure 4., Shift-VAE maps of peak parameters (σ, μ, δ) derived from the 50×50 grids learned manifold data projected to the spectral space. (d)–(f) VAE and shift-VAE decoded peak parameter errors as a function of the ground truth shift of the analyzing datasets. (g)–(i) Comparing the performance of VAE and shift-VAE by plotting the errors of decoded peaks as 2D histogram map.

shift into the offset variable and hence the reconstructed curves from shift-VAE centered at μ_0, which is the center of μ for the train data set (for this training data set with $\mu \in [-3, 3]$, the μ_0 is zero). In this case, the conventional latent variables only represent other variability factors, such as intensity (σ) and width (δ). Therefore, we observe gradual changes from left to right in σ map (figure 4(b)) and from top to bottom in δ map (figure 4(c)), indicating intensity is encoded in the first latent variable (horizontal) and width is encoded in the second latent variable (vertical). The additional analyses of VAE and shift-VAE for data sets with different ground truth parameters are given in supplementary materials figures S7 and S8.

To quantify the behavior of VAE and shift-VAE, we applied the trained VAE and shift-VAE models to analyze data sets with known ground truth values. These data sets can be generated by setting desired ground truth values, which can be done through the provided Jupyter Notebook. For example, in this analysis, we used 13 data sets with specified ground truth μ. For each data set, the ground truth μ is fixed at μ_s ($\mu_s \in [-6, 6]$ and μ_s is an integer). We first encode the raw data sets with trained VAE and shift-VAE models, then reconstruct the spectra using encoded latent vectors. Afterward, we fit the reconstructed spectra by Gaussian function and compare the fit parameters with ground truth values. The difference between fit parameters and ground truth values is taken as the error. Shown in figures 4(d)–(f) are the average errors of VAE and shift-VAE analyses on each data set as a function of the ground truth μ of the data set. Clearly, shift-VAE behaves well in encoding peak shift, where the shift error of shift-VAE is always close to zero, as shown in figure 4(e). In contrast, VAE only behaves well for the data set with ground truth shift μ near zero. The behavior of VAE and shift-VAE is also compared by plotting the 2D hist maps of errors, as shown in figures 4(g)–(i). It indicates that for all errors (μ, δ, and σ) the spread is larger in the VAE direction than the shift-VAE direction, suggesting the better performance of shift-VAE.
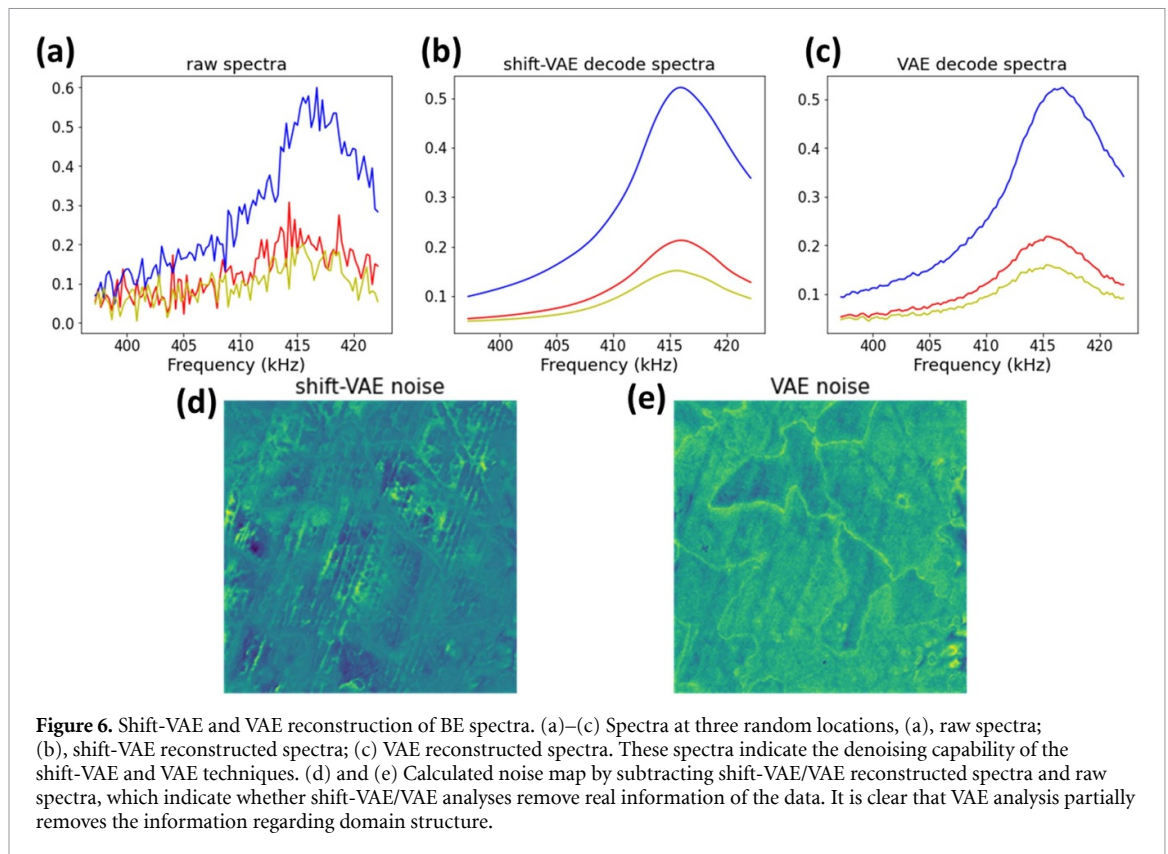
We further proceed to extend this analysis to experimental data sets. Shown in figures 5(a)–(d) is band excitation (BE) PFM data from bismuth ferrite (BFO) bulk ceramics, for which signals of 180° ferroelectric domains mainly show up in amplitude and phase images. In addition to ferroelectric domain contrast, frequency images also show some scratch stripes associated with the resonance frequency shift likely due to

Figure 5. BE-PFM data of BFO and shift-VAE analysis latent space. (a)–(d), BE-PFM results of a BFO sample show ferroelectric domain structure. (e) BE spectra from several random locations. (f) Latent space of shift-VAE analysis on this BE-PFM data. (g)–(i) Maps of the offset variable and latent variables of shift-VAE analysis.

crosstalk. In figure 5(e), we show several BE spectra from random locations (as marked in figure 5(a)), which indicate the variability of peak intensity, width, and position among these spectra.

The VAE and shift-VAE analyses are then performed on this BE-PFM data to disentangle information contained in BE spectra. Shown in figure S9 are the VAE results for latent dimensionality of two, for which the first latent variable well represents the ferroelectric domain structure, and the second latent variable represents the crosstalk scratch lines. Latent space starts to collapse when increasing the number of latent variables (i.e. latent space dimensionality), as shown in figures S10 and S11 for latent dimensionality equals of three and four. Nonetheless, the ferroelectric domain structure can always be disentangled into a non-collapsed latent variable. The results of shift-VAE analysis for latent dimensionality of two are shown in figure 5(f) and the maps of the offset and latent variables are depicted in figures 5(g)–(i). In this case, the offset variable map (figure 5(g)) mainly shows the scratch line structure. Similar to VAE, the latent space of shift-VAE starts to collapse with an increase in latent dimensionality, as shown in figures S12 and S13. This behavior is likely associated with the physical dimensionality of BE-PFM spectra, which ideally only contains information about resonance frequency (peak shift) and polarization amplitude (peak intensity). However, in a more realistic scenario, we can expect a slightly larger physical dimensionality of BE spectra due to measurement artifacts. Of importance is that for shift-VAE the offset variable always represents the crosstalk scratch lines. This matches our understanding of the physical mechanism of BE-PFM, for which crosstalk can induce resonance frequency changes (peak shift).

The remarkable aspect of VAE and shift-VAE analysis provides potential for systematic imputation on missing and high-noise data. For example, using a traditional SHO fit for data with the high noise leads to large uncertainty in resonance frequency position, necessitating the development of ad hoc criteria to mask the images based on amplitude or fit quality. Comparatively, VAE and shift-VAE approach allow to denoise the low-intensity curves using the information contained in high-intensity ones. This behavior is shown in

Figure 6. Shift-VAE and VAE reconstruction of BE spectra. (a)–(c) Spectra at three random locations, (a), raw spectra; (b), shift-VAE reconstructed spectra; (c) VAE reconstructed spectra. These spectra indicate the denoising capability of the shift-VAE and VAE techniques. (d) and (e) Calculated noise map by subtracting shift-VAE/VAE reconstructed spectra and raw spectra, which indicate whether shift-VAE/VAE analyses remove real information of the data. It is clear that VAE analysis partially removes the information regarding domain structure.

figures 6(a)–(c) of raw spectra, shift-VAE and VAE reconstructed spectra at three random locations. Clearly, both shift-VAE and VAE denoise the spectra, but shift-VAE shows a better performance. Even if for the spectra with low peak intensity (e.g. the yellow spectra), shift-VAE and VAE can reconstruct the spectra and reduce noise. In a denoising process, an important aspect we need to consider is whether any real information is removed during it. To evaluate this, we calculate the noise by subtracting reconstructed spectra and raw spectra, then show the noise as maps. Figure 6(d) and (e) show the noise maps calculated based on shift-VAE and VAE reconstructed spectra, respectively. Note that the VAE noise map shows ferroelectric domain structure, suggesting the denoising process of VAE removes a portion of real response. However, this does not occur in shift-VAE, again indicating the better performance of shift-VAE.

Finally, we extend this approach to analysis of the spectroscopic imaging in PFM, and by extension, more complex spectroscopies. As a model system, a lead titanate (PTO) sample is used to collect spectroscopic BE-PFM result. A bipolar triangular bias waveform is applied to switch the sample; basic PFM maps and switch loops are shown in figures S14 and S15. We note that this PFM data was used in our earlier work [28]; here we just use this data as a model system for shift-VAE analysis. When analyzing this data by shift-VAE, the shift-VAE reconstructed spectra match well with the corresponding raw spectra (figure S16), indicating the good performance of shift-VAE analysis on this data. Figures 7(a)–(c) shows the shift-VAE offset variable and conventional latent variables as 2D maps. We observe that domain structure is mainly captured by conventional VAE latent variables and the offset variable map only shows overall background information. Since this is spectroscopic data with switching information, further analysis is performed to look at the offset variable and latent variables as a function of the applied bias. In this analysis, the offset variable and latent variables are respectively averaged over the map at each bias step. Figures 7(d)–(f) shows the loops as a function of the applied bias. We observe that the offset variable shows a continuous drift over the switching process (figure 7(d)). This drift is likely related to the resonance frequency shift induced by the material change or surface charge accumulation during the application of bias. Then, latent variables show inverse butterfly loops, which are similar to the amplitude butterfly loop for switching of ferroelectric materials. This analysis indicates that shift-VAE disentangles the information related to material failure or electrostatic response from the information regarding the ferroelectric response, further strengthening the powerfulness of shift-VAE in learning spectra data.

To summarize, we introduced a shift-invariant VAE (shift-VAE) for analyzing 1D spectra data in a model-free unsupervised manner, which enables naturally accounting for the properties of spectral data. Using synthetic Gaussian peak data sets, we show that the shift-VAE latent variables derived from the
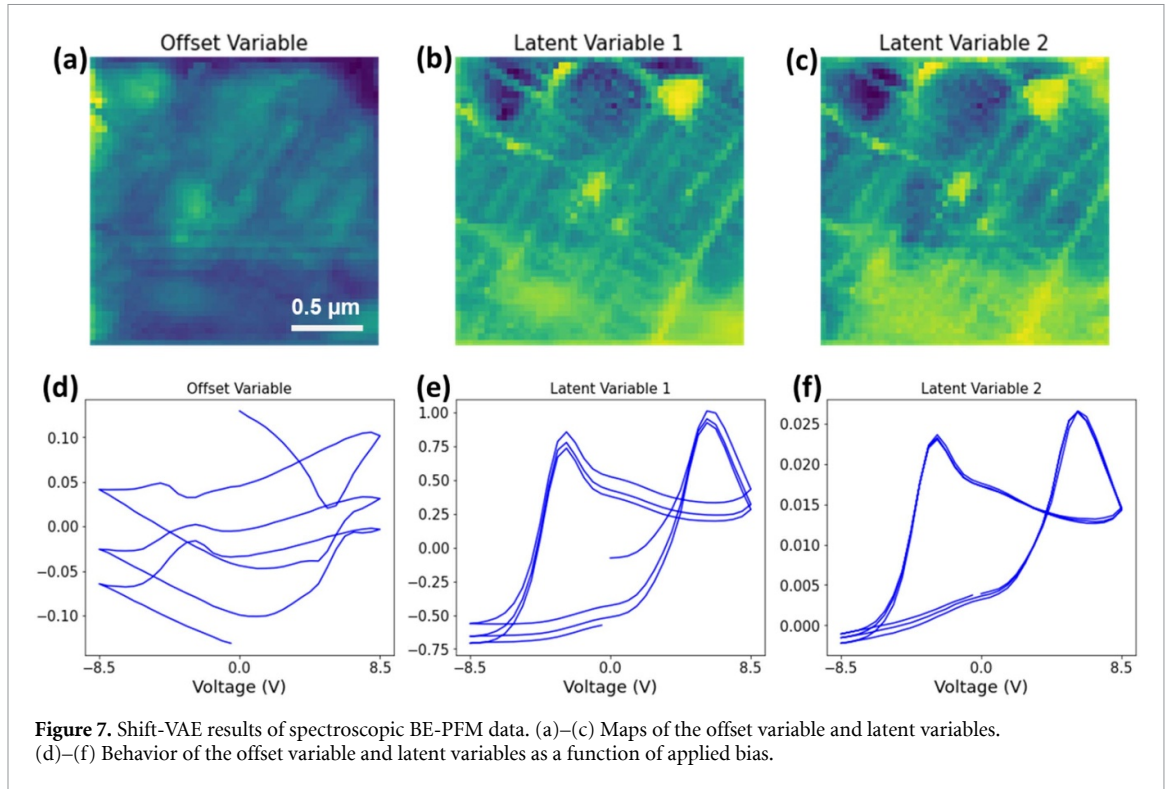
Figure 7. Shift-VAE results of spectroscopic BE-PFM data. (a)–(c) Maps of the offset variable and latent variables. (d)–(f) Behavior of the offset variable and latent variables as a function of applied bias.

unsupervised learning are linear functions of the ground truth parameters, disentangling physically relevant variables. The application of shift-VAE is illustrated for BE-PFM data. From two BE-PFM data sets, we consistently observed that shift-VAE successfully disentangles ferroelectric polarization and crosstalk-induced resonance frequency peak shift from BE curves. In addition, shift-VAE also shows the strength in denoising raw curves without sacrificing real response, allowing the recognition of real response (even if) for low signal-to-noise ratio data. Overall, shift-VAE proves to be a powerful technique for learning spectra, which should have very broad applications in many fields.

1. Materials and methods

1.1. Shift-VAE model

Overall, the generative process in shift-VAE is defined as

$$p(z) = \mathcal{N}(z|0, I); \ \ p(\Delta x) = \mathcal{N}(\Delta x|0, s^2_{\Delta x})$$
$$p_\theta(y|z, \Delta x) = \mathcal{N}(y|f(z, \Delta x)),$$

where $p(z)$ is a standard normal prior for the latent variable associated with the spectral structure (width, intensity, etc.), $p(\Delta x)$ is a normal prior for the shift latent variable with $s_{\Delta x}$ set by a user (here simply chosen to be I), and $\mathcal{N}(y|f(z, \Delta x))$ is a parametrized Gaussian likelihood function where f is a 'decoder' neural network and the variance (not shown) is fixed. Here it is assumed that the 'input layer' of the decoder shifts the coordinate grid by Δx and then concatenates it with the latent vector formed by z latent variables. The inference model outputs the approximate parameters of the posterior distribution over latent variables according to

$$q_\phi(z, \Delta x|y) = \mathcal{N}(z, \Delta x|\mu_\phi(y), \sigma^2_\phi(y)),$$

where $\mu_\phi(y)$ and $\sigma^2_\phi(y)$ correspond to the multi-head 'encoder' neural network. Hence, the shift-VAE can be formulated as an optimization problem where we learn the parameters of the encoder and decoder neural networks. The loss objective can be then written as

$$\mathcal{L}(y) = \mathrm{RE} + \beta_1(t) D_{\mathrm{KL}}(q(z|y)||p(z)) + \beta_2(t) D_{\mathrm{KL}}(q(\Delta x|y)||p(\Delta x)),$$

where RE is a reconstruction error (for the Gaussian likelihood, it is a mean-squared error), D_{KL} is a Kullback–Leibler divergence term, and β are optional 'time'-dependent regularization coefficients (set to 1 in

this work). The parameters (weights) of the encoder and decoder neural networks are trained using a mini-batch stochastic gradient descent with Adam optimizer [29]. Once trained, the encoder is used to produce the scatter plots of the encoded data in the latent space whereas the decoder is used to visualize the learned latent manifold by projecting it to the spectral space.

The shift-VAE was implemented using a home-built pyroVED package (https://github.com/ziatdinovmax/pyroVED). The details of shift-VAE and VAE analysis are available from Jupyter notebook at https://git.io/JOgFB.

1.2. Sample preparation and measurements

$BiFeO_3$ (BFO) ceramic was synthesized by a conventional solid-state reaction method using Bi_2O_3 and Fe_2O_3 power precursor (Alfa Aesar > 99.99%). The powder precursor is weighed in stoichiometric amounts. The powder mixtures were calcined at 760 °C for 1.5 h in air. After calcination, the powders were pressed using a cold isostatic method and then sintered at 780 °C for 1 h in air to form polycrystalline ceramics. The surface of the synthesized BFO bulk ceramic was polished for BE-PFM measurements. Pt-coated tips (Budget Sensors, ElectriMulti75-G; nominal resonance frequency, 75 kHz, 3 N m^{-1}) were used to apply voltages to the AFM tip. The used drive amplitude was 2.5 V (AC bias) for BE-PFM excitation. The frequency range for BE-PFM was centered 410 kHz. The measurements were taken over a grid of 256×256 pixels on the sample surface.

The PTO film was grown by chemical vapor deposition on a $SrRuO_3$ bottom electrode on a $KTaO_3$ substrate [30]. The BE-PFM imaging and spectroscopy were performed on a commercial microscope (Cypher, Asylum/Oxford Instruments) at room temperature with a Pt/Ir-coated tip (1 N m^{-1}). In-house developed LabVIEW code was used to acquire the band-excitation piezoforce spectroscopy data using National Instruments hardware. DC voltage ramping from -12.0 to $+12.0$ V was applied to the tip to measure the piezoresponse using band-excitation approach with a 1 V AC signal [28].

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://git.io/JOgFB.

Conflict of interest

The authors declare no conflict of interest.

Authors Contribution

S V K. conceived the project and M Z realized (shift-) VAE in Pyro probabilistic programming language. YL performed analyses. K K, D K, and R K V performed BE-PFM measurements. Y S synthesized BFO samples. S V K, M Z, and Y L wrote the manuscript. All authors contributed to discussions and the final manuscript.

ORCID iDs

Yongtao Liu ⬚ https://orcid.org/0000-0003-0152-1783
Rama K Vasudevan ⬚ https://orcid.org/0000-0003-4692-8579
Maxim Ziatdinov ⬚ https://orcid.org/0000-0003-2570-4592

References

[1] Kannan R, Ievlev A V, Laanait N, Ziatdinov M A, Vasudevan R K, Jesse S and Kalinin S V 2018 Deep data analysis via physically constrained linear unmixing: universal framework, domain examples, and a community-wide platform *Adv. Struct. Chem. Image* **4** 20

[2] Jesse S and Kalinin S V 2009 Principal component and spatial correlation analysis of spectroscopic-imaging data in scanning probe microscopy *Nanotechnology* **20** 085714

[3] Kingma D P and Welling M **2013** Auto-encoding variational bayes (arXiv:1312.6114)

[4] Binnig G, Quate C F and Gerber C 1986 Atomic force microscope *Phys. Rev. Lett.* **56** 930–3

[5] Nonnenmacher M, Oboyle M P and Wickramasinghe H K 1991 Kelvin probe force microscopy *Appl. Phys. Lett.* **58** 2921–3

[6] Cunningham S, Larkin I A and Davis J H 1998 Noncontact scanning probe microscope potentiometry of surface charge patches: origin and interpretation of time-dependent signals *Appl. Phys. Lett.* **73** 123–5

[7] Rabe U and Arnold W 1994 Acoustic microscopy by atomic-force microscopy *Appl. Phys. Lett.* **64** 1493–5

[8] Turner J A, Hirsekorn S, Rabe U and Arnold W 1997 High-frequency response of atomic-force microscope cantilevers *J. Appl. Phys.* **82** 966–79

[9] Kolosov O, Gruverman A, Hatano J, Takahashi K and Tokumoto H 1995 Nanoscale visualization and control of ferroelectric domains by atomic-force microscopy *Phys. Rev. Lett.* **74** 4309–12

[10] Gruverman A, Auciello O, Ramesh R and Tokumoto H 1997 Scanning force microscopy of domain structure in ferroelectric thin films: imaging and control *Nanotechnology* **8** A38–43

[11] Balke N, Bdikin I, Kalinin S V and Kholkin A L 2009 Electromechanical imaging and spectroscopy of ferroelectric and piezoelectric materials: state of the art and prospects for the future *J. Am. Ceram. Soc.* **92** 1629–47

[12] Hartmann U 1999 Magnetic force microscopy *Annu. Rev. Mater. Sci.* **29** 53–87

[13] Vatel O and Tanimoto M 1995 Kelvin probe force microscopy for potential distribution measurement of semiconductor-devices *J. Appl. Phys.* **77** 2358–62

[14] Tanimoto M and Vatel O 1996 Kelvin probe force microscopy for characterization of semiconductor devices and processes *J. Vac. Sci. Technol.* B **14** 1547–51

[15] Strelcov E, Jesse S, Huang Y L, Teng Y C, Kravchenko I I, Chu Y H and Kalinin S V 2013 Space- and time-resolved mapping of ionic dynamic and electroresistive phenomena in lateral devices *ACS Nano* **7** 6806–15

[16] Jesse S, Kalinin S V, Proksch R, Baddorf A P and Rodriguez B J 2007 The band excitation method in scanning probe microscopy for rapid mapping of energy dissipation on the nanoscale *Nanotechnology* **18** 435503

[17] Jesse S, Guo S, Kumar A, Rodriguez B J, Proksch R and Kalinin S V 2010 Resolution theory, and static and frequency-dependent cross-talk in piezoresponse force microscopy *Nanotechnology* **21** 405703

[18] Guo S L, Kalinin S V and Jesse S 2012 Half-harmonic Kelvin probe force microscopy with transfer function correction *Appl. Phys. Lett.* **100** 063118

[19] Guo S L, Kalinin S V and Jesse S 2012 Open-loop band excitation Kelvin probe force microscopy *Nanotechnology* **23** 125704

[20] Jesse S and Kalinin S V 2011 Band excitation in scanning probe microscopy: sines of change *J. Phys. D: Appl. Phys.* **44** 464006

[21] Collins L, Belianinov A, Proksch R, Zuo T T, Zhang Y, Liaw P K, Kalinin S V and Jesse S 2016 G-mode magnetic force microscopy: separating magnetic and electrostatic interactions using big data analytics *Appl. Phys. Lett.* **108** 19

[22] Kumar A, Ciucci F, Morozovska A N, Kalinin S V and Jesse S 2011 Measuring oxygen reduction/evolution reactions on the nanoscale *Nat. Chem.* **3** 707–13

[23] Collins L *et al* 2014 Probing charge screening dynamics and electrochemical processes at the solid-liquid interface with electrochemical force microscopy *Nat. Commun.* **5** 3871

[24] Vasudevan R K, Ziatdinov M, Vlcek L and Kalinin S V 2021 Off-the-shelf deep learning is not enough, and requires parsimony, Bayesianity, and causality *npj Comput. Mater.* **7** 1–6

[25] Vasudevan R K, Kelley K P, Eliseev E, Jesse S, Funakubo H, Morozovska A and Kalinin S V 2020 Bayesian inference in band excitation scanning probe microscopy for optimal dynamic model selection in imaging *J. Appl. Phys.* **128** 10

[26] Locatello F, Bauer S, Lucic M, Raetsch G, Gelly S, Schölkopf B and Bachem O 2019 *Challenging common assumptions in the unsupervised learning of disentangled representations Int. Conf. Machine Learning*, (PMLR) pp 4114–24

[27] Burgess C P, Higgins I, Pal A, Matthey L, Watters N, Desjardins G and Lerchner A **2018** Understanding disentangling in beta-VAE (arXiv:.03599)

[28] Kalinin S V, Kelley K, Vasudevan R K and Ziatdinov M 2021 Toward decoding the relationship between domain structure and functionality in ferroelectrics via hidden latent variables *ACS Appl. Mater. Interfaces* **13** 1693–703

[29] Kingma D P and Ba J *2015* Adam: a method for stochastic optimization (arXiv:1412.6980)

[30] Morioka H, Yamada T, Tagantsev A K, Ikariyama R, Nagasaki T, Kurosawa T and Funakubo H 2015 Suppressed polar distortion with enhanced Curie temperature in in-plane 90°-domain structure of a-axis oriented PbTiO3 Film *Appl. Phys. Lett.* **106** 042905