PAPER

Bayesian optimization in continuous spaces *via* virtual process embeddings†

Mani Valleti, [ID] *[a] Rama K. Vasudevan,[b] Maxim A. Ziatdinov[bc] and Sergei V. Kalinin*[d]

Automated chemical synthesis, materials fabrication, and spectroscopic physical measurements often bring forth the challenge of process trajectory optimization, *i.e.*, discovering the time dependence of temperature, electric field, or pressure that gives rise to optimal properties. Due to the high dimensionality of the corresponding vectors, these problems are not directly amenable to Bayesian Optimization (BO). Here we propose an approach based on the combination of the generative statistical models, specifically variational autoencoders, and Bayesian optimization. Here, the set of potential trajectories is formed based on best practices in the field, domain intuition, or human expertise. The variational autoencoder is used to encode the thus generated trajectories as a latent vector, and also allows for the generation of trajectories *via* sampling from latent space. In this manner, Bayesian optimization of the process is realized in the latent space of the system, reducing the problem to a low-dimensional one. Here we apply this approach to a ferroelectric lattice model and demonstrate that this approach allows discovering the field trajectories that maximize curl in the system. The analysis of the corresponding polarization and curl distributions allows the relevant physical mechanisms to be decoded.

Introduction

Automated experiments and autonomous laboratories are rapidly becoming the leading paradigm in areas ranging from materials science and condensed matter physics to chemistry. For synthesis, concepts such as fully autonomous robotic laboratories,[1–3] microfluidic systems,[4] and integrated human-high throughput synthesis and characterization workflows[5] have been proposed. In microscopy and scattering, automated experiment for rapid optimization of imaging parameters and discovery is rapidly gaining popularity,[6,7] the development stimulated by the emergence of the control interfaces by leading manufacturers. Notably, microscopic imaging over the composition spread libraries opens further opportunities for rapid materials discovery.

Common for materials discovery and optimization and imaging problems is the need to explore non-convex parameter spaces towards desired functionality.[8–11] For materials synthesis, this is often the compositional space of selected multicomponent phase diagram,[12,13] or the processing history.[14,15] For imaging, this is the image plane of the object[16] or the parameter space of the microscope control.[17] From the optimization perspective, critical considerations are the dimensionality and completeness of the parameter space, and the properties of the target function defined over it.

For compositional space in synthesis or instrumental tuning alike, the parameter space is generally low dimensional and complete, often allowing for the use of the classical Bayesian optimization based methods.[18–21] In cases where the function defined over the parameter space is discontinuous, the classical BO methods fail. However, if the character of these behaviors is partially known, the use of the structured Gaussian processing with the mean function tailored for specific behavior allows to address this problem.[22]

However, the many problems including molecular and materials discovery,[23] active learning of structure–property relationships,[24,25] and process optimization[26] require optimization in very large dimensional spaces, *e.g.*, the chemical space of the system,[27] possible processing trajectories, or local microstructures. It is well recognized that solution for this problem is possible if the behaviors of interest form low-dimensional manifolds in these high dimensional spaces, and hence active

[a]Bredesen Center for Interdisciplinary Research, University of Tennessee, Knoxville, TN 37916, USA. E-mail: svalleti@vols.utk.edu

[b]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

[c]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

[d]Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN 37916, USA. E-mail: sergei2@utk.edu

learning problem is indelibly linked to the discovery of this manifold.

Recently, we have demonstrated the use of the deep kernel learning methods for discovery of structure–property relationships.[28–31] Here, the algorithm has the access to the full structural information in the system and seeks to explore the relationship between the structure and functionalities by sequential measurements. The deep kernel learning parts build the correlative relationship between structure (*e.g.*, image patch) and properties (spectra), whereas the use of physics-based search criteria allows to guide the discovery towards specific objects of interest.

Here, we extend this concept towards optimization in high dimensional spaces for applications such as materials processing or adaptive imaging and spectroscopy in physical measurements. We propose that from a practical viewpoint, we aim to optimize not the full set of possible trajectories, but a practically relevant subset that contains sufficient variability to approach a desired target. These can be chosen based on domain expertise, prior published data, or domain-specific intuition. One way to achieve this will be to use the trajectories parametrized *via* a certain functional form. However, this approach leads to large parameters spaces as well. Hence, we introduce the concept of deep kernel learning on the virtual spaces and demonstrate it for the lattice phase model of

ferroelectric material. However, the proposed approach is universal and can be applied to other processing problems.

Simplified workflow of the manuscript

Fig. 1 shows the schematic of the workflow discussed in the manuscript and is a representative of how the data is handled between the different components discussed in this work. Fig. 1a shows the workings of variational autoencoder (VAE) on a simplified level. The input to VAE a large collection of trajectories whose dimensions are sequentially reduced (encoding), passed through a bottleneck layer (latent space), and then reconstructed back into the original space (decoding). The bottleneck layer ensues in low dimensional representation of the input curves. Trajectories in this work are 1-dimensional functions which for the most cases, are a function of time and control the dynamics of complex system (black box). The working of VAEs on the trajectories of different forms is extensively discussed in this work. As a domain specific example for the complex system (black box) that takes a trajectory as an input, a kinetic simulation, FerroSIM has been implemented. Fig. 1b delineates the workings of FerroSIM at a peripheral level. The input to this model is a time dependent electric field trajectory that is applied externally on a ferroelectric material. The simulation is then carried out and the polarization field
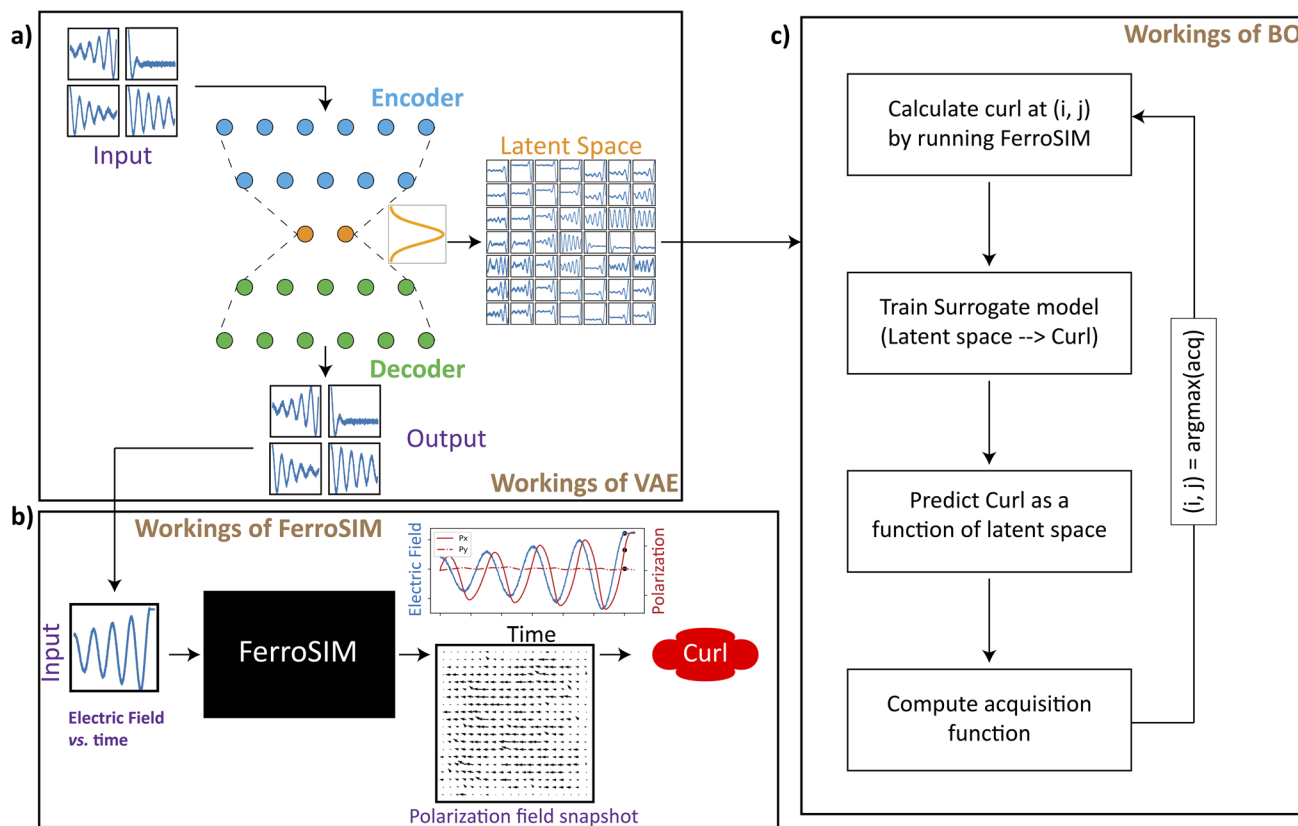

Fig. 1 Schematic diagram of the workflow discussed in this work and the dataflow between the components described in this manuscript. (a) The representational flowchart of how variational autoencoder (VAE) produces smoothly varying latent space from a given set of input curves. (b) The illustration of the workings of FerroSIM, a domain specific model discussed in this work. (c) Flowchart of Bayesian Optimization (BO) algorithm and how it is used to optimize a certain scalar quantity in a low dimensional smoothly varying latent space.

over the lattice at different points in time is the output of the simulation. In Fig. 1b, at the output side, the snapshot of the polarization map of the material at some random time along with the total polarization as a function of time are shown. The curl of the polarization map at the end of the simulation is the quantity we wish to optimize in this work. The input family of trajectories is the sinusoids whose amplitudes are modulated by exponential function. To optimize in the space of these trajectories, initially, we train a VAE whose latent space at the end of the training can be sampled and decoded to produce smoothly varying family of curves that act as inputs to FerroSIM. The latent space optimization is then performed using Bayesian Optimization (BO) algorithm that determines the maximum of curl in the latent space while bypassing the need to run Ferro-SIM simulations at every point in the latent space (Fig. 1c). A workflow to optimize in the space of trajectories is discussed in the domain-specific case of FerroSIM in this work. However, the same workflow can be applied to optimize any physical process that takes a 1-dimensional function as input and outputs a scalar quantity.

VAEs for reducing dimensionality of arbitrary functions

To illustrate the concept of dimensionality reduction, we first explore the salient features of low-dimensional representations of complex 1-dimensional functions. From here on, we will refer to these 1-dimensional functions as trajectories. Firstly, we explore a method based on variational autoencoder (VAE) to obtain a low dimensional representation these trajectories without considering the actual system. As an example, we consider the set of arbitrary trajectories, each represented by a linear combination of Legendre polynomials,

$$F(x) = \sum_{p=0}^{N} A_p L_p(x) \tag{1}$$

where $L_p(x)$ is the Legendre polynomial of degree p and A_p is its coefficient in the linear expansion of the function $F(x)$. The Legendre polynomials L_p form an orthogonal basis on the interval [0,1], and possess the properties that $L_p(0) = -1/+1$ for odd/even p, and $L_p(1) = 1$. As an orthogonal system over the unit interval, they form a convenient basis to expand any practically relevant function, and the properties of such expansions are well explored.

Here, as an illustration, we generate 5000 trajectories governed by eqn (1), where coefficients A'_p are sampled from a uniform distribution on [0, 1] and $N = 14$. These randomly sampled coefficients are then scaled in the form $A_p = A'_p/p$ to reduce the effects of higher-order polynomials and make them more practical. Twenty-five randomly sampled trajectories from the family generated are shown in Fig. 2a. The family of trajectories generated lie in a 15-dimensional space where p^{th} dimension corresponds to the corresponding Legendre polynomial L_p and by construction, will contain the trajectories that represent practically important processing functions, *e.g.*, temperature profile during field annealing, field evolution

during the poling of ferroelectric material, *etc.* However, optimization in such high-dimensional space is complex.

We subsequently explore the encoding of the generated functions into a low dimensional latent space *via* the variational autoencoder described in our previous works.[25,32–35] Here, the trajectories generated as described above act as the input to the VAE. The VAE then builds the smooth encoding of trajectories, where the trajectories are mapped to an n-dimensional continuous latent space. Each trajectory is then represented as an n-dimensional vector $(l_1, \ldots, l_n)$ in the latent space, where l_n is the nth component of the latent vector. The dimensionality of the latent space (n) is a user selected variable.

The dimensionality of the latent space is set to 2 for visualization purposes. The two-dimensional latent space is uniformly sampled and each sampled point in the latent space is then decoded back into the space of trajectories. This decoded latent space shown in Fig. 2b, helps in visualizing the trajectories encoded in different parts of the latent space. We will refer to such plots as the decoded latent space from now on. It can be observed from the decoded latent space that the trajectories are encoded in a continuous and smooth manner, which is the characteristic of the VAE. Each trajectory in the real space is represented as a scatter-point in the latent space and then colored using the root mean squared error (RMSE) between the original function and the output of VAE. The reconstruction error here is the average RMSE between the trajectories in the input space and the encoded–decoded functions, the outputs of the VAE. This plot is shown in Fig. 2c, and it can be observed from this figure that the reconstruction error is not a function of latent space, *i.e.*, there is no specific region in the latent space where the trajectories are encoded poorly by VAE. We will refer to such plots where trajectories are plotted as scatter points in the latent space as the latent distributions. Finally, to assess the effect of latent space dimensionality on the encoding, we plot the reconstruction error *vs.* the number of latent dimensions in Fig. 2d. Initially, the error decreases with the number of latent dimensions and increases slightly. This is because with a large number of latent dimensions, the latent space starts to encode the noise in the trajectories.

To obtain insights into the relationship between original parameters that generated the trajectories and the latent distribution, we plotted the latent distribution with colors corresponding to the parameters. Latent distributions where the trajectories are encoded into latent space, are color coded using the A_1, A_2, and A_4 are shown in Fig. 3a–c respectively. The latent space appears to have a strong dependence on A_1 (Fig. 3a) and A_2 (Fig. 3b) while it does not appear to have any correlation with A_4 (Fig. 3c). This is intuitively clear, as the first two coefficients in the linear expansion of the functions have larger weights by construction. The dependence of the latent space on the coefficients of the higher order polynomials diminishes with the order and can be explored further in the provided Jupyter notebook.

However, the latent space can be sampled to form trajectories in the real space with a continuous variability of higher-order parameters. To demonstrate this property, we uniformly sampled the latent space into 10 000 points where each latent
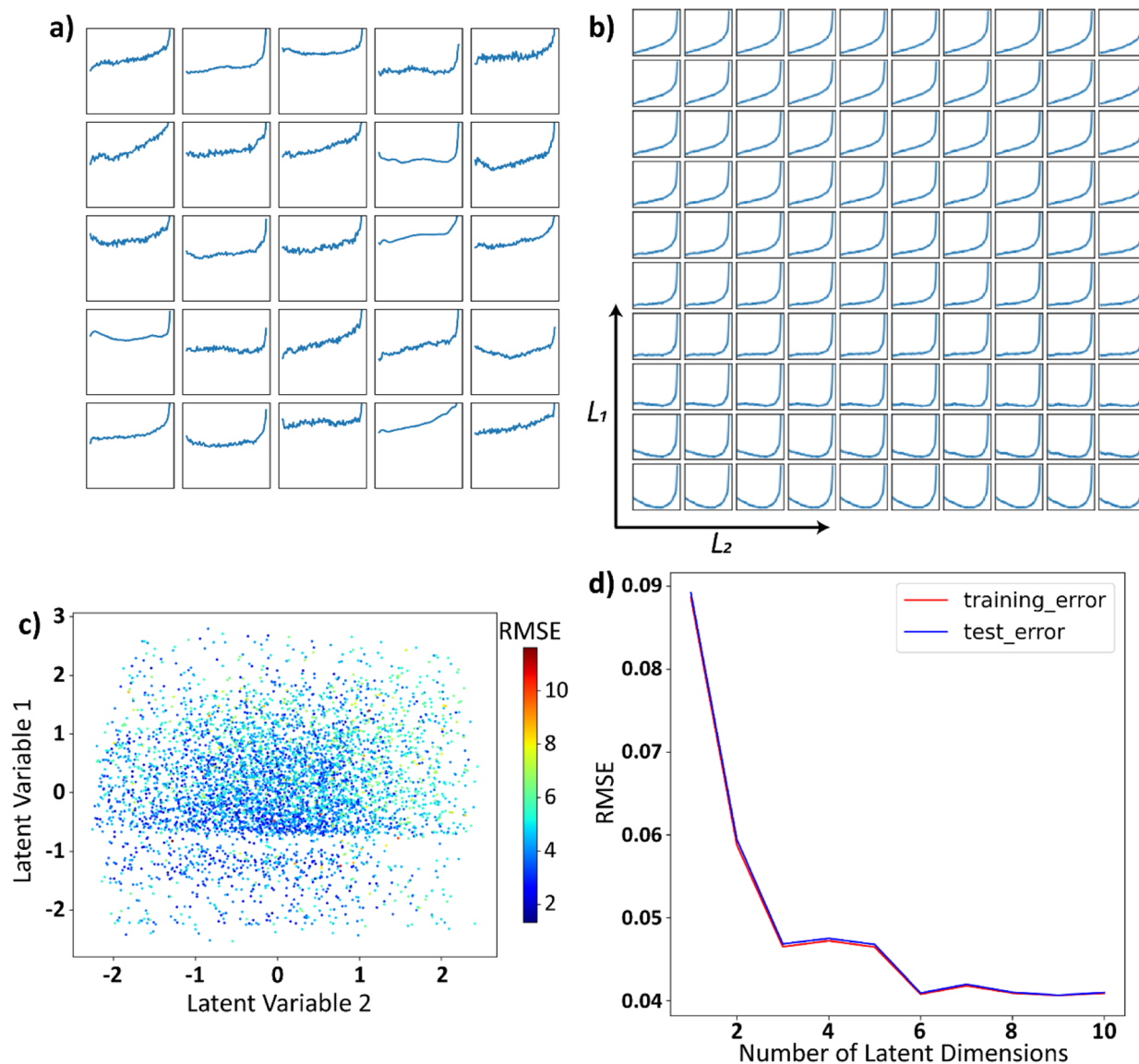
Fig. 2   (a) Twenty-five randomly selected functions generated by eqn (1), the y-axis limits and the x-axis limits are set to be $[-1, 1]$ for all the plots and are omitted from the plots for brevity. (b) The latent space is uniformly sampled between $[-1, 1]$ in each of the latent dimensions and these sampled points are decoded back into the space of functions. (c) Functions generated by eqn (1) are represented by their encodings in the latent space and are colored using RMSE, illustrating the reconstruction quality *vs.* latent encoding. (d) RMSE between the function in the input space and the functions that are decoded back to the input space after encoding as a function of number of latent dimensions.

dimension varies between $[-2, 2]$. These latent points are then decoded into the real space and are expanded to a Legendre polynomial series using the least squares method *i.e.*, (l_1, l_2) $\Rightarrow f(x)$ into Legendre series, $f(x) = \sum_{p=0}^{\infty} \tilde{A}_p L_p(x)$. With this, we plot the expansion coefficients in the latent space as shown in Fig. 3d and e. To show that the latent space has a smooth variation in both lower order and higher order polynomials, we used coefficients $\tilde{A}_1, \tilde{A}_2, \tilde{A}_3$ in Fig. 3d and $\tilde{A}_9, \tilde{A}_{10}, \tilde{A}_{11}$ in Fig. 3e. The value of these coefficients in the expanded Legendre series act as color channels for the RGB images. This would aid is noticing the variability of three coefficients per image. Note that while the resulting distributions are generally multi modal, they nonetheless show clear variability over latent space for lower

and higher order coefficients. In such a way, the process trajectories are effectively encoded in the latent space, and corresponding parameters form smooth fields over the latent space with a small number of minimum and maxima. Assuming that the relationship between the trajectory of the process and resultant materials functionality is smooth almost everywhere, this strongly suggests that the Bayesian optimization can be performed over the latent space, where the decoded trajectories are used as an input to the process, and resultant functionality defines the target function.

Prior to the optimizing the latent space with Bayesian optimization, we assess the effect of the presence of trajectories in the input dataset that are not part of the physical system on VAE. This can also be treated as a case where the input
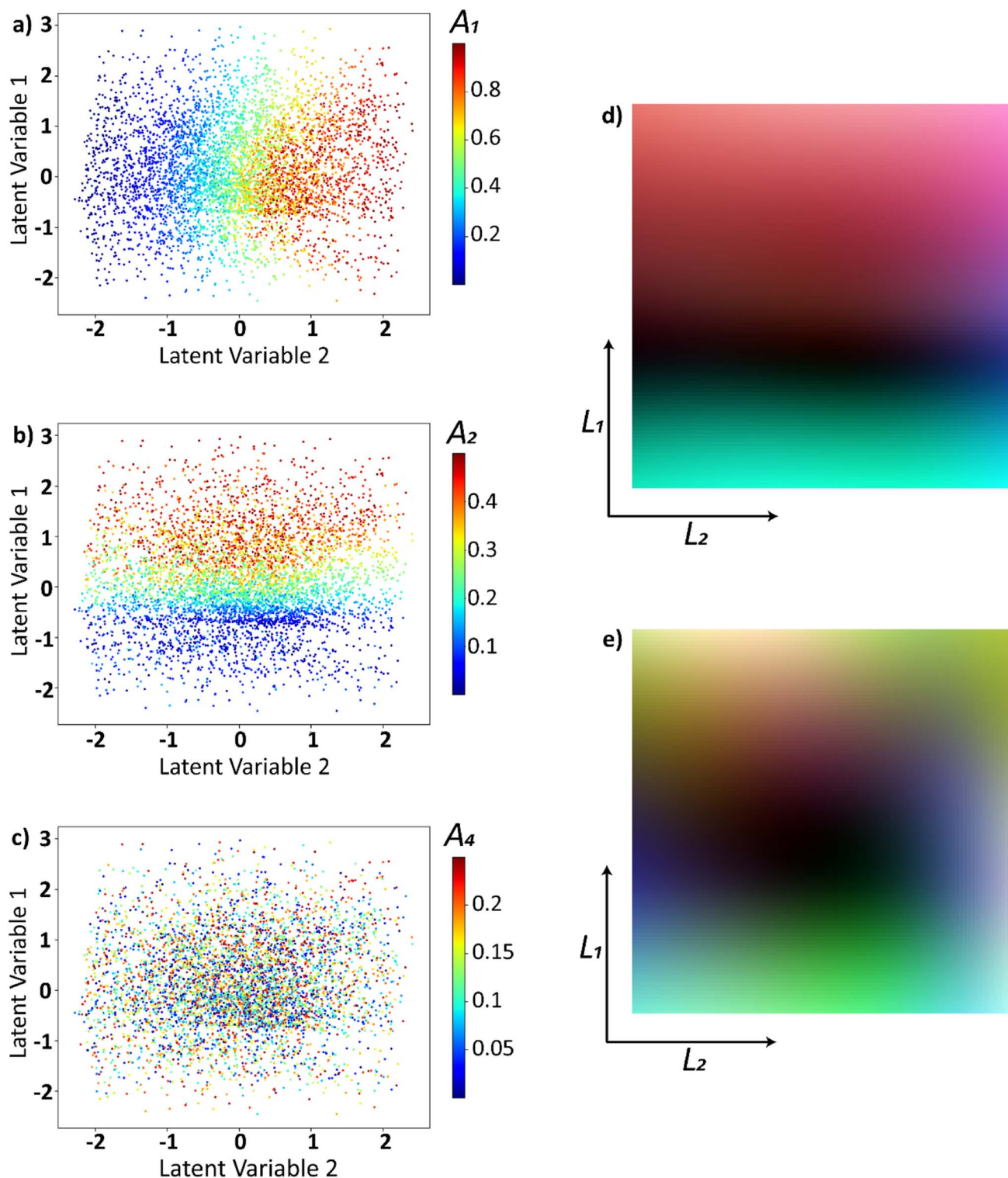
Fig. 3 Input functions represented as points in the latent space and are colored as a function of (a) A_1, (b) A_2, (c) A_4. Latent space is uniformly sampled and decoded into the space of input function which are then expanded into a Legendre series and the coefficients (d) $\tilde{A}_1$, $\tilde{A}_2$, and $\tilde{A}_3$, (e) $\tilde{A}_9$, $\tilde{A}_{10}$, and $\tilde{A}_{11}$ are plotted in the latent space as color channels in an RGB image.

trajectories are not necessarily limited to single family of curves. We generated a dataset where a fixed proportion of these trajectories are linear (which comply with the physical system), and the remaining are sinusoidal. The linear trajectories have variability in slope and y-intercept, which leads to variability in their range. For sinusoidal trajectories, only their frequency has been varied. In this section, we refer to the sinusoidal

trajectories as wrong trajectories, *i.e.*, the trajectories that do not belong to the original distribution.

We trained three different VAEs with varying proportions (1%, 10%, 30%) of wrong trajectories in the input data. The decoded latent space for each of the three different wrong trajectory proportions are shown in Fig. 4a (1%), Fig. 4b (10%), and Fig. 4c (30%). It can be observed from the results that the latent space is affected by the presence of wrong trajectories irrespective of the proportion of the wrong trajectories. Furthermore, all the input functions are plotted as points in the latent space where the linear trajectories are colored red, and the sinusoidal trajectories are colored green. This latent space representation of linear and wrong trajectories is shown in Fig. 4d (1%), Fig. 4e (10%), and Fig. 4f (30%). In all the three cases, the sinusoidal trajectories occupy a very tiny portion of the latent space. This is because the variability in the linear trajectories is much higher than the sinusoidal trajectories which are only varied in frequency. The Jupyter notebook provided in the data availability statement shows an example in which 90% of the curves in the input curves are sinusoidal and the discussion on latent space still holds true. The VAE latent space is expected to be smoothly varying and when decoded, certain areas should produce curves that have properties of both the families of curves. This ability of VAE to produce curves that are not part of the input dataset can be observed in decoded latent space in all the 3 cases (Fig. 4a–c) where the sinusoidal curves overlapped on the linear curves when moving radially

away from the center of the image. From this analysis, it can be deduced that the latent space is always affected even if there is a limited presence of wrong trajectories. The families of curves are grouped together in the latent space whose assigned area in the latent space is proportional to their variability.

With this efficient encoding of the complex multiparametric trajectories in low dimensional parameter space demonstrated, we further extend this approach towards the process optimization. Here, we implement the Bayesian optimization over pre-encoded set of trajectories. The latter in turn is generated based on domain specific intuition and prior knowledge or can be derived assuming some desirable factors of variability for a given set of functions. The latent vector is decoded to yield the process trajectory, and the experiment is run using this trajectory. The experimental output is used as a signal to guide the BO over the latent space.

Process optimization in a ferroelectrics simulation

As a model system, we implement the analysis of the domain structure evolution in the ferroelectric material. Here, we use the FerroSIM model, representing the discrete lattice of continuous spins as described in ref. [36] However, the choice of this specific example does not affect the generality of proposed approach. FerroSIM is a lattice model representing a ferroelectric material first introduced by Ricinschi *et al.*[37] In this model,
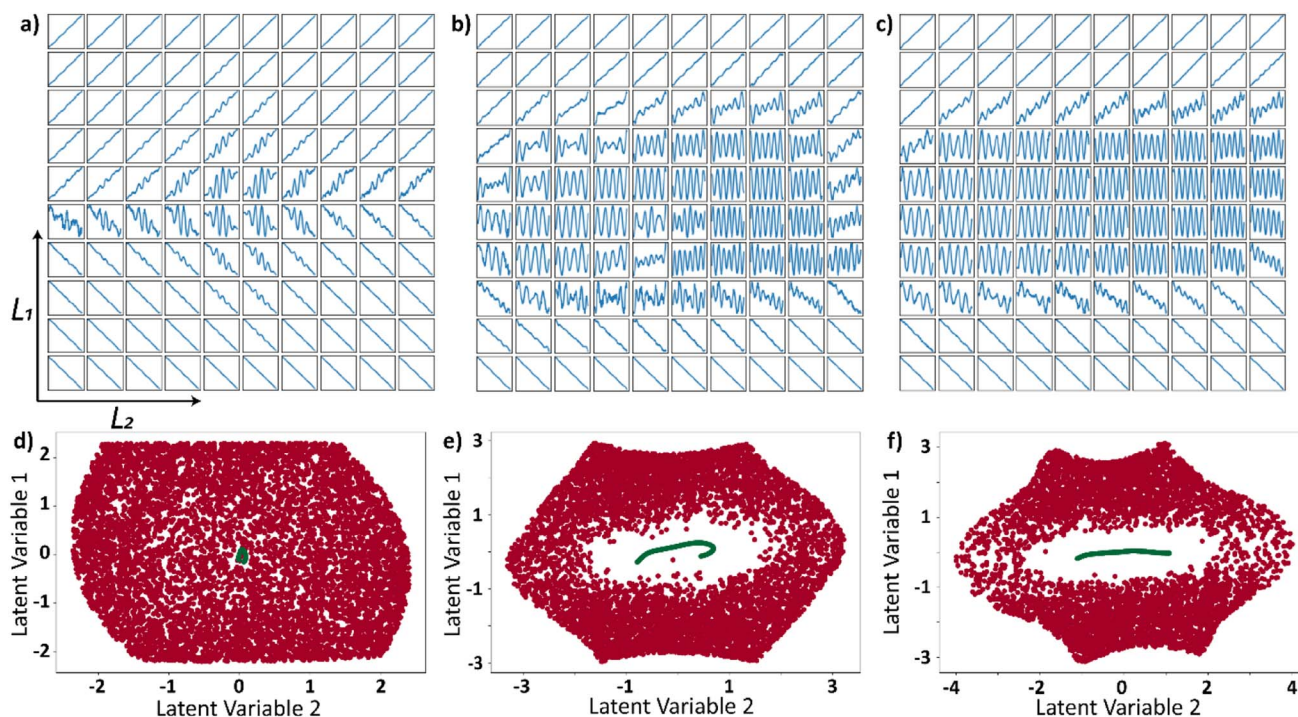

Fig. 4 Decoded latent space when (a) 1%, (b) 10%, and (c) 30% of the input trajectories (linear) are corrupted with the trajectories that do not follow the same physics of the system. Distribution of input dataset in the latent space for the presence of (d) 1%, (e) 10%, and (f) 30% wrong trajectories in the input system. The red dots represent the linear (right trajectories) and the green dots represent the sinusoids (wrong trajectories) in the latent space.

the order parameter, polarization in the case of a ferroelectric material is represented as a continuous quantity at every lattice site. The local free energy function is assumed to be of the Ginzburg–Landau–Devonshire (GLD) form and is given by eqn (2),

$$F_{ij} = \alpha_1 \left(p_{x_{ij}}^{\;2} + p_{y_{ij}}^{\;2} \right) + \alpha_2 \left(p_{x_{ij}}^{\;4} + p_{y_{ij}}^{\;4} \right) + \alpha_3 p_{x_{ij}}^{\;2} p_{y_{ij}}^{\;2} - E_{\mathrm{loc}_{x_{ij}}} p_{x_{ij}} - E_{\mathrm{loc}_{y_{ij}}} p_{y_{ij}}$$

(2)

where α_1, α_2, and α_3 are the GLD coefficients, (i, j) refers to the row and column index of a lattice site, p_x and p_y are the components of local polarization vector in x and y direction respectively at the lattice site referred by (i, j), $-E_{\mathrm{loc}} p$ term is the resultant free energy locally due to the coupling between polarization (p) and the electric field. E_{loc} is the local field at the lattice site and is given by the eqn (3). It is a vector sum of external field (E_{ext}), depolarization field (E_{dep}), and any random field disorders $(E_{\mathrm{d}}(i,j))$ present at the lattice site. The depolarization field is assumed to be proportional to the total polarization of the lattice and is given by eqn (4).

$$E_{\mathrm{loc}} = E_{\mathrm{ext}} + E_{\mathrm{dep}} + E_{\mathrm{d}}(i,j)$$ (3)

$$E_{\mathrm{d}} = -\alpha_{\mathrm{dep}} \langle p \rangle, \quad \text{where} \quad \langle p \rangle = \sum_N p$$ (4)

The total free energy of the lattice is then a sum of the local free energies and the interactions between the lattice sites and their neighbors. We will only be considering the nearest neighbor interactions for this analysis; complex models can be constructed from this model where the interactions last several unit cells. The total free energy of the lattice system is given by the eqn (5), where K is the coupling constant and the interactions are summed over the entire neighborhood *via* summation over k,l. Summation over i, j results in total interaction energy over the entire lattice.

$$F = \sum_{i,j}^{N} F_{ij} + K \sum_{k,l} \left(p_{x_{ij}} - p_{x_{i+k,j+l}} \right)^2 + K \sum_{k,l} \left(p_{y_{ij}} - p_{y_{i+k,j+l}} \right)^2 \quad (5)$$

Finally, the time dynamics of the system are given by the classical Landau–Khalatnikov equation and is given by eqn (6). The polarization at each time step is then updated by calculating the derivatives of the total free energy with respect to the polarization field at the previous time step and are given by eqn (7) and (8).

$$\gamma \frac{\mathrm{d} p_{i,j}}{\mathrm{d} t} = -\frac{\partial F}{\partial p_{i,j}}$$ (6)

$$\frac{\mathrm{d} p_{x_{ij}}}{\mathrm{d} t} = -\gamma^{-1} \left(2\alpha_1 p_{x_{ij}} + 4\alpha_2 p_{x_{ij}}^{\;3} + 2\alpha_3 p_{x_{ij}}^{\;2} p_{y_{ij}} \right.$$

$$\left. + 2K \sum_{k,l} \left(p_{x_{ij}} - p_{x_{i+k,j+l}} \right) - E_{\mathrm{loc}} \right)$$ (7)

$$\frac{\mathrm{d} p_{y_{ij}}}{\mathrm{d} t} = -\gamma^{-1} \left(2\alpha_1 p_{y_{ij}} + 4\alpha_2 p_{y_{ij}}^{\;3} + 2\alpha_3 p_{x_{ij}} p_{y_{ij}}^{\;2} \right.$$

$$\left. + 2K \sum_{k,l} \left(p_{y_{ij}} - p_{y_{i+k,j+l}} \right) - E_{\mathrm{loc}} \right)$$ (8)

Simulation results for different field-time stimuli

The results of an example simulation of FerroSIM are shown in shown in Fig. 5. Here the FerroSIM is run on a 20 × 20 lattice with impurities placed at 15% of these sites. These impurities are placed at randomly selected locations can create local field disturbances in either x or y directions. For simplicity, external electric field applied is applied in only x-direction and is shown in Fig. 5a. The electric field applied is a sinusoid with a fixed frequency whose amplitude is modulated with a positive exponential function. This external electric field is applied for 3 seconds divided into 900 times steps. The system is then equilibrated for a span of 50-timesteps at a constant field strength of electric field at $t = 3$ s. The resulting polarization in x and y components as a function of the applied field are shown in Fig. 5b and as a function of time are shown in Fig. 5c. Polarization in x-direction as a function of time has a similar shape as the applied electric field in x-direction and is lagging which is typical to a kinetic simulation of the ferroelectric materials. Polarization in y-direction oscillates with a very low amplitude due to impurities causing local fields in y-direction and the coupling between the polarizations in x and y directions. Polarization as a function of the applied electric field in x-direction encompasses the properties associated with the traditional ferroelectric materials like hysteresis loop, coercive fields and remnant polarization. Finally, polarization plots at different time steps are shown in Fig. 5d. These time steps are also marked with green lines in Fig. 5a and c. At lower applied electric field strengths, the local electric field is dominated by the impurities. Random field impurities do not allow the polarization to completely align with the external electric field at lower field strengths. The coupling mechanism between the lattice sites enforced in the FerroSIM propagates this behavior to the nearby neighbors. This results in formation of domains whose polarization is different from the overall polarization of the lattice. This phenomenon can be observed in the 8th snapshot of Fig. 5d. This results in a non-zero curl of the polarization. In contrast at higher electric field strengths, polarization vectors of all the lattice sites are aligned with the external electric field which would result in zero curl. The final curl of the lattice also depends on the history of the electric field. If the electric field in the recent past is large compared to the fields due to impurities, all the effects due to the impurities will be negated.
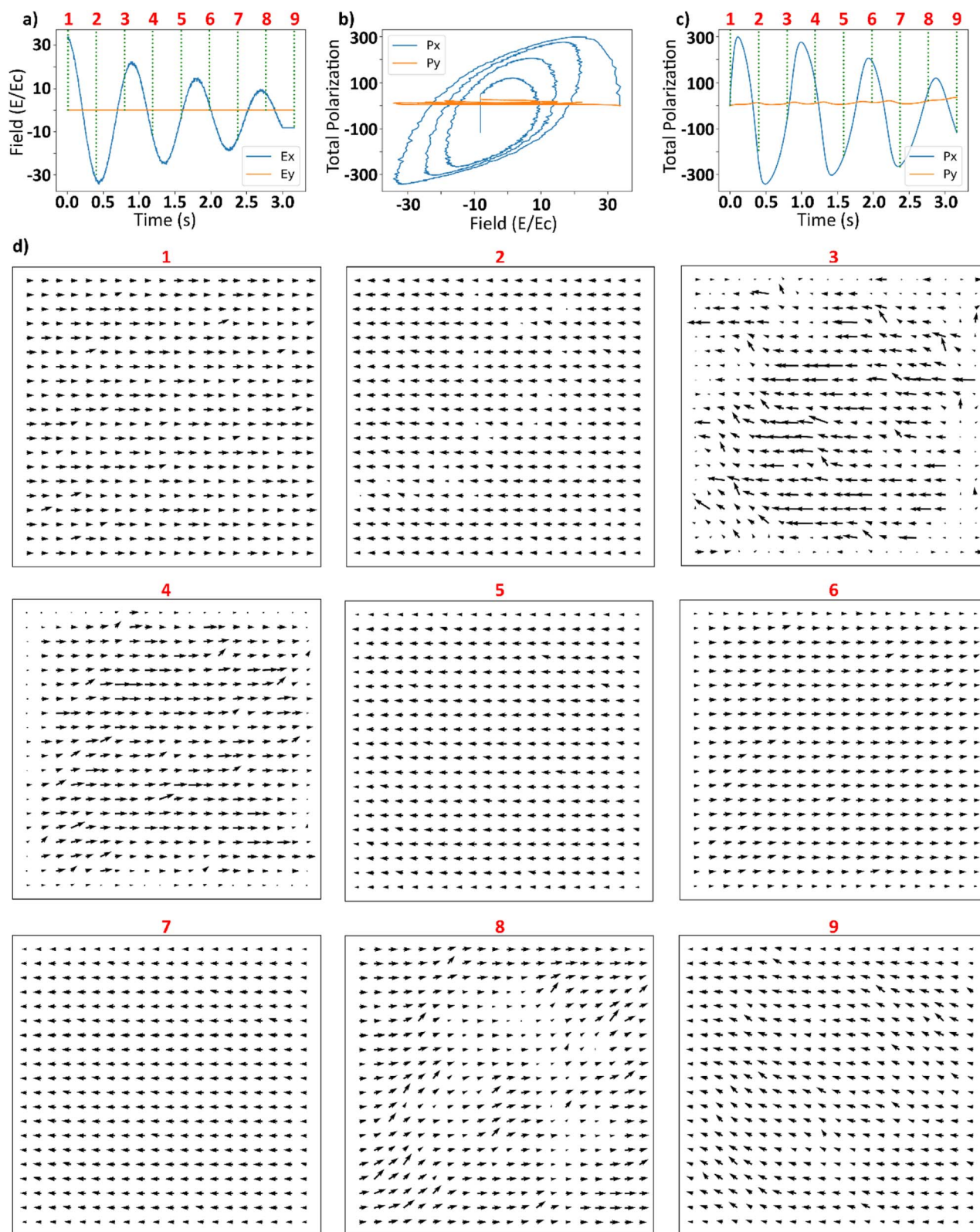
Fig. 5   (a) Applied external electric field as a proportion of coercive field of the system as a function of time. (b) Resulting polarization in x-direction and y-direction as a function of applied electric field in x-direction. (c) Resulting polarizations in x and y directions as a function of time. (d) The snapshots of polarization vectors at each lattice sites at uniform time intervals. Vertical green lines in (a) and (c) represent the time at which polarization maps in (d) are plotted.

Generating field trajectories and sampling in a reduced space

With the FerroSIM model illustrated, we proceed to apply the proposed optimization method for the process optimization, specifically the history of the electric field. As an initial family of possible trajectories, we chose family of sinusoids whose amplitudes are modulated by exponentials. To realize this on a 20×20 lattice at a given concentration of the impurities, we have created a set of electric field trajectories in the form of,
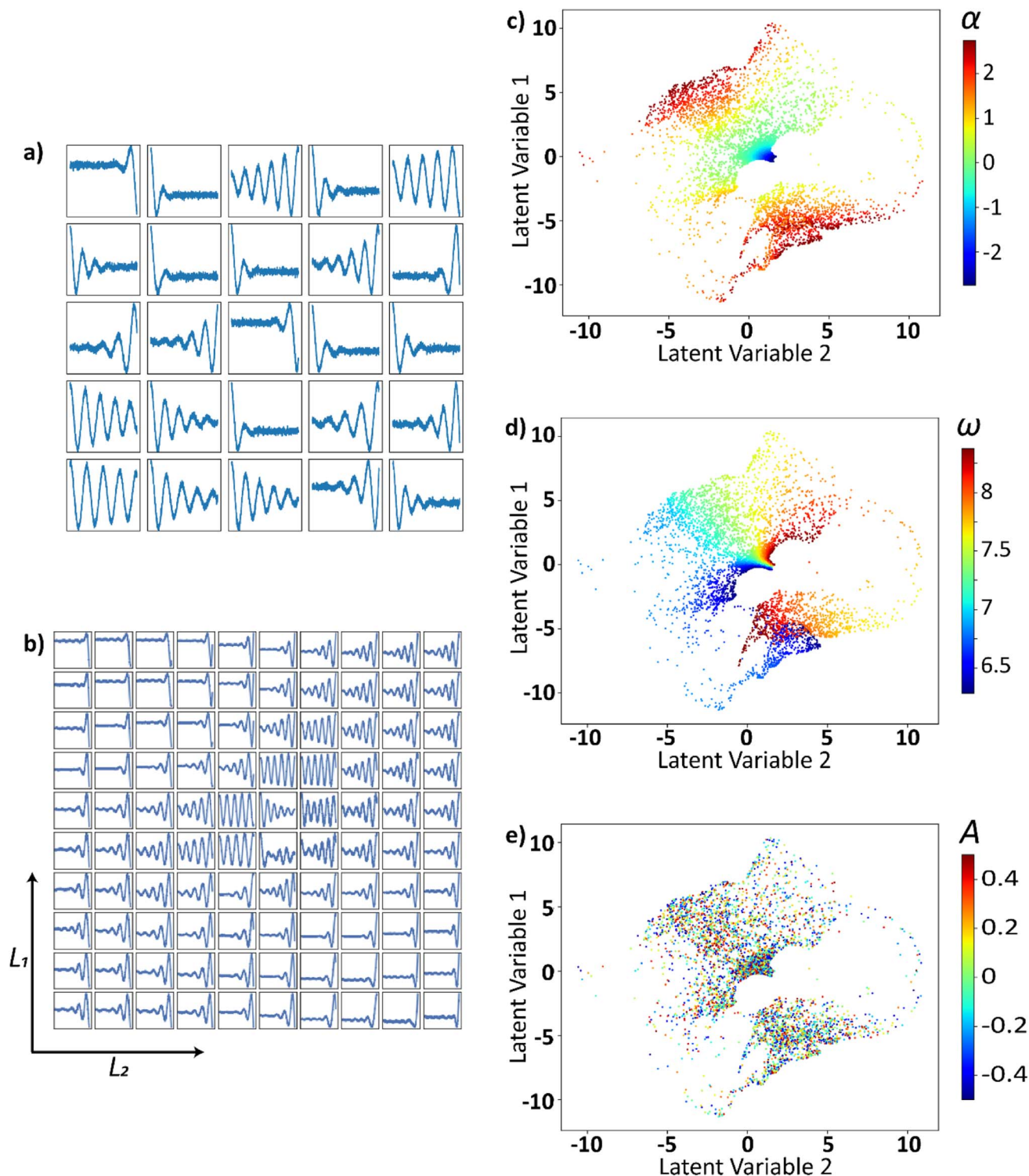

Fig. 6 (a) Twenty-five randomly drawn curves from the input space of electric fields to the VAE. (b) Uniformly sampled and decoded representation of the latent space. The input functions are represented as points in the latent space and are colored using their respective values of (c) α, (d) ω, and (e) A.

$$A \exp(\alpha t)\sin(\omega t) + B \qquad (9)$$

where $A \in [0, 0.75]$, $\alpha \in [-2.75, 2.75]$, $\omega \in [-2.75, 2.75]$. The parameters for each curve are randomly sampled from uniform distributions with the above-mentioned limits. The exponential factors are considered in such a way that the curve still has some visible sinusoidal structure whereas the frequency of the sinusoids is considered so that the electric field has 3–4 cycles in the time considered. The curves are then generated for 900 equidistant timesteps corresponding to 3 s using eqn (9) and then set to a constant value which is the final value of the field at the 900$^{\text{th}}$ timestep for the next 50-timesteps. This is done so that the system equilibrates at a constant external field for fifty timesteps before the curls are calculated. Each of these curves are normalized to fall in [−1, 1]. Since the curves with positive exponential coefficients always have their maxima at the end *i.e.*, at $t = 3$ s, after normalization, the final value of these curves at $t = 3$ s is always 1. This would mean that the half the curves generated equilibrate at the same electric field and might not result in a variation in curl. To negate this effect and to include the effect of the exponential factor on the final value of the electric field, we let curves develop for 1200 timesteps, normalized the curve, and then considered only the first 900 timesteps of the curve. The 50 equilibration time steps are then added later. We have performed this operation on the exponentially increasing and decreasing curves to be consistent.

These trajectories are then projected into a two-dimensional latent space using VAE as shown in Fig. 6. This smooth continuously varying latent space will be sampled to generate the electric field that belong to the family of curves described above. Twenty-five randomly drawn electric field trajectories from the family of electric field curves generated are shown in Fig. 6a. The two-dimensional latent space is uniformly sampled, decoded, and shown in Fig. 6b. The latent distributions *i.e.*, trajectories plotted as points in the 2-D latent space and are colored as a function of α in Fig. 6c, ω in Fig. 6d, and B in Fig. 6e. The latent space appears to be a strong function of α and ω, and they vary smoothly and continuously over the latent space. The latent space does not appear to be strongly correlated with A or B of the input curve. The effect of A and B on the latent space is lacking since each curve is normalized prior to the application of VAE.

The latent space is uniformly sampled into 10 000 points and then decoded back into the space of electric fields to study the variation of the sum of absolute curls. The sum of magnitude of the curl at each lattice site will be referred to as the curl of the system from hereon for brevity. These decoded electric field curves are also normalized by design and are multiplied by a constant (150) to make them practical for the simulations. This constant value is user selected and the discussions that follow depend on the value of this multiplication constant. Once a curve is decoded, only first 900-timesteps of the curve is used as an input to the FerroSIM, discarding the last 300-timesteps as mentioned earlier. The equilibration timesteps, where the external electric field is held at a constant value for 50-timesteps, are then added to all the curves. The FerroSIM
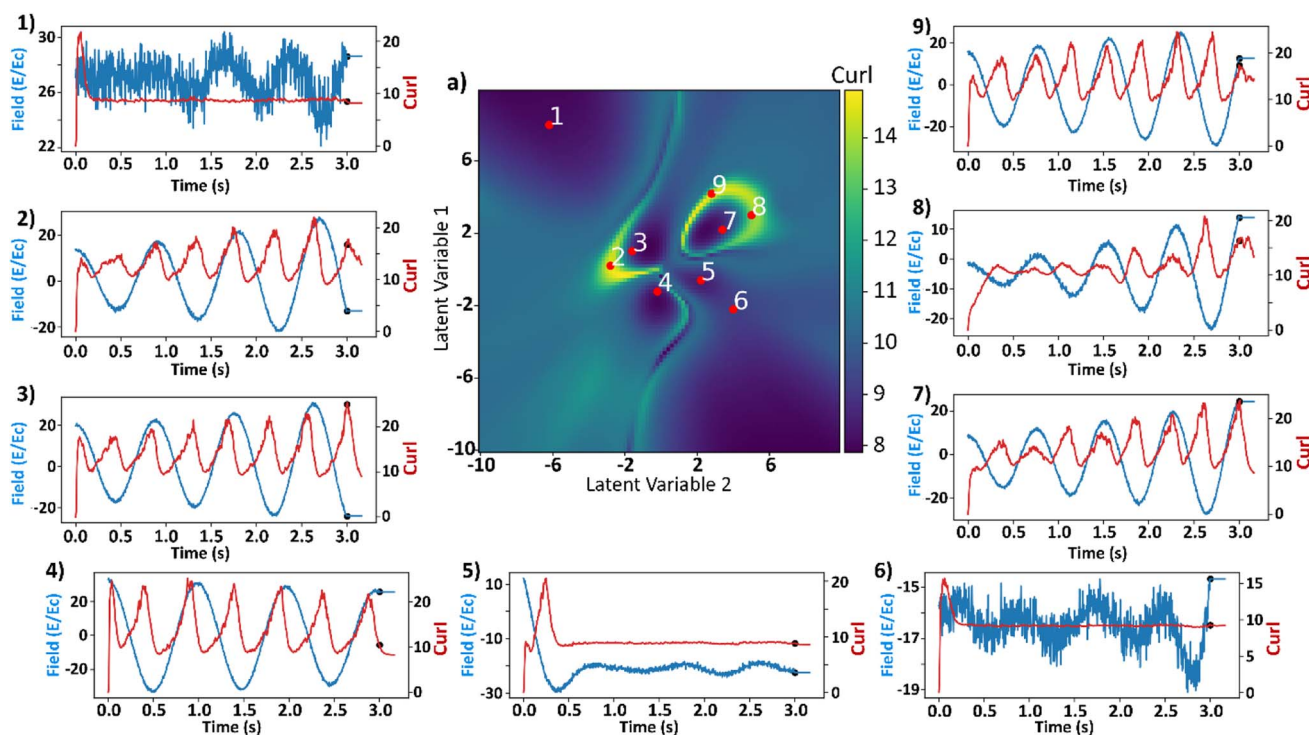

Fig. 7 (a) (Center of the figure) Sum of curl magnitudes at the end of the simulations as a function of latent space. Nine salient points are selected from the latent space and their respective histories of curl (red) and applied fields (blue) as a proportion of the coercive field are shown around the curl surface. The onset of equilibration region is marked as a scatter point in both curl and electric field histories.

simulations are then run for each of these electric fields corresponding to the 10 000 points in the latent space. The value of curl at the end of 950-timesteps is calculated for each point in the latent space and is shown in the middle in Fig. 7.

Mapping polarization curl as a function of input trajectories

The latent space distributions in Fig. 6c and d show that the encoding of the electric field curves is non-linear. Even with a linearly encoded latent space, the curl surface would be non-linear due to the abundance of parameters in the system that control the final value of curl. To obtain insights into the ground truth of the system, electric field and curl histories from different points in the latent space are plotted around the curl surface in Fig. 7. These points are assumed to represent their local neighborhood in the latent space. Some parts of the latent space are relatively straightforward to explain. For example, points 1 and 6 refer to the curves with a high absolute value of the electric field and low standard deviation. Electric fields with this behavior will lead to large magnitudes of polarization vectors as the system always experiences a high value of external electric field. Larger values of polarization vectors decrease the F_{loc} in eqn (2) by making the $-E_{loc}p$ term larger. At such large values of polarization fields, curl is not sustainable as any deviation from the average polarization vector will lead to a very high value of the coupling term in eqn (5) and thereby increasing the free energy of the system. This results in absence
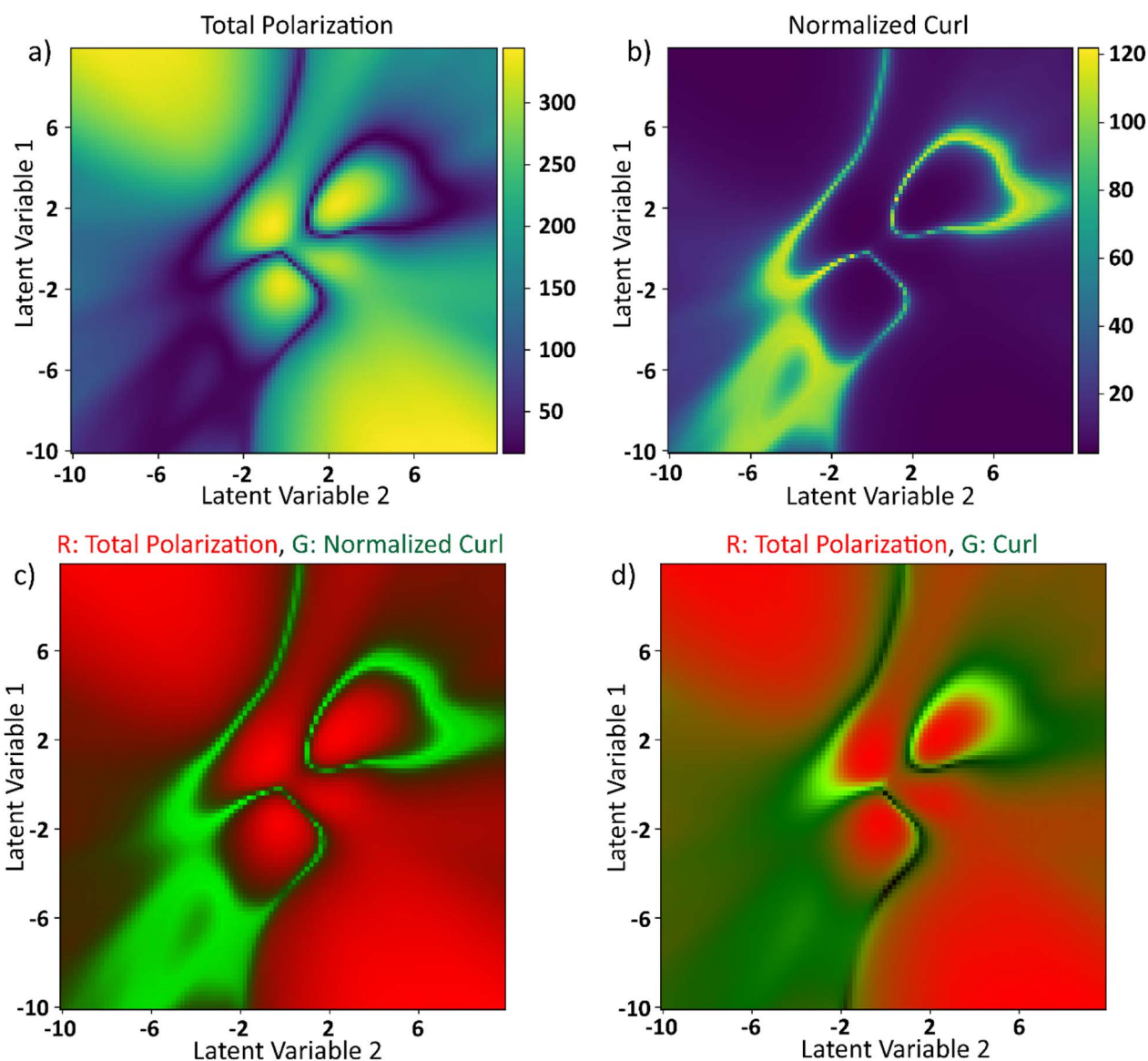

Fig. 8 (a) Total polarization magnitude as a function of latent space. (b) Unit curl: sum of absolute curls as a function of latent space when the polarization vectors at each lattice site are normalized at the end of 950-timesteps. (c) Total polarization magnitude and unit curl as a function of latent space are represented as red and green channels in an RGB image. (d) Total polarization magnitude and curl surface from Fig. 6 are represented as red and green channels respectively in an RGB image. Blue channel is set to be zeros in both (c) and (d).

and/or minimization of the curl at these locations. The rest of the points in the latent space exhibit two common trends: (i) the curl appears to be maximum at the local optima of the electric field, and (ii) in the equilibration region, the curl continuously decreases in magnitude. The apparent overlap of the maxima in curl and the local optima in electric fields seems counterintuitive. Polarization, lagging the applied electric field, is about to reach its own optimum at the optima of the applied electric fields. Furthermore, near the maximum value of polarization, as explained above, any deviation from the average polarization field of the lattice is not sustainable. For the second trend, the rate of decrease of curl seems to be proportional to the value of curl at the onset of the equilibration region. For example, comparing points 2 and 3, the value of curl at the onset of the equilibration region is ~25 and 17, respectively. The curl in the equilibration region corresponding to point 2 falls rapidly to a low value. The curl corresponding to point 3 falls slowly and belongs to one of the two regions that produce a local maximum in the final curl surface.

To explore effects of the final value of total polarization on the resulting curl of the system, we plotted the total polarization at the end of the simulation as a function of latent space in Fig. 8a. To negate the effects of magnitude of polarization on the final value of curl, the curl is recalculated after normalizing the polarization vectors at each lattice site at the end of the simulations. The simulations are however run using the actual polarization vectors for the entire duration. The recalculated curl (referred to as normalized curl) is shown in Fig. 8b. The normalized curl surface will provide insights on how much the polarization field rotates without accounting for the magnitude of polarization vectors. It can be observed from Fig. 8a and b that the polarization surface and normalized curl surface are

inversed with respect to each other *i.e.*, the normalized curl is maximum when the polarization is minimum and *vice versa*. To visualize this better, the total polarization surface is used as the R-channel and the normalized curl is used as the G-channel while the B-channel is left empty in an RGB image and is shown in Fig. 8c. It can be observed that most part of the image is either entirely green (large values of normalized curl) or entirely red (large values of polarization) with very little overlap. This is because when the magnitude of the polarization is small, any curl (deviation from the average polarization field) in the system will not lead to a substantial coupling term in free energy equation (eqn (5)). These deviations from the average polarization vector will lead to a large value of the normalized curl. However, since any such deviations at large polarization fields are not sustainable and any small deviations will not contribute much to the normalized curl, its value in these regions is going to be small. But the final value of curl is also dependent on the actual magnitude of the polarization vectors. A field with higher magnitudes of polarization vectors will result in higher value of curl compared to a field with smaller polarization vectors but a similar rotational tendency. To show this, the total polarization is used as the R-channel and the curl surface from Fig. 7 is used as the G-channel in an RGB image and is shown in Fig. 8d. It can be observed that there is some overlap between the green and red regions. So far from the discussion, it is apparent that neither a large nor a small value of the total polarization corresponds to the maximum curl. The polarization needs to be at an optimal value to create a large curl.

We carried out additional analyses to investigate the coincidence of local maxima of the curl surface and the local optima of electric field (Fig. 9). Points 2, 5, and 7 from Fig. 7 are
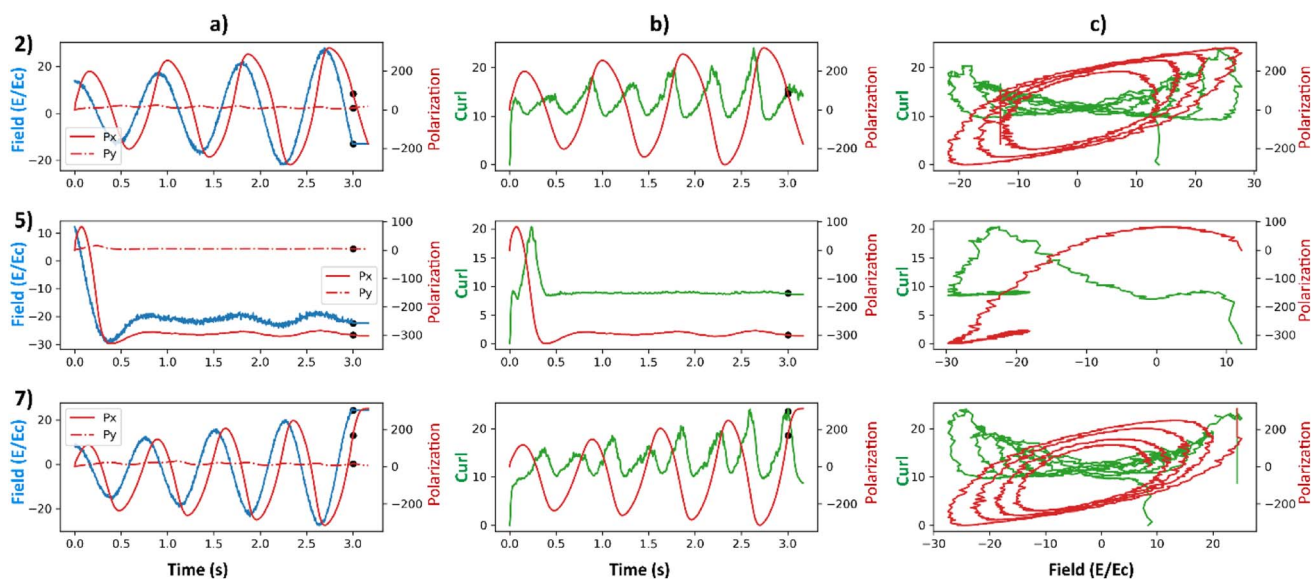

Fig. 9 (a) Total polarization in *x*-direction (red), total polarization in *y*-direction (red-dotted), and curl (blue) as a function of time. (b) Curl (green) and total polarization magnitude (red) as a function of time. (c) Curl (green) and total polarization magnitude (red) as a function of applied electric field. The applied electric field is represented in the units of coercive field of the system. The rows represent the points 2, 5, 7 in the curl surface from Fig. 6. The onset of equilibration region is marked with a black scatter point in (a) and (b).
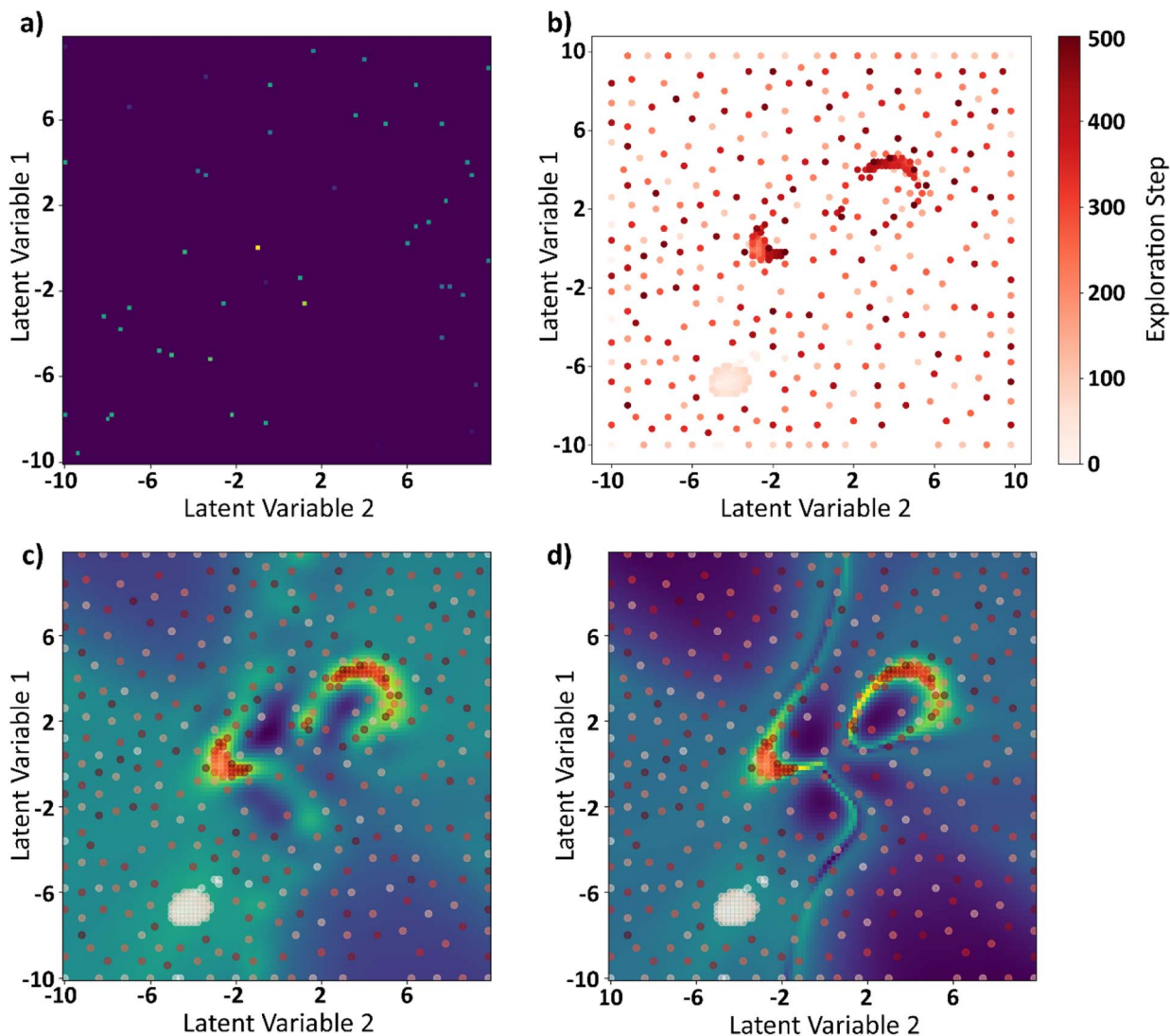
Fig. 10 (a) Randomly sampled 0.5% of the total points (100 × 100) in the latent space that are provided initially as the seed points to GP−BO, (b) points explored by BO in the latent space at the end of 500 iterations as function of the iteration number, (c) the curl surface approximated by BO at the end of 500 iterations, (d) actual curl surface obtained through brute force calculations. The red scatter points in (c) and (d) are the points explored by BO in 500 iterations.

considered for this analysis. Point 2 corresponds to a large value of final curl, point 7 corresponds to a large value of curl when the system starts to equilibrate which then falls rapidly to a very small value of curl. Finally, point-5 corresponds to a very large negative exponential factor (α) in the electric field which causes the electric field to decay to a constant value in less than a cycle. For these three points in the latent space, polarizations in x and y directions and the applied electric field as a function of time are plotted in Fig. 9a. Curl and polarization magnitudes as a function of time are plotted in Fig. 9b, and as a function of electric field in Fig. 9c. It can be observed from Fig. 9a that the polarization lags the electric field in every case. This is intuitive as the effect always lags the cause in any kinetic simulation (akin to learning rate in ML). From Fig. 9b it can be observed that the curl is maximum after a few timesteps once the polarization crosses zero. This is because the polarization

vectors are in a state of maximum normalized curl right after the coercive field. Subsequently, the polarization vectors start to grow in magnitude because of a high value in the applied electric field. It takes a few time steps in the simulation for the polarization to grow to a magnitude that constitute the maximum curl for the electric field's half cycle. In our simulation, the half cycle of electric field is approximately same as the time steps the curl takes to attain the local maximum. This results in apparent coincidence of the local optima of the electric fields with the local maxima of the curl which can also be observed in Fig. 9c. This local maximum of curl is not stable as the electric field has a very high absolute value at its optima and destroys the curl in the subsequent time steps. This destruction of curl is what results in a low final value of curl for point 7. For point 5, since the electric field is always a high value and so are the polarization fields, any small deviations in the

polarization field might result in some curl. The results so far depend on the various parameters that go into the FerroSIM simulation. Any change in these parameters will lead to a different curl surface but the underlying physics will still be same.

Process optimization to maximize curl of polarization

Finally, we demonstrate the Bayesian optimization of processing trajectories in the latent space *via* the maximization of the curl. Note that the chosen material model generally is not prone to the natural formation of polarization vortices, and hence non-zero curl can emerge only due to the interactions between the frozen disorder and polarization field for a given field history. The curl distribution in the latent space is obtained by a brute force search where the simulations are run for every point in the latent space. When we are only concerned about optimizing the magnitude of the curl and/or when the latent space is dimensionality is large making the process of running simulations at each latent point intractable, one can utilize the Bayesian optimization towards a guided search of the latent space.

Bayesian optimization is an extension of Gaussian Processes (GP) regression. GP regression is an approach of interpolating or learning a function f, given the set of observations $D = \{(x_1, y_1), \ldots, (x_N, y_N)\}$. It is assumed that the arguments x_i are known exactly, whereas the observations can have some Gaussian noise with zero mean in them *i.e.*, $y_i = f(x_i) + \varepsilon$. The key assumption of the GP method is that the function f has a prior distribution $f \sim \mathcal{GP}(0, K_f(x, x'))$, where K_f is a covariance function (kernel). The kernel function defines the relationship between the values of the function across the parameter space, and its functional form is postulated as a part of the fit. Here we implemented the Matern kernel function given by eqn (10)

$$k_{\text{Matern}}(x_1, \ x_2) = \sigma^2 \exp\left(-\sqrt{5} \times \frac{|x_1 - x_2|}{l} \right)\left(1 + \sqrt{5} \right.$$
$$\left. \times \frac{|x_1 - x_2|}{l} + \frac{5}{3} \times \frac{|x_1 - x_2|^2}{l} \right) \quad (10)$$

The learning is performed *via* Bayesian inference in a function space and the expected value of the function, corresponding uncertainties, and kernel hyperparameters are optimized simultaneously. The output of the GP process is then the predicted data set, uncertainty maps representing the quality of prediction, and kernel hyperparameters. The aspect that makes GP standout from the other extrapolation methods by providing not only the function value, but also the uncertainty associated with the prediction. Bayesian Optimization (BO) then samples the input space based on an inbuilt or user defined acquisition function which uses the predictions by GP. The acquisition function can be set to a pure exploration state where the point selected by BO is either random or proportional to the uncertainty of the prediction. It can also be set in an exploitation state

where the selection of next point in the input space is based on the target value of the function. In our case, the acquisition function is a linear combination of mean and standard deviation and is defined by eqn (11).

$$\text{acq}(x: \lambda) = \mu(x) + \lambda\sigma(x) \quad (11)$$

where μ and σ are the mean and standard deviation (uncertainty) of the GP predictions respectively and λ is the tunable hyperparameter which is set to 10 for the complete exploration of the latent space. The effects of the value of hyperparameter λ can be observed in the accompanying Jupyter notebook that runs the BO algorithm with 4 different values of λ *viz.*, 1, 2, 5, 10.

The applications of BO algorithm are ubiquitous and can be found in a plethora of fields including but not limited to materials science,[38–40] catalysis,[41,42] 3D printing,[43,44] chemistry,[20,45] solar cells,[5,46] pharmaceuticals,[47,48] While the application of BO in the latent space of the VAE is a recent development,[44,45,49–52] the optimization in the latent space of trajectories where the trajectories in the input dataset are constructed for a domain specific application is still not explored.

The curl surface obtained through brute force is a result of running 10 000 FerroSIM simulations. For BO based GP, we ran 0.5% of the simulations and the rest of the points are interpolated by GP. At each iteration, BO samples a point in the latent space based on exploration–exploitation trade off and the simulation corresponding to that point in the latent space is run and the actual value of curl is obtained. This point is now added to the input space to GP and the process is repeated for 500 iterations. The results of BO–GP analysis are shown in Fig. 10.

The initial set of points at the start of the BO process are shown in Fig. 10a. These are the 0.5% of the total points the latent space was divided into. Fig. 10b shows the points explored by the BO in 500 iterations progressively. GP interpolations of the curl surface can be plotted in the Jupyter notebook attached in the data availability statement. Finally, the explored points are plotted in the GP predicted curl surface at the end of 500 iterations in Fig. 10c and the brute force curl surface in Fig. 10d. BO does a sub optimal job in exploring one of the two local maxima that is because the curl surface is multimodal with multiple local optima and none of the initial seed points to the BO are a part of this region. However, a mere 5.5% of the points are explored at the end of this exercise and the BO was able to identify the two local maxima of the curl surface even with incomplete information.

Summary

To summarize, here we proposed the approach for optimization of processing trajectories towards target functionalities *via* Bayesian optimization over the low-dimensional latent space of variational autoencoder. The initial training set here is formed based on the domain knowledge or human intuition. The effects of the variability of the initial trajectory set and distribution of parameters in the latent space are explored.

This approach is further used to maximize specific functionalities of material represented *via* lattice Ginzburg–Landau

model. Here we have demonstrated that this approach can be used to tune the process towards maximization of the total curl of polarization field in the system, which is a symmetry breaking phenomenon inconsistent with the symmetry of the lattice or applied stimulus. The maximization of the curl is found to occur in the vicinity of the phase transition as a result of the spontaneous symmetry breaking in the vicinity of the defects.

While here this approach is applied for specific family of the trajectories and model, we believe this approach to be universally applicable to a broad range of optimization problems such as materials processing, ferroelectric poling, and so on. It should be noted that the choice of the initial trajectory set has to be based on domain specific consideration and chosen broad enough to have sufficient variability but at the same time sufficiently narrow to avoid overly localized minima.

Data availability

● The python code delineating the entire workflow of the paper is available at **https://github.com/saimani5/Notebooks_for_papers/blob/main/FerroSim_VAE_BO_code_for_paper.ipynb**. The datasets required to reproduce the results can be downloaded inside the Jupyter Notebook.

● The code for FerroSim model can be found at **https://github.com/ramav87/FerroSim/tree/rama_dev_updated**

● The code for Variational Autoencoder can be found at **https://github.com/ziatdinovmax/pyroVED**

● The code for Gaussian processes and Bayesian optimization can be found at **https://github.com/ziatdinovmax/GPim**

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 B. P. MacLeod, F. G. L. Parlane, T. D. Morrissey, F. Hase, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott, T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein and C. P. Berlinguette, *Sci. Adv.*, 2020, **6**(20), 8.

2 S. Langner, F. Hase, J. D. Perea, T. Stubhan, J. Hauch, L. M. Roch, T. Heumueller, A. Aspuru-Guzik and C. J. Brabec, *Adv. Mater.*, 2020, **32**(14), 6.

3 Z. Li, M. A. Najeeb, L. Alves, A. Z. Sherman, V. Shekar, P. Cruz Parrilla, I. M. Pendleton, W. Wang, P. W. Nega, M. Zeller, J. Schrier, A. J. Norquist and E. M. Chan, *Chem. Mater.*, 2020, **32**(13), 5650–5663.

4 R. W. Epps, M. S. Bowen, A. A. Volk, K. Abdel-Latif, S. Han, K. G. Reyes, A. Amassian and M. Abolhasani, *Adv. Mater.*, 2020, **32**(30), 2001626.

5 K. Higgins, S. M. Valleti, M. Ziatdinov, S. V. Kalinin and M. Ahmadi, *ACS Energy Lett.*, 2020, **5**(11), 3426–3436.

6 A. Krull, P. Hirsch, C. Rother, A. Schiffrin and C. Krull, *Commun. Phys.*, 2020, **3**(1), 8.

7 K. P. Kelley, Y. Ren, A. N. Morozovska, E. A. Eliseev, Y. Ehara, H. Funakubo, T. Giamarchi, N. Balke, R. K. Vasudevan, Y. Cao, S. Jesse and S. V. Kalinin, *ACS Nano*, 2020, **14**(8), 10569–10577.

8 J. Ling, M. Hutchinson, E. Antono, S. Paradiso and B. Meredig, *Integr. Mater. Manuf. Innov.*, 2017, **6**(3), 207–217.

9 L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, *npj Comput. Mater.*, 2016, **2**(1), 16028.

10 F. Yang, C. Rippe, J. Hodges and B. Lattimer, *Fire Mater.*, 2019, **43**(6), 694–706.

11 M. Ahmadi, M. Ziatdinov, Y. Zhou, E. A. Lass and S. V. Kalinin, *Joule*, 2021, **5**(11), 2797–2822.

12 A. G. Kusne, H. Yu, C. Wu, H. Zhang, J. Hattrick-Simpers, B. DeCost, S. Sarker, C. Oses, C. Toher, S. Curtarolo, A. V. Davydov, R. Agarwal, L. A. Bendersky, M. Li, A. Mehta and I. Takeuchi, *Nat. Commun.*, 2020, **11**(1), 5966.

13 L. Velasco, J. S. Castillo, M. V. Kante, J. J. Olaya, P. Friederich and H. Hahn, *Adv. Mater.*, 2021, **33**(43), 2102301.

14 E. J. Braham, R. D. Davidson, M. Al-Hashimi, R. Arróyave and S. Banerjee, *Dalton Trans.*, 2020, **49**(33), 11480–11488.

15 Y. Tian, D. Xue, R. Yuan, Y. Zhou, X. Ding, J. Sun and T. Lookman, *Phys. Rev. Mater.*, 2021, **5**(1), 013802.

16 S. M. P. Valleti, R. Ignatans, S. V. Kalinin and V. Tileli, *Small*, 2022, 2104318.

17 L. Scipioni, C. Sanford, E. van Veldhoven and D. Maas, *Microsc. Today*, 2011, **19**(3), 22–26.

18 O. Martin, *Bayesian Analysis with Python: Introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ*, Packt Publishing, 2nd edn, 2018.

19 B. Lambert, *A Student's Guide to Bayesian Statistics*, SAGE Publications Ltd, 1st edn, 2018.

20 R. J. Hickman, M. Aldeghi, F. Häse and A. Aspuru-Guzik, *Digital Discovery*, 2022, **5**, 732–744.

21 M. Rashidi and R. A. Wolkow, *ACS Nano*, 2018, **12**(6), 5185–5189.

22 M. Ziatdinov, Y. Liu, A. N. Morozovska, E. A. Eliseev, X. Zhang, I. Takeuchi and S. V. Kalinin, *Adv. Mater.*, 2021, **34**(20), 2201345.

23 T. Xie, A. France-Lanord, Y. Wang, Y. Shao-Horn and J. C. Grossman, *Nat. Commun.*, 2019, **10**(1), 2667.

24 A. S. Barnard and G. Opletal, *Nanoscale*, 2019, **11**(48), 23165–23172.

25 S. V. Kalinin, K. Kelley, R. K. Vasudevan and M. Ziatdinov, *ACS Appl. Mater. Interfaces*, 2021, **13**(1), 1693–1703.

26 Z. Liu, N. Rolston, A. C. Flick, T. W. Colburn, Z. Ren, R. H. Dauskardt and T. Buonassisi, *Joule*, 2022, **6**(4), 834–849.

27 H. Choubisa, J. Abed, D. Mendoza, Z. Yao, Z. Wang, B. Sutherland, A. Aspuru-Guzik and E. H. Sargent, *arXiv*, 2022, preprint, arXiv:2205.09007.

28 K. M. Roccapriore, O. Dyck, M. P. Oxley, M. Ziatdinov and S. V. Kalinin, *ACS Nano*, 2022, **16**(5), 7605–7614.

29 Y. Liu, M. Ziatdinov and S. V. Kalinin, *ACS Nano*, 2022, **16**(1), 1250–1259.

30 Y. Liu, K. P. Kelley, R. K. Vasudevan, H. Funakubo, M. A. Ziatdinov and S. V. Kalinin, *Nat. Mach. Intell.*, 2022, **4**(4), 341–350.

31 K. M. Roccapriore, S. V. Kalinin and M. Ziatdinov, *arXiv*, 2021, preprint, arXiv:2108.03290.

32 S. V. Kalinin, M. P. Oxley, M. Valleti, J. Zhang, R. P. Hermann, H. Zheng, W. Zhang, G. Eres, R. K. Vasudevan and M. Ziatdinov, *npj Comput. Mater.*, 2021, **7**(1), 181.

33 S. V. Kalinin, S. Zhang, M. Valleti, H. Pyles, D. Baker, J. J. De Yoreo and M. Ziatdinov, *ACS Nano*, 2021, **15**(4), 6471–6480.

34 R. Ignatans, M. Ziatdinov, R. Vasudevan, M. Valleti, V. Tileli and S. V. Kalinin, *Adv. Funct. Mater.*, 2022, 2100271.

35 M. Valleti, S. V. Kalinin, C. T. Nelson, J. J. P. Peters, W. Dong, R. Beanland, X. Zhang, I. Takeuchi and M. Ziatdinov, 2021, arXiv:2101.06892.

36 S. V. Kalinin, M. Ziatdinov and R. K. Vasudevan, *J. Appl. Phys.*, 2020, **128**(2), 024102.

37 D. Ricinschi, C. Harnagea, C. Papusoi, L. Mitoseriu, V. Tura and M. Okuyama, *J. Phys.: Condens. Matter*, 1998, **10**(2), 477–492.

38 P. I. Frazier and J. Wang, in *Information Science for Materials Discovery and Design*, ed. T. Lookman, F. J. Alexander and K. Rajan, Springer International Publishing, Cham, 2016, pp. 45–75.

39 L. Kotthoff, H. Wahab and P. Johnson, *arXiv*, 2021, preprint, arXiv:2108.00002.

40 Y. Zhang, D. W. Apley and W. Chen, *Sci. Rep.*, 2020, **10**(1), 4924.

41 W. Yang, T. T. Fidelis and W.-H. Sun, *ACS Omega*, 2020, **5**(1), 83–88.

42 A. S. Nugraha, G. Lambard, J. Na, M. S. A. Hossain, T. Asahi, W. Chaikittisilp and Y. Yamauchi, *J. Mater. Chem. A*, 2020, **8**(27), 13532–13540.

43 T. Erps, M. Foshey, M. K. Luković, W. Shou, H. H. Goetzke, H. Dietsch, K. Stoll, B. von Vacano and W. Matusik, *Sci. Adv.*, 2021, **7**(42), eabf7435.

44 T. Xue, T. J. Wallin, Y. Menguc, S. Adriaenssens and M. Chiaramonte, *Extreme Mech. Lett.*, 2020, **41**, 100992.

45 R.-R. Griffiths and J. M. Hernández-Lobato, *Chem. Sci.*, 2020, **11**(2), 577–586.

46 H. C. Herbol, W. Hu, P. Frazier, P. Clancy and M. Poloczek, *npj Comput. Mater.*, 2018, **4**(1), 51.

47 S. Sano, T. Kadowaki, K. Tsuda and S. Kimura, *J. Pharm. Innovation*, 2020, **15**(3), 333–343.

48 E. O. Pyzer-Knapp, *IBM J. Res. Dev.*, 2018, **62**(6), 2.

49 E. Siivola, A. Paleyes, J. González and A. Vehtari, *Appl. AI Lett.*, 2021, **2**(2), e24.

50 R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams and A. Aspuru-Guzik, *ACS Cent. Sci.*, 2018, **4**(2), 268–276.

51 M. J. Kusner, B. Paige and J. M. Hernández-Lobato, in *Proceedings of the 34th International Conference on Machine Learning*, ed. P. Doina and T. Yee Whye, PMLR, Proceedings of Machine Learning Research, 2017, vol. 70, pp. 1945–1954.

52 A. Tripp, E. Daxberger and J. M. Hernández-Lobato, in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Curran Associates Inc., Vancouver, BC, Canada, 2020, p. 945.