Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination

Wenjian Hu,[1,2] Rajiv R. P. Singh,[1] and Richard T. Scalettar[1]

[1]*Department of Physics, University of California Davis, Davis, California 95616, USA*
[2]*Department of Computer Science, University of California Davis, Davis, California 95616, USA*

We apply unsupervised machine learning techniques, mainly principal component analysis (PCA), to compare and contrast the phase behavior and phase transitions in several classical spin models—the square- and triangular-lattice Ising models, the Blume-Capel model, a highly degenerate biquadratic-exchange spin-1 Ising (BSI) model, and the two-dimensional XY model—and we examine critically what machine learning is teaching us. We find that quantified principal components from PCA not only allow the exploration of different phases and symmetry-breaking, but they can distinguish phase-transition types and locate critical points. We show that the corresponding weight vectors have a clear physical interpretation, which is particularly interesting in the frustrated models such as the triangular antiferromagnet, where they can point to incipient orders. Unlike the other well-studied models, the properties of the BSI model are less well known. Using both PCA and conventional Monte Carlo analysis, we demonstrate that the BSI model shows an absence of phase transition and macroscopic ground-state degeneracy. The failure to capture the "charge" correlations (vorticity) in the BSI model (XY model) from raw spin configurations points to some of the limitations of PCA. Finally, we employ a nonlinear unsupervised machine learning procedure, the "autoencoder method," and we demonstrate that it too can be trained to capture phase transitions and critical points.

I. INTRODUCTION

In the age of big data, machine learning has become an indispensable tool whose utility transcends scientific and academic boundaries [1]. From social networking [2,3] to object and image recognition [4,5], from advertising to finance [6], from engineering to medicine [7], from biological physics [8,9] to astrophysics [10], wherever there is a preponderance of information and real data, machine learning is helping to find and quantify patterns and even discover basic laws [11].

Condensed-matter physics is a relatively late arrival in this field. Although isolated applications of machine learning have been made over many years, it is only recently that a concerted effort toward the use of these methods for addressing problems in many-body physics has started to emerge [12–21]. On the one hand, machine-learning algorithms, such as deep learning [22,23], have profound connections to the foundations of statistical physics [24,25]. Scaling and renormalization [26] are core principles that underlie our ability to make sense of macroscopic phenomena, in particular their simplicity and universality despite incredible microscopic complexity. It is perhaps not surprising that the way forward for machines to learn from large data sets would incorporate similar principles.

On the other hand, machine learning is also being used, mostly in conjunction with Monte Carlo [27–29] data, to recognize phases and phase transitions [13,15,16,20,21,30], to maximize the accuracy of variational Monte Carlo methods in quantum many-body systems [12], to guide choices of Monte Carlo moves [31,32], to explore overcoming the famous sign-problem bottleneck [15], and to infer spectral functions by performing analytic continuation from imaginary-time to real-time data [33]. Both supervised learning [13,34], which depends on first training the system with trial datasets, and unsupervised learning [16], which works with no prior training, have been found useful.

In this paper, we focus further on Monte Carlo simulations and examine the extent to which unsupervised machine learning can succeed in deciphering and distinguishing different physics. We introduce a simple biquadratic-exchange spin-1 Ising model that has a macroscopic ground-state degeneracy that can be lifted by a small bilinear-exchange term. We contrast its behavior with better understood and studied cases—the Ising ferromagnet on a square lattice [35] and the Ising antiferromagnet on a triangular lattice [36–38], the Blume-Capel model [39–43] (which has both first- and second-order transitions to phases with true long-range order), and the two-dimensional XY model [44,45], which has a low-temperature power-law correlated phase.

Our primary results are as follows: (a) Simple machine-learning methods are good at recognizing order and symmetry breaking and the temperature region of sharpest change. Thus, it is possible to locate the transition temperature with reasonable accuracy. (b) Strong first-order transitions can be easily distinguished from second-order transitions by the evolution of the principal component distribution. (c) A crossover can be distinguished from a phase transition. (d) Principal components are related to Fourier modes and order parameters, and their size dependence can be used to extract critical exponents. (e) The principal components are particularly interesting in the fully frustrated triangular antiferromagnet, where they point to incipient orders upon perturbation that are not evident at first glance. (f) Finally, the analysis is sensitive to the nature of the data provided to the machine-learning algorithms. The use of bare spin configurations versus bond, triangle, or square-plaquette spin, or local vorticity configurations, provides complementary inferences, and they are not readily discovered from the other.

In the conclusion, we will discuss why automated machine learning can be an important part of our toolbox for studying phase transitions using large-scale numerical simulations.

II. MODELS

In this paper, we investigate several classical models of phase transitions. The simplest is the two-dimensional square-lattice Ising model:

$$H = -J \sum_{\langle i,j \rangle} S_i S_j, \tag{1}$$

where the "spin" $S_i = \pm 1$. Positive J corresponds to ferromagnetism, where the energy favors aligned spins, while for negative J antiferromagnetic configurations have lower energy. On a bipartite lattice with nearest-neighbor interactions, the thermodynamics of $J > 0$ and $J < 0$ are identical, and, for the square lattice, a phase transition to a magnetically ordered state occurs as the temperature T is reduced below the critical value $T_c/J = 2/\ln(1 + \sqrt{2}) \approx 2.269$ [46]. We also study the antiferromagnetic triangular lattice Ising model (TLIM) as an example of a fully frustrated model with no finite-temperature phase transition and macroscopic ground-state degeneracy.

The Blume-Capel model (BCM) is a generalization of Eq. (1), which allows three values $S_i = \pm 1,0$ with associated energy,

$$H = -J \sum_{\langle i,j \rangle} S_i S_j + \Delta \sum_i S_i^2. \tag{2}$$

Again, we choose the coupling $J = +1$ between nearest-neighbor spins on a square lattice with periodic boundary conditions. Qualitatively speaking, the BCM allows "site vacancies" ($S_i = 0$) in addition to nonzero magnetic moments ($S_i = \pm 1$). The parameter Δ of Eq. (2) controls the vacancy density. For $\Delta \to -\infty$, $S_i = 0$ is energetically unfavorable, and the BCM reduces to the Ising model. As Δ increases, the second-order Ising-like transition from ferromagnet to paramagnet becomes first order above a tricritical point $(T/J, \Delta/J) = [0.609(4), 1.965(5)]$ [47] and occurs with a discontinuous jump in magnetization and vacancy density. The BCM provides a description of systems ranging from metamagnets and ternary alloys to multicomponent spins and the dynamics of rough surfaces [41,48–52]. It has also been studied for nonuniform Δ_i [53] to understand situations in which the spin density varies across the lattice.

The BCM was further generalized by Blume, Emery, and Griffiths to

$$H = -J \sum_{\langle i,j \rangle} S_i S_j + K \sum_{\langle i,j \rangle} S_i^2 S_j^2 + \sum_i \Delta_i S_i^2 \tag{3}$$

to incorporate an additional biquadratic interaction $K \sum_{\langle i,j \rangle} S_i^2 S_j^2$, and it was used to study ^{3}He-^{4}He mixtures [41]. A variant of Eq. (3), which we call the biquadratic-exchange spin-1 Ising (BSI) model,

$$H = -J \sum_{\langle\langle i,k \rangle\rangle} S_i S_k + K \sum_{\langle i,j \rangle} S_i^2 S_j^2, \tag{4}$$

will be studied here. As with the BCM and BEG model, S_i takes values $+1$, 0, or -1. Our primary interest is in the purely biquadratic limit with $J = 0$. Because it has been less well studied, we discuss in this section some of the basic physics of the BSI model revealed by conventional Monte Carlo before turning to the results of machine-learning approaches.

Since $K > 0$, it is energetically unfavorable to have adjacent sites of the BSI model both "occupied" with $S_i = \pm 1$. Therefore, as T is lowered, one expects all occupied sites to be surrounded by vacancies. As long as $J = 0$, there is no preference for one $S_i = \pm 1$ orientation over the other. If one sublattice is occupied with $S_i = 0$, the other sublattice is free to take any spin value. Thus the $J = 0$ BSI model has macroscopic ground-state entropy. A question arises: Is there a phase transition to a checkerboard "charge-ordered" phase where the $S_i = \pm 1$ variables preferentially occupy one sublattice? Thus, along with the triangular antiferromagnet, this model provides machine learning with the challenge of dealing with high degeneracy and deciding between the presence and absence of a phase transition.

Simulations of this $J = 0$ case [Figs. 1(a)–1(c)] can be contrasted with a second-neighbor bilinear exchange interaction $J = 0.1$, where conventional Monte Carlo analysis reveals a sharp second-order phase transition at temperature $T/K = 0.163$. Figures 1(d)–1(f) show spin configuration snapshots that go from the ordered ferromagnetic phase at $T = 0.1$ to an intermediate degenerate regime at $T = 0.3$, to a high-temperature phase at $T = 2.0$.

Finally, the XY model in two dimensions,

$$H = -J \sum_{\langle ij \rangle} \cos(\theta_i - \theta_j), \tag{5}$$

in which sites are occupied by a continuous spin $\theta_i \in (-\pi, \pi]$, offers a situation in which the performance of machine learning for a Kosterlitz-Thouless (KT) [44,45] phase transition between exponentially decaying spin correlations and a line of critical points with power-law decaying correlations can be evaluated when provided with various sorts of real-space snapshots—the spin directions themselves or measures of local vorticity.

III. METHODS OF MACHINE LEARNING

Broadly speaking, machine learning can be divided into "unsupervised" and "supervised" approaches. In the latter case, a neural network, for example, can be "trained" by providing a set of patterns along with a desired "answer" (e.g., in the case of statistical mechanics whether the pattern is ordered or not). Thus the trained (supervised) network is able to "generalize" and distinguish order from disorder on patterns to which it had not been previously exposed. In the former case, patterns are given to the analysis procedure, but no "answers" are provided. In this paper, we utilize two methods for unsupervised machine learning, namely "principal component analysis" and the "autoencoder" approach. These two approaches have been described in [54–59]. We provide a brief review here, with further details given in the Appendix.

A. Principal component analysis

Principal component analysis (PCA) [54–56] is a strikingly simple method. In its implementation for our statistical models, one constructs a matrix **S**, each row of which is a snapshot of the instantaneous values of the degrees of freedom from a Monte Carlo simulation, e.g., a listing of spins on each lattice site. We denote by M the total number of such configurations,
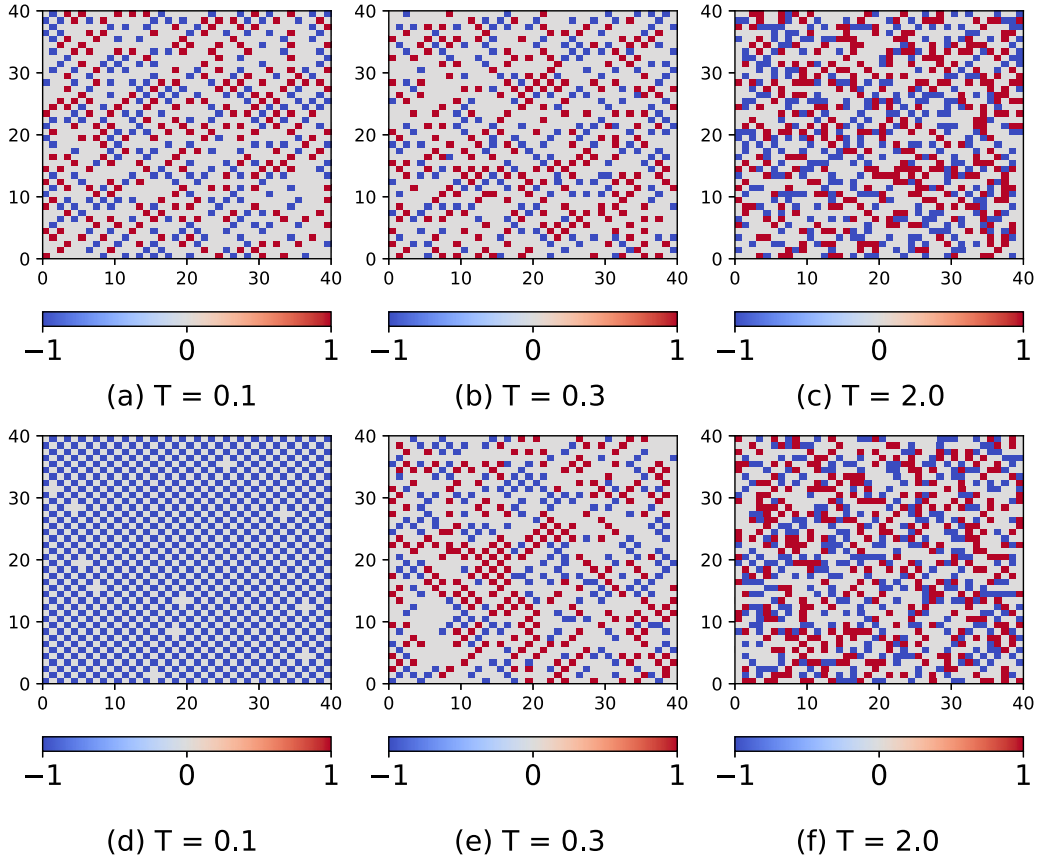
FIG. 1. Top row: Spin configuration snapshots from conventional Metropolis Monte Carlo simulations of the BSI model with fixed $K = 1$, $J = 0$ and linear lattice size $L = 40$. At high temperatures there is a random mix of the three spin possibilities $S_i = 0, \pm 1$. As T decreases below K, sites occupied by $S_i = \pm 1$ are gradually surrounded by near-neighbor vacancy sites $S_i = 0$. Bottom row: Similar spin configuration snapshots with $J = 0.1$. The evolution from $T = 2.0$ (f) to $T = 0.3$ (e) is similar to that at $J = 0.0$ [(c)–(b)]. However, in this case an ordered phase emerges (d) at the lowest $T = 0.1$. In all BSI calculations reported in this paper, K is fixed to 1 as the energy scale.

and therefore the number of rows of **S**. It is convenient to compute the mean value $m_j = (1/M) \sum_i S_{ij}$ of each column and subtract that value from the entries in the column to obtain the "centered data matrix" **X** with "columnwise zero empirical mean." To study phase transitions using PCA, we choose t evenly separated temperatures (or some other parameter), and for each, we generate n uncorrelated spin configurations using a Monte Carlo simulation. The dimension of **X** is $M = nt$.

PCA then extracts features from the data in **X** by performing an orthogonal transformation. The result is the conversion of a set of configurations of possibly correlated variables into a set of values of linearly uncorrelated variables, named principal components, such that the greatest variance of the data comes to lie on the first principal component, the second greatest variance on the second principal component, and so on [55]. Each succeeding component in turn has the largest variance possible under the constraint that it is orthogonal to the preceding components, and the resulting vectors are an uncorrelated orthogonal basis set.

To put this more formally, given the centered data matrix **X**, the transformation is defined by iteratively constructing a set of N-dimensional "weight vectors." The first is found by

$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}\|=1} \left\{ \sum_i (\mathbf{x}_i \cdot \mathbf{w})^2 \right\}, \tag{6}$$

where the arguments of the maxima (abbreviated as arg max) are the points of the domain of some function at which the function values are maximized. Succeeding weight vectors are obtained by subtracting the already calculated principal components from **X** and repeating Eq. (6). The full principal component decomposition of **S** can therefore be given as $\mathbf{P} = \mathbf{S}\mathbf{W}$, where $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$.

The inner products of the spin configuration $\mathbf{S}_i$, with the weight vectors $\mathbf{w}_j$, are termed its components p_{ij},

$$p_{ij} = \mathbf{S}_i \cdot \mathbf{w}_j. \tag{7}$$

The "quantified principal components" are defined as the averages,

$$\langle |p_j| \rangle = \frac{1}{n} \sum_i |p_{ij}|, \tag{8}$$

where the average is taken over some appropriate subset of all samples. In our statistical mechanics application of PCA, it is natural to do the quantification by averaging over all n spin configurations with the same temperature. The central power of PCA is in cases in which the nature of the system is well described by a small number of "principal components," e.g., p_{i1} and p_{i2} only, and their associated averages.

The principal component transformation can, equivalently, be thought of in terms of another matrix factorization method,
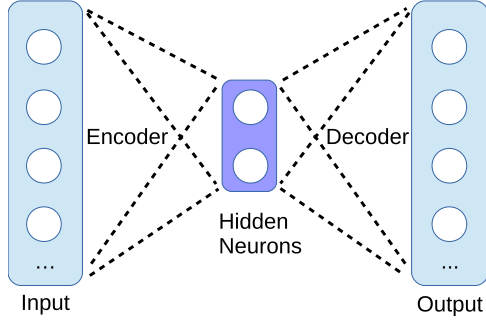
FIG. 2. Schematic structure of an autoencoder with two hidden neurons. CNNs are treated as the encoder and the decoder in this work.

the singular value decomposition (SVD). It can be shown that the PCA weight vectors are eigenvectors of the $N \times N$ real symmetric matrix $\mathbf{X}^T \mathbf{X}$, determined by [56,60]

$$\mathbf{X}^T \mathbf{X} \mathbf{w}_n = \lambda_n \mathbf{w}_n. \tag{9}$$

Eigenvalues returned from Eq. (9), sorted in descending order, $\lambda_1 \geqslant \lambda_2 \cdots \geqslant \lambda_N \geqslant 0$, represent variances of the input matrix $\mathbf{X}$ along corresponding weight vectors. It is convenient to denote the normalized eigenvalues $\tilde{\lambda}_n = \lambda_n / \sum_{i=1}^{N} \lambda_i$ as relative variances.

B. The autoencoder approach

PCA has been shown to be effective in extracting aspects of phase transitions [16,61]. As discussed above, however, PCA is based on linear transformations of the input data. Thus, it is natural to explore other, nonlinear transformation, methods.

In this spirit, we have also examined the use of an "autoencoder" [57–59], i.e., an artificial neural network used for unsupervised learning of efficient encodings by exploiting the discovery of compressed representations. The basic idea is to pass information, in our case spin configurations, through an intermediate layer whose number of "neurons" is considerably smaller than the number of bits needed to encode the original, full spin configuration. The goal is to reproduce the original spin configuration in the best way possible despite the constraint imposed by the reduced information storage capacity of the hidden layer. By attempting this accurate reconstruction, the network may discover important structures and patterns in the input data.

A schematic autoencoder structure with two hidden neurons is shown in Fig. 2. Although the autoencoder and PCA both use the idea of dimensional reduction [in the case of the PCA by focusing only on the one or two largest ("principal") components of the decomposition in Eq. (9)], the autoencoder approach is generally more powerful and more flexible than PCA. Rather than being restricted to linear transformations of the input data into principal components, the autoencoder can incorporate nonlinear representations. Given a set of spin configurations $\{\mathbf{S}_i\}$, in an autoencoder the weights in the neural network can be trained through, for example, backpropagation [1,62] to return target values equal to the inputs. In this work, we choose convolutional neural networks (CNNs) [23,63,64] as the encoder and the decoder networks. As an alternative,
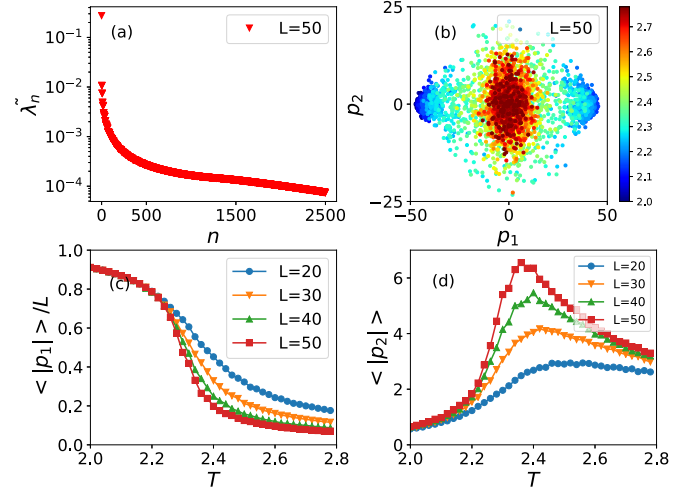

FIG. 3. PCA results for the 2D square-lattice Ising model. (a) Relative variances $\tilde{\lambda}_n$ obtained from the raw Ising configurations, with the horizontal axis indicating corresponding component labels. (b) Projection of the raw Ising configurations onto the plane of the two leading principal components. The color bar on the right indicates the temperature in units of J. (c) The normalized quantified first leading component as a function of temperature. (d) The quantified second leading component as a function of temperature.

artificial neural-network-based variational autoencoders can also be applied to tackle the problem, as reported in Ref. [30].

IV. RESULTS AND ANALYSIS

A. PCA results of the Ising model

By feeding a set of raw Ising spin configurations $\{\mathbf{S}_i\}$ into PCA, we generate results for the square-lattice Ising model, shown in Fig. 3. We have chosen $t = 40$ temperatures in the range from 2.0 to 2.8 with $\Delta T = 0.02$. These bracket the critical temperature $T_c \approx 2.269$. For each temperature, we generate $n = 10\,000$ uncorrelated spin configurations, so that $M = 400\,000$ configurations are used for calculating weight vectors and relative variances. In Fig. 3(a), a single dominant principal component is observed; the largest λ_k, obtained through Eq. (9), is more than an order of magnitude greater than the next largest. As we shall argue further below, the physical implication is that there is a single dominant spin pattern for the Ising model, corresponding to alignment of all spins (i.e., ferromagnetism).

Projecting raw spin configurations into the 2D plane spanned by the first two principal components p_{i1} and p_{i2}, we obtain Fig. 3(b). For clarity of presentation, only 100 scatter points for each temperature are shown. We see that above the critical temperature, the points lie in a single blob at the origin, but that below T_c a separation arises into two distinct regions.

In Fig. 3(c), we plot the quantified first leading component $\langle |p_1| \rangle / L$ versus temperature, normalized to the system size L. This is seen to mimic the behavior associated with a measurement of the (absolute value of) magnetization from a conventional Monte Carlo analysis. Figure 3(d) displays the quantified second leading component $\langle |p_2| \rangle$ versus
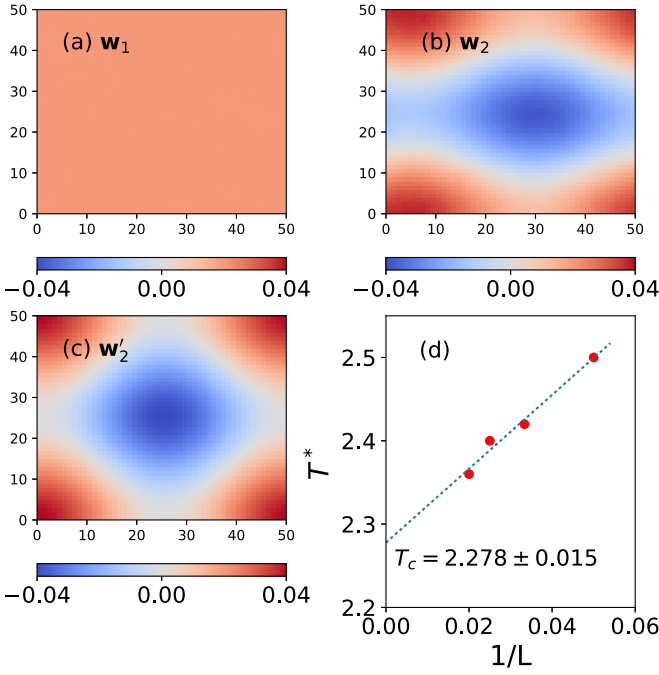
FIG. 4. (a) Visualization of the weight vector corresponding to the first leading component. (b) Visualization of the weight vector corresponding to the second leading component. (c) The vector $\mathbf{w}_2'$ defined by Eq. (10). In (a)–(c) the lattice size $L = 50$. (d) Peaks (T^*) returned from the quantified second leading component as a function of the inverse of the lattice dimension ($1/L$). The extrapolation was performed using a linear least-squares fit.

temperature. Its development is reminiscent of that of the susceptibility χ.

The weight vector corresponding to the first leading component is shown in Fig. 4(a). Ignoring statistical and numerical variations, this weight vector is seen to be roughly constant: $\mathbf{w}_1 = \frac{1}{L}[1, \ldots, 1]$. Since the first principal component p_{i1} is the inner product of $\mathbf{w}_1$ with $\mathbf{S}_i$ [Eq. (6)], the connection to magnetization is clear. The ability of an automatic machine learning algorithm to recognize this order parameter is not surprising as the order parameter of the model is so evident. We have checked that studying the variation of this with size at the critical temperature T_c can be related to known critical exponents for this 2D Ising system. This is also not a surprise given that this principal component is just the order parameter of the system.

Similarly, the weight vector corresponding to the second leading component is shown in Fig. 4(b). To understand the physical meaning of this weight vector, we compare it in Fig. 4(c) with

$$\mathbf{w}_2' = \frac{1}{L}[\cos(\mathbf{r}_1\mathbf{k}_1), \ldots, \cos(\mathbf{r}_N\mathbf{k}_1)]$$
$$+ \frac{1}{L}[\cos(\mathbf{r}_1\mathbf{k}_2), \ldots, \cos(\mathbf{r}_N\mathbf{k}_2)], \quad (10)$$

where $\mathbf{r}_i$ are the lattice sites, and $\mathbf{k}_1 = (0, \frac{2\pi}{L})$ and $\mathbf{k}_2 = (\frac{2\pi}{L}, 0)$ are the two Fourier wave vectors closest to the origin, $\mathbf{k}_0 = (0,0)$. There is a clear similarity between Figs. 4(b) and 4(c).
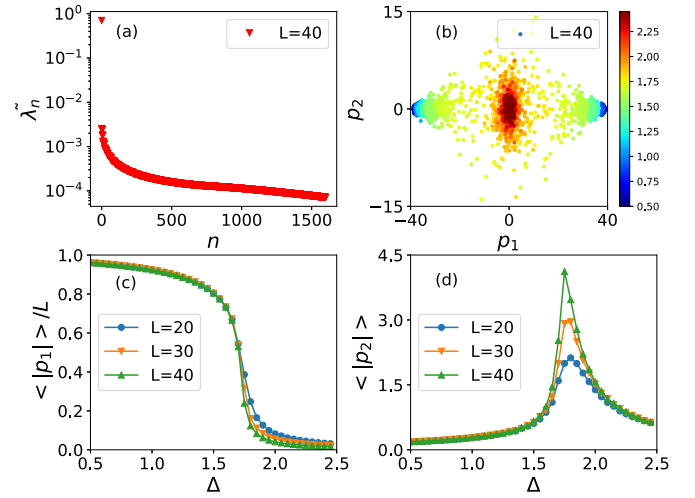

FIG. 5. PCA results for the BCM at fixed $J = 1.0, T = 1.0$ and sweeping Δ. (a) Relative variances obtained from the raw Blume-Capel configurations, with the horizontal axis indicating the corresponding component label. (b) Projection of the raw Blume-Capel configurations onto the plane of the leading two principal components. The color bar on the right indicates Δ in units of J. (c) The normalized quantified first leading component as a function of Δ. (d) The quantified second leading component as a function Δ. Peaks provide estimates of the transition point Δ_c.

We conclude that, for the ferromagnetic Ising model, PCA is building up weight vectors corresponding to the Fourier modes of the spin configuration. In its ordered phase, the physics of the Ising model is dominated by the single point $\mathbf{k}_0 = (0,0)$, and hence PCA reveals a single dominant eigenvalue. Subleading eigenvalues are associated with the next most ordered arrangements, i.e., with a single horizontal or vertical domain wall. When supplied with configurations of the ferromagnetic Ising model in the zero magnetization ensemble, PCA generates four large eigenvalues corresponding to horizontal or vertical domain walls [16].

Finally, Fig. 4(d) shows the peaks (T^*) returned from the quantified second leading component versus the inverse of the lattice dimension ($1/L$). A linear finite-size scaling fit to these peaks yields $T_c \sim 2.278 \pm 0.015$, which agrees reasonably well with the exact result $T_c/J \approx 2.269$.

B. PCA results of the Blume-Capel model

Following the same procedure, we feed a set of raw Blume-Capel spin configurations $\{\mathbf{S}_i\}$ into the PCA to generate the results shown in Figs. 5 and 6. For both plots, we choose $t = 40$ values of Δ and $n = 10\,000$ uncorrelated spin configurations for each Δ. As noted earlier, the phase diagram of the BCM differs from that of the Ising model in one important respect. It possesses both first- and second-order transitions, separated by a tricritical point at $(T/J, \Delta/J) = [0.609(4), 1.965(5)]$. In Fig. 5, the PCA was supplied with a set of spin configurations in a sweep of Δ in the range $0.5 < \Delta < 2.5$ at fixed $T = 1.0$, thus crossing the phase boundary in a second-order transition. Similarly, in Fig. 6 the sweep covers the range $1.0 < \Delta < 3.0$ at fixed $T = 0.4$, and it crosses the phase boundary in a first-order transition.
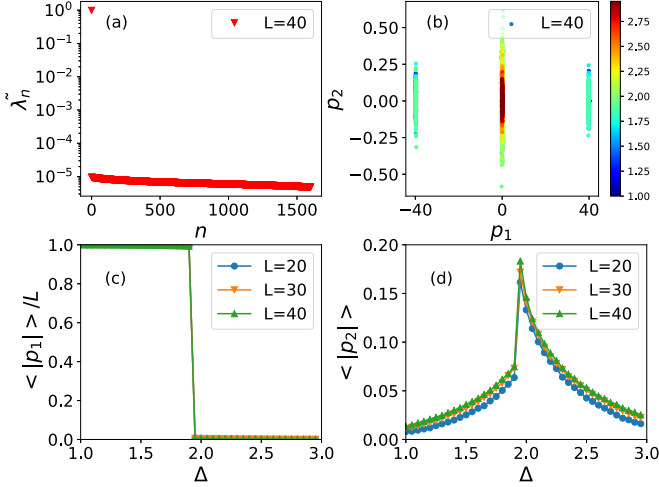
FIG. 6. PCA results for the BCM at fixed $J = 1.0, T = 0.4$. (a) Relative variances. (b) Projection of the raw Blume-Capel configurations onto the plane of two leading principal components. The color bar on the right indicates Δ in units of J. (c) The normalized quantified first leading component as a function of Δ. (d) The quantified second leading component as a function of Δ. Peaks provide estimates of the transition value Δ_c.

The panels of Fig. 5 are very similar to that of the Ising model, Fig. 3. Using the peaks returned from Fig. 5(d) and finite-size scaling, we can locate $\Delta_c \approx 1.70 \pm 0.01$ in the thermodynamic limit. This is close to the value $\Delta_c \approx 1.63$ at $T = 1.0$ reported in [53]. Of course, to locate the critical point more accurately, one could refine the Δ range and add more samples at each temperature, much as improvements in critical values are obtained with conventional analysis.

At fixed $T = 0.4$, the BCM undergoes a first-order transition at $\Delta_c \approx 1.996$ [65]. The corresponding PCA data are shown in Fig. 6. These first-order phase-transition results can be easily distinguished from those at the second-order transition, which were seen in Figs. 3(b) and 5(b) to spread out uniformly from the origin. In stark contrast, the scatter points in Fig. 6(b) spread out mostly in the second principal component, and there are few intermediate points between two phases. The latter observation is of course expected from a first-order transition. We see that PCA clearly behaves differently across first- and second-order transitions.

C. Monte Carlo and PCA results of the BSI model

To explore the thermodynamic properties of the BSI model beyond the preliminary results presented in Fig. 1, we first study its energy, specific heat, and the "checkerboard" structure factor, which are shown in Figs. 7(a), 7(b) and 7(c), respectively. Their definitions are as follows:

$$\langle E \rangle = (1/N') \sum_{\sigma_i} (E_{\sigma_i}/N),$$

$$\langle C \rangle = N \frac{\langle E^2 \rangle - \langle E \rangle^2}{T^2}, \quad (11)$$

$$\langle F \rangle = (1/N) \sum_{\mathbf{i,j}} e^{i\mathbf{Q}(\mathbf{i-j})} \left(\langle S_{\mathbf{i}}^2 S_{\mathbf{j}}^2 \rangle - \langle S_{\mathbf{i}}^2 \rangle \langle S_{\mathbf{j}}^2 \rangle \right),$$
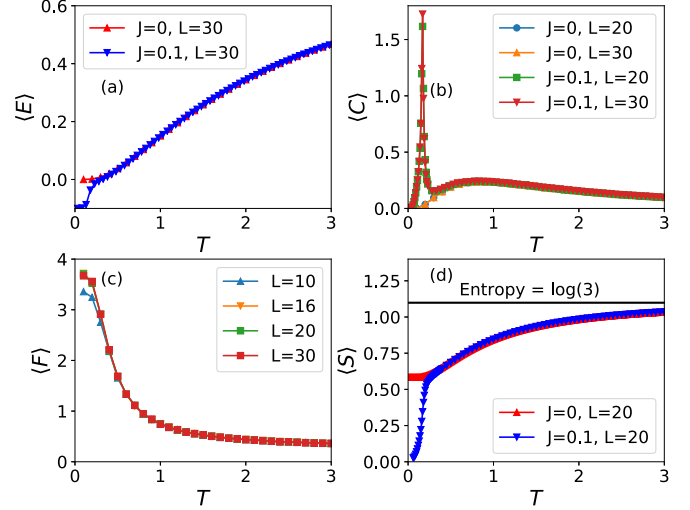

FIG. 7. (a) Energy as a function of temperature with (blue) or without (red) the J term. (b) Specific heat as a function of temperature with (green, brown) or without (blue, yellow) the J term. Finite-size effect is verified to be negligible. Sharp peaks agree well with the Binder ratio crossing in Fig. 8(c). (c) Structure factor as a function of temperature with different system sizes. J is fixed to 0. (d) Entropy as a function of temperature with (blue) or without (red) the J term. Ground-state entropy for the red curve is 0.584.

where σ_i is the spin configuration from the Monte Carlo simulation, E_{σ_i} is the total energy of the spin configuration σ_i, N' is the total number of spin configurations used for calculation, and $\mathbf{Q}$ is the ordering wave vector (π, π)

With only the biquadratic term, the specific heat of the BSI model has a peak at $T \approx 0.8$, however the intensity of the peak does not increase with system size, indicative of only short-range order. The structure factor is shown in Fig. 7(c). It shows a sharp increase around the same temperature as the specific-heat peak. However, once again, the peak does not grow with system size, indicating an absence of a phase transition for the $J = 0$ BSI model. If we turn on the next-nearest-neighbor bilinear interaction ($J = 0.1$), we can see a clear sharp peak in the specific heat at $T \approx 0.17$, which grows with system size, indicating a true phase transition. This peak is very sharp because the strength of the bilinear exchange interaction $J = 0.1$ is small compared with the leading energy scale $K = 1$.

For frustrated systems, an important quantity is the entropy of the system. In Fig. 7(d), we show the entropy, calculated from

$$S(T_b) - S(T_a) = \int_{T_a}^{T_b} \frac{C(T')}{T'} dT'$$

$$= \frac{E(T_b)}{T_b} - \frac{E(T_a)}{T_a} + \int_{T_a}^{T_b} \frac{E(T')}{T'^2} dT'. \quad (12)$$

Replacing E with the Monte Carlo average $\langle E \rangle$ and approximating $S(T_b)$ with $S(+\infty) - c/T_b^2$, where c is a fitting
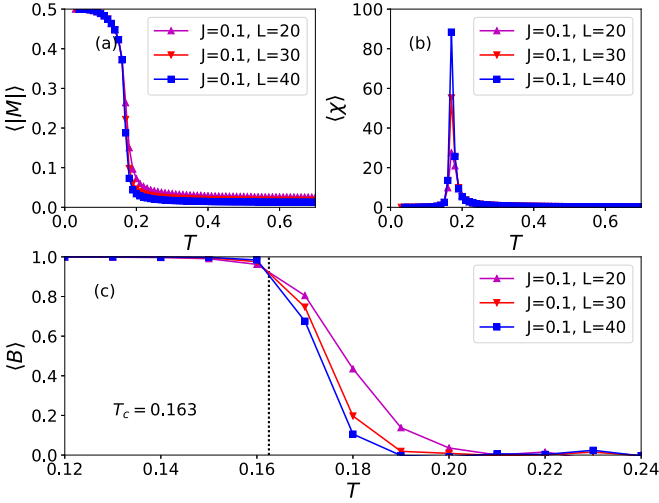
FIG. 8. Monte Carlo results for the BSI model with fixed $K = 1.0$, $J = 0.1$. (a) Absolute magnetization as a function of temperature. (b) Susceptibility as a function of temperature. Sharp peaks agree well with the Binder ratio crossing. (c) Binder ratio as a function of temperature with varied system sizes. The critical temperature is $T_c = 0.163$.

constant, we can estimate the low-temperature entropy with

$$\langle S(T) \rangle = \langle S(+\infty) \rangle - \frac{c}{T_b^2}$$
$$- \frac{\langle E(T_b) \rangle}{T_b} + \frac{\langle E(T_a) \rangle}{T_a} - \int_{T_a}^{T_b} \frac{\langle E(T') \rangle}{T'^2} dT'. \quad (13)$$

With only the biquadratic term, the system has a nonzero ground-state entropy of approximately 0.584, which clearly exceeds $\ln 3/2 \approx 0.5493$ obtained from fixing one sublattice to have $S_i = 0$ and allowing spins to take arbitrary values on the other sublattice. With a nonzero $J = 0.1$, below T_c the entropy quickly drops to zero. For the Ising antiferromagnet on a triangular lattice, we have checked that we obtain the correct ground-state entropy [36–38], with the value 0.323, as reported in Ref. [66].

To locate the critical temperature accurately in the latter case, we calculate the absolute magnetization $\langle |M| \rangle$, the susceptibility $\langle \chi \rangle$, and the Binder ratio [67] $\langle B \rangle$, which are shown in Fig. 8 and defined as

$$\langle |M| \rangle = (1/N') \sum_{\sigma_i} \left(\left| \sum_j s_j \right| \Big/ N \right),$$
$$\langle \chi \rangle = N \frac{\langle M^2 \rangle - \langle |M| \rangle^2}{T}, \quad (14)$$
$$\langle B \rangle = \frac{1}{2} \left(3 - \frac{\langle M^4 \rangle}{\langle M^2 \rangle} \right),$$

where N' is the total number of spin configurations used for the calculation. Similar to $\langle C \rangle$, both $\langle |M| \rangle$ and $\langle \chi \rangle$ have small finite-size effects, and their transition signals agree well with the peak of $\langle C \rangle$. In Fig. 8(c), the Binder ratio plot shows a crossing at $T_c = 0.163$, which gives the critical point for the $J = 0.1$ BSI model.

Next, we apply the PCA method to the BSI model and generate results as shown in Fig. 9. In the top row of Fig. 9, we display PCA results for the $J = 0$ case. We first feed raw spin configurations $\{\mathbf{S}_i\}$ into PCA, however no dominant principal components are observed (see the red symbols). Projecting raw spin configurations onto the leading principal component plane, we observe a complete mixture of points of all temperatures. This is one of the failures of an automated machine-learning algorithm, which we will see again in the study of the XY model. The physics of the problem relates to how the $S_i = \pm 1$ spin variables are arranged with respect to the $S_i = 0$ variables, not with respect to each other. However, PCA averages these ± 1 variables out to zero and fails to recognize any emergent order.

However, if we feed the squared spin configurations $\{\mathbf{S}_i^2\}$ into PCA, we can get dominant principal components (blue symbols). If we further project squared spin configurations into two leading principal components, as shown in Fig. 9(c), we are able to distinguish the low-temperature phase with the high-temperature phase. Actually, the first leading principal component is just the local "charge" order parameter, where we take $s_i = +1$ and -1 as "charge" 1 and $s_i = 0$ as "charge" 0. The charge density changes from $2/3$ at high temperatures to below 0.5 at low temperatures. PCA shows that this is a greater discriminant of different temperatures than the true emergent charge order in the checkerboard pattern. The checkerboard order forms only the second principal component. Although it has a larger relative change with temperature, in magnitude it is smaller than the extensive change in the total number of charges present.

The scatter points in Fig. 9(c) are different from the case of the first-order phase transition in Fig. 6(b) or the second-order phase transition in Fig. 5(b). The most evident difference is the absence of symmetry breaking, which in Figs. 6(b) and 5(b) shows up by the bifurcation of scatter points for the low-temperature data at left and right ends. This is a clear signature of a gradual crossover to low-temperature behavior and an absence of a phase transition.

Turning on the bilinear exchange interaction, by feeding either the raw spin configurations $\{\mathbf{S}_i\}$ or the squared spin configurations $\{\mathbf{S}_i^2\}$ into PCA, we observe dominant principal components, as shown in Fig. 9(d). There is only one dominant relative variance returned from $\{\mathbf{S}_i^2\}$ but two equally weighted dominant relative variances returned from $\{\mathbf{S}_i\}$. In Fig. 9(e), projecting raw spin configurations onto the plane of two leading principal components p_{i1} and p_{i2}, we see four distinct branches extending from the center, and interestingly, all of them correspond to temperatures lower than T_c. In this model, there exist four distinct ground states, which are shown in Fig. 10, and each distinct ground state is associated with a branch in Fig. 9(e). Thus we see that the number of dominant relative variances N_v is related to the number of ground states N_g by $N_g = 2^{N_v}$.

On the other hand, by feeding $\{\mathbf{S}_i^2\}$ into PCA and projecting data onto two leading principal components, we generate Fig. 9(f). The original four branches become two, because replacing s_i with s_i^2, will only leave us with two distinct ground states corresponding to the charge order. In this case, it is evident that s_i^2 cannot capture the full phase behavior of the model.
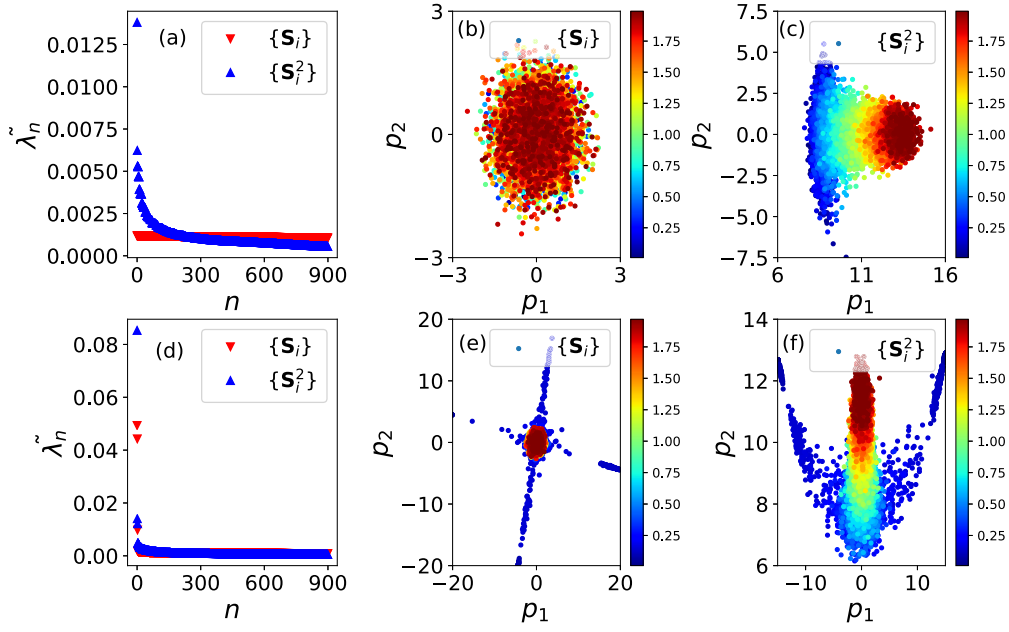
FIG. 9. PCA results for the BSI model at lattice size $L = 30$. Color bars of scatter plots indicate the temperature of the samples in units of K. Top row: The BSI model with fixed $K = 1.0$, $J = 0$. (a) Relative variances obtained from the raw BSI configurations $\{\mathbf{S}_i\}$ (red) and the squared spin configurations $\{\mathbf{S}_i^2\}$ (blue). (b) Projection of the raw BSI configurations $\{\mathbf{S}_i\}$ onto the plane of the leading two principal components. (c) Projection of the squared spin configurations $\{\mathbf{S}_i^2\}$ onto the plane of the leading two principal components. Bottom row: The BSI model with fixed $K = 1.0$, $J = 0.1$. (d), (e), and (f) follow the same definitions as (a), (b), and (c).

D. PCA results of the TLIM

We now turn to the PCA of the fully frustrated anti-ferromagnetic triangular-lattice Ising model (TLIM). This is a particularly interesting case as there is no long-range order in the model all the way down to $T = 0$. Instead, power-law spin correlations develop as temperature is lowered [37,38]. We feed raw spin configurations $\{\mathbf{S}_i\}$ into PCA. We observe two equally weighted relative variances, as shown in Fig. 11(a). If we project them into the leading two principal components, we see a separation of high-temperature scatter points, forming a central blob, from the low-temperature scatter points, forming the outside periphery of the circle, as shown in Fig. 11(b). Figures 11(c) and 11(d) show the growth of the projected principal components as the temperature is lowered. Figures 11(e) and 11(f) show the spatial patterns in the projected components.

It is interesting to observe that the two principal components correspond to patterns of ordering in which the three sublattices of the triangular-lattice form roughly an $(m, 0, -m)$ and an $(m, -m/2, -m/2)$ pattern, respectively. PCA shows an emergent continuous XY-type symmetry between these patterns. It is known [68,69] that these are the two incipient orders of the triangular antiferromagnet when either a weak transverse field is added to the system or if triangular-Ising layers are stacked on top of each other with a weak interlayer coupling. These two patterns are nearly degenerate. Within the Ginzburg-Landau approach, the degeneracy is only lifted by an irrelevant sixth-order anisotropy term. It is remarkable that machine learning can automatically point to such incipient order in the Monte Carlo data. Independently recognizing such incipient patterns may be one of the strengths of machine learning.
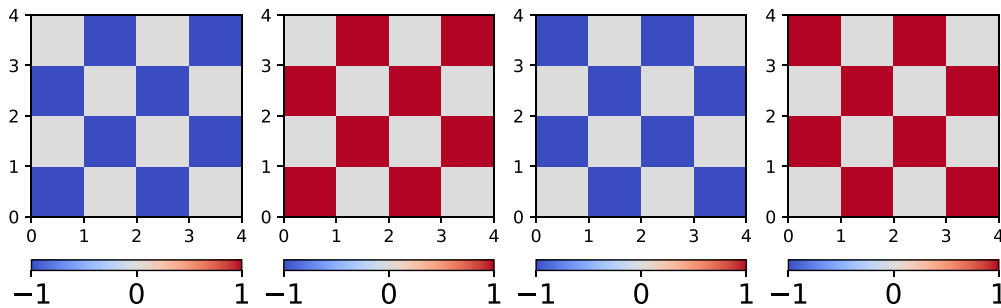

FIG. 10. Four distinct ground states for the BSI model, with fixed $K = 1$, $J = 0.1$, $N = 40 \times 40$. For clarity, we only display a 4×4 block for each ground state.
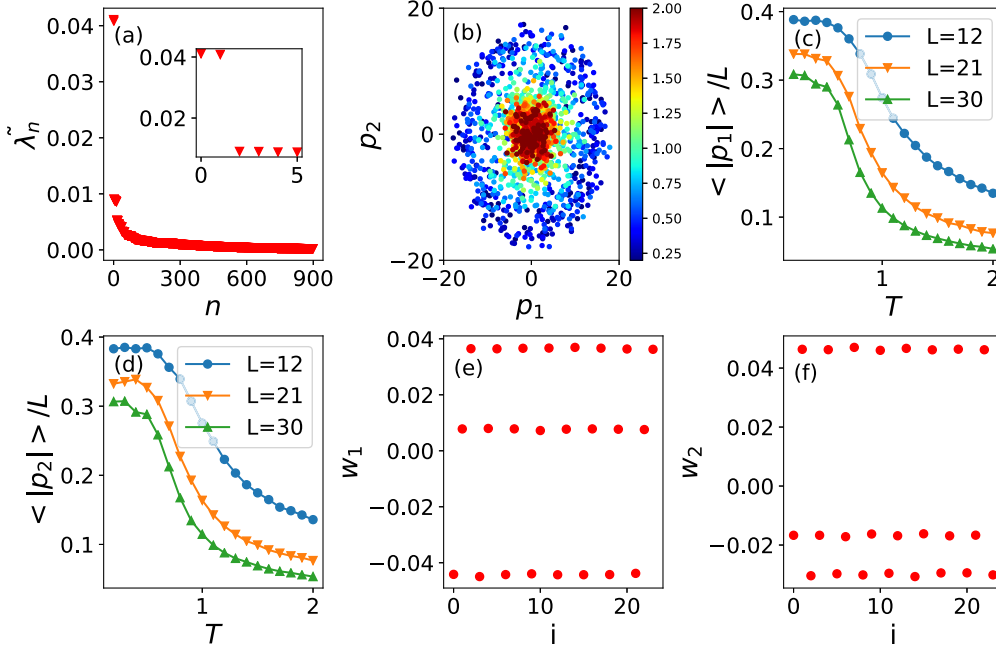
FIG. 11. Top row: PCA results for the antiferromagnetic (fully frustrated) TLIM, with fixed $J = -1$. (a) Relative variances. $L = 30$. (b) Projection onto the plane of the leading two principal components. The color bar indicates the temperature in units of $|J|$. $L = 30$. (c) The normalized quantified first leading component as a function of temperature. (d) The normalized quantified second leading component as a function of temperature. (e) The weight vector corresponding to the first leading component. (f) The weight vector corresponding to the second leading component. See the text for an interpretation.

E. PCA results of the XY model

We now focus on the PCA analysis of the 2D XY model by feeding various sorts of real-space snapshots—the spin directions themselves and measures of local vorticity—into PCA. To express spin directions, we define an angle configuration as $[\cos(\theta_1), \sin(\theta_1), \ldots, \cos(\theta_N), \sin(\theta_N)]$ and a collection of angle configurations simply as $\{\cos(\theta_i), \sin(\theta_i)\}$. By providing these projections of the XY spins along the x and y axes to PCA, we observe two equally weighted relative variances in Fig. 12(a) and concentric circles in Fig. 12(b), which essentially recovers the rotational symmetry of XY spins in the spin space. In Fig. 12(b), we can see that in the high-temperature phase, scatter points aggregate in the center and are separated from the low-temperature scatter points, which spread out in the periphery of the circle. In Fig. 12(c) we show a plot of $\langle \sqrt{p_1^2 + p_2^2} \rangle / L$, which is the order parameter for the system. Also shown in Fig. 12(c) by a dashed vertical line is the known KT transition temperature $T_c = 0.892$ [70]. The data show an abrupt change in the temperature dependence around T_c. Differentiating power-law correlations from long-range order is possible in PCA but will require careful finite-size study of the projected principal components, which will again take us back to conventional analysis.

The KT transition is known to be driven by an unbinding of vortex-antivortex pairs. Since PCA is a linear transformation method, it cannot deduce vortex-related information from the spin configurations themselves. However, we can first preprocess raw spin configurations to generate local vortex and antivortex configurations by the definition

$$\oint_C \Delta\theta \, dl = 2\pi k, \quad k = \pm 1, \pm 2, \pm 3, \ldots, \quad (15)$$

where k is the winding number, and the integral is done for each square plaquette. Explicitly, each spin difference along a plaquette bond is rescaled into the range $(-\pi, \pi]$. The spin differences are then summed over the plaquette to give the "integral" of Eq. (15). The winding number k is nonzero only when the integral is a multiple of 2π. Thus each square plaquette is expressed using its winding number, which is either $+1$, indicating a vortex, 0, indicating no vorticity, or -1, indicating an antivortex. However, when we feed the raw vorticity configurations into PCA, we do not observe any dominant principal components, as shown in Fig. 12(d) (red triangles). This again shows the limitation of the PCA method. If equally weighted positive and negative vorticities exist in one configuration, they will cancel out each other, and PCA fails to capture their proliferation with temperature. This is analogous to the failure of PCA in capturing the "charge" density in the BSI model when feeding $\{S_i\}$.

On the other hand, if we feed absolute vorticity configurations (defined to be $+1$ if there is a vortex or antivortex present and 0 otherwise) into PCA, a clear dominant principal component is observed. See the blue triangles in Fig. 12(d). The scatter points of the leading two principal components are shown in Fig. 12(e), which appears more like a crossover from previous examples than a true phase transition. We believe that seeing the vortex-antivortex unbinding transition in the 2D XY model may be difficult in PCA or any machine-learning technique. The change of the projected principal component with temperature is shown in Fig. 12(f), and it shows the proliferation of vortices with an increase in temperature. The dashed vertical line again shows the known transition temperature and is consistent with the temperature of onset for the proliferation of vortices.
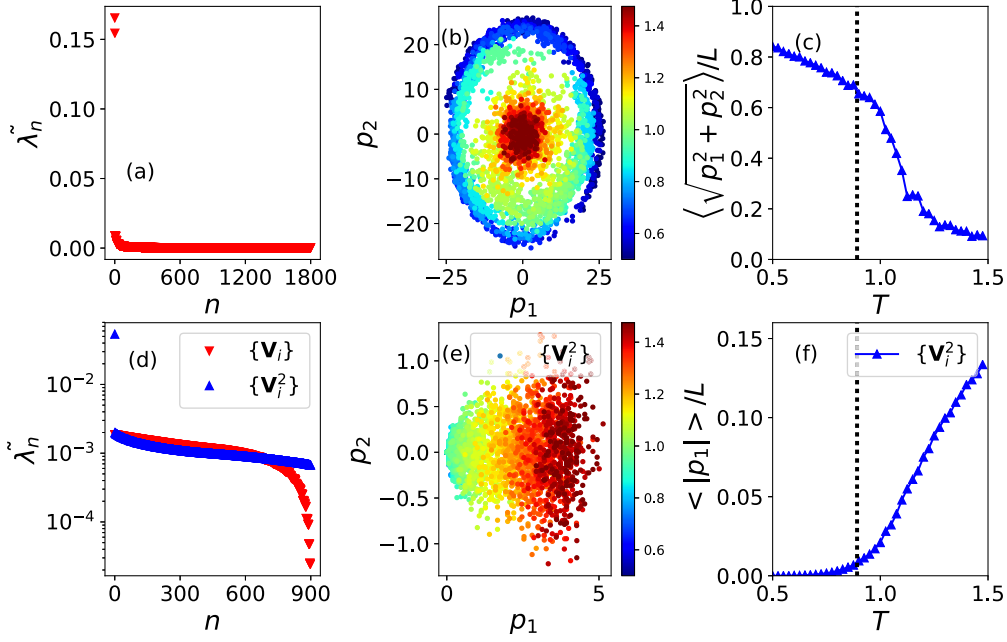
FIG. 12. PCA results for the *XY* model with fixed $J = 1.0$ at $L = 30$. Top row: Feed angle configurations $\{\cos(\theta_i), \sin(\theta_i)\}$ into PCA. (a) Relative variances. (b) Projection onto the plane of the leading two principal components. The color bar indicates the temperature in units of J. Each temperature has 100 scatter points. (c) $\langle\sqrt{p_1^2 + p_2^2}\rangle/L$ vs temperature. The dashed line marks the true T_c. Bottom row: (d) Relative variances. Red (blue) triangles, with the label $\{\mathbf{V}_i\}$ ($\{\mathbf{V}_i^2\}$), indicate results by feeding the raw vorticity (squared vorticity) configurations into PCA. (e) Projection of the squared vorticity configurations onto the leading two principal components. The color bar indicates temperature, and each temperature has 100 scatter points. (f) $\langle|p_1|\rangle/L$ vs temperature. The dashed line marks the true T_c.

A clear message from these studies is that PCA with spatial configurations of local variables is primarily a study of relevant patterns, their symmetry, and symmetry breaking in the model. The discrete symmetry breaking shows up in the twofold or fourfold patterns of a separation of principal component data. Continuous rotational symmetry in the spin or order parameter space is reflected in the circular pattern of scatter points.

F. Autoencoder results of the Ising model

The autoencoder method and architecture are explained in Sec. III B and the Appendix. In Fig. 13(a), we display reconstruction results using 200 hidden neurons. The top row is the raw Ising configurations at several temperatures, and the bottom row is the reconstructed Ising configurations. Although the reconstructed configurations lose detailed spin structures, the network succeeds in capturing essential information of the input data, such as the domain area, the domain position, and also the overall magnetization.

Next, we limit the autoencoder network to only two hidden neurons, and results have been shown in Fig. 13(b). Interestingly, we obtain similar results with the PCA method, as shown in Fig. 3, and the ferromagnetic phase is reasonably distinguished from the high-temperature phase.

Furthermore, we limit the autoencoder network with a single hidden neuron, and we still try to reconstruct original inputs, with results revealed in Fig. 13(c). From the plot, we can clearly observe that the transition temperature is around 2.3, above which the autoencoder network learns a trivial mapping function by projecting all high-temperature configurations into an almost constant. However, when
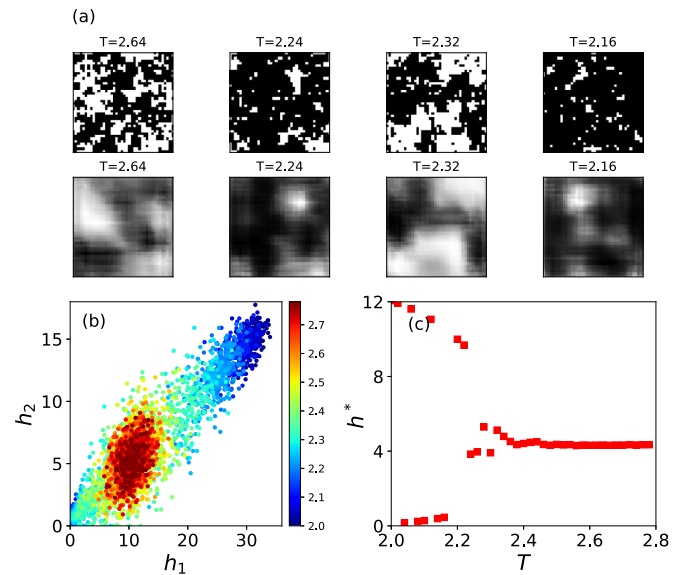

FIG. 13. Ising model autoencoder results. $J = 1.0$ is fixed and the original lattice size is $N = 40 \times 40$. (a) Spin configuration (above) and reconstructions (below) at the indicated temperatures using 200 hidden neurons. (b) Encoding of the raw Ising configurations $\{\mathbf{S}_i\}$ onto the plane of the two hidden neuron activations (h_1, h_2). Color bar indicates the temperature of the sample in units of J. Each temperature has 100 scatter points. (c) Encoding of the raw Ising configurations $\{\mathbf{S}_i\}$ using a single hidden neuron activation (h^*) as a function of temperature. Each data point is averaged over $n = 10\,000$ samples of the same temperature.

the temperature is below the transition temperature, the trained network is able to capture two distinct branches, which actually correspond to two distinct ground states of the Ising model. Thus, even with a single hidden neuron, the autoencoder network can be trained to learn key phase-transition information and help locate the critical point.

V. CONCLUSIONS

In this work, we have employed unsupervised machine learning techniques, such as PCA and autoencoder, to study phase transitions in the Ising, Blume-Capel, BSI, and XY models. One of our key goals was to examine critically what these methods are capturing from the Monte Carlo data and as a result where they succeed and where they fail. In the variety of models we studied here, some had a clearly discernible order parameter. In others it was a more subtle one. Some had no finite-temperature phase transition, just a gradual crossover from uncorrelated high-temperature behavior to strongly correlated low-temperature behavior, some had a power-law ordered phase, and others had a long-range ordered one. Some had a first-order phase transition, others a second-order one. Some had continuous symmetry, others a discrete symmetry. The number of ordered phases varied between models. It is clear from our study that even a straightforward implementation of PCA or the autoencoder-network machine-learning methods can be useful tools for studying phase transitions and can distinguish between many different scenarios.

Much of our focus was on the PCA method. It is clear that the PCA method, when one feeds in spin configurations from Monte Carlo data over a range of parameters, is primarily about recognizing spatial patterns of order and symmetry breaking. This is what discriminates one temperature from another. When we have a rather obvious order parameter, such as in the square-lattice Ising model, the dominant principal component must correspond to this order parameter. The subleading principal components are related to small-q behavior, and hence capture the domain walls and low-energy fluctuations in the system. Careful analysis of the temperature and size dependence of the projections onto these principal components can be used to obtain the transition temperature and the critical exponents, much like in conventional analysis.

In a fully frustrated system, such as the antiferromagnetic triangular-lattice Ising model (TLIM), there is no obvious order parameter, only subtle incipient ones. PCA automatically brings them out. This may be one of the real strengths of automated machine-learning methods. The TLIM has power-law correlations at $T = 0$. Furthermore, it has been shown [68,69] that a stack of triangular-lattice Ising models with arbitrarily weak interlayer coupling, or TLIM in a weak transverse quantum field, is susceptible to order in a three-sublattice pattern in which the three sublattices have $(m, 0, -m)$ or $(m, -m/2, -m/2)$ magnetization. These two patterns of order are nearly degenerate, leading to an emergent XY symmetry that is only broken by a higher-order anisotropy term. PCA picks out these as the principal components and the emergent XY symmetry between them. To fully understand these, one would clearly need further analysis. But, PCA could serve as a useful starting point to the analysis of Monte Carlo data.

PCA can distinguish between first- and second-order transitions and between crossovers and phase transitions. The most clear distinction comes in the way the projections onto the lowest few principal components spread out. The degeneracy or number of PCA components with large comparable variances and the groups in which the scatter points of data arrange themselves very clearly and nicely illustrate the number of ordered phases and the symmetries of the model.

One glaring weakness of PCA is the inability to recognize an order in s_i^2 when data on s_i are fed in. We saw that problem in the BSI model, where short-range order is associated with a checkerboard arrangement of charge (s_i^2), not the original spin (s_i) variables. We encountered the same problem in studying vortex-antivortex pairs in the 2D XY model. The PCA fails to discriminate between configurations where opposite sign charges equal in number are changing substantially with temperature. Only when the squared variables are fed into the PCA algorithms does it recognize charge order or proliferation of vortices. This suggests that it may generally be useful to apply PCA with different powers of the local spin variable to get a more complete picture.

The 2D XY model is particularly challenging for any numerical study. It would be too much to expect an automated machine-learning algorithm to learn about unbinding of vortex-antivortex pairs from the data on raw spin configurations. In the current form, the algorithms cannot even recognize the existence of vortices. Similarly, while we believe the power-law and exponentially decaying phases can be distinguished by studying the size dependence of the projected principal components, accurately locating the phase transition between them would be difficult. Designing machine learning algorithms that can deal with such a situation will be an important achievement.

An interesting question is whether automated machine-learning approaches bear any parallels with the study of entanglement entropies in quantum systems [71,72]. A major theme of condensed-matter physics over the past decade has been the exploration of phases with unusual, or perhaps even nonexistent, local patterns of order. The entanglement entropy [73–75] of quantum lattice models is touted as a quantity that, in principle, catches them all—conventional order, conformal invariance, Goldstone modes, Fermi surfaces, and topological order—although, in practice, it takes substantial effort to address each of these individually. It is clear from our study that the simple PCA, without any preprocessing of data, is more closely tied to conventional order. One can imagine defining a "discriminant entropy" in terms of the relative variances of the PCA method as a general characteristic of the complexity of the model. It remains to be seen if such a properly defined quantity can be helpful in studies of statistical models in an order-parameter independent manner. This may be an interesting avenue for future research.

Finally, we note that because they directly analyze the real-space configurations of the degrees of freedom, machine-learning algorithms might make useful connections to experimental situations in which the order varies over the sample. For example, in studies of cold atomic gases, the confining potential leads to a variation of density and energy scales as one moves outward from the cloud center. As a consequence, metallic, magnetic, and Mott insulating phases can coexist in different regions of a single sample. By allowing the machine-learning algorithm to be exposed to different spatial windows on real-space data obtained from quantum
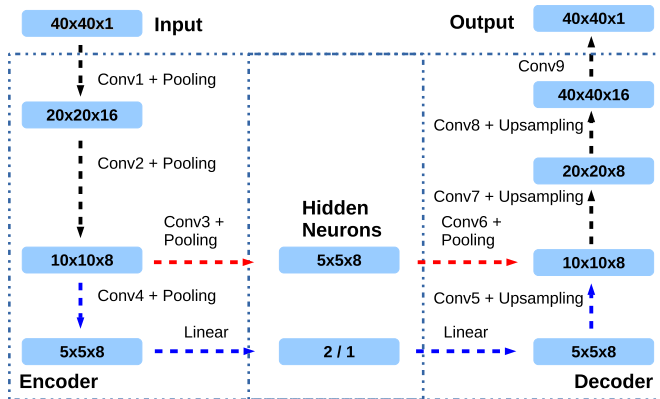
FIG. 14. Autoencoder architecture. To use 200 hidden neurons, we choose to follow the path from black to red to black, and to use 2 or 1 hidden neuron, we choose to follow the path from black to blue to black. We use raw spin configurations as the input data, with fixed size $N = 40 \times 40$.

microscopy [76–78], useful information about this coexistence and domain boundaries may be generated.

APPENDIX: AUTOENCODER ARCHITECTURE

In the preceding discussion, we demonstrated detailed results about the autoencoder method using Keras [79]. In Fig. 14, we show the detailed autoencoder architecture. In the encoder, which is surrounded by the left blue dashed rectangle, the convolutional layer 1, abbreviated as Conv1, uses 16 learnable filters, each of which is an array of weights, of size 3×3 with stride 1 and zero padding of 1. The filter is spatially small, but it extends through the full depth of the input volume. During the forward pass, each filter can move across the width and height of the input volume to compute dot products between the filter and the input. Following the convention, stride 1 stands for moving the filters one pixel at a time, and zero padding indicates padding the input volume with zeros around the border.

Similarly, Conv2, Conv3, and Conv4 use eight filters of size 3×3 with stride 1 and zero padding of 1. All pooling layers are using max pooling, which returns the maximum of the subregion covered by the filter, with filters of size 2×2 and stride 2. In the decoder, surrounded by the right blue dashed rectangle, Conv5, Conv6, and Conv7 use eight filters of size 3×3 with stride 1 and zero padding of 1. Conv8 (Conv9) uses 16 (1) filters of size 3×3 with stride 1 and zero padding of 1. Instead of using pooling layers, in the decoder we apply upsampling layers, which upsamples low-dimensional data to high-dimensional data through deconvolutional filters that can be learned just like normal convolutional kernels, with filters of size 2×2. The rectified linear unit (ReLU) function is activated after every convolutional layer, and the loss function used here is the binary cross entropy loss. More details about the convolutional neural network can be found in Ref. [80]. The network is trained for 100 epochs in each case [Figs. 13(a), 13(b) or 13(c)], and the training data are the raw spin configurations $\{S_i\}$ of the Ising model on a square lattice, with fixed size $N = 40 \times 40$.

[1] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann, Burlington, MA, 2016).

[2] K. Lee, J. Caverlee, and S. Webb, Uncovering social spammers: Social honeypots + machine learning, *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (ACM, Geneva, Switzerland, 2010).

[3] D. Liben-Nowell and J. Kleinberg, The link-prediction problem for social networks, J. Assoc. Inf. Sci. Technol. **58**, 1019 (2007).

[4] C. M. Bishop, Pattern recognition, Machine Learning **128**, 1 (2006).

[5] E. Rosten and T. Drummond, Machine learning for high-speed corner detection, *European Conference on Computer Vision* (Springer, Berlin, 2006).

[6] I. Bose and R. K. Mahapatra, Business data mining—A machine learning perspective, Inf. Manage. **39**, 211 (2001).

[7] R. D. King, S. Muggleton, R. A. Lewis, and M. J. Sternberg, Drug design by machine learning: The use of inductive logic programming to model the structure-activity relationships of trimethoprim analogues binding to dihydrofolate reductase, Proc. Natl. Acad. Sci. USA **89**, 11322 (1992).

[8] B. A. McKinney, D. M. Reif, M. D. Ritchie, and J. H. Moore, Machine learning for detecting gene-gene interactions, Appl. Bioinform. **5**, 77 (2006).

[9] N. P. Schafer, B. L. Kim, W. Zheng, and P. G. Wolynes, Learning to fold proteins using energy landscape theory, Isr. J. Chem. **54**, 1311 (2014).

[10] J. VanderPlas, A. J. Connolly, Ž. Ivezić, and A. Gray, Introduction to astroML: Machine learning for astrophysics, *2012 Conference on Intelligent Data Understanding (CIDU)* (IEEE, Piscataway, NJ, 2012).

[11] J. P. Crutchfield, Between order and chaos, Nat. Phys. **8**, 17 (2012).

[12] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[13] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, Nature Phys. **13**, 431 (2017).

[14] G. Torlai and R. G. Melko, Learning thermodynamics with Boltzmann machines, Phys. Rev. B **94**, 165134 (2016).

[15] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, Machine learning quantum phases of matter beyond the fermion sign problem, arXiv:1608.07848.

[16] L. Wang, Discovering phase transitions with unsupervised learning, Phys. Rev. B **94**, 195105 (2016).

[17] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, A structural approach to relaxation in glassy liquids, Nat. Phys. **12**, 469 (2016).

[18] Q. Wei, R. G. Melko, and J. Z. Chen, Identifying polymer states by machine learning, Phys. Rev. E **95**, 032504 (2017).

[19] G. Torlai and R. G. Melko, A neural decoder for topological codes, arXiv:1610.04238.

[20] E. P. L. van Nieuwenburg, Y. H. Liu, and S. D. Huber, Learning phase transitions by confusion, Nat. Phys. **13**, 435 (2017).

[21] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Machine learning phases of strongly correlated fermions, arXiv:1609.02552.

[22] G. E. Hinton, S. Osindero, and Y. Teh, A fast learning algorithm for deep belief nets, Neural Comput. **18**, 1527 (2006).

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems 25 (NIPS 2012)* (NIPS, 2012), pp. 1097–1105.

[24] A. Engel and C. Van den Broeck, *Statistical Mechanics of Learning* (Cambridge University Press, Cambridge, UK, 2001).

[25] P. Mehta and D. J. Schwab, An exact mapping between the variational renormalization group and deep learning, arXiv:1410.3831.

[26] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, UK, 1996), Vol. 5.

[27] U. Wolff, Collective Monte Carlo Updating for Spin Systems, Phys. Rev. Lett. **62**, 361 (1989).

[28] C. P. Robert, *Monte Carlo Methods* (Wiley, Hoboken, New Jersey, 2004).

[29] H. G. Katzgraber, Introduction to Monte Carlo methods, arXiv:0905.1629.

[30] S. J. Wetzel, Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders, arXiv:1703.02435.

[31] W. R. Gilks and C. Berzuini, Following a moving target—Monte Carlo inference for dynamic Bayesian models, J. R. Stat. Soc.: Ser. B **63**, 127 (2001).

[32] M. M. Graham and A. Storkey, Asymptotically exact inference in differentiable generative models, arXiv:1605.07826.

[33] L. F. Arsenault, R. Neuberg, L. A. Hannah, and A. J. Millis, Projected regression methods for inverting Fredholm integrals: Formalism and application to analytical continuation, arXiv:1612.04895.

[34] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, Self-learning Monte Carlo method, Phys. Rev. B **95**, 041101 (2017).

[35] K. Binder and D. P. Landau, Phase diagrams and critical behavior in Ising square lattices with nearest-and next-nearest-neighbor interactions, Phys. Rev. B **21**, 1941 (1980).

[36] G. H. Wannier, Antiferromagnetism. The triangular Ising net, Phys. Rev. **79**, 357 (1950).

[37] J. Stephenson, Ising-Model spin correlations on the triangular lattice, J. Math. Phys. **5**, 1009 (1964).

[38] J. Stephenson, Ising-model spin correlations on the triangular lattice. III. Isotropic antiferromagnetic lattice, J. Math. Phys. **11**, 413 (1970).

[39] M. Blume, Theory of the first-order magnetic phase change in UO_2, Phys. Rev. **141**, 517 (1966).

[40] H. W. Capel, On the possibility of first-order phase transitions in Ising systems of triplet ions with zero-field splitting, Physica **32**, 966 (1966).

[41] M. Blume, V. J. Emery, and R. B. Griffiths, Ising model for the λ transition and phase separation in He^3-He^4 mixtures, Phys. Rev. A **4**, 1071 (1971).

[42] A. K. Jain and D. P. Landau, Monte Carlo study of the fcc Blume-Capel model, Phys. Rev. B **22**, 445 (1980).

[43] C. J. Liu and H. B. Schüttler, Behavior of damage spreading in the two-dimensional Blume-Capel model, Phys. Rev. E **65**, 056103 (2002).

[44] E. Barouch and B. M. McCoy, Statistical mechanics of the XY model. II. Spin-correlation functions, Phys. Rev. A **3**, 786 (1971).

[45] J. M. Kosterlitz, The critical properties of the two-dimensional xy model, J. Phys. C **7**, 1046 (1974).

[46] L. Onsager, Crystal statistics. I. A two-dimensional model with an order-disorder transition, Phys. Rev. **65**, 117 (1944).

[47] J. C. Xavier, F. C. Alcaraz, D. P. Lara, and J. A. Plascak, Critical behavior of the spin-$\frac{3}{2}$ Blume-Capel model in two dimensions, Phys. Rev. B **57**, 11575 (1998).

[48] D. Mukamel and M. Blume, Ising model for tricritical points in ternary mixtures, Phys. Rev. A **10**, 610 (1974); the most general form of the Blume-Emery Griffiths model includes odd perturbation terms, $H \sum_{\mathbf{i}} S_{\mathbf{i}} + L \sum_{\langle ij \rangle} S_{\mathbf{i}} S_{\mathbf{j}}^2 + S_{\mathbf{i}}^2 S_{\mathbf{j}}$, where H, L are often both set to zero to prevent separate sublattices from forming.

[49] J. Lajzerowicz and J. Sivardiere, Spin-1 lattice-gas model. I. Condensation and solidification of a simple fluid, Phys. Rev. A **11**, 2079 (1975).

[50] K. E. Newman and J. D. Dow, Zinc-blende–diamond order-disorder transition in metastable crystalline $(GaAs)_{1-x}Ge_{2x}$ alloys, Phys. Rev. B **27**, 7495 (1983).

[51] H. Ez-Zahraouy and A. Kassou-Ou-Ali, Phase diagrams of the spin-1 Blume-Capel film with an alternating crystal field, Phys. Rev. B **69**, 064415 (2004).

[52] A. F. Brito, J. A. Redinz, and J. A. Plascak, Dynamics of rough surfaces generated by two-dimensional lattice spin models, Phys. Rev. E **75**, 046106 (2007).

[53] S. M. Pittman, G. G. Batrouni, and R. T. Scalettar, Monte Carlo study of an inhomogeneous Blume-Capel model: A case study of the local density approximation, Phys. Rev. B **78**, 214208 (2008).

[54] K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, London, Edinburgh, Dublin Philos. Mag. J. Sci. **2**, 559 (1901).

[55] Wikipedia contributors, Principal component analysis, Wikipedia, The Free Encyclopedia, 16 Mar. 2017.

[56] I. Jolliffe, *Principal Component Analysis* (Wiley, Hoboken, New Jersey, 2002).

[57] H. Bourlard and Y. Kamp, Auto-association by multilayer perceptrons and singular value decomposition, Biol. Cybern. **59**, 291 (1988).

[58] R. S. Zemel, *Autoencoders, Minimum Description Length and Helmholtz Free Energy* (NIPS, Denver, United States, 1994).

[59] G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science **313**, 504 (2006).

[60] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, Phys. Rev. Lett. **69**, 2863 (1992).

[61] S. Bradde and W. Bialek, PCA meets RG, J. Stat. Phys. **167**, 462 (2017).

[62] Y. Chauvin and D. E. Rumelhart, *Backpropagation: Theory, Architectures, and Applications* (Psychology Press, Hove, United Kingdom, 1995).

[63] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, Face recognition: A convolutional neural-network approach, IEEE Trans. Neural Netw. **8**, 98 (1997).

[64] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Netw. **61**, 85 (2015).

[65] W. Kwak, J. Jeong, J. Lee, and D. H. Kim, First-order phase transition and tricritical scaling behavior of the Blume-Capel model: A Wang-Landau sampling approach, Phys. Rev. E **92**, 022134 (2015).

[66] M. Rigol, T. Bryant, and R. R. P. Singh, Numerical linked-cluster algorithms. I. Spin systems on square, triangular, and kagomé lattices, Phys. Rev. E **75**, 061118 (2007).

[67] K. Binder, Applications of Monte Carlo methods to statistical physics, Rep. Prog. Phys. **60**, 487 (1997).

[68] S. V. Isakov and R. Moessner, Interplay of quantum and thermal fluctuations in a frustrated magnet, Phys. Rev. B **68**, 104409 (2003).

[69] D. Blankschtein, M. Ma, A. N. Berker, G. S. Grest, and C. M. Soukoulis, Orderings of a stacked frustrated triangular system in three dimensions, Phys. Rev. B **29**, 5250 (1984).

[70] P. Olsson, Monte Carlo analysis of the two-dimensional XY model. II. Comparison with the Kosterlitz renormalization-group equations, Phys. Rev. B **52**, 4526 (1995).

[71] J. Eisert, M. Cramer, and M. B. Plenio, Colloquium: Area laws for the entanglement entropy, Rev. Mod. Phys. **82**, 277 (2010).

[72] D. L. Deng, X. Li, and S. D. Sarma, Quantum Entanglement in Neural Network States, Phys. Rev. X **7**, 021021 (2017).

[73] T. J. Osborne and M. A. Nielsen, Entanglement in a simple quantum phase transition, Phys. Rev. A **66**, 032110 (2002).

[74] P. Calabrese and J. Cardy, Entanglement entropy and quantum field theory, J. Stat. Mech.: Theor. Exp. (2004) P06002.

[75] L. Cincio, J. Dziarmaga, M. M. Rams, and W. H. Zurek, Entropy of entanglement and correlations induced by a quench: Dynamics of a quantum phase transition in the quantum Ising model, Phys. Rev. A **75**, 052321 (2007).

[76] E. Haller, J. Hudson, A. Kelly, D. A. Cotta, B. Peaudecerf, G. D. Bruce, and S. Kuhr, Single-atom imaging of fermions in a quantum-gas microscope, Nat. Phys. **11**, 738 (2015).

[77] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner, A quantum gas microscope for detecting single atoms in a Hubbard-regime optical lattice, Nature (London) **462**, 74 (2009).

[78] J. F. Sherson, C. Weitenberg, M. Endres, M. Cheneau, I. Bloch, and S. Kuhr, Single-atom-resolved fluorescence imaging of an atomic Mott insulator, Nature (London) **467**, 68 (2010).

[79] F. Chollet, Building autoencoders in keras, May, 2016, https://blog.keras.io/building-autoencoders-in-keras.html

[80] Wikipedia contributors, Convolutional neural network, Wikipedia, The Free Encyclopedia, 16 Mar. 2017.