Discovering phase transitions with unsupervised learning

Lei Wang

Beijing National Lab for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China

(Received 6 June 2016; revised manuscript received 14 October 2016; published 2 November 2016)

Unsupervised learning is a discipline of machine learning which aims at discovering patterns in large data sets or classifying the data into several categories without being trained explicitly. We show that unsupervised learning techniques can be readily used to identify phases and phases transitions of many-body systems. Starting with raw spin configurations of a prototypical Ising model, we use principal component analysis to extract relevant low-dimensional representations of the original data and use clustering analysis to identify distinct phases in the feature space. This approach successfully finds physical concepts such as the order parameter and structure factor to be indicators of a phase transition. We discuss the future prospects of discovering more complex phases and phase transitions using unsupervised learning techniques.

DOI: 10.1103/PhysRevB.94.195105

Classifying phases of matter and identifying phase transitions between them is one of the central topics of condensed matter physics research. Despite an astronomical number of constituting particles, it often suffices to represent states of a many-body system with only a few variables. For example, a conventional approach in condensed matter physics is to identify order parameters via symmetry consideration or analyze low-energy collective degree of freedoms and use them to label phases of matter [1].

However, it is harder to identify phases and phase transitions in this way in an increasing number of new states of matter, where the order parameter may only be defined in an elusive, nonlocal way [2]. These developments call for different ways of identifying appropriate indicators of phase transitions.

To meet this challenge, we use machine learning techniques to extract information of phases and phase transitions directly from many-body configurations. In fact, the application of machine learning techniques to condensed matter physics is a burgeoning field [3–13].[1] For example, regression approaches are used to predict crystal structures [3], to approximate density functionals [6], and to solve quantum impurity problems [10]; artificial neural networks are trained to classify phases of classical statistical models [13]; neural network inspired variational wave functions are used to solve quantum many-body problems in and out of equilibrium [14]. Most of those applications use *supervised* learning techniques (regression and classification), where a learner needs to be trained with the previously solved data set (input/output pairs) before it can be used to make predictions.

On the other hand, in *unsupervised* learning, there is no such explicit training phase. The learner should by itself find out interesting patterns in the input data. Typical unsupervised learning tasks include cluster analysis and feature extraction. Cluster analysis divides the input data into several groups based on certain measures of similarities. Feature extraction finds a low-dimensional representation of the data set while still preserving the essential characteristics of the original data. Unsupervised learning methods have broad applications in data compression, visualization, online advertising, recommender systems, etc. They are often being used as a preprocessor of supervised learning to simplify the training procedure. In many cases, unsupervised learning also leads to better human interpretations of complex data sets.

In this paper, we explore the application of unsupervised learning in many-body physics with a focus on phase transitions. The advantage of unsupervised learning is that one assumes neither the presence of the phase transition nor the precise location of the critical point. Dimension reduction techniques can extract salient features such as order parameters and structure factors from the raw configuration data. Clustering analysis can then divide the data into several groups in the low-dimensional feature space, representing different phases. Our studies show that unsupervised learning techniques have great potentials for addressing the big data challenge in the many-body physics and making scientific discoveries.

As an example, we consider the prototypical classical Ising model

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

where the spins take two values $\sigma_i = \{-1, +1\}$. We consider the model (1) on a square lattice with periodic boundary conditions and set $J = 1$ as the energy unit. The system undergoes a phase transition at a temperature $T/J = 2/\ln(1+\sqrt{2}) \approx 2.269$ [15]. A discrete Z_2 spin inversion symmetry is broken in the ferromagnetic phase below T_c and is restored in the disordered phase at temperatures above T_c.

We generate 100 uncorrelated spin configuration samples using the Monte Carlo simulation with the Wolff cluster update [16] at temperatures $T/J = 1.6, 1.7, \ldots, 2.9$ each and collect them into an $M \times N$ matrix,

$$X = \begin{pmatrix} \uparrow & \downarrow & \uparrow & \cdots & \uparrow & \uparrow & \uparrow \\ & & & \vdots & & & \\ \downarrow & \uparrow & \downarrow & \cdots & \uparrow & \downarrow & \uparrow \end{pmatrix}_{M \times N}, \tag{2}$$

where $M = 1400$ is the total number of samples, and N is the number of lattice sites. The up and down arrows in the matrix denote $\sigma_i = \pm 1$. Such a matrix is the *only* data we feed to the unsupervised learning algorithm.

[1] We also note the application of physics ideas such as phase transition [30], renormalization group [31], tensor networks [32], and quantum computation [33] to machine learning.

Our goal is to discover a possible phase transition of the model (1) without assuming its existence. This is different from supervised learning tasks, where exact knowledge of T_c was used to train a learner [13]. Moreover, the following analysis does not assume any prior knowledge about the lattice geometry and the Hamiltonian. We are going to use the unsupervised learning approach to extract salient features of the data and then use this information to divide the samples into distinct phases. Knowledge about the temperature of each sample and the critical temperature T_c of the Ising model is used to verify the clustering.

Interpreting each row of X as a coordinate of an N-dimensional space, the M data points form a cloud centered around the origin of a hypercube.[2] Discovering a phase transition amounts to finding a hypersurface which divides the data points into several groups, each representing a phase. The task is akin to the standard unsupervised learning technique, cluster analysis [17], where numerous algorithms are available, and they classify the data based on different criteria.

However, directly applying clustering algorithms to the Ising configurations may not be very enlightening. The reasons are twofold. First, even if one manages to separate the data into several groups, clusters in high-dimensional space may not directly offer useful physical insights. Second, many clustering algorithms rely on a good measure of similarity between the data points. Its definition is, however, ambiguous without supplying domain knowledge, such as the distance between two spin configurations.

In fact, the raw spin configuration is a highly redundant description of the system's state because there are correlations among the spins. Moreover, as the temperature varies, there is an overall tendency in the raw spin configurations, such as lowering the total magnetization. In the following, we will first identify these crucial features as effective low-dimensional representations of the original data. The meaning of the distance between configurations will become more transparent in terms of these features. The separation of phases is also often clearly visible and comprehensible by the human in the reduced space spanned by these features. Therefore, feature extraction not only simplifies the subsequent clustering analysis but also provides an effective means of visualizing and offering physical insights. We denote the crucial features extracted by unsupervised learning as *indicators* of the phase transition. In general, they do not necessarily need to be the same as the conventional order parameters defined in condensed matter physics. This unsupervised learning approach nevertheless provides an alternative view of phases and phase transitions.

Principal component analysis (PCA) [18] is a widely used feature extraction technique. The principal components are mutually orthogonal directions along which the variances of the data decrease monotonically. PCA finds the principal components through a linear transformation of the original coordinates $Y = XW$. When applied to the Ising configurations in Eq. (2), PCA finds the most significant variations of the data changing with the temperature. We interpret them as relevant features in the data and use them as indicators of the phase transition, if there is any.

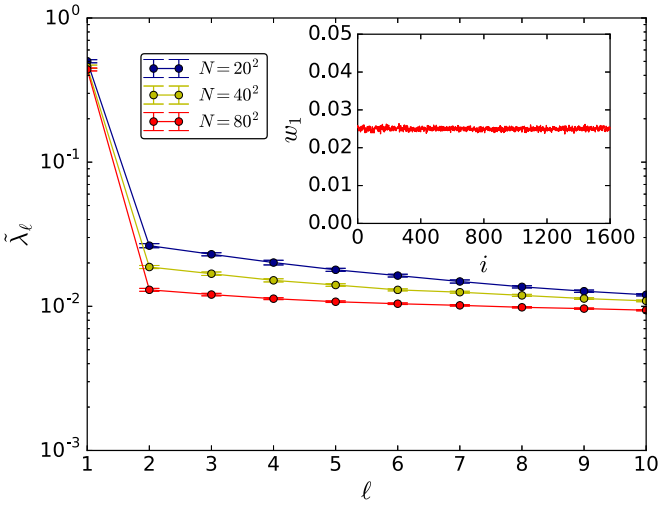

FIG. 1. The first few explained variance ratios obtained from the raw Ising configurations. The inset shows the weights of the first principal component on an $N = 40^2$ square lattice.

We write the orthogonal transformation into column vectors $W = (w_1, w_2, \ldots, w_N)$ and denote w_ℓ as weights of the principal components in the configuration space. They are determined by an eigenproblem [19][3]

$$X^T X w_\ell = \lambda_\ell w_\ell. \tag{3}$$

The eigenvalues are non-negative real numbers sorted in descending order, $\lambda_1 \geqslant \lambda_2 \cdots \geqslant \lambda_N \geqslant 0$. Using the terminology of PCA, we denote the normalized eigenvalues $\tilde{\lambda}_\ell = \lambda_\ell / \sum_{\ell=1}^{N} \lambda_\ell$ as the *explained variance ratio*. When keeping only the first few principal components, PCA is an efficient dimension reduction approach which captures most variations of the original data. Moreover, PCA also yields an optimal approximation of the data in the sense of minimizing the squared reconstruction error [19].

Figure 1 shows the first few explained variance ratios for various system sizes. Notably, there is only one dominant principal component. The error bars are estimated by dividing the data into four groups and repeating the calculation (3) for each group. As the temperature changes, the Ising configurations vary most significantly along the first principal component, whose weight is shown in the inset of Fig. 1. The flat distribution over all the lattice sites means the transformation actually gives the uniform magnetization $\frac{1}{N}\sum_i \sigma_i$. In this sense, PCA has identified the order parameter of the Ising model (1) upon a phase transition.

Next, we project the samples in the space spanned by the first two principal components, shown in Fig. 2. The

[2]Each column of X sums up to zero since on average each site has zero magnetization.

[3]In practice, this eigenproblem is often solved by a singular value decomposition of X. In fact, replacing the input data X (raw spin configurations collected at various temperature) by the wave function of a one-dimensional quantum system, the math here is identical to the truncation of Schmidt coefficients in density-matrix renormalization group calculations [34].

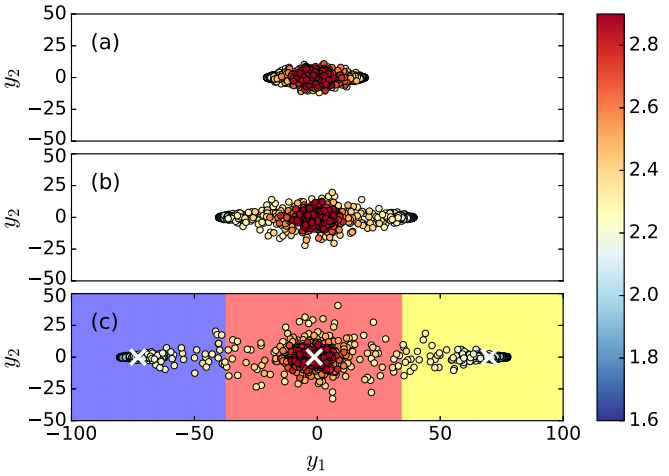

FIG. 2. Projection of the samples onto the plane of the two leading principal components. The color bar on the right indicates the temperature T/J of the samples. (a)–(c) are for $N = 20^2$, 40^2, and 80^2 sites, respectively. In (c) we perform k-means clustering to split the data into several phases. The white crosses denote the cluster centroids. Background colors indicate different predicted phases.

color of each sample indicates its temperature. The projected coordinates are given by the matrix-vector product

$$y_\ell = X w_\ell. \tag{4}$$

The variation of the data along the first principal axis y_1 is indeed much stronger than that along the second principal axis y_2. Most importantly, one clearly observes that as the system size enlarges, the samples tend to split into three clusters. The high-temperature samples lie around the origin while the low-temperature samples lie symmetrically at finite y_1. The samples at the critical temperature (light yellow dots) have a broad spread because of large critical fluctuations. We note that Ref. [13] presents a different low-dimensional visualization of the Ising configurations using the stochastic neighbor embedding technique.

When folding the horizontal axis of Fig. 2 into $\sum_i |\sigma_i|$ or $(\sum_i \sigma_i)^2$, the two clusters associated with the low-temperature phase merge together. With such a linearly separable low-dimensional representation of the original data, a cluster analysis algorithm such as k-means[4] can easily identify clusters corresponding to different phases. The vertical decision boundaries in Fig. 2(c) show that only y_1 affects the division. The clustering analysis also provides an estimate of the critical temperature $T_c/J \approx 2.3$. Notice that the unsupervised learning analysis not only discovers the phase transition and estimates the critical temperature but also offers insight into the difference between phases.

Having established the baseline of applying unsupervised learning techniques in a prototypical Ising model, we now turn to a more challenging case where the learner can make nontrivial findings. For this, we consider the same Ising model Eq. (1) with a conserved order parameter (COP) $\sum_i \sigma_i \equiv 0$. This model describes classical lattice gases [20], where the

[4]http://scikit-learn.org/stable/modules/clustering.html#k-means

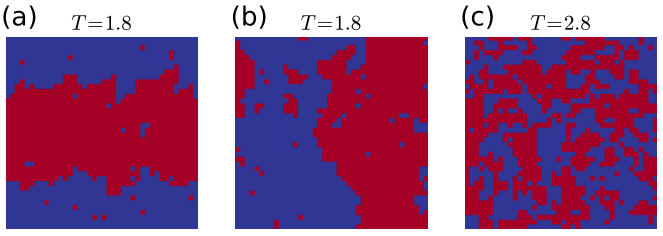

FIG. 3. Typical configurations of the COP Ising model (a), (b) below and (c) above the critical temperature. Red and blue pixels indicate up and down spins. Exactly half of the pixels are red/blue due to the constraint $\sum_i \sigma_i \equiv 0$.

occupation of each lattice site can be either one or zero and the particles interact via a short-range attraction. The conserved total magnetization corresponds to the constraint of a half-filled lattice.

On a square lattice with periodic boundary conditions, the spins tend to form two domains at low temperatures, shown in Figs. 3(a) and 3(b). The two domain walls wrap around the lattice either horizontally or vertically to minimize the domain wall energy [20]. Besides, the domains can also shift in space due to translational invariance. As the temperature increases, these domain walls melt and the system restores both the translational and rotational symmetries in the high-temperature phase shown in Fig. 3(c). At zero total magnetization, the critical temperature of such a solid-gas phase transition is the same as the Ising transition $T_c/J \approx 2.269$ [21]. However, since the total magnetization is conserved, simply summing up the spins as the ordinary Ising model cannot be used as an indicator to distinguish the two phases.

We perform the same PCA to the COP Ising configurations sampled with Kawasaki spin exchange Monte Carlo updates [20,22]. Figure 4 shows the first few explained variance ratios. Notably, there are four leading principal components instead of one. Their weights, plotted in the insets of Fig. 4, show a notable nonuniformity over the lattice sites. This

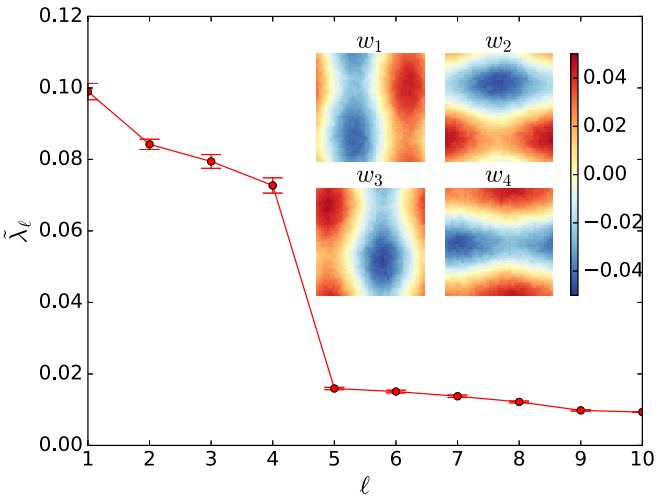

FIG. 4. Explained variance ratios of the COP Ising model. The insets show the weights corresponding to the four leading principal components.

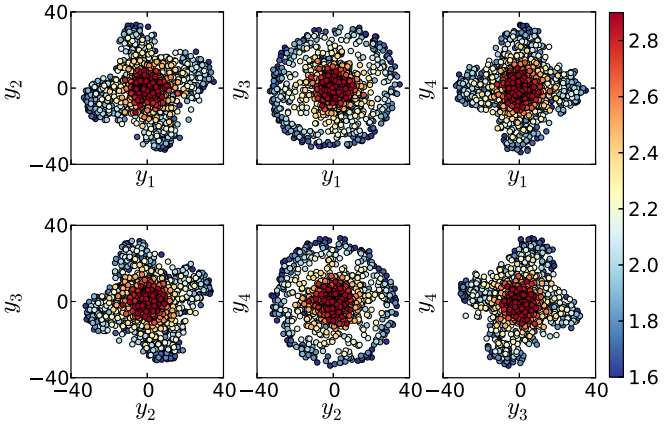

FIG. 5. Projections of the COP Ising samples to the four leading principal components.

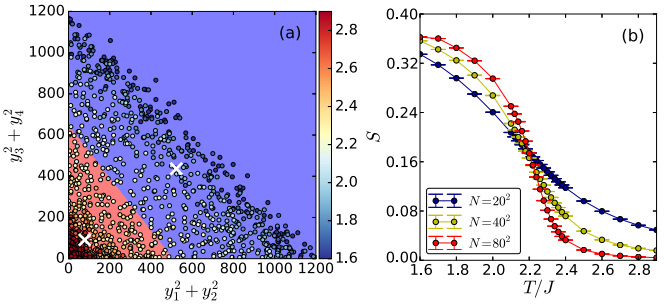

FIG. 6. (a) Further projection of the COP Ising samples to a two-dimensional space. The line between the red and blue background colors is the decision boundary found by the k-means algorithm. The white crosses denote the cluster centroids. (b) The structure factor Eq. (5) of the COP Ising model vs temperature for various system sizes.

indicates that in the COP Ising model the spatial distribution of the spins varies drastically as the temperature changes. Denote the Euclidean coordinate of site i as (μ_i,ν_i), where $\mu_i,\nu_i = 1,2,\ldots,\sqrt{N}$. The weights of the four leading principal components can be written as $\cos(\theta_i)$, $\cos(\phi_i)$, $\sin(\theta_i)$, $\sin(\phi_i)$, where $(\theta_i,\phi_i) = (\mu_i,\nu_i)\times 2\pi/\sqrt{N}$.[5] Note these four mutually orthogonal weights correspond to the two orientations of the domain walls shown in Figs. 3(a) and 3(b). Therefore, the PCA correctly finds out the rotational symmetry breaking caused by the domain wall formation.

To visualize the samples in the four-dimensional feature space spanned by the first few principal components, we plot two-dimensional projections in Fig. 5. In all cases, the high-temperature samples are around the origin while the low-temperature samples form a surrounding cloud. Motivated by the circular shapes of all these projections, we further reduce the data to a two-dimensional space via a nonlinear transformation $(y_1,y_2,y_3,y_4)\mapsto(y_1^2+y_2^2,y_3^2+y_4^2)$. As shown in Fig. 6(a), the line $\sum_{\ell=1}^4 y_\ell^2 \approx$ const (a four-dimensional sphere of a constant radius) separates the low- and high-temperature samples. This motivates a further dimensional reduction to a single variable $\sum_{\ell=1}^4 y_\ell^2$ as an indicator of the phase transition in the COP Ising model.

Substituting weights of the four principal components $\cos(\theta_i)$, $\cos(\phi_i)$, $\sin(\theta_i)$, $\sin(\phi_i)$, the sum $\sum_{\ell=1}^4 y_\ell^2$ is proportional to

$$S=\frac{1}{N^2}\sum_{i,j}\sigma_i\sigma_j[\cos(\theta_i-\theta_j)+\cos(\phi_i-\phi_j)]. \qquad (5)$$

Such an expression is equivalent to the sum of spin structure factors at momenta $(2\pi/\sqrt{N},0)$ and $(0,2\pi/\sqrt{N})$. Figure 6(b) shows the structure factor (5) versus temperature for various system sizes. It decreases as the temperature increases and clearly serves as a good indicator of the phase transition. We emphasis that the input spin configurations contain no information about the lattice geometry or the Hamiltonian. However, the unsupervised learner has by itself found out meaningful information related to the breaking of the orientational order. Therefore, even without the knowledge of the lattice and an analytical understanding of the structure factor Eq. (5), $\sum_{\ell=1}^4 y_\ell^2$ plays the same role as Eq. (5) to separate the phases in the projected feature space.

It is interesting to compare our analysis of phase transitions to standard image recognition applications. In the Ising model example, the learner essentially finds out the brightness of the imagine $\sum_i \sigma_i$ as an indicator of a phase transition, while in the COP Ising model example, instead of detecting the sharpness of the edges (melting of domain walls) following the ordinary image recognition routine, the PCA learner finds out the structure factor Eq. (5) related to symmetry breaking, which is a fundamental concept in phase transition and condensed matter physics.

Considering that PCA is arguably one of the simplest unsupervised learning techniques, the obtained results are rather encouraging. In essence, our analysis finds the dominant collective modes of the system related to the phase transition. The approach can be readily generalized to more complex cases such as models with emergent symmetry and order by disorder [23]. The unsupervised learning approach is particularly profitable in the case of hidden or multiple intertwined orders, where it can help to single out various phases.

Although *nonlinear* transformation of the raw configuration Eq. (5) was discovered via visualization in Fig. 5, PCA is, however, limited to linear transformations. Therefore, it remains challenging to identify more subtle phase transitions related to the topological order, where the indicators of the phase transition are nontrivial nonlinear functions of the original configurations. For this purpose, it would be interesting to see if a machine learning approach can comprehend concepts such as duality transformation [24], Wilson loops [25], and string order parameters [26]. A judicial application of kernel techniques [27] or neural network based deep autoencoders [28] may achieve some of these goals.

Furthermore, although our discussions focus on thermal phase transitions of the classical Ising model, unsupervised

[5]The weights shown in the inset of Fig. 4 are linear mixtures of them.

learning approaches can also be used to analyze quantum many-body systems and quantum phase transitions [29]. In these applications, diagnosing quantum states of matter without knowledge of the Hamiltonian is a useful paradigm for cases with access only to wave functions or experimental data.

The author thanks Xi Dai, Ye-Hua Liu, Yuan Wan, QuanSheng Wu, and Ilia Zintchenko for discussions and encouragement. The author also thanks Zi Cai for discussions and careful reading of the manuscript. L.W. is supported by the Ministry of Science and Technology of China under Grant No. 2016YFA0300603 and a startup grant of IOP-CAS.

[1] P. W. Anderson, *Basic Notions of Condensed Matter Physics* (Benjamin Cummings, San Francisco, 1984).
[2] X.-G. Wen, *Quantum Field Theory of Many-Body Systems* (Oxford University Press, Oxford, U.K., 2004).
[3] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder, Phys. Rev. Lett. **91**, 135503 (2003).
[4] O. S. Ovchinnikov, S. Jesse, P. Bintacchit, S. Trolier-McKinstry, and S. V. Kalinin, Phys. Rev. Lett. **103**, 157203 (2009).
[5] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, Chem. Mater. **22**, 3762 (2010).
[6] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, Phys. Rev. Lett. **108**, 253002 (2012).
[7] Y. Saad, D. Gao, T. Ngo, S. Bobbitt, J. R. Chelikowsky, and W. Andreoni, Phys. Rev. B **85**, 104104 (2012).
[8] E. LeDell, Prabhat, D. Y. Zubarev, B. Austin, and W. A. Lester, J. Math. Chem. **50**, 2043 (2012).
[9] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).
[10] L.-F. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, Phys. Rev. B **90**, 155136 (2014).
[11] G. Pilania, J. E. Gubernatis, and T. Lookman, Phys. Rev. B **91**, 214302 (2015).
[12] Z. Li, J. R. Kermode, and A. De Vita, Phys. Rev. Lett. **114**, 096405 (2015).
[13] J. Carrasquilla and R. G. Melko, arXiv:1605.01735.
[14] G. Carleo and M. Troyer, arXiv:1606.02318.
[15] L. Onsager, Phys. Rev. **65**, 117 (1944).
[16] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[17] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis* (Wiley, Hoboken, NJ, 2010).
[18] K. Pearson, Philos. Mag. **2**, 559 (1901).
[19] I. Jolliffe, *Principal Component Analysis* (Wiley, Chichester, U.K., 2002).
[20] M. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, Oxford, U.K., 1999).
[21] C. N. Yang, Phys. Rev. **85**, 808 (1952).
[22] K. Kawasaki, Phys. Rev. **145**, 224 (1966).
[23] R. Moessner and S. L. Sondhi, Phys. Rev. B **63**, 224401 (2001).
[24] F. J. Wegner, J. Math. Phys. **12**, 2259 (1971).
[25] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).
[26] M. den Nijs and K. Rommelse, Phys. Rev. B **40**, 4709 (1989).
[27] B. Schölkopf, A. Smola, and K. R. Müller, Neural Comput. **10**, 1299 (1998).
[28] G. E. Hinton and R. R. Salakhutdinov, Science **313**, 504 (2006).
[29] S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge, U.K., 2011).
[30] L. Saitta, A. Giordana, and A. Cornuéjols, *Phase Transitions in Machine Learning* (Cambridge University Press, Cambridge, U.K., 2011).
[31] P. Mehta and D. J. Schwab, arXiv:1410.3831.
[32] E. M. Stoudenmire and D. J. Schwab, arXiv:1605.05775.
[33] S. Lloyd, M. Mohseni, and P. Rebentrost, arXiv:1307.0411.
[34] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).