

Unsupervised learning of topological phase transitions using the Calinski-Harabaz index

Jielin Wang,¹ Wanzhou Zhang^{2,*}, Tian Hua,^{1,†} and Tzu-Chieh Wei^{3,‡}

¹College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China

²College of Physics and Optoelectronics, Taiyuan University of Technology, Taiyuan 030024, China

³Department of Physics and Astronomy, C.N. Yang Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3840, USA



(Received 30 December 2019; revised 16 October 2020; accepted 16 November 2020; published 22 January 2021)

Machine learning methods have been recently applied to learning phases of matter and transitions between them. Of particular interest is the topological phase transition, such as in the XY model, whose critical points can be difficult to be obtained by using unsupervised learning, such as the principal component analysis. Recently, the authors of [Nat. Phys. **15**, 790 (2019)] employed the diffusion map method for identifying topological orders and were able to determine the Berezinskii-Kosterlitz-Thouless (BKT) phase transition of the XY model, specifically via the intersection of the average cluster distance \bar{D} and the within-cluster dispersion parameter $\bar{\sigma}$ (when the different clusters vary from separation to mixing together). However, sometimes it is not easy to find the intersection if \bar{D} or $\bar{\sigma}$ does not change too much due to topological constraint. In this paper, we propose to use the Calinski-Harabaz (ch) index, defined roughly as the ratio $\bar{D}/\bar{\sigma}$, to determine the critical points at which the ch index reaches a maximum or minimum value or jumps sharply. We examine the ch index in several statistical models, including ones that contain a BKT phase transition. For the Ising model, the peaks of the quantity ch or its components are consistent with the position of the specific-heat maximum. For the XY model, both on the square and on the honeycomb lattices, our results of the ch index show the convergence of the peaks over a range of parameters $\varepsilon/\varepsilon_0$ in the Gaussian kernel. We also examine the generalized XY model with $q = 2$ and $q = 8$ and study the phase transition using the fractional $1/2$ -vortex or $1/8$ -vortex constraint, respectively. The global phase diagram can be obtained by our method, which does not use the label of configuration needed by supervised learning, nor a crossing from two curves \bar{D} and $\bar{\sigma}$. Our method is, thus, useful to both topological and nontopological phase transitions and can achieve accuracy as good as supervised learning methods previously used in these models and may be used for searching phases from experimental data.

DOI: [10.1103/PhysRevResearch.3.013074](https://doi.org/10.1103/PhysRevResearch.3.013074)

I. INTRODUCTION

Exploring phases of matter and phase transitions is a traditional but still active research direction in statistical physics [1,2], partly due to new phases of matter that have been uncovered. In recent years, this field of research has been revived thanks to the introduction of artificial intelligence and machine learning methods to recognize phases and transitions [3]. Among the various methods, supervised learning methods are used to train the network using prior labeled data generated by Monte Carlo (MC) methods in various classical systems, such as the Ising model and its variants [3–5], XY models [6–9] with Berezinskii-Kosterlitz-Thouless (BKT) phase transitions [10,11], Dzyaloshinskii-Moriya ferromagnets [12], skyrmions [13], Potts models [14], and percolation

models [9]. The properties of quantum systems, such as the energy spectrum, entanglement spectrum [15], or configuration in the imaginary time [16,17] are also used as inputs to the neural networks for the training. On the other hand, unsupervised learning methods can provide unbiased classification as they do not need prior knowledge of the system so as to classify phases and obtain the phase transition. These include methods, such as the principal component analysis [18], autoencoders [19], t -distributed stochastic neighboring ensemble [20], and clustering with quantum mechanics [21]. A key feature of these is to determine the sought-after properties (such as phase transition) by examining indicators from the scattered plot in the reduced space. Beyond equilibrium statistical physics, nonequilibrium and dynamical properties [22–24] are also obtained by machine learning methods. In addition, there are other newly developed schemes applied to the phases of matter, such as the adversarial neural networks [25], confusion method and its extension [26], the super-resolving method [27], and even applications to experimental data [28,29]. Other developments in this area can be found in Refs. [30–41] and the review article [42].

It is well known that in traditional continuous phase transitions global symmetry is broken and these transitions can be described by the Landau theory. However, the topological

*Corresponding author: zhangwanzhou@tyut.edu.cn

†Corresponding author: tianhua@tyut.edu.cn

‡Corresponding author: tzu-chieh.wei@stonybrook.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

phase transitions have no broken symmetry, and, therefore, it is of great interest to understand how the transitions emerge and how to locate the transition points. Recent developments of machine learning offer new tools for this endeavor [6–9], based on supervised learning. However, the learning-by-confusion scheme when applied to the XY model predicts a transition temperature set by the value located at the maximum of the specific heat [8]. In Ref. [7], it was found that significant feature engineering of the raw spin states is needed to relate vortex configurations and the transition. Moreover, single-hidden-layer fully connected networks could not correctly learn the phases in the XY model, whereas the convolutional neural networks were successfully employed to learn the BKT transition [7], and the classification was later extended to the generalized XY model [9]. However, for unsupervised learning with the principal component analysis (PCA) method, it was claimed to be difficult to identify the transition [6]. Recently, Rodriguez-Nieva and Scheurer (RNS) used the diffusion map method [43] invented by Coifman *et al.* [44] and related to quantum clustering [21] and devised an unsupervised learning method for identifying topological orders and successfully locating the BKT transition. They divided the configurations into several topological sectors with different winding numbers, then calculated the eigenvector Ψ and eigenvalues λ of the so-called diffusion matrix P (see Sec. II A). From the intersections of the average cluster distance \bar{D} and within-cluster dispersion $\bar{\sigma}$, or, equivalently, the intersection of $\Delta\lambda$ (the jump in the eigenvalues) and σ_λ (the standard deviation of the eigenvalues), the phase transition of the XY model on the square lattice can be obtained very well.

Our motivation of this paper is to examine whether or not the diffusion map (DM) method of the RNS method is suitable beyond the pure XY model, such as the generalized XY model. Indeed we find that the DM method works for some topological phase transitions, but it fails to locate the phase transition in the generalized XY model in other regimes. Specifically, the RNS method for determining the transition point relies on the intersection of the two curves (such as the average cluster distance \bar{D} and within-cluster dispersion $\bar{\sigma}$), and in the $q = 8$ generalized XY model as illustrated in Sec. V, we cannot find an intersection there. The thermal fluctuation is not strong enough, and the scattering clusters with different winding as numbers do not mix close to T_c , i.e., \bar{D} does not decrease substantially. The question arises: Are there other quantities that can be used to locate the phase-transition points?

In this paper, we mainly use the Calinski-Harabaz (ch) index score [45] defined by ch_t/ch_b , related to the ratio of $\bar{D}/\bar{\sigma}$, which means that if the variation of \bar{D} can be negligible due to strong topological constraints, the variation of $\bar{\sigma}$ itself can help to determine the phase-transition point. We also use the silhouette coefficient (sc) [46] or its components to compare with the results.

The outline of this paper is as follows. Section II shows the DM methods, the definition of the indices ch, sc, and their components, their advantage, and prior knowledge for using the indices. Section III shows the signature of ch and sc for the two-dimensional (2D) Ising model from configurations using the Swendsen-Wang algorithms [47]. In Sec. IV, the critical phase-transition points of the XY model on the square and

the honeycomb lattices are obtained using the DM method. In Sec. V, for the $q = 2$ and $q = 8$ generalized XY models, the global phase diagrams are obtained by the DM method assisted by PCA or kernel-PCA (k-PCA) methods. Other techniques are discussed in Sec. VI regarding the effect of higher dimensions considered in the k -means method and how to automatically find the hyperparameter $\varepsilon/\varepsilon_0$. Concluding comments are made in Sec. VII. In Appendix A, the simplest example one-dimensional (1D) XY model is discussed, and the comparison between PCA and k-PCA is shown in Appendix B. Finally, a total of 11 indices used in unsupervised learning are listed in Appendix C.

II. METHODS

A. The review of the diffusion map method

Here, we explain the DM method of Rodriguez-Nieva and Scheurer [43]. Assume that we have M configurations $\{x_l\}$, where $l = 1, \dots, M$. The connectivity between x_l and $x_{l'}$ is denoted by the elementary Gaussian kernel,

$$K_\varepsilon(x_l, x_{l'}) = \exp\left(-\frac{\|x_l - x_{l'}\|^2}{2N\varepsilon}\right). \quad (1)$$

The normalization of $K_\varepsilon(x_l, x_{l'})$ can be performed by performing the sum over l' ,

$$P_{l,l'} = \frac{K_\varepsilon(x_l, x_{l'})}{z_l}, \quad z_l = \sum_{l'=1}^m K_\varepsilon(x_l, x_{l'}). \quad (2)$$

We can also perform the normalization along the direction of l (i.e., the sum over l). The eigenvalue equation of the diffusion matrix $P_{l,l'}$ is $P\psi_k = \lambda_k\psi_k$, where ψ_k 's are the right eigenvectors with the corresponding eigenvalues $\lambda_k \leq 1$ for $k = 0, 1, \dots, m-1$.

To find the phase-transition point, Ref. [43] and its earlier preprint use two different ways, respectively:

(a) Intersection of the mean distance of cluster centers $\bar{D}/2n$, where n is the number of clusters and the dispersion $\bar{\sigma}$ of the data points in each cluster. The quantities $\bar{D}/2n$ and $\bar{\sigma}$ are obtained from the scatter plot, where n is the number of topological sectors.

After obtaining the eigenvectors of the P matrix given by

$$\Phi := [(\psi)_1, (\psi)_2, \dots, (\psi)_{m-1}], \quad (3)$$

the authors project them onto a two-dimensional space, namely, a two-column matrix, then $\bar{D}/2n$ and the dispersion $\bar{\sigma}$ can be obtained from the scatter plot of the two-column vectors or their modified version. The detailed application to the one-dimensional XY model is reproduced in Appendix A.

(b) Intersection of the mean fluctuation σ_λ and gap of eigenvalues $\Delta\lambda$, where

$$\sigma_\lambda = \frac{1}{n} \sum_{k=0}^{n-1} (\lambda_k - \bar{\lambda})^2, \quad \bar{\lambda} = \frac{1}{n} \sum_{k=0}^{n-1} \lambda_k, \quad (4)$$

and the gap of eigenvalues between the topological sectors n and $n-1$,

$$\Delta\lambda = \lambda_n - \lambda_{n-1}. \quad (5)$$

B. The indices ch and sc

We propose to use indices instead of intersections. Based on the first two leading eigenvectors Ψ_0 and Ψ_1 of the PCA, k-PCA, or the second and third vectors Ψ_1 and Ψ_2 of the DM method for a manually chosen cluster number k , the ch index is given in terms of the following ratio:

$$ch = \frac{ch_t}{ch_b} = \frac{tr(B_k) N - k}{tr(W_k) k - 1}, \tag{6}$$

where B_k is the between-group dispersion matrix and W_k is the within-cluster dispersion matrix and they are defined as follows:

$$W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q)(x - c_q)^T, \tag{7}$$

$$B_k = \sum_q n_q (c_q - c)(c_q - c)^T,$$

where N is the number of data points, C_q is the set of points in cluster q , c_q is the center of cluster q , c denotes the average center of all cluster centers $\{c_q\}$, and n_q denotes the number of points in cluster q . The sc index of the i th sample is

$$sc(i) = \frac{b(i) - a(i)}{\max[b(i), a(i)]}, \tag{8}$$

where $a(i)$ is the mean distance between sample i and all other data points in the same cluster, $b(i)$ is the mean dissimilarity of point i to some cluster C expressed as the mean of the distance from i to all points in C (where $C \neq C_i$). The mean $sc = \sum sc(i)/N$ over all points of a cluster is a measure of how tightly grouped all the points in the cluster are. Sometimes $sc_a = \sum a(i)/N$ or $sc_b = \sum b(i)/N$ are also very useful [48].

The performance of a total of 11 indices is shown in Appendix C, and the results of indices ch, dn, pbm, and Ii applied to the $q = 8$ generalized XY model are all reasonable choices. Here, we choose two representative indices ch and sc as examples.

C. Advantage and prior knowledge required using ch

Figure 1 draws the flowchart of the RNS method and the place of our modifications. Depending on whether or not we are considering topology, we perform dimensional reduction by DM or PCA (k-PCA), respectively, to get the features of the data, i.e., the eigenvectors of the diffusion matrix P or the covariance matrix. Then we apply k -means to cluster the two-dimensional or higher-dimensional scattering points. If we do not choose ch, then intersections of $\Delta\lambda$ and σ_λ , or the intersections of $\bar{D}/2n$ and $\bar{\sigma}$ [43] can be used instead.

To test whether the clustering is good or bad, if ch is chosen, then the peaks of ch or its components ch_t and ch_b will be used directly. For the topological phase transition, the index ch or its components ch_t and ch_b can give signatures of phase transition when it is not easy to determine the intersection of the RNS method or when there is no intersection.

The index ch can also be applied to nontopological phase transitions, such as the order-disorder phase transition of the Ising model. This does not require any prior knowledge except

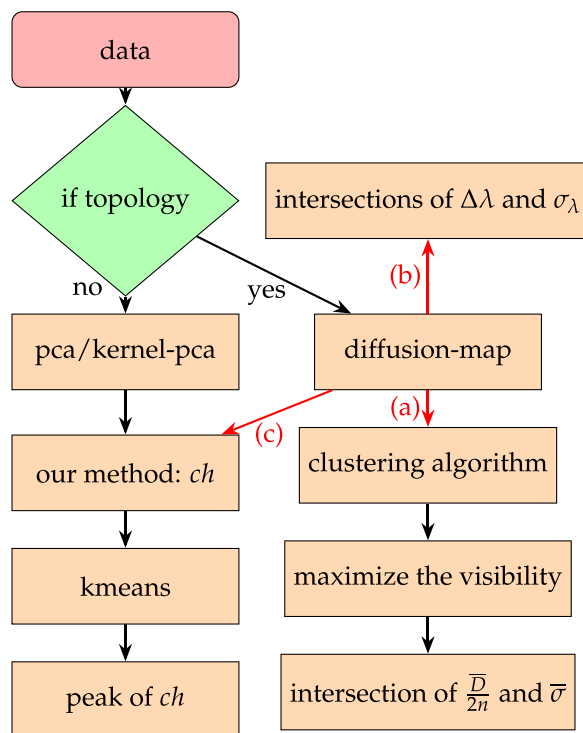


FIG. 1. Flow chart and main steps of the RNS method and our modifications. The main difference begins at how to use the DM method, (a) clustering algorithm in the RNS method, (b) intersections of $\Delta\lambda$ and σ_λ , and (c) our method using ch.

for the configurations generated by, e.g., Monte Carlo methods or from real experiments.

For topological phase transitions, such as the XY and generalized XY models, although this type of unsupervised learning is not similar to supervised learning (such as fully connected layers, or convolutional neural networks), it still needs labels of configurations, i.e., the topological winding number and the number of possible phases. However, the label of the topological winding number does not mean the label of phases, and essentially, this method is still an unsupervised learning method.

III. THE TWO-DIMENSIONAL ISING MODEL

To test the ability to locate the phase-transition point T_c , we calculate ch for the simplest model, i.e., the Ising model,

$$H = -J \sum_{(i,j)} S_i S_j, \tag{9}$$

where J is the ferromagnetic interaction between a pair of nearest-neighborhood spins and $S_i = \pm 1$. The unsupervised learning of the Ising model has been studied before, (see, e.g., Refs. [18,19]).

We use a two-dimensional 64×64 lattice and generate $N_s = 2000$ samples for each temperature T_i and analyze them by PCA, using the scattering data of the first two leading eigenvectors Ψ_0 and Ψ_1 . Moreover, we calculate $ch(k = 2)$ and $sc(k = 2)$ for each T_i according to Eqs. (6) and (8).

Figure 2 shows the main results for the Ising model. In Fig. 2(a), sc itself and sc_b have a sharp decrease whereas

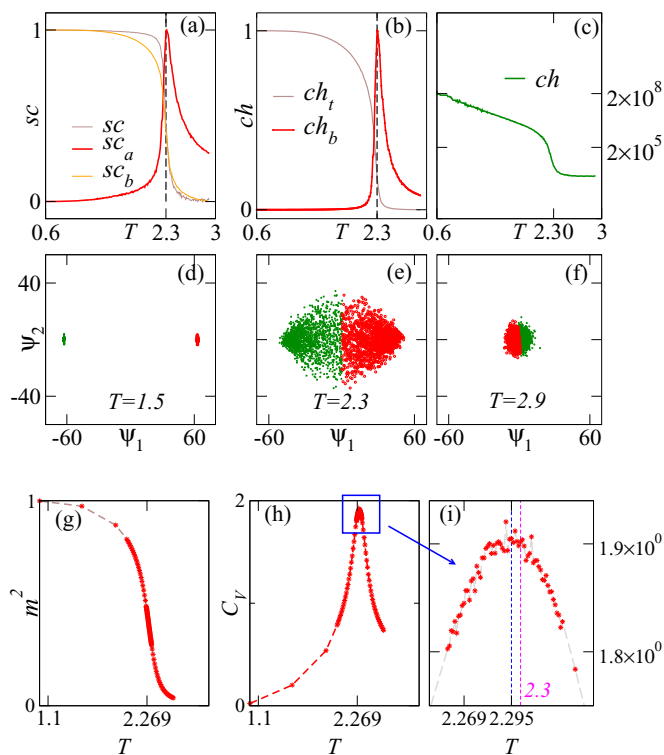


FIG. 2. For the Ising model with sizes $L = 64$, (a) sc , sc_a , sc_b , (b) ch_t and ch_b presented in the normalized range $[0,1]$, (c) ch (d)–(f) scatter plots with Ψ_1 and Ψ_2 at three typical temperatures $T < T_c$, $T \approx T_c$, and $T > T_c$, respectively, where different colors mean labels generated by k -means. (g)–(i) Monte Carlo results square of magnetization m , specific heat capacity C_V , and zoomed C_V . The peaks are at 2.295, consistent with 2.30.

sc_a has peaks around T_c (here we rescale the results so the maximum value is 1). In Figs. 2(b) and 2(c), the ch_b index is peaked around 2.3; ch_t and ch also have a sharp jump at the phase-transition points.

To understand these results, the scatter plots of Ψ_1 and Ψ_2 are shown in Figs. 2(d)–2(f) for temperatures $T = 1.5$, 2.3, and 2.9, respectively. At low temperatures, the clusters identified by the two colors separate from each other in the reduced space and finally mix together at high temperatures.

In general, the ch index is large when the clusters are well separated, and the points in each cluster are well aggregated. From this viewpoint in the low-temperature phase, the ch index becomes large because all up states and all down states can be easily distinguished if the analysis is successful.

The index ch and its components perform well in detecting the transition, and the position of the peak or if the jump is the largest at $T = 2.3$. Around the phase-transition point, configurations possess the properties from both phases (paramagnetic and ferromagnetic), and the fluctuation is the largest there.

Comparing to traditional Monte Carlo results with the same size $L = 64$ as shown in Figs. 2(g)–2(i), we find that the position of the peak is located at around 2.295 consistent with our ch or sc results with numerical intervals of 0.01. For the purpose of reference, the thermodynamic limit transition $T_c = 2.269$ is marked.

It should be noted that if we simulate the Ising model by the Metropolis Monte Carlo algorithm [49], which flips one spin each time in the low-temperature limit, almost all spins choose to sit in the initial state (i.e., 111111, spin up). The scatter plot will simply have one group. However, this is not wrong because the state still obeys the Boltzmann distribution. This is the reason for small ch at low temperatures of using the Metropolis algorithm to generate configurations. However, we can still observe the signature at the phase-transition point. Using the Swendsen-Wang algorithm instead [47], i.e., a global updating Monte Carlo method, the spin states will all be spin up or spin down in lower temperatures, which is not dependent of the initial state.

IV. THE 2D XY MODEL

The Hamiltonian of the classical XY model is given by

$$H = -J \sum \cos(\theta_i - \theta_j),$$

where $\langle i, j \rangle$ denotes a nearest-neighbor pair of sites i and j , and θ in $(0, 2\pi]$ is a classical variable defined at each site describing the angles of spin directions in a two-dimensional spin plane. The sum in the Hamiltonian is over nearest-neighbor pairs on the square lattice ($L \times L$) with the periodic boundary condition; other lattices can be also considered.

A. The 2D XY model on square lattices

Now we analyze the first example, i.e., the two-dimensional XY model on the square lattice. First, we generate the configurations with five fixed winding number pairs at $\nu = (\nu_x, \nu_y) = (0, 0), (0, 1), (1, 0), (0, -1),$ and $(-1, 0)$. For each fixed winding number pair, it should be noted that, for the two-dimensional geometry, the winding number component $\nu_x = 1$ means that the spins in each row form a winding number of 1 rather than just the spins on one row randomly selected. Cluster simulation algorithms [47,50] are not suitable to update the spins because the global flips break the topological winding number easily and, therefore, the Metropolis algorithm is used while trying to rotate the spin vector with a very small step each time so as to preserve the winding number.

For each topological sector, we generate $m = 500$ configurations. Combining all configurations $\{x_l, l = 1, \dots, 5m\}$ from all five sectors, we construct the kernel K_ε . The elements between $K_\varepsilon(x_l, x_{l'})$ are defined in Eq. (1) and the normalized matrix $P_\varepsilon(x_l, x_{l'})$ is obtained, obeying its eigenvalue equation $P\Psi_k = \lambda_k \Psi_k$. Using the scatter plot of the second and third leading eigenvectors Ψ_1 and Ψ_2 of P_ε , the ch indices are obtained and displayed in Figs. 3(a)–3(d).

There is a parameter ε in the above matrices and in order to deduce consistent results, we need to make sure the results are converging for a finite range of $\varepsilon/\varepsilon_0$. We see that this is indeed the case for different values of $\varepsilon/\varepsilon_0$ in Figs. 3(a)–3(d). For $\varepsilon/\varepsilon_0 = 2.5, 3, 3.5,$ and 4 , T_c is less than 0.9, and then becomes 0.93 when $\varepsilon/\varepsilon_0 = 4.5, 5, 5.5, 6, 6.5,$ and 7 . Finally, the peaks move left and deviate from 0.93 again when $\varepsilon/\varepsilon_0 = 8$ or larger. It should be noted that here 10 000 or more Monte Carlo steps are used in order to reach the equilibrium of the systems.

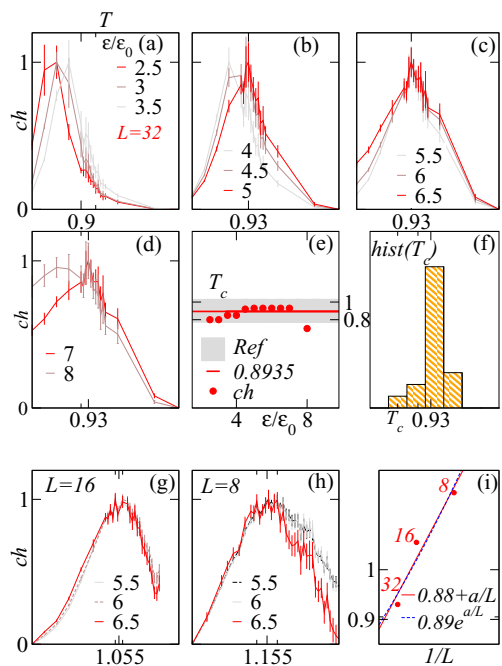


FIG. 3. The ch index for the XY model on a two-dimensional square lattice of size $L = 32$ with various values of $\varepsilon/\varepsilon_0$ (a) 2.5, 3, 3.5, (b) 4, 4.5, 5, (c) 5.5, 6, 6.5, and (d) 7 and 8. The error bars are calculated by using ten or more bins of data. (e) Comparing the T_c vs $\varepsilon/\varepsilon_0$ using ch , Ref. [43], and the latest MC result [51]. (f) Histogram of T_c obtained by using various $\varepsilon/\varepsilon_0$'s. (g)–(i) ch index for the XY model on a two-dimensional square lattice of sizes $L = 8$ and $L = 16$, and the scaling of estimated critical points.

The estimated points are labeled by the circles in Fig. 3(e), and they all distribute nearby or on the red lines representing the latest result critical temperature $T_c = 0.8935$ [51]. The intersections of \bar{D} and $\bar{\sigma}$ are labeled by the gray regions in the critical regimes $T_c = 0.9 \pm 0.1$ from Ref. [43]. It appears that it is easier to use ch to locate the phase transition as we only need to identify the peak location. The results from Ref. [43] have greater uncertainty than those by using the ch index.

The histogram of our estimated T_c is shown in Fig. 3(f), which helps to determine the hyperparameter $\varepsilon/\varepsilon_0$ (see Sec. VI).

In Figs. 3(g)–3(i), the finite-size effect of T_c is checked using the ch with two smaller sizes $L = 8$ and $L = 16$ with $\varepsilon/\varepsilon_0 = 5.5, 6$, and 6.5 . Combining the estimated $T_c(L)$ with $L = 8, 16, 32$, we use two different ways of (linear and exponential) extrapolation to get $T_c(\infty)$ in the thermodynamic limit. The results are $0.88(5)$ and $0.89(5)$, respectively.

B. The 2D XY model on honeycomb lattices

Here, we study the second example, i.e., the pure XY model on the honeycomb lattice as the critical point is known exactly at $T_c = 1/\sqrt{2} \approx 0.707$ [52].

The geometry of the honeycomb lattice is equivalent to the 8×8 brick-wall lattice shown in Fig. 4 where every spin has three nearest-neighbor spins. To initialize the configurations with a fixed winding number $(\nu_x, \nu_y) = (0, 1)$, the spins connected by solid gray lines are defined as forming $\nu_y = 1$ in the

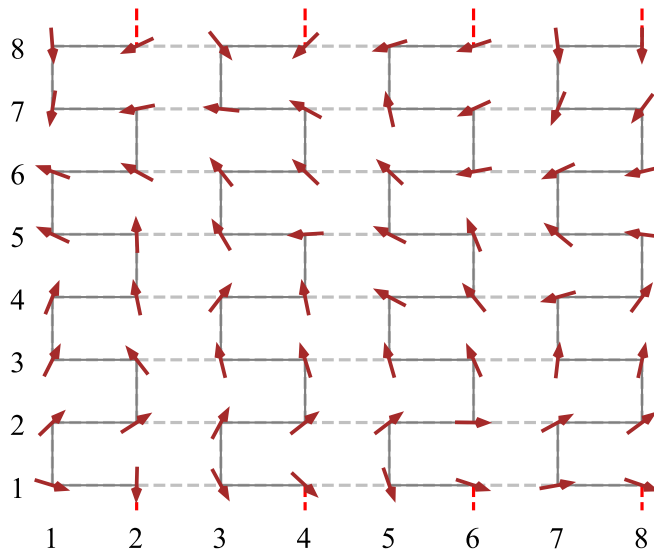


FIG. 4. An 8×8 honeycomb lattice equivalent to a brick-wall lattice. The gray solid line connects the spin forming the winding number in the y direction, i.e., $(\nu_x, \nu_y) = (1, 0)$.

vertical direction. Specifically, if we start a spin at position $(2,1)$ and then go left to $\rightarrow (1, 1) \rightarrow (1, 2) \cdots (2, 8)$ and go back to $(2,1)$ through the red dashed lines, connecting the sites at the boundaries for periodic boundary conditions, the spins sweep an angle of 2π counterclockwise.

It should be noted that two spins, such as $(1,1)$ and $(2,1)$, connected by the horizontal gray lines will contribute to the winding number ν_y , and they contribute to ν_x . This poses a problem when fixing ν_x and ν_y independently. Fortunately, this problem can be solved. Specifically, in the first row, labeled by 1 in the vertical (y) direction, the relation of angles obeys $\theta_{(2,1)} - \theta_{(1,1)} = -(\theta_{(1,1)} - \theta_{(2,1)}) \approx -(\theta_{(3,1)} - \theta_{(2,1)})$. During the simulation, small fluctuations are allowed if they do not break the winding number.

In Fig. 5, using configurations constrained in five topological sectors on the 32×32 lattices, we find that the peaks are located at the exact value of $T_c = \frac{1}{\sqrt{2}} \approx 0.7$ for several different values of $\varepsilon/\varepsilon_0$ in the interval $[2,6]$ with intervals of 0.5. We also calculate the intersections by $\Delta\lambda$ and σ_λ at $\varepsilon/\varepsilon_0 = 5-7$. The intersection can also arrive at 0.7 when $\varepsilon/\varepsilon_0 = 7$ but not 5 and 6. It indicates that the ch performs better than the intersection as the intersection method may not give high confidence for the transition.

V. THE 2D GENERALIZED XY MODELS

Here, we apply our method to the generalized XY (GXY) models [53,54], whose Hamiltonian is given by

$$H = - \sum [\Delta \cos(\theta_i - \theta_j) + (1 - \Delta) \cos(q\theta_i - \theta_j)],$$

where Δ is the relative weight of the pure XY model and q is another integer parameter that can drive the system to form a nematic phase. For both $\Delta = 0$ and $\Delta = 1$ the model reduces to the pure XY model (redefining q as 1 in the first case), and hence, the transition temperature is identical to that of the pure XY model. However, such a redefinition is not possible with

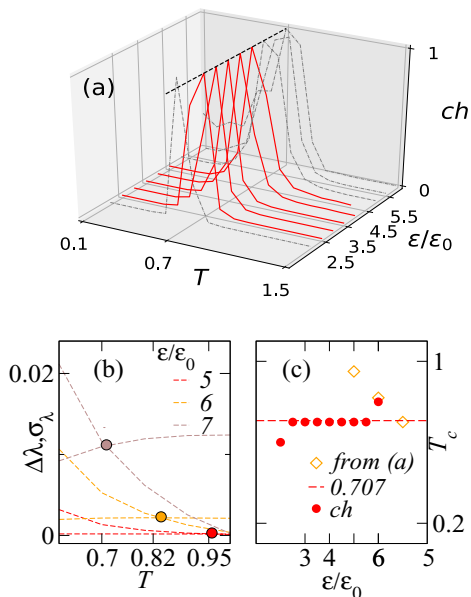


FIG. 5. Locating the phase-transition points of the pure XY model on the honeycomb lattices (a) ch index, (b) the intersections by $\Delta\lambda, \sigma_\lambda$, and (c) comparison results between (a) and (b) and the exact result $\sqrt{2}/2$.

$\Delta \neq 1$. The phase diagrams of the GXY models depend on the integer parameter q , and they have been explored extensively.

A. $q = 2$ GXY model

The $q = 2$ GXY model has a richer phase diagram than the XY model and has an additional new phase [9]. Thus, it is a good candidate model to test our method away from the pure XY limit. The phase diagram is illustrated in Fig. 6(a) on the Δ - T plane, and we show the results from our unsupervised method and those from PCA as comparison. The symbols N, F , and P represent nematic, ferromagnetic, and paramagnetic phases, respectively. The dashed lines are data from the MC simulations [53] mainly of $L = 50$ up to $L = 300$. The color indicates the value of the ch index obtained from simulation with the system size $L = 32$. We now discuss the F - P , N - F , and N - P transitions as follows.

(i) The F - P phase transition: Interestingly, the positions of the peaks by ch are consistent with the dashed line of the phase boundary in the whole region $\Delta > 0.4$. The index ch performs very well where Δ is away from the pure XY model limit. The essential nature of the F - P phase transition is still BKT.

(ii) The N - P phase transition: In the regime $\Delta = 0$, using integer-vortex constraint the ch peaks around $T = 0.7$, which is 0.2 less than 0.9. This discrepancy is likely due to the nature of half-vortex in the N phase. We can improve the result by limiting the configurations to have the half-winding number $\nu_{x(y)} = 1/2$ as the topological constraint in our Monte Carlo simulation. The half-vortex physics has been discussed in Ref. [53].

Thus, we only consider (ν_x, ν_y) in the four types of combinations $(\pm 1/2, 0)$ and $(0, \pm 1/2)$. To form $(\nu_x = 1/2, 0)$, the difference between a pair of spins located at the left-most and right-most boundaries is fixed as π , and we assign each spin

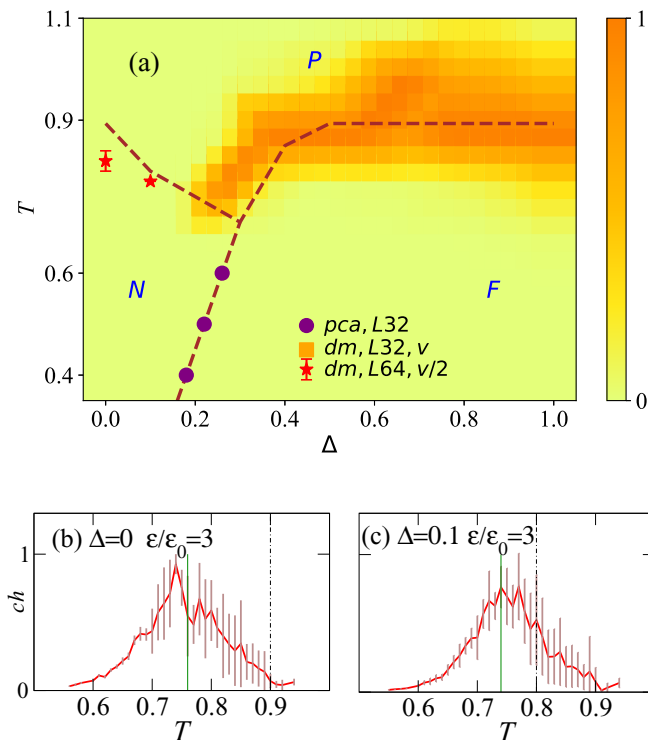


FIG. 6. (a) The phase diagram of the $q = 2$ GXY model, containing N, F , and P by calculating the ch index. The dashed lines are from Ref. [53] (b) N - P transition at fixed $\Delta = 0$: ch vs T . (c) N - P transition $\Delta = 0.1$: ch vs T . The peaks are closer to the known values using the half-vortex constraint with L up to 64 when compared with those using only integer vortices.

using Eq. (A2). With the half-vortex constraint, our results illustrated by the red symbols move closer to the dashed lines of the N - P transition in Fig. 6(a) than the results using the integer-value constraint.

Figures 6(b) and 6(c) show the details at $\Delta = 0$ and $\Delta = 0.1$, respectively. For $L = 64$, we generate two groups of data, the peaks almost converge in the interval $[0.82, 0.85]$, and the convergence is closer to 0.89 than the result from the integer-vortex constraint. For $\Delta = 0.1$, the half-vortex constraint gives better results at $T_c = 0.8$.

(iii) The N - F phase transition: When we realize that the topological constraint makes the peaks (features) of the distribution for the spin directions implicit in the F and P phases, we use the configurations without any constraint in this case. The results are labeled by purple symbols with the legend “pca, L32.”

The behavior of ch depends on the structure of the sample points, and sometimes ch emerges as a sharp jump at the phase-transition point, such as in the Ising model discussed in the main text. Sometimes it is a local minimum value at the phase-transition point. Take the $q = 2$ GXY model at $T = 0.5$ as an example in Fig. 7, both $\ln(ch_t)$ and $\ln(ch_b)$ increase as a function of Δ . However, the difference $\alpha = \ln(ch_t) - \ln(ch_b) = -\text{gap}$ decreases first and then increases around $\Delta_2 = 0.22$ where the gap has a positive value.

According to the following equation:

$$ch = \exp[\ln(ch_t) - \ln(ch_b)] = \exp(-\text{gap}), \quad (10)$$

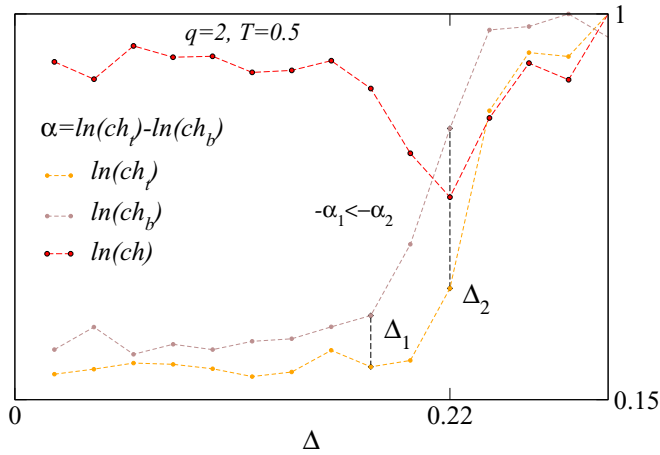


FIG. 7. For the $q = 2$ GXY model, ch , ch_t , and ch_b on the logarithmic-scale vs Δ without any constraint. ch is obtained by applying k-PCA and k -means.

clearly, $\min(ch) \Rightarrow \max(\text{gap})$ and the local minimum of the ch is at the location of Δ_2 .

For the N - F phase transition in Fig. 6(a), the PCA is stronger than the DM method. Here, we would like to give a physical discussion about this because such a difference is related to the advantage of the proposed method, and, thus, a more detailed discussion should be made.

The reasons for PCA being stronger for the N - F transition can be explained as follows:

(i) The N - F phase transition is not a topological phase transition [54]. The DM method designed here is to determine the topological phase transition. It is still interesting to see the ch of the DM by inputting Ising configurations without any topological constraint. In Fig. 8, using the k-PCA and the DM method with complete same configurations, the signatures of phase transition emerge and the results are consistent at $T_c = 2.3$. However, the signature of the k-PCA method is clearer.

(ii) From the view of the data, it is also understandable that PCA is stronger. Without any topological constraint in the nematic (N) phase the spins prefer two dominant directions, and their histograms obey a double peaks structure. In the ferromagnetic (F) phase, one main direction remains, and one peak emerges for the histograms. Figures 9(b1)–9(b3)

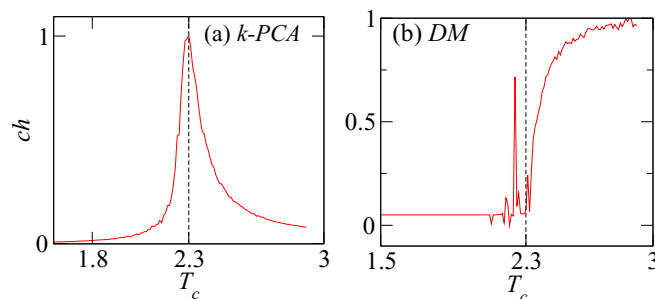


FIG. 8. The value of ch_b obtained by the k-PCA and the DM model for the complete same configurations of the 64×64 Ising model without any constraint.

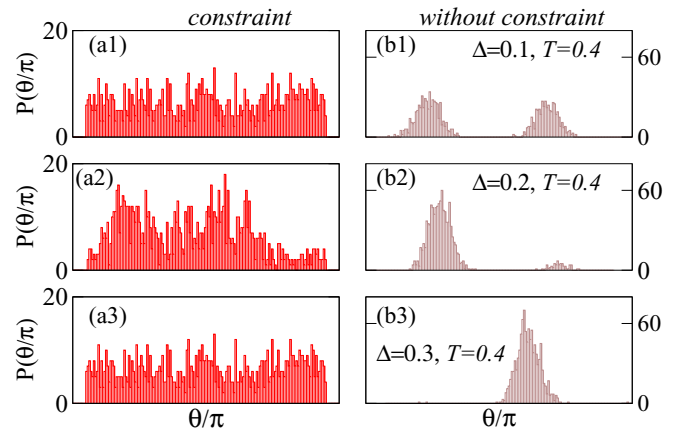


FIG. 9. In the N ($\Delta = 0.1$) and F ($\Delta = 0.3$) phases and critical points $\Delta = 0.2$, the distributions of spin vector of the $q = 2$ GXY model with (a1)–(a3) and without (b1)–(b3) constraints, respectively.

show the typical histograms in three phases. The main feature difference between the phases is obvious.

However, with a topological constraint, the spin directions are mainly distributed according to the winding numbers. For example, the spin angles in each row are $\theta_x = 2\pi x/L$ with additional fluctuations, where x and L are the number index and total number, respectively, in each row. Therefore, the distribution sits almost in the range $[0, 2\pi]$ as shown in Figs. 9(a1)–9(a3). The feature difference of spin angles disappears when using the constraint. Therefore, the DM method cannot get transition points using the configurations with constraints.

B. $q = 8$ GXY model

For the $q = 8$ GXY model [54], Fig. 10(a) shows the global phase diagram using the values of the ch index. The phases N , F_2 , F , and P are obtained, and the distributions are shown.

In the new F_2 phase, the distribution of the spin vectors has eight peaks but is dominated by four possible angles. The “X” shape dashed lines are from Refs. [53,54]. The orange color represents the values of ch by the DM method.

Let us first discuss the $F(F_2)$ - P transition. Clearly, for the F - P transition in the range $0.5 < \Delta < 1$, the peak positions of ch align well with the dashed line. In the center of X, the F_2 and P transitions are also consistent with MC result, i.e., $T_c = 0.5$.

However, we could not use the intersection of the cluster average distance \bar{D} and within-cluster dispersion $\bar{\sigma}$ as described in Ref. [43] because \bar{D} does not vary too much and there is no intersection as shown in Figs. 10(e)–10(g). The five different colors therein represent the five typical topological sectors. However, fortunately, when zooming in the figures in Figs. 10(g)–10(i), we find that, near the transition region, the shape of a fixed cluster shrinks because the data points gather closer together, and, hence, the within-cluster dispersion $\bar{\sigma}$ is smaller. The index $ch = \frac{ch_t}{ch_b}$, thus, develops a peak around $T_c = 0.5$ as shown in Fig. 10(c) with ch_t and ch_b also displayed in Fig. 10(d). In contrast, Fig. 10(b) shows that sc cannot be used to signal the transition temperature T_c because

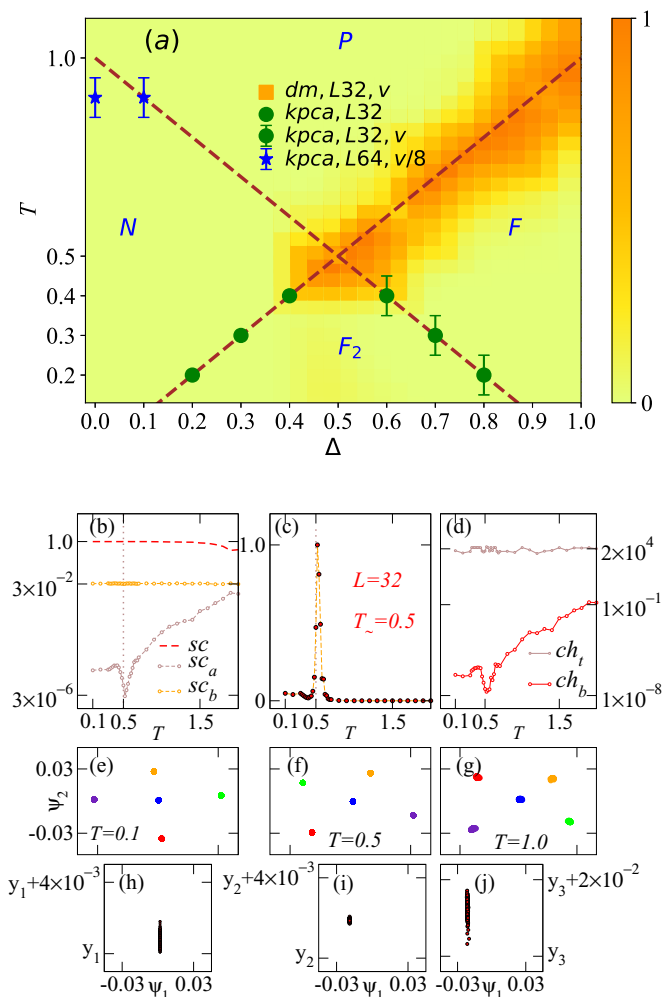


FIG. 10. (a) Phase diagram (with N , F , F_2 , and P phases) of the $q = 8$ GXY model and results from calculating the ch index, whose values are represented by colors. The dashed blue lines are MC results [54]. (b) The curves of sc , sc_a , sc_b vs T at $\Delta = 0.5$. (c) and (d) The curves of ch , ch_1 , and ch_b vs T at $\Delta = 0.5$ (and note that $T_c = 0.5$). (e)–(g) Scatter plot of Ψ_1 and Ψ_2 at $T = 0.1, 0.5$, and 1.0 . (h)–(j) Zoom-in view for a fixed group of data.

it evaluates the difference $sc_b - sc_a$, but $sc_a \ll sc_b$ despite the fact that sc_a has a local minimum.

It should be mentioned that for the N - P transition, the use of either the integer or the half-vortex constraint is not suitable. Instead the $v_{x(y)} = 1/8$ vortex constraint is needed in generating the configurations. Moreover, after comparison, we find that k-PCA works better if we use θ_i as the input into the k-PCA (using PCA does not work well). The kernel used for k-PCA defined as a radial basis functional kernel $\exp(-\gamma \|x/L - y/L\|^2)$, where $L = 32$ is the system size, x and y are the configurations $\{\theta_i\}$, and $\gamma = 1$ is the default value [55].

The other details of the F - P , F_2 - F , N - F_2 , and N - P transitions will be discussed as follows. For the $q = 8$ GXY model, Fig. 11 shows the distributions of spin vectors in the (a1) N , (a2) F_2 , and (a3) F phases with the integer-vortex constraint.

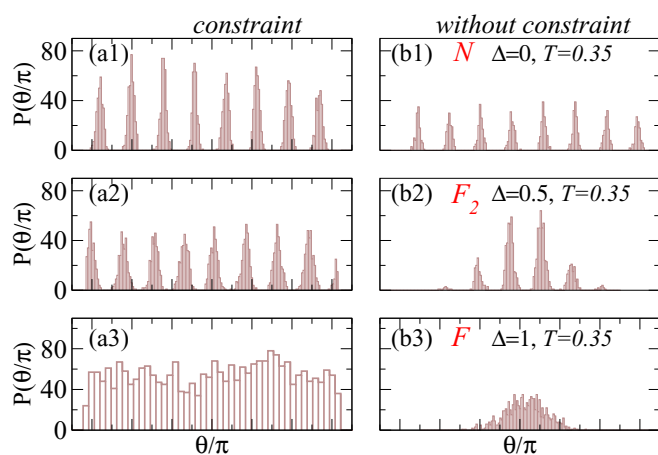


FIG. 11. In the N , F_2 , and F phases, the distributions of spin vector of the $q = 8$ GXY model with (a1)–(a3) and without (b1)–(b3) constraints, respectively.

the constraints are canceled, and the distributions are shown in Figs. 11(b1)–11(b3), respectively.

Figures 12(a)–12(d) show the detail of phase-transition F - P , F_2 - F , and N - F_2 , respectively. For the F - P transition, in Fig. 12(a), fixing Δ in the interval $[0.5, 1]$ at steps of 0.1 , the peaks of ch are located at 0.5 and 1 , respectively, completely matching the dashed lines in the global phase diagram in Fig. 10.

In Fig. 12(b), for the F_2 - F transition, by fixing $\Delta = 0.8$, the peaks of ch are located at 0.2 with $L = 8, 16, 32, 48$, and 64 . The other values of Δ are not shown. In Fig. 12(c), fixing $T = 0.2, 0.3$, and 0.4 , the positions Δ of the local minimum of ch are located at $0.2, 0.3$, and 0.4 , respectively.

In Fig. 12(d), for $\Delta = 0$, using the $1/8$ vortex constraint results in the most accurate critical points. The peak position is at 0.9 more closely to 1 than 0.7 by the integer-vortex constraint. The error bars are obtained by the bootstrap method using 400 randomly chosen configurations between the total 2000 configurations and the total of 20 bins.

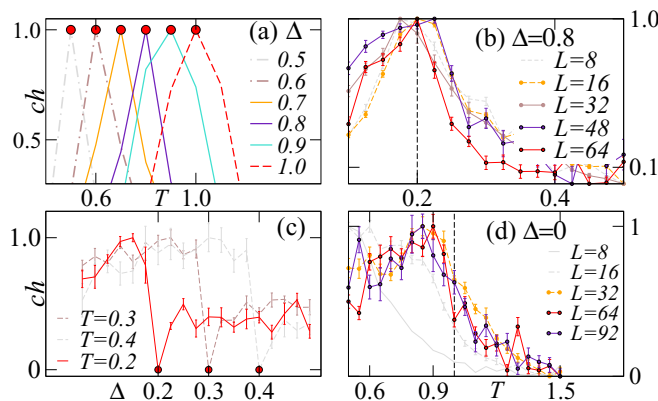


FIG. 12. (a) F - P : ch vs T at fixed Δ in the interval $[0.5, 1]$. The peaks are at $0.5, 0.6, 0.7, 0.8, 0.9$, and 1 , respectively. (b) F_2 - F : ch vs T , and T_c is 0.2 for $\Delta = 0.8$. (c) The numerator ch_1 and denominator ch_b of the ch index in Eq. (6). (d) N - F_2 : ch vs Δ for $T = 0.2, 0.3$, and 0.4 . (d) N - P : using the $1/8$ vortex constraint and k-PCA, ch vs T , and T_c is about 0.9 .

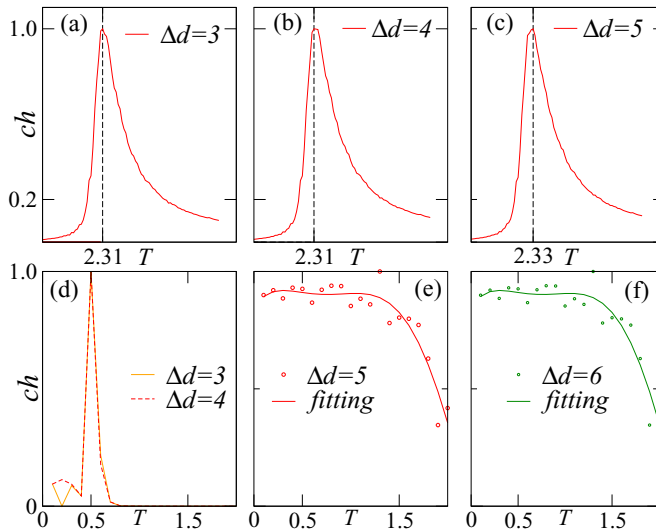


FIG. 13. The ch index as a function of T for various values of Δd (a) $\Delta d = 3$, (b) $\Delta d = 4$, and (c) $\Delta d = 5$ for Ising systems. ch vs T of the GXY model at fixed $\Delta = 0.5$ and varying T with the number of vectors considered (d) $\Delta d = 3$ and 4, (e) $\Delta d = 5$, and (f) $\Delta d = 6$, respectively. The results are best using two vectors $\Delta d = 2$.

VI. OTHER TECHNICAL MODIFICATIONS

In the above sections, during the use of the k -means method, we apply two output eigenvectors of the diffusion matrix P and then get the values of the ch index. The conclusion is that using two leading vectors leads to the best accuracy. At the same time, it is possible to design a way to determine the super parameter $\varepsilon/\varepsilon_0$ automatically. In this section, we will focus on such issues for the completeness of our method.

Here, the effects of higher-dimensional features will be discussed using k -means, namely, whether or not the accuracy is enhanced when including more features will be clarified. The answer is that retaining more dimensions of the eigenvectors, the result may deviate from the right critical points or even predict wrong critical points.

Assuming the covariance matrix of the PCA method is C , and its eigenvectors are $\{\Psi_d\}$, where $d = 1, \dots, d_{max}$ ($\Delta d = d_{max}$). For the Ising model in Sec. III, the first and second vectors $\{\Psi_d\}$ are used, where $d_{max} = 2$, or $\Delta d = 2$. Here, we consider vectors with $\Delta d = 3-5$. The difference is found between the results of two-dimensional vectors and higher-dimensional data as shown in Figs. 13(a)–13(c). The estimated position of the peaks are 2.31, 2.31, and 2.33, respectively. The result $T_c = 2.3$ for $\Delta d = 2$ is the closest to the peak of the specific heat.

For the generalized XY model ($q = 8$, $\Delta = 0.5$) with the DM method, the vectors with index $d = 2, \dots, d_{max}$ ($\Delta d = d_{max} - 1$) are used as the first vector ignored. In Fig. 13(d) with $\Delta d = 3$ or 4, the position of peaks are $T_c = 0.5$ as the same as that of $\Delta d = 2$. However, if higher-dimensional data are used, such as $\Delta d = 5$ and 6, the ch score does not yield the correct signature of the phase transition as shown in Figs. 13(e) and 13(f).

Another issue discussed here is whether one can design an automatic method to determine the hyperparameter $\varepsilon/\varepsilon_0$.

Since its value has been tried many times for some reasonable choices, one may expect to find a way to determine its optimal value automatically.

Here is such an automatic determination of the hyperparameter $\varepsilon/\varepsilon_0$. Here we devise a simple analysis from the statistics (histogram) of T_c obtained by various $\varepsilon/\varepsilon_0$'s:

(i) Calculate $ch(T)$ as a function of temperature T with various $\varepsilon/\varepsilon_0$'s in a range from $\varepsilon^{\min}/\varepsilon_0$ to $\varepsilon^{\max}/\varepsilon_0$ by using the configurations with topological constraints;

(ii) store the position of ch peaks, i.e., T_c and the times they appear as shown in Figs. 3(e) and 3(f).

The histogram of T_c vs $\varepsilon/\varepsilon_0$ can be used to give the best estimate for the transition. Therefore, corresponding to the highest position $T_c = 0.96$, the values of $4 \leq \varepsilon/\varepsilon_0 \leq 6$ are acceptable.

Another possible approach is to calculate a location-dependent σ for each data point instead of selecting a single scaling parameter [56]. Then, the kernel matrix between a pair of points can be written as

$$\omega(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_i \sigma_j}\right), \quad (11)$$

where σ_i and σ_j are the local scale parameters for x_i and x_j , respectively. The selection of the local scale σ_i is determined by the local statistics of the neighborhood of point x_i . For example, the scale can be set as

$$\sigma_i = \|x_i - x_K\|, \quad (12)$$

where x_K is the K th nearest neighbor. However, here, K is also a hyperparameter to be chosen. By comparing Eqs. (11) and (1), $2N\varepsilon = \sigma_i \sigma_j$, the obtained $\varepsilon/\varepsilon_0 \approx 30$ is about several times larger than the acceptable regimes.

VII. DISCUSSION AND CONCLUSION

To summarize, we use the ch index to successfully locate phase transitions of a few classical statistical models, including the Ising, XY, and the generalized XY models. From the scatter data for which we use the ch index to obtain the phase transitions for the Ising, XY, and GXY models on lattices. This combines the advantages of the PCA and DM methods as the scattering data can be based on either the first two leading eigenvectors of the PCA, kernel-PCA, or the second and third vectors from the DM method.

The advantages of using the ch index are less steps, wider applications, and better convergence. After k -means applied to the eigenvectors of the diffusion matrix P , using the ch index, we do not need to maximize the visibility of cluster as proposed in the RNS method. For some phase transitions, it may not be easy to find the intersections of \bar{D} and $\bar{\sigma}$. In Figs. 10(e)–10(g), the ch index could capture the signatures of both quantities. In Fig. 5, for the pure XY model on the honeycomb lattices, the exact critical point is at $T_c = 0.707$. Using ch when $\varepsilon/\varepsilon_0$ is in the interval [2,6], the estimated results of T_c are all close to 0.7.

In the pure XY limit, we have tested that ch can locate the phase transition for the XY model on both the square and the honeycomb lattices, similar to the DM method of Rodriguez-Nieva and Scheurer. For the $q = 2$ GXY model, the values of the ch index in the whole phase diagram matched the

TABLE I. Internal clustering validation measures.

	Measure	Notation	Definition
1	Calinski-Harabasz index	ch	$\frac{\sum_i n_i d^2(c_i, c)/(NC - 1)}{\sum_i \sum_{x \in C_i} d^2(x, c_i)/(n - NC)}$
2	Silhouette index	sc	$\left\{ \frac{1}{NC} \sum_i \left[\frac{1}{n_i} \sum_{x \in C_i} \frac{b(i) - a(i)}{\max[b(i), a(i)]} \right] \right\}$ $a(x) = \frac{1}{n_i - 1} \sum_{y \in C_i, y \neq x} d(x, y)$ $b(x) = \min_{j, j \neq i} \left[\frac{1}{n_j} \sum_{y \in C_j} d(x, y) \right]$
3	Davies-Bouldin index	db	$\left\{ \frac{1}{NC} \sum_i \max_{j, j \neq i} \frac{\frac{1}{n_i} \sum_{x \in C_i} d(x, c_i) + \frac{1}{n_j} \sum_{x \in C_j} d(x, c_j)}{d(c_i, c_j)} \right\}$
4	SDBw validity index	SDBw	$\{Scat(NC) + Dens_bw(NC)\}$ $\sum_i \ \sigma(C_i)\ $ $scat(NC) = \frac{1}{NC} \frac{\sum_i \ \sigma(D)\ }{\ \sigma(D)\ }$ $Dens_bw(N) = \frac{\sum_{x \in (C_i \cup C_j)} f(x, u_{ij})}{\sum_{x \in C_i} f(x, c_i) + \sum_{x \in C_j} f(x, c_j)}$ $= \frac{1}{NC(NC-1)} \sum_i \left[\frac{\sum_{x \in C_i} d^2(x, c_i)}{\max \left\{ \sum_{x \in C_i} f(x, c_i), \sum_{x \in C_j} f(x, c_j) \right\}} \right]$
5	Xie-Beni index	xb	$\left\{ \frac{\sum_i \sum_{x \in C_i} d^2(x, c_i)}{n \min_{i, j \neq i} d^2(c_i, c_j)} \right\}$
6	Dunn's indices	dn	$\left\{ \min_i \left[\min_j \left(\frac{\min_{x \in C_i, y \in C_j} d(x, y)}{\max_k \left\{ \max_{x, y \in C_k} d(x, y) \right\}} \right) \right] \right\}$
7	pbm	pbm	$\left\{ \frac{1}{K} \frac{\sum_{i=1}^K d(x_i, y_j)}{N} - \max_{i, j=1, 2, \dots, K} d(v_i, y_j) \right\}$
8	I index	Ii	$\left\{ \left[\frac{1}{NC} \frac{\sum_{x \in D} d(x, c)}{\sum_i \sum_{x \in C_i} d(x, c_i)} \max_{i, j} d(c_i, c_j) \right] \right\}$
9	Root-mean-square std dev	rmss	$\left\{ \left[\frac{\sum_i \sum_{x \in C_i} \ x - c_i\ ^2}{P \sum_i (n_i - 1)} \right]^{1/2} \right\}$
10	R^2	r^2	$\left\{ 1 - \frac{\sum_i \sum_{x \in C_i} \ x - c_i\ ^2}{\sum_{x \in D} \ x - c\ ^2} \right\}$
11	Modified Hubert ^a Γ statistic	mhgs	$\left\{ \frac{2}{n(n-1)} \sum_{x \in D} \sum_{y \in D} d(x, y) d_{x \in C_i, y \in C_j}(c_i, c_j) \right\}$

^a D : dataset; n : number of objects in D ; c : center of D ; P : attributes number of D ; NC : number of clusters; C_i : the i th cluster; n_i : number of objects in C_i ; c_i : center of C_i ; $\sigma(C_i)$: variance vector of C_i ; $d(x, y)$: distance between x and y ; $\|X_i\| = (X_i^T X_i)^{1/2}$.

boundary between the ferromagnetic phase to the paramagnetic phase very well even away from the pure XY limit. Close to the N - P transition, the accuracy can be improved by using the 1/2-vortex constraint in generating the Monte Carlo configurations. For the $q = 8$ GXY model, using intersections of \bar{D} and $\bar{\sigma}$ cannot be used to locate the transition point, such as at $\Delta = 0.5$, but the ch index can still work to locate the phase-transition point due to its incorporation of the fluctuation of

samples within each cluster. Moreover, the N - P transition can also be identified by using the 1/8-vortex constraint.

We have also compared the results from other indices (see Table I), but we find that the ch index works best overall for the models considered in this paper. In the future, it will be desirable to systematically study the utility of various parameters in models of statistical physics that exhibit different natures of transitions.

The development of applying machine learning to physics could develop new methods for studying unknown phases. For the Ising model or similar models, the transition point is well known. This situation will help to check our proposed method.

If we get the first-hand data from experiments and do not know the details of the Hamiltonian, our unsupervised method can help deduce the number of possible classes (phases) in the data. Furthermore, using the ch index we could identify phase-transition points. Therefore, the method of using the ch index is very useful for future nontopological phase transitions. For the XY model and Hamiltonians similar to the GXY model by using the ch index, topological phase-transition points were obtained without using the traditional method of measuring correlations [54] and spin stiffness [51]. Of course, some reasonable prior knowledge is needed, such as the possible winding numbers. The idea of DM can be applied to topological quantum systems (see Refs. [57,58]). In principle, the topological quantum phases and transitions between them may be probed using the ch or sc indices.

Note added. After we finished the revised version of the paper, we found the same topics by other methods [59–61].

ACKNOWLEDGMENTS

We acknowledge the valuable discussion with J. F. Rodriguez-Nieva, M. S. Scheurer, Y. Z. You, and T. Chotibut on simulations. We also thank T. C. Scott for his help. T.-C.W. was supported by the National Science Foundation under Grant No. PHY 1915165 and J.W. and H.T. were supported by NSFC under Grant No. 51901152. W.Z. was supported from Science Foundation for Youths of Shanxi Province under Grant No. 201901D211082.

APPENDIX A: THE 1D XY MODEL

The Hamiltonian of the pure XY model reads

$$H = -J \sum_{i=1}^N \cos(\theta_i - \theta_{i+1}), \quad (\text{A1})$$

where J is the coupling strength of the nearest-neighboring pair of spins $\langle i, i+1 \rangle$ and throughout this paper, we set $J = 1$ for simplicity and use the periodic boundary condition $\theta_{N+1} = \theta$. For simplicity following Ref. [43], we generate the spin vectors according to the following:

$$\theta_i^{(l)} = 2\pi\nu^{(l)}i/N + \delta\theta_i^{(l)} + \bar{\theta}^{(l)}, \quad (\text{A2})$$

where ν is the winding number and l is the identification number of the configuration $\{\theta_i\}$ with i varying from 1 to the total length N . The first term $2\pi\nu^{(l)}i/N$ is used to define the winding number $\nu = \sum_i \Delta_i/2\pi$ where Δ_i is in the range $[-\pi, \pi)$ by the so-called saw function [7] obtained by replacing Δ_i with $\Delta_i \pm 2\pi$ if it is not in the target range. The term $\delta\theta_i^{(l)}$ obeys the Gaussian fluctuation, and $\bar{\theta}^{(l)}$ is generated randomly between $[0, 2\pi)$.

We consider two types of generated configurations with winding numbers $\nu = 0$ and $\nu = 1$. The histogram of the values of the first component of the diffusion map ψ_1 is shown in Fig. 14(a). The histogram of ψ_1 has values at ± 1 , and ψ is a vector with size $m \times 1$, therefore, the values are equal

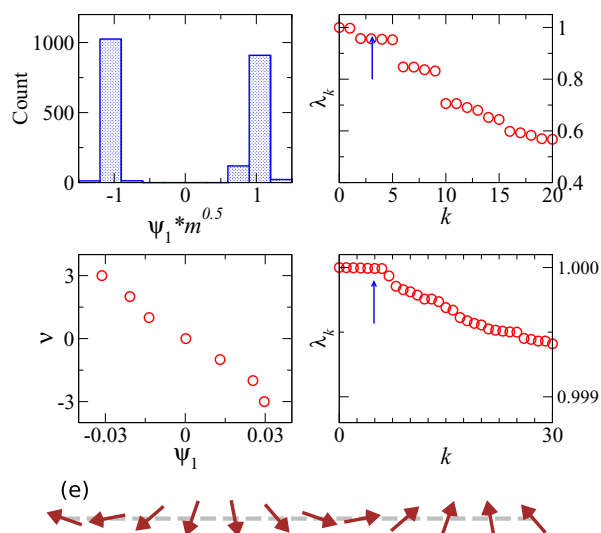


FIG. 14. (a) Histogram count of $\psi_1\sqrt{m}$. (b) The largest 20 eigenvalues of matrix P , and (c) ν vs ψ_1 for the configurations with seven topological sectors. (d) The largest 30 eigenvalues of matrix P . Clearly in (b) and (d), the leading two and seven eigenvalues occur, respectively.

to ± 1 when ψ_1 is rescaled by \sqrt{m} . Figure 14(b) shows the largest 20 eigenvalues of the transition probability matrix $P_{i,i'}$ in Eq. (2). Two maximum eigenvalues are found equal to unity. Following Ref. [43], we also test $\nu = 7$ according to the following equation:

$$\theta_i^{(l)} = 2\pi\nu^{(l)}i/N + \delta\theta_i^{(l)} + \bar{\theta}^{(l)} + \eta^{(l)}[1 - \cos(2\pi i/N)], \quad (\text{A3})$$

where $\nu = \{0, \pm 1, \pm 2, \pm 3\}$. We find that ψ_1 has seven distinct values ranging from -0.03 to 0.03 , corresponding to seven winding numbers marked by the symbols in Fig. 14(c). Eigen-

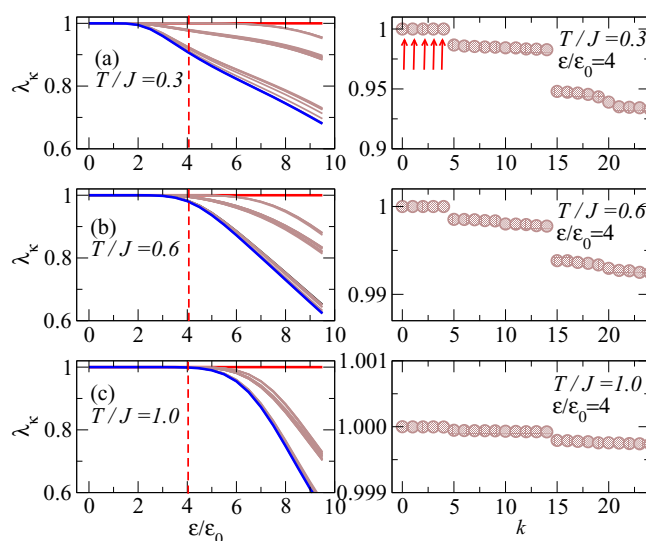


FIG. 15. The distributions of λ_k vs $\varepsilon/\varepsilon_0$ of the 2D XY model for three typical temperatures, (a) $T = 0.3$, (b) $T = 0.6$, and (c) $T = 1.0$. The right column: the corresponding λ_k vs k . The five arrows point to the eigenvalues of the leading five topological sectors.

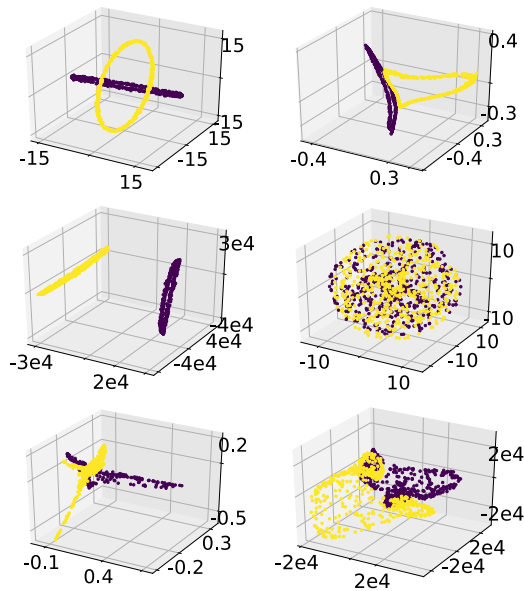


FIG. 16. PCA and kernel PCA with Gaussian and polynomial kernels on a simple dataset. PCA and kernel PCA with Gaussian and polynomial kernels for the generalized dataset.

values λ_k vs k are also shown in Fig. 14(d). Clearly, the plateau of eigenvalues appears in $k \leq 7$.

Figures 15(a)–15(c) show the largest 24 eigenvalues $\lambda_{k \leq 24}$ of $P_{l,l}$ as a function of $\varepsilon/\varepsilon_0$ for $T = 0.3, 0.6$, and 1.0 , respectively. The band of eigenvalues λ_k could not be distinguished for small values of $\varepsilon/\varepsilon_0$, and the reason can be seen from the matrix of P , which is a diagonal matrix in that limit. Increasing $\varepsilon/\varepsilon_0$, the bands of λ_k will separate away from each other.

The choice of $\varepsilon/\varepsilon_0$ is important. We choose $\varepsilon/\varepsilon_0 = 4$ marked by the dashed lines in the left column. The right column presents λ_k vs k ($k = 0, 1, \dots$). Clearly, the gap between the $k = 4$ eigenvalue and the $k = 5$ eigenvalue becomes smaller when increasing temperature and, subsequently, disappears when $T > T_c$.

APPENDIX B: THE PCA AND THE k-PCA

Figure 16 represents some results of PCA and k-PCA on simple and generalized datasets in the Appendix of Ref. [43]. The simple datasets are configurations of the one-dimensional XY model with size $N = 256$, $m = 1000$, and $\sigma_\theta = 0.3$ and

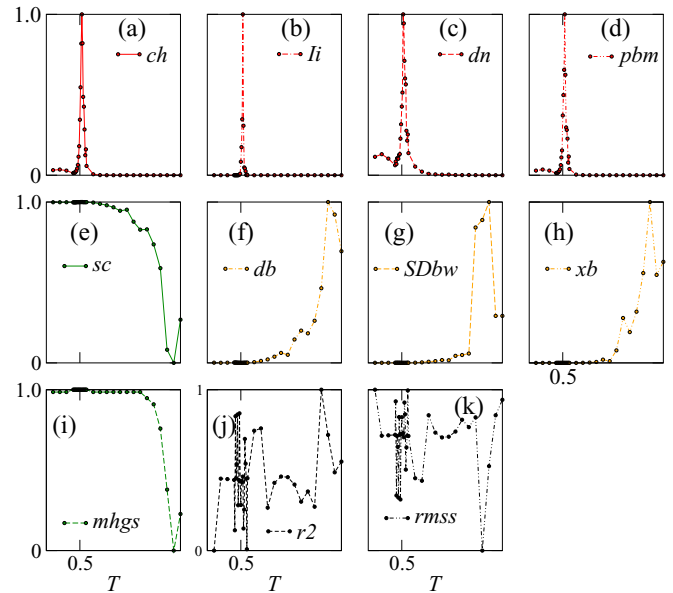


FIG. 17. In the first row (a)–(d), the indices, called ch , li , dn , and pbm , can distinguish the F_2 and P phases. (e)–(k) The remaining indices fail to present signals.

with topological winding numbers $\nu = 0$ and $\nu = 1$ according to Eq. (A2). Figure 16 shows results of linear PCA, PCA with Gaussian kernels, and PCA with polynomial kernels, respectively (the first three panels). Obviously, the classifications of PCA with a nonlinear kernel are much clearer for the XY models. The three-dimensional visualization is based on the three reduced components. However, the above method fails for data generated with slight modification of Eq. (A2) with an additional term $\eta^{(l)}[1 - \cos(2\pi i/N)]$, where $\eta^{(l)}$ is random in the range $[-\eta_0, \eta_0]$. The results are shown in Fig. 16 at the last three panels, respectively.

APPENDIX C: OTHER INDICES

As listed in Ref. [48], we check the 11 indices listed in Table I for validating the classifications. Take the $q = 8$ GXY model as an example, the value of the parameters, such as the temperature T and Δ are fixed as those in Fig. 10(b). We find only four indices, presented in Figs. 17(a)–17(d) produce signals at the critical points. These are ch , li , and pbm , whose full names are listed in Table I.

The other indices could not give correct signals to locate the phase-transition points.

- [1] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, New York, 1971).
- [2] S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge, UK, 1999).
- [3] J. Carrasquilla and R. Melko, Machine learning phases of matter, *Nat. Phys.* **13**, 431 (2017).

- [4] K. I. Aoki and T. Kobayashi, Restricted Boltzmann machines for the long range Ising models, *Mod. Phys. Lett. B* **30**, 1650401 (2016).
- [5] D. Kim and D. Kim, Smallest neural network to learn the Ising criticality, *Phys. Rev. E* **98**, 022138 (2018).
- [6] W. Hu, R. R. P. Singh, and R. T. Scalettar, Discovering phases, phase transitions, and crossovers through unsupervised machine

- learning: A critical examination, *Phys. Rev. E* **95**, 062122 (2017).
- [7] M. Beach, A. Golubeva, and R. Melko, Machine learning vortices at the Kosterlitz-Thouless transition, *Phys. Rev. B* **97**, 045207 (2018).
- [8] P. Suchsland and S. Wessel, Parameter diagnostics of phases and phase transition learning by neural networks, *Phys. Rev. B* **97**, 174435 (2018).
- [9] W. Zhang, J. Liu, and T.-C. Wei, Machine learning of phase transitions in the percolation and XY models, *Phys. Rev. E* **99**, 032142 (2019).
- [10] J. M. Kosterlitz and D. J. Thouless, Ordering, metastability and phase transitions in two-dimensional systems, *J. Phys. C* **6**, 1181 (1973).
- [11] V. L. Berezinskiĭ, Destruction of long-range order in one-dimensional and two-dimensional systems possessing a continuous symmetry group. II. Quantum systems, *Sov. J. Exp. Theor. Phys.* **34**, 610 (1972).
- [12] V. Singh and J. Han, Application of machine learning to two-dimensional Dzyaloshinskii-Moriya ferromagnets, *Phys. Rev. B* **99**, 174426 (2019).
- [13] I. A. Iakovlev, O. M. Sotnikov, and V. V. Mazurenko, Supervised learning approach for recognizing magnetic skyrmion phases, *Phys. Rev. B* **98**, 174411 (2018).
- [14] C. Li, D. Tan, and F. Jiang, Applications of neural networks to the studies of phase transitions of two-dimensional Potts models, *Ann. Phys. (NY)* **391**, 312 (2018).
- [15] E. van Nieuwenburg, Y. Liu, and S. Huber, Learning phase transitions by confusion, *Nat. Phys.* **13**, 435 (2017).
- [16] P. Broecker, J. Carrasquilla, R. Melko, and S. Trebst, Machine learning quantum phases of matter beyond the fermion sign problem, *Sci. Rep.* **7**, 8823 (2017).
- [17] X.-Y. Dong, F. Pollmann, and X.-F. Zhang, Machine learning of quantum phase transitions, *Phys. Rev. B* **99**, 121104(R) (2019).
- [18] L. Wang, Discovering Phase Transitions with Unsupervised Learning, *Phys. Rev. B* **94**, 195105 (2016).
- [19] S. J. Wetzel, Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders, *Phys. Rev. E* **96**, 022140 (2017).
- [20] K. Ch'ng, N. Vazquez, and E. Khatami, Unsupervised machine learning account of magnetic transitions in the Hubbard model, *Phys. Rev. E* **97**, 013306 (2018).
- [21] T. C. Scott, M. Therani, and X. M. Wang, Data clustering with quantum mechanics, *Mathematics* **5**, 5 (2017).
- [22] E. van Nieuwenburg, E. Bairey, and G. Refael, Learning phase transitions from dynamics, *Phys. Rev. B* **98**, 060301(R) (2018).
- [23] J. Venderley, V. Khemani, and E. Kim, Machine Learning Out-of-Equilibrium Phases of Matter, *Phys. Rev. Lett.* **120**, 257204 (2018).
- [24] C. Casert, T. Viejira, J. Nys, and J. Ryckebusch, Interpretable machine learning for inferring the phase boundaries in a nonequilibrium system, *Phys. Rev. E* **99**, 023304 (2019).
- [25] P. Huembeli, A. Dauphin, and P. Wittek, Identifying quantum phase transitions with adversarial neural networks, *Phys. Rev. B* **97**, 134109 (2018).
- [26] Y. Liu and E. P. L. van Nieuwenburg, Discriminative Cooperative Networks for Detecting Phase Transitions, *Phys. Rev. Lett.* **120**, 176401 (2018).
- [27] S. Efthymiou, M. J. S. Beach, and R. G. Melko, Super-resolving the Ising model with convolutional neural networks, *Phys. Rev. B* **99**, 075113 (2019).
- [28] W. Lian *et al.*, Machine Learning Topological Phases with a Solid-State Quantum Simulator, *Phys. Rev. Lett.* **122**, 210503 (2019).
- [29] A. Bohrdt *et al.*, Classifying snapshots of the doped Hubbard model with machine learning, *Nat. Phys.* **15**, 921 (2019).
- [30] P. Zhang, H. Shen, and H. Zhai, Machine Learning Topological Invariants with Neural Networks, *Phys. Rev. Lett.* **120**, 066401 (2018).
- [31] Y. Zhang, R. Melko, and E. Kim, Machine learning Z2 quantum spin liquids with quasiparticle statistics, *Phys. Rev. B* **96**, 245119 (2017).
- [32] Y. Zhang and E. A. Kim, Quantum Loop Topography for Machine Learning, *Phys. Rev. Lett.* **118**, 216401 (2017).
- [33] Y. Hsu, X. Li, D. Deng, and S. Das Sarma, Machine Learning Many-Body Localization: Search for the Elusive Nonergodic Metal, *Phys. Rev. Lett.* **121**, 245701 (2018).
- [34] Z. Cai and J. Liu, Approximating quantum many-body wave functions using artificial neural networks, *Phys. Rev. B* **97**, 035116 (2018).
- [35] R. Vargas-Hernández, J. Sous, M. Berciu, and R. Krems, Extrapolating Quantum Observables with Machine Learning: Inferring Multiple Phase Transitions from Properties of a Single Phase, *Phys. Rev. Lett.* **121**, 255702 (2018).
- [36] A. A. Shirinyan, V. K. Kozin, J. Hellsvik, and M. Pereiro, Self-organizing maps as a method for detecting phase transitions and phase identification, *Phys. Rev. B* **99**, 041108(R) (2019).
- [37] C. Giannetti, B. Lucini, and D. VDACCHINO, Machine Learning as a universal tool for quantitative investigations of phase transitions, *Nucl. Phys. B* **944**, 114639 (2019).
- [38] J. Greitemann, K. Liu, and L. Pollet, Probing hidden spin order with interpretable machine learning, *Phys. Rev. B* **99**, 060404(R) (2019).
- [39] S. S. Lee and B. J. Kim, Confusion scheme in machine learning detects double phase transitions and quasi-long-range order, *Phys. Rev. E* **99**, 043308 (2019).
- [40] P. Ponte and R. G. Melko, Kernel methods for interpretable machine learning of order parameters, *Phys. Rev. B* **96**, 205146 (2017).
- [41] Z. Li, M. Luo, and X. Wan, Extracting critical exponents by finite-size scaling with convolutional neural networks, *Phys. Rev. B* **99**, 075418 (2019).
- [42] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [43] J. F. Rodriguez-Nieva and M. S. Scheurer, Identifying topological order through unsupervised machine learning, *Nat. Phys.* **15**, 790 (2019).
- [44] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, *Proc. Natl. Acad. Sci. USA* **102**, 7426 (2005).
- [45] T. Caliński and J. Harabasz, A dendrite method for cluster analysis, *Commun. Statist.* **3**(1), 1 (1974).
- [46] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.* **20**, 53 (1987).

- [47] R. H. Swendsen and J. S. Wang, Nonuniversal Critical Dynamics in Monte Carlo Simulations, *Phys. Rev. Lett.* **58**, 86 (1987).
- [48] Y. Liu, Z. Li, H. Xiong, X. Gao, J. Wu, and S. Wu, Understanding and enhancement of internal clustering validation measures, *IEEE Trans. Cybern.* **43**, 3 (2013).
- [49] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087 (1953).
- [50] U. Wolff, Collective Monte Carlo Updating for Spin Systems, *Phys. Rev. Lett.* **62**, 361 (1989).
- [51] Y. Hsieh, Y. J. Kao, and A. W. Sandvik, Finite-size scaling method for the Berezinskii-Kosterlitz-Thouless transition, *J. Stat. Mech: Theory Exp.* (2013) P09001.
- [52] B. Nienhuis, Exact Critical Point and Critical Exponents of $O(n)$ Models in Two Dimensions, *Phys. Rev. Lett.* **49**, 1062 (1982); Y. Deng, T. Garoni, W. Guo, H. Blöte, and A. Sokal, Cluster Simulations of Loop Models on Two- Dimensional Lattices, *ibid.* **98**, 120601 (2007).
- [53] D. M. Hübscher and S. Wessel, Stiffness jump in the generalized XY model on the square lattice, *Phys. Rev. E* **87**, 062112 (2013).
- [54] G. A. Canova, Y. Levin, and J. J. Arenzon, Competing nematic interactions in a generalized XY model in two and three dimensions, *Phys. Rev. E* **94**, 032140 (2016).
- [55] F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* **12**, 2825 (2011).
- [56] L. Zelnik-Manor and P. Perona, Self-tuning spectral clustering, *Neural Inf. Process. Syst.* **17**, 1601 (2004); G. Mishne and I. Cohen, Multiscale anomaly detection using DFs, *IEEE J. Sel. Top. Sig. Proc.* **7**, 111 (2013).
- [57] M. S. Scheurer and R.-J. Slager, Unsupervised Machine Learning and Band Topology, *Phys. Rev. Lett.* **124**, 226401 (2020).
- [58] Y. Long, J. Ren, and H. Chen, Unsupervised Manifold Clustering of Topological Phononics, *Phys. Rev. Lett.* **124**, 185501 (2020).
- [59] L. Wang, Y. Jiang, L. He, and K. Zhou, Continuous-mixture Autoregressive Networks for efficient variational calculation of many-body systems, [arXiv:2005.04857](https://arxiv.org/abs/2005.04857).
- [60] T. Mendes-Santos, X. Turkeshi, M. Dalmonte, and A. Rodriguez, Unsupervised learning universal critical behavior via the intrinsic dimension [Phys. Rev. X (to be published)] (2021).
- [61] Y. Che, C. Gneiting, T. Liu, and F. Nori, Topological quantum phase transitions retrieved through unsupervised machine learning, *Phys. Rev. B* **102**, 134213 (2020).