


Application of the variational autoencoder to detect the critical points of the anisotropic Ising model

Anshumitra Baul 


Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA

Nicholas Walker

Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA

Juana Moreno and Ka-Ming Tam

Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA, and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA

 (Received 19 November 2021; revised 26 August 2022; accepted 8 March 2023; published 7 April 2023)

We generalize the previous study on the application of variational autoencoders to the two-dimensional Ising model to a system with anisotropy. Due to the self-duality property of the system, the critical points can be located exactly for the entire range of anisotropic coupling. This presents an excellent test bed for the validity of using a variational autoencoder to characterize an anisotropic classical model. We reproduce the phase diagram for a wide range of anisotropic couplings and temperatures via a variational autoencoder without the explicit construction of an order parameter. Considering that the partition function of $(d + 1)$ -dimensional anisotropic models can be mapped to that of the d -dimensional quantum spin models, the present study provides numerical evidence that a variational autoencoder can be applied to analyze quantum systems via the quantum Monte Carlo method.

DOI: [10.1103/PhysRevE.107.045301](https://doi.org/10.1103/PhysRevE.107.045301)

I. INTRODUCTION

Machine learning (ML) has become an indispensable tool to expand the limits of scientific understanding in the age of big data. An overflow of information is being analyzed using ML to quantify patterns in a wide variety of fields, including social networks, object and image recognition, advertising, finance, engineering, medicine, biological physics, and astrophysics, among others [1].

Machine learning is a data modeling approach that employs algorithms that favor strategies driven by statistical analysis and based on pattern extraction and inference. ML algorithms, such as deep learning, provide new avenues for understanding physical data. Opportunities for scientific investigations are being devised, particularly in numerical studies, which naturally involve large data sets and complex systems, where explicit algorithms are often challenging. A concerted effort is emerging to address large data problems in statistical mechanics and many-body physics using the ML approach [2–10]. The foundation of ML is deeply connected with statistical physics and hence is fruitful to combine ML techniques with numerical methods that involve predicting phase transition regions. Scaling and renormalization are the core principles of understanding macroscopic phenomena from microscopic properties. The way forward for machines to learn from large data sets would incorporate conceptually similar principles [2,10,11].

Changes in the macroscopic properties of a physical system occur in phase transitions, which often involve a symmetry-breaking process [12]. The theory of phase

transitions and symmetry breaking was formulated by Landau as a phenomenological model and was later devised from microscopic principles using the renormalization group [13]. Phases can be identified by an order parameter that is zero in the disordered phase and finite in the ordered phase. The order parameter is determined by the symmetry considerations of the underlying Hamiltonian. There are states of matter where the order parameter can only be defined in a complicated non-local way. These systems include topological states such as quantum spin liquids [14]. A major goal of the ML approach in complex statistical mechanics models or strongly correlated systems is to detect phase transitions from the data without explicitly constructing any order parameter [3].

The development of artificial neural networks to detect phase transitions is a major advance in the area of ML applications in physics. In earlier works, artificial neural networks were based on supervised learning algorithms [2,3,10]. Labeled data are used to train the supervised learning algorithm, from which the algorithm learns to assign labels to input data points [15,16]. Apart from supervised learning, another major type of ML is unsupervised learning, for which the input data have no labels. Conventional unsupervised learning algorithms, such as principal component analysis [17], find structure in unlabeled data without involving any artificial neural network. Here, data are classified into clusters, and labels can then be assigned to the data points accordingly [17].

The autoencoder is a new direction to utilize artificial neural networks in unsupervised machine learning. The first versions of the autoencoder were used to reduce the

dimensions of the data before feeding its output to other ML algorithms [18,19]. The autoencoder is created by encoding an artificial neural network, which outputs a latent representation of the given input data; this intermediate latent data are utilized as input to a decoding neural network that tries to accurately reconstruct the input data from the latent representation [20,21].

A major shortcoming of the autoencoder is the possibility of sharp changes in the latent representation with respect to small differences in the input data. Ideally, the latent representation should be a smooth function of the input data. The variational autoencoder (VAE) represents the latent representation in terms of probability distributions instead of a fixed set of numbers [22,23]. This probabilistic latent representation allows a smooth behavior of the model. Since 2013, VAEs have become one of the most successful unsupervised learning algorithms [22]. Promising results are shown in both data encoding and data reconstruction [2,22,23].

VAEs have recently been successfully applied to detect phase transitions in classical spin models [2,24,25]. The input data sets are given by the Monte Carlo method. Then, unsupervised machine learning, such as VAE, is used to decipher and distinguish different regimes of physics in the input data sets. After successfully applying this to classical models, a natural question is whether such an approach remains viable for quantum models. In particular, a VAE can be viable for distinguishing different quantum phases and transition regions on the basis of the data obtained from the quantum Monte Carlo method. Recently, various models from statistical mechanics, particularly the Ising model and the Potts model, have been investigated [26–28]. In this work, we investigate a rather simple quantum model, the one-dimensional transverse-field Ising model (TFIM), to address the capability and limitations of the autoencoder. Since the critical line of the model can be calculated analytically due to the self-dual property [29–31], the TFIM is an excellent test bed to address the various aspects of VAEs in quantum models.

This paper is organized as follows. In Sec. II, we briefly describe the TFIM and the Suzuki-Trotter formulation by mapping it to the anisotropic Ising model. In Sec. III, the Monte Carlo method and the VAE are presented. The results of the VAE are described in Sec. IV. We conclude and discuss the implication and possible future applications of the method developed in this study in Sec. V. The self-duality of the anisotropic two-dimensional Ising model and the details of the VAE are discussed in the Appendices.

II. TRANSVERSE FIELD ISING MODEL

We consider an Ising model with a transverse field [32–34]. The Hamiltonian is given as

$$H = - \sum_{\langle i,j \rangle} J_{ij} \sigma_i^z \sigma_j^z - \Gamma \sum_i \sigma_i^x, \quad (1)$$

where σ^α ($\alpha = x, y,$ and z) are the Pauli matrices which obey the commutation relation $[\sigma_i^\alpha, \sigma_j^\beta] = 2i\delta_{ij}\epsilon_{\alpha\beta\gamma}\sigma_i^\gamma$. J_{ij} is the coupling between the spins at sites i and j . Only nearest-neighbor coupling is considered in this study. Γ is the transverse field applied in the x direction and σ^z has

eigenvalues ± 1 . Their eigenvectors are symbolically denoted by $|\uparrow\rangle$ and $|\downarrow\rangle$, that is,

$$|\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (2)$$

and

$$|\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3)$$

The order parameter is given by the average magnetization $m = \sum_i \langle \sigma_i^z \rangle / N$ (N is the total number of sites), which characterizes the phase transition between the paramagnetic and ferromagnetic phases. Without specification, we consider only the one-dimensional TFIM with coupling limited to the nearest neighbors.

Instead of working with quantum spins directly, we follow the standard procedure of mapping a d -dimensional quantum Hamiltonian to a $d + 1$ -dimensional effective classical Hamiltonian by using the Suzuki-Trotter transformation [35–37]. The details are discussed in Appendix A. We define the longitudinal spin-coupling terms and the transverse-field terms as follows:

$$\begin{aligned} H_0 &\equiv - \sum_{\langle i,j \rangle} J_{ij} \sigma_i^z \sigma_j^z, \\ V &= -\Gamma \sum_{i=1} \sigma_i^x, \\ H &= V + H_0. \end{aligned} \quad (4)$$

The effective classical Hamiltonian after the Suzuki-Trotter transformation is

$$\begin{aligned} H_{\text{eff}}(\{\sigma\}) &= \sum_{\langle i,j \rangle} \sum_{k=1}^M \left[\frac{-J_{ij}}{M} \sigma_{i,k} \sigma_{j,k} \right. \\ &\quad \left. - \frac{1}{(2\beta)} \ln \left(\coth \frac{\beta\Gamma}{M} \right) \sigma_{i,k} \sigma_{i,k+1} \right]. \end{aligned} \quad (5)$$

The effective Hamiltonian shows the spin system in the $1 + 1$ -dimensional lattice with an additional label k for each spin variable. Each single quantum spin variable σ_i in the original Hamiltonian now carries an array of M classical spins $\sigma_{i,k}$. This new (timelike) dimension along which these classical spins are spaced is called the Trotter dimension.

In this paper, we assume that the model only has nearest-neighbor coupling on a square lattice. The coupling along the x direction and the y direction are denoted as $J_x = \frac{J_{ij}}{M}$ and $J_y = \frac{1}{(2\beta)} \ln \left(\coth \frac{\beta\Gamma}{M} \right)$, respectively. We also define $K_x = 1/(\beta J_x)$ and $K_y = 1/(\beta J_y)$. We set $N = M$ in this study.

The critical points can be analytically obtained for the two-dimensional classical Ising model due to the self-dual property. The detail of the dual transformation is shown in Appendix B.

III. METHODOLOGY

A. Monte Carlo sampling

The spin configurations are generated by using the single spin-flip Metropolis algorithm. We first flip the spin of a single

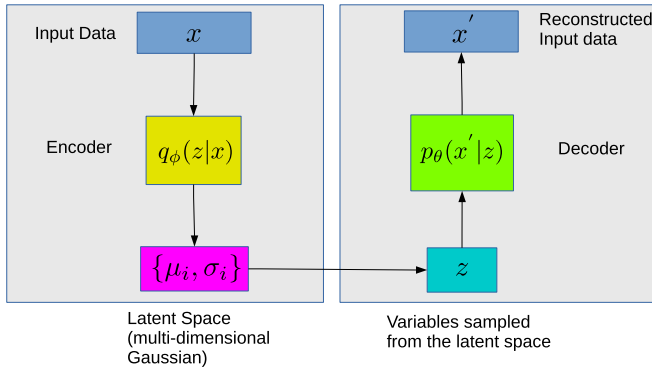


FIG. 1. Diagram depicting the structure of the VAE. The left-hand side is the encoding part, and the right-hand side is the decoding part. x is the input Ising spin configuration, $q_\phi(z|x)$ is the encoder neural network, and μ and σ are the latent means and standard deviations of the latent space distribution z . $p_\theta(x'|z)$ is the decoder neural network, and x' is the reconstructed Ising spin configuration.

lattice site and calculate the change in energy ΔE . Then, we use the resulting change in energy to calculate the Metropolis criterion $\exp(-\Delta E/T)$. If a randomly generated number in the range $[0,1)$ is less than or equal to the Metropolis criterion, the configuration is accepted as the new configuration. The simulation code is written in PYTHON using the NumPy library [38–41]. We note that using a generative neural network, instead of the Monte Carlo method, for sampling has also been recently proposed for one-dimensional (1D) quantum spin models [42,43].

B. Variational autoencoder

The VAE is in the category of generative models. New data can be produced by learning the input data distribution [22,44]. We use an encoder-decoder architecture where the encoder maps the input to a latent representation in terms of some chosen distribution. The latent distribution is mapped back to reconstruct the input data using the decoder. The latent representation in a well-trained model can be used to generate new samples that resemble the original training data [45].

This study only considers a multidimensional Gaussian distribution for the latent space representation. We briefly explain the main concepts of the VAE in the following.

Encoder. The encoder is a neural network that takes the higher-dimensional input data into a lower-dimensional latent space. It is also called a probabilistic encoder of the VAE. For example, the spin configuration of a lattice of size $32 \times 32 = 1024$ can be converted into a vector of dimension 8. The encoder is a method for dimensionality reduction. The neural network representing the encoder maps each sample to a distribution. The encoder, denoted as $q_\phi(z|x)$, is a distribution that maps an input sample x (Ising spin configurations) to provide a latent representation z (the parameters in the multidimensional Gaussian distribution). ϕ is a set of learnable parameters in the neural network that are varied to produce the encoder output (see Fig. 1).

Latent space. The latent space is the input for the decoder network, and it is outputted from the encoder network. For

the VAE, the latent space is represented by a multidimensional Gaussian distribution. As the Gaussian distribution is completely specified by its mean and standard deviations, the dimension of the latent space is twice that of the dimension of the multidimensional Gaussian distribution. An encoded sample is denoted by z .

The latent space is regularized and then penalized for deviating from the prior multidimensional Gaussian distribution by the Kullback-Leibler (KL) divergence term [46].

Decoder. The decoder in VAE converts compressed samples in the latent space back to input samples [47] (see Fig. 1). It is represented as $p_\theta(x'|z)$, a distribution that produces reconstructed samples x' conditioned on latent representation z . θ is a set of neural network learning parameters that can be varied to produce a different output.

Loss functions in the VAE. The VAE contain two sets of trainable parameters, θ and ϕ , for the neural networks of the encoder and the decoder, respectively. They are trained by minimizing the loss function. Loss functions in the VAE consist of two terms. The first one measures the “similarity” of the inputs and the reconstructed outputs. The second one measures the difference between the designated prior distribution, chosen to be a multidimensional Gaussian distribution, and the actual distribution of the inputs.

For the first term of the loss function, the standard reconstruction loss measures the error between the samples generated by the decoder and the original input samples. We measured this error by the binary cross-entropy between the encoder input and the decoder output. This is expressed as $L_{RC} = -E_{z \sim q_\phi(z|x)}[\log p_\theta(x|z)]$. The expectation E is over the representations z with respect to the encoder’s distribution.

The second is the Kullback-Leibler divergence (KLD) of the latent representation. The KLD measures the divergence between the chosen latent representation $p(z)$ and the approximated distribution from the output of the encoder $q_\phi(z|x)$.

The KLD is defined as

$$L_{KLD} = D_{KL}[q_\phi(z|x)||p(z)] = - \sum_z q_\phi(z|x) \log \left(\frac{q_\phi(z|x)}{p(z)} \right). \quad (6)$$

It is minimized to optimize the latent representation of the encoder $q_\phi(z|x)$ to resemble the latent representations of $p(z)$.

The total loss is the sum of the reconstruction loss and the KLD [48]:

$$L(\phi, \theta; x, z) = -E_{z \sim q_\phi(z|x)}[\log p_\theta(x|z)] + D_{KL}[q_\phi(z|x)||p(z)] = L_{RC} + L_{KLD}. \quad (7)$$

The two losses in the VAE are optimized simultaneously. The linear combination of the reconstruction loss and the KLD is often denoted as a variational lower bound or evidence lower bound loss function since both the reconstruction loss and the KLD are non-negative. Minimizing the loss, $\min_{\theta, \phi} L(\theta, \phi; x, z)$, maximizes the lower bound of the probability of generating new samples [45].

C. β -total correlation VAE

The VAE can be further refined by decomposing the KLD into three parts. The three parts describe the index-code mutual information, the total correlation, and the dimensionwise Kulback-Leibler divergence and are denoted as α , β , and γ , respectively. β is the most important one for obtaining optimal results [48,49].

The β -total-correlation VAE (β -TCVAE) is defined as the VAE with $\alpha = \gamma = 1$, and β is a tuning parameter. Further details are given in Appendix B. This approach is well suited for learning patterns [49]. In this study, we fixed the decomposition parameters with $\alpha = \gamma = 1$ and $\beta = 8$; these values were found to be reliable in finding the phase diagram of the two-dimensional (2D) isotropic Ising model [25,50].

Our goal is to map the raw Ising spin configurations to a reduced set of descriptors that discriminate between the samples using the criterion inferred by the β -total correlation VAE [25,50]. The encoder and the decoder are implemented as deep convolution neural networks (CNN) to preserve the spatially dependent two-dimensional structure of the Ising spin configurations [51]. A scaled exponential linear unit (SELU) activation function is used in each convolution layer. The output of the final convolution layer is flattened and fed into two two-dimensional dense layers. Then, the output of the encoder CNN is used as the input layer for the decoder CNN. The decoder network is simply the opposite of the encoder network, with convolution transpose layers instead of standard convolution layers [50–52]. The final output layer of the decoder network is reproduced from the original input configurations obtained from the encoder network, which uses a sigmoid activation function. The loss term consists of the reconstruction loss and the β -total-correlation KLD (β -TCKLD) term with $\alpha = \gamma = 1$ and $\beta = 8$ [46,50]. We employ minibatch stratified sampling on the given data while training.

To optimize the loss, Nesterov-accelerated adaptive moment estimation was used, which efficiently minimizes the loss during training of the β -TCVAE model [50,53]. The default parameters provided by the Keras library and a learning rate of 0.000 01 were chosen. Training was carried out over 100 epochs with a batch size of $33 \times 33 = 1089$ for both lattice sizes $N = 64$ and $N = 128$ with a number of samples of 1024 per phase point. The reduced descriptors of the 2D Ising spin configuration are given by the latent variables [50]. The β -TCVAE model used in this work was implemented using the Keras ML library with TensorFlow as the supporting backend [54].

D. Principal component analysis on the latent space

The principal component analysis (PCA) is applied to the latent means and standard deviations obtained after fitting the β -TCVAE using the sci-kit-learn package [17,55]. The PCA performs an orthogonal transformation into a new basis of linearly uncorrelated features, principal components. Each principal component encompasses the largest possible variance across the sample space under an orthogonality constraint [17].

The latent representation characterizes the structure of the Ising configurations, but the principal components of the

latent representation show greater discrimination between the different structural characteristics of the configurations compared to the raw variable representation [50]. The rationale for using the PCA is to provide a more compact representation that characterizes the different phases of the Ising model. As we show in the results section, the first and second components already distinguish between the different phases of the anisotropic Ising model.

IV. RESULTS

As discussed in the previous section, for VAE models, the samples are drawn from a multidimensional Gaussian distribution parametrized by a vector of means, μ_i , and standard deviations, σ_i , where i is the index of the dimension of the distribution. All plots in this paper were generated with the MATPLOTLIB package using a perceptually uniform color map [56]. In each plot, the coloration of each square pixel represents the average ensemble value of the principal components of the mean or variance of the latent space. We study two different system sizes, $N = 64$ and 128 . We focus on the first two principal components of the latent variance and the second principal component of the latent mean. We denote them by τ_0 , τ_1 , and ν_1 , respectively. The first principal component of the latent mean ν_0 does not capture a clear distinction between the ferromagnetic and paramagnetic phases of the system. See Appendix D for the plot of ν_0 .

Figure 2 displays ν_1 , the second principal component of the latent mean. It is apparent that ν_1 resembles the magnetization m of snapshots of Ising spin configurations. We note that it is not expected that any of the latent variables have the same value as any physical quantity, such as magnetization. Nonetheless, the plot of ν_1 clearly discriminates the ferromagnetic phase from the paramagnetic phase. The phase transition line in white corresponds to the analytical solution in the thermodynamic limit, Eq. (B14) in Appendix B. Since the magnetization can be seen as the order parameter for the two-dimensional Ising model, a reasonable representation of the order parameter is seen to be possibly extracted from the VAE. We note that the simulations and the VAE are performed on finite-size systems where a truly broken symmetry does not occur. This is the reason for the seemingly random values of ν_1 in the ferromagnetic phase. We show that other latent variables from the VAE have structures similar to the amplitude of the magnetization. As magnetization is a linear feature of the Ising spin configuration, a simpler linear model would be sufficient for extracting the magnetization.

The first principal component of the latent variance τ_0 is plotted in Fig. 3. The white line is the analytical solution for the phase transition line. The value of τ_0 remains very small in the upper right region of the figures. This can be considered as the reflection of the small changes in the amplitude of the energy or the magnetization in the magnetic phase. Once the system approaches the critical line from the upper right region, the value of τ_0 increases sharply. This behavior is again consistent with the larger range of values of energy or magnetization.

In particular, we consider the case for the isotropic limit ($K_x = K_y$), that is, the classical ($\Gamma = 0$) limit. The critical

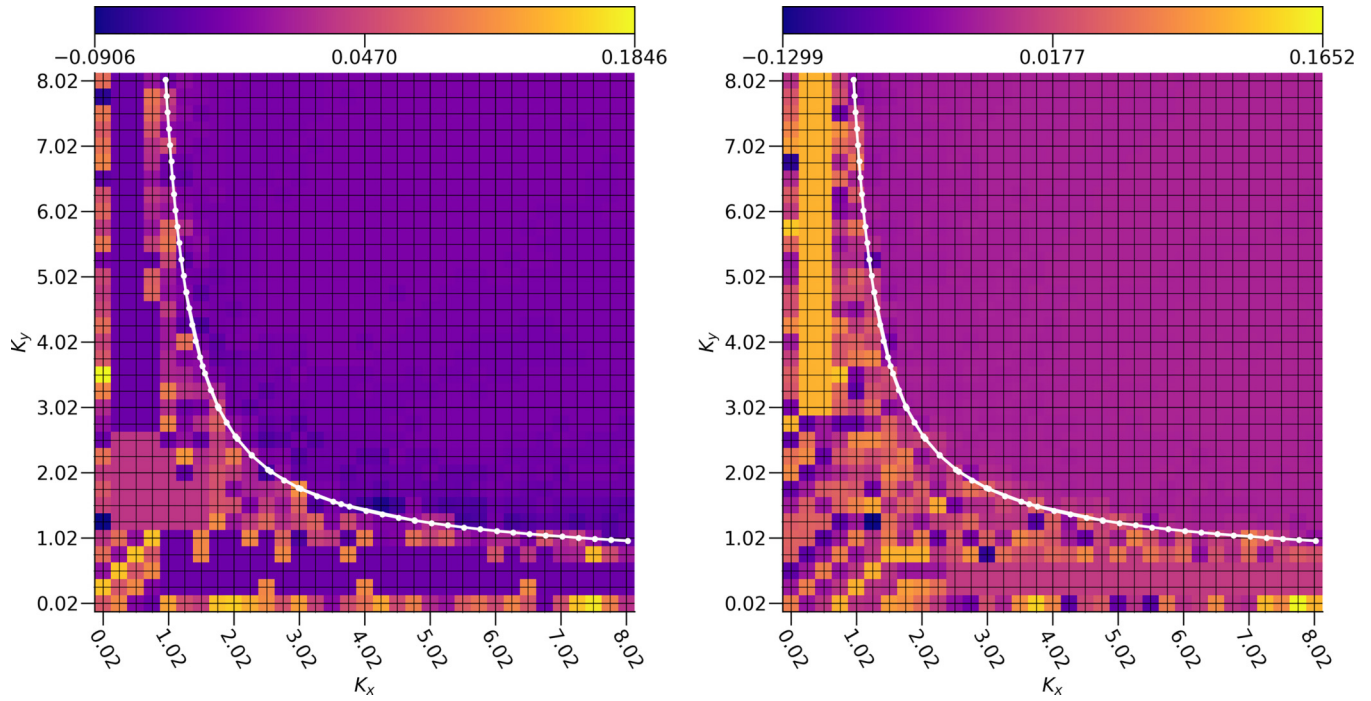


FIG. 2. The second principal component of the latent mean, v_1 , with respect to model parameters K_x [$1/(\beta J_x)$] and K_y [$1/(\beta J_y)$] for the 2D square lattice. The left panel and the right panel are for system sizes $N = M = 64$ and $N = M = 128$, respectively. The white line is the analytical phase transition line.

point is given as $K_c (= K_x = K_y) = \frac{2}{\ln[1+\sqrt{2}]} \sim 2.2721$ [57]. Correspondingly, in Fig. 3, we see a sharp change in τ_0 around this value of K_c .

It should be noted that the paramagnetic samples are essentially noisy due to entropic contributions. Therefore, these are easy to discriminate from the rest of the samples using

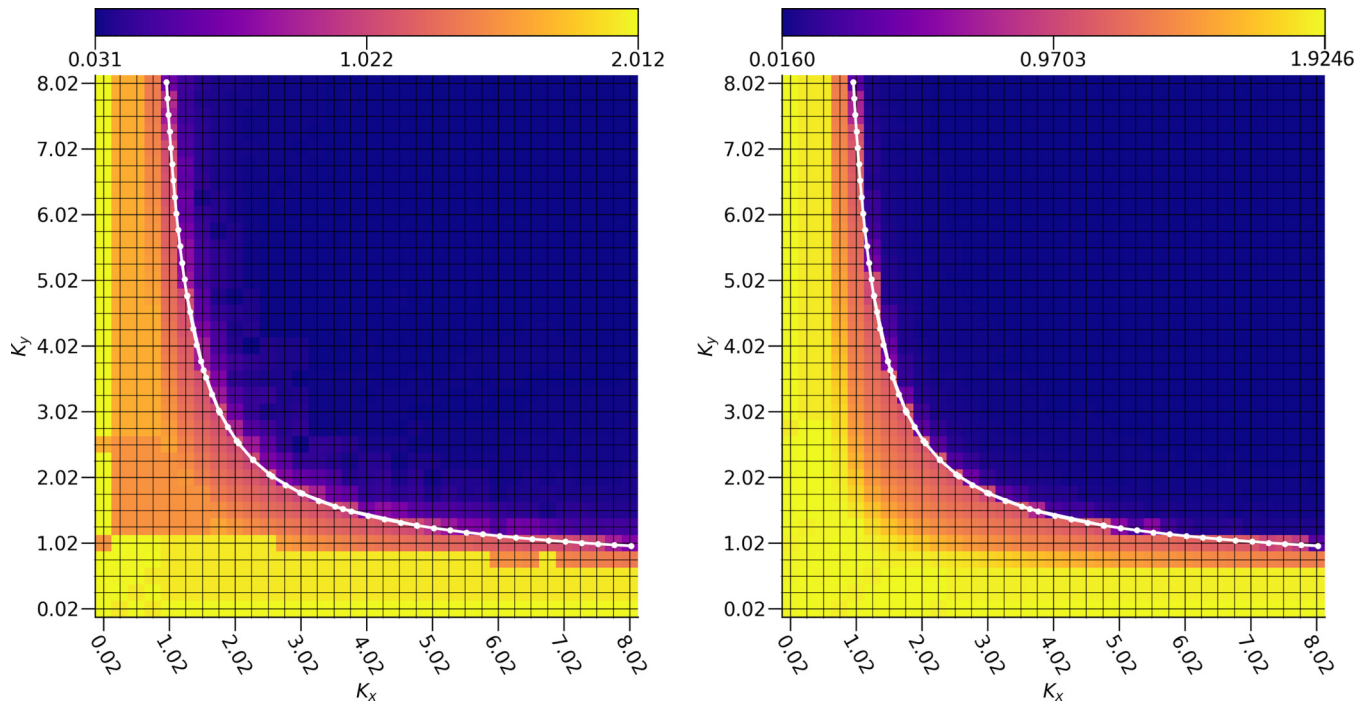


FIG. 3. The first principal component of the latent variance, τ_0 , with respect to model parameters K_x [$1/(\beta J_x)$] and K_y [$1/(\beta J_y)$] for the 2D square lattice. The left and right panels are for the system sizes $N = M = 64$ and $N = M = 128$, respectively. The white line is the analytical phase transition line.

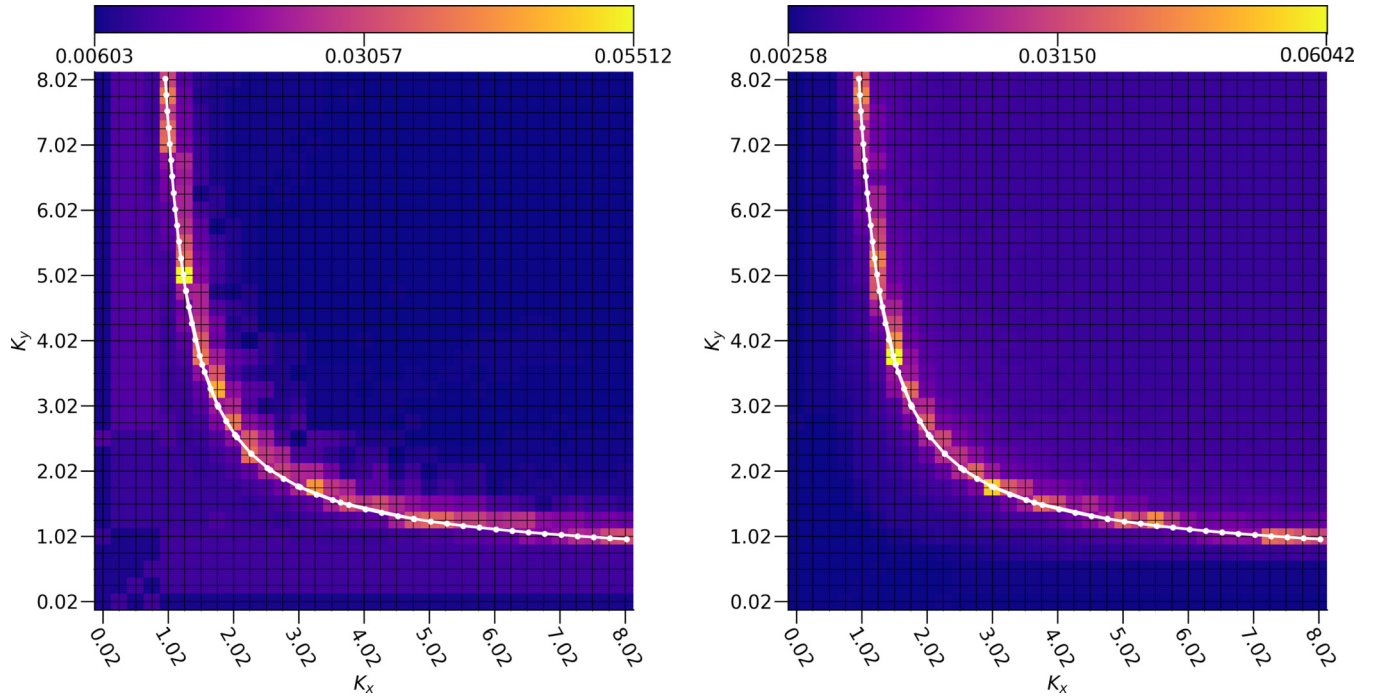


FIG. 4. The second principal component of the latent variance, τ_1 , with respect to model parameters K_x [$1/(\beta J_x)$] and K_y [$1/(\beta J_y)$] for the 2D square lattice. The left and right panels are for the system sizes $N = M = 64$ and $N = M = 128$, respectively. The white line is the analytical phase transition line.

a β -TCVAE model. Given this fact, the samples with ν_1 values corresponding to nearly zero magnetization and relatively high values for τ_0 resemble Gaussian noise with no notable order preference. An interesting question is to look for a quantity that may resemble a response function, such as susceptibility or heat capacity.

We plot the second principal component of the latent variance τ_1 in Fig. 4. The phase-transition line is again included for comparison. The values of τ_1 along the analytical critical line are much higher than for any other combination of K_x and K_y . Thus, τ_1 has a substantial similarity for magnetic susceptibility.

In brief, from the plots of τ_0 , τ_1 , and ν_1 , we find that the VAE can generate quantities that resemble the magnetization, the amplitude of the magnetization or the energy, and the magnetic susceptibility or the heat capacity, respectively. We can infer from these quantities the phase transition line between the paramagnetic and ferromagnetic phases.

The quality of fitting the Ising spin configurations to the present VAE model can be quantified by the values of the loss functions. The three losses represented in Fig. 5 are the VAE loss, the reconstruction loss, and the latent loss. The reconstruction loss converges to about 0.5. The latent loss is obtained from the β -TCKLD term. This loss quickly converges to a value close to 0. The total β -TCVAE loss for the two-dimensional anisotropic Ising model for both lattice sizes, $N = 64$ and $N = 128$, also settles quickly to a value around 0.5.

V. DISCUSSION AND CONCLUSIONS

We used a β -TCVAE model to extract structural information from the raw Ising spin configurations. It exposes

interesting derived descriptors of the configurations which are used to identify the second-order phase transition line. This is done by studying the behavior of latent variable mappings of the Ising spin configurations with respect to the anisotropic coupling (J_x, J_y) associated with temperature.

We find that ν_1 , the second principal component of the latent mean, reflects the magnetization of the two-dimensional anisotropic Ising model. Hence, ν_1 is interpreted as an indicator of the magnetization exhibited by the configurations. In contrast, τ_0 and τ_1 , the first two principal components of the latent variance, can be understood as indicators of the amplitude of the magnetization or the energy and the magnetic susceptibility or the heat capacity, respectively. Thus both τ_0 and τ_1 can also provide a reasonable estimate of the second-order phase transition line. In the lack of a proper ansatz to perform a finite-size scaling, the data generated by machine learning cannot be used to extract the critical exponents corresponding to the isotropic limit. In addition, it is a crude assumption to assume that the data obtained from the VAE are within the thermodynamic limit.

Since the $d + 1$ -dimensional anisotropic Ising model is equivalent to the d -dimensional quantum spin system through the Suzuki-Trotter transformation, this method can be trivially extended to other 1D quantum systems [58]. Various quantum Monte Carlo methods map correlated fermions on the lattice to an effective classical Ising model via a discrete Hubbard-Stratonovich transformation [59–61]. The approach presented here has the same data structure as those quantum Monte Carlo methods, which can be readily adapted to analyze the data from these methods.

Moreover, methods for strongly correlated systems, such as the dynamical mean-field theory (DMFT) and their cluster

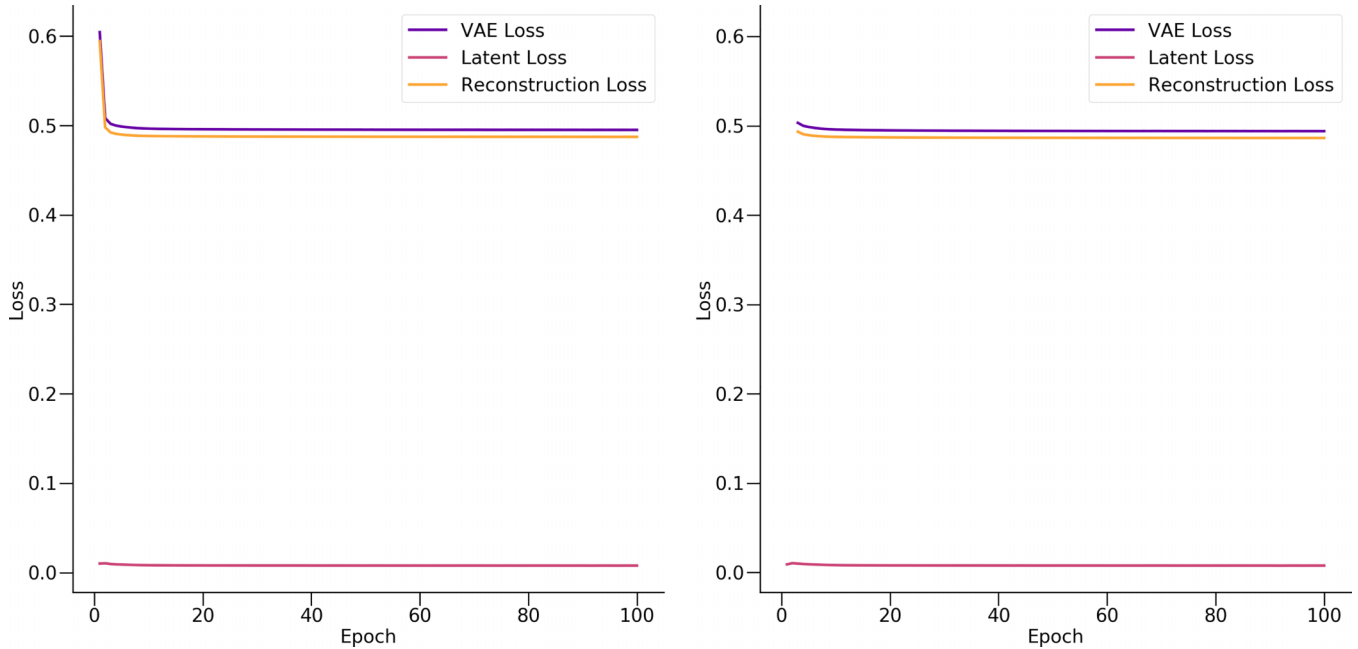


FIG. 5. β -TCVAE loss and the reconstruction loss settle to about 0.5 and the latent loss settles close to 0 for 100 epoch cycles. The left and right panels are for the square lattice of sizes $N = M = 64$ and $N = M = 128$, respectively.

generalizations—the dynamical cluster approximation (DCA) and the cellular dynamical mean-field theory (CDMFT)—all have a very similar data structure when using the Hirsch-Fye quantum impurity solver [60,62,63]. The dynamical variables for the quantum impurity solver are represented by Ising-like variables since the Hubbard-Stratonovich field to decouple the local Hubbard interaction can be simplified to binary variables for fermionic operators [60]. For the single quantum impurity in the DMFT, the data structure is given by a set of $1 \times N_T$ binary variables, where N_T is the number of Suzuki-Trotter steps. On the other hand, for quantum clusters with N impurities in DCA and CDMFT, the data structure is given by a set of $N \times N_T$ binary variables. It is notoriously difficult to obtain the putative quantum critical point from the paramagnetic solution of the Hubbard model, as there is no simple quantity to track the transition [64,65]. The method presented here can readily be adapted to the study of these quantum systems. Therefore, this method would be an essential tool for analyzing data from DMFT, DCA, and CDMFT.

There are many opportunities to develop this method further, not only by investigating more complex systems but also by introducing improvements beyond the scope of this work. Finite-size scaling has become an important approach to address limitations in using finite-sized systems to investigate regions near critical phenomena [66,67]. To find a correspondence between the VAE encodings of different system sizes is a challenging problem, as different VAE structures need specific training for each system size, which in turn demands different hyperparameters and training iterations [66]. Numerical difficulties arise when performing finite-size scaling analysis since the variation of properties concerning the system size is hard to isolate from the systemic variation because of the use of different neural networks trained with different hyperparameters to extract the specific macroscopic properties. Solving this issue would play a

significant role in improving the VAE characterization of critical phenomena.

Another interesting direction is to use the generative adversarial network [68] instead of the VAE as the generative model. Promising results have been obtained for the isotropic 2D Ising model [69].

An important topic that has not been studied in the present work is the estimation of the critical exponents. Conventional numerical methods estimate exponents by calculating the corresponding critical quantities for numerous system sizes. A finite-size scaling ansatz based on the theory of renormalization group is used to fit the exponents. The fundamental assumption is that the correlation length of a finite-size system is cut off by the system size and thus ends up in the finite-size scaling ansatz. It is not well studied whether the same approach can be applied on the neural network. In particular, we do not know whether a neural network can be critical. Even if it is critical, it is unknown whether different training data could tune the neural network to different universality classes. These are crucial questions that must be addressed before a proper estimation of the critical exponents can be obtained.

ACKNOWLEDGMENTS

A.B. and N.W. are supported by NSF Grant No. DMR-1728457. J.M. and K.-M.T. are supported by the U.S. Department of Energy, Office of Science and Office of Basic Energy Sciences, under Award No. DE-SC0017861. This work used high-performance computational resources provided by the Louisiana Optical Network Initiative [70] and HPC@LSU computing.

APPENDIX A: SUZUKI-TROTTER FORMALISM

We apply the Suzuki-Trotter transformation to the transverse-field Ising model. The longitudinal spin-coupling

term and the transverse-field term are defined as follows:

$$\begin{aligned} H_0 &= - \sum_{\langle i,j \rangle} J_{ij} \sigma_i^z \sigma_j^z, \\ V &= -\Gamma \sum_{i=1} \sigma_i^x, \\ H &= V + H_0. \end{aligned} \quad (\text{A1})$$

The partition function of H reads

$$Z = \text{Tr}(e^{-\beta(H_0+V)}), \quad (\text{A2})$$

where β is the inverse temperature. The Trotter formula approximates the exponential of a sum,

$$\exp(A_1 + A_2) = \lim_{M \rightarrow \infty} [\exp(A_1/M) \exp(A_2/M)]^M, \quad (\text{A3})$$

when $[A_1, A_2] \neq 0$. Using the Trotter formula, we have the following:

$$Z = \sum_i \lim_{M \rightarrow \infty} \langle s_i | [\exp(-\beta H_0/M) \exp(-\beta V/M)]^M | s_i \rangle, \quad (\text{A4})$$

where s_i is the i th spin configuration of the entire system, and the above summation runs over all possible 2^N configurations denoted by i . We use M identity operators to decouple the terms. The identity operator formed from the spin operators is given as

$$\begin{aligned} I &= \sum_i^{2^N} |s_{i,k}\rangle \langle s_{i,k}| \\ &= \sum_{\sigma_{i,k}=-1}^1 |\sigma_{i,k}, \dots, \sigma_{N,k}\rangle \langle \sigma_{i,k}, \dots, \sigma_{N,k}|, \end{aligned} \quad (\text{A5})$$

where $k = 1, 2, \dots, M$. Hence, Z becomes the product of M exponentials:

$$Z = \lim_{M \rightarrow \infty} \text{Tr} \prod_{k=1}^M \langle \sigma_{i,k}, \dots, \sigma_{N,k} | \quad (\text{A6})$$

$$\exp\left(-\frac{\beta H_0}{M}\right) \exp\left(-\frac{\beta V}{M}\right) | \sigma_{i,k+1}, \dots, \sigma_{N,k+1} \rangle. \quad (\text{A7})$$

Applying the periodic boundary conditions $\sigma_{N+1,p} = \sigma_{1,p}$, we arrive at the following expression of the partition function,

$$Z = C^{\frac{NM}{2}} \text{Tr}_\sigma [-\beta H_{\text{eff}}(\{\sigma\})], \quad (\text{A8})$$

where $C = \frac{1}{2} \sinh \frac{2\beta\Gamma}{M}$ and the effective classical Hamiltonian is

$$\begin{aligned} H_{\text{eff}}(\{\sigma\}) &= \sum_{\langle i,j \rangle} \sum_{k=1}^M \left[\frac{-J_{ij}}{M} \sigma_{i,k} \sigma_{j,k} \right. \\ &\quad \left. - \frac{1}{(2\beta)} \ln \left(\coth \frac{\beta\Gamma}{M} \right) \sigma_{i,k} \sigma_{i,k+1} \right], \end{aligned} \quad (\text{A9})$$

where $\sigma_{i,k}$ are the eigenvalues of σ^z and, therefore, there is no noncommuting part in H_{eff} .

Each single quantum spin variable σ_i in the original Hamiltonian is now represented by an array of M numbers of classical spins $\sigma_{i,k}$. Therefore, the partition function of the 1D quantum Ising model is mapped to that of the 2D Ising model.

This new (timelike) dimension k , along which these classical spins are spaced, is called the Trotter dimension.

APPENDIX B: SELF-DUALITY OF THE TWO-DIMENSIONAL ISING MODEL

In this Appendix, we summarize the derivation of the critical line for the anisotropic two-dimensional Ising model via the self-duality property [29,71,72]. We closely follow the lecture notes by Muramatsu [72].

Taking into account the partition function of the Ising model (where $K = \beta J$),

$$\begin{aligned} Z &= \sum_{\{S_j\}} e^{K \sum_{\langle j,l \rangle} S_j S_l} = \sum_{\{S_j\}} \prod_{\langle j,l \rangle} e^{K S_j S_l} \\ &= \sum_{\{S_j\}} \prod_{\langle j,l \rangle} \sum_{r=0}^1 C_r(K) (S_j S_l)^r, \end{aligned} \quad (\text{B1})$$

where $C_0(K) = \cosh K$ and $C_1(K) = \sinh K$. Applying a simple transformation, for each bond $\langle j, l \rangle$, a new Z_2 variable r is introduced. We label the new variable as r_μ with $\mu \equiv (i, \langle i, j \rangle)$, labeling it with the site i from which the bond $\langle i, j \rangle$ emanates. The partition function thus follows

$$Z = \sum_{\{S_j\}} \sum_{\{r_\mu\}} \prod_{\langle j,l \rangle} C_{r_\mu}(K) \prod_i S_i^{\sum_{\langle i,j \rangle} r_\mu}. \quad (\text{B2})$$

Grouping all products of spins at site i together, $\sum_{\langle i,j \rangle} r_\mu$, contains all four contributions resulting from the bonds connected to site i . Further, we explicitly perform the sum over all spin configurations:

$$\begin{aligned} Z &= \sum_{\{r_\mu\}} \prod_{\langle j,l \rangle} C_{r_\mu}(K) \prod_i \sum_{S_i=\pm 1} S_i^{\sum_{\langle i,j \rangle} r_\mu} \\ &= \sum_{\{r_\mu\}} \prod_{\langle j,l \rangle} C_{r_\mu}(K) \prod_i 2\delta \left[\text{mod}_2 \left(\sum_{\langle i,j \rangle} r_\mu \right) \right]. \end{aligned} \quad (\text{B3})$$

We define a dual lattice in which the vertices of the dual lattice are set in the center of the plaquettes defined by the original lattice. We have vanishing contributions as a result of the presence of the Kronecker δ in many configurations. By defining the new Z_2 variables $\sigma_i = \pm 1$ on the sites of the dual lattice, we can associate with each link of the original lattice a pair of σ_i 's (e.g., on the sites i and j on the dual lattice). Therefore, the variable r_μ is expressed as

$$r_\mu = \frac{1}{2}(1 - \sigma_i \sigma_j), \quad (\text{B4})$$

where the sites i and j in the dual network are those where the link crosses r_μ . The sum of r_μ is over the four nearest neighbors of a site i , and we have

$$\sum_{\langle i,j \rangle} r_\mu = 2 - \frac{1}{2}(\sigma_1 \sigma_2 + \sigma_2 \sigma_3 + \sigma_3 \sigma_4 + \sigma_4 \sigma_1). \quad (\text{B5})$$

There are 2^4 possible configurations for the four variables $\sigma_1, \dots, \sigma_4$. They are grouped into four cases, and all of the cases lead to an even number for the sum of r_μ over the nearest-neighbor bonds. The choice of variables needs to

satisfy the δ function. The partition function becomes

$$Z = \frac{1}{2} 2^N \sum_{\{\sigma_i\}} \prod_{\langle j,l \rangle} C_{[(1-\sigma_j\sigma_l)/2]}(K), \quad (\text{B6})$$

where N is the number of sites in the lattice, and the product is now in the bonds associated with the dual lattice. The expression of the partition function shows that the weight for each configuration of the σ 's is given by the coefficients $C(K)$. Hence, rewriting $C(K)$ to a form that resembles a Boltzmann weight,

$$\begin{aligned} C_r(K) &= \cosh(K)\{1 + r[\tanh(K) - 1]\} \\ &= \cosh(K) \exp(\ln\{1 + r[\tanh(K) - 1]\}) \\ &= \cosh(K) \exp[r \ln \tanh(K)] \\ &= \cosh(K) \exp\left[\frac{1}{2}(1 - \sigma_i\sigma_j) \ln \tanh(K)\right] \\ &= [\cosh(K) \sinh(K)]^{1/2} \exp\left(-\frac{1}{2} \ln \tanh K \sigma_i\sigma_j\right), \end{aligned} \quad (\text{B7})$$

the partition function becomes

$$Z = 1/2(\sinh 2\tilde{K})^{-N} \sum_{\{\sigma_i\}} \exp\left(\tilde{K} \sum_{\langle j,l \rangle} \sigma_j\sigma_l\right). \quad (\text{B8})$$

There are $2N$ bonds, and we define the new coupling constant on the dual lattice as follows:

$$\tilde{K} \equiv -\frac{1}{2} \ln \tanh(K), \quad (\text{B9})$$

where $K = \beta J$. The Ising model is self-dual, since the duality transformations bring it into itself. We consider the free energy per site

$$f(Z) = -\frac{1}{N} \ln Z. \quad (\text{B10})$$

According to the relation between the partition functions of the original and the dual model, we can write

$$f(K) = \ln \sinh(2\tilde{K}) + f(\tilde{K}). \quad (\text{B11})$$

This is a strong constraint on free energy. Since $\sinh(2\tilde{K})$ is an analytic function, a singularity in $f(K)$ implies a singularity in $f(\tilde{K})$. $\tilde{K}(K)$ is a monotonous function of K ; hence, it holds $\tilde{K}_c = K_c$ and we have the following:

$$K_c = \frac{1}{2} \ln(1 + \sqrt{2}). \quad (\text{B12})$$

Self-duality has allowed us to calculate the exact value of the critical temperature in the two-dimensional isotropic Ising model ($K_x = K_y$), where $K_x = \beta J_x$ and $K_y = \beta J_y$. Generalizing the results obtained in the isotropic to the anisotropic one, that is, when couplings $K_x \neq K_y$ in the respective directions, the anisotropic case is as follows:

$$\tilde{K}_x \equiv -\frac{1}{2} \ln \tanh(K_x), \quad \tilde{K}_y \equiv -\frac{1}{2} \ln \tanh(K_y). \quad (\text{B13})$$

Given K_x and K_y , there is only one critical point, with the following condition for a critical line that separates the ordered phase from the disordered phase in the anisotropic Ising

model:

$$\sinh(2K_{xc}) \sinh(2K_{yc}) = 1. \quad (\text{B14})$$

APPENDIX C: β -TCVAE LOSS

The expression of the total loss for the VAE is given by

$$L(\phi, \theta; x, z) = L_{\text{RC}} + L_{\text{KLD}}, \quad (\text{C1})$$

where the reconstruction error (RC) and the Kullback-Leibler divergence (KLD) are defined as

$$L_{\text{RC}} = -E_{z \sim q_\phi(z|x)}[\log p_\theta(x|z)] \quad (\text{C2})$$

and

$$L_{\text{KLD}} = D_{\text{KL}}[q_\phi(z|x)||p(z)], \quad (\text{C3})$$

respectively. Suppose the prior distribution of the latent representation is Gaussian. In that case, the VAE provides disentangled factors in the latent representation, which means the significant dimensions of the latent space are largely independent of each other. In β -total correlation VAE (β -TCVAE), we try to improve the disentanglement of factors in the representation by decomposing the KLD term and apply the tuning parameters independently [49,50]. Each training sample is identified with a unique integer index $n \in 1, 2, \dots, N$ and assigned a uniform random variable in this decomposition. The aggregated posterior $q_\phi(z) = \sum_n q_\phi(z|n)p(n)$ captures the aggregate structure of the latent variables under the distribution of the input, where $q_\phi(z|n) = q_\phi(z|x_n)$ and $q_\phi(z, n) = q_\phi(z|n)p(n) = \frac{1}{N}q_\phi(z|n)$. The decomposition is given as

$$\begin{aligned} I(z; x) + D_{\text{KL}}\left[q_\phi(z)||\prod_j q_\phi(z_j)\right] \\ + \sum_j D_{\text{KL}}[q_\phi(z_j)||p(z_j)]. \end{aligned} \quad (\text{C4})$$

The first term is the index-code mutual information, $I(z; x) = D_{\text{KL}}[q_\phi(z, n)||q_\phi(z)p(n)]$, between the input and the latent variable, which is based on the empirical input distribution $q_\phi(z, n)$. The second term measures the dependence between the latent variables and the total correlation (TC). It is essential to produce representations that penalize the total correlation and force the model to discover statistically disentangled factors in the input distribution. The third term prevents the individual latent variables in the representation from deviating far from their priors. It is called dimensionwise KLD [50]. After adding the tuning parameters to the decomposition, the β -TC-modified KLD term becomes

$$\begin{aligned} L_{\beta\text{-TC}} &= \alpha I(z; x) + \beta D_{\text{KL}}\left[q_\phi(z)||\prod_j q_\phi(z_j)\right] \\ &+ \gamma \sum_j D_{\text{KL}}[q_\phi(z_j)||p(z_j)]. \end{aligned} \quad (\text{C5})$$

Modulating only the parameter β shows the most significant effect on disentanglement in the latent representation given by empirical evidence [49].

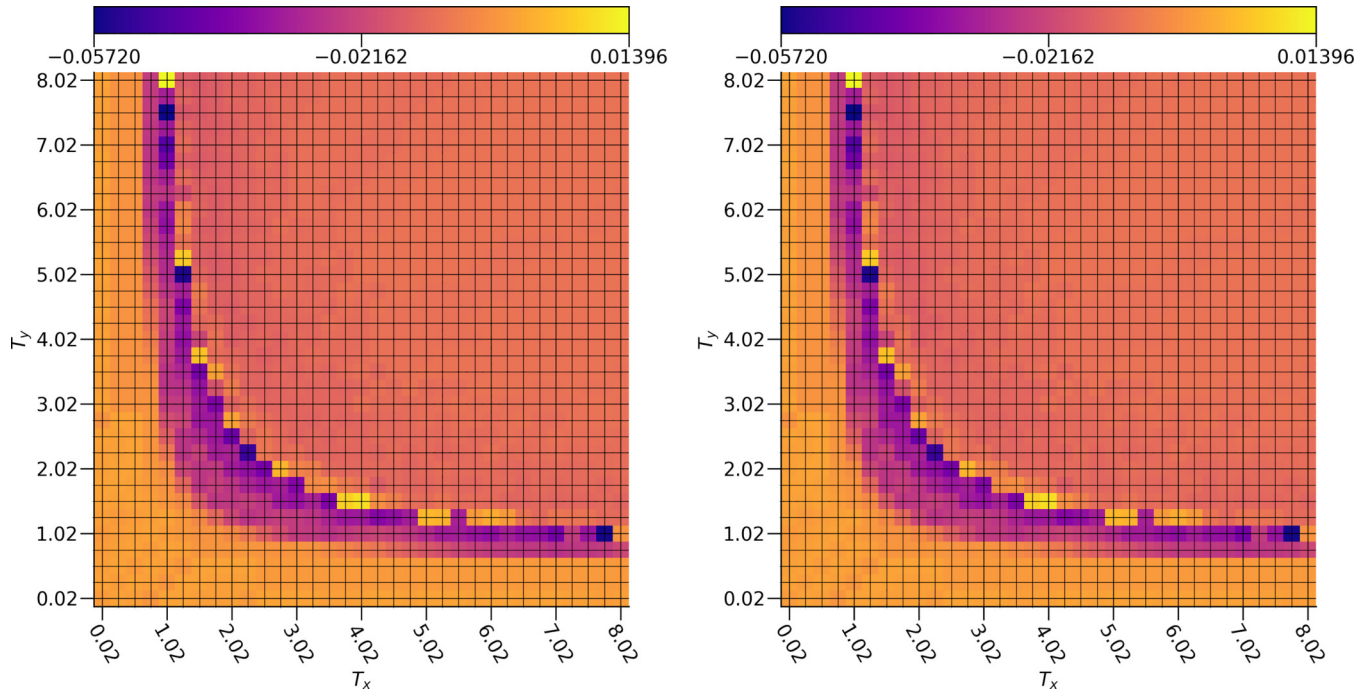


FIG. 6. The first principal component of the latent mean, v_0 , with respect to the different K_x and K_y [$1/(\beta J_x)$ and $1/(\beta J_y)$] for the 2D square lattice. The left and right panels are for the system sizes $N = M = 64$ and $N = M = 128$, respectively.

APPENDIX D: FIRST PRINCIPAL COMPONENT OF THE LATENT MEAN

The first principal component of the latent mean v_0 is shown in Fig. 6. As the samples in the paramagnetic phase

are noisy from thermal fluctuations that destroy any possible order, the v_0 remains somewhat uniform in the whole phase-space diagram. However, it varies slightly near the phase transition region without clearly demarcating the two phases, as seen in Fig. 6.

-
- [1] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. V. Essen, A. A. S. Awwal, and V. K. Asari, *Electronics* **8**, 292 (2019).
- [2] S. J. Wetzel, *Phys. Rev. E* **96**, 022140 (2017).
- [3] J. Carrasquilla and R. G. Melko, *Nat. Phys.* **13**, 431 (2017).
- [4] G. Pilania, J. E. Gubernatis, and T. Lookman, *Phys. Rev. B* **91**, 214302 (2015).
- [5] N. Walker, K.-M. Tam, B. Novak, and M. Jarrell, *Phys. Rev. E* **98**, 053305 (2018).
- [6] L. Wang, *Phys. Rev. B* **94**, 195105 (2016).
- [7] W. Hu, R. R. P. Singh, and R. T. Scalettar, *Phys. Rev. E* **95**, 062122 (2017).
- [8] S. J. Wetzel and M. Scherzer, *Phys. Rev. B* **96**, 184410 (2017).
- [9] L. Huang and L. Wang, *Phys. Rev. B* **95**, 035105 (2017).
- [10] G. Torlai and R. G. Melko, *Phys. Rev. B* **94**, 165134 (2016).
- [11] P. Mehta and D. J. Schwab, *arXiv:1410.3831*.
- [12] L. D. Landau, *Phys. Z. Sowjetunion* **11**, 26 (1937).
- [13] M. E. Fisher, *Rev. Mod. Phys.* **46**, 597 (1974).
- [14] L. Savary and L. Balents, *Rep. Prog. Phys.* **80**, 016502 (2016).
- [15] J. Schmidhuber, *Neural Networks* **61**, 85 (2015).
- [16] T. O. Ayodele, *New Advances in Machine Learning* (Yagang Zhang, Rijeka, 2010), pp. 19–48.
- [17] K. Pearson, *London Edinburgh Philos. Mag. J. Sci.* **2**, 559 (1901).
- [18] Y. Wang, H. Yao, and S. Zhao, *Neurocomputing* **184**, 232 (2016).
- [19] J. Wang, H. He, and D. V. Prokhorov, *Procedia Comput. Sci.* **13**, 120 (2012).
- [20] J. Almotiri, K. Elleithy, and A. Elleithy, in *2017 IEEE Long Island Systems, Applications and Technology Conference (LISAT)* (IEEE, New York, 2017), pp. 1–5.
- [21] E. Plaut, *arXiv:1804.10253*.
- [22] D. P. Kingma and M. Welling, *arXiv:1312.6114*.
- [23] D. P. Kingma and M. Welling, *Found. Trends Mach. Learn.* **12**, 307 (2019).
- [24] C. Alexandrou, A. Athenodorou, C. Chrysostomou, and S. Paul, *Eur. Phys. J. B* **93**, 226 (2020).
- [25] N. Walker, K.-M. Tam, and M. Jarrell, *Sci. Rep.* **10**, 13047 (2020).
- [26] D. Kim and D.-H. Kim, *Phys. Rev. E* **98**, 022138 (2018).
- [27] A. Morningstar and R. G. Melko, *J. Mach. Learn. Res.* **18**, 1 (2018).

- [28] K. Shiina, H. Mori, Y. Okabe, and H. K. Lee, *Sci. Rep.* **10**, 2177 (2020).
- [29] H. A. Kramers and G. H. Wannier, *Phys. Rev.* **60**, 263 (1941).
- [30] G. H. Wannier, *Rev. Mod. Phys.* **17**, 50 (1945).
- [31] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Elsevier, Amsterdam, 2016).
- [32] P. Pfeuty, *Ann. Phys. (NY)* **57**, 79 (1970).
- [33] R. B. Stinchcombe, *J. Phys. C: Solid State Phys.* **6**, 2459 (1973).
- [34] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University, Cambridge, England, 1995).
- [35] M. Suzuki, *Quantum Monte Carlo Methods in Equilibrium and Nonequilibrium Systems*, Springer Series in Solid-State Sciences (Springer, Berlin, Heidelberg, 1984), Vol. 74.
- [36] M. Suzuki, *Prog. Theor. Phys.* **46**, 1337 (1971).
- [37] B. K. Chakrabarti and A. Das, *Quantum Annealing and Other Optimization Methods* (Springer, Berlin, 2005), pp. 1–36.
- [38] G. Van Rossum and F. L. Drake, Jr., *Python Reference Manual* (Centrum voor Wiskunde en Informatica, Amsterdam, 1995).
- [39] S. van der Walt, S. C. Colbert, and G. Varoquaux, *Comput. Sci. Eng.* **13**, 22 (2011).
- [40] Dask Development Team, Dask: Library for dynamic task scheduling (2016), <https://dask.org>.
- [41] S. K. Lam, A. Pitrou, and S. Seibert, in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, LLVM '15* (Association for Computing Machinery, New York, 2015), pp. 7:1–7:6.
- [42] K. A. Nicoli, S. Nakajima, N. Strodthoff, W. Samek, K.-R. Müller, and P. Kessel, *Phys. Rev. E* **101**, 023304 (2020).
- [43] J. Vielhaben and N. Strodthoff, *Phys. Rev. E* **103**, 063304 (2021).
- [44] C. Doersch, [arXiv:1606.05908](https://arxiv.org/abs/1606.05908).
- [45] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, in *International Conference on Machine Learning* (PMLR, 2016), pp. 1558–1566.
- [46] S. Kullback and R. A. Leibler, *Ann. Math. Stat.* **22**, 79 (1951).
- [47] L. Weng, [lilianweng.github.io/lil-log](https://github.com/lilianweng/lil-log), 2018.
- [48] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, [arXiv:1804.03599](https://arxiv.org/abs/1804.03599).
- [49] R. T. Q. Chen, X. Li, R. Grosse, and D. Duvenaud, in *Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18* (Curran Associates Inc., Red Hook, NY, 2018), p. 2615.
- [50] N. Walker, Ph.D. thesis, Louisiana State University, 2020.
- [51] W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, *Appl. Opt.* **29**, 4790 (1990).
- [52] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17* (Curran Associates Inc., Red Hook, NY, 2017), p. 972.
- [53] S. Ruder, [arXiv:1609.04747](https://arxiv.org/abs/1609.04747).
- [54] F. Chollet *et al.*, Keras, <https://keras.io>, 2015.
- [55] L. van der Maaten and G. Hinton, *J. Mach. Learn. Res.* **9**, 2579 (2008).
- [56] J. D. Hunter, *Comput. Sci. Eng.* **9**, 90 (2007).
- [57] L. Onsager, *Phys. Rev.* **65**, 117 (1944).
- [58] H. Betsuyaku, *Prog. Theor. Phys.* **73**, 319 (1985).
- [59] J. E. Hirsch, *Phys. Rev. B* **28**, 4059 (1983).
- [60] J. E. Hirsch and R. M. Fye, *Phys. Rev. Lett.* **56**, 2521 (1986).
- [61] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, *Phys. Rev. D* **24**, 2278 (1981).
- [62] T. A. Maier, M. Jarrell, T. Prushke, and M. Hettler, *Rev. Mod. Phys.* **77**, 1027 (2005).
- [63] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996).
- [64] N. S. Vidhyadhiraja, A. Macridin, C. Şen, M. Jarrell, and M. Ma, *Phys. Rev. Lett.* **102**, 206407 (2009).
- [65] S. Kellar, K.-M. Tam, and J. Moreno, *Crystals* **13**, 106 (2023).
- [66] J. Cardy, *Finite-Size Scaling*, Current Physics (North-Holland, Amsterdam, 1988).
- [67] M. E. Fisher and M. N. Barber, *Phys. Rev. Lett.* **28**, 1516 (1972).
- [68] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Commun. ACM* **63**, 139 (2020).
- [69] N. Walker and K.-M. Tam, *Mach. Learn.: Sci. Technol.* **2**, 025001 (2021).
- [70] Louisiana Optical Network Initiative, <http://www.loni.org>.
- [71] J. B. Kogut, *Rev. Mod. Phys.* **51**, 659 (1979).
- [72] A. Muramatsu, The Ising model, duality, and transfer matrix, in Lattice Gauge Theory lecture notes (unpublished).