Recognizing the topological phase transition by Variational Autoregressive Networks

Lingxiao Wang1,2, Yin Jiang1,3, Lianyi He†,1 and Kai Zhou§2

1Department of Physics, Tsinghua University and Collaborative Innovation Center of Quantum Matter, Beijing 100084, China.
2Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany
3Department of Physics, Beihang University, Beijing 100191, China.

(Dated: May 12, 2020)

A generative model, the Variational Autoregressive Networks(VANs) is introduced to recognize the Kosterlitz-Thouless phase transition on a periodic square lattice. Vortices as the quasi-Long Range Order(LRO) are accurately detected by the ordering neural networks. By learning the microscopic possibility distributions from macroscopic thermal distribution, the free energy is directly calculated and vortices(anti-vortices) pair emerge with temperature rising. As a more precise estimation, the helicity modulus is evaluated to anchor the transition temperature to be at $\beta_c = 1.101$. Although time-consuming of the training process is inevitably increasing with the lattice size, the training time remains unchanged around the transition temperature.

**Introduction**—Machine Learning(ML) techniques are attracting a widespread interests in different fields, since its power to extract and express structures inside complex data. The machine learning is permeating all fields of physics as projects including the classification, regression and generating patterns [11][2]. Remarkably, it is found that the neural networks can classify phase structures in both condensed matter and high energy physics [3][4]. As for the regression, the event selection in a large data set e.g. from the LHCb [5][6], the spinodal decomposition in heavy ion collisions [7], and the molecular structure prediction [9] are successful applications. Furthermore, the machine learning sheds light on the innovation of the first principle calculation, e.g. the many-body system computation has been spurred by its development. The Restricted Boltzmann Machine(RBM) was applied in solving quantum many-body system firstly [10], and a deep neural network was constructed to derive solutions of the many-electron schrödinger equation [11], where proper neural networks can work as an Ansatz to represent systems as efficient as possible. Even an open quantum many-body system could be represented by the neural networks [12][13]. Modifying the classical algorithms with machine learning is another potential direction [14][16][17], which is routinely adopted to improve or assist the conventional calculations.

In addition, it’s natural to set the lattice simulation as a platform for applying neural networks, since they share similar discrete architectures. The advantage of the lattice computing is that it can discretize the problems on finite sites and get the correct physical results under the continuous limit eventually. There are some meritorious attempts [17][21] with training data generated by the standard Markov Chain Monte Carlo(MCMC) method, which constrains its expandability and efficiency since the critical slowing down near the critical point [21] and the sign-problem [20]. Recently, a new method based on an autoregressive neural network was proposed and applied in discrete spin systems [22][23]. By decomposing the macroscopic probability distribution onto the microscopic lattice site by site to be the variational ansatz, it achieved a higher accuracy in solving several Ising-type systems. As a matter of fact, the Variational Autoregressive Networks(VANs) is a typical generative model which has extraordinary potentialities in the lattice calculation as it already shows in the image processing projects [24].

Inspired by the successful application of machine learning, the topological phase transitions can be explored with the state-of-the art ML approaches. Different from the classical phase transition, the topological phase transition occurs with topological defects emerging, which has been continuously attracting the attention from various fields of physics [25]. Some related efforts in both supervised learning and unsupervised learning have been attempted [20][29]. The winding numbers were recognized by a supervised trained neural network for one-dimensional insulator model [28]. By generating the configurations with MCMC sampling and supplying feature engineered vortex configurations as the input, neural networks could detect the topological phase transition from well-preprocessed configurations [27]. In another attempt [30] the Variational Autoencoder was employed with augmented objective function, even the bulk magnetizations are not quantitatively captured by the configurations generated from their decoder.

In this letter, we apply the VANs to the task of recognizing the topological phase transition with continuous variables in unsupervised manner. As a concise and reference example is the two-dimensional XY model, which exhibits a Kosterlitz-Thouless (KT) phase transition [31][32][33]. In the autoregressive neural networks, the microscopic state on each lattice site is probabilistically modeled in order, which constructs a joint probability from
these conditional probability distribution for the whole configuration [22, 23]. In the following parts, a generic VANs framework is introduced and the neural networks suitable for the XY model is constructed. The signal of the KT phase transition, vortices, are automatically generated by the neural networks. Correspondingly, a more accurate estimation of the transition temperature for the KT phase transition in 2-D XY model is given by calculating the helicity modulus [33]. As for the time-consuming, despite of the increasing trend with lattice size enlarging, the training time remains unchanged around the transition point. Considering the advantages of the VANs, the latent applications to the other physical systems are proposed in the final part.

Variational Autoregressive Networks.—The Hamiltonian of the 2-dimensional XY model on the lattice is expressed by spins living on the lattice sites with nearest-neighbor interactions

\[ H = -J \sum_{<i,j>} s_i s_j = -J \sum_{<i,j>} \cos(\phi_i - \phi_j) \] (1)

where \(<i,j>\) indicates that the sum is taken over all nearest-neighbor pairs and the angle \(\phi_i \in [0, 2\pi]\) denotes the spin orientation on site \(i\). The Mermin-Wagner theorem states that a long-range ordered (LRO) phase cannot exist in two dimensional systems with continuous degrees of freedom, since the fluctuations breaks the order [34]. Nevertheless the formation of topological defects (i.e. vortices/anti-vortices) in the XY model brings in a quasi-LRO phase, which characterizes the global property of the many-body system.

To detect the KT phase transition in the XY model, in statistical mechanics the free energy \(F = -(1/\beta) \ln Z\) should be concerned, where \(\beta \equiv 1/T\) is the inverse temperature. The free energy is constructed by the partition function \(Z \equiv \sum_s \exp(-\beta E(s))\), which contains all information about the system. The summation covers all possible configurations \(\{s\}\) of the system. Monte Carlo algorithms are routinely applied to generate the configurations, which can achieve the proper relative importance among configurations. However the free energy can not be computed directly from the algorithm. Variational approaches were proposed to solve the problem and with recent extension of the variational Ansatz to autoregressive neural networks as advanced in VAN [22], which is an effective approximate method widely applied in the many-body systems. In this paper, the variational target function is the joint probability of the configurations, which follows the Boltzmann distribution \(p(s) = e^{-\beta E(s)}/Z\). The configurations \(s = \{s_1, s_2, ..., s_N\}\) with continuous spins are defined on the lattice with \(N\) sites. As a variational Ansatz, the joint distribution \(q_\theta(s)\) are parametrized by variational parameters \(\theta\) and tuned to approach the target distribution \(p(s)\).

The Kullback-Leibler (KL) divergence [35] between the variational and the target distribution, \(D_{KL}(q_\theta||p) \equiv \mathbb{E}_{s \sim q_\theta}(-\log p + \log q_\theta)\), provides the measure of the closeness from \(q_\theta\) to \(p\). The corresponding variational free energy derives from \(D_{KL}(q_\theta||p) = \beta \left( F_\theta - F \right)\), that’s

\[ F_\theta \equiv \langle 1/\beta \rangle \sum_s q_\theta(s) \left[ \beta E(s) + \ln q_\theta(s) \right] \] (2)

Since \(D_{KL}(q_\theta||p)\) is non-negative, the variational free energy \(F_\theta\) is always an upper bound of the true free energy. Thus the minimization on \(D_{KL}(q_\theta||p)\) and the variational free energy \(F_\theta\) are equivalent. Meanwhile, as several works pointed out [10, 22, 23], it’s directly to map the parameters onto the weights of an Artificial Neural Network (ANN), then the variational free energy is the loss function.

\[ F_\theta = \sum_{i=1}^{N} q_\theta(s_i|s_1, \ldots, s_{i-1}) \] (3)

which provides the variational Ansatz by parametrizing each conditional probability as neural networks. In XY model, considering the fact that the orientation of spin changes continuously, each conditional probability factor \(q(s_i)\) should be continuous on each site. As a proper choice, we propose to take the mixture of beta distribution \(X \sim Beta(a, b)\) which ensures that the random variables distribute in a finite interval. Furthermore, in the Bayesian inference the beta distribution is the conjugate
prior probability distribution of the Bernoulli distribution, which has been proven to be a proper distribution for Ising model case. The Beta(a, b) is continuously defined in a finite interval with two positive shape parameters (a, b). Thus the hidden layers of neural networks are designed to be two channels type for each Beta component, and the conditional probabilities are derived as

$$q_\theta(s_i|s_1, ..., s_{i-1}) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} s_i^{a_i-1}(1-s_i)^{b_i-1}$$  \hspace{1cm} (4)$$

where \(\Gamma(a)\) is the gamma function, and \(s_i = \theta_i/2\pi \in [0, 1]\). \((a_i, b_i) > 0\). The outputs of the hidden layers are \((a, b)\), which can be realized by Fig. 1.

To match the two dimensional structures for the system, the PixelCNN was employed which can preserve naturally the locality and the translational symmetry. In addition, the autoregressive property is guaranteed by putting a mask on the convolution kernel, so that the weights are non-zero for half of the kernel, and each conditional probability \(q_\theta(s_i)\) is independent of \(s_j\) with \(j < i\) for a prechosen ordering. As Fig. 1 shows, the input layer takes in the configurations \(s\) on the lattice, and after passing it through several masked convolution layers, the parameters of the Beta distribution on each site are obtained in the output layer and thus the configuration probability \(q_\theta(s)\) can be derived, with which the variational free energy can be further calculated via Eq. 2 after such a forward propagation on a batch of independent configurations. Specifying channels in the convolution layers to represent parameters of each Beta component is found to be effective in saving the training time and speeding up the sampling later. Training the neural networks here is the key to perform the variational approach. With a classical back-propagation algorithm, the gradient of the loss function (i.e. variational free energy) with respect to network parameters is needed, which after employing the log-derivative trick \[37\] reads

$$\beta \nabla_\theta \mathcal{F}_Q = \mathbb{E}_{x \sim q_\theta(s)} \left\{ [\beta E(s) + \ln q_\theta(s)] \nabla_\theta \ln q_\theta(s) \right\}$$  \hspace{1cm} (5)$$

where the gradient \(\nabla_\theta \ln q_\theta(s)\) is weighted by the reward signal \(\beta E(s) + \ln q_\theta(s)\).

The nuts-and-bolts VANs computation is implemented by the following procedures: With the randomly initialized network, sample independently a batch of configurations to be the training set; Forward pass the training set to evaluate their log-probability and the variational free energy \(F_Q\); Estimate the gradient \(\beta \nabla_\theta \mathcal{F}_Q\) and update the network weights via back-propagation; With the updated network re-sample a batch of configurations to be the new training set, which actually follow the current joint probability \(q\); Repeat the above until the loss function is convergent; Sample ensemble of configurations from \(q\) independently site by site at once; Calculate the thermodynamic observables. Although the criteria of the convergence \(\epsilon\) is not rigorously defined, the difference of the energy between only one site changing or not could be an approximate superior limit: \(\epsilon \ll 4J/N\) on the square lattice with \(N\) sites.

Rediscovery KT Phase Transition.—As mentioned earlier, the thermodynamic observables in 2-d XY model have been well-computed in numerous MCMC works.\[33\]\[35\]\[39\]. Since that, it’s necessary to compare the results in the VANs and the MCMC. In the following calculations, the default setup of the network we adopted in VANs is with width and depth as \((32, 3)\). Here 12 multi-channel is used to construct a mixture Beta function, which helps the network to be more expressive \[40\]. The Adam optimizer is applied to minimize the loss function in PyTorch. The implementation of the VANs is available at Ref. \[41\]. The corresponding hyperparameters are consistent with the former works.\[22\] As for computation time-consuming reference: with batch size 1000 for a 16 x 16 square lattice with periodic boundary, a typical training step cost 0.198 second on a single NVIDIA RTX 2080 GPU. On the other side, the MCMC was implemented in a classical algorithm \[42\] \[43\] with 50000 warm-up steps to reach the equilibrium, and the energy was computed from 1000 configurations sampled from each 10000 steps in equilibrium.

The variational free energy per site (density) is presented in Fig. 2 where the results are divided into four different lattice sizes, \(L = 4, 8, 16\). With the lattice size increasing, the results of \(L = 8\) and \(L = 16\) indicates that the free energy converges rapidly. This ensures that the size effect can be avoided later for larger size \(L = 16\) in the following discussion. It should be added that although it’s difficult to calculate the free energy directly in the MCMC, the similar variational methods can also be applied in the XY model with a MCMC updating process, which has been discussed in the different models\[41\].

![Fig. 2: The free energy per site of the XY model on a square lattice with periodic boundary condition.](image)
algorithm. At the low temperature area (\(\beta > 1\)), the VANs results are in line with the MCMC even containing the statistical error from thermal fluctuations. At the high temperature area (\(\beta < 1\)), the results are not perfectly matched with the MCMC. It’s understandable by inspecting the vortices as shown in Table I where the vortices (anti-vortices) pair density \(n = v/(2L^2)\) with the vorticity \(v = (1/2\pi) \oint_C \nabla \phi(r) \cdot dr\) are evaluated and compared between VAN and MCMC. Since the vortex emerges with higher entropy in XY model, configurations with more vortices have lower free energy than the less. That’s the reason why the energy in VANs is higher than it in MCMC at \(\beta < 0.7\), for more vortices balance the free energy. And the situation is reversed at \(0.7 < \beta < 1\) for the same reason.

![Energy density distribution](image)

**FIG. 3:** The energy density distributes with the inverse temperature \(\beta\) in the MCMC method and the VANs. The size of lattice is \(16 \times 16\), and the shadow areas label the standard deviation from thermal fluctuations. The vortices pair density extracted in the VANs is posted at the northeast corner.

| Inverse temperature | MCMC | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 |
|---------------------|------|-----|-----|-----|-----|-----|
| Energy density      | MCMC | -0.424 | -0.682 | -0.996 | -1.336 | -1.502 |
|                     | VANs | -0.376 | -0.577 | -1.025 | -1.359 | -1.498 |
| Vortices \(n\)      | MCMC | 0.114 | 0.080 | 0.042 | 0.010 | 0.002 |
|                     | VANs | 0.122 | 0.096 | 0.035 | 0.004 | 0.001 |

**TABLE I:** Vortices. The vortices (anti-vortices) pair density \(n\) can be extracted from the well-trained networks. After the variational free energy converging, 1000 configurations are sampled from the networks, and the following values are from ensemble average. The non-zero value of the vortices density at high temperature suggests that there is a topological phase transition in \(\beta = 1 \sim 1.2\) interval.

From other works \[31, 39, 45\], a relative accurate transition temperature is reported as \(\beta_{KT} = 1/T_{KT} \approx 1.12\), above which the dominate degrees of freedom in the system are vortices. In our work, the implemented VANs captures the global property of the spin system since the trained neural networks helps achieving a good evaluation for the free energy. Nevertheless, the disorder of configurations in the thermal fluctuation results in a slight mismatching at high temperature. As the temperature rises, the long range correlation becomes exponentially attenuating \[32\], which slightly weakens the expressive ability of VANs in finite size case. Additionally, it must be emphasized that the elementary conditional distribution on each site \(q_\theta(s_i|s_{<i})\) should be carefully chosen, for the spin in the XY model has a periodic value rather than an infinite interval. We found as a common test that the choice of the normal distribution for \(q_\theta(s_i|s_{<i})\) hardly converge the loss function in a reasonable value.

**Phase Transition Point.**—To recognize the transition point of the Kosterlitz-Thouless phase transition quantitatively, the spin stiffness \(\rho_s\) could be introduced, which reflects the change in the ground state of a spin system as a result of importing a slow twist \(\delta \phi\) on the spins. In the continuous limitation, it is \(\rho_s = [\partial^2 F(\delta \phi)/\partial (\delta \phi)^2]|_{\delta \phi=0}\). For the sake of easing the computing demands, the helicity modulus \[31, 33, 38, 39\] is a proper substitute, and they are equivalent in the small angular limitation.

\[
\gamma(L) = -\frac{E}{2L^2} - \frac{J\beta}{L^2}\left(\sum_{<i,j>} \sin(\phi_i - \phi_j)\vec{e}_{ij}\vec{x}\right)^2
\]

where \(L\) is the size of the square lattice, \(\vec{e}_{ij}\) is the vector pointing from site \(j\) to site \(i\), \(\vec{x}\) is an unit vector of a fixed direction in the lattice plane (the trivial choice is \(x, y\) on the square lattice). The Kosterlitz renormalization-group \[32\] predicts that \(\gamma(L \to \infty)\) jumps from the value \(2T_c/\pi\) to zero at the critical temperature, thus the helicity modulus gives a reliable estimation of the phase transition point.

![Helicity modulus distribution](image)

**FIG. 4:** The helicity modulus distributes with the inverse temperature \(\beta\) in the VANs, and the cross point is \(\beta_c \approx 1.101\) in lattice size \(L = 16\) with 100 channels mixture Beta function.

In Fig. 4 the evaluated helicity modulus from the VANs is shown for lattice size \(L = 16\) with multi-channel
Beta function respectively. The markers are the numerical results from the VANs with multi-channel mixture Beta function and the dashed lines show their fitting curve which is used to anchor the crossing point with the $2/(\pi \beta)$ line. We observe that the crossing point for $L = 16$ is $\beta_c \simeq 1.101$, and with channels increasing the point moves towards the $\beta_{KT}$. Since the helicity modulus depends on the correlation function which needs a higher order statistics than the energy, small-sized lattices are not considered here given the larger sized results can give more precise evaluation, which is observed from standard Monte Carlo simulation [27, 39]. For the time-consuming in the case $L = 16$, it remains 0.198 seconds as per training step independent of the temperature values. It also perfectly support parallel sampling from the trained network on GPU. These suggest that the Critical Slowing Down(CSD) is hopefully avoided.

Even though the burden due to the increased auto-correlation time [21, 46] in MCMC doesn’t appear in the VANs, the training cost with increasing lattice size should be mentioned. As shown in Fig. 5, the circles are the training time per step for different lattice size calculated via networks with width and depth $(32, 3)$ near the transition temperature $\beta = 1.12$. The dashed line is the fitting curve with the form $t(L) = a L^b$, and the red dot ($L=24$) is not used in fitting. Here the cost of bypassing the CSD is introducing a training time which has polynomial dependence on $L$ approximately but one-off since the following sampling can take on parallel advantage of GPU for large ensemble generation. A more powerful GPU can reduce the time-consuming. In our case, the results in Fig. 5 are obtained on a Nvidia RTX 2080 GPU, which is 4 times faster than on a Nvidia RTX 2070 Max-Q GPU approximately.

![FIG. 5: The time-consuming increases with the lattice size $L$ at inverse temperature $\beta = 1.12$.](image)

**Discussions.**—In this work, an autoregressive neural network is designed to be an Ansatz of the variational approach for investigating the topological phase transition with continuous variables within 2-d XY model. The Variational Autoregressive Networks learn to construct microscopic states of the spin system accurately, in which vortices (anti-vortices) emerge automatically. The energy and vorticity density are calculated with configurations generated from the networks, furthermore the comparison with MCMC algorithm indicates that the VANs tends to extract dominant collective degrees of freedom in XY model, vortices, which are crucial at high temperature. The autoregressive structure of the neural networks is beneficial to mine the long range correlation even beyond the phase transition point. It brings an opportunity that more latent topological structures can be investigated, such as in a coupled XY model [45] or in a twisted bilayer graphene [47], where new long range correlations emerge. Besides, a straightforward estimation to the transition point of the KT phase transition in the VANs is shown to be consistent with prior works. Although the time-consuming with size increasing is unavoidable, with the help of the powerful GPU computation, searching critical point becomes more economical in the limit of generating big ensemble of configurations. In complicated many-body system, the critical slowing down problem in MCMC is expected to be alleviated, which will help i.e., Lattice QCD to reach possible critical end point region. The $\phi^4$ model could be a practical step [21], in which the calculation accuracy and practicability should be rigorously examined.

In a short summary, the machine learning approach, especially with well-designed neural networks can match with specific physical problems, such as the VANs to the topological phase transitions shown in this work, the RNN to the system with time reversal symmetry [48] and the DNN to the renormalization group approaches [49]. This inspire us to explore the Machine Learning techniques in a more physical viewpoint, which will help us open the black box of machine learning and nature.

We thank Giuseppe Carleo and Junwei Liu for useful discussions. The work of this research is supported by the BMBF under the ErUM-Data project (K. Z.), by the AI grant of SAMSON AG, Frankfurt (K. Z.), by the National Natural Science Foundation of China, Grant No. 11875002(Y.J.) and No.11775123 (L.H. and L.W.), by the Zhoubai Program of Beihang University(Y.J.), by the National Key R&D Program of China, Grant No. 2018YFA0306503 (L.H.). K.Z. also thanks the donation of NVIDIA GPUs by NVIDIA Corporation for research.

**Bibliography**

[1] M. Buchanan, *Nat. Phys.* **15**, 1208 (2019)
[2] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Rev. Mod. Phys. 91, 045002 (2019)
[3] L. Wang, Phys. Rev. B 94, 195105 (2016)
[4] J. Carrasquilla and R. G. Melko, Nat. Phys. 13, 431 (2017)
[5] L.-G. Pang, K. Zhou, N. Su, H. Petersen, H. Stöcker, and X.-N. Wang, Nat. Commun. 9, 210 (2018)
[6] E. M. Metodiev and J. Thaler, Phys. Rev. Lett. 122, 241602 (2018)
[7] G. Kasieczka, T. Plehn, A. Butter, K. Cranmer, D. Debnath, M. D. Dillon, M. Fairbairn, D. A. Faroughy, W. Fedorko, C. Gay, L. Gouskos, J. F. Kamenik, P. Komiske, S. Leiss, A. Lister, S. Macaluso, E. Metodiev, L. Moore, B. Nachman, K. Nordström, J. Parkes, H. Qu, Y. Rath, M. Rieger, D. Shih, J. Thompson, and S. Varma, SciPost Phys. 7, 014 (2019)
[8] S. Background, Topological Phase Transitions and Topological Phases of Matter, Tech. Rep. (the Royal Swedish Academy of Sciences, 2016).
[9] C. Wang and H. Zhai, Phys. Rev. B 96, 144432 (2017)
[10] G. Carleo and M. Troyer, Science 355, 602 (2017)
[11] D. Pfau, J. S. Spencer, A. G. d. G. Matthews, and W. M. C. Foulkes, ArXiv190300804 Cond-Mat Physicshep-Th (2019), arXiv:1903.00804 [cond-mat, physics:hep-th].
[12] A. Nagy and V. Savona, Phys. Rev. Lett. 122, 250502 (2019)
[13] R. Gupta, J. DeLapp, G. G. Batrouni, G. C. Fox, C. F. Bailleil, and J. Apostolakis, Phys. Rev. Lett. 61, 1996 (1988)
[14] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Phys. Rev. B 97, 094505 (2018)
[15] H. Shen, J. Liu, and L. Fu, Phys. Rev. B 97, 205140 (2018)
[16] Y. Mori, K. Kashiwa, and A. Ohnishi, Prog Theor Exp Phys 2018 (2018), 10.1093/ptep/ptx191
[17] K. Zhou, G. Endrödi, L.-G. Pang, and H. Stöcker, Phys. Rev. D 100, 011501 (2019)
[18] B. Nachman, K. Nordström, J. Pearkes, H. Qu, Y. Rath, M. Rieger, D. Shih, J. Thompson, and S. Varma, SciPost Phys. 7, 014 (2019)
[19] D. Wu, L. Wang, and P. Zhang, Phys. Rev. Lett. 122, 080602 (2019)
[20] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, Phys. Rev. Lett. 124, 020503 (2020)
[21] Z. Ou, ArXiv180801630 Cs Stat (2019), arXiv:1808.01630 [cs, stat]
[22] S. Background, Topological Phase Transitions and Topological Phases of Matter, Tech. Rep. (the Royal Swedish Academy of Sciences, 2016).
[23] C. Wang and H. Zhai, Phys. Rev. B 96, 144432 (2017)