Condensed Matter Physics, 2018, Vol. 21, No 3, 33602: 1–11
DOI: 10.5488/CMP.21.33602
http://www.icmp.lviv.ua/journal

# A machine learning approach to the Berezinskii-Kosterlitz-Thouless transition in classical and quantum models

M. Richter-Laskowska[1], H. Khan[2], N. Trivedi[2], M.M. Maśka[1]

[1] Institute of Physics, University of Silesia, 75 Pułku Piechoty 1, 41-500 Chorzów, Poland

[2] Department of Physics, The Ohio State University, 191 W. Woodruff Ave., Columbus, Ohio 43210, USA

Received May 31, 2018, in final form August 16, 2018

The Berezinskii-Kosterlitz-Thouless transition is a very specific phase transition where all thermodynamic quantities are smooth. Therefore, it is difficult to determine the critical temperature in a precise way. In this paper we demonstrate how neural networks can be used to perform this task. In particular, we study how the accuracy of the transition identification depends on the way the neural networks are trained. We apply our approach to three different systems: (i) the classical XY model, (ii) the phase-fermion model, where classical and quantum degrees of freedom are coupled and (iii) the quantum XY model.

**Key words:** *phase transitions, topological defects, XY model, artificial neural networks, machine learning*

**PACS:** *64.60.-i, 05.70.Fh, 07.05.Mh*

## 1. Introduction

In many cases, thermodynamic phase transitions are clearly visible with well defined and easily identifiable critical points. In the Landau picture, we define a macroscopic order parameter that is nonzero in the ordered phase and vanishes when we cross the critical temperature. Typically, the transition is signaled by a discontinuity or divergence of some thermodynamic quantities such as specific heat or magnetic susceptibility. There also exist unconventional phase transitions which are much more difficult to find. One example are topological phase transitions, connected to the formation of topological defects, such as dislocations in two-dimensional crystals, vortices in two-dimensional superconductors and so on. The proliferation of these defects leads to the Berezinskii-Kosterlitz-Thouless (BKT) phase transition [1–4] where thermodynamic quantities behave smoothly. Therefore, traditional detection methods based on, e.g., the divergence of the specific heat or the spin susceptibility, cannot be used. In numerical approaches, the main difficulty with the identification of the BKT transition stems from the fact that the interaction between the topological charges depends logarithmically on the spatial separation. Therefore, numerical results converge very slowly with the size of the system, and a precise determination of the critical temperature is a computationally challenging task. The standard approach is based on the scaling properties of, e.g., the spin stiffness or superfluid density. This is particularly difficult for quantum models, where numerical methods are usually very involved and memory- and time-consuming.

Besides the traditional methods which rely on the idea of the order parameter or quantities accessible in experiments, such as the specific heat or magnetic susceptibility, attempts to use artificial neutral networks to identify phases in condensed matter and transition between them have been recently undertaken [5–20]. These new methods have proven to be accurate and reliable for classical Ising-type models [5–7]. Less spectacular progress has been made for quantum systems [15, 16] and for systems with topological phase transitions [17–20]. In the latter case, the main difficulty comes from the fact that the topologically ordered phase is not described by a local order parameter [1–4]. Instead, its formation is connected with suppression of non-local topological defects which are difficult to identify.

In this paper, we demonstrate the application of machine learning approaches to identify topological transitions in a few different types of two-dimensional classical and quantum systems. In particular, we study the classical XY (c-XY) model, the phase-fermion (PF) [21] where the interaction is only between quantum and classical degrees of freedom, and the fully quantum XY (q-XY) model.

The first of these models has already been thoroughly analyzed in [17]. The authors have shown there that treating spin configurations as raw images in the case of a feed-forward network does not lead to the correct value of the critical temperature. Instead, they propose to preprocess the spin configurations into vorticity and then use the results to train two kinds of artificial neural networks (ANN): a one-layer feed-forward network and a deep convolutional network. In both cases, the results are scaled with the system size towards the correct value of the critical temperature, but the one-layer network performed poorly for large systems. In the present approach, we do not have convolutional layers, but we use a deep feed-forward network composed of four fully-connected layers (we learned that the choice of particular meta-parameters is not crucial for the network performance). What is important is that, instead of using raw spin configurations where each spin is represented by a number from 0 to 2π (which gives rather inaccurate results, as demonstrated in [17]), we represent the configurations as vectors of sines and cosines of the spin angles, which reflects the system's symmetry.

## 2. Models

The first model describes classical spins of unit length in a square lattice with a nearest neighbor interaction given by the following Hamiltonian

$$H_{\text{c-XY}} = -J \sum_{\langle i,j\rangle} \cos\left(\theta_i - \theta_j\right), \tag{2.1}$$

where $J$ is the coupling between the spins, and $\theta_i$ describes the direction of spin $i$. It is known that this model exhibits the BKT phase transition, and precise finite-size scaling gives the critical temperature $T_{\text{BKT}} \approx 0.8935$ in units of $J$ [22].

In the next model, classical spins $\theta_i$ interact with fermions. The Hamiltonian of the PF model is given by

$$H_{\text{PF}} = -t \sum_{\langle i,j\rangle,\sigma} \hat{c}^{\dagger}_{i\sigma}\hat{c}_{j\sigma} + g\sum_i \left(\mathrm{e}^{\mathrm{i}\theta_i}\hat{c}_{i\uparrow}\hat{c}_{i\downarrow} + \text{h.c.}\right), \tag{2.2}$$

where $\hat{c}^{\dagger}_{i\sigma}$ ($\hat{c}_{i\sigma}$) is an operator that creates (annihilates) spin-$\sigma$ electron at the lattice site $i$, and $g$ describes the strength of the interaction between the classical and quantum degrees of freedom. We set the hopping integral $t$ as the energy unit ($t = 1$). In this model, the fermions mediate an effective interaction between the classical spins $\theta_i$ which also leads to the BKT phase transition. The critical temperature is a function of $g$, and for $g = 2$ Monte Carlo (MC) simulations give $T_{\text{BKT}} \approx 0.12$ [21]. The PF model can be treated as an approximation of the boson-fermion model [23], valid when fluctuations of the number of bosons at a lattice site can be neglected.

The Hamiltonian of the last model, the quantum XY model, is given by

$$H_{\text{q-XY}} = \frac{E_{\text{c}}}{2}\sum_i \hat{n}_i^2 - J\sum_{\langle i,j\rangle}\cos\left(\hat{\theta}_i - \hat{\theta}_j\right), \tag{2.3}$$

where $\hat{n}_i$ is the number operator that is canonically conjugate to the quantum phase operator $\hat{\theta}_i$, and $E_{\text{c}}$ is the charging energy.

To train neural networks and to classify phases, one needs extensive sets of spin configurations generated at different temperatures. They were produced with the help of the MC simulations. In the case of the classical XY model, we directly used the Metropolis algorithm. For the PF model, we also used the Metropolis algorithm, but in each MC step (i.e., for each generated spin configuration) we needed to diagonalize the fermionic Hamiltonian (2.2) [24]. For the quantum XY model, we used a quantum MC method in which the Hamiltonian was mapped to a classical action of spins on an effective (2+1)D lattice. We then used a Wolff cluster algorithm to simulate these spins. In all cases, the simulations were performed on 16 × 16 systems.

In all these models, the helicity modulus $\Upsilon$, at finite temperature defined as the second derivative of the free energy with respect to an externally imposed global twist across the sample [25], has a universal jump at the BKT transition. In the thermodynamic limit, it drops at $T_{\mathrm{BKT}}$ from $\frac{2}{\pi}T_{\mathrm{BKT}}$ to zero. However, in finite systems it evolves smoothly and converges only logarithmically to the thermodynamic limit, as shown in figures 1 (a), 1 (c), and 1 (e). A finite size scaling of the temperature at which $\Upsilon(T)$ crosses $\frac{2}{\pi}T$ can give a rough estimate of $T_{\mathrm{BKT}}$. This is demonstrated in figures 1 (b), 1 (d), and 1 (f). It is also possible to determine $T_{\mathrm{BKT}}$ in a more precise way. At the critical temperature, the helicity modulus scales with the system size according to the Kosterlitz renormalization group (RG) equations [26],

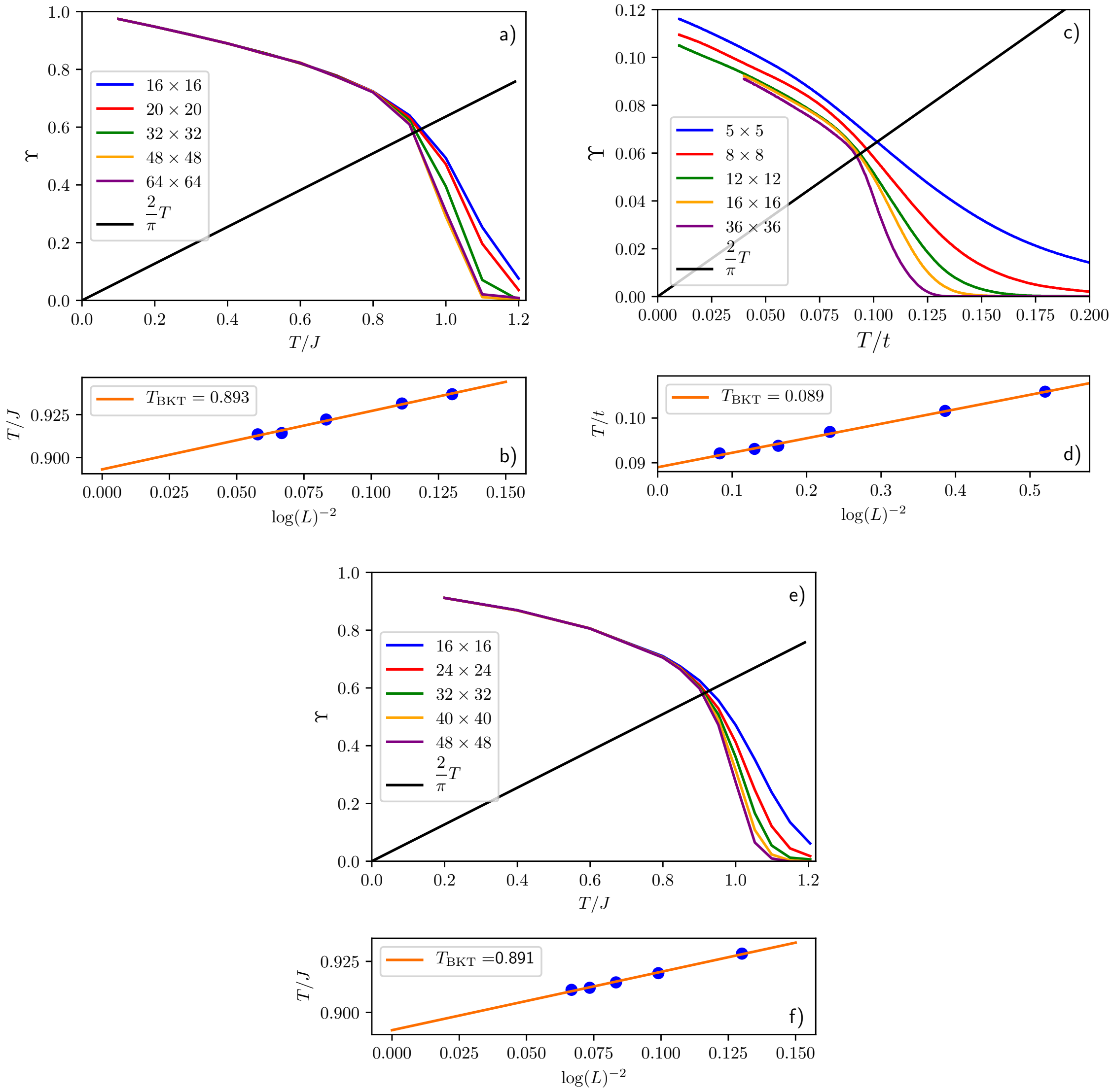

**Figure 1.** (Colour online) Helicity modulus for the c-XY model (a), PF model for $g = 4t$ (c), and q-XY model for $E_c = 0.1J$ (e). The corresponding critical temperatures extrapolated to the thermodynamic limit are presented in panels (b), (d), and (f). Since $T^*(L) - T_{\mathrm{BKT}} \propto \log(L)^{-2}$, where $T^*(L)$ is the temperature at which $\Upsilon$ drops the most rapidly in a $L \times L$ system [22, 27, 28], the temperature is presented in these panels as a function of $\log(L)^{-2}$.

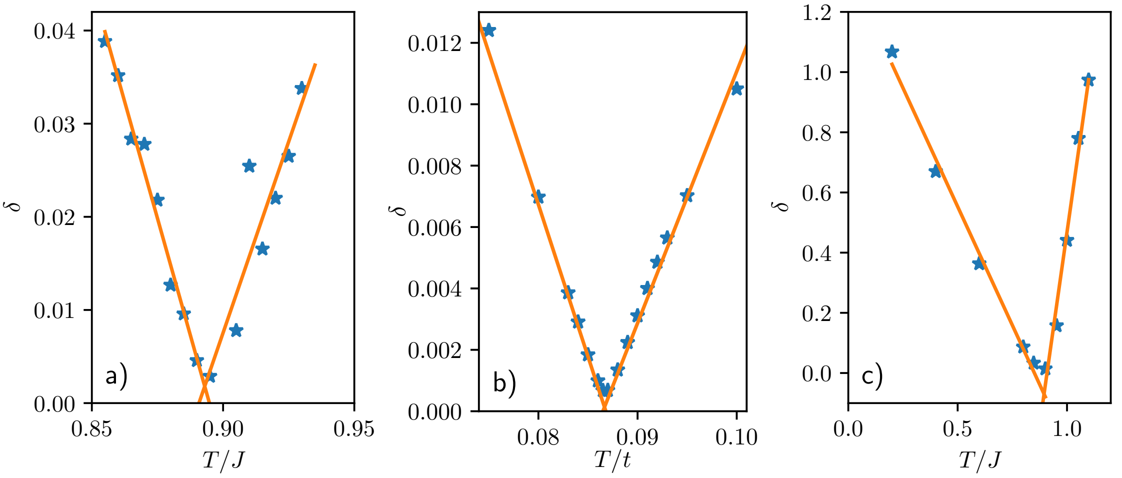

**Figure 2.** (Colour online) The root-mean-square error $\delta$ for fitting MC results for the c-XY model (a), the PF model (b), and the q-XY model (c) to the RG predictions given by equation (2.4). The sharp minimum of the fitting errors indicates $T_{\mathrm{BKT}}$. The model parameters are the same as in figure 1.

$$\Upsilon_L = \Upsilon_{L\to\infty}\left[1+\frac{1}{2}\frac{1}{\ln(L)+C}\right], \tag{2.4}$$

where $L$ is a linear size of the system and $C$ is a constant. Therefore, if one fits $\Upsilon_L(T)$ calculated in MC simulations for different system sizes $L$ to the RG predictions, the fitting errors drop almost to zero exactly at $T_{\mathrm{BKT}}$ [29]. In figure 2 one can see the procedure.

## 3. Artificial neural network

Here, however, we study how the critical temperature can be determined with the help of ANN trained to identify the low- and high-temperature phases. When a system approaches the critical temperature, thermal fluctuations increase drastically so that the spin configurations can be very different from configurations generated at very low temperatures. However, as long as $T$ is below $T_{\mathrm{BKT}}$, the system is still in the same low-temperature phase. The problem with systems where the BKT transition takes place, such as the c-XY, PF or q-XY models, is that in the low-temperature phase, MC simulations on finite clusters show finite magnetization, which according to the Mermin-Wagner theorem cannot exist in the thermodynamic limit. Since the MC results are used to train the ANN, it is possible that the network will learn finite-size features of the spin configurations, such as magnetization. It was proposed in [17] that this difficulty can be overcome by adding a convolutional layer designed to identify topological defects in spin configurations. Then, the network learns configurations of vortices and antivortices instead of raw spin configurations.

Our aim here is to show how the features that can be used to identify phases depend on the distance from the critical temperature $T_{\mathrm{BKT}}$. The standard ANN-based approach to the problem of finding a phase transition is to train the ANN to recognize features of the low- and high-temperature phases. Within the scheme of *supervised learning*, one needs to feed the ANN with labelled spin configurations generated at low and high temperatures. The network is supposed to extract characteristic patterns and to learn how to classify configurations which were not used at the training stage. Usually, this is an easy task for configurations generated at very low or at very high temperatures, since they are clearly distinctive. However, to precisely determine the critical temperature, the ANN must distinguish configurations generated slightly below $T_{\mathrm{BKT}}$ and slightly above $T_{\mathrm{BKT}}$. Thermal fluctuations in this regime make those configurations very different from the fully ordered low-temperature configurations and from completely random high-temperature configurations. The question then is, whether an ANN trained at extreme temperatures will be capable of classifying the phases close to $T_{\mathrm{BKT}}$? In other words, do the distinctive features learned by the ANN at the extremes balance each other out just right in the evaluation process such that the correct $T_{\mathrm{BKT}}$ is predicted? To answer this question we trained the ANN at different distances from $T_{\mathrm{BKT}}$ and checked how the distance $|T - T_{\mathrm{BKT}}|$ affects the accuracy of finding $T_{\mathrm{BKT}}$. To be precise, we used MC simulations to generate sets of configurations $\{C_1\}$, $\{C_2\}$, $\ldots$, $\{C_N\}$ at temperatures $T_1$, $T_2$, $\ldots$, $T_N$ ($T_1 < T_2 < \ldots < T_N$) below and above $T_{\mathrm{BKT}}$. For the c-XY, PF, and q-XY models, the ranges of temperatures were from $0.1J$ to $1.6J$, from $0.02t$ to $0.2t$, and from $0.1J$ to $1.5J$, respectively. Then, we generated sets of configurations representing different low- ($\mathcal{L}_m$) and high-temperature ($\mathcal{H}_m$) ranges:

$$\mathcal{L}_m = \bigcup_{i=1}^{m} C_i \quad \text{and} \quad \mathcal{H}_m = \bigcup_{i=N-m+1}^{N} C_i\,, \tag{3.1}$$

where $m < N/2$ is the number of temperatures in each range. We randomly removed from $\mathcal{L}_m$ and $\mathcal{H}_m$ some number of configurations to keep the cardinality of these sets fixed. In this way, our study, on how the critical temperature predicted by the ANN depends on the range of temperatures used to train it, was not affected by a different number of configurations for different ranges. In order to connect $m$ with a temperature range used in the trainings, we introduce $\tau$ as a measure of the relative temperature range:

$$\tau = \frac{T_{\mathrm{high}} - T_{\mathrm{low}}}{T_{\mathrm{BKT}} - T_{\mathrm{low}}}\,, \tag{3.2}$$

where $T_{\mathrm{low}}$ and $T_{\mathrm{high}}$ are the lowest and the highest temperatures lower than $T_{\mathrm{BKT}}$ which were used to train the ANN.

Configurations from $\mathcal{L}_m$ and $\mathcal{H}_m$ were mixed together and shuffled and then they were used to teach the ANN to identify the low- and high-temperature phases.

The spin configurations generated in MC are stored as numbers $\theta_i$ from 0 to $2\pi$. However, in order to take into account the character of classical two-dimensional spins, the configurations were rewritten as an array composed of cosines and sines: $\cos\theta_1, \ldots, \cos\theta_N, \sin\theta_1, \ldots, \sin\theta_N$, where $N = L \times L$ is the number of lattice sites. This is equivalent to a representation by complex numbers and has the advantage that almost parallel spins are represented by close numbers, which is not the case for the original representation by the angles $\theta_i$.

Since the generation of extensive sets of spin configurations for different system sizes and at different temperatures is a time-consuming task, especially for the PF and q-XY models, we used a technique known from machine learning-based image recognition to increase the number of configurations without additional MC simulations. Namely, we transformed the original configurations according to symmetries of the system. We used periodic boundary conditions, which allowed us to apply translations to create new configurations. Other transformations were reflections and rotations. Figure 3 schematically shows how the ANN is used to classify spin configurations.

The ANN was implemented using the KERAS package with TENSORFLOW as the computational backend. We used a deep feedforward network with four fully connected hidden layers with 512, 192,

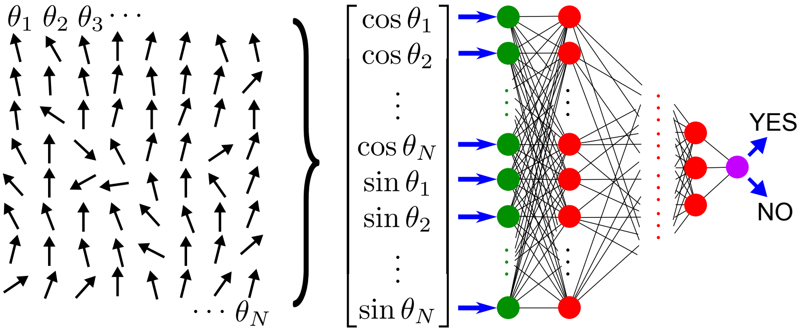

**Figure 3.** (Colour online) A scheme of a binary classification process in a feedforward ANN: a given spin configuration $\{\theta_i\}$, $i = 1, \ldots, N$, is rewritten as a length-$2N$ vector $[\cos\theta_1, \ldots, \cos\theta_N, \sin\theta_1, \ldots, \sin\theta_N]$ and is presented to the input layer of the ANN (green circles). Then, activations flow across fully connected hidden layers (red circles) to the output layer composed of only one neuron (violet circle).

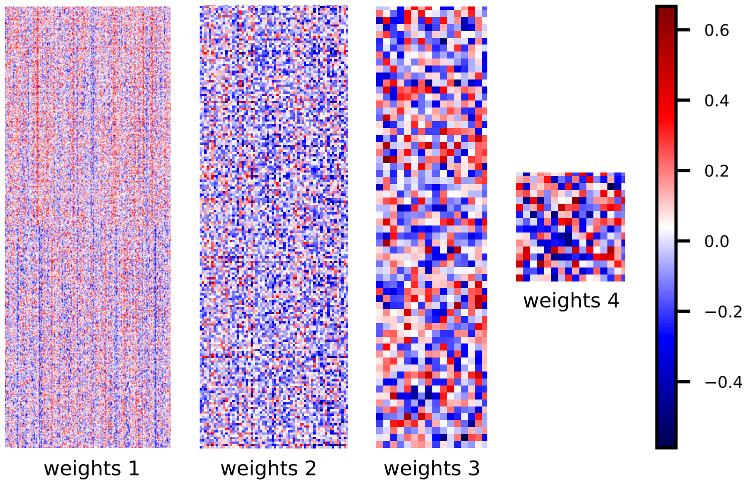

**Figure 4.** (Colour online) Visualization of the activations in a deep ANN composed of hidden layers of 512, 192, 64, 16, and 16 neurons. White colour corresponds to zero weight, blue to negative weight and red to positive weight. "Weights 1" connect the output of layer 1 to layer 2, so they are represented by a rectangle with $512 \times 192$ colour squares. Similarly, "weights 2" are represented by a $192 \times 64$ rectangle, etc. The network was initiated with all weights close to zero (we assumed a finite value $|w_i| < 10^{-3}$ to avoid the *vanishing gradient problem* [30]), so blue and red parts indicate neurons which were activated during training.

64, 16, and 16 neurons. Such a structure was a balance between the number of training epochs required for convergence and the time needed for one epoch. It turns out, however, that the metaparameters are not crucial for the ANN performance. As the activation function, a rectified linear unit (ReLU) was used in the hidden layers and sigmoid function in the output layer. The network was trained to minimize the distance between the MC data and the model predictions defined by the binary cross entropy. The loss function is given by

$$L = -\sum_i \left[y_i \log p_i + (1 - y_i) \log(1 - p_i)\right], \tag{3.3}$$

where $y_i$ are labels and $p_i$ are the corresponding predictions.

We found that with the Tikhonov regularization (L2) [31], the network performs better in the classification of the low- and high-temperature phases. Figure 4 shows an example of learned weights used to classify phases of a $16 \times 16$ system. One can see that despite a rather large size of the network, most of the neurons are activated.

## 4. Results

Figure 5 shows how the predictions for $T_{\rm BKT}$ of the c-XY model calculated by the ANN depend on the way the network was trained. In each case, we trained the neural network using a 10-fold cross validation technique [32] and repeated this procedure 10 times. As a result, we obtained 100 possible values of probability $P$ that a given configuration belongs to the high-temperature phase (the probability that it belongs to the low-temperature phase is $1 - P$). Each time, the ANN was initialized with random weights and biases, and a larger spread of the predictions indicates a greater difficulty in an unambiguous classification of the phase. The same method of multiple trainings starting from different random weights and biases was used in [13] to determine the standard error of the predicted critical temperature. One can see in figure 5 that even if the ANN was trained only at the extreme temperatures ($T = 0.1$ and $T = 1.6$, $m = 1$) corresponding to a fully ordered and completely random configurations, the average predicted $T_{\rm BKT}$ is not far from the actual value. This could be an accidental coincidence because with an increasing $m$ the deviation slightly increases, but always remains below 10%, which can be seen in figure 5 (d). For $m = 16$, the deviation is smaller than the line width. The problem, however, is that in

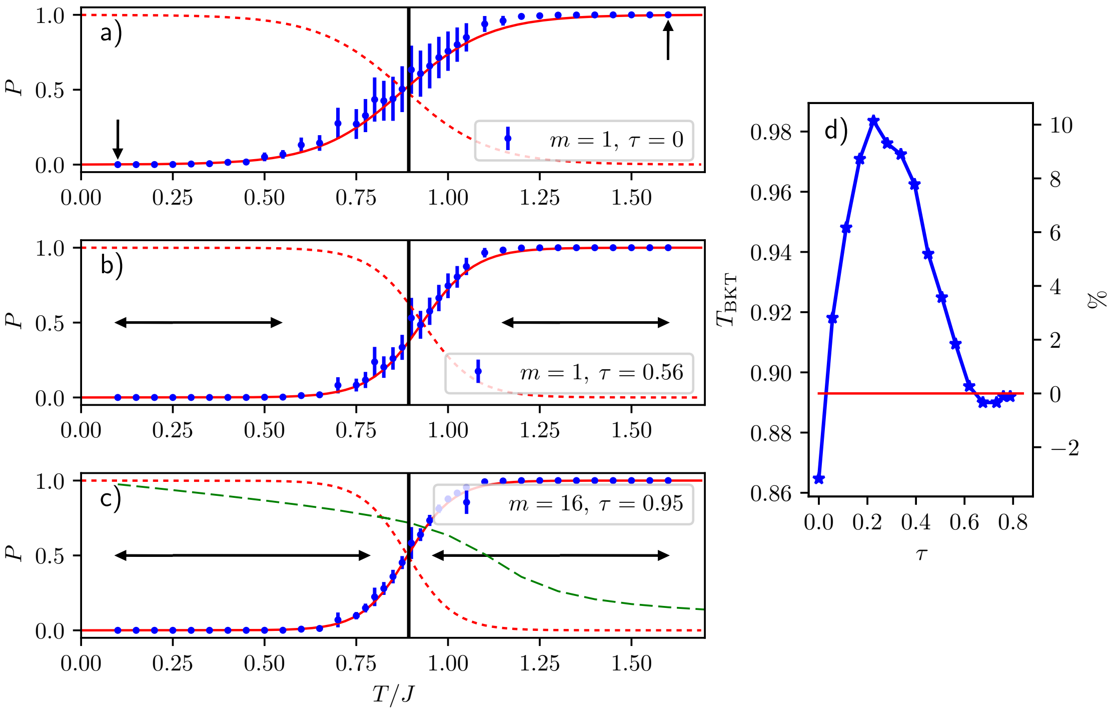

**Figure 5.** (Colour online) Calculated by the ANN, the probability that a given configuration belongs to the high-temperature phase of the c-XY model. The network was trained on labelled data generated according to equation (3.1) for $m = 1$ (a), $m = 10$ (b), and $m = 16$ (c). Then, the network determined the probabilities 100 times for configurations generated at different temperatures each time starting from different weights and biases. The vertical error bars show the standard deviation. The solid red line is the best fit $P(T) = 0.5 \tanh\left[\alpha\left(T - T_{\mathrm{BKT}}\right)\right] + 0.5$, where $\alpha$ and $T_{\mathrm{BKT}}$ are fitting parameters. The dashed red line shows $1 - P(T)$ which is the probability of being classified as a low-temperature phase. The black arrows indicate the temperatures (a) or the ranges of temperatures (b), (c) used to train the network. The black vertical line indicates $T_{\mathrm{BKT}}$ determined from the RG equations. The dashed green line in panel (c) shows magnetization. Comparing panels (a)–(c) one can see that the for the c-XY model, the average critical temperature does not change significantly with an increasing range of training temperatures, but the spread of the results shrinks. The temperatures at which the ANN was trained and tested were (in units of $J$): from 0.10 to 0.70 and from 1.10 to 1.60 with stepsize 0.05 and from 0.750 to 1.050 with stepsize 0.025. Panel (d) shows estimated $T_{\mathrm{BKT}}$ as a function of $\tau$. The horizontal red line shows the critical temperature determined from fitting the MC results to the RG equation (2.4). The right-hand vertical axis shows $(T_{\mathrm{BKT}} - T^{0}_{\mathrm{BKT}})/T^{0}_{\mathrm{BKT}} \times 100\%$, where $T_{\mathrm{BKT}}$ is the average ANN prediction and $T^{0}_{\mathrm{BKT}}$ is the actual critical temperature.

this case, the uncertainty is large. So, for a precise determination of $T_{\mathrm{BKT}}$, the averaging over a large number of configurations generated at different temperatures is necessary. For example, for $m = 1$, the average prediction for 1000 statistically independent configurations is less than 1% off the actual $T_{\mathrm{BKT}}$, but individual predictions are spread within the range of ±18% around this value. This means that if one saves the computational time required to generate configurations for training, more configurations should be generated for the evaluation stage. With an increasing width of the temperature range used for training, the spread of the calculated probabilities decreases significantly. Figure 5 (d) shows how the average $T_{\mathrm{BKT}}$ depends on the number of different temperatures used in training. On the right-hand vertical axis, showing the relative error, one can see that the error is always below 10%, and with an increasing number $m$, it converges to the actual value of $T_{\mathrm{BKT}}$ for the c-XY model.

In the case of the PF model, the results are different. As can be seen in figure 6, the spread of the calculated probabilities is less dependent on $m$, but the average critical temperature strongly depends on $m$. This means that for the PF model, an increase of the number of configurations used at the stage of phase classification will not guarantee a more precise estimation of $T_{\mathrm{BKT}}$. Instead, for this model, a sufficiently

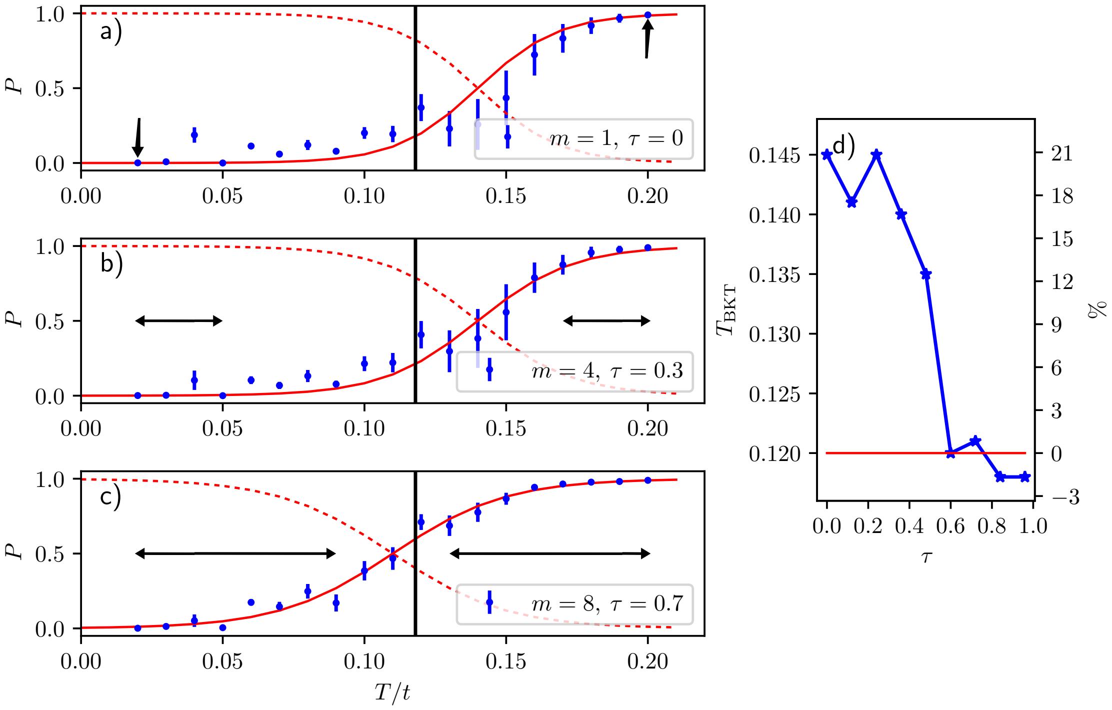

**Figure 6.** (Colour online) The same as in figure 5, but for the PF model. In this case, the critical temperature is affected by the width of the range training temperatures and a relatively wide range is necessary to obtain precise $T_{\mathrm{BKT}}$ [c.f. figure 6 (b)]. The temperatures at which the ANN was trained and tested were (in units of $t$): from 0.02 to 0.20 with stepsize 0.01.

wide range of temperatures at which configurations for training are generated is necessary. From the physical point of view, this means that in the PF model, extremely low-temperature configurations and extremely high-temperature configurations are more different from the configurations close to the critical point than in the case of the c-XY model. It can be seen in figure 6 (d) that for the PF model, the relative error for $m = 1$ extends to more than 20%.

The distribution of probabilities calculated for the q-XY model is similar to that for its classical counterpart. It is presented in figure 7. Though for $m = 1$, the estimated critical temperature differs from its real value by 14% [see figure 7 (d)], the difference decreases very quickly with an increasing $m$ and already for $m \geqslant 3$, the relative error is around 2%.

The results show that in the case of the PF model, a much richer set of configurations is required to properly train the ANN than for the XY models. The reason can be connected to a different character of this model. In both the classical and quantum XY models, the interaction range is limited to nearest neighbors. On the other hand, fermions in the PF model mediate the effective interactions between arbitrarily spaced lattice sites. This effect is seen in figure 1: at low temperature the helicity modulus in the c-XY and q-XY models converges even for very small systems. This is not the case for the PF model, where even at very low temperatures (i.e., in an almost fully ordered state) the energy per lattice site depends on the system size. This results from the delocalization of fermions which in the ordered state are similar to quantum particles in an infinite quantum well, with their energies strongly dependent on the size of the well.

## 5. Summary

We have demonstrated how the accuracy of finding the BKT transition in three different models depends on the range of temperatures at which the ANN was trained. We used a simple feedforward network with densely connected hidden layers. We did not perform any feature engineering of the spin

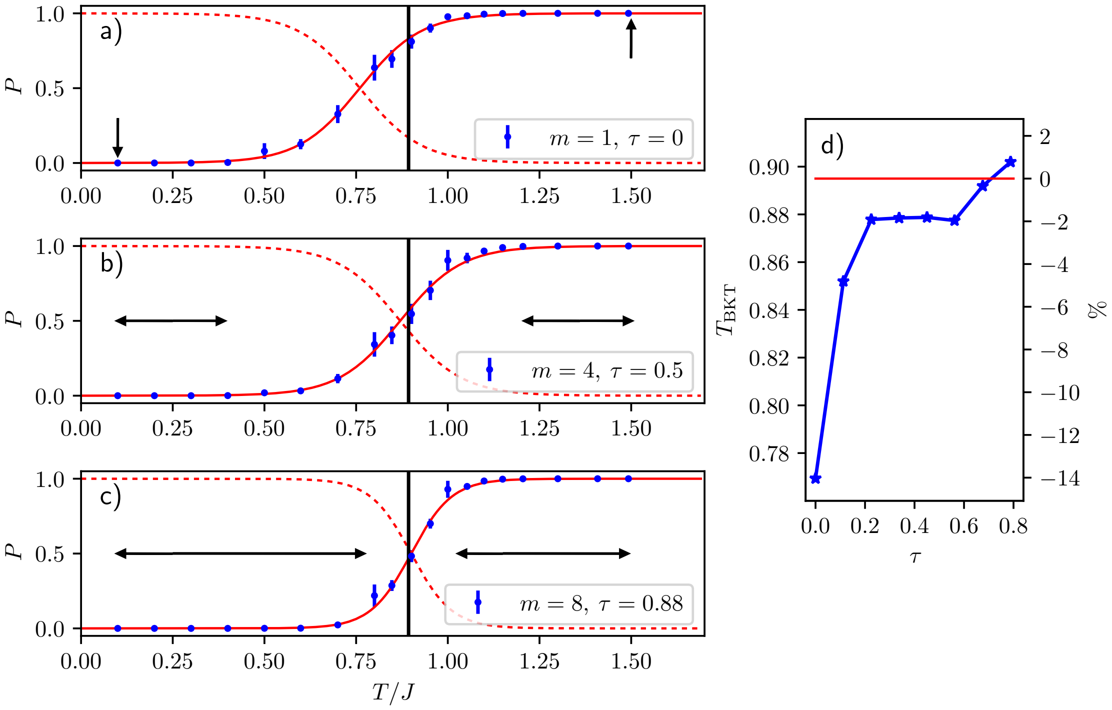

**Figure 7.** (Colour online) The same as in figure 5, but for the q-XY model. Similarly to the c-XY model, also here the critical temperature is rather insensitive to the width of the training temperatures, but the spread of the results decreases with an increasing $m$. The temperatures at which the ANN was trained and tested were (in units of $J$): from 0.1 to 0.8 and from 1.2 to 1.5 with stepsize 0.1 and from 0.85 to 1.15 with stepsize 0.05.

configurations generated in MC simulations and we did not use convolutional layers. Therefore, the phase classification was based on raw spin configurations, not on the explicitly extracted vortices as in [17]. Nevertheless, in figure 5 (c) we compare the calculated probabilities and magnetization that results from the finite size of the system. One can see there that the section of $P(T)$ that indicates the BKT transition is much steeper than the temperature dependence of the magnetization, even if the network was trained on the extreme temperatures [figure 5 (a)]. Therefore, we believe that the ANN learns not only the magnetization (which would vanish in the thermodynamic limit), but also some topological features connected with the BKT transition. One also cannot exclude that the ANN is capable of learning the character of the spin-spin correlations which change their behavior at the BKT transition. To confirm this, however, at least a finite size analysis of the ML results would be necessary, which has not been performed here. However, our aim was different — we wanted to demonstrate how the critical temperature determined by the ANN depends on the composition of the training set. As one can expect, the larger is the variety of the configurations representing the low-temperature and high-temperature phases, the better is the accuracy of the critical temperature. We found, however, that for the c-XY and q-XY model, the average $T_{\mathrm{BKT}}$ was close to the actual one even if the ANN was trained relatively far from the critical point. Increasing the range of temperatures at which the network was trained, only slightly improves the numerical accuracy (i.e., the difference between the average $T_{\mathrm{BKT}}$ determined by the ANN and the value found from the RG equations), but significantly reduces the uncertainty. The situation is different for the PF model, where the numerical accuracy is strongly dependent on the temperature range used at the training stage. We attribute this behavior to the long-range effective interaction present in the PF model which can lead to a longer range of the spin-spin correlations and their different temperature dependence.

Despite the difference in the results for the XY models and the PF model, in all cases it is important to train the ANN not only at very low and at very high temperatures, but also as close as possible to the critical temperature. The main problem in training at temperatures close to $T_{\mathrm{BKT}}$ is that for supervised learning, the configurations must be labelled, so one should know at least an approximate value of the critical temperature. One of the ways to overcome this difficulty is to use the *learning by confusion* approach [14] based on a combination of supervised and unsupervised techniques.

## Acknowledgements

M.M.M. acknowledges support by NCN (Poland) under grant 2016/23/B/ST3/00647. H.K. and N.T. acknowledge funding from grant no. NSF DMR 1629382.

## References

1. Berezinskii V.L., Sov. Phys. JETP, 1971, **32**, 493.
2. Berezinskii V.L., Sov. Phys. JETP, 1972, **34**, 610.
3. Kosterlitz J.M., Thouless D.J., J. Phys. C: Solid State Phys., 1972, **5**, L124, doi:10.1088/0022-3719/5/11/002.
4. Kosterlitz J.M., Thouless D.J., J. Phys. C: Solid State Phys., 1973, **6**, 1181, doi:10.1088/0022-3719/6/7/010.
5. Carrasquilla J., Melko R.G., Nat. Phys., 2017, **13**, 431, doi:10.1038/nphys4035.
6. Morningstar A., Melko R., J. Mach. Learn. Res., 2018, **18**, 163.
7. Ponte P., Melko R., Phys. Rev. B, 2017, **96**, 205146, doi:10.1103/PhysRevB.96.205146.
8. Zhang Y., Kim E.-A., Phys. Rev. Lett., 2017, **118**, 216401, doi:10.1103/PhysRevLett.118.216401.
9. Wang L., Phys. Rev. B, 2016, **94**, 195105, doi:10.1103/PhysRevB.94.195105.
10. Hu W., Singh R.R.P., Scalettar R.T., Phys. Rev. E, 2017, **95**, 062122, doi:10.1103/PhysRevE.95.062122.
11. Wetzel S.J., Phys. Rev. E, 2017, **96**, 022140, doi:10.1103/PhysRevE.96.022140.
12. Broecker P., Carrasquilla J., Melko R.G., Trebst S., Sci. Rep., 2017, **7**, 8823, doi:10.1038/s41598-017-09098-0.
13. Ch'ng K., Carrasquilla J., Melko R.G., Khatami E., Phys. Rev. X, 2017, **7**, 031038, doi:10.1103/PhysRevX.7.031038.
14. Van Nieuwenburg E.P.L., Liu Y.-H., Huber S.D., Nat. Phys., 2017, **13**, 435, doi:10.1038/nphys4037.
15. Torlai G., Mazzola G., Carrasquilla J., Troyer M., Melko R., Carleo G., Nat. Phys., 2018, **14**, 447, doi:10.1038/s41567-018-0048-5.
16. Carleo G., Troyer M., Science, 2017, **355**, 602, doi:10.1126/science.aag2302.
17. Beach M.J.S., Golubeva A., Melko R.G., Phys. Rev. B, 2018, **97**, 045207, doi:10.1103/PhysRevB.97.045207.
18. Deng D.-L., Li X., Sarma S.D., Phys. Rev. B, 2017, **96**, 195145, doi:10.1103/PhysRevB.96.195145.
19. Zhang W., Liu J., Wei T.-C., Preprint arXiv:1804.02709, 2018.
20. Rodriguez-Nieva J.F., Scheurer M.S., Preprint arXiv:1805.05961, 2018.
21. Maśka M.M., Trivedi N., Preprint arXiv:1706.04197, 2017.
22. Hsieh Y.-D., Kao Y.-J., Sandvik A.W., J. Stat. Mech.: Theory Exp., 2013, **2013**, P09001, doi:10.1088/1742-5468/2013/09/P09001.
23. Micnas R., Ranninger J., Robaszkiewicz S., Rev. Mod. Phys., 1990, **62**, 113, doi:10.1103/RevModPhys.62.113.
24. Maśka M.M., Czajka K., Phys. Rev. B, 2006, **74**, 035109, doi:10.1103/PhysRevB.74.035109.
25. Fisher M.E., Barber M.N., Jasnow D., Phys. Rev. A, 1973, **8**, 1111, doi:10.1103/PhysRevA.8.1111.
26. Kosterlitz J.M., J. Phys. C: Solid State Phys., 1974, **7**, 1046, doi:10.1088/0022-3719/7/6/005.
27. Schultka N., Manousakis E., Phys. Rev. B, 1994, **49**, 12071, doi:10.1103/PhysRevB.49.12071.
28. Tomita Y., Okabe Y., Phys. Rev. B, 2002, **65**, 184405, doi:10.1103/PhysRevB.65.184405.
29. Weber H., Minnhagen P., Phys. Rev. B, 1988, **37**, 5986(R), doi:10.1103/PhysRevB.37.5986.
30. Sussillo D., Abbott L.F., Preprint arXiv:1412.6558v3, 2015.
31. Ng A.Y., In: Proceedings of the Twenty-First International Conference on Machine Learning (Banff, Canada, 2004), ACM, New York, 2004, 78, doi:10.1145/1015330.1015435.
32. Gunasegaran T., Cheah Y.-N., In: Proceedings of the 8th International Conference on Information Technology (Amman, Jordan, 2017), IEEE, 2017, 89–95, doi:10.1109/ICITECH.2017.8079960.

# Застосування машинного навчання до переходу Березинського-Костерліца-Таулесса в класичних і квантових моделях

М. Ріхтер-Лясковська[1], Г. Хан[2], Н. Тріведі[2], М.М. Маська[1]

[1] Інститут фізики, Сілезький університет, вул. 75-го піхотного полку, 1, 41-500 Хожів, Польща

[2] Факультет фізики, університет штату Огайо, просп. В. Вудраффа, 191, Колумбус, Огайо 43210, США

Перехід Березинського-Костерліца-Таулесса є дуже специфічним фазовим переходом, при якому всі термодинамічні величини є неперервними. Тому важко точно визначити критичну температуру. У цій статті нами показано, як можна використати нейронні мережі для розв’язання цього завдання. Зокрема, досліджено, до якої міри точність розпізнавання переходу залежить від способу навчання нейронних мереж. Ми застосовуємо наш підхід до трьох різних систем: (i) класична XY модель, (ii) фазово-ферміонна модель із взаємодією між класичними й квантовими ступенями вільності та (iii) квантова XY модель.

**Ключові слова:** *фазові переходи, топологічні дефекти, XY модель, штучні нейронні мережі, машинне навчання*