National Research University Higher School of Economics

as a manuscript

Chertenkov Vladislav Igorevich

Universality classes and machine learning

Dissertation summary

for the purpose of obtaining academic degree

Doctor of Philosophy in Applied Mathematics HSE

Academic supervisor:

Doctor of Sciences in Physics and Mathematics

professor Shchur L.N.

Moscow – 2024

Relevance

The application of machine learning (ML) techniques to the study of phase transitions has become a promising tool. Machine learning algorithms can solve problems with large amounts of data. Neural networks (NN) are widely used in natural language processing (NLP), computer vision (CV) tasks for object recognition, time series analysis and engineering applications. To solve problems in various domains, algorithms do not require an a priori understanding of the nature of the data, only a sufficient number of training examples. For instance, in order to understand natural language, a NN do not need to know the principles of morphology and semantics of the language.

An analysis of the number of scientific publications from 2015 to 2023, according to Google Scholar, shows the growing interest in ML tools in condensed matter physics. The application of machine learning to the study of phase transitions will make it possible to combine classical methods of statistical physics and advanced approaches to modeling complex processes. The interdisciplinary approach will enrich the current theoretical framework with new methods for a deeper understanding of the complex dynamics of phase transitions and the interpretation of physical systems.

In addition to the theoretical significance, the practical value consists in the development of new tools, software packages and methods for dealing with phase transitions in real systems. The widespread adoption of ML methods create an opportunities for new applications in materials science, particle physics, and other related fields.

Degree of problem development

Paper [1] proposes an approach to analyze phase transitions by the supervised ML method. The critical temperature $T_c^* = 2.266(2)$ and critical exponent $\nu = 1.0(2)$ are numerically extracted through the collapse of the NN output data, by solving the phase classification problem for the Ising model on a square lattice. Transfer learning approach on the square lattice allows to extract $T_c^* = 3.65(1)$ and $\nu = 1.0(3)$ for Ising model on the triangular lattice [2]. The phase transition is transferred from the Ising model to the q-state Potts ($q \in [2; 10]$) model with the accuracy of T_c^* estimation to 3 decimal places, by solving the regression problem for predicting the temperature of spin configurations [3].

Generative-adversarial networks (GAN) are used [4] to generate new examples on the lattice while preserving the distribution of thermodynamic quantities statistics. The critical temperature $T_c^* = 2.266(4)$ in the Ising model is extracted [5] using unsupervised dimensionality reduction ML methods. ML methods are used [6–9] to study the phase transitions in percolation problem and the BKT transition in the XY model and q-state clock model.

In order to compare the accuracy of methods for extracting W values, let us introduce the relative error ϵ, % of the method (equation 1). For example, if the numerical estimate of the critical temperature $\hat{T}_c = 2.258(5)$ for the exact solution $T_c = 2.269$, then relative error is $\epsilon = 100 \cdot \max(|2.269 - 2.258|, 0.005)/2.269 = 100 \cdot \max(0.011, 0.005)/2.269 = 0.5\%$.

$$\epsilon = 100 \cdot \max(|W - \hat{W}|, W_e)/W. \tag{1}$$

Numerical extraction of the critical temperature T_c^* using ML methods has been repeatedly reported in various papers. The relative error of T_c^* extraction (see eq. 1) is less than 1% in the majority of published papers. The correlation length critical exponent ν is extracted less consistently and with larger relative error. The numerical estimates of ν in the original paper [1] have a relative error of 20% for the square lattice and 30% for the triangular lattice of the Ising model. Various works use the data collapse method, for which the numerical solution of ν lies in a wide range. In the paper [9] a different method to extract the critical exponents is applied, but there is no independent confirmation of the obtained results.

A number of questions arise: with what accuracy NN can extract critical behaviour of spin lattice models; what factors affect the accuracy of critical properties estimation; and whether it is possible to create a method that works for models in different universality classes.

Aims of the study

Development of a method for analyzing phase transitions in lattice spin models using supervised machine learning methods.

Tasks of the study

- Development of a method for analyzing phase transitions in lattice spin models using neural network supervised machine learning methods.
- Development of a software for the study of phase transitions in lattice spin models using classical methods and neural networks.
- Application of the method to study phase transitions in lattice spin models which belongs to the Ising model and 4-state Potts model universality classes.
- Measurement of the developed method accuracy and investigation of the factors influencing the accuracy.
- Development of a transfer learning method for analyzing phase transitions in lattice spin models.

The scientific novelty of the study

1. A method for analyzing phase transitions in lattice spin models is proposed, which is based on the variance output scaling of the neural network trained to solve supervised binary classification problem. From the output data of the neural network, an estimate of the critical temperature and the critical correlation length exponent are systematically extracted with high accuracy.
2. The proposed method was used to extract the critical correlation length exponent and critical temperature for the Ising, Baxter-Wu, and 4-state Potts models. For

the latter two models, the critical exponents are extracted for the first time using ML methods.

3. The influence of neural network architectures, hyperparameters of training and methods of input data encoding on the accuracy of extracted critical exponents of lattice spin models were investigated for the first time.

4. A new method of input data encoding is proposed for transfer learning of the phase transitions between the Ising and the 4-state Potts model universality classes.

Methodology and methods

Markov chain Monte Carlo methods and finite-size scaling analysis have been applied to investigate lattice spin models by conventional approaches. The developed modeling method applies deep machine learning, computer vision, neural networks, supervised learning, and optimization methods. The general research methods are statistical analysis and numerical approximation methods.

Summary

The **first chapter** describes two-dimensional lattice spin models: the Ising [10] model on a square lattice, the Baxter-Wu model on a triangular lattice, and the 4-state Potts model on a square lattice. A second order phase transitions is between the ordered ferromagnetic and disordered paramagnetic phases at the critical point T_c. There are exact analytical solutions: the Ising model is solved by L. Onsager [11] in 1944, $T_c \approx 2.269$, the Baxter-Wu model is solved by R.Baxter and F.Wu [12] in 1973, $T_c \approx 2.269$, 4-state Potts model is solved by R.Potts [13] in 1952, $T_c \approx 0.91$.

The models have different Hamiltonian, lattice topology, and ground state symmetry. Models can be categorized into two classes of universality: the Ising model, which belongs to the universality class named after itself, the Baxter-Wu and 4-state Potts models belonging to the 4-state Potts model universality class. The universality class is defined by a set of critical exponents of heat capacity α, magnetization β, susceptibility γ, correlation length ν, and others.

A number of relations [14–16] between critical exponents follow from the scaling and universality hypotheses [17, 18]

$$2\beta + \gamma = 2 - \alpha = d\nu, \tag{2}$$

where d is space dimensionality. The Ising model universality class: $\alpha = 0, \beta = 1/8, \gamma = 7/4, \nu = 1$, the 4-state Potts model universality class: $\alpha = 2/3, \beta = 1/12, \gamma = 7/6, \nu = 2/3$. Critical exponent ν reflects the dependence of the correlation length on the linear lattice size L. The critical exponent ν and the critical temperature T_c will be further determined by analyzing NN output.

The conventional method of phase transitions investigation is Markov chain Monte-Carlo (MC) methods. Images of instant spin configurations ("snapshots") are simulated by the MC method according to the recommendations [19]. The MC simulation parameters are selected in order to generate the uncorrelated snapshots of the spin configurations: i) relaxation time at "hot start" to minimize systematic error, ii) correlation time τ_{corr} at thermodynamic equilibrium to minimize the statistical error. Consider two classes of algorithms: Metropolis [20, 21] with single spin-flip updates and Swendsen-Wang [22], Wollf [23] with cluster updates. In the Metropolis algorithm, a decision is made at each step to choose a new spin orientation, and for cluster algorithms – a new orientation for the group (cluster) of spins. The Swendsen-Wang and Wollff algorithms have different clustering algorithms and efficiency. It is known that all algorithms suffer from critical slowing down in the critical region, in which the correlation time grows rapidly $\tau_{corr} \sim \min(L, \xi)^{d+z}$, where ξ is a correlation length, z is a dynamic critical exponent. Wollf's algorithm is the most efficient, but its implementation in the Baxter-Wu model leads to a shift of the cluster percolation from the critical point to the low-temperature region and thereby distort the critical behavior [24]. The Metropolis algorithm, which is the least efficient, was chosen to retain a unified approach to data generation.

The **second chapter** focuses on machine learning methods. A general formulation of the binary classification problem in terms of ML – for each object $x_i \in X$ predict a discrete value y_i called a class label: $F(x_i) \to y_i$, where $y_i \in \{0, 1\}$, F is a decision function. The quality of the trained classifier can be measured by the set of metrics: precision, recall, f-score and area under the ROC curve (AUCROC).

The decision function F is a neural network. The NN consists of consecutive blocks called layers. The number and order of layers in the NN determines its architecture. Each layer is a differentiable function. A layer has an input, an output and can consist of linear and nonlinear operators. Common types of NN building blocks are fully connected, convolutional, pooling layers and activation functions. Three NN architectures are proposed: fully connected NN (FCNN), convolutional network (CNN) and deep convolutional network ResNet [25].

Not only the basic building blocks of NN are described, but also the mechanisms to adjust network parameters, the stages of training, validation and testing, as well as heuristics affecting the quality of training. The NN is trained using optimization methods based on gradient descent [26] and the backpropagation of errors [27, 28] to update NN adjustable weights. The NN training process includes the stages of splitting the data, feature encoding, training protocol analysis, and metrics evaluation. To improve quality metrics and speed up the learning process, heuristic approaches are used: random weights initialization, batch learning, regularization of the input data (augmentation).

The **third chapter** describes the developed method for analyzing phase transitions in lattice spin models. We solve a supervised ML binary classification problem. To investigate the critical behavior of lattice spin models, spin configurations are generated for different linear lattice sizes using the Monte Carlo method by the Metropolis algorithm, with 1500 images per temperature point. The output layer of NN is represented by two neurons for the ferromagnetic and paramagnetic phases. Separate NN is trained for each lattice size L of spin model. Loss function is a binary cross-entropy (BCE), and the quality of the training process is evaluated at multiple epochs using plots of the training and validation curves. To analyze critical behaviour of the lattice spin model, test set of data (dataset) is used on the inference stage of the NN weights state after the first epoch of training.

The NN output variance function $V(T)$ of the ferromagnetic neuron f_i^T is measured for each linear size L, the function is defined at each temperature point T

$$V(T) = \frac{1}{N}\sum_{i=1}^{N}\left(f_i^T\right)^2 - \left(\frac{1}{N}\sum_{i=1}^{N} f_i^T\right)^2, \tag{3}$$

where $i \in \{1, 2, ..N\}$ are uncorrelated snapshots of spin configurations at T. We study the dependence of lattice size L of the variance $V(T)$.

We use fitting procedure of the function $V(T)$ for each value of lattice size L. The approximation of $V(T)$ vs T is performed with an unnormalized Gaussian-like ansatz $V(T) \sim \exp\left(-(T-T_*)^2/2\sigma^2\right)$ to extract parameters σ and T_*. Width σ of the variance curve $V(T)$ is then fitted with power-law ansatz $\sigma(L) \sim 1/L^{1/\nu}$ to extract the critical exponent ν. Peak position T_* of the variance curve $V(T)$ is fitted with Ferdinand-Fisher ansatz [29] $T_*(L) - T_c \sim 1/L^{1/\nu}$ with fixed value of ν to extract critical temperature $\hat{T}_c$ of the infinite volume system $L \to \infty$. For the Ising, Baxter-Wu, and 4-state Potts models, numerical extraction is performed to estimate critical temperature $\hat{T}_c$ and critical exponent ν.

The remainder of the chapter is devoted to three groups of experiments. In the first one, the influence of the size of training dataset on the accuracy of the extracted values $\hat{T}_c, \nu$ is conducted by reducing the training sample size by a factor of two and a factor of four. In the second, the effect of the number of NN training epochs is investigated. As long as training continues, NN is better at solving the classification problem, which is reflected in the accuracy of the extracted critical exponents. In the third, a numerical extraction of the critical temperature estimation $\hat{T}_c$ and critical exponent ν is performed using transfer learning between lattice spin models.

In **chapter four**, the influence of input data encoding methods is investigated on the accuracy of extracted critical behaviour. In contrast to the previous chapter, where the NN is trained in the space of spin configurations, two other approaches of encoding are applied: into the space of spin correlators and into the space of bond energies.

To encode spin configuration into the correlator space, a matrix of size $L \times L$ is created, each cell contains spin correlator. A priori knowledge of the lattice topology is used to fill the matrix. Each spin in a lattice vertex interacts with other spins at distance $L/2$ according to equation

$$g_{x,y}(L/2) = \frac{1}{D}\sum_{d=1}^{D} s_{x,y}s_*, \tag{4}$$

where $g_{x,y}(L/2)$ is a value of correlator in the vertex (x, y), $s_{x,y}$ is a spin value at the same vertex, s_* is a spin value located at a distance $L/2$ from the spin $s_{x,y}$ along axis d, D is the total number of the directions of spin interactions.

To encode spin configuration into the bond energies space, two matrices of size $L \times L$ is created to store vertical $\{\sigma_V\}$ and horizontal $\{\sigma_H\}$ bonds. A priori knowledge of the interactions between spins is used to fill the matrices:

$$\begin{aligned}
\{\sigma_H\} &= \{\sigma_{i,j}\sigma_{i,j+1}, \sigma_{i,j+1}\sigma_{i,j+2}, ..., \sigma_{i+L-1,j+L-1}\sigma_{i+L-1,j}\}, \\
\{\sigma_V\} &= \{\sigma_{i,j}\sigma_{i+1,j}, \sigma_{i+1,j}\sigma_{i+2,j}, ..., \sigma_{i+L-1,j+L-1}\sigma_{i,j+L-1}\}.
\end{aligned} \tag{5}$$

Numerical extraction from output variance $V(T)$ scaling is performed to estimate critical temperature $\hat{T}_c$ and critical exponent ν in the spaces of correlators and bond energies. Transfer learning is conducted between every combination of lattice spin models in the space of bond energies.

Thesis statements submitted for defense

- A method for analyzing phase transitions in lattice spin models is introduced, which is based on the variance output scaling of the neural network trained to solve supervised binary classification problem. The following aspects of the method are described: the neural network training process, the parameters influencing the accuracy of extraction of critical exponent ν and estimation of critical temperature $\hat{T}_c$, recommended values of training hyperparameters.

- It was established that the variance function of the output of the NN trained using the method developed in the thesis carries information about the critical correlation length exponent ν and the critical temperature $\hat{T}_c$ of lattice spin model. Relative error of $\hat{T}_c$ extraction is 0.1-0.2%, ν is 1-3%.

- The proposed method was used to extract the correlation length critical exponent ν and the estimation of critical temperature $\hat{T}_c$ for the Ising $1/\nu = 1.02(1)$, $\hat{T}_c = 2.270(5)$; Baxter-Wu $1/\nu = 1.49(2)$, $\hat{T}_c = 2.2691(1)$; and 4-state Potts $1/\nu = 1.49(4)$, $\hat{T}_c = 0.9101(1)$ models.

- The method for analyzing the critical properties of lattice spin models is verified using several neural network architectures. The method works for a fully connected network with a single hidden layer, a shallow convolutional network, and a deep convolutional architecture ResNet. The accuracy of each architecture depends on the properties of the lattice spin model and hyperparameters of training.

- The method for analyzing the critical properties of lattice spin models is sensitive to hyperparameters of training. Growth in the number of iterations of NN training leads to an improvement in the quality of classification, but worsens the accuracy of critical exponent ν extraction. Increasing the size of the training set of data does not affect the accuracy of the extracted critical properties when a certain level of classification quality metrics is reached. Input data encoding method affects the accuracy of the extracted critical properties ν and $\hat{T}_c$.

- A new method of input data encoding based on the lattice spin model bond energies is proposed for transfer learning of the phase transitions. Extraction of ν and $\hat{T}_c$ was performed for Ising, Baxter-Wu and 4-state Potts models by learning with knowledge transfer within and outside of the native universality class. The results agree with the exact solution, however the method is unsystematic and sensitive to the parameters of lattice spin models, NN architectures and hyperparameters of training.

Personal contribution of the author

The hypotheses, ideas and method were developed jointly with the academic supervisor. The stages of data generation by Monte Carlo method, network training and testing, analysis of neural network modeling results, validation and comparison with classical methods were made personally by the author. In the publications the author's contribution is determinant. Sectional reports at conferences were made personally by the author.

General findings of the study

- A method for analyzing phase transitions in lattice spin models is developed, which is based on the variance output scaling $V(T)$ of the neural network. NN is trained with supervised learning to solve binary classification problem. The proposed method was used to extract the correlation length critical exponent ν and the estimation of critical temperature $\hat{T}_c$ for the Ising, Baxter-Wu and 4-state Potts models. Relative error of $\hat{T}_c$ extraction is within 0.2%, ν is within 3%, which outperforms other ML approaches in lattice spin models. The values of ν and $\hat{T}_c$ within statistical error coincide with the exact analytical solutions for these models in the Ising and 4-state Potts universality classes.

- The analysis of NN **architectures**, from shallow to deep ones, has shown that the method works systematically. However, it is impossible to choose the best one among the architectures, due to their differences in accuracy which depends on the properties of the studied lattice spin models, approaches of input data encoding, and the method of critical behaviour analysis (with or without transfer learning). When trained in the space of spin configurations, shallow architectures FCNN and CNN provide more accurate results of $\hat{T}_c, \nu$ extraction, which is consistent with the observations of [30]. The FCNN architecture faces underfitting, when the input data is encoded into the space of bond energies, while the accuracy of CNN and ResNet does not differ from each other.

- Influence of the training **number of epochs** shows that growth in the number of iterations is leading to an improvement in the quality of classification, but worsens the accuracy of critical exponent extraction. At later epochs NN turns into a perfect classifier without the ability to extract ν. The **size** of the training set of data in NN can affect the accuracy of the extracted critical properties $\hat{T}_c, \nu$ until a certain level of classification quality metrics is reached. The Ising model FCNN when trained on a quarter of the original training dataset size, faces underfitting on the AUCROC metric, resulting in a worsening of ν extraction.

- Extraction of $\hat{T}_c$ and ν was performed for Ising, Baxter-Wu and 4-state Potts models by learning with knowledge transfer within and outside of the native universality class. The transfer learning method is less systematic then native inferencing. The critical temperature $\hat{T}_c$ is extracted systematically for all combinations of lattice spin models, except for the knowledge transfer from the square lattice to the triangular lattice in the space of spin configurations. Relative error of $\hat{T}_c$ extraction is 0.1-1.5% in the space of spin configurations, and 0.1-5%

in the space of bond energies. The largest relative error of extraction $\hat{T}_c$ occurs when transferring knowledge from the 4-state Potts model: 0.3-4% within the class of universality, 7-16% outside the universality class. The critical exponent ν is extracted less systematically. In the space of spin configurations, knowledge transfer is possible only from the 4-state Potts universality class to the Ising universality class, and is not possible from a square lattice to a triangular lattice. Encoding into the space of bond energies provide knowledge transfer of the phase transitions from the Ising model to the Baxter-Wu model with the same level of accuracy as from native model extraction. When moving from the spin configurations space to the space of bond energies, NN do not need to "understand" the spatial arrangement of spins on the lattice, a problem encountered by networks when transferring knowledge from a square lattice to a triangular lattice. The encoding of spin configurations into the space of correlators and bond energies involves incorporating additional information into the NN regarding the lattice topology and the Hamiltonian governing the interactions within the model.

Approbation of the results

The list of articles:

- Chertenkov V .I. Universality classes and machine learning / Shchur L.N. // Journal of Physics: Conference Series. 2021. N 1740. P. 1-5. (Scopus Q4)
- Chertenkov V .I. Deep machine learning investigation of phase transitions / Burovski E.A., Shchur L.N. // Lecture Notes in Computer Science. 2022. N 13708. P. 397-408. (Scopus Q2)
- Chertenkov V .I. Finite-size analysis in neural network classification of critical phenomena / Burovski E.A., Shchur L.N. // Physical Review E - Statistical, Nonlinear, and Soft Matter Physics. 2023. T. 108. N 3. P. 1-5. (Scopus Q1)
- Sukhoverhova D.D. Validity and limitations of supervised learning for phase transition research / Chertenkov V .I., Burovski E.A., Shchur L.N. // Lecture Notes in Computer Science. 2023. N 14389. P. 314-329. (Scopus Q2)

The list of conferences:

- IV International Conference "Computer Simulation in Physics and beyond", Russia, Moscow, 12-16 October 2020, "Universality classes and machine learning".
- International Conference «Russian Supercomputing Days», Russia, Moscow, 26-27 September 2022, "Deep machine learning investigation of phase transitions".
- Annual Interuniversity Scientific and Technical Conference of students, postgraduates and Young specialists named after E.V.Armensky, Russia, Moscow, 27 February – 7 March 2023, "Investigation of spin models using machine learning methods".

- National Supercomputing Forum NSCF-2023, Russia, Pereslavl-Zalessky, 28 November – 1 December 2023, “Influence of learning protocols on deep learning studies of phase transitions”.
- Research Seminar Computational Wednesdays, Moscow, HSE, 14 November 2023, “Unsupervised learning of phase transitions via modified anomaly detection with autoencoders”.
- International Conference «Russian Supercomputing Days», Russia, Moscow, 23-24 September 2024, “Supervised and Transfer Learning for Phase Transition Research”.

Certificate of state registration of software:

- “A system for studying phase transitions in lattice spin models”.

References

[1] Juan Carrasquilla and Roger G Melko. “Machine learning phases of matter”. In: *Nature Physics* 13.5 (2017), pp. 431–434.

[2] Gordon Frank Newell. “Crystal statistics of a two-dimensional triangular Ising lattice”. In: *Physical Review* 79.5 (1950), p. 876.

[3] Kimihiko Fukushima and Kazumitsu Sakai. “Can a CNN trained on the Ising model detect the phase transition of the q-state Potts model?” In: *Progress of Theoretical and Experimental Physics* 2021.6 (2021), 061A01.

[4] Nicholas Walker and Ka-Ming Tam. “InfoCGAN classification of 2D square Ising configurations”. In: *Machine Learning: Science and Technology* 2.2 (2020), p. 025001.

[5] Constantia Alexandrou et al. “The critical temperature of the 2D-Ising model through deep learning autoencoders”. In: *The European Physical Journal B* 93 (2020), pp. 1–15.

[6] Wanzhou Zhang, Jiayu Liu, and Tzu-Chieh Wei. “Machine learning of phase transitions in the percolation and X Y models”. In: *Physical Review E* 99.3 (2019), p. 032142.

[7] Yusuke Miyajima et al. “Machine learning detection of Berezinskii-Kosterlitz-Thouless transitions in q-state clock models”. In: *Physical Review B* 104.7 (2021), p. 075114.

[8] Kenta Shiina et al. “Machine-learning studies on spin models”. In: *Scientific reports* 10.1 (2020), p. 2177.

[9] Dimitrios Bachtis, Gert Aarts, and Biagio Lucini. “Mapping distinct phase transitions to a neural network”. In: *Physical Review E* 102.5 (2020), p. 053306.

[10] Ernst Ising. “Beitrag zur theorie des ferro-und paramagnetismus”. PhD thesis. Grefe & Tiedemann Hamburg, Germany, 1924.

[11] Lars Onsager. “Crystal statistics. I. A two-dimensional model with an order-disorder transition”. In: *Physical Review* 65.3-4 (1944), p. 117.

[12] RJ Baxter and FY Wu. “Exact solution of an Ising model with three-spin interactions on a triangular lattice”. In: *Physical Review Letters* 31.21 (1973), p. 1294.

[13] Renfrey Burnard Potts. “Some generalized order-disorder transformations”. In: *Mathematical proceedings of the cambridge philosophical society.* Vol. 48. 1. Cambridge University Press. 1952, pp. 106–109.

[14] Benjamin Widom. “Equation of state in the neighborhood of the critical point”. In: *The Journal of Chemical Physics* 43.11 (1965), pp. 3898–3905.

[15] Michael E Fisher. “The theory of equilibrium critical phenomena”. In: *Reports on progress in physics* 30.2 (1967), p. 615.

[16] Leo P Kadanoff et al. “Static phenomena near critical points: theory and experiment”. In: *Reviews of Modern Physics* 39.2 (1967), p. 395.

[17] Leo P Kadanoff. “Scaling laws for Ising models near T c”. In: *Physics Physique Fizika* 2.6 (1966), p. 263.

[18] АЗ Паташинский and ВЛ Покровский. “О поведении упорядочивающихся систем вблизи точек фазового перехода”. In: *ЖЭТФ* 50.2 (1966), pp. 439–447.

[19] Alan Sokal. “Monte Carlo methods in statistical mechanics: foundations and new algorithms”. In: *Functional integration: Basics and applications.* Springer, 1997, pp. 131–192.

[20] W Keith Hastings. “Monte Carlo sampling methods using Markov chains and their applications”. In: (1970).

[21] Nicholas Metropolis et al. “Equation of state calculations by fast computing machines”. In: *The journal of chemical physics* 21.6 (1953), pp. 1087–1092.

[22] Robert H Swendsen and Jian-Sheng Wang. “Nonuniversal critical dynamics in Monte Carlo simulations”. In: *Physical review letters* 58.2 (1987), p. 86.

[23] Ulli Wolff. “Collective Monte Carlo updating for spin systems”. In: *Physical Review Letters* 62.4 (1989), p. 361.

[24] Lev N Shchur and Wolfhard Janke. “Critical amplitude ratios of the Baxter–Wu model”. In: *Nuclear Physics B* 840.3 (2010), pp. 491–512.

[25] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016, pp. 770–778.

[26] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[27] Paul Werbos. “Beyond regression: New tools for prediction and analysis in the behavioral sciences”. In: *PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA* (1974).

[28] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. “Learning representations by back-propagating errors”. In: *nature* 323.6088 (1986), pp. 533–536.

[29] Arthur E Ferdinand and Michael E Fisher. “Bounded and inhomogeneous Ising models. I. Specific-heat anomaly of a finite lattice”. In: *Physical Review* 185.2 (1969), p. 832.

[30] Alan Morningstar and Roger G Melko. "Deep learning the ising model near criticality". In: *Journal of Machine Learning Research* 18.163 (2018), pp. 1–17.