

# Machine Learning of Phases and Structures for Model Systems in Physics

Djenabou Bayo<sup>1,2</sup>, Burak Çiviticioğlu<sup>2</sup>, Joseph J. Webb<sup>1</sup>, Andreas Honecker<sup>2</sup>, and Rudolf A. Römer<sup>1\*</sup>

<sup>1</sup>*Department of Physics, University of Warwick, Coventry, CV4 7AL, United Kingdom*

<sup>2</sup>*Laboratoire de Physique Théorique et Modélisation, CNRS UMR 8089, CY Cergy Paris Université, 95302 Cergy-Pontoise, France*

The detection of phase transitions is a fundamental challenge in condensed matter physics, traditionally addressed through analytical methods and direct numerical simulations. In recent years, machine learning techniques have emerged as powerful tools to complement these standard approaches, offering valuable insights into phase and structure determination. Additionally, they have been shown to enhance the application of traditional methods. In this work, we review recent advancements in this area, with a focus on our contributions to phase and structure determination using supervised and unsupervised learning methods in several systems: (a) 2D site percolation, (b) the 3D Anderson model of localization, (c) the 2D  $J_1$ - $J_2$  Ising model, and (d) the prediction of large-angle convergent beam electron diffraction patterns.

## 1. Introduction

Identification of critical points separating distinct phases of matter is a central pursuit in condensed matter and statistical physics.<sup>1,2)</sup> This task requires a thorough understanding of the global behavior of the many-body system because phenomena may emerge that are very difficult to derive from microscopic rules.<sup>3)</sup> Traditional analytic methods and numerical simulations have proven effective in understanding these complex systems,<sup>4,5)</sup> but they often come with limitations, particularly in high-dimensional parameter space.<sup>6)</sup>

Machine-learning (ML) methods, particularly supervised<sup>7)</sup> and unsupervised learning techniques,<sup>8)</sup> have in the last years appeared in physics as a novel strategy to bypassing some of these limitations.<sup>9–11)</sup> Convolutional neural networks (CNN), a class of deep, i.e., multi-layered, neural networks (DNNs) in which spatial locality of data values is retained during training, have, when coupled with a form of residual learning,<sup>12)</sup> shown to allow astonishing precision when classifying images, e.g., of animals<sup>13)</sup> and handwritten characters,<sup>14)</sup> or when predicting numerical values, e.g., of market prices.<sup>15)</sup> These *supervised* learning strategies similarly yield promising predictions in identifying critical points or phases in parameter space,<sup>16–22)</sup> providing an alternative and potentially more efficient way of exploring complex systems. By now, the evidence in favour of supervised machine-learning methods' efficacy in identifying different phases of a physical system appears compelling.<sup>16–22)</sup> *Unsupervised* learning and semi-supervised learning approaches have also demonstrated the ability to reconstruct the outlines of a system's phase diagram.<sup>23–28)</sup> The potential to identify structural changes within a system further supports the significance of these techniques in modern scientific exploration.<sup>29)</sup>

Among the various models studied in the context of machine learning and *statistical physics*, the Ising model on the square lattice has served as an important benchmark<sup>16, 23, 25–27, 30–42)</sup> due to the simplicity of its two thermal phases, the low-temperature ferromagnet and the high-temperature paramagnet, and the ready availability of its exact solution<sup>43)</sup> with exactly known critical temperature. We

note that the use of ML to determine phases from just the spin configurations suggests that these themselves should contain sufficient information to identify phases, providing a level of physical insight that was, while not unknown, at least not as clear as it now seems. We also mention related work on multi-layer<sup>44)</sup> and Potts models,<sup>45–47)</sup> where the latter include the Ising model as the  $q = 2$  case.

Percolation can be considered as the  $q \rightarrow 1$  limit of the Potts model<sup>48)</sup> and yields another class of paradigmatic models to which machine-learning techniques have been applied to identify the non-spanning and spanning phases.<sup>49–54)</sup> Previous ML studies have mostly used supervised learning in order to find the two phases via ML classification.<sup>49, 51)</sup> An estimate of the critical exponent of the percolation transition has also been given.<sup>49)</sup> The task of determining the transition threshold,  $p_c$ , was further used to evaluate different ML regression techniques.<sup>54)</sup> For unsupervised and generative learning, less work has been done.<sup>49–51)</sup> While some successes have been reported,<sup>50, 55)</sup> other works show the complexities involved when trying to predict percolation states.<sup>51)</sup>

Disordered electron systems provide *quantum systems* with similarly rich phase diagrams. Examples are given by the Anderson insulator,<sup>56)</sup> diffusive metals,<sup>57)</sup> the quantum Hall<sup>58, 59)</sup> and quantum anomalous Hall insulators,<sup>60, 61)</sup> Weyl semimetals,<sup>62–64)</sup> as well as topological insulators.<sup>65, 66)</sup> In these cases, the thermal states investigated for Ising-type models are replaced by quantum mechanical eigenfunctions, or variations thereof such as the local density of states (LDOS). These have specific features in each phase but, due to the random nature of these systems, precisely determining a phase from an LDOS is difficult.<sup>67, 68)</sup> Recent supervised learning work on the Anderson model of localization, capturing the features of eigenfunctions across the delocalization–localization transition,<sup>69)</sup> as well as further transfer-learning approaches to the disordered Chern insulator–Anderson insulator transition,<sup>70)</sup> have shown to allow a seemingly accurate description of phases and phase boundaries.

The power of *generative* machine learning has not yet been harnessed to the same extent. This is partly because it is still a relatively novel machine learning strategy.<sup>71)</sup> In brief, the difference to the supervised methods lies in the generative meth-

\*r.roemer@warwick.ac.uk

ods being able to seemingly create novel predictions which do not appear in any of the provided data. For example, in computer vision, generative networks construct previously non-existent high-resolution images, conditional on information from other images.<sup>72,73</sup> Here, we will show how to use such generative ML strategies to study the phases for the  $J_1 - J_2$  Ising model, an extension of the Ising model that incorporates competing interactions across the diagonals of the Ising squares and presents a more challenging 3-phase structure. As ML generator, we shall use a so-called variational autoencoder (VAE), a type of neural network that reconstructs a given predicted state after being trained on a selected set of states.<sup>74</sup>

The application of ML to structure determination via electron diffraction has also blossomed in the last decade.<sup>75</sup> ML strategies have been used to reduce the data flow in single-molecule data classification,<sup>76</sup> CNNs were shown to help with phase reconstruction for convergent-beam electron diffraction (CBED)-based scanning transmission electron microscopy (TEM)<sup>77</sup> while molecular structure imaging was found to benefit from such CNNs as well.<sup>78</sup> At the core of the deep learning (DL) methods employed in these works lie the same supervised DL techniques as used for phase determination. Again, generative ML for electron diffraction is not so common. Here, we will show how a so-called conditional generative adversarial network (cGAN) can be used to make accurate predictions of large-angle CBED (LACBED) images from just standard crystal information as encoded, e.g., in the usual text information<sup>79</sup> given in the Inorganic Crystal Structure Database (ICSD),<sup>80</sup> the world's largest such database.

## 2. A brief recap of the ML approach to phases and structures

### 2.1 Classification and regression

ML differs from traditional programming in that it does not rely on explicit rules to solve tasks. Instead, the network is expected to develop a strategy based on the input dataset to accomplish the required task. There are three primary types of learning: supervised learning, unsupervised learning, and reinforcement learning.<sup>71</sup> Here, we will focus mainly on the first two.<sup>9</sup>

Supervised learning aims to discover the optimal strategy for performing a task by using a labeled dataset. Within supervised learning, two key tasks can be identified: classification and regression. In classification, the ML model learns to divide data into distinct categories. Essentially, it finds an optimal representation of the dataset that separates samples into different classes. In regression, the algorithms are trained to understand the relationship between inputs and labels, enabling them to make continuous predictions for new, unseen labels based on the given inputs. This sets regression apart from classification, as it allows the model to predict values for data not encountered during training.

The second type of learning is unsupervised learning. In this approach, the ML algorithm processes unlabeled data and is expected to uncover hidden patterns or correlations without any external guidance. Unsupervised learning is further divided into three categories: clustering, dimensionality reduction, and association learning. Clustering aims to group similar samples within the dataset. Dimensionality reduction seeks to simplify the data representation while retaining its es-

sential characteristics. Association learning explores relationships between different samples in the dataset. Unsupervised learning has a wide range of applications. It can be used as a preprocessing step to reveal the structure of a dataset before supervised learning begins.<sup>49</sup> It also powers generative methods, such as VAEs and GANs, which create new data samples.

### 2.2 Generative ML: VAEs and cGANs

A VAE represents a relatively recent deep learning architecture that integrates standard compression techniques with the regularization strategies of machine learning, functioning simultaneously as a generative model.<sup>74,81-83</sup> In essence, a VAE comprises an *encoder*, which is a multilayered neural network trained on input data to generate output parameters for a variational distribution. These parameters define a low-dimensional probabilistic distribution, referred to as the *latent space*. The *decoder*, another deep neural network architecture, then reconstructs the output data from the latent space, drawing samples from this space rather than selecting deterministic points.

When the latent space dimensionality,  $d$ , is significantly smaller than the information content of the input data, some degree of information loss is inevitable. Thus, the goal is to design the encoder and decoder in such a way that maximizes the preservation of information during encoding while minimizing the error in the reconstructed data during decoding.

To effectively train a VAE, two primary loss functions are utilized. The *reconstruction loss*  $\ell_\epsilon$  measures the discrepancy between the input and reconstructed output during training. Additionally, the Kullback-Leibler divergence,<sup>84</sup> which serves as a regularization term, ensures that the latent space approximates a standard normal distribution.<sup>74</sup> In practice, the training process involves minimizing a total loss  $\ell$ , which is a combination of the reconstruction loss  $\ell_\epsilon$  and the Kullback-Leibler loss  $\ell_{KL}$ , such that  $\ell = \ell_\epsilon + c\ell_{KL}$ , where  $c$  is a hyperparameter that balances the two components.<sup>74</sup>

GANs have emerged as a highly popular architecture for image-to-image translation tasks.<sup>82,83</sup> While VAEs are known to struggle with producing high-fidelity outputs, often resulting in blurriness,<sup>85</sup> GANs inherently avoid this issue by design.<sup>72</sup> An absence of blurriness is particularly critical in quantitative electron diffraction, where clarity is essential. For this reason, we focus on cGANs,<sup>86</sup> which are well-suited to our image-to-image task involving the learning of a mapping from an input image  $x$  and random noise vector  $z$  to a target image  $y$ , denoted as  $G : x, z \rightarrow y$ . In this context,  $G$  represents the *generator*. GANs also introduce a second component, the *discriminator*, denoted as  $D$ . The discriminator is trained to differentiate between 'real' images from the dataset and 'fake' images generated by  $G$ . This adversarial setup ensures that the generator improves over time, as the discriminator learns to recognize blurry images as fake, thereby driving the generator to produce sharper outputs. Unlike VAEs, which rely on a predefined loss function, GANs instead learn a loss function for the desired task, solving another problem: deciding which loss function to use for comparing diffraction patterns is not a priori clear and can vary between different applications.

## 3. Predicting percolating clusters with CNNs

This section reviews work done previously,<sup>52,53</sup> where we showed that standard CNNs, usually employed in im-

age recognition ML tasks, also work very well for classifying site percolation states according to occupation probability  $p$  as well as for regression when determining  $p$  from such states. However, analyzing in detail whether spanning clusters at  $p < p_c$  or non-spanning clusters at  $p > p_c$  are correctly identified, we found that the same CNNs consistently fail to reflect the ground truth. Rather, it appears that the CNNs use  $p$  as a proxy measure to inform their classification predictions — a strategy that is obviously false for the percolation problem.

### 3.1 The physics model of “percolation”

The percolation problem is well-known with a rich history across the natural sciences.<sup>87–92</sup> It provides the usual statistical characteristics across a second-order transition such as, e.g., critical exponents, finite-size scaling, renormalization and universality.<sup>88</sup> Briefly, on a percolation lattice of size  $L \times L$ , individual lattice sites  $\vec{x} = (x, y)$ ,  $x, y \in [1, L]$ , are randomly occupied with *occupation probability*  $p$  such that the state  $\psi$  of site  $\vec{x}$  is  $\psi(\vec{x}) = 1$  for occupied and  $\psi(\vec{x}) = 0$  for unoccupied sites. We say that a connection between neighboring sites exists when these are side-to-side nearest-neighbors on the square lattice, while diagonal sites can never be connected. A group of these connected occupied sites is called a *cluster* (cf. Fig. 1(a)). Such a cluster then *percolates* when it spans the whole lattice either vertically from the top of the square to the bottom or, equivalently, horizontally from the left to the right. Obviously, for  $p = 0$ , all sites are unoccupied and no spanning cluster can exist while for  $p = 1$  the spanning cluster trivially extends throughout the lattice. In Fig. 1(a), we show examples of percolation clusters generated for various  $p$  values. The *percolation threshold* is at  $p = p_c(L)$ , such that for  $p < p_c(L)$  most clusters do not span while for  $p > p_c(L)$  there is at least one spanning cluster. This can be expressed via the quantities  $P(p)$ ,  $Q(p) = 1 - P(p)$  that denote the probabilities of the presence or absence of the spanning cluster at a given  $p$ , respectively (cf. Fig. 1(b)). We note that  $P$  is a finite- $L$  version of  $\psi$  in the notation of Ref.<sup>92</sup> We will occasionally emphasize this point using  $P_L$  and, likewise,  $Q_L$ . For an infinite system ( $L \rightarrow \infty$ ), one finds the emergence of an infinite spanning cluster at  $p_c = 0.59274605079210(2)$ . This estimate has been determined numerically evermore precisely over the preceding decades<sup>93</sup> while no analytical value is yet known.<sup>92</sup>

### 3.2 The ML approach to the percolation problem and the generation of ML “data”

Several ML studies on the percolation model have been published, mostly using supervised learning in order to identify the two phases via ML classification.<sup>49–51,54,55</sup> In order to facilitate the recognition of percolation with image recognition tools of ML, we have generated finite-sized  $L \times L$ , with  $L = 100$ , percolation states, denoted as  $\psi_i(p)$ , for the 31  $p$ -values  $0.1, 0.2, \dots, 0.5, 0.55, 0.555, 0.556, \dots, 0.655, 0.66, 0.7, \dots, 0.9$ . For each such  $p$ ,  $N = 10000$  different random  $\psi_i(p)$  have been generated. Each state  $\psi_i(p)$ ,  $i = 1, \dots, N$ , is of course just an array of numbers with 0 denoting unoccupied and 1 occupied sites. Nevertheless, we occasionally use for convenience the term “image” to denote  $\psi_i(p)$ . The well-known Hoshen-Kopelman algorithm<sup>94</sup> is employed to identify and label clusters from which we deter-

mine the presence or absence of a spanning cluster. Correlation measures have also been calculated but are not shown here for brevity.<sup>52,53</sup>

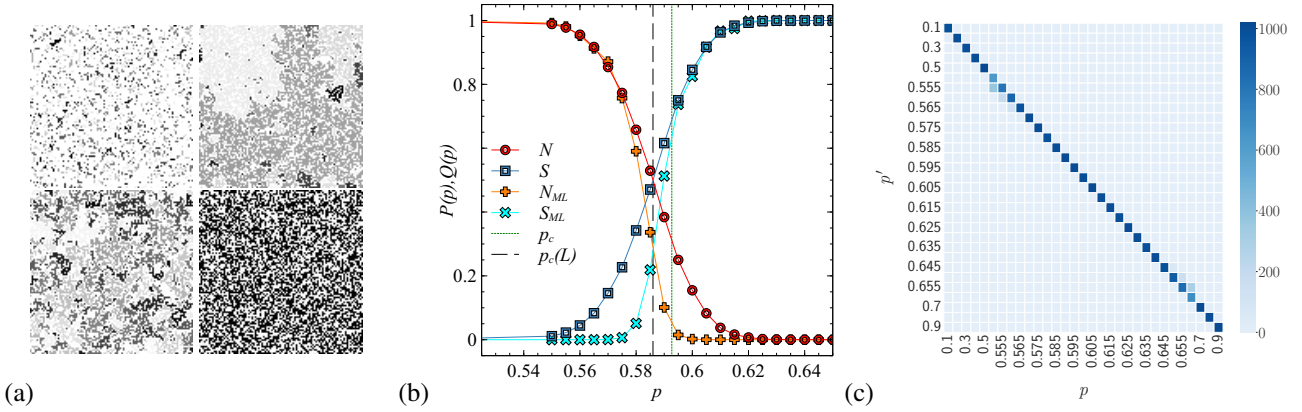
We emphasize that in the construction, we took care to only fabricate states such that for each  $p$ , the number of occupied sites is exactly  $N_{\text{occ}} = p \times L^2$  and hence  $p$  can be used as exact label for the supervised learning approach. Hence  $p = N_{\text{occ}}/L^2$  can also be called the *percolation density*. For the ML results discussed below, it will also be important to note that the spacing between  $p$  values reduces when  $p$  reaches 0.5 with the next  $p$  value given by 0.55 and then 0.555. Similarly, the  $p$  spacing increases as 0.655, 0.66, 0.7. We will later see that this results in some deviations from perfect classification/regression. For reference, we now have 12 values  $p = 0.1, \dots, 0.58 < p_c(100)$  and 18 values  $p = 0.59, \dots, 0.9 > p_c(100)$ . We also note that the training set contains 92.7% of states without a spanning cluster below  $p_c$  and 94.8% are spanning above  $p_c$ . We have also generated similar training and test sets for  $L = 200$ ; our results do not change significantly.<sup>53</sup> Last, all our ML results have been obtained from ten training, validation and test cycles allowing us to quote ML indicators, such as losses, accuracies, in terms of averages and their errors.<sup>53</sup> Our CNN uses the ResNET18 implementation of PyTorch.<sup>95</sup>

### 3.3 Results for ML classification according to spanning or non-spanning properties

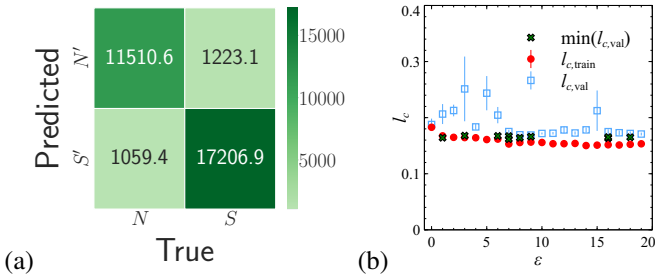
The hallmark of the percolation transition is the existence of a spanning cluster which determines whether the system is percolating or not.<sup>88</sup> We now want to check this and label all states only according to whether they are spanning or non-spanning, i.e., removing all density-related labels and various possible feature leakages. From Fig. 1(b), it is immediately clear that for finite-sized systems considered here, there are a non-negligible number of states which appear already spanning even when  $p < p_c$  and, vice versa, are still non-spanning when  $p > p_c$ . Furthermore, we note that for such  $L$ , the difference between  $p_c$  and  $p_c(L)$  is large enough to be important and we hence use  $p_c(L) = 0.585(5)$  as the appropriate value to distinguish the two phases.

Fig. 2 shows the averaged results after  $\epsilon = 20$  with a validation loss of  $\min_{\epsilon}[l_{c,\text{val}}] = 0.165 \pm 0.001$  (corresponding to a maximal validation accuracy  $\max_{\epsilon}[a_{c,\text{val}}] = 92.702\% \pm 0.001$ ). At first glance, the figure seems to indicate a great success: from the 31000 states present in  $\tau$ , 11510.6 have been correctly classified as non-spanning (i.e.,  $N \rightarrow N'$ ), and 17206.9 as spanning ( $S \rightarrow S'$ ) while only 1223.1 are wrongly labeled as non-spanning ( $S \rightarrow N'$ ) and 1059.4 as spanning ( $N \rightarrow S'$ ) (We note that these numbers are not integers since they are computed as averages over 10 independent training runs<sup>53</sup>). Overall, we would conclude that 92.6% of all test states are correctly classified while 7.4% are wrong.

However, from the full percolation analysis for  $\tau$ , we can compute that there are 11127 states (92.7%) without a spanning cluster below  $p_c(L)$  while 873 states (7.3%) already contain a spanning cluster. Similarly, for  $p > p_c(L)$ , 94.9% of states, equivalent to 17075 states, are spanning and 5.1% are not, corresponding to 925 states. At  $p_c(L) = 0.585$ , we furthermore have 482 spanning and 518 non-spanning states. Hence in total, we expect 2280 wrongly classified states. Since the last number is very close to the actual number of



**Fig. 1.** (Color online) (a) Examples of four percolation clusters of size  $L^2 = 100^2$ , obtained for  $p = 0.2 < p_c$ ,  $p = 0.6 > p_c$  in the top row and  $p = 0.5$ , i.e., just below  $p_c$ , in the bottom row. Occupied sites are marked by small dots while empty sites are left white. Each cluster of connected sites has been identified through the Hoshen-Kopelman algorithm. While individual clusters have been highlighted with different gray scales for the first three images, the bottom right image with  $p = 0.5$  shows all occupied sites in black only, irrespective of cluster identity. This latter representation is used below for the ML approach. (b) The square ( $\square$ ) symbols ( $\circ$  symbols) show the probability to have a spanning,  $P(p)$  (non-spanning  $Q(p)$ ) sample in the training dataset. The cross ( $\times$ ) symbols ( $+$  symbols) denote the corresponding ML prediction for the probability to have a spanning (non-spanning) sample, according to the trained network. The lines connecting the symbols are only a guide to the eye. The vertical lines indicate the percolation thresholds as given in the legend. (c) Average confusion matrix for classification according to  $p$ . The dataset used is the test data  $\tau$  and the models used for predictions are those corresponding to a minimal  $l_{c,\text{val}}$ . True labels for  $p$  are indicated on the horizontal axis while the predicted labels are given on the vertical axis. The shading scale represents the number of samples in each matrix entry.



**Fig. 2.** (Color online) (a) Average confusion matrix for classification according to spanning/non-spanning. The dataset used is the test data  $\tau$  and the models used for predictions are those corresponding to a minimal  $l_{c,\text{val}}$ . The true labels for  $N$  and  $S$ , are indicated on the horizontal axis while the predicted labels are given on the vertical axis. (b) Dependence of losses  $l_{c,\text{train}}$  and  $l_{c,\text{val}}$ , averaged over ten independent training seeds, on the number of epochs  $\epsilon$  for classification according to spanning/non-spanning. The circles (solid  $\bullet$ ) denote  $l_{c,\text{train}}$  while the squares (open  $\square$ ) show  $l_{c,\text{val}}$ . The crosses ( $\times$ ) indicate the minimal  $l_{c,\text{val}}$  for each of the ten trainings.

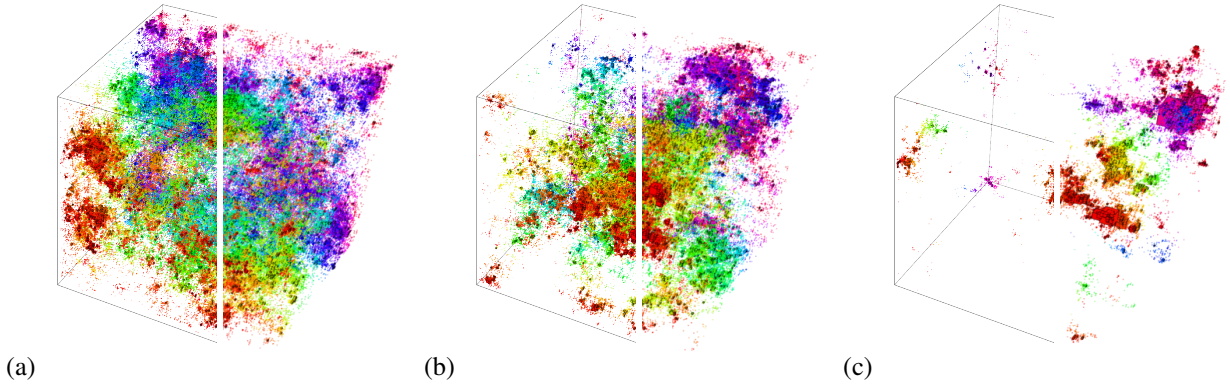
2282.5 of misclassified states, this suggests that it is precisely the spanning states below  $p_c(L)$  and the non-spanning ones above  $p_c(L)$  which the DL network is unable to recognize. Let us rephrase for clarity: the CNN, when trained in whether a cluster is spanning or non-spanning, completely disregards this information in its classification outputs. We show that this is indeed the case by a detailed analysis of the clusters around  $p_c$ ,<sup>53)</sup> a bespoke training just at fixed  $p_c$ <sup>96)</sup> and also using test sets which have been constructed to allow testing for the existence of the spanning cluster.<sup>53)</sup>

In summary, when looking at  $p$ , classification and regression techniques for percolation states allow us to obtain good recognition with near-perfect  $\langle a_{c,\text{val}} \rangle = 99.323\% \pm 0.003$  for classification (cf. also Fig. 1(c)) and near-zero  $\langle l_{c,\text{val}} \rangle = 0.000062 \pm 0.000012$  average mean-square loss for regression.<sup>52)</sup> On the other hand, the DL network completely ignores whether a cluster is spanning or non-spanning, essentially missing the underlying physics of the percolation prob-

lem — it seems to still use  $p$  as its main ordering measure. We believe that the root cause of the failure to identify the spanning clusters, or their absence, lies in the fundamentally *local* nature of the CNN: the filter/kernels employed in the ResNets span a few local sites only. Hence it is not entirely surprising that such a CNN cannot correctly identify the essentially *global* nature of spanning clusters. But it is of course exactly this global percolation that leads to the phase transition. This should serve as a warning to enthusiastic proponents of the ML approach not to ignore the physics undeservedly.

#### 4. Resolving disorder strengths from images of the 3D Anderson model

One of the hardest challenges in modern eigenvalue computation is the numerical solution of large-scale eigenvalue problems, in particular those arising from quantum physics.<sup>97)</sup> Typically, these problems require the computation of some eigenvalues and eigenvectors for systems which have up to several million unknowns due to their high spatial dimensions. Here, the Anderson model of localization<sup>98)</sup> is a particularly paradigmatic model as its underlying structure involves random perturbations of matrix elements which invalidates simple preconditioning approaches based on the graph of the matrices.<sup>99)</sup> Its physical importance comes from the prediction of a spatial confinement of the electronic motion upon increasing the disorder – the so-called Anderson localization.<sup>100)</sup> When the model is used in three spatial dimensions, it exhibits a metal-insulator transition in which the disorder strength  $w$  mediates a change of transport properties from metallic behavior at small  $w$  via critical behavior at the transition  $w_c \sim 16.57$  to insulating behavior and strong localization at larger  $w > w_c$ .<sup>101)</sup> The 3D Anderson model hence provides us with a physically meaningful quantum problem in which to use ML strategies to distinguish its two phases, namely the metallic phase with *extended* states at  $w < w_c$  and the insulating phase with *localized* states at  $w > w_c$  (Occasionally, one might want to also study  $w \approx w_c$  as a 3rd phase), while avoiding the many challenges of fully interacting quan-



**Fig. 3.** (Color online) Extended (left), critical (center) and localized (right) wave function probabilities  $|\psi(\vec{r})|^2$  for the 3D Anderson model with periodic boundary conditions at  $E = 0$  with  $N = 100^3$  and  $w = 14, 16.5$  and  $19$ , respectively. Every site with probability  $|\psi(x, y, z)|^2$  larger than the average  $1/N^3$  is shown as a box with volume  $N|\psi_{E=0}(x, y, z)|^2$ . Boxes with  $N|\psi(x, y, z)|^2 > \sqrt{1000}$  are plotted with black edges. The shading of each cube distinguishes between different slices of the system along the axis into the page. In each panel, the left half is the originally constructed image while the right half shows the image in its converted PNG form with  $500 \times 500$  pixel resolution. Obviously, upon conversion, the black edges around the large  $|\psi(x, y, z)|^2$  become less prominent and the overall black frames are also removed.

tum systems.<sup>102)</sup> In this sense, it can be seen as the quantum ML test partner to complement the classical statistical physics tests available via the percolation and Ising-type models. Similarly to the percolation model, previous ML studies have already been performed and showed good success for ML classification with CNNs to identify the two phases of the system.<sup>69,70,103,104)</sup> Here, we show that not only phases but also disorder strengths can be recovered from eigenstates of the 3D Anderson model.

#### 4.1 The formulation of the Anderson model in 3D

In its usual form, the localization problem in 3D with coordinates  $x, y, z$  corresponds, in the absence of a magnetic field, to a Hamilton operator in the form of a real symmetric matrix  $H$ , with quantum mechanical energy levels given by the eigenvalues  $E_n$ . The respective wave functions are simply the eigenvectors of  $H$ , i.e., vectors  $\psi_n(\vec{r}) \in \mathbb{C}$  for  $\vec{r} = (x, y, z)$ . With  $N = M^3$  sites, the quantum mechanical (stationary) Schrödinger equation is equivalent to the eigenvalue equation  $H\psi_n = E_n\psi_n$ , which in site representation reads as

$$\sum_{\sigma=\pm} \psi_n(\vec{r} + \sigma\vec{a}) + \psi_n(\vec{r} + \sigma\vec{b}) + \psi_n(\vec{r} + \sigma\vec{c}) = [E_n - \varepsilon(\vec{r})] \psi_n(\vec{r}), \quad (1)$$

with  $\vec{a} = (1, 0, 0)$ ,  $\vec{b} = (0, 1, 0)$  and  $\vec{c} = (0, 0, 1)$  denoting the lattice vectors of a periodic, simple cubic lattice. The disorder usually<sup>105)</sup> enters the matrix on the diagonal, where the entries  $\varepsilon_n(\vec{r})$  correspond to a spatially varying disorder potential and are selected randomly according to a suitable distribution.<sup>97)</sup> Here, we shall use the standard box distribution  $\varepsilon(\vec{r}) \in [-w/2, w/2]$  such that  $w$  parameterizes the aforementioned disorder strength. For disorders  $w \ll w_c$ , most of the eigenvectors are *extended*, i.e.,  $\psi_n(\vec{r})$  fluctuating from site to site, but the envelope  $|\psi_n|$  is approximately a nonzero constant. For large disorders  $w > w_c$ , all eigenvectors are *localized* such that the envelope  $|\psi_n|$  of the  $n$ th eigenstate may be approximately written as  $\sim \exp[-|\vec{r} - \vec{r}_n|/\xi(w)]$  with  $\xi(w)$  denoting the *localization length* of the eigenstate. Directly at  $w = w_c$ , the last extended states at  $E = 0$  vanish. The wave function vector  $\psi_{E=0}(\vec{r})$  appears simultaneously extended and localized

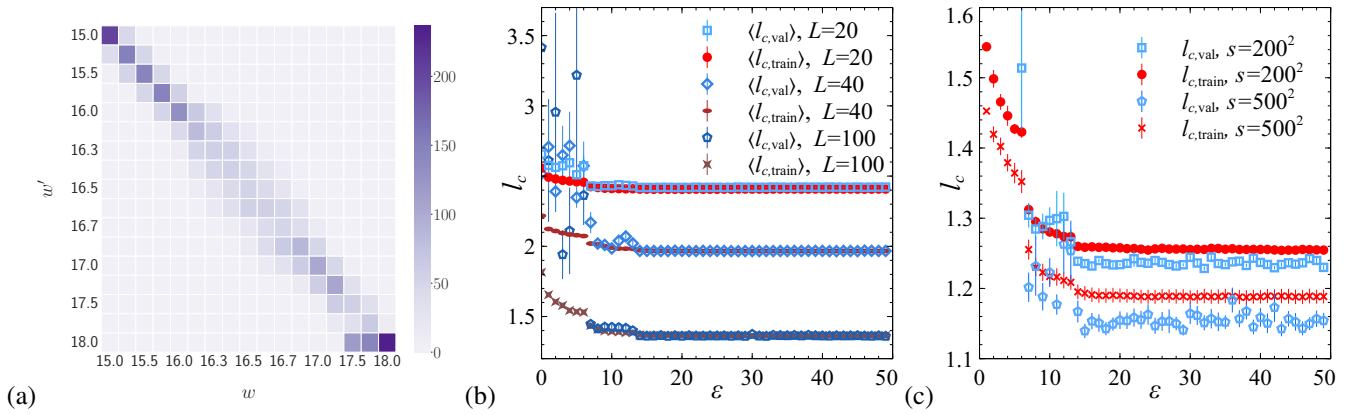
and has multifractal properties.<sup>67,68)</sup> In Fig. 3, we show examples of such states.

In order to numerically distinguish the two (or three) phases mentioned before, one usually needs to (i) go to rather large system sizes of order  $N = 10^6$  to  $10^8$  and (ii) average over many different realizations of the disorder, i.e., compute eigenvalues or eigenvectors for many matrices with different diagonals.<sup>67,68,100,101)</sup> In the present work, we concentrate on the computation of a few eigenvalues and corresponding eigenvectors for the physically most interesting case around the critical disorder  $w_c$  and in the center of the spectrum  $\sigma(H)$ , i.e., at  $E = 0$ , for large system sizes.

#### 4.2 Hamiltonian eigenfunctions as data

The square-normalized eigenstates  $\psi_n = \sum_{x,y,z} \psi_n(x, y, z)|x, y, z\rangle$  have been numerically obtained using the JADAMILU library.<sup>106)</sup> The  $|x, y, z\rangle$  indicate the orthonormal Wannier basis in the usual tight-binding formulation. For the 17 disorders  $w = 15, 15.25, \dots, 16, 16.2, \dots, 17, 17.25, \dots, 18$  we consider for training and validation a previously used dataset<sup>67,68)</sup> with 5000 disorder realization for each disorder and system sizes  $N = 20^3, 30^3, \dots, 100^3$ . For all the data,<sup>67,68)</sup> we have considered a single eigenstate per sample (disorder realization) with energy close to  $E = 0$ . This is costly in terms of computing time but essential to avoid the noticeable correlations that exist between eigenstates of the same sample.<sup>67,68)</sup> In addition, we have generated, for each of the disorders, 500 independent test wave functions at  $E = 0$ , i.e., using random numbers with different seeds.

In order to be able to use standard 2D image recognition machine learning tools, we represent the  $\psi_n$  graphically as in Fig. 3. We remove the black box and the color scale before using the images for training, validation and testing purposes. Furthermore, the images are converted from their original postscript, using the IMAGEMAGICK set of routines, and rendered as portable network graphics (PNG) in the pixel resolutions of  $s = 100 \times 100, 200 \times 200$  and  $500 \times 500$ . This conversion results in some changes in the visual presentation as shown in Fig. 3.



**Fig. 4.** (Color online) (a) Average confusion matrix for *image* classification of the 17 disorders  $w = 15, 15.25, \dots, 16, 16.2, \dots, 17, 17.25, \dots, 18$  for system size  $N = 100^3$  and image resolution  $100 \times 100$ . The dataset used is the test data  $\tau$  and the models used for predictions are those corresponding with a minimal  $l_{c,\text{val}}$ . (b) Dependence of losses  $l_{c,\text{train}}$  and  $l_{c,\text{val}}$  on the number of epochs  $\epsilon$  for disorder classification and the three system sizes  $N = 20^3, 40^3$  and  $100^3$  (a). The squares (open  $\square$ ) denote  $l_{c,\text{train}}$  while the circles (solid  $\bullet$ ) show  $l_{c,\text{val}}$ . (c) Epoch dependence of  $l_{c,\text{train}}$  and  $l_{c,\text{val}}$  for  $N = 100^3$  and two image resolutions,  $200 \times 200$  and  $500 \times 500$ . Symbols are as in (b),  $s$  indicates the size of the images.

### 4.3 ML models and results

Previous ML studies of the Anderson model use CNNs composed of 6 convolutional layers and a fully connected layer to identify the extended and localized phases from the  $|\psi(x, y, z)|^2$ .<sup>69,70,104</sup> Here, our goal is to expand on these results and show that a RESNET18, as used in section 3, can also recover the value of  $w$  used in images made from these  $|\psi|^2$ .

We first establish the capacity of the RESNET18 to identify the two phases of the 3D Anderson model of localization from images (not shown here). Here, we want to train a network to identify individual disorder values. Following a similar strategy as in section 3, we train our network for 17 disorder values  $w = 15, 15.25, \dots, 16, 16.2, \dots, 17, 17.25, \dots, 18$  for fixed  $N = 20^3, 40^3, 100^3$  and  $s = 100^2$ . After training the 17 disorder values for  $N = 20^3$ , we obtain a  $\min_{\epsilon}[\langle l_{c,\text{val}} \rangle] = 2.408 \pm 0.003$  (corresponding to an accuracy of  $\max_{\epsilon}[\langle a_{c,\text{val}} \rangle] = 15.9\% \pm 0.2$ ). At first, the performance of the network on this system appears to be rather limited. From the confusion matrix obtained after training (not shown), we notice that only the smallest and largest disorders, i.e.,  $w = 15$  and  $w = 18$ , are perfectly classified. We increase the size of the system and train our network for  $N = 40^3$ . Following the training we reach  $\min_{\epsilon}[\langle l_{c,\text{val}} \rangle] = 1.951 \pm 0.004$  (corresponding to an accuracy of  $\max_{\epsilon}[\langle a_{c,\text{val}} \rangle] = 25.7\% \pm 0.2$ ). Looking at the metrics in Fig. 4 (b), we notice the decrease of  $\langle l_{c,\text{val}} \rangle$  and  $\langle l_{c,\text{train}} \rangle$  between the training for  $N = 20^3$  and  $N = 40^3$ . Still, the apparent improvement in the performance of the network is not yet convincing. We finally train for  $N = 100^3$  and  $s = 100 \times 100$ . We obtain  $\min_{\epsilon}[\langle l_{c,\text{val}} \rangle] = 1.327 \pm 0.006$  (corresponding to an accuracy of  $\max_{\epsilon}[\langle a_{c,\text{val}} \rangle] = 43.3\% \pm 0.3$ ). This is an increase of almost 18%. Even though the accuracy is still less than 50%, the network seems to be getting better at recognizing the  $w$  values. The confusion matrix obtained after this training is given in Fig. 4(a). Clearly, the matrix is heavily diagonally dominant: misclassifications appear to exist mostly between directly adjacent disorder values. Thus, while the training does not result in a perfect recognition of  $w$ 's, it is nevertheless already very good in recognizing the vicinity of each  $w$ , even very close to the metal-insulator transition. Increasing the size of the input images to  $s = 200 \times 200$  does not help to provide significant improvement. After train-

ing we obtain  $\min_{\epsilon}[\langle l_{c,\text{val}} \rangle] = 1.216 \pm 0.003$  (corresponding to an accuracy of  $\max_{\epsilon}[\langle a_{c,\text{val}} \rangle] = 47.9\% \pm 0.2$ ). Furthermore, training for such a large input leads to a substantial increase in training time.

In summary, we find that even using images of eigenstates allows to distinguish the phase of the 3D Anderson model well, while the classification of  $w$  values proceeds with nearly the same accuracy as in the case of classifying  $p$  for percolation in section 3. Furthermore, increasing the system size from  $N = 20^3$  to  $100^3$  improves the predictions considerably. Such finite-size effects remind us rather reassuringly that the ML strategies are obviously subject to the same physics constraints as standard approaches.

## 5. Predicting phases of the $J_1$ - $J_2$ Ising model with VAEs

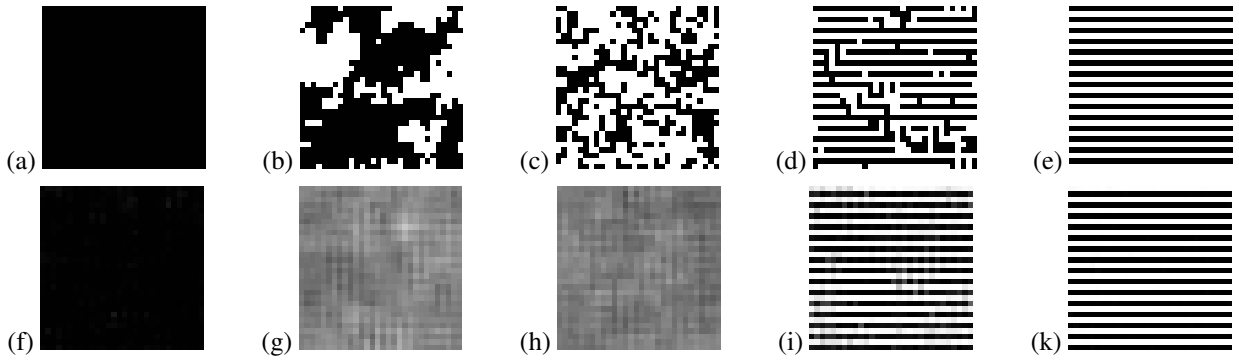
The  $J_1$ - $J_2$  Ising model serves as a still relatively simple system to illustrate an already more complex 3-phase behavior.<sup>107–118</sup> With  $J_1$  denoting the nearest-neighbor interaction, the competing second-neighbor interaction  $J_2$  gradually suppresses the ordering temperature, until it vanishes completely when  $J_2 = |J_1|/2$ .<sup>107,108</sup> Furthermore, beyond this point, a new ordered “superantiferromagnetic phase” appears. The universality class of the transition into the superantiferromagnetic phase has been investigated early on,<sup>107–109</sup> but still continues to attract attention since its nature remains controversial.<sup>116–118</sup> There is at least also one investigation of this model on the D-wave quantum annealer<sup>119</sup> and a small number of machine-learning investigations.<sup>28,38</sup>

### 5.1 Definition of the $J_1$ - $J_2$ Ising model

The Hamiltonian of the  $J_1$ - $J_2$  Ising model can be expressed as

$$H_{J_1 J_2} = -J_1 \sum_{\langle i, j \rangle} s_i s_j + J_2 \sum_{\langle\langle i, j \rangle\rangle} s_i s_j, \quad (2)$$

where  $s_i$  represents the spin at site  $i$ , which can be either up (+1) or down (-1);  $\langle i, j \rangle$  refers to nearest-neighbor pairs,  $\langle\langle i, j \rangle\rangle$  denotes next-nearest neighbor pairs, while  $J_1, J_2 \geq 0$  signify the interaction strengths between the nearest and next-nearest neighbors, respectively. Our chosen sign conventions in Eq. (2) lead to a ferromagnetic coupling for  $J_1$  pairs while



**Fig. 5.** (a-e) Five illustrative spin configurations of the  $J_1$ - $J_2$  Ising model on a periodic  $30 \times 30$  square lattice. (f-k) Predicted spin configurations from (a-e), respectively, using our VAE learning. Panels (a, b, c) keep  $J_2 = 0.1$  fixed while increasing  $T$  from (a) ferromagnetic at  $T = 0.1$  to (b)  $T = 1.975$  near the ferromagnetic-to-paramagnetic transition and (c) a configuration deep in the paramagnetic phase at  $T = 4.0$ . Panels (d, e) have  $J_2 = 0.8$  and then decrease  $T$  from (d)  $T = 1.575$  near the paramagnetic-to-superantiferromagnetic transition to (e) a superantiferromagnetic configuration at  $T = 0.1$ . In (f-k), the parameters are as in (a-e). In all cases, the black squares correspond to up spins while white is for down spins. In (f-k), the values in the interval  $[-1, 1]$  are denoted by the gray squares in the panels.

next-nearest neighbors prefer to align in an antiferromagnetic structure.<sup>120</sup> The three distinct phases of the model correspond to (i) a low-temperature, low- $J_2$  *ferromagnet*, (ii) a low-temperature, high- $J_2$  *superantiferromagnet* and (iii) the high-temperature *paramagnet*. We illustrate spin configurations representative of these phases and close-to-phase transitions in Fig. 5. In this work, we review recent work aiming to predict the three phases with a generative VAE, using a spin-adapted mean-squared error  $\varepsilon$  as ML cost function.<sup>120</sup>

## 5.2 Generating states as ML training data via the Metropolis Monte-Carlo approach

To generate the necessary input data for the training of the VAE, we utilize the Metropolis algorithm, a well-established method for simulating statistical models at finite temperature.<sup>121–124</sup> In the present investigation, we initially focus on a system size of  $30 \times 30$  with periodic boundary conditions.<sup>28,125</sup> In order to assess the influence of the size of the system, we also investigate  $60 \times 60$  and  $120 \times 120$  square lattices. Equilibration of the model can be difficult, in particular in the regime of  $J_2 \approx |J_1|/2$ .<sup>115</sup> We assure proper thermalization by successively cooling our configurations for fixed  $J_2/|J_1|$ .<sup>120</sup> We set the energy scale with  $J_1 = 1$ .

For  $J_2 = 0$ , we are back to the nearest-neighbor Ising model with known critical temperature  $T_{c,\text{Ising}} \approx 2.269$ .<sup>126</sup> We can therefore confidently start our exploration of the as-yet unknown phase diagram by choosing an initial temperature range of  $0 \leq T \leq 4 \approx 2 \times T_{c,\text{Ising}}$  containing  $T_{c,\text{Ising}}$ . We also know that the ferromagnetic-to-superantiferromagnetic transition is at  $J_2 = 1/2$ .<sup>115</sup> Hence we choose a range for  $J_2$  from 0 to 1.5. Should we later see that these ranges do not suffice to capture all phases, we could further increase the maximal  $T$  and  $J_2$  values. Using  $\Delta T = 0.025$ , we thus proceed with a set  $\mathcal{T}$  of  $|\mathcal{T}| = 157$  temperatures with  $T \in [0, 4]$  for  $T \in \mathcal{T}$ . The Monte-Carlo construction is repeated with different random numbers until we have  $C = 40$  configurations for each temperature at the given values of  $J_2$ . Let  $\mathcal{J}_2 = \{0, 0.1, 0.2, 0.3, 0.4, 0.45, 0.48, 0.49, 0.495, 0.5, 0.505, 0.51, 0.52, 0.55, 0.6, 0.65, 0.7, 0.8, 0.9, 1, 1.2, 1.5\}$  denote the  $|\mathcal{J}_2| = 22$  chosen distinct values. In total, this results in a dataset containing  $|\mathcal{T}| \times |\mathcal{J}_2| \times C = 157 \times 22 \times 40 = 138\,160$  independent

configurations for a given system size.

## 5.3 Reconstruction of the phase diagram using single-region VAEs

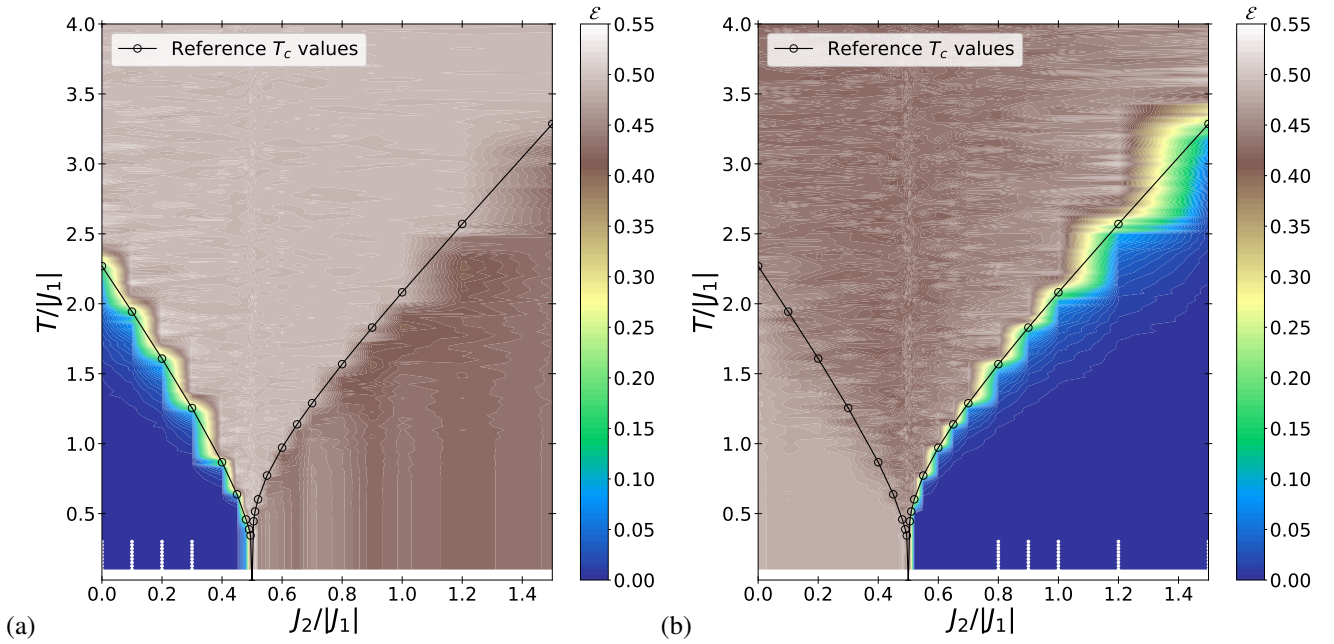
We can now use the VAE architecture to identify the phases of the  $J_1$ - $J_2$  model as a function of  $T$  and  $J_2$  for constant  $J_1 = 1$ . Details of the VAE implementation can be found elsewhere.<sup>120</sup> We start the training of the VAE for  $T \ll T_{c,\text{Ising}}$  in two distinct regions, namely (i)  $J_2 < 1/2$  and (ii)  $J_2 > 1/2$ . Consequently, we have two restricted training data regions  $\rho_{\text{low-}J_2}$  and  $\rho_{\text{high-}J_2}$ . In order to have a reasonable amount of training data, we use all 40 values for each  $(T, J_2)$  in each training region. For the results underlying Fig. 6(a), this amounts to 1440 training configurations in  $\rho_{\text{low-}J_2}$ , while for Fig. 6(b), we have 1800 configurations in  $\rho_{\text{high-}J_2}$ .

From Fig. 6 we see that indeed two distinct regions emerge. The low- $T$ , low- $J_2$  region shown in (a) is clearly separated from the rest of the  $(T, J_2)$  plane. Similarly, panel (b) establishes a low- $T$ , high- $J_2$  region. We note that in both cases, the  $\varepsilon$  values in the low/high- $J_2$  regions are close to zero, while in the other regions we have  $\varepsilon \approx 0.5$ . This value suggests that in both cases, the out-of-region configurations have about 50% of spins different, in agreement with the behavior in the known phases. We can therefore conclude the existence of two low- $T$  regions identified in Fig. 6. By exclusion, the third region corresponds to  $\varepsilon \approx 0.5$  from both trainings.

Indeed, these regions agree very well with the previously established phase boundaries shown in Fig. 6. The  $\varepsilon$  values of 0, 0.25, and 0.5 indicate best, random, and worst reconstruction possible, respectively, compatible with the spin configurations in each phase. Clearly, the regions with  $\varepsilon \approx 0$  correspond to the ordered ferro- and superantiferromagnetic phases in Fig. 6 (a) and (b), respectively. Further results with similar ML strategies as well as on a direct comparison of states can be found elsewhere.<sup>120</sup>

## 6. Microscopy with GANs

CBED<sup>127,128</sup> is a TEM technique with unparalleled sensitivity.<sup>129</sup> Its origins date back nearly 100 years to pioneering work<sup>130</sup> and its modern applications include crystal symmetry classification,<sup>131–133</sup> lattice parameter determination,<sup>134–136</sup> strain & defect analysis,<sup>137–140</sup> and more.<sup>141</sup>



**Fig. 6.** (Color online) Minimal averaged error  $\varepsilon$  for the VAE-based reconstruction of the  $J_1$ - $J_2$  model's phase diagram. The results correspond to  $L = 30$ . Panel (a) represents the in-phase learning from the low- $J_2$  region  $\rho_{\text{low-}J_2}$  and (b) gives results for the in-phase learning from the high- $J_2$  region  $\rho_{\text{high-}J_2}$ . The  $(T, J_2)$  data points of various training regions are indicated by small white dots for each  $(T, J_2)$  pair (usually these are closely spaced and hence appear as vertical lines). In all panels,  $\circ$  symbols connected by black lines denote known reference phase boundaries.<sup>115)</sup>

However, CBED sees the majority of its use in symmetry determination<sup>131)</sup> and charge density refinement<sup>142)</sup> and is still lacking in popularity when compared to the more established structure solution and refinement methods of X-ray and neutron diffraction.<sup>143,144)</sup> Collecting the necessary amount of high-quality diffraction data from a TEM, to construct a LACBED image, is one of the inherent challenges of the method. Here, modern computer-controlled TEM setups offer a clear advantage and can make the task near automatic.<sup>129)</sup> A perhaps even more constraining challenge lies in the fact that the complexity introduced by multiple scattering of electrons as they propagate through the specimen<sup>127)</sup> requires sophisticated modelling techniques to construct the theoretical predictions to compare with TEM results. To make CBED a more accessible approach, there have been two major computational methods developed: (i) the Bloch-wave method,<sup>127,144–146)</sup> and (ii) Multislice.<sup>146–151)</sup> Whilst both have seen success in accurately generating CBED patterns, they even today remain computationally resource- and time-intensive, often well beyond what a standard desktop computer can provide.<sup>152)</sup>

### 6.1 Selection of training data

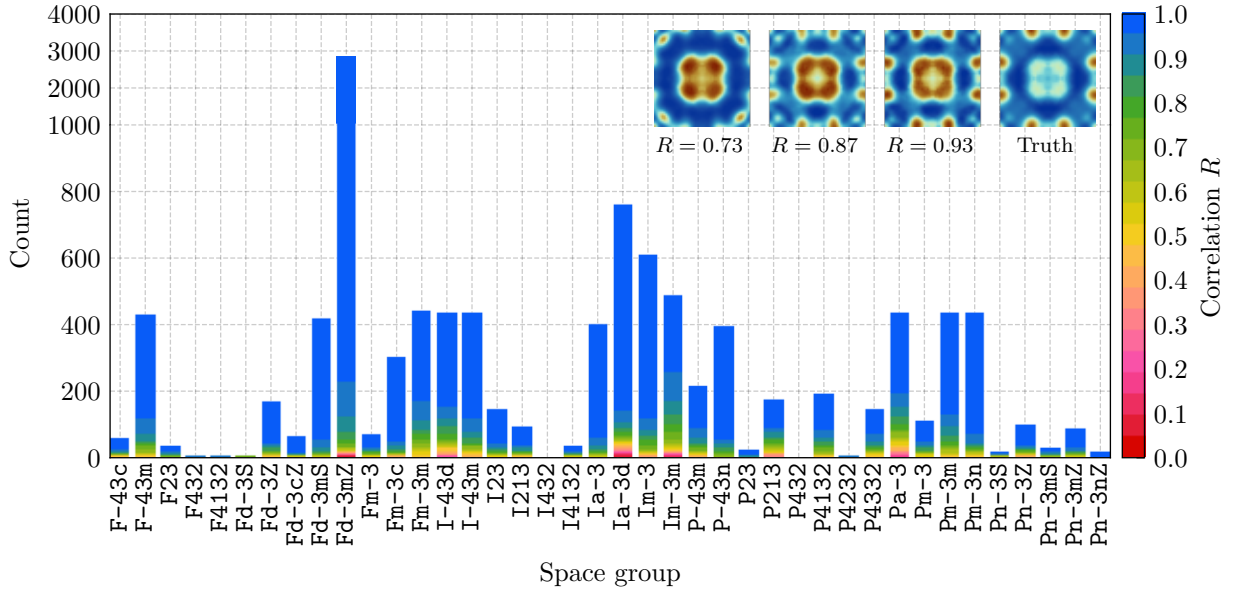
For our aim to generate LACBED patterns via ML, we require a large body of data in which crystal structure information has been paired with corresponding bright field LACBED images. Experience from previous such machine learning tasks in computer vision<sup>71,153)</sup> and related applications,<sup>52,53,69,70,125)</sup> as well as in the previous sections, suggests that often more than 10,000 such training pairs are needed. On the scale necessary for a successful model, it is infeasible to use experimental data for such patterns. Fortunately, the ICSD,<sup>80)</sup> as the world's largest such database, provides ready access to the full structural information for more

than 240,000 crystals in the form of a 'Crystallographic Information File' (CIF), a standard text file format.<sup>79)</sup>

Direct training with structured textual data, as available in the CIFs, is still a major challenge for machine learning tasks.<sup>154,155)</sup> Even small changes in, e.g., numerical values of the lattice parameters, can have major changes in the resulting CBED patterns. On the other hand, existing image-to-image translation tasks have been primarily optimized for 2D data. To harness this knowledge, we need a feasible way to represent the CIF information in 2D image form as well. Fortunately, the projected atomic potential  $\rho$  is a convenient such image representation. Using the structure factors  $F(\mathbf{g})$  of the crystal, obtained from FELIX,<sup>152)</sup> we generate the projected potential as  $\rho(\mathbf{r}) \propto \sum_{\mathbf{g}} F(\mathbf{g}) \cdot \exp[-2\pi i \mathbf{g} \cdot \mathbf{r}]$ . Here, the  $\mathbf{g}$  are the lattice vectors of the unit cell in reciprocal space. We normalize the resulting image of atomic potential strength and also restrict their size to  $128 \times 128$ . Note that in requiring all of these inputs to be the same image dimensions for the machine learning model, we lose information regarding the size of the crystal.

Next, we need a method to construct the LACBED patterns corresponding to each CIF and projected electron potential  $\rho$ . We employ FELIX, an open-source software implementation of the Bloch-wave method<sup>145,146)</sup> for generating LACBED images originally developed in part by one of us.<sup>152)</sup> FELIX has been shown to provide atomic coordinate refinements with sub-picometer accuracy,<sup>143,144)</sup> and can accurately simulate LACBED patterns.<sup>129)</sup> The software takes as input a CIF, beam parameters, microscope settings, crystal settings, and the desired beam direction. Most of these values were calibrated previously.<sup>143)</sup> In our simulations, we only consider the  $(0, 0, 0)$  beam direction for simplicity. The other simulation parameters used here are provided in the code accompanying the present work.<sup>156)</sup> We use FELIX to generate LACBED





**Fig. 7.** (Color online) The distribution of crystal structure data entries from the 10 accumulated test sets in each of the 36 (+4 from alternative origin choices) cubic space groups as obtained from the ICSD. The shading scale denotes the cross-correlation index  $R$ , see (3), obtained for each structure in the indicated space group using the trained cGAN, sorted from overall highest (top) to overall lowest (bottom)  $R$ . We note that the vertical scale above 1000 has been compressed for clarity. The LACBED images shown on the top correspond to different  $R$  values to give a more intuitive interpretation (using the same intensity scale as in Fig. 8).

images of size  $128 \times 128$ . Such sizes are sufficient for many computer-vision-based machine learning tasks,<sup>71)</sup> whilst remaining small enough to allow the generation of results on a large scale for our dataset.

Our strategy in generating the necessary input from the information provided in the ICSD is then as follows: (i) We convert the textual information provided by each crystal's CIF into the normalized projected electronic potential, i.e., a 2D image. (ii) we compute, via the Bloch-wave code FELIX, the corresponding bright-field LACBED images. While the construction of the electron potential images is very fast, generating the LACBED dataset takes a few weeks using bespoke high-performance computing architecture. In the current proof-of-principle work, we focus on the 12454 CIFs each corresponding to a unique *cubic* crystal. We ultimately have a dataset of 12454 image pairs of size  $128 \times 128$ , with pixel values between 0 and 255. Each pair consists of a crystal's projected electron potential and its simulated  $(0, 0, 0)$  LACBED diffraction pattern. This dataset is publicly hosted.<sup>157)</sup>

Whilst we use as many cubic crystals as we can, since after all, a much higher number is still desired, we encounter significantly imbalanced data in many areas. For example, when resolved according to their space group classification, we find that the ICSD data is highly imbalanced. As shown in Fig. 7, some space groups contain less than 10 ICSD entries while others have many thousands. It is well known that machine learning methods, and in particular our chosen adversarial network architecture, suffer in their predictive strength when using imbalanced data.<sup>158)</sup> Hence it could be worthwhile in future studies to include other crystal data from the ICSD beyond the cubic ones.

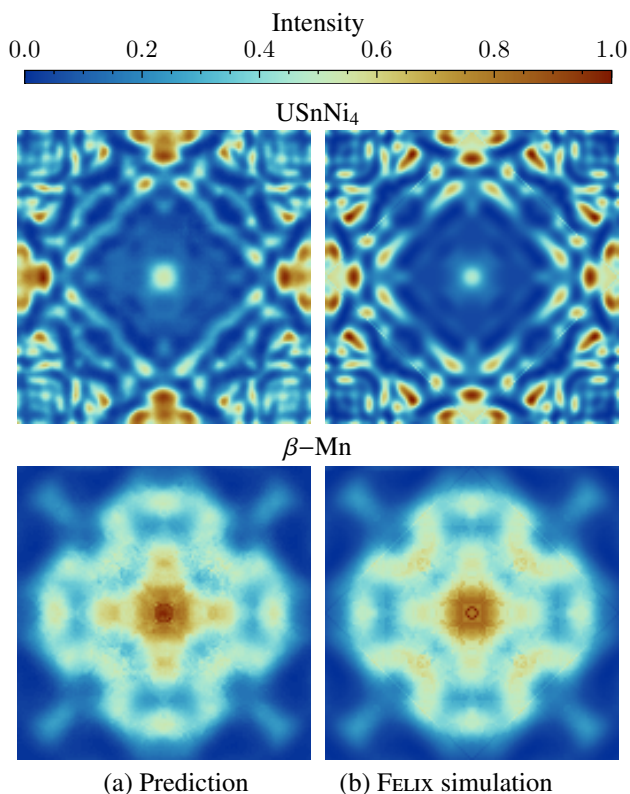
## 6.2 Results

We train a cGAN to create LACBED patterns by providing the projected electron density as input. We repeat each training 10 times and use a fresh 80%/20% split of our data into training/test data sets for each such learning cycle, respectively.<sup>156)</sup> The images given in Fig. 8 show a comparison between typical LACBED predictions of the cGAN and the expected behavior from the FELIX results. Before going into details on how we create these images, we start by noting that the ground truth LACBED images shown in the figure, with whom we compare our predictions, needed about 400 seconds each to be constructed by the Bloch-wave method on a high-performance compute cluster while our predicted LACBED images arrived within 20 milliseconds on a modern, i.e., GPU-supported, desktop.

In Fig. 8, we find an overall good qualitative agreement, in particular w.r.t. the underlying two-fold symmetries. While we use the standard mean-squared error in training the cGAN, we report a *shifted* zero-mean normalized cross-correlation fit index  $R$  for pixel intensities,

$$R(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{2} + \frac{1}{2n^2} \sum_{i,j} \frac{y_{ij} - \langle \mathbf{y} \rangle}{\sigma(\mathbf{y})} \cdot \frac{\hat{y}_{ij} - \langle \hat{\mathbf{y}} \rangle}{\sigma(\hat{\mathbf{y}})}, \quad (3)$$

which has often been used in CBED image comparison studies.<sup>129)</sup> Here,  $y_{ij}$  and  $\hat{y}_{ij}$  are the pixel intensities at  $(i, j)$ ,  $i, j = 1, \dots, n$ , for a cGAN-predicted and FELIX-simulated image, respectively, while the  $\langle \cdot \rangle$  denotes mean intensities and  $\sigma$  their standard deviation. In this normalization, the value  $R = 1$  corresponds to a perfect fit, while  $R = 0$  is perfectly anti-correlated. The values  $R = 0.5$  and  $0.75$  emerge when  $1/2$  or  $3/4$  of the image pixels are correlated and  $1/2$  and  $1/4$  are anti-correlated, respectively. Also,  $R = 0.5$  corresponds to two images with uncorrelated intensities. In Fig. 8 we show two typical results, namely example crystals with  $R$  values



**Fig. 8.** (Color online) Comparison of (left column) ML-predicted LACBED patterns for the [001] projected electron potential of a unit cell with (right column) the FELIX Bloch-wave simulation (“ground truth”). The cross-correlation fit index  $R$ , cf. Eq. (3), for the complete test data set has a mean of  $R = 0.934 \pm 0.011$  and median of  $0.990 \pm 0.002$ . The two unseen crystals displayed here have been selected as typical examples from the test data set. In the top row, we show  $\text{USnNi}_4$  (54390), from space group  $F43m$ . Its  $R$  value is close to the mean. In the bottom row, we display  $\beta\text{-Mn}$  (163414) from  $P4132$  with an  $R$  value close to the median. Computation of each simulated image via FELIX takes about 400 seconds with 48 CPU cores whilst the ML result is generated in under 20 milliseconds on a standard desktop PC+GPU. All images are  $128 \times 128$  pixels, have normalized amplitudes and use the same shading scale as indicated at the top of the image.

close to the mean and median values of  $R$  obtained for the full test data set. Our resulting typical  $R$  values for the LACBED image reconstruction, as given in the caption of Fig. 8, and the colored bars in Fig. 7, indicate an overall very good quantitative agreement.

## 7. Conclusions and Outlook

The learning aspects of DL networks are often referred to as “black boxes”, highlighting that it appears occasionally surprising how a DNN arrives at its classification, regression or generative predictions. On the other hand, it is exactly this lack of apriori imposed basic descriptors that allows a DL architecture to variationally construct its own set of descriptors to achieve an optimal prediction. So when ML succeeds in classifying states of Ising-type, percolation, and Anderson models, this also shows that the phase information must be encapsulated directly in the states alone, even for those relatively close to the phase boundaries as shown by the overall good reconstruction of these phases. While this was not unknown before or unexpected,<sup>67,68,101,159</sup> it is nevertheless an interesting qualitative insight to have re-emphasized. Conversely, this also suggests that simply comparing states with

each other, by mean-squared deviations,  $R$  correlation or otherwise, might also be an alternative quantitative method for phase diagram construction - as already demonstrated.<sup>120</sup>

The caveat discovered when studying the globally spanning cluster for the percolation problem with locally focused CNNs, i.e., the failure of such CNNs to correctly identify the percolating cluster,<sup>53</sup> furthermore suggests that even the power of modern ML approaches can fail when the underlying physics is ignored.<sup>102</sup> In this context it is also important to mention that the cGAN predictions for the outcome of electron interference experiments, i.e., the LACBED intensities, do not somehow circumvent the quantum mechanical measurement problem. Rather, they simply provide a good interpolation to the various diffraction solutions of the electron dynamical scattering problem provided by the Bloch wave calculations the cGAN was trained on.

Last, the review given here clearly reflects the prejudices and preferences of its authors in selecting the applications of ML to physics. Many other applications and application areas have been ignored such as Boltzmann machines<sup>9</sup> and the extremely interesting approaches to finding states of many-body systems.<sup>160–164</sup>

**Acknowledgments** J.J.W. and R.A.R. gratefully acknowledge discussions with Richard Beanland, Warwick, on the results presented here for the LACBED reconstructions. This work forms part of an ongoing project of all three of us. D.B., A.H., and R.A.R. are grateful for co-tutelle funding via the EUTOPIA Alliance. R.A.R. also very gratefully acknowledges the CY Initiative of Excellence (grant “Investissements d’Avenir” ANR-16-IDEX-0008). The computations presented here used the University of Warwick’s Research Technology Platform (RTP Scientific Computing) and the Sulis Tier 2 HPC platform hosted by the RTP. Sulis is funded by EPSRC Grant EP/T022108/1 and the HPC Midlands+ consortium. UK data statement: Data can be accessed at the links provided in references.<sup>53,120,152,156</sup>

**Djenabou Bayo** graduated from the Sorbonne Université in 2017. Following this, she obtained a Master’s degree from CY Cergy Paris Université in 2019. As part of the Master’s program, she did an internship under the supervision of Vita Ilakovac on “Calculation of vibrations in core excited molecules” at the Laboratoire de Chimie Physique-Matière et Rayonnement (Sorbonne Université). In late 2019, she joined Römer at the University of Warwick as a co-tutelle PhD student with Honecker at CY Cergy Paris Université. She defended her PhD thesis<sup>96</sup> in April 2024.

**Burak Çivitcioglu** was born in Antalya, Türkiye. After a first degree at Koç University / Istanbul, he obtained his Master degrees from the Université Paris Diderot and CY Cergy Paris Université. In 2020, he started a PhD there with Honecker, informally co-supervised by Römer. He defended his PhD thesis<sup>165</sup> in late 2024 and has recently joined aivancity School of AI & Data for Business & Society (France), Cachan.

**Joe Webb** was born in Essex in the UK. He recently received a BSc in Mathematics and Physics (1st hon.) from the University of Warwick. There he founded and published two editions of the Poincaré student magazine and collaborated with Römer on ML applications to LACBED patterns. He is currently a scholar at Worcester College, Oxford, studying for an MSc, with plans to pursue a PhD.

**Andreas Honecker** was born in Tübingen, Germany in 1967 and obtained first his diploma and then in 1995 his PhD degree in physics from the University of Bonn under the supervision of Werner Nahm and Günter von Gehlen, respectively. He went on his post-doctoral journey to the FU Berlin, SISSA

Trieste, and ETH Zürich. After that, he worked at the TU Braunschweig where he obtained his Habilitation in 2003 under the direction of Wolfram Brenig. He continued on to the Georg-August-Universität Göttingen, with partial support via a Heisenberg fellowship, where he was awarded a professor title (apl.). He held a number of further fixed-term professorial appointments at Hannover, Strasbourg, and Lyon before he became full professor at CY Cergy Paris Université, France, in 2014. Beyond broad scientific interests in statistical physics and condensed matter theory, he also contributes to administrative obligations. In particular, he served as director of the physics department and is presently serving as adjoint director of the Institut des Sciences et Techniques of CY Cergy Paris Université. He proudly supports new publishing initiatives such as SciPost Physics as editor.

**Rudo Römer** was born in Gedern, Hessen, Germany in 1966. He attended the Wolfgang-Ernst-Gymnasium in Büdingen, studied with Robert Schrader for a Dipl.-Phys. at the FU Berlin, obtained a PhD with Bill Sutherland at the University of Utah in 1994, postdoc'ed as Feodor-Lynen fellow with Srimam Shastry at the IISc in Bangalore, Dieter Vollhardt at the RWTH Aachen, achieved his Habilitation qualification with Michael Schreiber at the TU Chemnitz in 2000 and was awarded a Heisenberg fellowship in early 2002. In late 2002 he was appointed at the University of Warwick, UK, where he heads the disordered quantum systems research group. Since then he has spent research stays at the MPIPES in Dresden, the ICCMP at UN Brasilia (now in Natal), the Chinese Academy of Sciences at the Wuhan Institute of Physics and Mathematics, was appointed Lotus (Fu Rong) Visiting Professor at Xiangtan University, Xiangtan, Hunan province, China, and as Senior Fellow at CY Cergy Paris Université. He is on a number of editorial and advisory boards and editor-in-chief for *Physica E*.

- 1) N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Saunders College Publishing, Fort Worth, 1976. <https://archive.org/details/AshcroftSolidState>.
- 2) H. T. Diep, *Frustrated Spin Systems*. World Scientific, 2020.
- 3) P. W. Anderson, "More Is Different," *Science* **177** no. 4047, (1972) 393–396.
- 4) Y.-C. Yu, Y.-Y. Chen, H.-Q. Lin, R. A. Römer, and X.-W. Guan, "Dimensionless ratios: Characteristics of quantum liquids and their phase transitions," *Phys. Rev. B* **94** no. 19, (11, 2016) 195129.
- 5) C.-Y. Wang, T.-G. Zhou, Y.-N. Zhou, and P. Zhang, "Distinguishing Quantum Phases through Cusps in Full Counting Statistics," *Phys. Rev. Lett.* **133** no. 8, (8, 2024) 083402.
- 6) H. Gomez, M. Bures, and A. Moure, "A review on computational modelling of phase-transition problems," *Phil. Trans. R. Soc. A* **377** no. 2143, (2019) 20180203.
- 7) E. Alpaydin, *Introduction to Machine Learning, fourth edition*. Adaptive Computation and Machine Learning series. MIT Press, 2020. <https://mitpress.mit.edu/9780262043793/introduction-to-machine-learning/>.
- 8) G. Hinton and T. J. Sejnowski, *Unsupervised Learning: Foundations of Neural Computation*. The MIT Press, 8, 1999.
- 9) P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, "A high-bias, low-variance introduction to Machine Learning for physicists," *Phys. Rep.* **810** (5, 2019) 1–124.
- 10) G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, "Machine learning and the physical sciences," *Rev. Mod. Phys.* **91** no. 4, (12, 2019) 045002.
- 11) I. Schurov, D. Alforov, M. Katsnelson, A. Bagrov, and A. Itin, "Invariant multiscale neural networks for data-scarce scientific applications". <https://arxiv.org/abs/2406.08318v1>.
- 12) K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-Decem* (2016) 770–778.
- 13) M. A. Tabak, M. S. Norouzzadeh, *et al.*, "Machine learning to classify animal species in camera trap images: Applications in ecology," *Methods in Ecology and Evolution* **10** no. 4, (2019) 585–590.
- 14) R. Zhang, Q. Wang, and Y. Lu, "Combination of ResNet and Center Loss Based Metric Learning for Handwritten Chinese Character Recognition," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, pp. 25–29. IEEE, 2017.
- 15) Y. Zhao and M. Khushi, "Wavelet Denoised-ResNet CNN and Light-GBM Method to Predict Forex Rate of Change," in *2020 International Conference on Data Mining Workshops (ICDMW)*, vol. 2020-Novem, pp. 385–391. IEEE, 11, 2020.
- 16) J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," *Nat. Phys.* **13** no. 5, (2017) 431–434.
- 17) K. Ch'ng, N. Vazquez, and E. Khatami, "Unsupervised machine learning account of magnetic transitions in the Hubbard model," *Phys. Rev. E* **97** no. 1, (1, 2018) 013306.
- 18) A. Tanaka and A. Tomiya, "Detection of Phase Transition via Convolutional Neural Networks," *J. Phys. Soc. Jpn.* **86** no. 6, (6, 2017) 63001.
- 19) P. Huembeli, A. Dauphin, P. Wittek, and C. Gogolin, "Automated discovery of characteristic features of phase transitions in many-body localization," *Phys. Rev. B* **99** no. 10, (3, 2019) 104106.
- 20) X.-Y. Dong, F. Pollmann, and X.-F. Zhang, "Machine learning of quantum phase transitions," *Phys. Rev. B* **99** no. 12, (3, 2019) 121104.
- 21) A. Canabarro, F. F. Fanchini, A. L. Malvezzi, R. Pereira, and R. Chaves, "Unveiling phase transitions with machine learning," *Phys. Rev. B* **100** no. 4, (7, 2019) 045129.
- 22) K. Shinjo, K. Sasaki, S. Hase, S. Sota, S. Ejima, S. Yunoki, and T. Tohyama, "Machine Learning Phase Diagram in the Half-filled One-dimensional Extended Hubbard Model," *J. Phys. Soc. Jpn.* **88** no. 6, (6, 2019) 065001.
- 23) L. Wang, "Discovering phase transitions with unsupervised learning," *Phys. Rev. B* **94** no. 19, (11, 2016) 195105.
- 24) K. Kottmann, P. Huembeli, M. Lewenstein, and A. Acín, "Unsupervised Phase Discovery with Deep Anomaly Detection," *Phys. Rev. Lett.* **125** no. 17, (10, 2020) 170603.
- 25) C. Alexandrou, A. Athenodorou, C. Chrysostomou, and S. Paul, "The critical temperature of the 2D-Ising model through deep learning autoencoders," *Eur. Phys. J. B* **93** no. 12, (12, 2020) 226.
- 26) F. D'Angelo and L. Böttcher, "Learning the Ising model with generative neural networks," *Phys. Rev. Res.* **2** no. 2, (6, 2020) 023266.
- 27) K. Shiina, H. Mori, Y. Okabe, and H. K. Lee, "Machine-Learning Studies on Spin Models," *Sci. Rep.* **10** no. 1, (2, 2020) 2177.
- 28) I. Corte, S. Acevedo, M. Arlego, and C. A. Lamas, "Exploring neural network training strategies to determine phase transitions in frustrated magnetic models," *Comput. Mater. Sci.* **198** (2021) 110702.
- 29) N. Walker, K.-M. Tam, B. Novak, and M. Jarrell, "Identifying structural changes with unsupervised machine learning methods," *Phys. Rev. E* **98** (11, 2018) 053305.
- 30) A. Morningstar and R. G. Melko, "Deep Learning the Ising Model Near Criticality," *J. Mach. Learn. Res.* **18** no. 163, (2018) 1–17. <http://jmlr.org/papers/v18/17-527.html>.
- 31) P. Suchsland and S. Wessel, "Parameter diagnostics of phases and phase transition learning by neural networks," *Phys. Rev. B* **97** no. 17, (2018) 1–13.
- 32) S. Efthymiou, M. J. S. Beach, and R. G. Melko, "Super-resolving the Ising model with convolutional neural networks," *Phys. Rev. B* **99** no. 7, (2, 2019) 075113.
- 33) N. Walker, K.-M. Tam, and M. Jarrell, "Deep learning on the 2-dimensional Ising model to extract the crossover region with a variational autoencoder," *Sci. Rep.* **10** no. 1, (8, 2020) 13047.
- 34) S. Goel, "Learning Ising and Potts Models with Latent Variables," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, S. Chiappa and R. Calandra, eds., vol. 108 of *Proceedings of Machine Learning Research*, pp. 3557–3566. PMLR, 8, 2020. <https://proceedings.mlr.press/v108/goel20a.html>.
- 35) J. Wang, W. Zhang, T. Hua, and T.-C. Wei, "Unsupervised learning of topological phase transitions using the Calinski-Harabaz index," *Phys. Rev. Res.* **3** (Jan, 2021) 013074.
- 36) J. Arnold, F. Schäfer, M. Žonda, and A. U. J. Lode, "Interpretable and unsupervised phase classification," *Phys. Rev. Res.* **3** (Jul, 2021) 033052.
- 37) B. Çivitcioğlu, R. A. Römer, and A. Honecker, "Machine Learning the Square-Lattice Ising Model," *J. Phys. Conf. Ser.* **2207** no. 1, (3, 2022) 12058.
- 38) P. Basu, J. Bhattacharya, D. P. S. Jakka, C. Mosomane, and V. Shukla, "Machine learning of Ising criticality with spin-shuffling". <http://arxiv.org/abs/2203.04012>.

- 39) D. W. Tola and M. Bekele, "Machine learning of nonequilibrium phase transition in an Ising model on square lattice," *Condensed Matter* **8** no. 3, (2023) 83.
- 40) V. Chertakov, E. Burovski, and L. Shchur, "Finite-size analysis in neural network classification of critical phenomena," *Phys. Rev. E* **108** (Sep, 2023) L032102.
- 41) A. Naravane and N. Mathur, "Semi-supervised learning of order parameter in 2D Ising and XY models using Conditional Variational Autoencoders," <http://arxiv.org/abs/2306.16822>.
- 42) G. L. G. Pavioni, M. Arlego, and C. A. Lamas, "Minimalist neural networks training for phase classification in diluted Ising models," *Comput. Mater. Sci.* **235** (2024) 112792.
- 43) B. M. McCoy and T. T. Wu, *The Two-Dimensional Ising Model*. Harvard University Press, 12, 1973.
- 44) W. Rządowski, N. Defenu, S. Chiacchiera, A. Trombettoni, and G. Bighin, "Detecting composite orders in layered models via machine learning," *New J. Phys.* **22** no. 9, (9, 2020) 93026.
- 45) K. Fukushima and K. Sakai, "Can a CNN trained on the Ising model detect the phase transition of the q-state Potts model?," *Prog. Theor. Exp. Phys.* **2021** no. 6, (8, 2021) 061A01.
- 46) D. Giataganas, C.-Y. Huang, and F.-L. Lin, "Neural network flows of low q-state Potts and clock models," *New J. Phys.* **24** no. 4, (4, 2022) 43040.
- 47) A. Tirelli, D. O. Carvalho, L. A. Oliveira, J. P. de Lima, N. C. Costa, and R. R. dos Santos, "Unsupervised machine learning approaches to the q-state Potts model," *Eur. Phys. J. B* **95** no. 11, (11, 2022) 189.
- 48) F. Y. Wu, "The Potts model," *Rev. Mod. Phys.* **54** no. 1, (1, 1982) 235–268.
- 49) W. Zhang, J. Liu, and T.-C. Wei, "Machine learning of phase transitions in the percolation and XY models," *Phys. Rev. E* **99** no. 3, (3, 2019) 032142.
- 50) W. Yu and P. Lyu, "Unsupervised machine learning of phase transition in percolation," *Physica A* **559** (12, 2020) 125065.
- 51) J. Shen, W. Li, S. Deng, and T. Zhang, "Supervised and unsupervised learning of directed percolation," *Phys. Rev. E* **103** no. 5, (2021) 052140.
- 52) D. Bayo, A. Honecker, and R. A. Römer, "Machine learning the 2D percolation model," *J. Phys.: Conf. Ser.* **2207** no. 1, (2022) 12057.
- 53) D. Bayo, A. Honecker, and R. A. Römer, "The percolating cluster is invisible to image recognition with deep learning," *New J. Phys.* **25** no. 11, (11, 2023) 113041.
- 54) S. Patwardhan, U. Majumder, A. D. Sarma, M. Pal, D. Dwivedi, and P. K. Panigrahi, "Machine Learning as an Accurate Predictor for Percolation Threshold of Diverse Networks," <http://arxiv.org/abs/2212.14694>.
- 55) S. Cheng, F. He, H. Zhang, K.-D. Zhu, and Y. Shi, "Machine Learning Percolation Model," <http://arxiv.org/abs/2101.08928>
- 56) J. Li, R. L. Chu, J. K. Jain, and S. Q. Shen, "Topological Anderson insulator," *Phys. Rev. Lett.* **102** no. 13, (3, 2009) 136806.
- 57) Z. Balogh and G. Schmitz, "Diffusion in Metals and Alloys," *Physical Metallurgy: Fifth Edition 1* (1, 2014) 387–559.
- 58) J. Oswald and R. A. Römer, "Imaging of Condensed Quantum States in the Quantum Hall Effect Regime," *Physics Procedia* **75** (2015) 314–325.
- 59) J. Oswald and R. A. Römer, "Microscopic details of stripes and bubbles in the quantum Hall regime," *Phys. Rev. B* **102** no. 12, (9, 2020) 121305(R).
- 60) F. D. Haldane, "Model for a Quantum Hall Effect without Landau Levels: Condensed-Matter Realization of the "Parity Anomaly"," *Phys. Rev. Lett.* **61** no. 18, (10, 1988) 2015.
- 61) R. Yu, W. Zhang, H. J. Zhang, S. C. Zhang, X. Dai, and Z. Fang, "Quantized anomalous Hall effect in magnetic topological insulators," *Science* **329** no. 5987, (7, 2010) 61–64.
- 62) J. Pixley, P. Goswami, and S. Das Sarma, "Anderson Localization and the Quantum Phase Diagram of Three Dimensional Disordered Dirac Semimetals," *Phys. Rev. Lett.* **115** no. 7, (2015) 076601.
- 63) L. Meng, J. Wu, J. Zhong, and R. A. Römer, "A type of robust superlattice type-I Weyl semimetal with four Weyl nodes," *Nanoscale* **11** no. 39, (2019) 18358–18366.
- 64) E. Prodan, H. Schulz-Baldes, "Bulk and Boundary Invariants for Complex Topological Insulators: From K-Theory to Physics," *Mathematical Physics Studies Part F1117*, 1–201 (2015).
- 65) M. Z. Hasan and C. L. Kane, "Topological Insulators," *Physics* **82** no. 4, (2010) 23.
- 66) E. Rotenberg, "Topological insulators: The dirt on topology," *Nat. Phys.* **7** no. 1, (2011) 8–10. <http://dx.doi.org/10.1038/nphys1869>.
- 67) A. Rodriguez, L. J. Vasquez, K. Slevin, and R. A. Römer, "Critical Parameters from a Generalized Multifractal Analysis at the Anderson Transition," *Phys. Rev. Lett.* **105** no. 4, (7, 2010) 046403.
- 68) A. Rodriguez, L. J. Vasquez, K. Slevin, and R. A. Römer, "Multifractal finite-size scaling and universality at the Anderson transition," *Phys. Rev. B* **84** no. 13, (10, 2011) 134209.
- 69) T. Ohtsuki and T. Ohtsuki, "Deep Learning the Quantum Phase Transitions in Random Two-Dimensional Electron Systems," *J. Phys. Soc. Jpn.* **85** no. 12, (12, 2016) 123706.
- 70) T. Ohtsuki and T. Mano, "Drawing Phase Diagrams of Random Quantum Systems by Deep Learning the Wave Functions," *J. Phys. Soc. Jpn.* **89** no. 2, (2020) 022001.
- 71) A. Wichert and L. Sa-Couto, *Machine Learning — A Journey to Deep Learning*. World Scientific, 2, 2021.
- 72) P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017-January* (11, 2017) 5967–5976.
- 73) T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (12, 2018) 8798–8807.
- 74) D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," <http://arxiv.org/abs/1312.6114>.
- 75) J. M. Ede, *Advances in Electron Microscopy with Deep Learning*, PhD thesis, University of Warwick (2021). <http://arxiv.org/abs/2101.01178>.
- 76) S. Matinyan, B. Demir, P. Filipcik, J. P. Abrahams, and E. van Genderen, "Machine learning for classifying narrow-beam electron diffraction data," *Acta Crystallographica Section A Foundations and Advances* **79** no. 4, (7, 2023) 360–368.
- 77) T. Friedrich, C.-P. Yu, J. Verbeeck, and S. Van Aert, "Phase Object Reconstruction for 4D-STEM using Deep Learning," *Microsc. Microanal.* **29** no. 1, (2, 2023) 395–407.
- 78) X. Liu, K. Amimi, A. Sanchez, B. Belsa, T. Steinle, and J. Biegert, "Machine learning for laser-induced electron diffraction imaging of molecular structures," *Commun. Chem.* **4** no. 1, (11, 2021) 154.
- 79) S. R. Hall, J. D. Westbrook, N. Spadaccini, I. D. Brown, H. J. Bernstein, and B. McMahon, "Specification of the Crystallographic Information File (CIF)," in *International Tables for Crystallography*, pp. 20–36. International Union of Crystallography, Chester, England, 10, 2006.
- 80) Igor Levin, "NIST Inorganic Crystal Structure Database (ICSD)," 2018. <https://icsd.nist.gov/>
- 81) J. Chou, "Generated Loss and Augmented Training of MNIST VAE," <http://arxiv.org/abs/1904.10937>.
- 82) Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- 83) I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, eds. Curran Associates, Inc., 27 ed., 6, 2014.
- 84) S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Statist.* **22** no. 1, (3, 1951) 79–86.
- 85) D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544. IEEE, 6, 2016.
- 86) M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," <http://arxiv.org/abs/1411.1784>.
- 87) S. R. Broadbent and J. M. Hammersley, "Percolation processes: I. Crystals and mazes," *Math. Proc. Camb. Philos. Soc.* **53** no. 3, (1957) 629–641.
- 88) D. Stauffer and A. Aharony, *Introduction To Percolation Theory*. Taylor & Francis, 12, 2018.
- 89) R. J. Elliott, B. R. Heap, D. J. Morgan, and G. S. Rushbrooke, "Equivalence of the Critical Concentrations in the Ising and Heisenberg Mod-

- els of Ferromagnetism,” *Phys. Rev. Lett.* **5** no. 8, (10, 1960) 366–367.
- 90) P. J. Flory, *Principles of Polymer Chemistry*. Cornell University Press, 1953. [https://books.google.fr/books/about/Principles\\_of\\_Polymer\\_Chemistry.html?id=CQ0EbEkT5R0C&redir\\_esc=y](https://books.google.fr/books/about/Principles_of_Polymer_Chemistry.html?id=CQ0EbEkT5R0C&redir_esc=y)
  - 91) B. Derrida and D. Stauffer, “Corrections To Scaling and Phenomenological Renormalization for 2-Dimensional Percolation and Lattice Animal Problems,” *J. Phys. (Paris)* **46** no. 10, (1985) 1623–1630.
  - 92) G. Grimmett, *Percolation*. Springer Verlag, Berlin, 1989. <https://link.springer.com/book/10.1007/978-3-662-03981-6>.
  - 93) J. L. Jacobsen, “High-precision percolation thresholds and Potts-model critical manifolds from graph polynomials,” *J. Phys. A: Math. Theor.* **47** no. 13, (2014) 135001.
  - 94) J. Hoshen and R. Kopelman, “Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm,” *Phys. Rev. B* **14** no. 8, (10, 1976) 3438–3445.
  - 95) A. Paszke, S. Gross, et al., “PyTorch: An imperative style, high-performance deep learning library,” *Advances in Neural Information Processing Systems* **32** no. NeurIPS, (2019). <http://arxiv.org/abs/1912.01703>.
  - 96) D. Bayo, *Phase Determination by Deep Learning for Disordered Systems*. PhD thesis, University of Warwick (home) & CY Cergy Paris Université (host), 2024. <https://theses.fr/2024CYUN1280>.
  - 97) O. Schenk, M. Bollhöfer, and R. A. Römer, “On Large-Scale Diagonalization Techniques for the Anderson Model of Localization,” *SIAM Review* **50** no. 1, (1, 2008) 91–112.
  - 98) P. W. Anderson, “Absence of Diffusion in Certain Random Lattices,” *Phys. Rev.* **109** no. 5, (3, 1958) 1492–1505.
  - 99) U. Elsner, V. Mehrmann, F. Milde, R. A. Römer, and M. Schreiber, “The Anderson Model of Localization: A Challenge for Modern Eigenvalue Methods,” *SIAM Journal on Scientific Computing* **20** no. 6, (1, 1999) 2089–2102.
  - 100) T. Brandes and S. Kettmann, *Anderson Localization and Its Ramifications*, vol. 630 of *Lecture Notes in Physics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. <http://link.springer.com/10.1007/b13139>.
  - 101) F. Evers and A. D. Mirlin, “Anderson transitions,” *Rev. Mod. Phys.* **80** no. 4, (2008) 1355–1417. <http://link.aps.org/doi/10.1103/RevModPhys.80.1355>.
  - 102) H. Théveniaut and F. Alet, “Neural network setups for a precise detection of the many-body localization transition: finite-size scaling and limitations,” *Phys. Rev. B* **100** no. 22, (2019) 224202.
  - 103) T. Mano and T. Ohtsuki, “Phase diagrams of three-dimensional Anderson and quantum percolation models using deep three-dimensional convolutional neural network,” *J. Phys. Soc. Jpn.* **86** no. 11, (2017) 113704.
  - 104) T. Čadež, B. Dietz, D. Rosa, A. Andreev, K. Slevin, and T. Ohtsuki, “Machine learning wave functions to identify fractal phases,” *Phys. Rev. B* **108** no. 18, (2023) 184202.
  - 105) R. A. Römer, “Numerical methods for localization,” in *Reference Module in Materials Science and Materials Engineering*. Elsevier, 7, 2022.
  - 106) M. Bollhöfer and Y. Notay, “JADAMILU: a software code for computing selected eigenvalues of large sparse symmetric matrices,” *Comput. Phys. Commun.* **177** no. 12, (2007) 951–964.
  - 107) R. H. Swendsen and S. Krinsky, “Monte Carlo Renormalization Group and Ising Models with  $n \geq 2$ ,” *Phys. Rev. Lett.* **43** no. 3, (7, 1979) 177–180.
  - 108) D. P. Landau, “Phase transitions in the Ising square lattice with next-nearest-neighbor interactions,” *Phys. Rev. B* **21** no. 3, (2, 1980) 1285–1297.
  - 109) K. Binder and D. P. Landau, “Phase diagrams and critical behavior in Ising square lattices with nearest- and next-nearest-neighbor interactions,” *Phys. Rev. B* **21** no. 5, (3, 1980) 1941–1962.
  - 110) D. P. Landau and K. Binder, “Phase diagrams and critical behavior of Ising square lattices with nearest-, next-nearest-, and third-nearest-neighbor couplings,” *Phys. Rev. B* **31** no. 9, (5, 1985) 5946–5953.
  - 111) J. L. Morán-López, F. Aguilera-Granja, and J. M. Sanchez, “First-order phase transitions in the Ising square lattice with first- and second-neighbor interactions,” *Phys. Rev. B* **48** no. 5, (8, 1993) 3519–3522.
  - 112) A. Malakis, P. Kalozoumis, and N. Tyraskis, “Monte Carlo studies of the square Ising model with next-nearest-neighbor interactions,” *Eur. Phys. J. B* **50** no. 1, (2, 2006) 63–67.
  - 113) J. L. Monroe and S.-Y. Kim, “Phase diagram and critical exponent  $\nu$  for the nearest-neighbor and next-nearest-neighbor interaction Ising model,” *Phys. Rev. E* **76** no. 2, (8, 2007) 021123.
  - 114) R. A. dos Anjos, J. Roberto Viana, and J. Ricardo de Sousa, “Phase diagram of the Ising antiferromagnet with nearest-neighbor and next-nearest-neighbor interactions on a square lattice,” *Phys. Lett. A* **372** no. 8, (2008) 1180–1184.
  - 115) A. Kalz, A. Honecker, S. Fuchs, and T. Pruschke, “Phase diagram of the Ising square lattice with competing interactions,” *Eur. Phys. J. B* **65** no. 4, (10, 2008) 533.
  - 116) H. Watanabe, Y. Motoyama, S. Morita, and N. Kawashima, “Non-monotonic behavior of the Binder parameter in discrete spin systems,” *Prog. Theor. Exp. Phys.* **2023** no. 3, (8, 2023) 033A02.
  - 117) K. Yoshiyama and K. Hukushima, “Higher-order tensor renormalization group study of the  $J_1$ - $J_2$  Ising model on a square lattice,” *Phys. Rev. E* **108** no. 5, (11, 2023) 054124.
  - 118) S.-W. Li and F.-J. Jiang, “A Comprehensive Study of the Phase Transitions of the Frustrated  $J_1$ - $J_2$  Ising Model on the Square Lattice,” *Prog. Theor. Exp. Phys.* **2024** no. 5, (8, 2024) 053A06.
  - 119) H. Park and H. Lee, “Frustrated Ising Model on D-wave Quantum Annealing Machine,” *J. Phys. Soc. Jpn.* **91** no. 7, (6, 2022) .
  - 120) B. Çivitiçioğlu, R. A. Römer, and A. Honecker, “Phase determination with and without deep learning,” <http://arxiv.org/abs/2403.09786>.
  - 121) N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines,” *J. Chem. Phys.* **21** no. 6, (1953) 1087–1092.
  - 122) M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics*. Oxford University Press, 1999. <https://global.oup.com/academic/product/monte-carlo-methods-in-statistical-physics-9780198517979>.
  - 123) B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis*. World Scientific, 10, 2004.
  - 124) D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 4 ed., 2014.
  - 125) S. Acevedo, M. Arlego, and C. A. Lamas, “Phase diagram study of a two-dimensional frustrated antiferromagnet via unsupervised machine learning,” *Phys. Rev. B* **103** no. 13, (2021) 134422.
  - 126) H. A. Kramers and G. H. Wannier, “Statistics of the two-dimensional ferromagnet. part I,” *Phys. Rev.* **60** (Aug, 1941) 252–262.
  - 127) J. C. H. Spence and J. M. Zuo, *Electron Microdiffraction*. Springer US, Boston, MA, 1992.
  - 128) M. Tanaka, “Convergent-beam electron diffraction,” *Acta Crystallogr. A* **50** no. 3, (5, 1994) 261–286.
  - 129) R. Beaman, P. J. Thomas, D. I. Woodward, P. A. Thomas, and R. A. Roemer, “Digital electron diffraction – seeing the whole picture,” *Acta Crystallogr. A* **69** no. 4, (7, 2013) 427–434.
  - 130) W. Kossel and G. Möllenstedt, “Elektroneninterferenzen im konvergenten Bündel,” *Annalen der Physik* **428** no. 2, (1, 1939) 113–140.
  - 131) B. F. Buxton, J. A. Eades, J. W. Steeds, and G. M. Rackham, “The symmetry of electron diffraction zone axis patterns,” *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **281** no. 1301, (3, 1976) 171–194.
  - 132) M. Tanaka, R. Saito, and H. Sekii, “Point-group determination by convergent-beam electron diffraction,” *Acta Crystallogr. A* **39** no. 3, (5, 1983) 357–368.
  - 133) M. Tanaka, H. Sekii, and T. Nagasawa, “Space-group determination by dynamic extinction in convergent-beam electron diffraction,” *Acta Crystallogr. A* **39** no. 6, (11, 1983) 825–837.
  - 134) M. Saunders, A. G. Fox, and P. A. Midgley, “Quantitative zone-axis convergent-beam electron diffraction (CBED) studies of metals. II. Debye-Waller-factor measurements,” *Acta Crystallogr. A* **55** no. 3, (5, 1999) 480–488.
  - 135) J. M. Zuo, M. Kim, and R. Holmestad, “A new approach to lattice parameter measurements using dynamic electron diffraction and pattern matching,” *J. Electron Microsc. Tech.* **47** no. 2, (1, 1998) 121–127.
  - 136) U. Kaiser, K. Saitoh, K. Tsuda, and M. Tanaka, “Application of the CBED method for the determination of lattice parameters of cubic SiC films on 6H SiC substrates,” *J. Electron Microsc. Tech.* **48** no. 3, (1, 1999) 221–233.
  - 137) A. Armigliato, R. Balboni, G. P. Carnevale, G. Pavia, D. Piccolo, S. Frabboni, A. Benedetti, and A. G. Cullis, “Application of convergent beam electron diffraction to two-dimensional strain mapping in

- silicon devices,” *Appl. Phys. Lett.* **82** no. 13, (3, 2003) 2172–2174.
- 138) S. Krämer, J. Mayer, C. Witt, A. Weickenmeier, and M. Rühle, “Analysis of local strain in aluminium interconnects by energy filtered CBED,” *Ultramicroscopy* **81** no. 3-4, (4, 2000) 245–262.
- 139) D. Cherns and A. R. Preston, “Convergent beam diffraction studies of interfaces, defects, and multilayers,” *J. Electron Microsc. Tech.* **13** no. 2, (10, 1989) 111–122.
- 140) J. Morniroli and D. Cherns, “Analysis of grain boundary dislocations by large angle convergent beam electron diffraction,” *Ultramicroscopy* **62** no. 1-2, (1, 1996) 53–63.
- 141) P. A. Midgley and M. Saunders, “Quantitative electron diffraction: From atoms to bonds,” *Contemp. Phys.* **37** no. 6, (11, 1996) 441–456.
- 142) J. M. Zuo, M. Kim, M. O’Keeffe, and J. C. H. Spence, “Direct observation of d-orbital holes and Cu–Cu bonding in  $\text{Cu}_2\text{O}$ ,” *Nature* **401** no. 6748, (9, 1999) 49–52.
- 143) R. Beanland, A. Hubert, and R. Roemer, “Refinement of crystal structure using ‘digital’ large angle convergent beam electron diffraction,” *Microsc. Microanal.* **27** no. S1, (8, 2021) 1282–1284.
- 144) A. J. M. Hubert, R. Römer, and R. Beanland, “Structure refinement from ‘digital’ large angle convergent beam electron diffraction patterns,” *Ultramicroscopy* **198** (3, 2019) 1–9.
- 145) J. Zuo and A. Weickenmeier, “On the beam selection and convergence in the Bloch-wave method,” *Ultramicroscopy* **57** no. 4, (3, 1995) 375–383.
- 146) B. G. Mendis, ed., *Electron Beam-Specimen Interactions and Simulation Methods in Microscopy*. Wiley, 5, 2018.
- 147) J. M. Cowley and A. F. Moodie, “The scattering of electrons by atoms and crystals. I. A new theoretical approach,” *Acta Crystallographica* **10** no. 10, (10, 1957) 609–619.
- 148) D. Van Dyck and W. Coene, “The real space method for dynamical electron diffraction calculations in high resolution electron microscopy,” *Ultramicroscopy* **15** no. 1-2, (1, 1984) 29–40.
- 149) A. Chuvilin and U. Kaiser, “On the peculiarities of CBED pattern formation revealed by multislice simulation,” *Ultramicroscopy* **104** no. 1, (8, 2005) 73–82.
- 150) U. Kaiser and A. Chuvilin, “Prospects of the multislice method for CBED pattern calculation,” *Int. J. Mater. Res.* **97** no. 7, (7, 2006) 912–919.
- 151) E. J. Kirkland, *Advanced Computing in Electron Microscopy*. Springer US, Boston, MA, 2010.
- 152) R. Beanland, K. Evans, and R. A. Roemer, “Felix: Bloch wave method diffraction pattern simulation software.” 2023. <https://github.com/WarwickMicroscopy/Felix>.
- 153) “Kaggle: Your Home for Data Science.” 2023. <https://www.kaggle.com/>.
- 154) A. Halevy, P. Norvig, and F. Pereira, “The Unreasonable Effectiveness of Data,” *IEEE Intelligent Systems* **24** no. 2, (3, 2009) 8–12.
- 155) I. Spasic and G. Nenadic, “Clinical Text Data in Machine Learning: Systematic Review,” *JMIR Medical Informatics* **8** no. 3, (3, 2020) e17984.
- 156) J. J. Webb, “GitHub - wephy/ai-diffraction.” 2, 2024. <https://github.com/wephy/ai-diffraction>.
- 157) J. J. Webb and R. Beanland, “Felix Diffraction Patterns.” V2, 2023. <https://www.kaggle.com/datasets/wephy/felix-diffraction-patterns>.
- 158) M. Saini and S. Susan, “Tackling class imbalance in computer vision: a contemporary review,” *Artif. Intell. Rev.* **56** no. S1, (10, 2023) 1279–1335.
- 159) S. F. Edwards and P. W. Anderson, “Theory of spin glasses,” *J. Phys. F: Met. Phys.* **5** no. 5, (5, 1975) 965.
- 160) G. Carleo and M. Troyer, “Solving the quantum many-body problem with artificial neural networks,” *Science* **355** no. 6325, (2, 2017) 602–606.
- 161) G. Carleo, K. Choo, D. Hofmann, J.E.T. Smith, T. Westerhoute, F. Alet, E.J. Davis, S. Efthymiou, I. Glasser, S-H. Lin, M. Mauria, G. Mazzola, C.B. Mendl, E. van Nieuwenburg, O. O’Reilly, H. Théveniaut, G. Torlai, F. Vicentini, A. Wietek, “NetKet: A machine learning toolkit for many-body quantum systems,” *SoftwareX* **10** (2019) 100311.
- 162) T. Szoldra, P. Sierant, K. Kottmann, M. Lewenstein, and J. Zakrzewski, “Detecting ergodic bubbles at the crossover to many-body localization using neural networks,” *Phys. Rev. B* **104** no. 14, (10, 2021) L140202.
- 163) Z. Zhu, M. Mattheakis, W. Pan, and E. Kaxiras, “HubbardNet: Efficient predictions of the Bose-Hubbard model spectrum with deep neural networks,” *Phys. Rev. Res.* **5** no. 4, (10, 2023) 043084.
- 164) D. Pfau, S. Axelrod, H. Sutterud, I. v. Glehn, and J. S. Spencer, “Accurate computation of quantum excited states with neural networks,” *Science* **385** no. 6711, (8, 2024) .
- 165) B. Çivıtcioğlu, *Phase Determination with and without Deep Learning for the  $J_1$ - $J_2$  Ising Model*. PhD thesis, CY Cergy Paris Université, 2024.