

# Distinguishing an Anderson insulator from a many-body localized phase through space-time snapshots with neural networks

Florian Kotthoff<sup>1</sup>, Frank Pollmann<sup>1,2</sup> and Giuseppe De Tomasi<sup>3,4</sup>

<sup>1</sup>*Department of Physics, Technische Universität München, 85747 Garching, Germany*

<sup>2</sup>*Munich Center for Quantum Science and Technology (MCQST), Schellingstrasse 4, D-80799 München, Germany*

<sup>3</sup>*Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom*

<sup>4</sup>*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801-3080, USA*

(Received 25 August 2021; accepted 6 December 2021; published 27 December 2021)

Distinguishing the dynamics of an Anderson insulator from a many-body localized (MBL) phase is an experimentally challenging task. In this work we propose a method based on machine learning techniques to analyze experimental snapshot data to separate the two phases. We show how to train three-dimensional convolutional neural networks (CNNs) using space-time Fock-state snapshots, allowing us to obtain dynamic information about the system. We benchmark our method on a paradigmatic model showing MBL ( $t$ - $V$  model with quenched disorder), where we obtain a classification accuracy of  $\approx 80\%$  between an Anderson insulator and an MBL phase. We underline the importance of providing temporal information to the CNNs and we show that CNNs learn the crucial difference between an Anderson localized and an MBL phase, namely the difference in the propagation of quantum correlations. Particularly, we show that the misclassified MBL samples are characterized by an unusually slow propagation of quantum correlations, and thus the CNNs label them wrongly as Anderson localized. Finally, we apply our method to the case with quasiperiodic potential, known as the Aubry-André model (AA model). We find that the CNNs have more difficulties in separating the two phases. We show that these difficulties are due to the fact that the MBL phase of the AA model is characterized by a slower information propagation for numerically accessible system sizes.

DOI: [10.1103/PhysRevB.104.224307](https://doi.org/10.1103/PhysRevB.104.224307)

## I. INTRODUCTION

Advancements of controlled experimental techniques, such as ultracold atoms in optical lattices, trapped ions, and superconducting q-bits have led to considerable interest in the out-of-equilibrium dynamics of isolated quantum many-body systems [1–5]. In particular, it became possible to provide experimental evidence of many-body localization (MBL) [6–8], and therefore to shed light upon the emergence of the laws of statistical mechanics in the quantum realm.

MBL generalizes Anderson localization to the interacting case and has emerged as a novel paradigm for ergodicity breaking of generic many-body systems subjected to strong disorder [9–13]. An MBL phase is best understood in terms of an emergent form of integrability, meaning that the system is fully described by an extensive number of quasilocal integrals of motion (LIOMs), which are adiabatically connected to the noninteracting ones. As a result, due to the local nature of the LIOMs, transport in the system is strongly hindered [14–17]. Furthermore, unlike an Anderson insulator, interactions weakly couple the LIOMs, producing a dephasing mechanism that yields a slow logarithmic spread of entanglement [18–20]. Among the propagation of entanglement entropy, other dynamical indicators have been proposed to distinguish an Anderson insulator from an MBL phase, i.e., propagation of quantum mutual information [21], quantum Fisher information [7,14], temporal fluctuations [14,22], or spin noise spectroscopy [23,24]. However, it remains a challenge to reliably distinguish the two phases

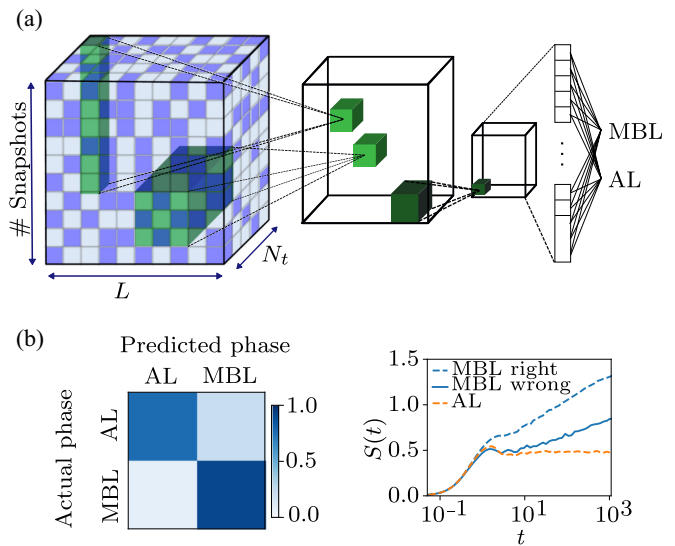


FIG. 1. (a) Three-dimensional (3D) structure of the input data ( $\# \text{Snapshots} \times L \times N_t$ ) with a pictorial representation of the CNN used to distinguish an MBL phase from an Anderson localized one. (b) The left panel shows the confusion matrix for the classification task. The CNN classifies 83.9% of AL and MBL samples correctly. The right panel shows the entanglement entropy  $S(t)$  averaged over MBL samples ( $V = 1$ ), right and wrong classified separately, and  $S(t)$  for the noninteracting case ( $V = 0$ ). As one can notice for the wrongly classified MBL samples,  $S(t)$  has an unusual slower growth of the entanglement.

experimentally. These difficulties come mainly from the fact that the distinction requires the measurement of nonlocal quantum correlations.

The aim of this work is to show how to use machine learning toolboxes and, in particular, convolutional neural networks (CNNs) to analyze experimental snapshot data and distinguish an MBL phase from an Anderson insulator.

Machine learning techniques have proven to be a useful tool in characterizing and understanding correlations in quantum phases of matter [25–38]. Recently, several works used machine learning to investigate the MBL transition which separates a thermal phase from the localized one [35,39–47]. In particular, in Ref. [48] a CNN was trained using experimentally Fock-space snapshots data to distinguish an MBL from a thermal phase. However, the distinction of an Anderson insular from an MBL phase is more subtle than the separation of an ergodic/extended from an MBL phase.

In both, the Anderson insulator and MBL phase, degrees of freedom are frozen and the only difference is found in the propagation of quantum correlations, which are harder to be measured. By using *space-time* Fock-space snapshots as input data, we show that CNNs are able to capture the important spatial and temporal correlations to distinguish the two phases. This approach should be opposed to the ones used to separate an ergodic from an MBL phase [48], in which only Fock-space snapshots at a single time are sufficient to distinguish the two phases.

Concretely, in quantum simulation platforms such as ultracold atoms in optical lattices, Fock-space snapshots are accessible following a time evolution with the use of a quantum microscope [49,50]. The computation of local observables at a specific target time is then found by a proper average over the ensemble of snapshots. Here we show how to construct and train a CNN (see Fig. 1) to extract dynamic properties and therefore to distinguish the two phases. Importantly, we provide numerical evidence that only a reasonable amount of snapshots is needed to separate the two phases, which bounds the number of experimental measurements. Moreover, we show that the MBL samples that are wrongly classified are characterized by an atypical slow propagation of information, i.e., entanglement entropy and particle number fluctuation, see Fig. 1(b). This supports the idea that the CNN learns the important aforementioned features to distinguish an MBL phase from an Anderson localized one.

This work is structured as follows. In Sec. II we introduce the utilized models. In Sec. III we explain the structure of the neural network and the generation of the snapshots. Section IV is dedicated to examining the network performance on classifying an Anderson insulator and an MBL phase. In particular, we show that it is necessary to consider snapshots from different points in time and thus gain dynamic information to improve the classification accuracy. In Sec. V we test the stability of our neural network by tuning the chain length and interaction strength of the input data. The observed stability of the trained neural network is an important component of our work. It provides an indication that our method can be applied to real experimental data. With the aim to understand the high performance of our network, in Sec. VI we show that

the wrongly classified MBL samples are characterized by an unusually slow growth of entanglement, see Fig. 1(b). Thus, our CNNs are able to detect the important features that distinguish an Anderson insulator from an MBL phase. Finally, we apply our method to the case in which the disorder is generated by a quasiperiodic potential (Aubry-André model), which is particularly relevant for experiments [6,51–55].

## II. MODEL

We study the  $t$ - $V$  disordered spinless fermionic chain with periodic boundary conditions

$$H = -\frac{t}{2} \sum_{j=1}^L c_j^\dagger c_{j+1} + \text{H.c.} + \sum_{j=1}^L h_j \left( n_j - \frac{1}{2} \right) + V \sum_{j=1}^L \left( n_j - \frac{1}{2} \right) \left( n_{j+1} - \frac{1}{2} \right), \quad (1)$$

where  $c_j^\dagger$  ( $c_j$ ) is the fermionic creation (annihilation) operator at site  $j$ .  $t = 1$  and  $V$  are the hopping and the interaction strength, respectively, and  $\{h_j\}$  are random fields which are uniformly distributed between  $[-W, W]$ .  $L$  is the length of the chain and  $N = L/2$  is the number of fermions (half-filling).

For  $V = 0$ , all the single-particle wave functions are exponentially localized for any amount of disorder [56–58]. The interacting case ( $V \neq 0$ ) is the paradigmatic model which is believed to have an MBL transition. Several numerical works have shown that the critical value of the transition is  $W_c \approx 3.5$  for  $V = 1$  [21,22,59–64] ( $W < W_c$  ergodic and  $W > W_c$  localized).

Additionally, in the final part of our work, we will test our method on the quasiperiodic case, known as the Aubry-André model (AA model) [65]. The AA model is obtained from Eq. (1) by setting  $t = 1$ ,  $h_j = W \cos(2\pi j\phi + \alpha)$ , where  $\phi = \frac{1+\sqrt{5}}{2}$  is the golden ratio and  $\alpha$  is a random phase uniformly distributed between  $[0, 2\pi]$ . For  $V = 0$ , the AA model has a metal insulator transition at  $W_c = 1$  [65], extended for  $W < W_c$  and localized for  $W > W_c$ . For  $V = 1$  the AA model is believed to show MBL at strong disorder ( $W_c > 4$ ) [66].

In order to be in the strongly localized regime, in both models, we consider  $W \in \{6, 7, 8\}$ .

## III. METHOD

In this section we introduce the numerical methods and the structure of the CNN [67]. As shown schematically in Fig. 1(a), the first layer is an adapted version of an inception layer [68], followed by a convolution layer and two fully connected layers. As we will show, this architecture enables the network to achieve good results in classifying the snapshots. A detailed description of the architecture and the hyperparameters can be found in the Appendix. The CNNs are trained to distinguish 3D Fock-state snapshots of the Anderson and MBL phases. Those snapshots have been obtained from the out-of-equilibrium dynamics of the system using exact diagonalization. In particular, with the aim to get as close as

possible to an experimental setup, we follow the dynamics of a global quench starting from the Néel state  $|\psi\rangle = \prod_j^N c_{2j}^\dagger |0\rangle$ , which is easy to realize in experiments and belongs to the middle of the spectrum of  $H$ . During the dynamics, we compute the probability to be in the Fock-state  $|\underline{n}\rangle = \prod_i^L (c_i^\dagger)^{n_i} |0\rangle$  with  $n_i \in \{0, 1\}$ ,

$$P_{\underline{n}}(t) = |\langle \underline{n} | e^{-iHt} | \psi \rangle|^2. \quad (2)$$

Finally, at each target time  $t$ , a given number of Fock-state snapshots  $\underline{n} = (n_1, n_2, \dots, n_L)$  are randomly sampled from  $P_{\underline{n}}(t)$ . Doing so, we simulate the collapse of the wave function, as done in an experimental protocol using a quantum microscope [49,50].

We organize the input data for the neural network in a three-dimensional (3D) structure as depicted in Fig. 1(a):  $\# \text{Snapshots} \times L \times N_t$ . A chosen number of Fock-state snapshots are taken at fixed target times and then stacked together to form two-dimensional slices. Along the  $t$  axis, slices from different times are then concatenated to form the 3D input blocks for CNNs [see Fig. 1(a)]. The resulting 3D block structure contains information that can be measured in experiments, namely the location of the fermions at a given time. If not stated otherwise, we will consider  $\# \text{Snapshots} \times L \times N_t = 30 \times 16 \times 11$ , where 30 snapshots were taken at 11 different times from systems of size  $L = 16$ . The target points in time are distributed linearly in log scale up to a final time  $t_f$ .

Finally, we repeat the procedure above to obtain one snapshot block for each random instance of the fields  $\{h_j\}$ , where we generate 1000 random configurations for  $W \in \{6, 7, 8\}$  and  $V \in \{0, 1\}$ . The resulting 6000 snapshot blocks are randomly divided into a training set containing 5000 samples and a validation and test set, both containing 500 samples. The network is trained using only data from the training set and adjusts its weights depending on the snapshots it uses in the training. To avoid overfitting on the training set, the network classifies the validation set after each training epoch. The CNN that achieves the best accuracy on the validation set is used after training to classify the test set [69]. This procedure is repeated for 100 CNNs (if not stated otherwise) to get an averaged classification accuracy.

#### IV. NEURAL NETWORK PERFORMANCE

In this section we study the performance of our neural network in distinguishing the MBL phase from the Anderson localized one. Furthermore, we inspect the dependence on the input data parameters, i.e.,  $\# \text{Snapshots}$ , system size  $L$  and final time  $t_f$ . Particular attention is given to the 3D structure of our data and on the fundamental importance of the time axis, which provides dynamic information of the system.

The performance of our CNN is quantified by the accuracy  $\in [0, 1]$ , which is given by the number of exactly classified samples within the test set divided by the total number of samples in the test set. Thus, the CNN fails to classify the two phases if the accuracy is  $\approx 1/2$  since it is equivalent to making a decision using a fair coin. Otherwise, if the accuracy is equal to one, the neural network has a perfect performance. Classifying ergodic/thermal and MBL phases with neural

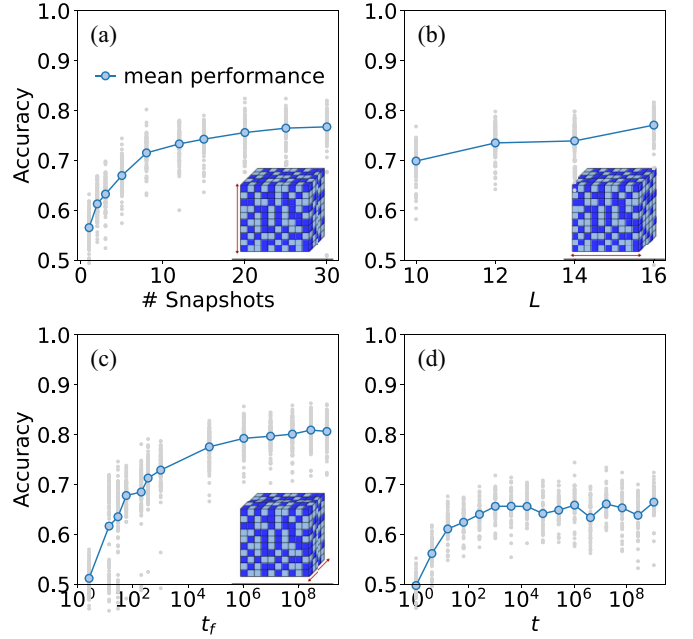


FIG. 2. Dependence of the classification accuracy on variations of the snapshot block parameters  $\# \text{Snapshots}$ ,  $L$ ,  $t_f$ , and  $t$ , where we tune each parameter separately. (a)  $L = 16$ ,  $t_f = 10^3$ , and we tune  $\# \text{Snapshots}$ . (b)  $\# \text{Snapshots} = 30$ ,  $t_f = 10^3$ , and  $L$  is varied. (c)  $\# \text{Snapshots} = 30$ ,  $L = 16$  are fixed and the final time  $t_f$  is varied. (d) Unlike the other panels where the CNNs were trained using snapshots at different times in the range  $[0, t_f]$ , in this panel, the CNNs were trained using only snapshots at a single target time  $t$ . In all panels, the gray dots are the performance of single CNNs and  $N_t = 11$  is fixed.

networks can be done with high accuracy  $\approx 1$  [40,45,46,48] since the two phases are fundamentally distinct. However, one expects that distinguishing an MBL from the Anderson phase is a much harder problem, since the two phases are both localized and differ only in terms of the information propagation [18,22].

First, we study the dependence of our classification accuracy on the value of  $\# \text{Snapshots}$  used in the CNN. This is particularly relevant in an experimental setup, where one would like to reduce the number of measurements. Figure 2(a) shows the mean test accuracy for  $t_f = 10^3$  and  $L = 16$  as a function of  $\# \text{Snapshots}$ , where the average was performed over 100 CNNs. For completeness, in Fig. 2, the performance of individual CNNs is also reported (dashed vertical dots), which quantifies the error. As expected, the accuracy increases with the number of snapshots used and reaches its maximum at  $\approx 0.75$ . Importantly, the accuracy reaches an almost constant value already for  $\# \text{Snapshots} = 15$ , meaning that only a limited number of measurements are needed. In Fig. 2(b) the accuracy of our CNNs is shown as a function of the system size  $L$ . Although only slightly, the mean accuracy increases with  $L$ , providing evidence that in an experimental setup, where hundreds of sites can be probed, the CNNs could perform even better.

Another important parameter of our input data is given by the final time  $t_f$ , which is the latest time in the dynamics. In Fig. 2(c) we study the dependence of  $t_f$  on our results. In

particular, we train our CNNs always taking  $N_t = 11$  target times in  $t \in (0, t_f]$  [70]. This ensures that the amount of input data is independent of the final time  $t_f$ , and, consequently, we can fairly compare the performance of our CNNs at different  $t_f$ . An increase in accuracy as a function of  $t_f$  is observed [see Fig. 2(c)], reaching accuracies  $\approx 0.8$  for  $t_f \approx 10^8$ . Though these large times are still not affordable in an experimental setup, where the longest times that have been simulated are of order  $\approx 10^3$  hopping units [71], the results in Fig. 2(c) provide us a hint of the kind of information learned by the neural network. We remind the reader that at such long timescales ( $t_f \approx 10^8$ ), in an MBL phase, local degrees of freedom are frozen and the only relevant dynamics are induced by the dephasing mechanism producing a logarithmic slow propagation of entanglement [18–20]. As a result, we can fairly assume that our CNNs are able to distinguish an MBL phase from an Anderson insulator by learning dynamic correlations between the snapshots induced by the dephasing mechanism.

Finally, we would like to emphasize the important role of the time axis in the 3D structure of the input data. Figure 2(d) shows the accuracy as a function of individual time  $t$  of our CNNs, where the CNNs were trained using input data only from time  $t$ . Thus, the CNNs are not trained with the full dynamic range  $t \in [0, t_f]$ , but only on single time slices. To fairly compare the results, we want to give the same amount of information to our network and hence stack the  $N_t = 11$  slices of snapshots from the same time  $t$  to form a 3D block. As expected, at short times  $t \approx 1$ , the neural network is not able to distinguish the two phases (accuracy  $\approx 1/2$ ), since interactions do not play any relevant role yet. Using larger times, a distinction is possible, though the performance is clearly worse in comparison to the performance of the CNNs trained with the full dynamic range. We conclude that the temporal correlation plays a major role in distinguishing the two phases.

## V. NETWORK ROBUSTNESS

Having tested the performance of our method, we now focus on the robustness of the CNNs. The goal of our work is to construct a neural network that can classify experimental data into an MBL and Anderson insulator. Ideally, one would like to train a CNN using numerical data and then classify experimental data. This opens up the issue that the data from training and testing are originated from the same source, namely numerical simulations. Moreover, experimental data can have several forms of imperfections, i.e., the Hamiltonian's parameters are known only up to some precision. In the following two subsections we show that the performance of our CNN is robust if tested on data produced with different Hamiltonian parameters. Then we train our CNN using Fock-space snapshots taken only from a small block of the system, allowing particle fluctuations. We show that information extracted from the small subsystem is already good enough to reliably distinguish the two phases [72].

### A. Robustness towards Hamiltonian perturbations

In the following we present an argument to provide evidence that our CNN approach is robust when tested on

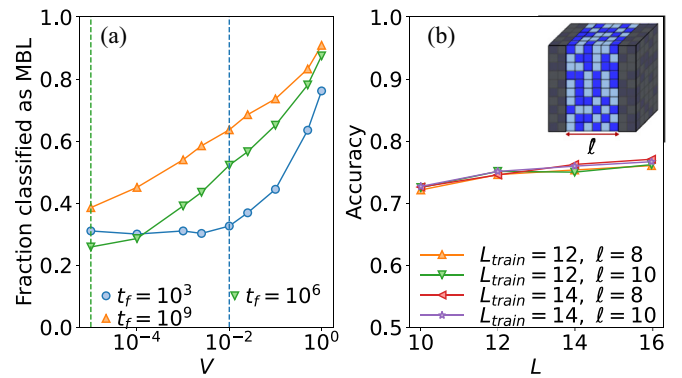


FIG. 3. (a) Performance of the CNNs on classifying data with  $V \in [10^{-5}, 1]$ , where the CNNs were only trained on different interaction strengths  $V \in \{0, 1\}$ . Different curves represent the performance of the CNNs with different final times  $t_f$ . (b) Average accuracy for CNNs trained on subblock snapshots of length  $\ell$  in a system of system size  $L_{\text{train}} \geq \ell$  and tested on subblock snapshots always of length  $\ell$ , but in a system of size  $L$ .

imperfect data sets and hence could be used in an experimental setup. Figure 3(a) shows the fraction of correctly classified MBL [73] as a function of interaction strength  $V \in [10^{-5}, 1]$  for systems of size  $L = 16$ . Importantly, the CNNs have been trained using only the noninteracting samples ( $V = 0$ ) and the interacting ones with  $V = 1$ . Moreover, Fig. 3(a) also shows the results for different final times  $t_f \in \{10^3, 10^6, 10^9\}$ . A few considerations are in order. As expected, the timescale  $t_f$  plays an important role. In agreement with the results in Fig. 2(c), the accuracy increases with  $t_f$ , since the network is trained in a longer dynamic range. As expected, if  $t_f \ll V^{-1}$  the performance is poor, since interactions have not shown their effects yet. This is manifested in the plateau of the blue curve ( $t_f = 10^3$ ) that exists only until  $t_f V \approx 10$  [dashed vertical line in Fig. 3(a)], where the accuracy starts to increase. In Sec. VI we will see in detail that features like the logarithmic growth of the entanglement entropy  $S(t) \sim \log tV$  are consistent with the information extracted by the networks. Thus, the rise of accuracy at  $\sim t_f V^{-1}$  could be explained by the onset of the information growth [19]. Importantly, the performance of our CNN is robust if we moderately perturb the Hamiltonian's parameters, as can be seen for interaction strengths close to the value used for training ( $V = 1$ ).

### B. Robustness towards subsystem classifications

Now we focus on the question of whether our method is stable if the CNN is trained using only Fock-state snapshots from a subsystem. The idea is to use cut-out techniques by training the CNN with Fock-state snapshots of length  $\ell$  taken from a small subsystem ( $n_{L_{\text{train}}/2-\ell/2}, \dots, n_{L_{\text{train}}/2+\ell/2}$ ) with  $n_i \in \{0, 1\}$  and  $L_{\text{train}}$  the length of the system used to generate the snapshot blocks, see Fig. 3(b) [74]. Hence the CNN learns how to make a distinction while only seeing a small subsystem of length  $\ell$  out of the entire system of length  $L_{\text{train}}$ . Figure 3(b) shows the test accuracy for several combinations of  $\ell$ ,  $L_{\text{train}}$ , and the size of the system used to test the networks  $L$ . Remarkably, the performance of the network shows almost no dependence on  $L_{\text{train}}$ ,  $\ell$ , and, in agreement

with the results in Fig. 2(b), the accuracy increases with  $L$ . This increase could be explained by the decreases of finite size effects with increasing  $L$ . The stability on  $L_{\text{train}}$  and  $\ell$  is a direct consequence of the fact that we are at strong disorder and therefore the localization length  $\xi_{\text{loc}}$  is shorter than the dimension of the subsystem  $\ell$ . As a result, one might train the CNN using finite-size numerics and then use the network to classify data from experiments, which are usually done on larger system sizes. In particular, subblock Fock-space snapshots have no particle number conservation, therefore we also tested the efficiency of our method in the case of a fluctuating number of particles.

In summary, we have found that the CNN is stable to small perturbations of the Hamiltonian's parameters. Moreover, the CNN can have a high performance even if trained using only Fock-state snapshots taken from a smaller subsystem. As a consequence, only measurements in a finite portion of the system are needed to distinguish the two phases. Second, this opens the possibility to train our CNN using exact diagonalization techniques, which are limited to small system sizes.

## VI. WHAT DOES THE NEURAL NETWORK LEARN?

Understanding how a neural network makes a distinction and on which patterns it focuses is usually a challenging task [69]. The aim of this section is to evaluate which information the neural network uses to distinguish the two phases.

Several methods have been proposed to solve this issue which are based on the examination of the kernels of the convolutional layer. For example, Schindler *et al.* [43] used the so-called dreaming mechanism [75], where a pretrained network modifies random input data until this data would be classified in one of the two phases. By doing so, the network dreams about new data, meaning it produces new data that has the features of the phases which are critical for the classification.

Here we take the more pragmatic approach of taking a closer look at the correctly classified and misclassified samples. In Sec. V we have provided evidence that the neural network might learn features connected to the interaction-induced dephasing mechanism and consequentially to the information propagation through the system. Thus, motivated by the last observation, we compare the dynamics of the correctly and wrongly classified samples using two different probes to distinguish the two phases. First, we compute the time evolution of the bipartite half-chain entanglement entropy

$$S(t) = -\text{Tr}[\rho_{L/2} \log \rho_{L/2}], \quad (3)$$

where  $\rho_{L/2}(t)$  is the half-chain reduced density matrix of the evolved states  $|\psi(t)\rangle$ . In an MBL phase, after an initially short ballistic propagation,  $S(t)$  spreads logarithmically slow in time and finally reaches a nonthermal volume law steady state for asymptotically long times [ $S(\infty) \sim O(L)$ ]. Instead, in an Anderson insulator phase entanglement does not propagate, and thus after the initial transient propagation,  $S(t)$  saturates to an area law value [ $S(\infty) \sim O(L^0)$ ]. The second dynamic

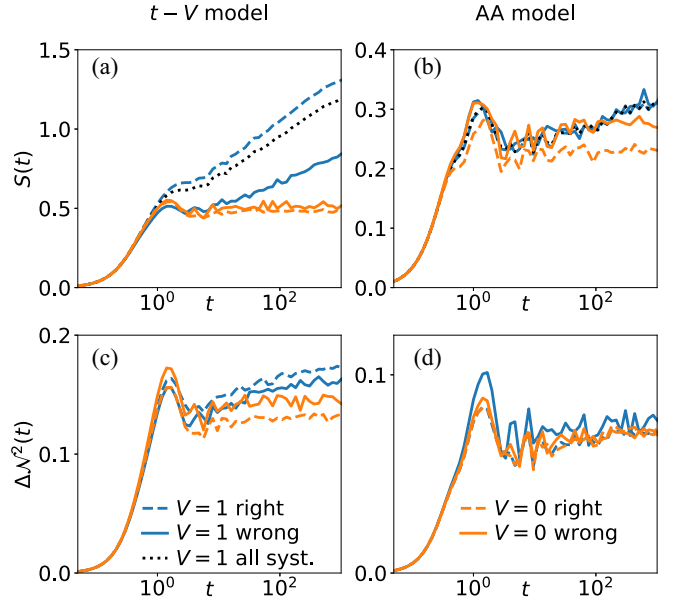


FIG. 4. (a) and (c) The growth of the entanglement entropy  $S(t)$  and particle fluctuations  $\Delta\mathcal{N}^2(t)$  averaged separately over the correctly and wrongly classified Anderson ( $V = 0$ ) and MBL ( $V = 1$ ) samples for the disordered  $t$ - $V$  model. (b) and (d) The same information as in the adjacent panels, but for the AA model with quasiperiodic potential.

indicator we evaluate is the particle number fluctuation

$$\Delta\mathcal{N}^2(t) = \langle O^2(t) \rangle - \langle O(t) \rangle^2, \quad O = \sum_{j=1}^{L/2} n_j. \quad (4)$$

In an interacting localized phase,  $\Delta\mathcal{N}^2(t)$  has an extremely slow propagation over several orders of magnitude in time, which is consistent with  $\Delta\mathcal{N}^2(t) \propto \log \log t$  or  $\propto \log^\nu t$  with  $\nu < 1$  [14,76,77], unlike in an Anderson localized phase,  $\Delta\mathcal{N}^2(t \gg 1) \sim O(L^0)$ . Although it is debated if the found propagation is only transient or persistent for asymptotically long times [76,78],  $\Delta\mathcal{N}^2(t)$  still remains a useful and experimentally accessible probe to distinguish the two phases [7] at finite timescales.

During the time evolution, we compute both the entanglement entropy  $S(t)$  and the particle fluctuation  $\Delta\mathcal{N}^2(t)$  and simultaneously sample randomly the Fock-state snapshots  $\underline{n} = (n_1, \dots, n_L)$  from the probability distribution  $P_{\underline{n}}(t)$  in Eq. (2). Finally, we test our CNN on this data, and in order to understand which features the CNN learns, we take a closer look at  $S(t)$  and  $\Delta\mathcal{N}^2(t)$  for the samples that have been correctly and wrongly classified.

Figures 4(a) and 4(c) show the entanglement entropy  $S(t)$  and the particle number fluctuation  $\Delta\mathcal{N}^2(t)$  averaged separately over correctly and wrongly classified samples (dashed lines), respectively. For the interacting case,  $t$ - $V$  model with  $V = 1$ , a clear separation between the correctly and wrongly classified snapshots is visible. Crucially, the samples which the network wrongly classified as Anderson localized exhibit a late onset and a slower propagation of information, thus affirming our main assumption.

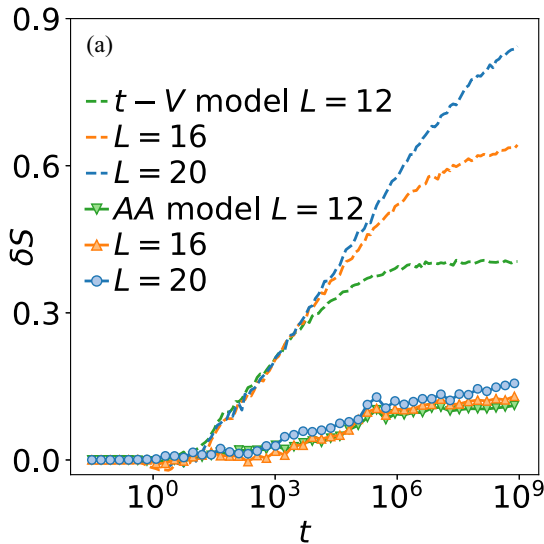


FIG. 5. (a) The entanglement entropy after subtracting the non-interacting value  $V = 0$ . The dynamics is computed using the  $\ell$ -bit Hamiltonian  $H^{\text{eff}}$  with  $W = 6$  and  $V = 0.1$ , for the  $t$ - $V$  (dashed line) and the AA (solid line) models.

The situation changes drastically when we consider the interacting model having quasiperiodic potential (AA model), defined in Sec. II [66,79,80]. In order to have a fair comparison to the  $t$ - $V$  model, we train our CNNs with the same 3D structure of the input data  $\#\text{Snapshots} \times L \times N_t$  with  $\#\text{Snapshots} = 30$ ,  $L = 16$ ,  $N_t = 11$ , final time  $t_f = 10^3$ , and  $W \in \{6, 7, 8\}$  [81].

In the case of the  $t$ - $V$  model, the CNN reaches an accuracy of  $\approx 76\%$ , while for the AA model the network is on average only able to classify  $\approx 66\%$  of the samples correctly (see the Appendix). The difficulty for the CNN to distinguish the two phases for the AA model is connected to an atypically slow information propagation through the system. Figures 4(b) and 4(d) show  $S(t)$  and  $\Delta\mathcal{N}^2(t)$  averaged only over correctly and wrongly classified samples, respectively. As one can observe, in the AA model the information propagation is much slower than in the case with the random potential [see Figs. 4(a) and 4(c)]. Particularly,  $S(t)$  has only a growth of  $\approx 30\%$  at timescale  $t_f = 10^3$ , compared to the noninteracting case, and  $\Delta\mathcal{N}^2(t)$  does not show any significant growth, as shown in Figs. 4(a) and 4(c). As a consequence, our CNNs have difficulties in distinguishing the two phases.

This anomalous propagation is traceable to the fractal nature of the single-particle spectrum of the AA model [82]. Indeed, it is known that the energy spectrum of the AA model is multifractal and thus its spectrum hosts quasidegenerate single-particle energy [82]. As a consequence, these quasidegenerate frozen degrees of freedom with energies  $\epsilon_x$  and  $\epsilon_y$  will need a time proportional to  $(\epsilon_x - \epsilon_y)^{-1}$  to dephase. We test this idea by studying the EE growth for a minimal-model constructed from  $H$  in Eq. (1) using perturbation theory in the interaction strength  $V$ . In the limit of weak interactions  $V \ll 1$  and strong disorder  $W \gg 1$ ,  $H$  can be approximated as  $H^{\text{eff}} = \sum_l \epsilon_l \eta_l^\dagger \eta_l + V \sum_{l,m} \mathcal{B}_{lm} \eta_l^\dagger \eta_l \eta_m^\dagger \eta_m$ , where  $\eta_l^\dagger (\eta_l)$  is the creation (annihilation) operator for the single-particle orbital  $\phi_l(x)$  at energy  $\epsilon_l$ , and  $\mathcal{B}_{lm}$  can be found by using perturbation

theory (for details see Ref. [14]). In this limit,  $H^{\text{eff}}$  represents the paradigmatic model known as  $\ell$ -bits Hamiltonian [13,15,83], describing an MBL phase at strong disorder.

Figure 5 shows  $\delta S$ , which is the entanglement entropy after subtracting the noninteracting values ( $V = 0$ ).  $\delta S$  is computed using  $H^{\text{eff}}$  constructed from the  $t$ - $V$  model (dashed lines) and the AA model (solid lines). As one can observe and in agreement with the results in Fig. 4, in the MBL phase for the AA model  $S(t)$  has an anomalous slow propagation if compared to the  $t$ - $V$  model case. We checked numerically that the dephasing couplings  $\mathcal{B}_{l,m}$  in  $H^{\text{eff}}$  do not present significant differences between the two models. Thus, the important difference between the two models is only given by the single-particle energies  $\{\epsilon_l\}$ . Indeed, in the case of the  $t$ - $V$  model the  $\{\epsilon_l\}$  are Poissonian distributed and their density of states can be approximated with a box function at strong disorder, whereas in the AA model the  $\{\epsilon_l\}$  are almost degenerate and the density of states is a Cantor set. As a result, this provides numerical evidence that the slower propagation of information for the AA model is due to the fractal nature of its single-particle spectrum [82].

## VII. CONCLUSION

In this work we addressed the question of how to distinguish an Anderson insulator from an MBL phase using snapshot data by formulating a method based on machine learning tools, i.e., CNNs. We trained a CNN using a 3D structure for the input data, which contains a fixed amount of space-time Fock-state snapshots. A particular focus was given to this 3D structure and on the fundamental importance of having dynamic/temporal information. Unlike the case of distinguishing an ergodic phase from a localized one, where only a fixed amount of snapshots is important [48], here it is crucial to analyze snapshots at different times. Thus, this work provides a novel method to analyze accessible experimental data (e.g., from cold-atoms setups) and to separate the two localized phases.

We benchmarked our method on the paradigmatic model exhibiting an MBL phase ( $t$ - $V$  model with quenched disorder) and showed that the CNN reaches accuracies of  $\approx 80\%$  in distinguishing the two phases. We studied the stability of our method and provided evidence that it can be used to analyze real experimental data. Importantly, we found an upper bound to the number of snapshots needed in an experimental setup to distinguish the two phases.

In order to understand what kind of features of the input data are used by the CNN to classify the two phases, we took a closer look at the dynamics of the entanglement entropy and the particle fluctuation. Both quantities are known to be dynamical indicators of an MBL phase. From this analysis we provided evidence that the CNN makes use of the difference in the information propagation between an MBL phase and an Anderson localized one. In fact, we showed that the interacting samples misclassified as Anderson localized are characterized by an unusually slow entanglement propagation.

Finally, we applied our method to a model with quasiperiodic potential (AA model). Like the  $t$ - $V$  model, this model has been used several times in cold-atoms experiments to study the MBL transition. In this case, the neural networks

had difficulties when separating the phases, and the reached accuracy is significantly lower than in the case of fully random potentials. This is due to a quantitative slower propagation of information in the AA model compared to the  $t$ - $V$  model. Indeed, we showed that the MBL phase of the AA model has a much slower growth of entanglement and the particle fluctuation fails to be a dynamical indicator to distinguish the two phases. These results indicate that separating an MBL phase from an Anderson insulator in the case of quasiperiodic is potentially more challenging than for a disordered system.

## ACKNOWLEDGMENTS

We would like to thank F. Meissen for his insightful comments. We also express our gratitude to S. Bera, R. Hardman Carter, S. Kim, and S. Lechner for the critical reading of the manuscript. F.P. acknowledges support from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (Grant Agreement No. 771537). F.P. acknowledges the support of the Deutsche Forschungs-gemeinschaft (DFG, German Research Foundation) under Germanys Excellence Strategy EXC-2111-390814868 and TRR 80. G.D.T. acknowledges the support from the EPIQS Program of the Gordon and Betty Moore Foundation.

## APPENDIX A

### 1. Network architecture and hyperparameters

The CNN used in this work has a rather simple architecture, which allows extracting complicated dynamic patterns from the 3D snapshot input. It consists of two convolution layers with 3D kernels and two fully connected layers. Figure 6 shows a schematic representation of the neural network. The first layer is an adaptation of the inception layer [68], where three different kernels extract information from the input data. One kernel of size  $(1 \times 29 \times 1)$  is able to see many snapshots of one site at a time, allowing it to compute averages over the snapshots. The second kernel of size  $(5 \times 5 \times 5)$  gets information from different sites, snapshots, and times. The third kernel of size  $(3 \times 3 \times 7)$  is specially designed to extract dynamic information, hence it has access to seven points in time. All kernels have a stride equal to one and a padding is chosen to match the dimension of the output after the convolution layer to the dimension of the input (known as *same padding*). A pooling layer reduces the amount of information, before the feature map is given to a second convolution layer with a kernel size of  $(2 \times 2 \times 2)$ , followed by another pooling layer. The last two layers are fully connected layers with 50 and 2 neurons to make the classification.

The optimizer used in training is Adam [84]. Note that we use a smaller learning rate of  $10^{-4}$  since the input data heavily depends on the random collapse of the wave function. The loss function we use is the cross-entropy loss, additional hyperparameters are listed in Table I. The CNNs were implemented in Pytorch [85].

The training was performed on GPUs supporting CUDA platform, namely a GeForce GTX 960, GeForce GTX 1050

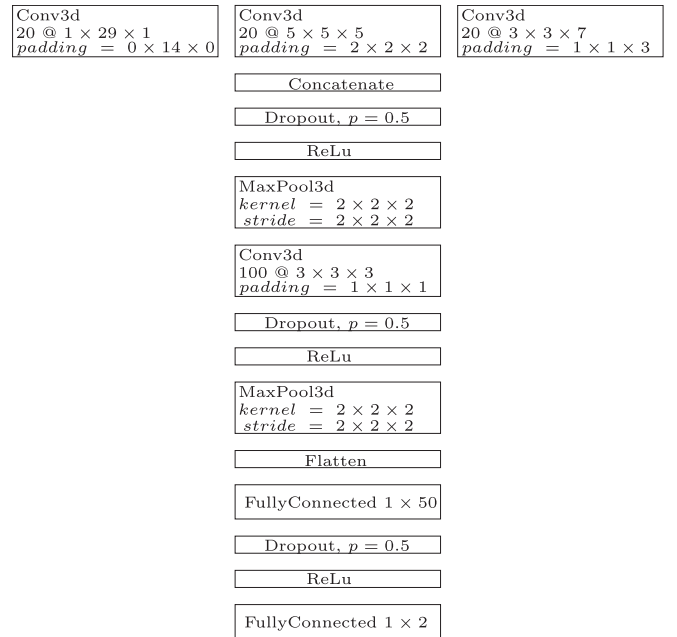


FIG. 6. Schematic representation of the CNN. The first layer is an adaptation of an inception layer [68], followed by a simple convolutional layer. The final classification is done by two fully connected layers. Two three-dimensional MaxPool layers shrink down the data size after the inception layer and after the convolutional layer. Non-linearities are introduced by the rectified linear unit (ReLU). Dropout is included to increase the classification accuracy.

Ti, and GeForce GTX 1650 with a minimum memory of 4 GB.

### 2. Quantum Fisher information

As in Sec. VI, we use the quantum Fisher information (QFI) to compare the dynamics of the right and wrong classified samples. The QFI is defined by

$$\mathcal{F}(t) = 4[\langle \hat{O}^2 \rangle - \langle \hat{O} \rangle^2], \quad \hat{O} = \sum_i (-1)^i \hat{n}_i. \quad (\text{A1})$$

In Fig. 7(a) a clear distinction between right and wrong classifications of the two phases with  $V = 0, 1$  cannot be made when examining  $\mathcal{F}(t)$ , indicating that the network does not focus on patterns that are comparable to  $\mathcal{F}(t)$ . For the quasiperiodic AA model in (b), we see that  $\mathcal{F}(t)$  is not a good quantity even to distinguish  $V = 0$  and  $V = 1$  since it exhibits no different dynamics for small times  $t < 10^3$ .

TABLE I. Hyperparameters we use for implementation and training. Parameters not listed are the standard parameters proposed by Pytorch.

Hyperparameter	Value
Learning rate	$10^{-4}$
Learning rate decay	0.995
Weight decay	0
Number of epochs	150

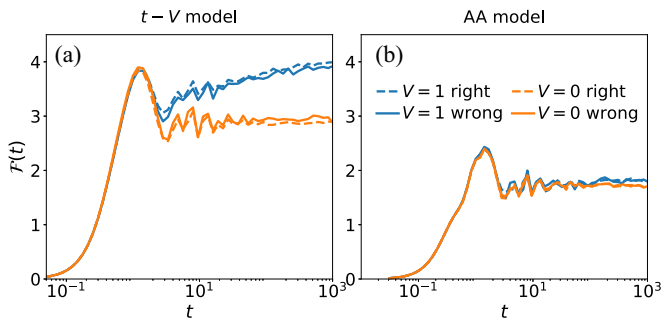


FIG. 7. Both panels show the QFI  $\mathcal{F}(t)$  averaged separately over the correctly and wrongly classified MBL and Anderson insulating samples for the  $t$ - $V$  model (a) and the AA model (b).

### 3. Network performance and robustness for the AA model

In this section we test the robustness of the CNNs and their general performance to the case with the quasiperiodic potential (AA model), just as we did in Secs. IV and V for the  $t$ - $V$  model. We find out that the neural network performs worse in the case of quasiperiodic potentials.

In Fig. 8(a) we test the dependence of the network performance on the shape of the snapshot blocks in the AA model, comparable to Fig. 2. The classification accuracy rises with rising #Snapshots up to a saturation value of about 65%, which is roughly 10% less than in the  $t$ - $V$  model. The snapshot blocks are taken from systems with  $L = 16$ ,  $t_f = 10^3$ , the result is averaged over 50 networks, gray dots show the performance of single CNNs.

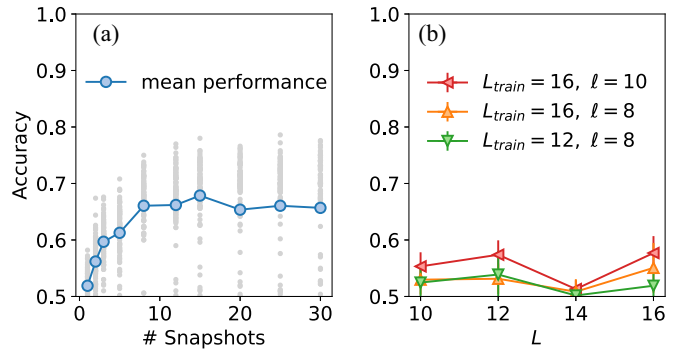


FIG. 8. (a) Dependence of the network performance on #Snapshots in the AA model. The blue line indicates the averaged performance over 50 networks, gray dots mark the performance of single CNNs. (b) Average accuracy for CNNs trained on subblock snapshots of length  $\ell$  in a system of size  $L_{\text{train}} \geq \ell$  and tested on subblock snapshots always of length  $\ell$ , but in a system of size  $L$  in the AA model.

In Fig. 8(b) we also test the robustness of our network when we change the chain length  $L$  in the AA model. Therefore, we use cutoffs of length  $\ell$  defined in Sec. V. For the  $t$ - $V$  model we saw good classification results when we trained our model using smaller blocks of length  $\ell$ , see Fig. 3(b). However, the very same idea and network architecture fail for the AA model, producing classification accuracies  $< 60\%$  which is only slightly better than tossing a fair coin.

Hence, we conclude that the dynamic patterns the CNN can extract from the 3D snapshot blocks are less apparent in the quasiperiodic AA model than in the  $t$ - $V$  model.

- [1] C. Gross and I. Bloch, Quantum simulations with ultracold atoms in optical lattices, *Science* **357**, 995 (2017).
- [2] C. Monroe, W. C. Campbell, L.-M. Duan, Z.-X. Gong, A. V. Gorshkov, P. W. Hess, R. Islam, K. Kim, N. M. Linke, G. Pagano, P. Richerme, C. Senko, and N. Y. Yao, Programmable quantum simulations of spin systems with trapped ions, *Rev. Mod. Phys.* **93**, 025001 (2021).
- [3] R. Blatt and C. F. Roos, Quantum simulations with trapped ions, *Nat. Phys.* **8**, 277 (2012).
- [4] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, *Annu. Rev. Condens. Matter Phys.* **11**, 369 (2020).
- [5] I. Bloch, J. Dalibard, and W. Zwerger, Many-body physics with ultracold gases, *Rev. Mod. Phys.* **80**, 885 (2008).
- [6] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, Observation of many-body localization of interacting fermions in a quasirandom optical lattice, *Science* **349**, 842 (2015).
- [7] J. Smith, A. Lee, P. Richerme, B. Neyenhuis, P. W. Hess, P. Hauke, M. Heyl, D. A. Huse, and C. Monroe, Many-body localization in a quantum simulator with programmable random disorder, *Nat. Phys.* **12**, 907 (2016).
- [8] J.-y. Choi, S. Hild, J. Zeiher, P. Schauß, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, Exploring the many-body localization transition in two dimensions, *Science* **352**, 1547 (2016).
- [9] R. Nandkishore and D. A. Huse, Many-body localization and thermalization in quantum statistical mechanics, *Annu. Rev. Condens. Matter Phys.* **6**, 15 (2015).
- [10] E. Altman and R. Vosk, Universal dynamics and renormalization in many-body-localized systems, *Annu. Rev. Condens. Matter Phys.* **6**, 383 (2015).
- [11] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, Colloquium: Many-body localization, thermalization, and entanglement, *Rev. Mod. Phys.* **91**, 021001 (2019).
- [12] F. Alet and N. Laflorencie, Many-body localization: An introduction and selected topics, *C. R. Phys.* **19**, 498 (2018).
- [13] J. Z. Imbrie, V. Ros, and A. Scardicchio, Local integrals of motion in many-body localized systems, *Ann. Phys.* **529**, 1600278 (2017).
- [14] G. De Tomasi, F. Pollmann, and M. Heyl, Efficiently solving the dynamics of many-body localized systems at strong disorder, *Phys. Rev. B* **99**, 241114(R) (2019).
- [15] D. A. Huse, R. Nandkishore, and V. Oganesyan, Phenomenology of fully many-body-localized systems, *Phys. Rev. B* **90**, 174202 (2014).
- [16] M. Serbyn, Z. Papić, and D. A. Abanin, Local Conservation Laws and the Structure of the Many-Body Localized States, *Phys. Rev. Lett.* **111**, 127201 (2013).



- [17] A. Chandran, I. H. Kim, G. Vidal, and D. A. Abanin, Constructing local integrals of motion in the many-body localized phase, *Phys. Rev. B* **91**, 085425 (2015).
- [18] J. H. Bardarson, F. Pollmann, and J. E. Moore, Unbounded Growth of Entanglement in Models of Many-Body Localization, *Phys. Rev. Lett.* **109**, 017202 (2012).
- [19] M. Serbyn, Z. Papić, and D. A. Abanin, Universal Slow Growth of Entanglement in Interacting Strongly Disordered Systems, *Phys. Rev. Lett.* **110**, 260601 (2013).
- [20] M. Žnidarič, T. c. v. Prosen, and P. Prelovšek, Many-body localization in the Heisenberg XXZ magnet in a random field, *Phys. Rev. B* **77**, 064426 (2008).
- [21] G. De Tomasi, S. Bera, J. H. Bardarson, and F. Pollmann, Quantum Mutual Information as a Probe for Many-Body Localization, *Phys. Rev. Lett.* **118**, 016804 (2017).
- [22] M. Serbyn, Z. Papić, and D. A. Abanin, Quantum quenches in the many-body localized phase, *Phys. Rev. B* **90**, 174302 (2014).
- [23] D. Roy, R. Singh, and R. Moessner, Probing many-body localization by spin noise spectroscopy, *Phys. Rev. B* **92**, 180205(R) (2015).
- [24] M. Serbyn, M. Knap, S. Gopalakrishnan, Z. Papić, N. Y. Yao, C. R. Laumann, D. A. Abanin, M. D. Lukin, and E. A. Demler, Interferometric Probes of Many-Body Localization, *Phys. Rev. Lett.* **113**, 147204 (2014).
- [25] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Learning phase transitions by confusion, *Nat. Phys.* **13**, 435 (2017).
- [26] E. van Nieuwenburg, Y. Baum, and G. Refael, From Bloch oscillations to many-body localization in clean interacting systems, *Proc. Natl. Acad. Sci. USA* **116**, 9269 (2019).
- [27] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, *Nat. Phys.* **13**, 431 (2017).
- [28] S. J. Wetzell, Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders, *Phys. Rev. E* **96**, 022140 (2017).
- [29] K. Chng, J. Carrasquilla, R. G. Melko, and E. Khatami, Machine Learning Phases of Strongly Correlated Fermions, *Phys. Rev. X* **7**, 031038 (2017).
- [30] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [31] P. Zhang, H. Shen, and H. Zhai, Machine Learning Topological Invariants with Neural Networks, *Phys. Rev. Lett.* **120**, 066401 (2018).
- [32] A. Bohrdt, C. S. Chiu, G. Ji, M. Xu, D. Greif, M. Greiner, E. Demler, F. Grusdt, and M. Knap, Classifying snapshots of the doped Hubbard model with machine learning, *Nat. Phys.* **15**, 921 (2019).
- [33] M. S. Scheurer and R.-J. Slager, Unsupervised Machine Learning and Band Topology, *Phys. Rev. Lett.* **124**, 226401 (2020).
- [34] T. Neupert, M. H. Fischer, E. Greplova, K. Choo, and M. Denner, Introduction to machine learning for the sciences, [arXiv:2102.04883](https://arxiv.org/abs/2102.04883).
- [35] 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**, L140202 (2021).
- [36] T. Ohtsuki and T. Ohtsuki, Deep learning the quantum phase transitions in random two-dimensional electron systems, *J. Phys. Soc. Jpn.* **85**, 123706 (2016).
- [37] T. Ohtsuki and T. Mano, Drawing phase diagrams of random quantum systems by deep learning the wave functions, *J. Phys. Soc. Jpn.* **89**, 022001 (2020).
- [38] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, Machine learning quantum phases of matter beyond the fermion sign problem, *Sci. Rep.* **7**, 8823 (2017).
- [39] E. V. H. Doggen, F. Schindler, K. S. Tikhonov, A. D. Mirlin, T. Neupert, D. G. Polyakov, and I. V. Gornyi, Many-body localization and delocalization in large quantum chains, *Phys. Rev. B* **98**, 174202 (2018).
- [40] Y.-T. Hsu, X. Li, D.-L. Deng, and S. Das Sarma, Machine Learning Many-Body Localization: Search for the Elusive Non-ergodic Metal, *Phys. Rev. Lett.* **121**, 245701 (2018).
- [41] 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**, 104106 (2019).
- [42] W.-J. Rao, Machine learning the many-body localization transition in random spin systems, *J. Phys.: Condens. Matter* **30**, 395902 (2018).
- [43] F. Schindler, N. Regnault, and T. Neupert, Probing many-body localization with neural networks, *Phys. Rev. B* **95**, 245134 (2017).
- [44] R. Kausar, W.-J. Rao, and X. Wan, Learning what a machine learns in a many-body localization transition, *J. Phys.: Condens. Matter* **32**, 415605 (2020).
- [45] 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**, 224202 (2019).
- [46] W. Zhang, L. Wang, and Z. Wang, Interpretable machine learning study of the many-body localization transition in disordered quantum Ising spin chains, *Phys. Rev. B* **99**, 054208 (2019).
- [47] E. van Nieuwenburg, E. Bairey, and G. Refael, Learning phase transitions from dynamics, *Phys. Rev. B* **98**, 060301(R) (2018).
- [48] A. Bohrdt, S. Kim, A. Lukin, M. Rispoli, R. Schittko, M. Knap, M. Greiner, and J. Léonard, Analyzing Nonequilibrium Quantum States through Snapshots with Artificial Neural Networks, *Phys. Rev. Lett.* **127**, 150504 (2020).
- [49] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner, A quantum gas microscope for detecting single atoms in a Hubbard-regime optical lattice, *Nature (London)* **462**, 74 (2009).
- [50] J. F. Sherson, C. Weitenberg, M. Endres, M. Cheneau, I. Bloch, and S. Kuhr, Single-atom-resolved fluorescence imaging of an atomic Mott insulator, *Nature (London)* **467**, 68 (2010).
- [51] F. A. An, K. Padavić, E. J. Meier, S. Hegde, S. Ganeshan, J. H. Pixley, S. Vishveshwara, and B. Gadway, Interactions and Mobility Edges: Observing the Generalized Aubry-André Model, *Phys. Rev. Lett.* **126**, 040603 (2021).
- [52] H. P. Lüschen, P. Bordia, S. Scherg, F. Alet, E. Altman, U. Schneider, and I. Bloch, Observation of Slow Dynamics Near the Many-Body Localization Transition in One-Dimensional Quasiperiodic Systems, *Phys. Rev. Lett.* **119**, 260401 (2017).
- [53] P. Bordia, H. Lüschen, S. Scherg, S. Gopalakrishnan, M. Knap, U. Schneider, and I. Bloch, Probing Slow Relaxation and Many-Body Localization in Two-Dimensional Quasiperiodic Systems, *Phys. Rev. X* **7**, 041047 (2017).
- [54] P. Bordia, H. P. Lüschen, S. S. Hodgman, M. Schreiber, I. Bloch, and U. Schneider, Coupling Identical One-Dimensional Many-Body Localized Systems, *Phys. Rev. Lett.* **116**, 140401 (2016).

- [55] T. Kohlert, S. Scherg, X. Li, H. P. Lüschen, S. Das Sarma, I. Bloch, and M. Aidelsburger, Observation of Many-Body Localization in a One-Dimensional System with a Single-Particle Mobility Edge, *Phys. Rev. Lett.* **122**, 170403 (2019).
- [56] P. W. Anderson, Absence of diffusion in certain random lattices, *Phys. Rev.* **109**, 1492 (1958).
- [57] F. Evers and A. D. Mirlin, Anderson transitions, *Rev. Mod. Phys.* **80**, 1355 (2008).
- [58] N. F. Mott and W. Twose, The theory of impurity conduction, *Adv. Phys.* **10**, 107 (1961).
- [59] D. J. Luitz, N. Laflorencie, and F. Alet, Many-body localization edge in the random-field Heisenberg chain, *Phys. Rev. B* **91**, 081103(R) (2015).
- [60] S. Bera, H. Schomerus, F. Heidrich-Meisner, and J. H. Bardarson, Many-Body Localization Characterized from a One-Particle Perspective, *Phys. Rev. Lett.* **115**, 046603 (2015).
- [61] M. Serbyn, Z. Papić, and D. A. Abanin, Criterion for Many-Body Localization-Delocalization Phase Transition, *Phys. Rev. X* **5**, 041047 (2015).
- [62] A. Pal and D. A. Huse, Many-body localization phase transition, *Phys. Rev. B* **82**, 174411 (2010).
- [63] V. Oganesyan and D. A. Huse, Localization of interacting fermions at high temperature, *Phys. Rev. B* **75**, 155111 (2007).
- [64] S. Vardhan, G. De Tomasi, M. Heyl, E. J. Heller, and F. Pollmann, Characterizing Time Irreversibility in Disordered Fermionic Systems by the Effect of Local Perturbations, *Phys. Rev. Lett.* **119**, 016802 (2017).
- [65] S. Aubry and G. André, Analyticity breaking and Anderson localization in incommensurate lattices, *Ann. Isr. Phys. Soc.* **3**, 18 (1980).
- [66] S. Iyer, V. Oganesyan, G. Refael, and D. A. Huse, Many-body localization in a quasiperiodic system, *Phys. Rev. B* **87**, 134202 (2013).
- [67] W. Rawat and Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, *Neural Comput.* **29**, 2352 (2017).
- [68] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, [arXiv:1409.4842](https://arxiv.org/abs/1409.4842).
- [69] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, 2016).
- [70] The target times are log-linear distributed.
- [71] S. Scherg, T. Kohlert, P. Sala, F. Pollmann, M. Bharath H., I. Bloch, and M. Aidelsburger, Observing non-ergodicity due to kinetic constraints in tilted Fermi-Hubbard chains, *Nat. Commun.* **12**, 4490 (2021).
- [72] This argument can work only if the localization length of the system is larger than the subsystem.
- [73] Fraction of correctly classified MBL =  $\frac{\# \text{MBL classified}}{\# \text{MBL}}$ .
- [74] Additionally, we exploit the fact that we can produce more cutouts from larger systems, simply by using subsystems  $(n_0, \dots, n_{\ell-1})$  for the first cutout, then  $(n_1, \dots, n_{\ell})$  for the second, and so on until we reach the end of the chain  $(n_{L-\ell}, \dots, n_{L-1})$ . In testing, all cutouts of one snapshot block are classified and labeled as one of the two phases. After going through all cutouts of one snapshot block of length  $L$ , one assigns the category to the whole snapshot block which was ascribed to a majority of its cutouts. We call this procedure a voting mechanism, since each classification of cutouts gets one vote, and the majority vote decides which label is assigned to the whole system.
- [75] A. Mordvintsev, C. Olah, and M. Tyka, Inceptionism: Going deeper into neural networks, Google AI Blog (2015).
- [76] M. Kiefer-Emmanouilidis, R. Unanyan, M. Fleischhauer, and J. Sirker, Slow delocalization of particles in many-body localized phases, *Phys. Rev. B* **103**, 024203 (2021).
- [77] M. Kiefer-Emmanouilidis, R. Unanyan, J. Sirker, and M. Fleischhauer, Bounds on the entanglement entropy by the number entropy in non-interacting fermionic systems, *SciPost Phys.* **8**, 83 (2020).
- [78] D. J. Luitz and Y. B. Lev, Absence of slow particle transport in the many-body localized phase, *Phys. Rev. B* **102**, 100202(R) (2020).
- [79] M. Lee, T. R. Look, S. P. Lim, and D. N. Sheng, Many-body localization in spin chain systems with quasiperiodic fields, *Phys. Rev. B* **96**, 075146 (2017).
- [80] H. Singh, B. Ware, R. Vasseur, and S. Gopalakrishnan, Local integrals of motion and the quasiperiodic many-body localization transition, *Phys. Rev. B* **103**, L220201 (2021).
- [81] For the details of the CNNs see the Appendix.
- [82] G. Casati and B. Chirikov, *Quantum Chaos: Between Order and Disorder* (Cambridge University Press, Cambridge, 1995).
- [83] V. Ros, M. Müller, and A. Scardicchio, Integrals of motion in the many-body localized phase, *Nucl. Phys. B* **891**, 420 (2015); [Corrigendum. **900**, 446 (2015)].
- [84] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [85] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Red Hook, NY, 2019), pp. 8026–8037.