Neural networks for modeling electron transport properties of mesoscopic systems

Kangyuan Li, Junqiang Lu,* and Feng Zhai†
Department of Physics, Zhejiang Normal University, Jinhua 321004, China

(Received 17 December 2019; revised 19 July 2020; accepted 27 July 2020; published 13 August 2020)

In this work, we train neural networks to predict transport properties of a quasi-one-dimensional tight-binding model with disordered on-site energies. In the case where the on-site energies of all sites in the scattering region pick up values randomly and independently from a given interval (model I), it is found that the performance of the neural network depends strongly on the system size. For a small system size, the neural network can predict accurately the conductance at the Fermi energy. With an increase in the system size, the mean absolute error (MAE) increases. This tendency persists under an increase in the number of training samples and the number of neurons in the neural network. The physical mechanism behind this phenomenon can be ascribed to the universal conductance fluctuation. To reduce the effective dimension of features fed into the neural network, we consider the case of substitutional doping where the impurities have a fixed on-site energy, random position distribution, and variable concentration (model II). It is found that even for a relatively large system size, the trained neural network in model II performs much better in predicting the conductance at the Fermi energy. When the prediction target is changed to the average conductance under a finite bias (average value of 10 conductances at energies near the Fermi energy), the MAE can be reduced by almost a half. The small fluctuation of the average conductance reduces further the prediction difficulty of the neural network. The trained neural network can predict quickly and accurately the disorder-averaged transport properties for model II.

DOI: 10.1103/PhysRevB.102.064205

I. INTRODUCTION

In recent years, machine learning has achieved great success in industrial applications. The application of neural network technology in some fields is more attractive, such as e-commerce website recommendation, image recognition, and speech recognition [1–4]. Its application in scientific research fields such as biology, chemistry, and physics is emergent. Machine learning methods including neural networks have been implemented to solve some inverse problems [5] in material design, i.e., predicting the structures of materials with a given physical property. Neural networks have been used to analyze experimental data in high-energy physics [6,7], identify phases and phase transitions in many-body systems [8–15], find the atomization energy, ionization potential, and excitation energy of different organic molecules from known molecular structure data [16–18], and capture slowly changing collective variables and long-time-scale kinetic features of biomolecular processes [19,20]. For a wide range of amino acid–based ionic liquids [21], robust and reliable models have been created by artificial neural networks to estimate their thermophysical properties. For a glassy system without any handcrafted feature, the graph neural network has been used to determine its long-time evolution solely from the initial particle positions [22]. In these developments, computational physicists have demonstrated that machine-learning techniques can be used to speed up the simulation of complex systems and circumvent expensive computational costs [23–26].

It is a challenging and interesting problem to use some information on transport systems (such as the electronic energy and molecular structure) as input in machine learning models to predict the conductance [27–29]. A ridge regression model with a standard Laplacian kernel has been trained to predict the transmission coefficient of a transport system based on some tight-binding parameters [27]. The trained model performs remarkably in capturing the complexity of interference phenomena. The support vector machine model has been trained to establish the connection between molecular structures and corresponding electrical conductances determined from scanning tunneling microscopy of molecular break junctions [28]. By means of proper input features, a shallow neural network trained on short DNA nanojunctions can accurately predict the electrical conductance of millions of long DNA sequences [29]. These studies show that machine learning can assist the study of transport properties.

For the transport properties of mesoscopic systems, disorder plays an important role, which can lead to the metal-insulator transition [30,31] and drive the topological phase transition in quantum spin-Hall systems [32], three-dimensional strong topological insulators and superconductors [33,34], fractional quantum Hall systems [35], and weak AIII topological insulators [36]. By means of controllable disorder realization in ultracold atoms, a topological Anderson insulator has been observed in disordered atomic wires [37]. In a 10-qubit network subject to engineered static disorder, a crossover from Anderson localization to environment-assisted quantum transport has been observed [38]. Both the energy

*jlu@zjnu.cn
†fzhai@zjnu.cn

and the concentration of the impurity have a rich influence on quantum transport. In some cases, their small changes can alter significantly the conductance values. Therefore, it is necessary to do a lot of numerical calculations to study the effects of the impurity energy and impurity concentration on transport properties [39–41]. It is desirable to reduce the computational cost of this aspect through the popular machine-learning technology. In Ref. [27], the number of parameters affecting transport properties of the disordered system is no more than five. It is necessary to examine the feasibility of utilizing machine learning methods to predict the conductance of transport systems with a large number of impurity points or with a tunable impurity concentration.

In this work, we use the neural network to study the conductance of quasi-one-dimensional quantum wires based on a tight-binding model with random on-site energies. Two kinds of impurity models are considered: (i) the all-impurity model (model I), where all sites in the scattering region have random on-site energies uniformly distributed over an interval; and (ii) the substitution-doping model (model II), where the impurities have a fixed on-site energy, random position distribution, and variable concentration. For model I, when the system size is small, the prediction of the neural network can approach a high accuracy. With increasing system size, the prediction accuracy will drop quickly even if the size of the training data set and neural network increases. The prediction fails for a moderate system size. For model II, even for a relatively large system, the mean absolute error (MAE) between the conductance predicted by the neural network and the true conductance can reach $0.2G_0$. Here $G_0 = 2e^2/h$ is the conductance quantum, e is the fundamental charge, and h is the Planck constant. If the prediction is made for the average conductance (mean of 10 conductances at energies closest to the Fermi energy), the MAE can decrease to $0.1G_0$. For model II, our trained neural network can predict the disorder-averaged transport properties quickly and accurately.

II. MODEL AND METHODS

A. Model and formula

The system under consideration is depicted in Fig. 1, which is a quasi-one-dimensional quantum wire of uniform width W. Its central scattering region has a length L and connects with the left and right semi-infinite leads. A first-neighbor tight-binding Hamiltonian is used to describe the motion of electrons in the system,

$$H = \sum_i \epsilon_i \hat{a}_i^+ \hat{a}_i + \sum_{\langle i,j \rangle} V_{ij} \hat{a}_i^+ \hat{a}_j, \tag{1}$$

where ϵ_i and $\hat{a}_i$ are the on-site energy and annihilation operator of an electron at site i, and V_{ij} is the hopping energy of electrons from site j to a nearest-neighbor site i. All hopping energies are fixed at $V_{ij} = -t$. We take t as the energy unit and use the lattice constant of the square lattice as the length unit. For each site in the left or right leads, the on-site energy is fixed at 0. In the scattering region with $L \times W$ lattice sites, the on-site energy varies randomly and independently.

At a given Fermi energy E_F and a given configuration of $L \times W$ on-site energies, the zero-temperature conductance of the considered system in the linear-response regime can be calculated from the Landauer-Büttiker formula,

$$G = G_0 T_{\mathrm{RL}}(E_F). \tag{2}$$

Here $T_{\mathrm{RL}}(E)$ is the sum of the probability that an electron with energy E is transmitted from each eigenmode of the left lead to each eigenmode of the right lead. We use the numerical package Kwant [42] to calculate the transmission coefficient $T_{\mathrm{RL}}(E)$. The current under a source-drain bias V is given by $I = \overline{G}V$, where the average conductance $\overline{G}$ is calculated from

$$\overline{G} = G_0 \int_{-0.5}^{0.5} T_{\mathrm{RL}}(E_F + \alpha eV) d\alpha. \tag{3}$$

Here we assume that the voltages applied to the source and drain are $V/2$ and $-V/2$ and bring only a change of the corresponding electrochemical potentials [43].

We consider two models for the on-site energy of impurity sites. In model I, the on-site energy at each site in the scattering region is randomly selected from the interval $[-w_{\mathrm{dis}}/2, w_{\mathrm{dis}}/2]$ with a uniform distribution. Here w_{dis} is the disorder strength, which is set at 4 without specification. A typical configuration of on-site energies for this model is shown in Fig. 1(a). In model II, the on-site energy at each site in the scattering region is randomly assigned as 5 or 0 with a

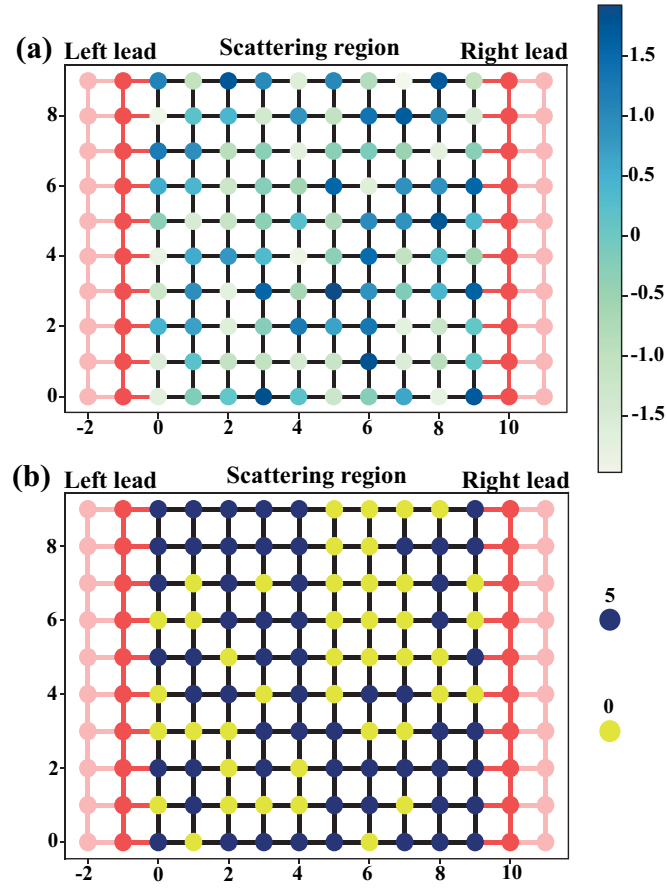

FIG. 1. (a) Model I: the on-site energy of each site in the scattering region is randomly selected from the interval $[-2, 2]$. (b) Model II: the impurities in the scattering region have a random position distribution, variable concentration, and fixed on-site energy 5.

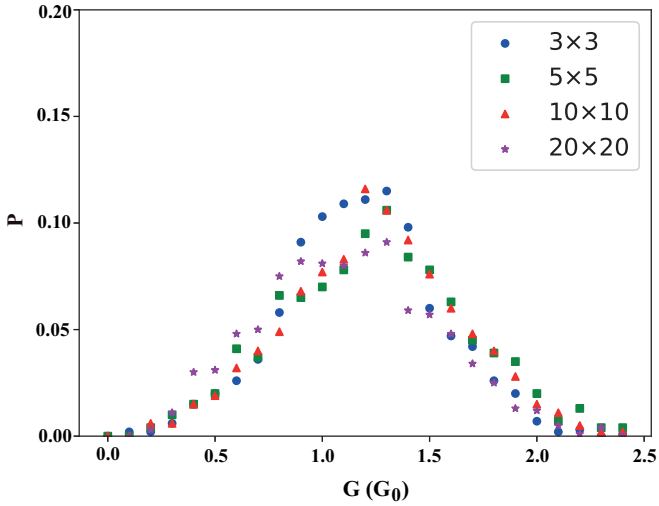

FIG. 2. Frequency distribution of conductance for the all-impurity model (model I) with system sizes 3×3, 5×5, 10×10, and 20×20. The on-site energy at each site in the scattering region takes random values in the interval $[-2, 2]$.

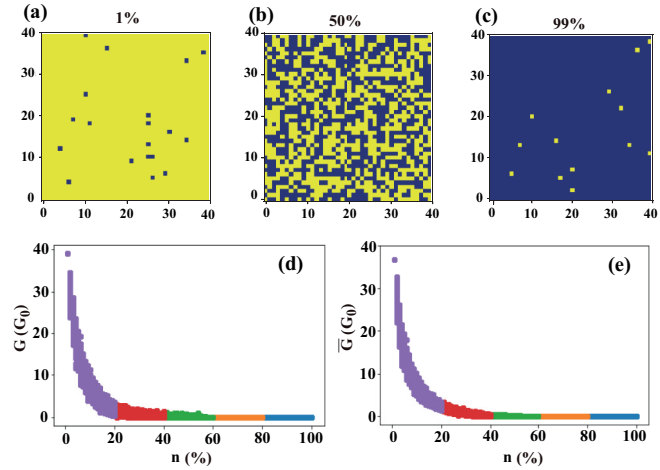

FIG. 3. Typical on-site energy configuration of model II with impurity concentrations (a) 1%, (b) 50%, and (c) 99%. The system size is 40×40 and the impurity sites are shown in blue. (d, e) Plot of the conductance G at the Fermi energy and the average conductance $\overline{G}$ for this system as a function of the impurity concentration n.

probability n or $1 - n$, where n is the impurity concentration. For this model, Fig. 1(b) presents a typical configuration of on-site energies.

B. System descriptor and training data set

An effective deep-learning model is closely related to the neural network architecture and the choice of descriptors. Descriptors have the most important impact on the effectiveness of neural networks. The adopted descriptors must be unique to each disordered transport system and must be closely related to the conductance [44]. We choose the on-site energies at all sites in the scattering region as the descriptor. This descriptor and the corresponding conductance constitute a sample in the training or testing data set. As shown in Figs. 1(a) and 1(b), each descriptor can be thought as a gray image, i.e., a grid of floating-point numbers.

For model I, we consider the system sizes 3×3, 5×5, 10×10, and 20×20, corresponding to 9, 25, 100, and 400 sites in the scattering region, respectively. For a given system size, the on-site energies of all sites in the scattering region are taken randomly and independently from the interval $[-2, 2]$. In this way, N_C configurations of on-site energies are generated independently at random. For each configuration of on-site energies, we calculate the conductance G at the Fermi energy $E_F = 0$ for the corresponding scattering problem. The frequency distribution of conductance is $P(G) = N_G/N_C$, where N_G is the number of configurations with conductance in the interval $[G - \bigtriangleup G,\ G + \bigtriangleup G]$ for $\bigtriangleup G = 0.1G_0$. As shown in Fig. 2, the frequency distribution of conductance for each system size is asymmetric around its peak and has a large peak width.

For model II, the system size is chosen as 40×40. The on-site energy is set at 5 for impurity sites and 0 otherwise. The 1600 sites in the scattering region are occupied by impurity randomly with a probability n. Here n is the impurity concentration. In Figs. 3(a)–3(c), we plot the configuration of on-site energies for the impurity concentrations 1%, 50%, and 99%, where the impurity sites are shown in blue. We generate N_C configurations of on-site energies by randomly and independently changing the impurity concentration and position and calculate the corresponding conductance spectrum. The label for each sample is chosen as either the conductance G at the Fermi energy $E_F = 0$ or the average conductance $\overline{G}$ at a fixed source-drain bias $V = 0.36\,t/e$. Here $\overline{G}$ in Eq. (3) is calculated as the mean of the 10 conductances at energies closest to the Fermi energy, $\overline{G} = G_0 \sum_{k=0}^{9} T_{\mathrm{RL}}[(-0.5 + k/9)eV + E_F]/10$. In Figs. 3(d) and 3(e), we plot the variation of G and $\overline{G}$ with the impurity concentration n. As the impurity concentration increases, both of them approach 0. Under a given n, G has a larger extension than $\overline{G}$.

C. Architecture of the adopted neural networks

The multilayer neural network architecture under consideration is constructed by Keras [45]. The basic units of a neural network are neurons. Neuron i receives an input vector $X^{(i)} = (x_1, x_2, x_3, \ldots, x_d)^T$, then generates an output vector $a_i(X^{(i)})$. Two steps are performed in neuron i. The first step is to make an affine transformation $Z^{(i)} = W^{(i)}X^{(i)} + b^{(i)}$, where $W^{(i)}$ and $b^{(i)}$ are the weight and bias. The second step is to feed $Z^{(i)}$ into the activation function $\sigma(x)$ to generate the neuron output $a_i(X^i) = \sigma(Z^{(i)})$. Each layer of a neural network is composed of many such neurons and the output of the previous layer serves as the input to the next layer [46–49]. For a sample with label k and the calculated conductance value G_k, the neural network takes its descriptor as the input signal and yields the prediction label y_k. The latter is a function of the weight and bias of each neuron. We consider the MAE and the mean square error (MSE) for the training data set or testing data set,

$$\mathrm{MAE} = \sum_{k \in S} \frac{|G_k - y_k|}{|S|}, \quad \mathrm{MSE} = \sum_{k \in S} \frac{(G_k - y_k)^2}{|S|}, \tag{4}$$

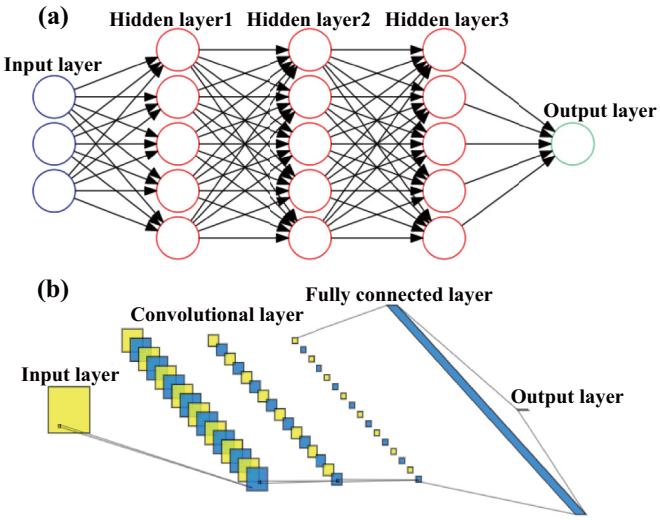

FIG. 4. (a) Architecture of a fully connected neural network with three hidden layers. (b) Architecture of a convolutional neural network with three convolutional layers and one fully connected layer. Each convolutional layer has 16 kernels of size 2×2 and stride 2.

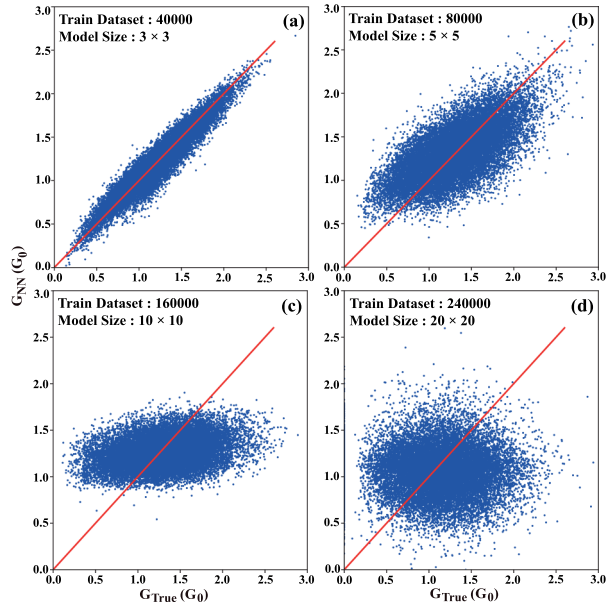

FIG. 5. Conductance value predicted by the DNN G_{NN} plotted as a function of the corresponding theoretical value G_{True} for all 20 000 testing samples. (a, b, c, d) The system size is 3×3, 5×5, 10×10, and 20×20, respectively; the number of training samples is 40 000, 80 000, 160 000, and 240 000; and each hidden layer in the DNN contains 20, 60, 150, and 600 neurons.

where $|S|$ is the number of samples in the data set S. When training the neural network, we take S as the training set and update the neuron weights and biases successively until the MSE (also called the loss) reaches the minimum. The optimization is performed by means of the Adam algorithm [50] at a learning rate 0.001. When using the neural network for prediction, we take S as the testing set and evaluate the accuracy by the corresponding MAE. A perfect neural network model can accurately predict the target values of all sample in the testing set, corresponding to MAE $= 0$.

For samples with a small number of features, we use a fully connected deep neural network [48] (DNN), whose architecture is shown in Fig. 4(a). The leftmost layer of the DNN is the input layer whose neurons accept feature values of samples. The rightmost layer of DNN is the output layer with only one neuron. Each middle layer of the DNN is called a hidden layer. The three hidden layers in Fig. 4(a) have the same numbers of neurons. For samples with a large feature space, the training of the DNN is computationally expensive. In this case, the convolutional neural network (CNN) is used; its architecture is shown in Fig. 4(b). The leftmost layer of the CNN is the input layer, which is a grid of floating-point numbers. The middle layers of the CNN are convolutional layers consisting of convolutional kernels. A proper size and stride of convolutional kernels are helpful to reduce the effective dimension of the feature space [49]. Each of the three convolutional layers shown in Fig. 4(b) has 16 convolutional kernels of size 2×2 and stride 2. A fully connected layer of width 1024 is connected to the final convolutional layer and the output layer with only one neuron. The activation function of all neurons in the middle layers and fully connected layers is selected as the rectified linear unit $\sigma(x) = \max(0, x)$.

III. RESULTS AND DISCUSSION

For the all-impurity model (model I), we present first the performance of the DNN in a small system, size 3×3. The training data set consists of 40 000 statistically independent samples and the testing data set contains 20 000 random samples. Each hidden layer in the DNN contains 20 neurons. For each sample, the conductance value at the Fermi energy predicted by the neural network (G_{NN}) is compared with the corresponding theoretical value (G_{True}). The prediction of the trained DNN on the testing data set has an MAE $\approx 0.1G_0$. In Fig. 5(a), every testing sample is represented by a blue point with a horizontal coordinate (G_{True}) and vertical coordinate (G_{NN}). The performance of the DNN is characterized by the average distance of all representative points to the straight red line $y = x$. It can be seen that the representative points are located densely around the line $y = x$. This indicates that the trained DNN can predict accurately the conductance of a small disordered system.

We then examine the performance of the neural network in predicting the conductance for a model system of a larger size. For systems of sizes 5×5, 10×10, and 20×20, the system parameters affecting transport properties are 25, 100, and 400. For the three values of system size, the number of samples in the training data set is chosen as 80 000, 160 000, and 240 000, while the number of neurons in the DNN is set at 60,150, and 600. The number of testing samples is fixed at 20 000. The representative points of testing samples are plotted in Figs. 5(b)–5(d). It is evident that more and more representative points are far away from the red line $y = x$ with increasing system size. The mean absolute error for the three system sizes are MAE $= 0.24G_0$, $0.31G_0$, and $0.38G_0$.

Although the number of training samples and neurons in the hidden layers increases almost linearly with the system size, the performance of the DNN in the conductance prediction becomes worse as the system size increases.

Such a system-size-dependent performance of the DNN in conductance prediction can be understood from the universal conductance fluctuation [51,52]. For an electron traversing the disordered system, there are many reflection paths. The probability waves propagating along different reflection paths interfere with each other, leading to universal conductance fluctuations. The number of reflection paths increases exponentially with the system size. With an increasing system size, the number of conductance patterns increases so quickly that the DNN with a limited number of neurons cannot predict accurately the conductance with a remarkable fluctuation. It is the complex interference behaviors of probability waves that makes the conductance prediction with the DNN extremely difficult.

We can also understand the performance of neural networks in the all-impurity model from the curse of dimensionality. For N_{train} training samples with a uniform random distribution, its coverage in the feature space of dimension $D = L \times W$ is estimated as $C \approx N_{\text{train}}(\varepsilon_S/w_{\text{dis}})^D$, where ε_S is the size of a box in the feature space occupied by a sample, which is controlled by the tolerance of conductance prediction and the conductance fluctation ΔG. In the regime of universal conductance fluctation [51], ΔG is a constant. For a fixed $\varepsilon_S/w_{\text{dis}} < 1$, the coverage C decreases exponentially with the dimension of feature space. The sparseness of training samples will make it difficult to modulate the parameters of neural networks correctly, leading to a poor performance. For the all-impurity model, the dimension of the feature space increases rapidly with the system size so that the neural network falls easily into the curse of dimensionality. In Appendix A, we demonstrate the influence of the DNN architecture and the number of training samples on the MAE. With an increase in the system size, the reduction of the converged MAE needs a more complicated DNN architecture (with more neurons) and a much larger training set (with exponentially increasing scale). We believe that the failure of neural networks in predicting the conductance for the large-scaled model I is an intrinsic problem of the method. As pointed out in Refs. [53] and [54], it is still difficult for neural networks to make a quantitative analysis or positive prediction in complicated physical systems.

In comparison with the results in Ref. [27], our model I with a larger system size has more parameters affecting the transport properties and thus the conductance prediction is more difficult. Our trained neural network captures the complexity of conductance fluctuations in disordered systems. The ridge regression method used in Ref. [27] assumes that the condutance curve consists of a series of peaks. Our DNN method removes this assumption. It performs much better than the ridge regression method when the number of training samples is sufficient. The disadvantages of the DNN method are that (i) the optimized architecture varies with the system size and (ii) the training set for a converged MAE is huge so that the training process is relatively time-consuming.

In the following we turn to model II, where the on-site energy of impurities is fixed at 5 and the system size is fixed at 40×40. The CNN is trained to predict the conductance at the Fermi energy by means of 80 000 training samples. The performance of the CNN is demonstrated by 20 000 testing samples. In Fig. 6, we plot the conductance G_{NN} predicted by the CNN for each testing sample as a function of its corresponding theoretical value G_{True}. The color scale of representative points indicates the absolute difference $|G_{\text{NN}} - G_{\text{True}}|$. It can be seen that for the 40×40 system of model II, the neural network performs much better in predicting the conductance at the Fermi energy than for that of model I. The reason is that the feature vector of samples in model II has much lower effective dimensions than that in model I, which reduces the difficulty of predicting the conductance by the neural network. As shown in Fig. 6(a), there are still many points where the predicted value has a relatively large absolute deviation from the theoretical value. The inset in Fig. 6(a) shows the frequency distribution P_{True} and P_{NN} of the theoretical conductance values and the predicted values for all testing samples. There is only a minor difference between P_{True} and P_{NN} except near the conductance region $[38, 40]\ G_0$. The prediction accuracy of the CNN can be improved by increasing the number of training samples N_{train}. For the given testing data set, Fig. 6(b) shows the MAE as a function of N_{train}. It can be seen that MAE/G_0 decreases gradually from 0.195 to 0.18 as N_{train} varies from 30 000 to 80 000.

We then change the target for prediction to the average conductance $\overline{G}$. The CNN architecture and the configurations of on-site energies in both the training and the testing data set are the same as in Fig. 6(a). The performance of the retrained CNN is shown in Fig. 6(c). In comparison with Fig. 6(a), the representative points of testing samples are concentrated more densely around the line $y = x$. Furthermore, the number of points with a large deviation $|G_{\text{NN}} - G_{\text{True}}|$ is obviously reduced. The improvement of the CNN performance is also

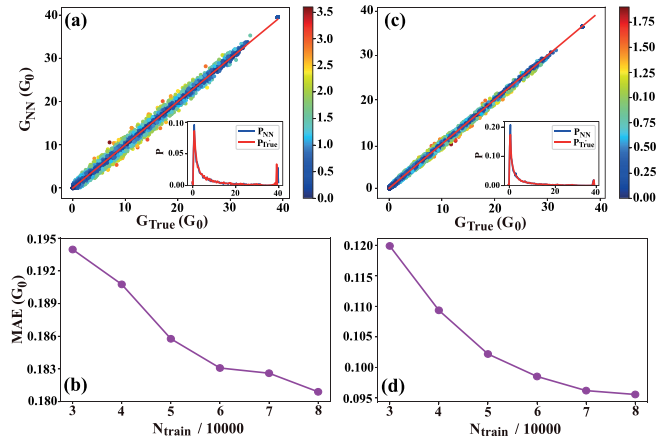

FIG. 6. (a) Scatter plot of representative points with coordinate $(G_{\text{True}}, G_{\text{NN}})$ of all testing samples. The color scale of any point indicates the absolute difference $|G_{\text{True}} - G_{\text{NN}}|$ between the theoretical value G_{True} of conductance at the Fermi energy and the corresponding predicted value G_{NN} from the CNN. (b) Mean absolute error (MAE) as a function of the number N_{train} of training samples. (c, d) The same as (a) and (b), but for predicting the average conductance $\overline{G}$. Insets in (a) and (c): Frequency distribution P_{True} and P_{NN} for the theoretical values and predicted values of conductance over all testing samples.

reflected in the frequency distribution P_{NN} of predicted values of the average conductance $\overline{G}$, which is shown in the inset in Fig. 6(c). The frequency distribution P_{True} of theoretical $\overline{G}$ values almost coincides with P_{NN} over the whole considered conductance region. As shown in Fig. 6(d), the MAE decreases almost exponentially with the number of training samples and converges more quickly than that in Fig. 6(b). For $N_{\mathrm{train}} = 80\,000$, the MAE$/G_0 \approx 0.095$ for the prediction target $\overline{G}$ is almost half that for the target G. The improvement of the prediction accuracy is due to the small fluctuation of the mean conductance $\overline{G}$, which further reduces the difficulty of conductance prediction for the neural network.

Finally, we examine the prediction ability and speed of the trained neural network in Fig. 6(c). For model II with a system size 40×40 and at a given impurity concentration n, we generate randomly $N_C = 1000$ on-site energy configurations and calculate the average conductance $\overline{G}_k$ for each configuration k. We define the ensemble-averaged conductance $\overline{G}_{\mathrm{ensemble}}$ and variance $\overline{G}_{\mathrm{variance}}$ of the system at this impurity concentration as

$$\overline{G}_{\mathrm{ensemble}} = \sum_{k=1}^{N_C} \overline{G}_k / N_C,$$

$$\overline{G}_{\mathrm{variance}} = \sqrt{\sum_{k=1}^{N_C} (\overline{G}_k - \overline{G}_{\mathrm{ensemble}})^2 / N_C}. \tag{5}$$

We use an Intel Core i5-7500 CPU with an eight-core processor working at 3.40 GHz to calculate $\overline{G}_{\mathrm{ensemble}}$ and $\overline{G}_{\mathrm{variance}}$ at a given concentration. The numerical calculation made directly by the Kwant package takes approximately 8885 s, while the time required to make a prediction by the trained neural network is approximately 32 s. The theoretical and predicted values of $\overline{G}_{\mathrm{ensemble}}$ and $\overline{G}_{\mathrm{variance}}$ are plotted in Fig. 7 as a function of the impurity concentration n. When the impurity concentration changes from 0% to 100%, the predicted value of $\overline{G}_{\mathrm{ensemble}}$ agrees well with the theoretical values. The predicted variance of the average conductance differs slightly from the corresponding theoretical values. This comparison demonstrates that the trained neural network can predict the disorder-averaged transport properties for the considered mesoscopic system quickly and accurately.

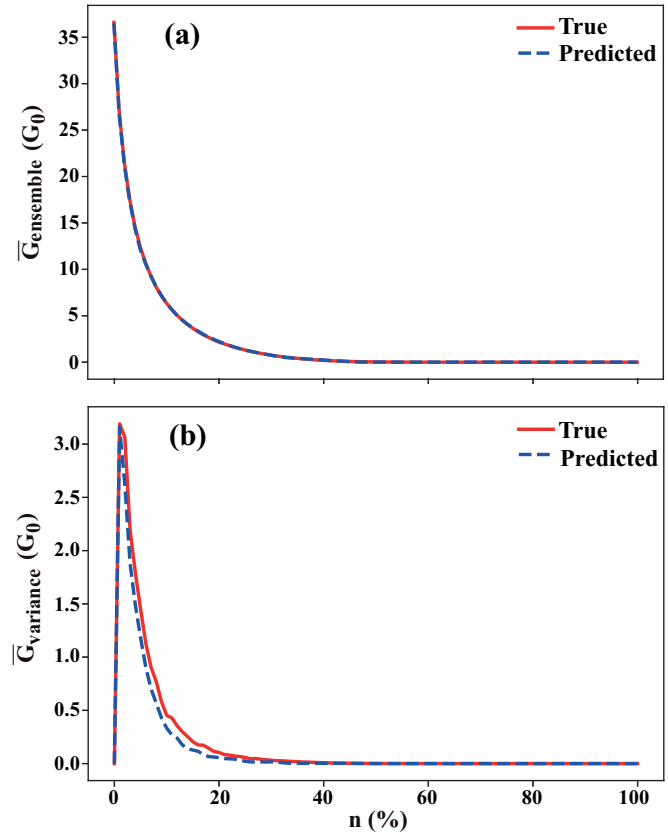

FIG. 7. (a) Ensemble-averaged conductance $\overline{G}_{\mathrm{ensemble}}$ and (b) variance $\overline{G}_{\mathrm{variance}}$ as a function of the impurity concentration n. The solid red line represents the results calculated directly from the Kwant package. The dashed blue line shows the results predicted by the same trained neural network as in Fig. 6(c).

IV. CONCLUSIONS AND REMARKS

In summary, we have trained neural networks to predict the electron transport properties of a disordered system which is described by a tight-binding model. For all sites in the scattering region of size $L \times W$, two types of disorders in the on-site energies are considered: (i) the on-site energies take values randomly and independently from a given interval (model I), and (ii) the on-site energies take one of two given values randomly (model II). For model I with a small system size (3×3), the trained neural network performs well in conductance prediction (MAE $\approx 0.1G_0$). With increasing system size, the MAE increases even under the enlargement of the training data set and the number of neurons in the neural network. It is the universal conductance fluctuation that results in the prediction failure of the neural network for model I with a large system size. For model II, the trained neural network performs much better even for a relatively large system size, 40×40, where MAE $\approx 0.2G_0$. If the conductance at a fixed bias (average of 10 conductances at energies closest to the Fermi energy) is selected as the prediction target, the MAE drops rapidly to $\approx 0.1G_0$. The trained neural network can predict the disorder-averaged transport properties for model II quickly and accurately.

The disordered system under consideration can be viewed qualitatively as a site percolation model [55] where each site belongs to one of N_c categories. All sites with an on-site energy in a given narrow region are in the same category. Model II is transformed into a site percolation with two categories ($N_c = 2$). The number of categories N_c in the site percolation corresponding to model I is large and depends on the disorder strength and system size. In Appendix B, we examine the effect of the number of categories (which is the number of impurity species plus 1) on the prediction accuracy of the neural network for a system of size 40×40. It is found that the MAE increases with N_c and has a weak dependence on the architecture of the neural network. This tendency partly reflects the difficulty of using the neural network to predict the conductance in model I.

Our work has revealed the bottleneck of neural networks in predicting the quantum conductance of a disordered

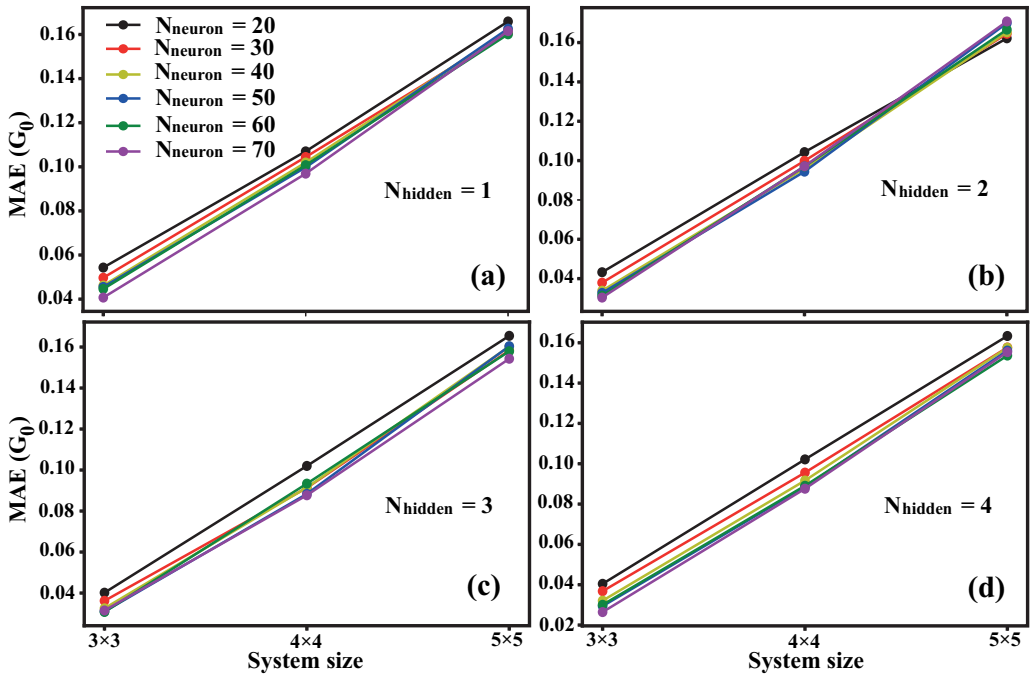

FIG. 8. Variation of the MAE with the system size of model I for a DNN under several values of N_{neuron}. (a) $N_{\text{hidden}} = 1$, (b) $N_{\text{hidden}} = 2$, (c) $N_{\text{hidden}} = 3$, and (d) $N_{\text{hidden}} = 4$.

mesoscopic system. For quasi-one-dimensional tight-binding systems of a relatively large size D, the performance of neural networks depends on the number (N_c) of different values of on-site energies (and the disorder strength). When $N_c \ll D$, the trained neural network shows a satisfactory accuracy and thus has great potential for conductance prediction. In the case of $N_c \sim D$, the performance of neural networks is poor for relatively large system sizes (and strong disorder) due to the presence of universal conductance fluctuation and the curse of dimensionality.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 11774314). F.Z. appreciates hospitality and stimulating discussions from Prof. Kai Chang at Institute of Semiconductors. Computational resources used in this work were provided by Tencent Cloud.

APPENDIX A: PERFORMANCE OF NEURAL NETWORKS IN THE ALL-IMPURITY MODEL

We take the DNN with $N_{\text{neuron}} \times N_{\text{hidden}}$ neurons as an example to demonstrate the curse of dimensionality. Here N_{hidden} and N_{neuron} are the number of hidden layers and the number of neurons in each hidden layer. For each considered system size, we randomly select 2000 samples as the testing set and set $w_{\text{dis}} = 2$.

First, the number of training samples is fixed at $N_{\text{train}} = 10\,000$. Variation of the MAE with the system size is plotted

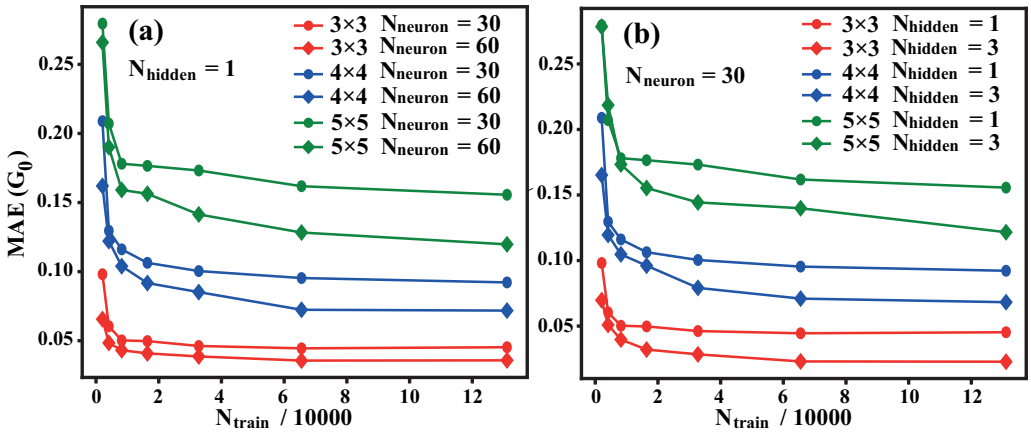

FIG. 9. Variation of the MAE with the number of training samples N_{train} for system sizes 3×3, 4×4, and 5×5. (a) $N_{\text{hidden}} = 1$ and $N_{\text{neuron}} \in \{30, 60\}$. (b) $N_{\text{neuron}} = 30$ and $N_{\text{hidden}} \in \{1, 3\}$.

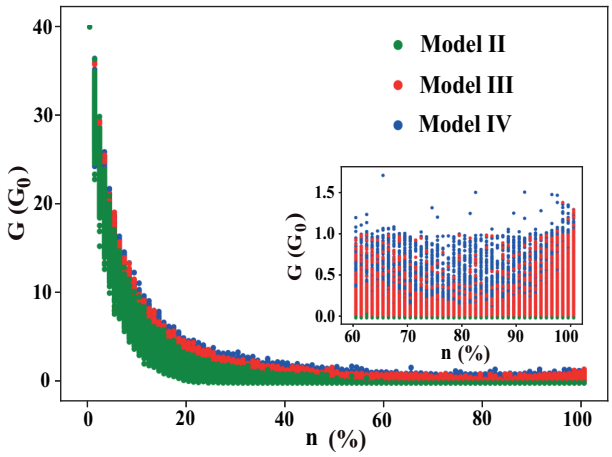

FIG. 10. Variation of the conductance G of models II, III, and IV with impurity concentration n. Inset: Enlarged section where n ranges from 60% to 100%.

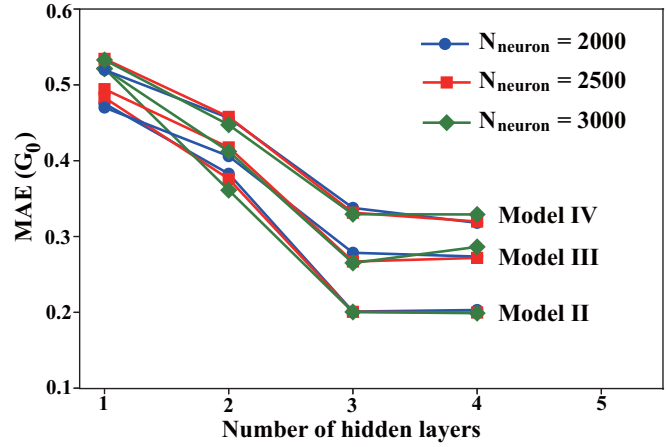

FIG. 11. Variation of the MAE with the number of hidden layers N_{hidden} of a DNN for models II, III, and IV. The number of neurons in each hidden layer is set at $N_{neuron} = 2000$, 2500, and 3000.

in Fig. 8 under several values of N_{neuron} and N_{hidden}. Usually, the MAE decreases with N_{neuron}. Under $N_{hidden} = 4$ [Fig. 8(d)], one can observe the convergence tendency of the MAE with an increase in N_{neuron}. Obviously, the value of the MAE increases quickly with the system size. This fact tells us that under a fixed N_{train} and various architectures of the DNN, the performance of the DNN becomes worse and worse as the system size increases. For a given system size and fixed N_{train}, we also observe that the MAE increases with the disorder strength (not shown here).

Next we examine the variation of the MAE with $N_{train} = 2^i$ $(11 \leqslant i \leqslant 18)$ under different DNN architecture parameters, which is shown in Fig. 9. Under $N_{hidden} = 1$ and $N_{neuron} = 30$, MAE/G_0 for the 3×3 system approaches a constant 0.04 under $N_{train} = 2^{16}$ [see Fig. 9(a)]. For the 4×4 system, MAE/G_0 tends to the value 0.09 under $N_{train} = 2^{18}$. It is difficult to lower the MAE for the 4×4 system to that for the 3×3 system only by means of increasing N_{train}. A remarkable reduction in the MAE can be realized by increasing N_{neuron} [see Fig. 9(a)] or N_{hidden} [see Fig. 9(b)]. For a DNN with a larger number of neurons, one usually needs more training samples to ensure convergence of the MAE. With the optimized architecture of the DNN for each system size, the MAE still increases quickly with the system size, as shown in Fig. 5. As the dimension of the feature space increases, the number of required training samples with coverage near 1 will increase exponentially. For the simplest 1×1 system, one can reduce MAE/G_0 to 0.006 by means of the DNN with 1×5 neurons and $N_{train} = 2^7$ training samples. For the 1×2 system, to reduce MAE/G_0 to the same value, 0.006, one needs a DNN with 1×12 neurons and $N_{train} = 2^{16}$ training samples.

APPENDIX B: MODELS WITH SEVERAL IMPURITY SPECIES

The neural network trained for model II has a satisfactory prediction accuracy. In model II, only one kind of impurity is considered. One may wonder how the number of impurity species affects the performance of the neural network. To this end, we design model III with two impurity species and model IV with three impurity species. In model III, the on-site energy of impurities is set at either 5/2 or 5 with probability 1/2. In model IV, the on-site energy of impurities picks up a value from $\{5/3, 10/3, 5\}$ with probability 1/3. The system size is fixed at 40×40.

In Fig. 10 the linear-response conductance G is plotted as a function of the impurity concentration n for models II–IV. It can be seen that at most values of n the fluctation of G increases with the number of impurity species. As n varies from 60% to 100%, the fluctation of G for model III is much larger than that for model II and is comparable to that for model IV. It is evident that the model complexity depends on the number of impurity species.

The DNN with N_{hidden} hidden layers is trained to predict the conductance of models II–IV. All hidden layers have the same number of neurons N_{neuron}. For each model, there are 80 000 training samples and 20 000 testing samples. For each model, the MAE of conductance prediction is plotted in Fig. 11 as a function of N_{hidden} under $N_{neuron} = 2000$, 2500, and 3000. One can observe that the MAE depends on both N_{hidden} and

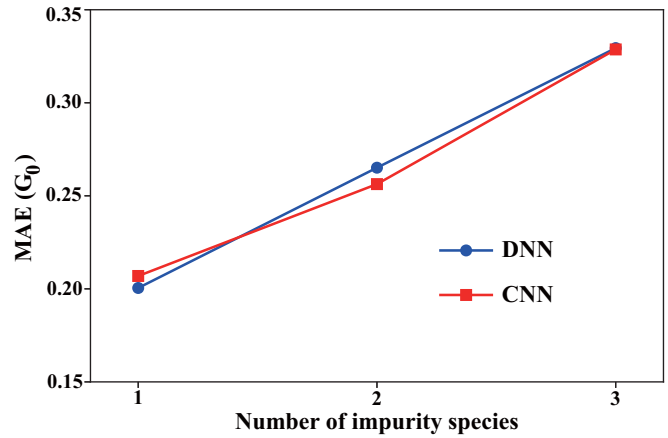

FIG. 12. Variation of the MAE with the number of impurity species for a DNN and a CNN.

N_{neuron}. For $N_{\text{neuron}} = 3000$, the neural network with three hidden layers provides a nearly optimal performance. The converged MAE for models II, III, and IV are, respectively, $0.2G_0$, $0.26G_0$, and $0.32G_0$. This indicates that the prediction accuracy of the DNN decreases with the number of impurity species.

As shown in Fig. 12, the tendency of the MAE to increase with the number of impurity species persists when the convolutional neural network is adopted instead to predict the conductance of the three models. Here the CNN has three convolutional layers and 16 convolutional kernels of size 2×2 and stride 2. Its performance is compared with that of the DNN with $N_{\text{hidden}} = 3$ and $N_{\text{neuron}} = 3000$. Both neural networks have almost the same prediction accuracy for the three models. The training cost of the CNN is much lower than that of the DNN.

[1] C. Farabet, C. Couprie, L. Naiman, and Y. L. Cun, IEEE Trans. Pattern Anal. Machine Intell. **35**, 1915 (2013).
[2] Y. L. Cun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
[3] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, and N. Jaitly, IEEE Signal Process. Mag. **29**, 82 (2012).
[4] T. Mikolvo, A. Deoras, D. Povey, L. Burget, and J. Cernocky, in *IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU), 2011* (IEEE, New York, 2011), p. 196.
[5] V. Mlinar, Ann. Phys. (Berlin) **527**, 187 (2015).
[6] P. Baldi, P. Sadowski, and D. Whiteson, Nat. Commun. **5**, 4308 (2014).
[7] S. Whiteson and D. Whiteson, Eng. Appl. Artif. Intell. **22**, 1203 (2009).
[8] L. Wang, Phys. Rev. B **94**, 195105 (2016).
[9] J. Carrasquilla and R. G. Melko, Nat. Phys. **13**, 431 (2017).
[10] E. P. L. Nieuwenburg, Y.-H. Liu, and S. D. Huber, Nat. Phys. **13**, 435 (2017).
[11] K. Chng, J. Carrasquilla, R.-G. Melko, and E. Khatami, Phys. Rev. X **7**, 031038 (2017).
[12] P.-F. Zhang, H.-T. Shen, and H. Zhai, Phys. Rev. Lett. **120**, 066401 (2018).
[13] T. Mano and T. Ohtsuki, J. Phys. Soc. Jpn. **86**, 113704 (2017).
[14] W.-J. Rao, Z.-Y. Li, Q. Zhu, M.-X. Luo, and X. Wan, Phys. Rev B **97**, 094207 (2018).
[15] P. Suchsland and S. Wessel, Phys. Rev. B **97**, 174435 (2018).
[16] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).
[17] G. Montavon, M. Rupp, V. Gobre, A. V. Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. Lilienfeld, New J. Phys. **15**, 095003 (2013).
[18] R. M. Balabin and E. I. Lomaking, J. Chem. Phys. **131**, 074104 (2009).
[19] C. Wehmeyer and F. Noe, J. Chem. Phys. **148**, 241703 (2018).
[20] A. Mardt, L. Pasquali, H. Wu, and F. Noe, Nat. Commun. **9**, 5 (2018).
[21] J. C. Cancilla, A. Perez, K. Wierzchos, and J. S. Torrecilla, Phys. Chem. Chem. Phys. **18**, 7435 (2016).
[22] V. Bapst, T. Keck, A. Grabska-Barwińska *et al.*, Nat. Phys. **16**, 448 (2020).
[23] K. Mills, M. Spanner, and I. Tamblyn, Phys. Rev. A **96**, 042113 (2017).
[24] S. Pilati and P. Pieri, Sci. Rep. **9**, 5613 (2019).
[25] N. Sun, J.-M. Yi, P.-F. Zhang, H.-T. Shen, and H. Zhai, Phys. Rev. B **98**, 085402 (2018).
[26] L.-F. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, Phys. Rev. B **90**, 155136 (2014).
[27] A. Lopez-Bezanilla and O. A. von Lilienfeld, Phys. Rev. B **89**, 235411 (2014).
[28] N. A. Lanzillo and C. M. Breneman, J. Appl. Phys. **120**, 134902 (2016).
[29] R. Korol and D. Segal, J. Phys. Chem. B **123**, 2801 (2019).
[30] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[31] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979).
[32] J. Li, R.-L. Chu, J. K. Jain, and S.-Q. Shen, Phys. Rev. Lett. **102**, 136806 (2009).
[33] S. Ryu and K. Nomura, Phys. Rev. B **85**, 155138 (2012).
[34] M. Brahlek, N. Koirala, M. Salehi, J. Moon, W. Zhang, H. Li, X. Zhou, M. G. Han, L. Wu, T. Emge *et al.*, Phys. Rev. B **94**, 165104 (2016).
[35] W. Zhu and D. N. Sheng, Phys. Rev. Lett. **123**, 056804 (2019).
[36] J. Claes and T. L. Hughes, Phys. Rev. B **101**, 224201 (2020).
[37] E. J. Meier, F. Alex An, A. Dauphin, M. Maffei, P. Massignan, T. L. Hughes, and B. Gadway, Science **362**, 929 (2018).
[38] C. Maier, T. Brydges, P. Jurcevic, N. Trautmann, C. Hempel, B. P. Lanyon, P. Hauke, R. Blatt, and C. F. Roos, Phys. Rev. Lett. **122**, 050501 (2019).
[39] L. Sheng, D. N. Sheng, and C. S. Ting, Phys. Rev. Lett. **94**, 016602 (2005).
[40] M. Onoda, Y. Avishai, and N. Nagaosa, Phys. Rev. Lett. **98**, 076802 (2007).
[41] D. N. Sheng, Z. Y. Weng, L. Sheng, and F. D. M. Haldane, Phys. Rev. Lett. **97**, 036808 (2006).
[42] C. W. Groth, M. Wimmer, A. R. Akhmerov, and X. Waintal, New J. Phys. **16**, 063065 (2014).
[43] J. B. Hong, J. Q. Lu, and F. Zhai, 2D Mater. **6**, 045053 (2019); B. Gaury, J. Weston, M. Santin, M. Houzet, C. Groth, and X. Waintal, Phys. Rep. **534**, 1 (2014).
[44] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, Phys. Rev. Lett. **114**, 105503 (2015).
[45] F. Chollet, Keras, https://github.com/fchollet/keras.
[46] P. Methta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schead, Phys. Rep. **810**, 1 (2019).
[47] N. Kriegeskorte and T. Golan, Curr. Biol. **29**, R231 (2019).
[48] J. Schmidhuber, Neural Netw. **61**, 85 (2015).
[49] M. A. Arbib, *The Handbook of Brain Theory and Neural Networks* (MIT Press, Cambridge, MA, 2003), p. 276.
[50] D. P. Kingma and J. Ba, arXiv:1412.6980 (2014); in *Proceedings of the 3rd International Conference on Learning Representations* (Ithaca, NY: arXiv.org, San Diego, CA, 2015).
[51] P. A. Lee and A. D. Stone, Phys. Rev. Lett. **55**, 1622 (1985).
[52] P. A. Lee, A. D. Stone, and H. Fukuyama, Phys. Rev. B **35**, 1039 (1987).
[53] A. Zunger, Nature **566**, 447 (2019).
[54] H. Theveniaut and F. Alet, Phys. Rev. B **100**, 224202 (2019).
[55] D. J. Thouless, Phys. Rep. **13**, 93 (1974).