

## Structure of the impurity band in heavily doped nonmagnetic semiconductors

Hongwei Chen<sup>1,2,3,4</sup> and Zi-Xiang Hu<sup>4,\*</sup>

<sup>1</sup>*Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA*

<sup>2</sup>*Stanford Institute for Materials and Energy Sciences, Stanford University, Stanford, California 94305, USA*

<sup>3</sup>*Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, California 94720, USA*

<sup>4</sup>*Department of Physics and Chongqing Key Laboratory for Strongly Coupled Physics, Chongqing University, Chongqing 401331, People's Republic of China*



(Received 18 January 2023; accepted 12 April 2023; published 25 April 2023)

We study the properties of the impurity band in heavily doped nonmagnetic semiconductors using the Jacobi-Davidson algorithm and the supervised deep learning method. The disorder averaged inverse participation ratio and Thouless number calculation show us the rich structure inside the impurity band. A convolutional neural network model, which is trained to distinguish the extended/localized phase of the Anderson model with high accuracy, shows us the results in good agreement with the conventional approach. Together, we find that there are three mobility edges in the impurity band for a specific on-site impurity potential, which means the presence of the extended states while filling the impurity band.

DOI: [10.1103/PhysRevB.107.134204](https://doi.org/10.1103/PhysRevB.107.134204)

### I. INTRODUCTION

The effect of disorder has been extensively studied since Anderson's seminal paper [1]. Diluted magnetic semiconductors (DMS) doped with a small concentration of charged impurities constitute an interesting magnetic system that has a number of novel features for study by numerical simulation [2]. Much of the research has been focused on II-VI (such as CdTe or ZnSe) and III-V (such as GaAs) compound semiconductors doped with a low concentration ( $x \sim 1\text{--}8\%$ ) of manganese (Mn) impurities. Of particular interest in this field is  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  which has been shown to exhibit ferromagnetic behavior above 100 K [3]. In these samples, the manganese is substituted with gallium and acts as an acceptor (donating one hole to the crystal), so that the material is  $p$  type. The holes bind to the impurities with an energy of around 130 meV around  $x \sim 10\%$  [4]. Since  $x \ll 1$ , the overlap between different impurity states can be ignored, thus the interaction between the charge carriers can be neglected. The system can be simply described by a noninteracting tight-binding model. When the system contains only one impurity, and the binding energy is large enough, an impurity state appears below the conduction band (we assume the impurity potential is attractive). It is locally distributed in space near the impurity potential within a localization length  $\zeta$ . As increasing the concentration  $x$ , the overlap between different impurity states extends the single impurity energy to an impurity band in the density of state (DOS) and eventually merges with the conduction band. Simultaneously, the states in the impurity band are expected to become more and more extended and ultimately regain their bandlike character [5]. However, the details inside the impurity band are rarely studied.

One reason for lacking such a study is the computation difficulty even in the noninteracting case. Generally, the percentage of the state in the impurity band in the total number of states is about 10% at the concentration we are interested in. Taking a three-dimensional Anderson model with lattice size  $30 \times 30 \times 30$  as an example, the number of states which we need to know in the impurity band is about 3000. The exact diagonalization [6] for such a system is very difficult due to the large dimension. On the other hand, we have to do a large number of sample averages. The sparse matrix diagonalization, such as the Lanczos method [7], can be adapted to obtain a few lowest-lying states or a few states nearby special energy [the simplest way is diagonalizing  $(H - \epsilon I)^2$  by using the original Lanczos diagonalization method].

Machine learning methods have recently emerged as a valuable tool to study the quantum many-body physics problems [8–25]. Its ability to process high-dimensional data and recognize complex patterns has been utilized to determine phase diagrams and phase transitions [26–37]. In particular, the convolutional neural network (CNN) [38] model, which initially is designed for image recognition, was widely used to study different kinds of phase transition problems including the Bose-Hubbard model [39], spin-1/2 Heisenberg model [40], quantum transverse-field Ising model [35], etc. The power of using machine learning to recognize quantum states lies in their ability to finish tasks without the knowledge of physics background or the Hamiltonian of the system. Even if the neural network is trained in a small energy region of the system, it can be used to obtain the whole phase diagram [27,29]. Also, it can discriminate quantum states with high accuracy even if they are trained from a totally different Hamiltonian. This special feature of machine learning inspires us to try to identify the delocalized states in the impurity band.

In this paper, we develop a method to obtain the correct DOS and other localization properties, such as inverse

\*zxhu@cqu.edu.cn

participation ratio (IPR) [41] and Thouless number [42], by using Jacobi-Davidson sparse matrix diagonalization [43] with an importance sampling statistics method. Meanwhile, we train a three-dimensional CNN model using the data generated from the Anderson model, and then the trained model is used to identify the existence of extended states in the impurity band. This paper is organized as follows: In Sec. II we describe the tight-binding model on the cubic lattice and numerical methods; Sec. III demonstrates the effect of heavy doping studied by studying the IPR and Thouless number; Sec. IV demonstrates the implementation of the deep learning approach and the results from the trained neural network model. Finally, we close with a conclusion.

## II. MODEL AND METHODS

We consider a tight-binding model on a  $D$ -dimensional hypercubic lattice with the nearest-neighbor hopping  $t$ , and on-site energies  $\epsilon_i$ :

$$H = -t \sum_{\langle i,j \rangle} (\hat{c}_j^\dagger \hat{c}_i + \text{H.c.}) + \sum_i \epsilon_i \hat{c}_i^\dagger \hat{c}_i. \quad (1)$$

The hopping term simulates the itinerant electrons and the on-site energy has a bimodal distribution  $\epsilon_i = -W$  with probability  $x$ , and  $\epsilon_i = 0$  with probability  $(1-x)$ . This model has a host lattice with a single relevant band, with a fraction  $x$  of substitutional impurities. For one-dimensional ( $d = 1$ ) free electrons, the energy-momentum dispersion relation is  $E(k) = 2t \cos(k)$ , and it is easy to get the DOS with the formula

$$\rho(E) = \left( \frac{1}{2\pi} \right)^d \int \frac{dS}{\nabla_k E}. \quad (2)$$

The result for one dimension (1D) is

$$\rho_{1d}(E) = \frac{1}{\sqrt{4t^2 - E^2}}. \quad (3)$$

There is no analytic solution for higher dimensional systems; however, an approximation that is accurate to roughly 2% was given by Andres *et al.* [44]. Instead, the DOS can be calculated numerically by exact diagonalization as shown in Fig. 1 where  $t$  has been set to the unit. After introducing the impurities, all states become localized in 1D and two dimensions (2D) based on the scaling theory of localization [45]. In three dimensions, part of the states become localized and develop into an impurity band at the edge of the conducting band. To determine the localized/extended state, namely, the location of the mobility gap, we calculate the IPR [41],

$$\text{IPR} = \frac{\sum_i |\psi_i^4|}{(\sum_i |\psi_i^2|)^2} \quad (4)$$

for each state, where the  $\psi_i$  is the weight of an eigen wave function on the  $i$ th site. Heuristically, if we compare two trivial states with wave functions for a  $N$ -site system:

$$\Psi_{\text{extended}} = \sum_i (\psi_i = 1/\sqrt{N}) = \sqrt{N} \quad (5)$$

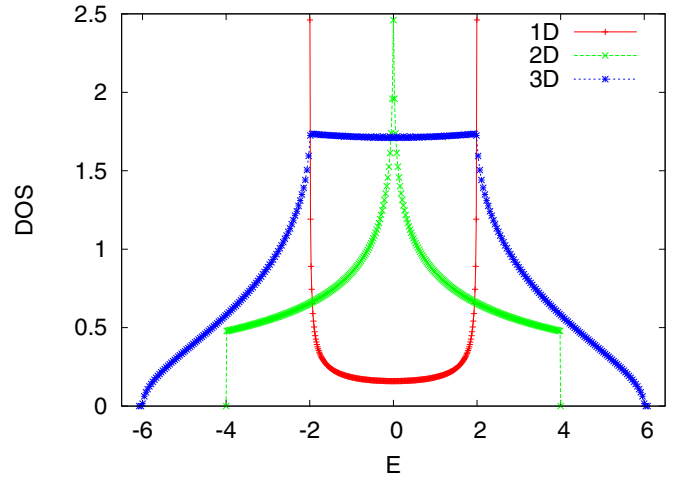


FIG. 1. The density of states for free electrons in the tight-binding model in one, two, and three dimensions. Here  $t$  has been set as the unit.

and

$$\Psi_{\text{localized}}(j) = \sum_i (\psi_i = \delta_{ij}) = 1, \quad (6)$$

where  $\Psi_{\text{extended}}$  is an extended state which has equal weight on each site and  $\Psi_{\text{localized}}(j)$  is a localized state which only has weight on the  $j$ th site. It is easy to see that the IPR of  $\Psi_{\text{extended}}$  decreased with the order of  $\frac{1}{N}$  and a constant for  $\Psi_{\text{localized}}(j)$ . On the other hand, the Thouless number [42] is defined as

$$g(E) = \frac{\langle |\Delta E| \rangle}{\langle \delta E \rangle}, \quad (7)$$

where  $\delta E$  is the energy difference while the boundary condition changes from periodic boundary condition to antiperiodic boundary condition and the  $|\Delta E|$  is the average energy distance around  $E$ . Since only the extended states are sensitive to the change of boundary condition,  $g(E)$  grows linearly as a function of the system size for the extended state, and conversely, it reduces for the localized state. In this work, we determine the localization properties by systematically studying the IPR and Thouless number for different system sizes, and the crossover points of the Thouless number give us a hint of the mobility edge.

For a three-dimensional cubic lattice of size  $L$ , the Hamiltonian matrix has a dimension of  $L^3$ . General full exact diagonalization methods, such as Lapack library [46], can only deal with small system sizes. The computation time of diagonalizing one matrix with size  $L^3$  grows dramatically as a function of the system size. According to Ref. [5], while the doping concentration  $x > 5\%$ , the maximum height and breadth of the impurity band DOS seems to saturate, thus we treat 5% as the typical value of the heavily doped concentration. As shown in Fig. 2, we deal with a system with size  $19 \times 20 \times 21$  with doping concentration  $x = 5\%$ , after averaging thousands of samples, we obtained the DOS for different doping energies. It is shown that a peak emerges gradually near the band edge as increasing the doping energy  $W$ . This peak becomes more prominent around  $W \sim 4.5$  at

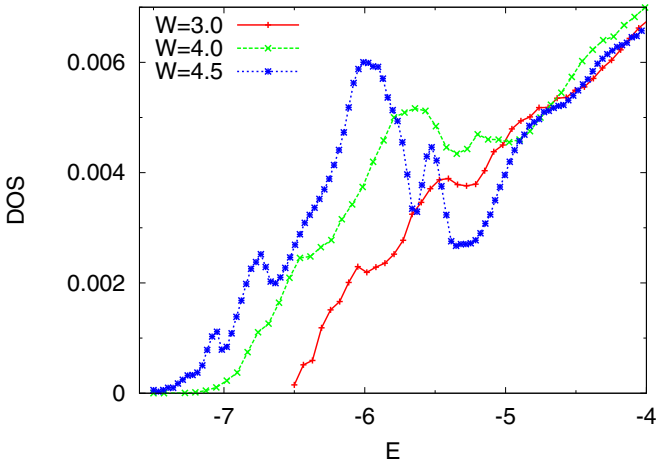


FIG. 2. The evolution of the DOS at the band edge with different doping strengths. The system has size  $19 \times 20 \times 21$  and can be fully diagonalized.

which an obvious depletion is developed at the junction between the impurity band and the conduction band.

Since only the developed impurity band is the interesting part, we are focusing on that. The number of states in the impurity band is about the lowest 10% of states in the whole band, thus we do not have to fully diagonalize the Hamiltonian. On the other hand, we just need to calculate the DOS, IPR, and Thouless number for those lowest 10% states after averaging thousands of samples. For our needs, we use the sparse matrix diagonalization with the Jacobi-Davidson (JADA) method [43] which can search a few (10–20) states efficiently around a specific energy. For a given sample at fixed doping strength, we randomly distribute the reference points (30–50 points) in the impurity band; taking  $W = -4.5$  as an example, the reference points are picked randomly in the region  $[-8: -4]$ ; about 10–20 states can be obtained by JADA around each reference point. The reference points could also be picked by importance sampling based on the DOS for a small system from the full diagonalization. We collect all these energies for each reference point in one sample. After thousands of sample averaging, we obtain the same DOS as that from the full diagonalization for a small system. It is obvious that the JADA method can easily go beyond the limit of the full exact diagonalization. At least on the same price of the computation time, we can nearly double the system size compared to the Lapack method. In this work, we calculate the properties for system sizes up to  $40^3$  sites by using the JADA method.

### III. THE EFFECT OF HEAVY DOPING

As analyzed in the previous section, with typical doping concentration  $x = 5\%$ , we find that a clear impurity band in the DOS is developed at about  $W = -4.5$ . We plot the DOS and IPR together for different system sizes as shown in Fig. 3. The line of the DOS for different system sizes collapses to a single curve and it is the same as that from exact diagonalization (ED) as shown in Fig. 2, which tells us that we have already obtained the essential information of the

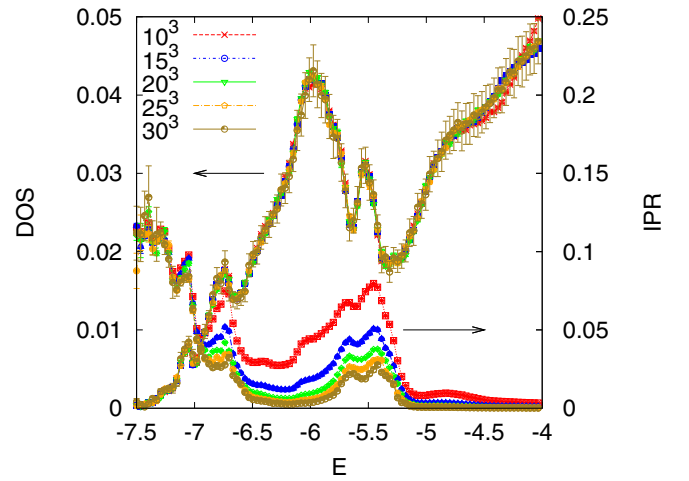


FIG. 3. The DOS and IPR for the 5% doped system with  $W = -4.5$ . The results are obtained from exact diagonalization. The number of configurations ranges from 1000 for system  $14 \times 15 \times 16$  to 50 for  $29 \times 30 \times 31$ . The DOS is almost system-size independent. The IPR drops in the center of the impurity band.

impurity band. As increasing the system size, the IPR does not change on the edge of the band, which means the states on the edge of the whole band are localized. The IPR in the bulk decreases as enlarging the system; especially at the center of the impurity band ( $E \sim -6.2$ ), the IPR drops to zero, which is the same as in the system bulk ( $E \sim -4.0$ ). However, there is a small peak near  $E \sim -5.5$ , which is at the right edge of the impurity band. The IPR in the vicinity of this point tends to saturate to a finite value as increasing the system size. The nonzero saturation of the IPR at this energy means another possible mobility edge existing near the junction between the conduction band and the impurity band.

In order to justify our conjecture, we systematically study the value of the IPR for several system sizes. As shown in Fig. 4(a), we choose four points from the knowledge of the DOS and IPR. (1)  $E = -4.2$  is in the bulk of the conduction band, at which the state is extended. (2)  $E = -5.4$  is at the right edge of the impurity band. The state here is localized according to our conjecture. (3)  $E = -6.2$  is in the bulk of the impurity band, which is extended according to its zero IPR value in large  $L$  limit. (4)  $E = -6.8$  is on the left edge of the impurity band and thus at the edge of the whole energy band. The state at the band edge is supposed to be localized. In Fig. 4(b) we again compare the DOS from JADA with that from Lapack, which shows a convergence in large system size. According to the way of choosing these four points, (1) and (3) should have similar behavior as increasing the system size, and vice versa for (2) and (4). Figure 4(c) shows the IPR for these four energies in different system sizes. We plot the data in logarithmic scale and fit it by function

$$\ln(\text{IPR}) = A + B \ln(L) + C \ln(L)^2. \quad (8)$$

The sign of the curvature  $C$  tells us whether the state is localized or not. For (1) and (3),  $C < 0$  means they are extended states, and oppositely  $C > 0$  for localized states at points (2) and (4).

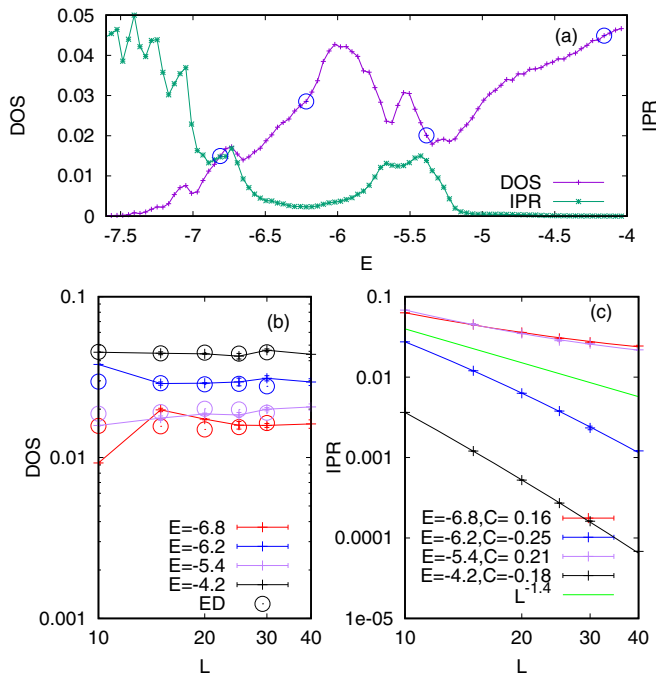


FIG. 4. The IPR/DOS as a function of system size for fixed energies. In (c), we fit the data by using a function  $\ln(\text{IPR}) = A + B \ln(L) + C \ln(L)^2$ . The curvature  $C$  is labeled in the figure.

As another criterion, we calculate the Thouless number  $g(E)$  for different system sizes. The results are shown in Fig. 5 in which we plot the DOS together with the same horizontal axis. The impurity band has been divided into several regions at the crossover of  $g(E)$  for different sizes. We label these regions by “L” (localized) and “E” (extended) to demonstrate different behavior  $g(E)$ . As increasing the system size, it is obvious that the  $g(E)$  increases in the E region and decreases in the L region. The energies with vertical lines are the locations

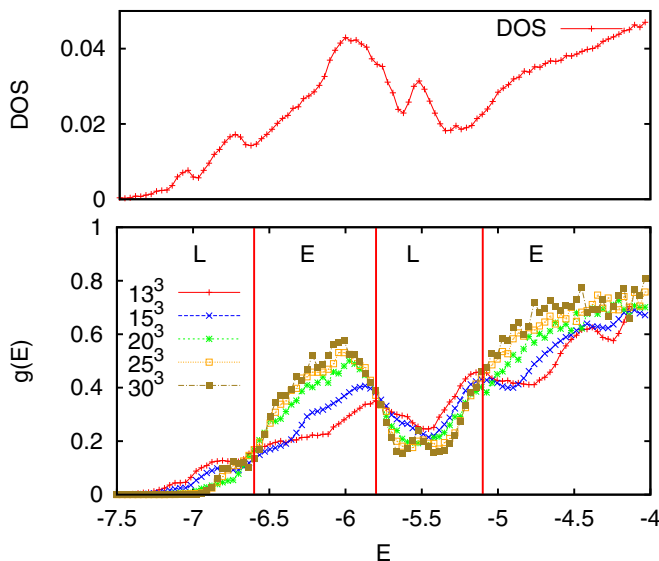


FIG. 5. The Thouless number  $g(E)$  as a function of energy for finite systems using exact diagonalization.

of the mobility edges, or the boundaries between the localized states and the extended states.

#### IV. DEEP LEARNING APPROACH

The CNN, which is originally designed for two-dimensional image recognition, has been widely adopted in studying phase transition and achieves high accuracy in recognition. A standard image recognition model can be used for a three-dimensional (3D) electron system by integrating the 3D electron density in one direction. But the drawback of this approach is that the information of the electron density along one direction is lost during integration. So, we design a 3D CNN model for our 3D lattice model. To distinguish the localized and delocalized state, the CNN model will return two real numbers to represent the probability of the extended state  $P$  and the localized state  $(1 - P)$  for the given wave function. If the probability of the extended state is larger than 0.5, we think the eigenstate is delocalized, and vice versa. Due to the limitation of the graphics memory (8 GB) of our graphics card (NVIDIA GTX 1080), we consider a 3D  $20 \times 20 \times 20$  lattice. The hidden layers in the CNN model consist of convolutional layers, max-pooling layers, and fully connected layers. The loss function is defined by the cross entropy  $H(x) = -\sum_x p(x) \ln p(x)$ . During the training, we use the RMSPropOptimizer solver defined in TENSORFLOW [47] as the stochastic gradient descent solver to minimize the loss function. The details of the neural network model are in Appendix.

The training data for different phases are sampled from the three-dimensional Anderson model using different disorder parameters. It’s well known that the critical disorder at  $E = 0$  for the 3D Anderson model is  $16.54 \pm 0.01$  [48–50]. When the disorder strength  $W$  is larger than the critical value, the wave functions are exponentially localized and the system behaves as an insulator. Otherwise, the wave functions are delocalized and the system behaves as a metal. This phenomenon is known as metal-insulator transition (MIT) [1]. We get 4000 eigenstates from  $W \in [14.0, 16.0)$  as the delocalized phase and 4000 eigenstates from  $W \in [17.0, 19.0)$  as the localized phase by steps of 0.1. For each  $W$ , we prepare 40 different realizations of randomness and for each realization, we take five eigenstates around  $E = 0$ . For the validation data set, we get another 600 eigenstates from  $W \in [10.0, 16.0)$  and 600 eigenstates from  $W \in [17.0, 23.0)$  in steps of 0.1. During each step of the training, we randomly select 256 eigenstates from the training data set as the input and calculate the gradient of the loss function with respect to the parameters in the CNN model and update them. After every 50 steps, we test the prediction accuracy on the validation data set and save the model with the highest prediction accuracy.

To show the prediction accuracy for different disorder parameters  $W$ , we generate another 16 000 eigenstates sampled from the Anderson model using  $W \in [0.1, 16.0)$  and  $W \in [17.0, 33.0)$ . The prediction accuracy for different disorder strengths  $W$  is shown in Fig. 6(a), and the overall accuracy is 99.0%. The lowest prediction accuracy around the critical disorder  $0.8W_c < W < 1.2W_c$  is about 83%. We also test our trained model by producing the phase transition diagram of the 3D Anderson model. The testing data are sampled from

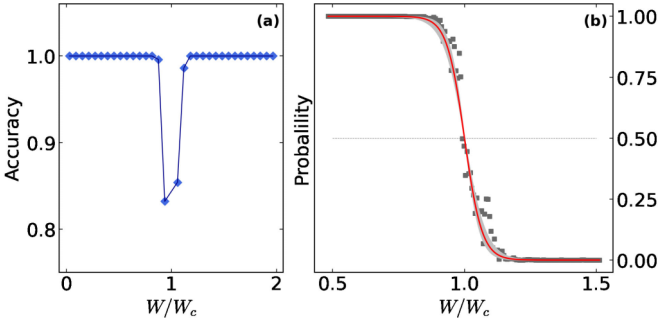


FIG. 6. The performance of the trained neural network on the Anderson model with different disorder parameters  $W$ .  $W_c = 16.54$  is the critical disorder for  $E = 0$ . (a) The classification accuracy of the trained neural network model. (b) The probability that the wave function is considered as an extended state by the trained neural network model.

$W \in [8.0, 25.0]$  by steps of 0.1. In each realization of the same disorder parameter  $W$ , we pick five eigenstates around the band center ( $E = 0$ ) as input data and use the averaged delocalized probability of the five eigenstates as the delocalized probability of this realization. We prepare five random realizations for each  $W$  and average the delocalized probability. The phase diagram calculated using our trained CNN model is shown in Fig. 6(b). From Fig. 6(b), we see that the trained CNN model successfully captures the MIT.

Owing to its excellent classification accuracy, the trained neural network model is ready to find the extended state in the impurity band. We generate 1000 random realizations for the Hamiltonian in Eq. (1) with doping probability  $x = 5\%$  and disorder parameter  $W = -4.5$ , and obtain all eigenstates using the exact diagonalization method in LAPACK. These quantum states are used as the input data for our trained CNN model to calculate the delocalized probability. We average the probability over 1000 realizations and the result is shown in Fig. 7. The trained CNN model confirms that delocalized states exist in the impurity band, which is in good agreement with the results obtained via IPR or Thouless number.

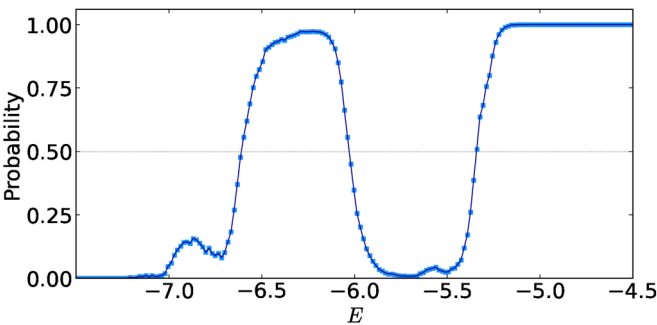


FIG. 7. The probability that the corresponding wave function for different eigenenergies is considered as an extended state by the trained neural network model. The input wave functions are generated from the Hamiltonian in Eq. (1) using exact diagonalization. Averages over 1000 realizations are taken.

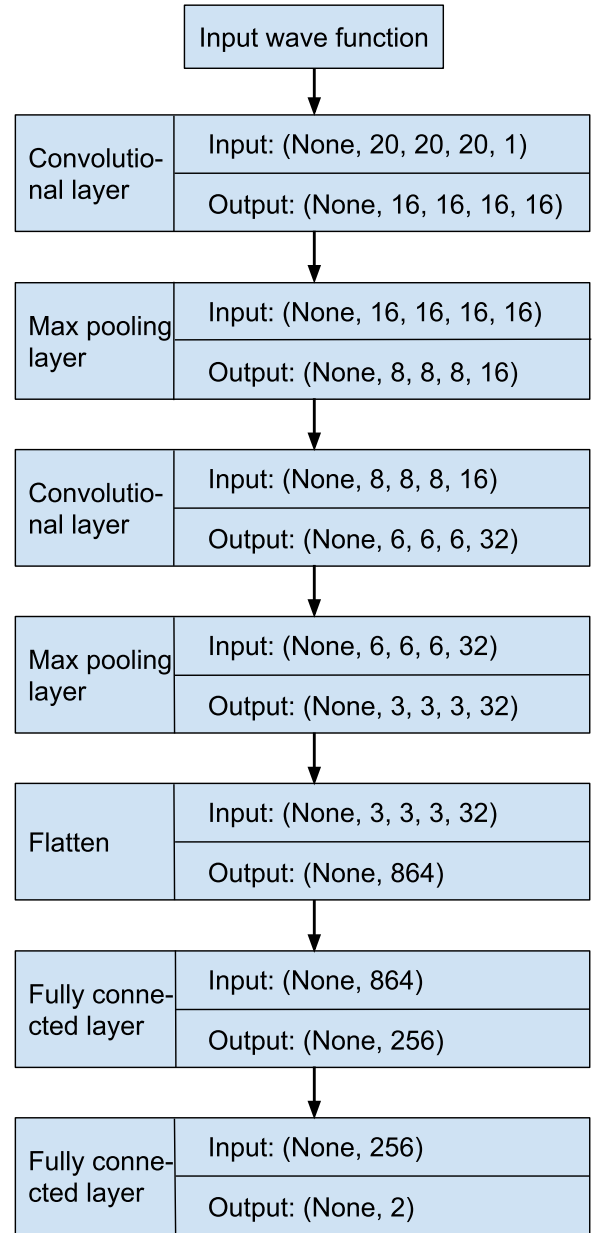


FIG. 8. The architecture of the 3D CNN model employed in this paper for the 3D  $20 \times 20 \times 20$  lattice. “None” in the figure represents the batch size during training or evaluation, which is not a fixed number.

## V. CONCLUSIONS

In this work, we numerically investigate the properties of the states in the “impurity band” of heavily doped nonmagnetic semiconductors. By using general full exact diagonalization and sparse matrix diagonalization with the JADA method, we find that with a typical doping probability  $x = 5\%$ , the impurity band in the DOS is developed at about  $W = -4.5$ . We calculate the IPR, Thouless number, and DOS together for different system sizes and study the relationship between them. The data fitting of IPR and system size on four points suggests the existence of the extended states in the impurity band. The Thouless number calculation supports

the same conclusion and gives the exact location of mobility edges.

In addition, we also utilize the supervised deep learning method, which is the state-of-the-art method in pattern recognition, to distinguish the extended and localized states in the impurity band. We train a 3D CNN model using the data generated from the Anderson model and then apply the trained neural network model to classify the states in the “impurity band.” Our trained neural network model achieves high accuracy (99.0%) in classifying different states in the Anderson model. The prediction of our trained model on impurity band also supports the finding from the relationship between IPR, Thouless number, and system size though the predicted locations of mobility edges have small discrepancies. Our calculation gives direct evidence that there are three mobility edges in the impurity band for a specific on-site impurity potential in heavily doped nonmagnetic semiconductors.

### ACKNOWLEDGMENTS

Z.-X.H. is supported by National Natural Science Foundation of China Grants No. 11974064 and No. 12147102, the Chongqing Research Program of Basic Research, and Frontier Technology Grant No. cstc2021jcyjmsxmX0081, Chongqing Talents: Exceptional Young Talents Project No. cstc2021ycjh-bgzxm0147, and the Fundamental Research

Funds for the Central Universities Grant No. 2020CDJQY-Z003. H.C. is supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award No. DE-SC0022216, as well as under Contract No. DE-AC02-76SF00515 for the Materials Sciences and Engineering Division.

### APPENDIX: NEURAL NETWORK MODEL ARCHITECTURE AND HYPERPARAMETERS

The 3D CNN model used in this paper has a similar architecture to “AlexNet” [38] and “VGGNet” [51], but with a smaller number of convolutional, max-pooling, and fully connected layers. This is because we are dealing with a 3D lattice and the edges in the lattice have a much smaller length compared to the images. The architecture of our model is shown in Fig. 8, and the input and output dimension of each layer is also listed in the figure.

The size of the convolution kernel applied in the first and second convolutional layers is  $5 \times 5 \times 5$  and  $3 \times 3 \times 3$ , respectively. The activation function, rectified linear unit [52], is performed after the convolutional layer and the fully connected layer except for the last layer, which is activated by the softmax [53] function. Bias parameters are included for artificial neurons. Dropout [54] is performed with probability  $p = 0.5$  after the first fully connected layer to avoid overfitting and increase the evaluation accuracy.

- 
- [1] P. W. Anderson, *Phys. Rev.* **109**, 1492 (1958).
  - [2] M. Avérous and M. Balkanski, *Semimagnetic Semiconductors and Diluted Magnetic Semiconductors* (Springer, New York, 1991).
  - [3] H. Ohno, *Science* **281**, 951 (1998).
  - [4] B. Beschoten, P. A. Crowell, I. Malajovich, D. D. Awschalom, F. Matsukura, A. Shen, and H. Ohno, *Phys. Rev. Lett.* **83**, 3073 (1999).
  - [5] A. M. Cook and M. Berciu, *Phys. Rev. B* **85**, 235130 (2012).
  - [6] A. Weiße and H. Fehske, *Computational Many-Particle Physics* (Springer, New York, 2008), pp. 529–544.
  - [7] I. U. Ojalvo and M. Newman, *AIAA J.* **87**, 1234 (1970).
  - [8] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
  - [9] J. Carrasquilla and R. G. Melko, *Nat. Phys.* **13**, 431 (2017).
  - [10] K. Ch’Ng, J. Carrasquilla, R. G. Melko, and E. Khatami, *Phys. Rev. X* **7**, 031038 (2017).
  - [11] E. P. Van Nieuwenburg, Y.-H. Liu, and S. D. Huber, *Nat. Phys.* **13**, 435 (2017).
  - [12] J. Venderley, V. Khemani, and E.-A. Kim, *Phys. Rev. Lett.* **120**, 257204 (2018).
  - [13] S. J. Wetzel, *Phys. Rev. E* **96**, 022140 (2017).
  - [14] J. F. Rodriguez-Nieva and M. S. Scheurer, *Nat. Phys.* **15**, 790 (2019).
  - [15] A. Lidiak and Z. Gong, *Phys. Rev. Lett.* **125**, 225701 (2020).
  - [16] Y.-T. Hsu, X. Li, D.-L. Deng, and S. D. Sarma, *Phys. Rev. Lett.* **121**, 245701 (2018).
  - [17] D. Hendry, H. Chen, P. Weinberg, and A. E. Feiguin, *Phys. Rev. B* **104**, 205130 (2021).
  - [18] K. Choo, T. Neupert, and G. Carleo, *Phys. Rev. B* **100**, 125124 (2019).
  - [19] D. Pfau, J. S. Spencer, A. G. Matthews, and W. M. C. Foulkes, *Phys. Rev. Res.* **2**, 033429 (2020).
  - [20] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, *Phys. Rev. Lett.* **124**, 020503 (2020).
  - [21] D. Hendry, H. Chen, and A. Feiguin, *Phys. Rev. B* **106**, 165111 (2022).
  - [22] H. Chen, D. Hendry, P. Weinberg, and A. Feiguin, in *Advances in Neural Information Processing Systems*, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, Inc., Red Hook, NY, 2022), Vol. 35, pp. 7490–7503.
  - [23] A. Mannodi-Kanakkithodi, M. Y. Toriyama, F. G. Sen, M. J. Davis, R. F. Klie, and M. K. Chan, *npj Comput. Mater.* **6**, 39 (2020).
  - [24] V. Fung, P. Ganesh, and B. G. Sumpter, *Chem. Mater.* **34**, 4848 (2022).
  - [25] M. P. Polak, R. Jacobs, A. Mannodi-Kanakkithodi, M. K. Chan, and D. Morgan, *J. Chem. Phys.* **156**, 114110 (2022).
  - [26] L. Wang, *Phys. Rev. B* **94**, 195105 (2016).
  - [27] T. Ohtsuki and T. Ohtsuki, *J. Phys. Soc. Jpn.* **86**, 044708 (2017).
  - [28] A. Tanaka and A. Tomiya, *J. Phys. Soc. Jpn.* **86**, 063001 (2017).
  - [29] T. Mano and T. Ohtsuki, *J. Phys. Soc. Jpn.* **86**, 113704 (2017).
  - [30] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, *Sci. Rep.* **7**, 8823 (2017).
  - [31] F. Schindler, N. Regnault, and T. Neupert, *Phys. Rev. B* **95**, 245134 (2017).
  - [32] Z. Li, M. Luo, and X. Wan, *Phys. Rev. B* **99**, 075418 (2019).

- [33] X.-Y. Dong, F. Pollmann, and X.-F. Zhang, *Phys. Rev. B* **99**, 121104(R) (2019).
- [34] F. Kotthoff, F. Pollmann, and G. De Tomasi, *Phys. Rev. B* **104**, 224307 (2021).
- [35] W. Zhang, L. Wang, and Z. Wang, *Phys. Rev. B* **99**, 054208 (2019).
- [36] W. Zhang, J. Liu, and T.-C. Wei, *Phys. Rev. E* **99**, 032142 (2019).
- [37] N. Käming, A. Dawid, K. Kottmann, M. Lewenstein, K. Sengstock, A. Dauphin, and C. Weitenberg, *Mach. Learn.: Sci. Technol.* **2**, 035037 (2021).
- [38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, in *Advances in Neural Information Processing Systems*, edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger (Curran Associates, Inc., Red Hook, NY, 2012), Vol. 25.
- [39] A. Bohrdt, S. Kim, A. Lukin, M. Rispoli, R. Schittko, M. Knap, M. Greiner, and J. Léonard, *Phys. Rev. Lett.* **127**, 150504 (2021).
- [40] H. Théveniaut and F. Alet, *Phys. Rev. B* **100**, 224202 (2019).
- [41] J. Brndiar and P. Markoš, *Phys. Rev. B* **74**, 153103 (2006).
- [42] J. Edwards and D. Thouless, *J. Phys. C* **5**, 807 (1972).
- [43] M. Bollhöfer and Y. Notay, *Comput. Phys. Commun.* **177**, 951 (2007).
- [44] K. Andres, R. N. Bhatt, P. Goalwin, T. M. Rice, and R. E. Walstedt, *Phys. Rev. B* **24**, 244 (1981).
- [45] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, *Phys. Rev. Lett.* **42**, 673 (1979).
- [46] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney *et al.*, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999).
- [47] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org, <https://www.tensorflow.org/>
- [48] A. MacKinnon and B. Kramer, *Phys. Rev. Lett.* **47**, 1546 (1981).
- [49] A. MacKinnon and B. Kramer, *Z. Phys. B* **53**, 1 (1983).
- [50] B. Kramer and A. MacKinnon, *Rep. Prog. Phys.* **56**, 1469 (1993).
- [51] K. Simonyan and A. Zisserman, in *3rd International Conference on Learning Representations*, edited by Y. Bengio and Y. LeCun (ICLR, San Diego, CA, USA, 2015).
- [52] V. Nair and G. E. Hinton, in *Proceedings of the 27th International Conference on International Conference on Machine Learning* (Omnipress, Madison, WI, USA, 2010), ICML-10, pp. 807–814.
- [53] J. S. Bridle, in *Proceedings of the 2nd International Conference on Neural Information Processing Systems* (MIT Press, Cambridge, MA, USA, 1989), NIPS-89, pp. 211–217.
- [54] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *J. Mach. Learn. Res.* **15**, 1929 (2014).