

ARTICLE OPEN



Machine learning method for tight-binding Hamiltonian parameterization from ab-initio band structure

Zifeng Wang¹, Shizhuo Ye¹, Hao Wang¹, Jin He¹, Qijun Huang¹ and Sheng Chang¹✉

The tight-binding (TB) method is an ideal candidate for determining electronic and transport properties for a large-scale system. It describes the system as real-space Hamiltonian matrices expressed on a manageable number of parameters, leading to substantially lower computational costs than the ab-initio methods. Since the whole system is defined by the parameterization scheme, the choice of the TB parameters decides the reliability of the TB calculations. The typical empirical TB method uses the TB parameters directly from the existing parameter sets, which hardly reproduces the desired electronic structures quantitatively without specific optimizations. It is thus not suitable for quantitative studies like the transport property calculations. The ab-initio TB method derives the TB parameters from the ab-initio results through the transformation of basis functions, which achieves much higher numerical accuracy. However, it assumes prior knowledge of the basis and may encompass truncation error. Here, a machine learning method for TB Hamiltonian parameterization is proposed, within which a neural network (NN) is introduced with its neurons acting as the TB matrix elements. This method can construct the empirical TB model that reproduces the given ab-initio energy bands with predefined accuracy, which provides a fast and convenient way for TB model construction and gives insights into machine learning applications in physical problems.

npj Computational Materials (2021)7:11 | <https://doi.org/10.1038/s41524-020-00490-5>

INTRODUCTION

New materials with attractive properties are springing up, sparking the exploration of their potential for electronics. During exploration, it is necessary to determine the band structures and transport properties of these systems. Fortunately, both of these factors can be derived from the Hamiltonian of the system^{1–3}.

For a realistic large-scale system, especially with limited periodicity, the calculation of the corresponding Hamiltonian is often intractable in a conventional ab-initio study. Additionally, the resultant Hamiltonian usually has a large number of basis functions, thus increasing the temporal complexity of transport property calculations in which many matrix diagonalization operations are involved. In this regard, the tight-binding (TB) method becomes a practicable approach, describing a system as real-space TB Hamiltonian matrices expressed on a manageable number of parameters^{4,5} to reduce the subsequent computing time. The choice of the TB parameters thus determines the reliability of the TB calculations. Traditionally, considering the amount of time required to obtain the reasonable TB parameters on one's own, the TB parameters are often obtained from the published TB parameter sets to construct the empirical TB models for the desired systems. However, these published empirical parameters are usually obtained by fitting to the ab-initio results of certain materials with fixed geometries and boundary conditions. As a result, the TB models constructed from these parameters can hardly reproduce the ab-initio band structures of the materials with different geometries and boundary conditions quantitatively, which becomes a source of the unreliability of this typical empirical TB method in quantitative research. In recent years, the reliability of the TB method has been greatly improved by the introduction of several ab-initio TB methods, which are based on the projection of the extended Bloch states obtained from the ab-initio calculations onto a much smaller set of localized

orbitals^{6–10}. Such methods drive the TB parameters directly from the ab-initio results of the desired material systems without the fitting process. The resulting ab-initio TB models are compatible with the typical TB form and reproduce the selected ab-initio energy bands with high accuracy. However, though successfully adopted in numerical studies of a variety of materials and devices^{11–15}, these projection-based methods have their own challenges.

First, these methods require full knowledge of the eigenenergies and eigenfunctions calculated from the ab-initio methods. The corresponding time-consuming ab-initio calculations must be performed before the TB Hamiltonian construction, which hinders massive high-throughput investigations. Additionally, the abundant experience is needed when selecting TB basis functions in these methods. An accurate representation of the ab-initio bands in the much smaller TB basis requires the good projectability¹⁶ of the corresponding ab-initio Bloch states onto the finite Hilbert space spanned by these TB basis functions. For methods based on non-iterative projection schemes, such as the direct projection¹⁰ and Quasi-atomic orbitals⁹, the TB basis is predetermined as a specific set of atomic or atomic-like orbitals, so the choice of the ab-initio bands to be reproduced is limited to those with satisfactory projectability on the specific Hilbert space considered. Though improved by a series of studies^{16–18}, such methods are still not good at dealing with the projection of unoccupied states far above the Fermi level unless increasing the richness of the basis. For iterative methods such as the maximally localized Wannier functions (MLWFs)^{19,20} approach and muffin-tin orbitals of arbitrary order (NMTO)⁶ approach, any bundle of the ab-initio bands can be selected for projection, whereas the TB basis functions should be iteratively optimized to span a suitable Hilbert space, which requires detailed knowledge of the underlying system and sufficient time for trial-and-error procedures²¹. Finally,

¹Key Laboratory of Artificial Micro- and Nano-Structures of Ministry of Education and School of Physics and Technology, Wuhan University, 430072 Wuhan, Hubei, People's Republic of China. ✉email: changsheng@whu.edu.cn

even well-projected orbitals may produce Hamiltonian matrix elements associated with long-range neighbor interactions²². Suppose the desired TB representation (like the nearest-neighbor TB model) does not consider such long-range interactions. In that case, these elements need to be truncated, and the resulting TB representation will suffer a loss of accuracy⁷.

This work explores another way for TB Hamiltonian parameterization, with machine learning (ML) techniques. This method uses a fast scheme to fit the ab-initio energy bands by focusing directly on adjusting the TB matrix elements. Specifically, a neural network (NN) model is introduced to fill the preselected real-space TB Hamiltonian matrices with the neurons by treating them as the matrix elements. The values of these neurons, which determine the TB matrices, will be flexibly updated to achieve a satisfactory match between the produced TB bands and the ab-initio bands during the training phase by a back-propagation algorithm^{23,24}. This method assigns an individual NN model to every desired system for TB parametrization, which is affordable because of the fast fitting scheme. The resulting one-to-one empirical TB model can achieve better reliability on the chosen system, than the typical empirical model constructed from TB parameter sets, because of the concentration on the reproduction of the given bands without considering the transferability to other systems. Additionally, the matrix element adjustment scheme is free from the issues mentioned above that both the projection-based methods. Within this method, the ab-initio band structure data of the desired material are used as the only training data, and no other input, such as atomic coordinate information or eigenfunction data, is needed. Therefore, the vast existing resources of band structures in public databases can be used, bypassing the ab-initio calculations. This approach implicitly defines the TB basis functions by directly parameterizing the TB Hamiltonian matrices, so no prior knowledge of the functions is involved. Furthermore, this method can in advance exclude the matrix elements that need to be truncated according to the desired TB model so that the truncation error can be avoided. Most importantly, this method is applicable to any material system with accessible and reliable band structure data. In the following article, this method is described in detail, and its merits are verified.

RESULTS

Parameterized TB Hamiltonian matrix

The empirical TB method works by writing the eigenstates of the Hamiltonian \hat{H} in a basis set of atomic or atomic-like orbitals, $|\phi_i\rangle$, and replacing the exact many-body Hamiltonian operator with a parametrized Hamiltonian matrix H , which can then be used to compute the desired electronic and transport properties of the given system. Typically, the basis set is not explicitly constructed but defined by the empirical parameters used to form H . Generally, the parametrized TB Hamiltonian matrix H can be written as follows:

$$H = \sum_i \epsilon_i n_i + \sum_{i \neq j} t_{ij} c_i^\dagger c_j, \quad (1)$$

where i and j run over the considered basis orbitals, ϵ_i denotes the energy of the electron at site i , t_{ij} is the hopping energy between the sites i and j , and c_i^\dagger (c_j) are the creation (annihilation) operator of electrons at site i (j). Since these on-site and hopping terms are usually obtained from the existing TB parameter sets, specific optimizations have to be performed according to the studied system at hand, or the numerical accuracy should be significantly lower than the ab-initio calculations. We address these problems with our tight-binding Hamiltonian construction neural network (TBHCNN), which uses the neurons to represent the TB matrix elements. The number and values of the neurons are adjusted continuously during the training phase with the ab-initio bands as references. After training, the neurons can be used to form the

parameterized TB Hamiltonian, which conforms to the matrix form as Eq. (1) and at the same time reproduces the reference bands.

Workflow of the TBHCNN

Figure 1 schematically illustrates the workflow of using the TBHCNN model for TB Hamiltonian parametrization, taking the examples of a one-dimensional (1D) periodic system and a uniform 1D non-periodic system.

For the periodic system, we first obtain its band structure data by searching in the online database if possible or by performing the ab-initio calculations on it, and then we select the energy bands of interest as the training set. When preparing the ab-initio calculations, although the unit cells of arbitrary sizes can be chosen, we advocate keeping the unit cell as small as possible. As the unit cell becomes larger, the corresponding first Brillouin zone shrinks and the calculated band structure becomes gradually dense energy levels²⁵. It is hard to extract useful information from such a heavily folded band structure²⁶, and thus hard to determine the selection of the energy bands as references. Also, more computing time is required with a larger unit cell for the ab-initio calculations. After selecting the unit cell and the ab-initio energy bands as references, we then determine the interaction range (i.e., the distances range in which two orbitals are considered to have hopping terms) considered in the desired TB model. Since our method does not assume the knowledge of the atomic coordinate information, we determine the interaction range by deciding on the real-space Hamiltonian matrices used in the chosen unit cell representation. For example, we may just use H^{0-1} , H^{00} , and H^{01} , which denote the real-space Hamiltonian matrices between the unit cell of the lattice vector $\mathbf{R} = \mathbf{0}$ and the cells of the lattice vector $\mathbf{R} = -\mathbf{1}$, $\mathbf{0}$, and $\mathbf{1}$, to construct the TB model where the interactions beyond the nearest unit cells are made negligible. Of course, other choices can be made on-demand. Then, we initialize the TBHCNN model according to the selected energy bands and real-space Hamiltonian matrices. Specifically, the Hamiltonian matrix size will default to the number of reference bands to ensure that the eventual TB model contains enough bands, and the proper number of neurons will be added to the TBHCNN model to fill in these matrices as their elements. These neurons have the initial values sampled from a standard normal distribution and will remain real numbers during training. Note to ensure that H^{00} should maintain a symmetric form, and the real-space Hamiltonian matrices with opposite lattice vectors should transpose each other. Now, we get a randomly initialized TB model. To obtain the TB band structure, ensure the orthogonality of the TB basis and diagonalize the reciprocal-space Hamiltonian,

$$H^{\text{TB}}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} H^{\text{OR}}, \quad (2)$$

where \mathbf{R} runs over the lattice vectors of the selected real-space Hamiltonian matrices, for desired \mathbf{k} vectors in the Brillouin zone to obtain the band energies $\epsilon_{n,\mathbf{k}}^{\text{TB}}$ and the eigenvectors $\psi_{n,\mathbf{k}}^{\text{TB}}$,

$$H^{\text{TB}}(\mathbf{k})\psi_{n,\mathbf{k}}^{\text{TB}} = \epsilon_{n,\mathbf{k}}^{\text{TB}}\psi_{n,\mathbf{k}}^{\text{TB}}, \quad (3)$$

where n is the band index. The TB band structure can be obtained by assembling all the band energies $\epsilon_{n,\mathbf{k}}^{\text{TB}}$ for all the \mathbf{k} vectors considered. The quality of the TB model can be evaluated by comparing the TB bands with the chosen ab-initio bands. As a measure of their mismatch, the mean squared error (MSE) between the energy eigenvalues is adopted (in units of (eV)²):

$$\Delta_E = \frac{1}{N_k} \sum_{i=1}^N \sum_{\mathbf{k}} (\epsilon_{i,\mathbf{k}}^{\text{TB}} - \epsilon_{i,\mathbf{k}}^{\text{Ab}})^2, \quad (4)$$

where N_k is the number of \mathbf{k} -points sampled in the reference bands. The procedure mentioned above is the forward pass, and Δ_E serves as the loss function Loss. In the backward pass, the

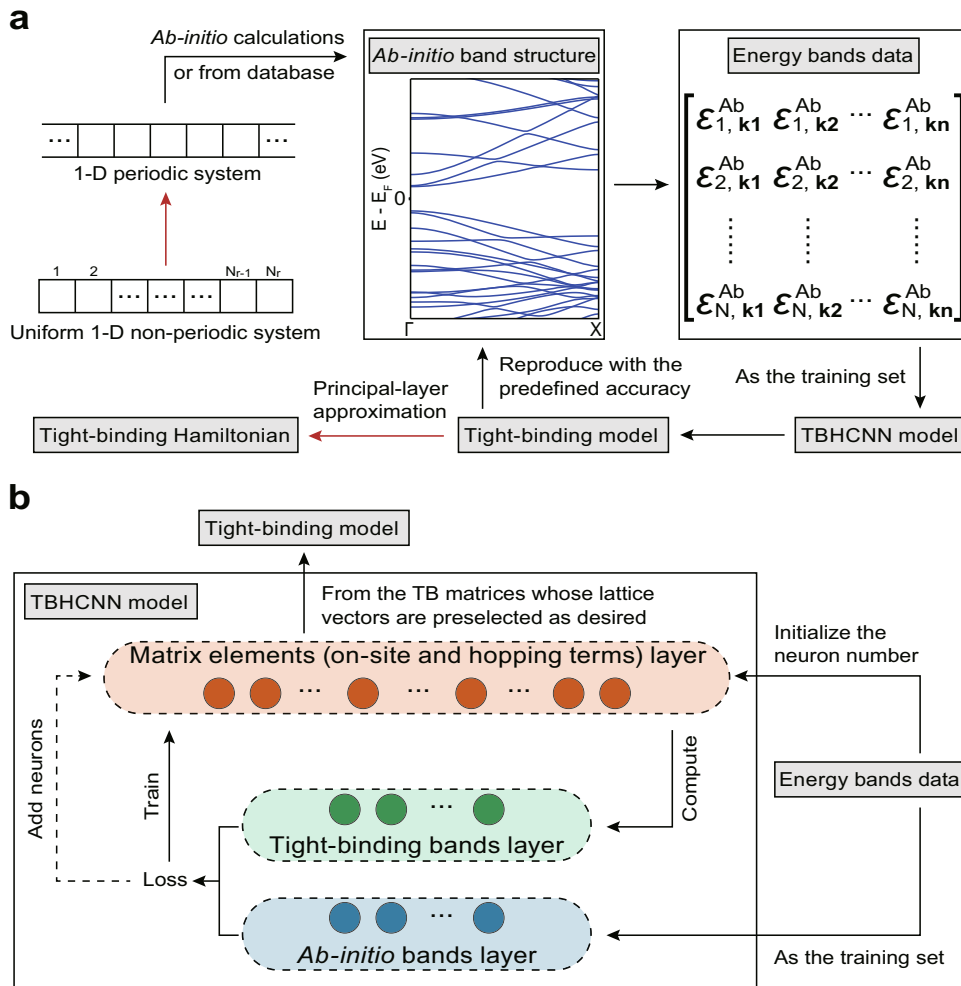


Fig. 1 Workflow of the TBHCNN. **a** The workflows of obtaining the TB model for a 1D periodic system and constructing the TB Hamiltonian matrix for a uniform 1D non-periodic system. There are two additional steps for the latter, which are marked with the red arrows. **b** Structure diagram of the TBHCNN model. The matrix elements layer in the TBHCNN will be initialized according to the number of the reference ab-initio bands and the real-space TB Hamiltonian matrices considered in the desired TB model. The reference bands data are used as the training set, of which the eigenenergies $\epsilon_{i,k}^{Ab}$ are encoded within the ab-initio bands layer for computing the loss function by comparing with the TB results encoded in the tight-binding bands layer. The loss function value will be backpropagated to train the value of the neurons in the matrix element layer, which will be used as the matrix elements to construct the considered real-space Hamiltonians. When the loss function cannot touch down the predefined threshold, the TBHCNN model will add new neurons to the matrix element layer and reinitialize the whole layer to start a new round of training.

derivative of the loss function with respect to the elements of the real-space TB Hamiltonian H^{OR} , denoted by \bar{H}^{OR} using standard algorithmic differentiation terminology²⁷, will be computed as

$$\bar{H}^{OR} = \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}} \Psi_{\mathbf{k}} \bar{E}_{\mathbf{k}} \Psi_{\mathbf{k}}^T, \quad (5)$$

with

$$\bar{E}_{\mathbf{k}} = \text{diag} \left(\frac{\partial \text{Loss}}{\partial \epsilon_{1,\mathbf{k}}^{TB}}, \frac{\partial \text{Loss}}{\partial \epsilon_{2,\mathbf{k}}^{TB}}, \dots, \frac{\partial \text{Loss}}{\partial \epsilon_{N,\mathbf{k}}^{TB}} \right), \quad (6)$$

and

$$\Psi_{\mathbf{k}} = \left(\psi_{1,\mathbf{k}}^{TB}, \psi_{2,\mathbf{k}}^{TB}, \dots, \psi_{N,\mathbf{k}}^{TB} \right). \quad (7)$$

Then, the matrix elements in the Hamiltonian can be updated by using the gradient descent algorithm

$$H_{n+1}^{OR} = H_n^{OR} - \alpha \bar{H}_n^{OR}, \quad (8)$$

where α is the learning rate, and the subscripts of the matrices represent the number of training steps in the TBHCNN.

Through the back-propagation process, the values of the neurons are continuously adjusted to minimize the loss function, leading to an improved match between the resultant TB bands and the ab-initio references. The numerical threshold for the loss function and the maximum number of training steps should be predefined as the criteria for ending the training. Once the loss function value touches down the preset threshold, the TBHCNN will end the training, and its neurons will be used for TB Hamiltonian parameterization, resulting in the TB model reproducing the reference bands with the MSE being the value of the preset threshold. However, if the loss function value remains higher than the preset threshold after the maximum number of training steps, the TBHCNN will add extra neurons and increase the basis size of the real-space Hamiltonian by a predefined number to enlarge the basis set; then, the whole network will be reinitialized, and a new round of training will begin. It can be seen that the TBHCNN model is in fact a dynamic network with a variable number of neurons. The additional bands induced by the increase of basis will be put above or below the set of reference bands and will not be used for loss function computation. These

procedures will be repeated until a satisfactory agreement between the produced TB bands and references is achieved. No manual intervention is needed. We hasten to add that the selection of the hyperparameters mentioned above should not be strictly fixed but rather a choice on demand. The loss function threshold should be determined according to the accuracy desired; the setting of the maximum number of training steps depends on the convergence of the loss function on the selected reference bands; the neurons added, i.e., the basis functions increased, should be chosen considering the tradeoff between the training time and the basis size of the resulting Hamiltonian. A large number may result in the TB model consisting of more basis functions than is actually needed to reproduce the reference bands, whereas a small number might lead to more training time.

For the uniform 1D system with limited periodicity, which cannot be represented by a set of real-space Hamiltonian matrices $H^{\mathbf{0R}}$ labeled with lattice vector \mathbf{R} but rather a whole real-space Hamiltonian matrix containing all the considered on-site and hopping terms, we present a simple TB Hamiltonian construction scheme with the principal-layer (PL) approximation²⁸. Compared with the 1D periodic system, two additional steps are required within the construction scheme. One is to select a fragment⁷ of this system. The fragment should be a repeating structural unit of the system. We then use the fragment as a unit cell of the corresponding periodic system to obtain its band structure. For the same reasons we stated above, the size of the fragment should be as small as possible. Then we can obtain the TB model of the corresponding periodic system using the same procedures as we demonstrated above. There comes the other additional step where the Hamiltonian matrix of the 1D non-periodic system is constructed by the produced TB model. Using the PL approach, we let the PL consist of one unit cell and make the TB model consider the real-space Hamiltonian H^{0-1} , H^{00} , and H^{01} only. In this case, the Hamiltonian of the desired 1D non-periodic system can be constructed by these matrices as

$$\begin{pmatrix} H^{00} & H^{01} & 0 & 0 & 0 \\ H^{0-1} & H^{00} & H^{01} & 0 & 0 \\ 0 & H^{0-1} & H^{00} & H^{01} & 0 \\ 0 & 0 & H^{0-1} & H^{00} & \ddots \\ 0 & 0 & 0 & \ddots & \ddots \end{pmatrix}_{(N_f \times N_f) \times (N_f \times N_f)}, \quad (9)$$

where N_f is the final size of the trained real-space Hamiltonian matrices and N_f is the number of fragments required to rebuild the 1D non-periodic system.

Theoretically, the TBHCNN applies to systems of any dimension. The 1D models above are used as examples for the sake of simplicity and brevity. To extend the above procedures to arbitrary periodic systems, take into account the real-space Hamiltonian matrices which describe the interactions along every periodic direction. And these matrices can be used to construct TB Hamiltonians of the corresponding uniform non-periodic systems in the same way as stated above.

We must emphasize that, within the method presented above, only the ab-initio energies are used for training the TBHCNN without the involvement of any information of the basis functions and atomic coordinate of the given system. Therefore, the method in its current form cannot ensure the match of the symmetry characters between the predicted TB bands and reference bands, nor the resulting Hamiltonian elements could reflect the geometrical symmetries. These should be considered as the limitations of the proposed method. However, we will show the validity of the resulting TB Hamiltonian parameterization for transport property calculations in the following text. We will also present a variation of this method to deal with the limitations by introducing and

exploiting the additional information of geometry and the symmetry of reference bands in later sections.

Additionally, the construction scheme that is shown in Eq. (9) with the PL approximation is suitable for the presented uniform system but would not apply to a general inhomogeneous system. In this case, the common practices are to divide the inhomogeneous system into several homogeneous subsystems, and then to perform the TB Hamiltonian parameterizations for these subsystems, respectively. And the proper stitching of these TB models should be performed to construct the Hamiltonian of the whole inhomogeneous system, which is also an important research aspect where several research findings have been published^{14,29} but would be out of the scope of this paper. Therefore, we do not expand on the stitching schemes in detail but recommend the cited references to the interested readers.

Application and validation of the proposed method

For novel material systems, the experience with and knowledge of the basis functions are not sufficient, so relevant TB models are rare, which severely hampers transport analyses with and applications of these systems. For example, 2D-InSe nanosheet is a prospective system since this typical III–VI semiconductor has many attractive properties, such as high electron mobility^{30–32} and good ohmic contact³³. Few-layer and monolayer InSe nanosheets have been successfully synthesized³⁴, but there have been few direct theoretical studies on quantum transport in 2D-InSe-based devices for the above reason. Here, the simulation of an InSe nanoribbon metal-oxide-semiconductor field-effect transistor (MOSFET) is performed with the TB Hamiltonian generated by the proposed ML method to illustrate the capacity of this method to accurately and efficiently solve cutting-edge problems.

The device geometry is shown in Fig. 2. A 13-atom-wide InSe nanoribbon with a hydrogen-passivated boundary is sandwiched between 2-nm-thick oxide layers with a relative dielectric constant of 3.9. The length of the source and drain is 5 nm, and the channel length is 10 nm. Both the source and the drain regions are doped with a molar fraction of fully ionized donors of 5×10^{-3} . The channel region is undoped, and the double gates cover the whole channel. The gates and oxide layers are not modeled atomically but are introduced to change the potential of the channel region and to act as the dielectric layers with a desired dielectric permittivity, respectively. To eliminate the impact of the work function values of different gate metals, we located the Fermi level of the channel in the middle of the bandgap in the following calculations.

To verify the merits of the proposed method, two TB models for the periodic InSe nanoribbon using the TBHCNN model and the MLWFs method, the most popular projection-based method, were obtained. Their corresponding device Hamiltonians were constructed according to Eq. (9), considering the interactions between adjacent unit cells only. During training, the TBHCNN model added the proper number of neurons to increase the TB basis size by 2 every 10000 training steps until that $\text{Loss} \leq 1.0 \times 10^{-5}$ was achieved, resulting in the TB model with a basis size of 18. And the

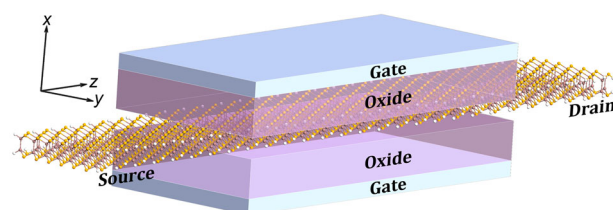


Fig. 2 Sketch of the simulated InSe nanoribbon-based transistor. Only the InSe nanoribbon is modeled atomically, whereas the gate metals and oxide layers are not introduced as entities but used to change the gate voltage and provide the dielectric layers.

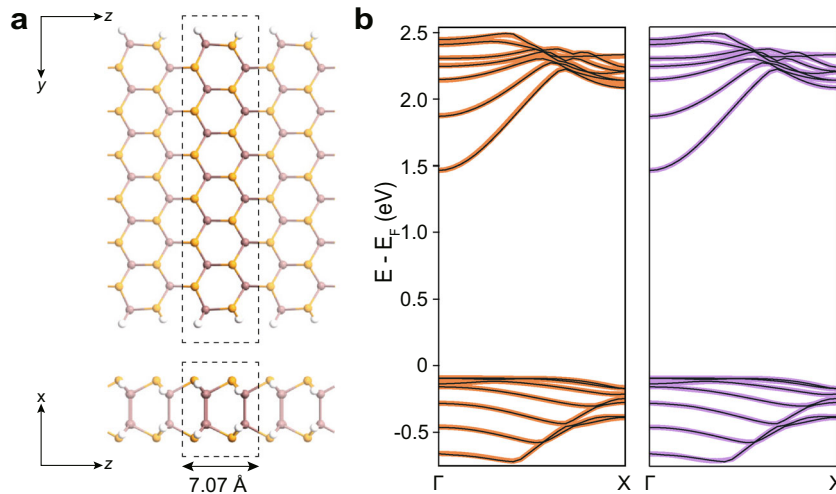


Fig. 3 Geometrical and electronic structure of the InSe nanoribbon. **a** Optimized structure of the 13-atom-wide InSe nanoribbon. The orange, brown, and white balls represent Se, In, and H atoms, respectively. The unit cell used is marked with a dotted rectangle. **b** Comparison between the ab-initio band structure (black) and the band structures calculated from the WTb model (orange) and the MLTB model (lilac) for the InSe nanoribbon. Only the 14 bands selected for reference are plotted for clarity.

MLWFs method obtained the TB model containing 208 basis functions, taking into account the outermost s and p orbitals of each In and Se atom in the unit cell, which is a common choice in existing research on layered InSe materials³⁵. In the following sections, the two TB models are called the machine learning tight-binding (MLTB) model and the Wannier tight-binding (WTB) model, and the corresponding device Hamiltonians are called the MLTB Hamiltonian and the WTB Hamiltonian, respectively. The band structures of the two TB models were calculated and compared with the ab-initio band structure, as shown in Fig. 3. Only the 7 conduction bands and 7 valence bands around the Fermi level with a $1 \times 1 \times 26$ uniform \mathbf{k} -points sampling along the \mathbf{k} -path from high symmetry points Γ (0, 0, 0) to X (0, 0, 1/2), which reflect the key physical properties and notably influence the determination of transport properties, were selected as references. That is to say, there were $14 \times 26 = 364$ ab-initio energies as the training data. There was little difference between these two TB models in the reproduction of the ab-initio band structure; both of them fit the selected ab-initio bands with high accuracy. However, considering their respective TB basis sizes, the complexities of our method and the MLWFs method were completely different. The automated creation of the MLTB model was finished in 110 s within our method. Figure 4 shows the convergence of the loss function over the training time; the loss function value quickly decreases before the predefined threshold is reached. The MLWFs method required over 30 h to obtain the 208-orbital WTB model in the same computing environment, even without considering the time required to find the optimal parameter values, which in fact took up much time. Furthermore, even with a smaller basis and lower computational cost, the band structure generated by the proposed MLTB model is more accurate ($\Delta E = 1.0 \times 10^{-5}$) than that derived from the WTB model ($\Delta E = 1.7 \times 10^{-5}$).

In addition to the ab-initio band structure reproduction, the transport properties of the InSe nanoribbon-based MOSFET were also investigated. With the MLTB Hamiltonian and the WTB Hamiltonian, the corresponding quantum transport properties were obtained by self-consistently coupling the Schrödinger and Poisson equations using the NanoTCAD ViDES software package³⁶. Figure 5 plots the $I_{GS} - V_{GS}$ and $I_{GS} - V_{DS}$ curves. Each pair of curves under different voltage biases is almost the same in trends while slightly different in values. Considering the widespread acceptance of the reliability of the MLWFs method, the consistency here reflects the excellent applicability of our method

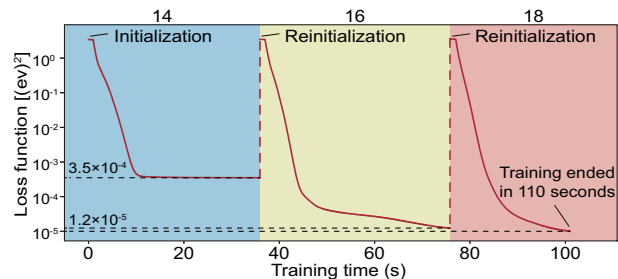


Fig. 4 Loss function for the TBHCNN. The red line represents the value of the loss function. The TBHCNN added neurons to increase the TB basis functions by 2 to restart training every 10,000 training steps until the ending criterion $\text{Loss} \leq 1 \times 10^{-5}$ was achieved. The blue, yellow, and red blocks denote the time regions when the TB basis contains 14, 16, and 18 basis functions, respectively.

to the device-performance level. Similar to the energy band computation, the device simulation with the MLTB Hamiltonian was more efficient than that with the WTB Hamiltonian. Specifically, the above device simulation was finished within 1 h using the MLTB Hamiltonian but more than 2 days with the WTB Hamiltonian.

As can be seen from the applications above, unlike previous ML methods introduced to handle the TB models^{37–46}, our algorithm does not require a large training set to make predictions. These predictive methods are introduced in an attempt to build a general mapping from the input data to the output data, using the input–output pairs within the training set. The predictive power of those methods will decrease when data considerably different from the training data are considered; as a result, these methods are not suitable for applications with limited datasets. Our method, instead, provides a generic way to construct the one-to-one TBHCNN model, which will be trained on the selected energy bands of the desired system and apply exclusively to the system. Hence, the required training set consists of the corresponding ab-initio energy bands only. The TBHCNN model here is introduced to represent the TB Hamiltonian elements directly rather than to act as the inferred function for mapping an external input to the Hamiltonian elements. Once a set of energy bands need to be reproduced, we can use them as the training data to initialize and train a TBHCNN model to obtain the TB model specifically suitable for the system with such energy bands.

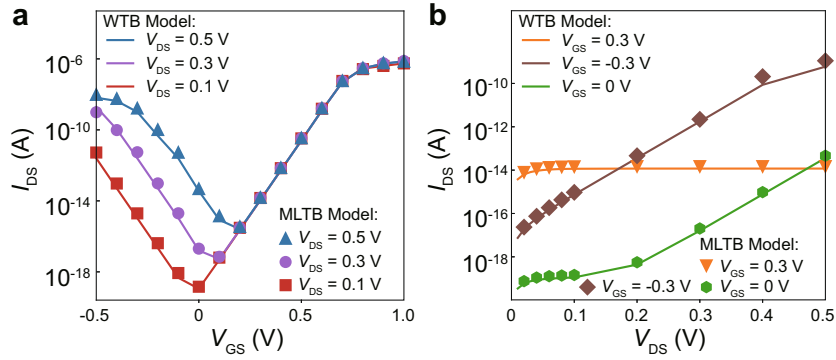


Fig. 5 Transfer characteristics of the InSe MOSFET. **a** Transfer characteristics of the device with the drain-source voltage V_{DS} being 0.5, 0.3, and 0.1 V. **b** Transfer characteristics of the device with the gate voltage V_{GS} being -0.3 , 0.0 , and 0.3 V.

We hope that our method provides insight into applying ML tools to physical problems, especially when data-driven methods are inaccessible.

To verify the generalization ability of the proposed method on different systems, extra simulations on 13-atom-wide graphene nanoribbon (13-AGNR), 2D MoS₂, and the devices based on these two materials were performed. The simulation results showed that our method could apply to different material systems and achieve better ab-initio band structure reproduction than the MLWFs method (Supplementary Note 1). Furthermore, to showcase the capability of the TBHCNN model of being trained on previously published data, we performed the TB Hamiltonian parameterizations for Si of the diamond structure and GaN of the wurtzite structure using the band structure data from the online data set Materials Project⁴⁷ (Supplementary Note 2). Additionally, we studied the impacts of the initialization of the TBHCNN model and the \mathbf{k} -point sampling used in the loss function on the quality of the resulting TB Hamiltonian for bands reproduction and transport property calculations. We found that, as long as the training hyperparameters were fixed, different initializations of the TBHCNN model would result in the TB models with different TB parameters that could reproduce the reference bands with the same accuracy. And the utilization of the device Hamiltonians from these different TB models on device simulation obtained very close results (Supplementary Note 3).

Two variations of the proposed method

Since the TBHCNN employed within our method can be conveniently customized, our method has the potential to be modified to meet the requirements of personalized analysis. Here, we develop two variations of the proposed method, which extend the applications of the TBHCNN to the fields outside the energy-space transport property calculations.

Variation I is to optimize a given TB model by fine-tuning the Hamiltonian elements. The basic version of our method focuses on constructing an accurate TB Hamiltonian with a minimal basis for the desired system to perform band structure and I - V curve calculations. However, it can also be modified to optimize a given TB model without breaking down the real-space Hamiltonian matrix structures.

Real-space information and time-consuming searches for optimal parameter values are not needed in Variation I by initializing the values of the neurons in the TBHCNN based on the matrix elements of the unoptimized TB model. Such a TB model can be obtained from projection-based approaches or the existing empirical TB parameter sets; both methods should retain the symmetry characters of the atomic basis to some extent. Then, a regularization term Δ_R , which penalizes the deviation of the Hamiltonian matrix elements from the original values, is added to

the loss function.

$$\Delta_R = \lambda \sum_i (V_i - U_i)^2, \quad (10)$$

where λ sets the magnitude of the penalty, for which the sum is over all matrix elements of the real-space Hamiltonians involved; V_i is the current value of the i th matrix element, and U_i is the value of the corresponding element in the unoptimized TB model. During training, the TBHCNN will hold the number of neurons but adjust their values to obtain an improved description of the selected ab-initio energy bands. By doing so, we can maintain the initial form of the given TB model to a large extent while greatly increase its accuracy for band reproduction. Variation I would also be very suitable for truncating a given TB model to have a smaller number of real-space Hamiltonian matrices while maintaining the band reproduction accuracy. The loss of accuracy caused by the truncated TB matrices, which describe the interactions in a longer range than is needed, can be mitigated by adjusting the elements of the TB matrices considered to implicitly include the impact of the truncated interactions.

To test the validity of Variation I, a 2D phosphorene system was tested. The sp TB model obtained by the MLWFs method was employed as the initial template, which could ensure the completeness of the TB basis⁴⁸. However, as illustrated in the left panel of Fig. 6a, the band structure obtained by this TB model deviates substantially from the corresponding ab-initio band structure. This sp TB model, deemed the WTB model here still, consisted of the real-space Hamiltonian matrices describing the interactions within a unit cell and between this unit cell with its 8 neighbor cells, as shown in Fig. 6b. This deviation indicated that the hopping terms describing the interactions in a longer range should be considered. Using the proposed variation, we constructed the optimized TB model on top of the WTB model, which is deemed the MLTB model here. It can be seen that the MLTB model succeeds in including the impact of longer-range interactions by modifying the elements in considered Hamiltonian matrices, as it reduced the MSE in-band energies from $\Delta_E = 0.14$ to $\Delta_E = 8.8 \times 10^{-7}$ when reproducing the 8 bands around the Fermi level. And the average absolute deviation per element of the 9 considered matrices was less than 0.04 eV. The difference matrices between these matrices before and after the optimization are plotted in Fig. 6c. It can be shown that Variation I has the potential to optimize the given TB model in the desired basis of atomic orbitals and maintain the symmetry characters of the basis to a large extent.

Variation II can construct the Slater-Koster TB Hamiltonian for the desired system from scratch by changing the mapping from the neurons to Hamiltonian elements.

To achieve this goal, the values of neurons are no longer assigned to the Hamiltonian elements directly. Instead, neurons are now used as adjustable parameters to calculate the

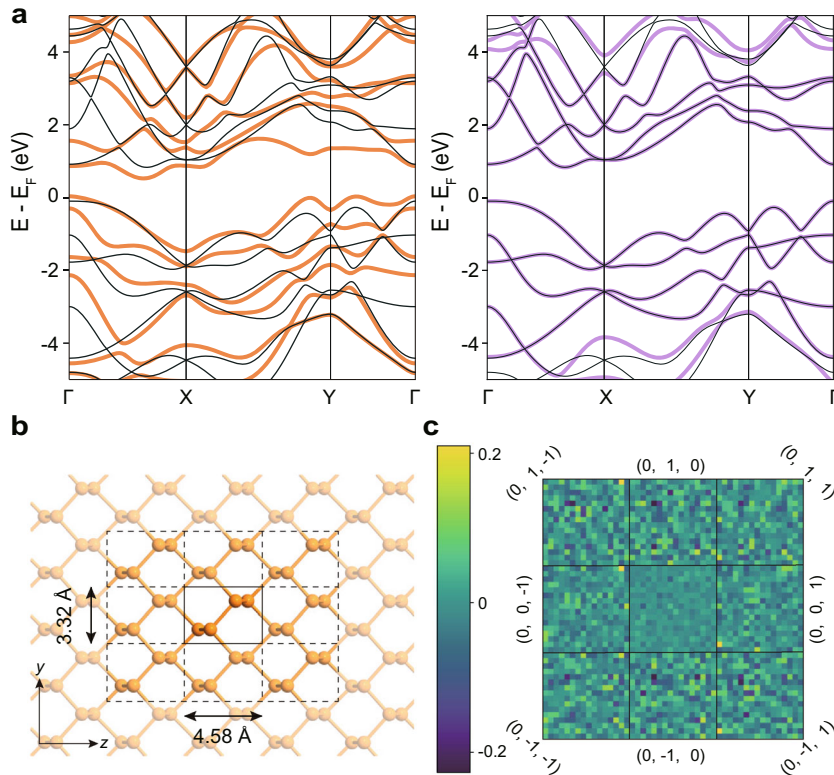


Fig. 6 The optimization for the *sp* TB model for the 2D phosphorene system using Variation I. **a** Comparison of the ab-initio band structure (black lines) and those calculated from the unoptimized WTb model (orange lines) and MLTB model (lilac lines) for the phosphorene. **b** The geometrical structure of the phosphorene. The unoptimized WTb model comprises the real-space Hamiltonian matrices describing the interactions within a unit cell (marked with the solid rectangle) and between this unit cell with its 8 neighbor cells (marked with the dotted rectangles). **c** The difference matrices between the considered real-space TB Hamiltonians before and after the optimization. The difference matrix of H^{00} is placed in the center while others are labeled with the lattice vectors alongside.

Slater–Koster parameters, such as $V_{ss\sigma}$, $V_{sp\sigma}$, and $V_{pp\pi}$, to obtain the matrix elements, using the two-center or three-center approximations proposed by Slater and Koster⁵. Variation II requires that the reference bands be analyzed in advance for choosing the types of atomic orbitals used for fitting the bands and thus for determining the fitting formulae. And the atomic coordinate information of the studied system needs to be obtained for providing the distance information used in the fitting formulae. This form is like the traditional methods for the determination of the Slater–Koster parameters. However, the gradient descent algorithm makes this method faster than the least-squares process used in traditional methods. During training, the types and number of the basis orbitals will be fixed, i.e., the fitting formulae and Hamiltonian matrix size will be held. And the Slater–Koster parameters will be optimized by adjusting the neuron values to better describe the ab-initio bands.

Taking the 13-atom-wide armchair graphene nanoribbon (13-AGNR) presented in Supplementary Note 1 as an example, we considered the p_z orbital of each C atom in the unit cell, and the real-space Hamiltonian matrices H^{00} , H^{0-1} , and H^{01} , which describe the interactions within a unit cell and between the unit cell with its 2 neighbor cells. With two-center approximations, the hopping parameters are expressed as⁴⁹:

$$V_{pp\pi}(r_{ij}) = a_1 r_{ij}^{-a_2} \exp(-a_3 r_{ij}^{a_4}), \quad (11)$$

where r_{ij} is the distance between atom i and j , which can be read directly from the real-space structure of the 13-AGNR, and a_1 , a_2 , a_3 , and a_4 are variable parameters, which will be represented by four different neurons, respectively. Still, the loss function was the

mean squared error between the ab-initio energy eigenvalues and the TB ones. After training, the optimal values of these parameters were obtained, and the TB Hamiltonian matrices calculated by these parameters were in the desired Slater–Koster form on the basis of p_z orbitals. Since much fewer neurons were used in Variation II than in the basic TBHCNN method, the produced TB Hamiltonian reproduced the ab-initio bands qualitatively. However, we advocate using Variation II for TB Hamiltonian parameterization when the information of the geometry is accessible, and the symmetry characters of reference bands are known, for analyzing the properties on which the symmetry characters have a significant impact. Variation II can be considered as a complement to our basic method to deal with its limitations on the match of the symmetry characters of the reference bands and on the reflection of geometric symmetry of the given system.

Figure 7 plots the comparison between the TB models obtained by using Variation II and the well-known empirical TB parameterization for graphene systems, which employs a value of -2.7 eV as the hopping parameter between the nearest-neighbor C atoms and 0 eV for otherwise. The Hamiltonian matrices of these two TB models have the same structures and slightly different matrix elements, and their TB band structures have similar accuracy.

Figure 8 shows the charge distribution of the 13-AGNR-MOSFET, which is presented in Supplementary Note 1. The device simulations were performed with the device Hamiltonians constructed by the produced TB model and a Wannier TB model from the MLWFs method. The comparison between the simulation results showed that the produced TB model could be used for qualitatively accurate real-space transport analysis.

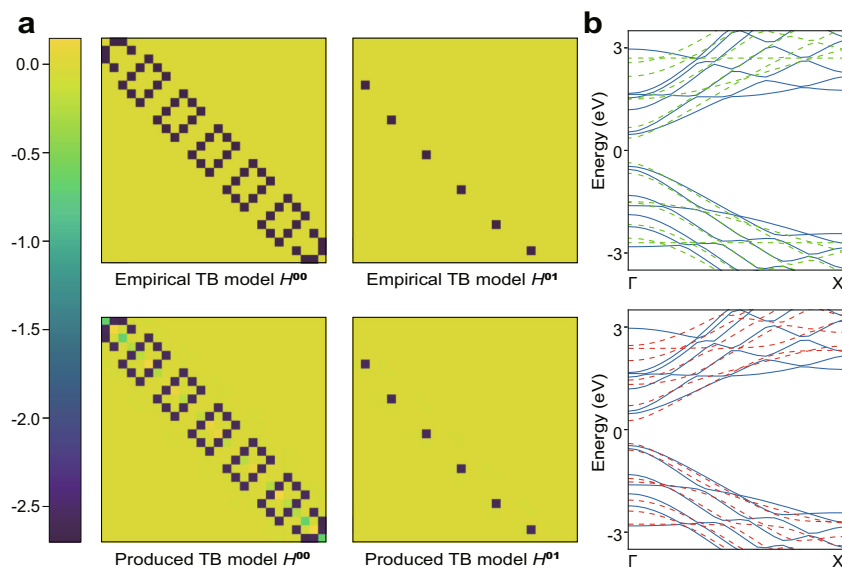


Fig. 7 The TB Hamiltonians for 13-AGNR in Slater–Koster form constructed from Variation II. **a** Comparison of the Hamiltonian elements of the produced TB model in Slater–Koster form and the empirical TB model. **b** Comparison of the ab-initio band structure (blue lines) and those calculated from the produced TB model (red dashed lines) and empirical TB model (green dashed lines) for 13-AGNR.



Fig. 8 Comparison of the charge distributions of the 13-AGNR-MOSFET using the WTB device Hamiltonian (above) and the produced TB device Hamiltonian (below). The gate voltage V_{GS} is set to be 0 V, and the drain-source voltage V_{DS} is set to be 0.2 V. The gray balls represent C atoms, whereas H atoms on the edges are not plotted.

DISCUSSION

In summary, we developed a generic method for TB Hamiltonian parameterization from the ab-initio band structure with ML algorithms. Our method's validity was tested through the calculations of the electronic structure of layered InSe and the transport characteristics of the InSe-based MOSFET. In the proposed case, our method surpassed the MLWFs method in both efficiency and accuracy. As an approach for TB model construction, the proposed method is free from prior knowledge requirements and truncation error. Additionally, the introduction of the one-to-one dynamic NN model can provide insights into applying ML methods to practical problems when it is difficult or even impossible to gather sufficient training data to build a general ML model. Also, two variations of our method were presented, which can to some extent deal with the mentioned limitations of our basic method, showing the flexibility of the proposed TBHCNN model. We believe our method cannot only promote the development of materials and device research but also help to improve the combination of ML techniques in the physics, chemistry, and materials science research communities.

METHODS

Neural network training

The proposed ML algorithm and neural network architecture can be conveniently achieved using mainstream ML platforms. In this work, we choose the widely used TensorFlow⁵⁰ framework, and employ its built-in Adam optimizer⁵¹ to perform relevant automatic differentiation tasks with an initial learning rate of $\alpha = 0.001$.

Ab-initio band structure calculation details

The ab-initio band structure calculations on the InSe nanoribbon, the 2D phosphorene, and the 13-AGNR, were carried out using the open-source ab-initio package QUANTUM ESPRESSO⁵². The exchange and correlation interactions between valence electrons are described by the Perdew–Burke–Ernzerhof (PBE) functional within the generalized gradient approximation (GGA)⁵³. The ultrasoft pseudopotential is used. The kinetic energy cutoff of the wave functions is 40 Ry, and the estimated energy error is less than 1×10^{-6} Ry.

For the InSe nanoribbon, in the chosen unit cell, the vacuum space is 20 Å, to ensure the interaction between periodic images can be safely avoided. The Brillouin zone is sampled with $1 \times 1 \times 10$ and $1 \times 1 \times 200$ Monkhorst–Pack⁵⁴ \mathbf{k} -points for the structure relaxation as well as self-consistent calculations and for the band structure calculations, respectively. The geometry is fully relaxed using the BFGS quasi-newton algorithm, with the criteria being that all components of all forces are smaller than 1×10^{-3} Ry/au, and the total energy changes are $< 1 \times 10^{-4}$ Ry.

For the 2D phosphorene, in the chosen unit cell, the vacuum space is 15 Å, to ensure the interaction between periodic images can be safely avoided. The Brillouin zone is sampled with $1 \times 10 \times 10$ Monkhorst–Pack \mathbf{k} -points for structure relaxation and self-consistent calculations. And 100 \mathbf{k} -points per \mathbf{k} -path are sampled uniformly for the band structure calculations on the three \mathbf{k} -paths: from Γ (0, 0, 0) to X (0, 1/2, 0), from X (0, 1/2, 0) to Y (0, 0, 1/2), and from Y (0, 0, 1/2) to Γ (0, 0, 0).

For the 13-AGNR, in the chosen unit cell, the vacuum space is 15 Å, to ensure the interaction between periodic images can be safely avoided. The Brillouin zone is sampled with $1 \times 1 \times 10$ and $1 \times 1 \times 50$ Monkhorst–Pack \mathbf{k} -points for the self-consistent calculations and the band structure calculations, respectively.

Here, we declare that the utilization of the calculation methods within the density functional theory (DFT) framework is not a strict request. When the effective single-particle band structures derived from DFT are qualitatively wrong, other more accurate methods for electronic structure calculations (e.g., DFT's many-body extensions like GW approximation) may be considered in place of the DFT method for obtaining the band structures. And It will not affect the existing framework of the proposed method since the TBHCNN model does not require a specific source of the energy band data.

MLWFs transformation details

The MLWFs method was implemented using Wannier90 software package⁵⁵.

For the InSe nanoribbon, the number of MLWFs is set to 208, which are initialized to the outermost s and p orbitals of In and Se atoms within the

unit cell. The outer energy window is set to $[-8 \text{ eV}, 6 \text{ eV}]$ and the inner energy window is set to $[-2.6 \text{ eV}, 0.4 \text{ eV}]$. The transformation is performed on top of the corresponding ab-initio calculation results with a $1 \times 1 \times 50$ Monkhorst–Pack \mathbf{k} -points sampling.

For the 2D phosphorene, the number of MLWFs is set to 16, which are initialized to the outermost s and p orbitals of P atoms within the unit cell. The outer energy window is set to $[-11.5 \text{ eV}, 6 \text{ eV}]$ and the inner energy window is set to $[-2.0 \text{ eV}, 0.0 \text{ eV}]$. The transformation is performed on top of the corresponding ab-initio calculation results with a $1 \times 10 \times 10$ Monkhorst–Pack \mathbf{k} -points sampling.

For the 13-AGNR, the number of MLWFs is set to 26, which are initialized to the p_z orbitals of C atoms within the unit cell. The outer energy window is set to $[-15.0 \text{ eV}, 9.0 \text{ eV}]$ and the inner energy window is set to $[-4.0 \text{ eV}, -2.0 \text{ eV}]$. The transformation is performed on top of the corresponding ab-initio calculation results with a $1 \times 1 \times 50$ Monkhorst–Pack \mathbf{k} -points sampling.

DATA AVAILABILITY

All the input files necessary to reproduce the ab-initio calculation results and transport property results presented in this paper are available on <https://github.com/whu-maple/tbhcnn>.

CODE AVAILABILITY

The codes developed in this paper are available on <https://github.com/whu-maple/tbhcnn>.

Received: 5 August 2020; Accepted: 28 December 2020;

Published online: 25 January 2021

REFERENCES

- Marin, E. G., Perucchini, M., Marian, D., Iannaccone, G. & Fiori, G. Modeling of electron devices based on 2-D materials. *IEEE Trans. Electron Devices* **65**, 4167–4179 (2018).
- Bloch, F. Über die Quantenmechanik der Elektronen in Kristallgittern. *Z. f.ür. Phys.* **52**, 555–600 (1929).
- Datta, S. Nanoscale device modeling: the Green's function method. *Superlattices Microstruct.* **28**, 253–278 (2000).
- Goringe, C. M., Bowler, D. R. & Hernandez, E. Tight-binding modelling of materials. *Rep. Prog. Phys.* **60**, 1447–1512 (1997).
- Slater, J. C. & Koster, G. F. Simplified LCAO method for the periodic potential problem. *Phys. Rev.* **94**, 1498–1524 (1954).
- Andersen, O. K. & Saha-Dasgupta, T. Muffin-tin orbitals of arbitrary order. *Phys. Rev. B* **62**, 16219–16222 (2000).
- Marzari, N., Mostofi, A. A., Yates, J. R., Souza, I. & Vanderbilt, D. Maximally localized Wannier functions: theory and applications. *Rev. Mod. Phys.* **84**, 57 (2012).
- Lu, W. C., Wang, C. Z., Chan, T. L., Ruedenberg, K. & Ho, K. M. Representation of electronic structures in crystals in terms of highly localized quasiatomic minimal basis orbitals. *Phys. Rev. B* **70**, 4 (2004).
- Qian, X. F. et al. Quasiatomic orbitals for ab initio tight-binding analysis. *Phys. Rev. B* **78**, 22 (2008).
- Agapito, L. A., Ferretti, A., Calzolari, A., Curtarolo, S. & Nardelli, M. B. Effective and accurate representation of extended Bloch states on finite Hilbert spaces. *Phys. Rev. B* **88**, 7 (2013).
- Fiori, G. & Iannaccone, G. Multiscale modeling for graphene-based nanoscale transistors. *Proc. IEEE* **101**, 1653–1669 (2013).
- Rudenko, A. N., Yuan, S. J. & Katsnelson, M. I. Toward a realistic description of multilayer black phosphorus: from GW approximation to large-scale tight-binding simulations. *Phys. Rev. B* **92**, 9 (2015).
- Lv, Y. W., Chang, S., Wang, H., He, J. & Huang, Q. J. Energy gap tunable graphene antidot nanoribbon MOSFET: a uniform multiscale analysis from band structure to transport properties. *Carbon* **101**, 143–151 (2016).
- Calogero, G., Papior, N., Koleini, M., Larsen, M. H. L. & Brandbyge, M. Multi-scale approach to first-principles electron transport beyond 100 nm. *Nanoscale* **11**, 6153–6164 (2019).
- Lv, Y. W. et al. Prediction of stable and high-performance charge transport in zigzag tellurene nanoribbons. *IEEE Trans. Electron Devices* **66**, 2365–2369 (2019).
- Agapito, L. A., Ismail-Beigi, S., Curtarolo, S., Fornari, M. & Nardelli, M. B. Accurate tight-binding Hamiltonian matrices from ab initio calculations: minimal basis sets. *Phys. Rev. B* **93**, 9 (2016).
- Agapito, L. A. et al. Accurate tight-binding Hamiltonians for two-dimensional and layered materials. *Phys. Rev. B* **93**, 8 (2016).
- D'Amico, P. et al. Accurate ab initio tight-binding Hamiltonians: effective tools for electronic transport and optical spectroscopy from first principles. *Phys. Rev. B* **94**, 10 (2016).
- Marzari, N. & Vanderbilt, D. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* **56**, 12847–12865 (1997).
- Souza, I., Marzari, N. & Vanderbilt, D. Maximally localized Wannier functions for entangled energy bands. *Phys. Rev. B* **65**, 13 (2002).
- Damle, A. & Lin, L. Disentanglement via entanglement: a unified method for Wannier localization. *Multiscale Model. Simul.* **16**, 1392–1410 (2018).
- Gresch, D. et al. Automated construction of symmetrized Wannier-like tight-binding models from ab initio calculations. *Phys. Rev. Mater.* **2**, 15 (2018).
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
- Hu, R. H., Chang, S., Wang, H., He, J. & Huang, Q. J. Efficient multispike learning for spiking neural networks using probability-modulated timing method. *IEEE Trans. Neural Netw. Learn. Syst.* **30**, 1984–1997 (2019).
- Huang, H. et al. A general group theoretical method to unfold band structures and its application. *N. J. Phys.* **16**, 033034 (2014).
- Ku, W., Berlijn, T. & Lee, C.-C. Unfolding first-principles band structures. *Phys. Rev. Lett.* **104**, 216401 (2010).
- Giles, M. B. Collected matrix derivative results for forward and reverse mode algorithmic differentiation. *Adv. Autom. Differ.* **64**, 35–44 (2008).
- Lee, D. H. & Joannopoulos, J. D. Simple scheme for surface-band calculations. *Phys. Rev. B* **23**, 4988–4996 (1981).
- Lihm, J. M. & Park, C. H. Reliable methods for seamless stitching of tight-binding models based on maximally localized Wannier functions. *Phys. Rev. B* **99**, 16 (2019).
- Sucharitakul, S. et al. Intrinsic electron mobility exceeding $10(3) \text{ cm}^2(\text{V s})$ in multilayer InSe FETs. *Nano Lett.* **15**, 3815–3819 (2015).
- Bandurin, D. A. et al. High electron mobility, quantum Hall effect and anomalous optical response in atomically thin InSe. *Nat. Nanotechnol.* **12**, 223–227 (2017).
- Wu, M. et al. Modulation of electronic and magnetic properties in InSe nanoribbons: edge effect. *Nanotechnology* **29**, 205708 (2018).
- Feng, W., Zhou, X., Tian, W. Q., Zheng, W. & Hu, P. A. Performance improvement of multilayer InSe transistors with optimized metal contacts. *Phys. Chem. Chem. Phys.* **17**, 3653–3658 (2015).
- Lei, S. D. et al. Evolution of the electronic band structure and efficient photo-detection in atomic layers of InSe. *ACS Nano* **8**, 1263–1272 (2014).
- Magorrian, S. J., Zolyomi, V. & Fal'ko, V. I. Electronic and optical properties of two-dimensional InSe from a DFT-parametrized tight-binding model. *Phys. Rev. B* **94**, 20 (2016).
- Fiori, G. & Iannaccone, G. Simulation of graphene nanoribbon field-effect transistors. *IEEE Electron Device Lett.* **28**, 760–762 (2007).
- Peano, V., Sapper, F. & Marquardt, F. Rapid exploration of topological band structures using deep learning. Preprint at <https://arxiv.org/abs/1912.03296> (2019).
- Hegde, G. & Bowen, R. C. Machine-learned approximations to density functional theory Hamiltonians. *Sci. Rep.* **7**, 11 (2017).
- Kranz, J. J., Kubillus, M., Ramakrishnan, R., von Lilienfeld, O. A. & Elstner, M. Generalized density-functional tight-binding repulsive potentials from unsupervised machine learning. *J. Chem. Theory Comput.* **14**, 2341–2352 (2018).
- Zhu, J. M., Vuong, V. Q., Sumpter, B. G. & Irlé, S. Artificial neural network correction for density-functional tight-binding molecular dynamics simulations. *MRS Commun.* **9**, 867–873 (2019).
- Dral, P. O., von Lilienfeld, O. A. & Thiel, W. Machine learning of parameters for accurate semiempirical quantum chemical calculations. *J. Chem. Theory Comput.* **11**, 2120–2125 (2015).
- Li, H., Collins, C., Tanha, M., Gordon, G. J. & Yaron, D. J. A density functional tight binding layer for deep learning of chemical Hamiltonians. *J. Chem. Theory Comput.* **14**, 5764–5776 (2018).
- Welborn, M., Cheng, L. & Miller, T. F. Transferability in machine learning for electronic structure via the molecular orbital basis. *J. Chem. Theory Comput.* **14**, 4772–4779 (2018).
- Stöhr, M., Medrano Sandonas, L. & Tkatchenko, A. Accurate many-body repulsive potentials for density-functional tight binding from deep tensor neural networks. *J. Phys. Chem. Lett.* **11**, 6835–6843 (2020).
- Panosetti, C., Engelmann, A., Nemeč, L., Reuter, K. & Margraf, J. T. Learning to use the force: fitting repulsive potentials in density-functional tight-binding with gaussian process regression. *J. Chem. Theory Comput.* **16**, 2181–2191 (2020).
- Huran, A. W., Steigemann, C., Frauenheim, T., Aradi, B. & Marques, M. A. L. Efficient automatized density-functional tight-binding parametrizations: application to group IV elements. *J. Chem. Theory Comput.* **14**, 2947–2954 (2018).

47. Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
48. Rudenko, A. N. & Katsnelson, M. I. Quasiparticle band structure and tight-binding model for single- and bilayer black phosphorus. *Phys. Rev. B* **89**, 5 (2014).
49. Tang, M. S., Wang, C. Z., Chan, C. T. & Ho, K. M. Environment-dependent tight-binding potential model. *Phys. Rev. B* **53**, 979–982 (1996).
50. Abadi, M. et al. TensorFlow: a system for large-scale machine learning. In *Proc. 12th USENIX Conference on Operating Systems Design and Implementation* Vol. 16, 265–283 (2016).
51. Kingma, D. & Ba, J. Adam: a method for stochastic optimization. In *Proc. 3rd Int. Conf. Learn. Represent* (2015).
52. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condes. Matter* **21**, 19 (2009).
53. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
54. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).
55. Mostofi, A. A. et al. wannier90: A tool for obtaining maximally-localised Wannier functions. *Comput. Phys. Commun.* **178**, 685–699 (2008).

ACKNOWLEDGEMENTS

We thank the supercomputing system in the Supercomputing Center of Wuhan University, on which the numerical calculations in this paper have been performed. We acknowledge support from the National Natural Science Foundation of China (61874079, 62074116, 81971702, and 61774113), the Wuhan Research Program of Application Foundation (2018010401011289), and the Luojia Young Scholars Program.

AUTHOR CONTRIBUTIONS

Z.W. and S.C. designed the algorithms of this work. Z.W. and S.Y. coded the TBHCNN. All authors discussed the results, commented on the manuscript, and edited the manuscript before submission.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41524-020-00490-5>.

Correspondence and requests for materials should be addressed to S.C.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021