Article

Neural network representation of electronic structure from *ab initio* molecular dynamics

Qiangqiang Gu [a], Linfeng Zhang [b], Ji Feng [a,c,*]

[a] *International Center for Quantum Materials, School of Physics, Peking University, Beijing 100871, China*
[b] *Department of Mathematics and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA*
[c] *Collaborative Innovation Center of Quantum Matter, Beijing 100871, China*

ARTICLE INFO

ABSTRACT

Despite their rich information content, electronic structure data amassed at high volumes in *ab initio* molecular dynamics simulations are generally under-utilized. We introduce a transferable high-fidelity neural network representation of such data in the form of tight-binding Hamiltonians for crystalline materials. This predictive representation of *ab initio* electronic structure, combined with machine-learning boosted molecular dynamics, enables efficient and accurate electronic evolution and sampling. When it is applied to a one-dimension charge-density wave material, carbyne, we are able to compute the spectral function and optical conductivity in the canonical ensemble. The spectral functions evaluated during soliton-antisoliton pair annihilation process reveal significant renormalization of low-energy edge modes due to retarded electron-lattice coupling beyond the Born–Oppenheimer limit. The availability of an efficient and reusable surrogate model for the electronic structure dynamical system will enable calculating many interesting physical properties, paving the way to previously inaccessible or challenging avenues in materials modeling.

© 2021 Science China Press. Published by Elsevier B.V. and Science China Press. All rights reserved.

1. Introduction

Ab initio molecular dynamics (AIMD) is an atomistic simulation method based on *ab initio* electronic structures, and a versatile tool for quantitatively understanding thermodynamic and kinetic behaviors of a wide range of matter, including molecules, liquids, and solids [1,2]. In an AIMD simulation, a huge number of atomic configurations have to be sampled for meaningful statistics, during which the Born–Oppenheimer (BO) electronic structure as a function of atomic trajectory is evaluated at *ab initio* level. Although a number of electronic properties can be computed on-the-fly, the electronic structure data, generated at high volumes and significant computation costs, are generally under-utilized, despite exceedingly rich information in these data. This is precisely a situation where a concise and predictive representation of the *ab initio* electronic structure data to harvest the electronic information discarded after AIMD is expected to be highly rewarding. This will constitute a reusable surrogate model, which can significantly expedite the sampling of configuration-dependent electronic structure.

The above problem falls squarely in the realm of the deep neural network, which has become a powerful and versatile paradigm for computerized analysis and interpretation of massive data [3]. Through supervised observation of labeled data, a neural network automatically recognizes patterns and regularities in the data and hence becomes predictive regarding the features it is devised to represent. In turn, the predictive neural network can substitute the originally time-consuming calculation with significant speedup. For example, a neural network representation of the interatomic potential has been developed and can reproduce the atomic trajectory without time-consuming *ab initio* calculations [4–8]. In these methods, the total potential energy is decomposed into a sum of atomic energy contributions, which is represented by neural networks through the atom and its neighbors (local chemical environment). However, the electronic information available in an AIMD is entirely lost in this approach. Therefore, a predictive representation of electronic structure to pick up the rich electronic structure information in AIMD simulation is highly desirable, which will enable many previous inaccessible or challenging tasks in modeling the electronic properties of a large and dynamical system.

Recently, progress has been made in surveying electronic properties using machine learning (ML) techniques, focusing exclusively on specific materials properties from structural

---

information, such as the electronic density of states [9] and charge density [10], or focusing on electronic Hamiltonians for small molecular systems [11]. A general ML approach that predicts a microscopic electronic Hamiltonian for crystalline materials from structural information is still lacking, which shall enable direct computation of physical properties with the high efficiency offered by ML techniques. In this paper, we present a neural network architecture to efficiently predict a tight-binding Hamiltonian (dubbed TBworks hereafter) to represent the electronic structure data from AIMD simulations. This architecture will employ a deep feedforward network to process atom trajectory and band structures, to produce a representation that predicts a tight-binding Hamiltonian for a given snapshot of atomic configuration. The main challenge in our method is that the neural network predicted tight-binding Hamiltonian is not uniquely determined by its band structure due to the $U(N)$ gauge freedom, which calls for much effort in designing the neural network architecture and optimization process. It is shown that by the elaborate design of the neural network architecture, the gauge freedom can be reduced. After introducing the architecture and algorithm, the method is tested using a one-dimension charge-density-wave (CDW) material, carbyne. It is shown that the *ab initio* electronic structure can be faithfully reproduced by the predicted Hamiltonian.

As an example, large-scale molecular dynamics (MD) simulations are performed on carbyne based on the ML force field, during which the TBworks is applied to sample the electronic structures and derived physical properties. This enables us to sample the electronic states in the dynamical annihilation of a soliton-antisoliton pair to compute the time-dependent spectral function, beyond the Born–Oppenheimer approximation. Optical conductivity is also obtained by sampling the current–current correlation function in a path-integral molecular dynamics. These results demonstrate that our TBworks, combined with ML force field, realizes a paradigm for expedited electronic structure sampling at *ab initio* accuracy, which can capture the non-adiabatic effects in transient processes by large-scale and long-time simulations.

2. Neural Network Architecture

In an AIMD simulation, a trajectory of ion positions is generated at a discrete set of time points, $\boldsymbol{u}_i(t) = \boldsymbol{r}_i(t) - \bar{\boldsymbol{r}}_i$, where $\boldsymbol{r}_i(t)$ and $\bar{\boldsymbol{r}}_i$ are respectively the instantaneous and the reference ion positions. Periodic boundary conditions are imposed on the simulation box, such that the *ab initio* eigenstates are Bloch functions. An AIMD yields a set of instantaneous electronic eigenstates for an ionic configuration $\boldsymbol{r} = \{\boldsymbol{r}_i\}$, with corresponding energy spectrum (bands) $\hat{\varepsilon}(\boldsymbol{r}) = \{\hat{\varepsilon}_{n\mathbf{k}}\}$. The aim is then to use $\boldsymbol{r}$ and $\hat{\varepsilon}(\boldsymbol{r})$ to train a neural network as a surrogate model for the *ab initio* electronic structure, through a tight-binding electronic Hamiltonian as a function of $\boldsymbol{r}$.

$$H(\boldsymbol{r}) = \sum_{i \neq j} V_{ij}(\boldsymbol{r})c_i^\dagger c_j + \sum_i V_{ii}(\boldsymbol{r})c_i^\dagger c_i + \text{H.c.}, \tag{1}$$

where $V_{ij}(\boldsymbol{r})$ $(i \neq j)$ is a hopping matrix element and $V_{ii}(\boldsymbol{r})$ is the on-site energy. Assuming that there is only a single orbital on each ion for notational brevity, i and j are also site indices.

Providing the existence of Wannier functions [12–14], it suffices to keep hopping matrix elements between pairs of orbitals within a finite range. Therefore, we introduce a *local chemical environment* for each site, as shown in Fig. 1. For site i, the ions lying within a sphere of radius R_{cut} centered at $\mathbf{r}_i$ form its local chemical environment $L_i = \{j \mid |\boldsymbol{r}_j - \boldsymbol{r}_i| < R_{\text{cut}}\}$, which assumes non-negligible influence on the Wannier functions on site i. Numerically, the local environment L_i is represented by the set $D_i = \{\boldsymbol{r}_{ji}|j \in L_i\}$, where $\boldsymbol{r}_{ji} = \boldsymbol{r}_j - \boldsymbol{r}_i$. When fed to neural network, the elements in set D_i are sorted by distances $|\boldsymbol{r}_{ji}|$ in ascending order. Therefore, the hop-

ping matrix element V_{ij} will depend on $\bar{\boldsymbol{r}}_{ji}, \boldsymbol{u}_i, \boldsymbol{u}_j, D_i$, and D_j, where $\bar{\boldsymbol{r}}_{ji} = \bar{\boldsymbol{r}}_j - \bar{\boldsymbol{r}}_i$. The on-site energy V_{ii} can be fully determined by D_i. A naïve neural network for our purpose is composed of two maps (see Fig. 1): $V_{ij} = \Phi(X; \theta)$ and $V_{ii} = \Psi(Y; \xi)$, where $X = (\bar{\boldsymbol{r}}_{ji}, \boldsymbol{u}_i, \boldsymbol{u}_j, D_i, D_j), Y = D_i$. θ and ξ are trainable parameters of the neural network. The tight-binding Hamiltonian $H(\bar{\boldsymbol{r}})$ for reference configuration (usually an equilibrium crystal structure) is easily accessible with standard Wannierization method [15], which is used to calibrate the neural network in a refined architecture,

$$V_{ij} = \overline{V}_{ij} + \Phi(X; \theta) - \Phi(\overline{X}; \theta), V_{ii} = \overline{V}_{ii} + \Psi(Y; \xi) - \Psi(\overline{Y}; \xi). \tag{2}$$

Obviously, when $\boldsymbol{r} = \bar{\boldsymbol{r}}, \Phi(X; \theta) - \Phi(\overline{X}; \theta)$ and $\Psi(Y; \xi) - \Psi(\overline{Y}; \xi)$ equal zero for the neural network to produce exactly $H(\bar{\boldsymbol{r}})$. Although we only show the neural network maps for the case of single orbital on each ion for brevity, generalization to multi-orbital case is straightforward by allocating individual maps for hoppings between different orbitals.

The maps Φ and Ψ are fulfilled through deep feedforward neural networks with n hidden layers neurons, as shown in Fig. 1. Take the map Φ as an example, $\Phi = \Phi_n \circ \Phi_{n-1} \circ \ldots \circ \Phi_1$. The μ-th hidden layer, Φ_μ takes the output vector of the preceding layer $x_{\mu-1}$, and generates a new vector through the component-wise hyperbolic tangent activation, $x_\mu = \tanh(w_\mu x_{\mu-1} + b_\mu)$, with $x_0 = X$, and $V_{ij} = x_n$. Here the matrix w_μ reweights $x_{\mu-1}$ and the vector b_μ provides a bias, both of which are adjustable parameters and determined via the Adam stochastic gradient descent method [16] employed to minimize the loss function during a back-propagation process. The loss function is the root-mean-square (rms) deviation between energy spectra $\varepsilon(\boldsymbol{r})$ computed from predicted $H(\boldsymbol{r})$ and training labels from the *ab initio* $\hat{\varepsilon}(\boldsymbol{r})$. In principle, $H(\boldsymbol{r})$ is not uniquely determined by energy spectra $\hat{\varepsilon}(\boldsymbol{r})$ alone due to the existence of gauge freedom. If a different gauge is chosen at different $\boldsymbol{r}, H(\boldsymbol{r})$ as a function of $\boldsymbol{r}$ may not necessarily be smooth and continuous. Because of the discontinuity, the predicted $H(\boldsymbol{r})$ may not interpolate well for unseen configurations, not to mention extrapolations. Fortunately, the continuity of $H(\boldsymbol{r})$ is fulfilled automatically in our neural network implementation, since the relationships between $\boldsymbol{r}$ and $H(\boldsymbol{r})$ are represented by Φ and Ψ, which are smooth and continuous composite maps. A simple test of whether our method is plagued by the gauge uncertainty is its ability to generalize. As will be shown next, our TBworks generalizes accurately on all the independent testing data sets.

A couple of remarks are warranted at this point, on how the method outlined above is different from some of the related existing techniques. First, our TBworks and ML force fields methods [4–8] offer completely different surrogate models for AIMD simulations. Whereas ML force fields predict the interactions potential energies and forces between ions, TBworks predicts electronic Hamiltonians for given ion configurations. Therefore, although both methods are designed to remove the need for time-consuming AIMD, ML force fields are fashioned to accelerate molecular dynamics, while our TBworks will enable fast sampling of electronic structure, for instance, in a ML-force-field accelerated MD. Second, TBworks is not the semi-empirical density-functional tight-binding method [17,18]. The latter is designed to speed up a density-functional theory calculation with the help of an approximate tight-binding Hamiltonian. Our TBworks, on other hand, produces a very accurate representation of Kohn–Sham electronic structure itself.

To sum up, we have outlined the essential idea of TBworks in this section. In the following sections, we will present an exemplary application of the TBworks on a simple yet nontrivial material, to demonstrate that TBworks can accurately and efficiently predict electronic structure of a dynamical system to facilitate
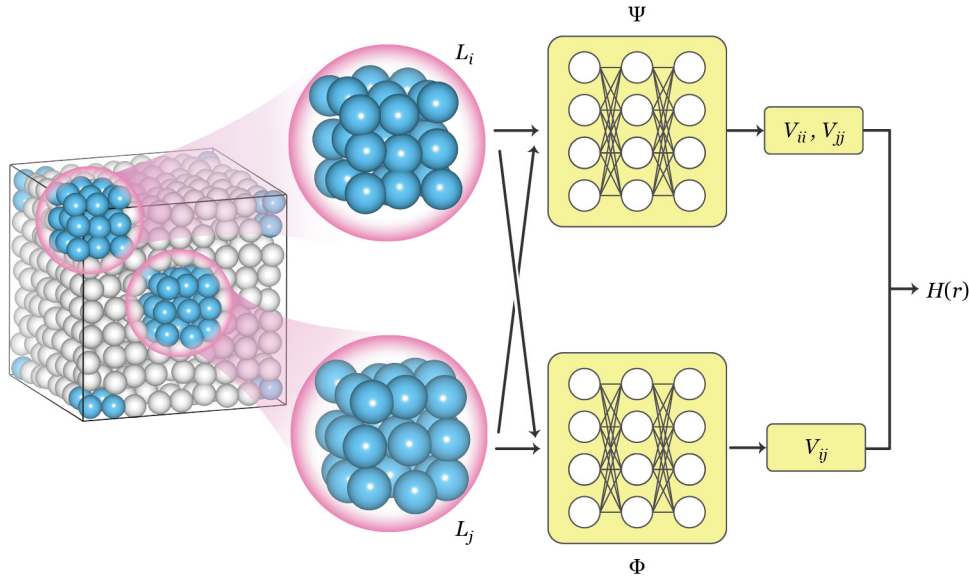
Section S-3 and S-4. The generalizability and transferability of TBworks are validated against multiple unlearned data sets (5000 AIMD snapshots each) from multiple T and N. We use the term generalizability to denote the ability of TBworks to predict accurately for unseen data for the same chemical compound, under the same conditions (system size N and temperature T) as training data. In contrast, by transferability we mean the robustness of TBworks for the same compound but under varied conditions (N or T, in this context).

In carbyne, the C–C σ bonds are formed by the overlap of sp^1 hybridized orbitals while the π bonds are formed by the other two half-filled p orbitals perpendicular to the carbon chain. The s and p bands within the energy window ($-22 - 0$ eV) are picked out as label data from 5000 AIMD snapshots of $N = 32$ at $T = 150$ K to train the 3-layer TBworks with the [200, 200, 200] neurons. Fig. 2b displays a faithful reproduction of the band structure of polyyne by the trained TBworks, including the CDW gap $\Delta \sim 0.27$ eV. In Fig. 2c, the TBworks predicted spectra for snapshots unseen in training are plotted against *ab initio* spectra, where samples are drawn from AIMD with multiple temperatures (10 – 200 K) and a range of simulation box sizes. The testing error (rms deviation) of these predictions is ~ 1 meV and the coefficient of determination $R^2 = 0.99999994$. The s-p separation of ~ 3.72 eV and the CDW gap ~ 0.27 eV are also well reproduced despite their disparate energy scales. As is also shown in Fig. 2d, the generalization errors of the TBworks trained at a single T and N are on the order of 1 meV, when tested against data from multiple temperatures and simulation box sizes.

Further generalization tests have been performed on a long-chain carbyne with $N = 256$, without or with a soliton pair. The order parameter values of several snapshots from the MD trajectory at $T = 10$ K based on the ML force fields [8] are shown in Fig. 3a. Three archetypal structures, polyyne (top), isolated soliton/antisoliton (middle) and coupled soliton-antisoliton pair (bottom), are selected for the tests. The band structures and density of states from TBworks and *ab initio* benchmarks for the three representative configurations are respectively shown in Fig. 3b–d. Evidently, the TBworks faithfully reproduces full energy spectra in the low-energy regime. In particular, the CDW gap in polyyne structure (Fig. 3b), the zero-energy soliton mode in isolated soliton/antisoliton structure (Fig. 3c) are accurately reproduced. So are the incipient in-gap modes with a finite splitting when the soliton and antisoliton experience coupling at short separations, as shown in Fig. 3d. Therefore, we conclude that the TBworks trained at a single T and N offers a high fidelity representation of the *ab initio* electronic structure, which is also highly transferable to a range of temperatures, system sizes, and especially, structural variety, unseen in the initial learning process.

With the accuracy perfectly preserved, we further investigate the efficiency of the TBworks model. The CPU times are obtained on a compute node equipped with 2 AMD EPYC-7452 CPUs (64 cores). As shown in Fig. 4, due to the inevitable diagonalization procedure, the computation cost of both the TBworks and *ab initio* calculations scale roughly as $\sim N^3$, where N denotes the system size. However, the prefactor is different by more than 5 orders of magnitude. This is because in *ab initio* calculations, diagonalization happens in solving the Kohn–Sham equations [29], where the one-electron wavefunctions are expanded over a large basis set of size proportional to but way larger than N. In contrast, the size of the TBworks Hamiltonian is the number of Wannier orbitals, which is generally of the same order of N. As such, the dramatic speed-up in TBworks makes the accessible system size around 2 orders of magnitude larger than *ab initio* calculations.

Further insights into the performance of the TBworks are gained by visualizing the training process on a reduced information plane
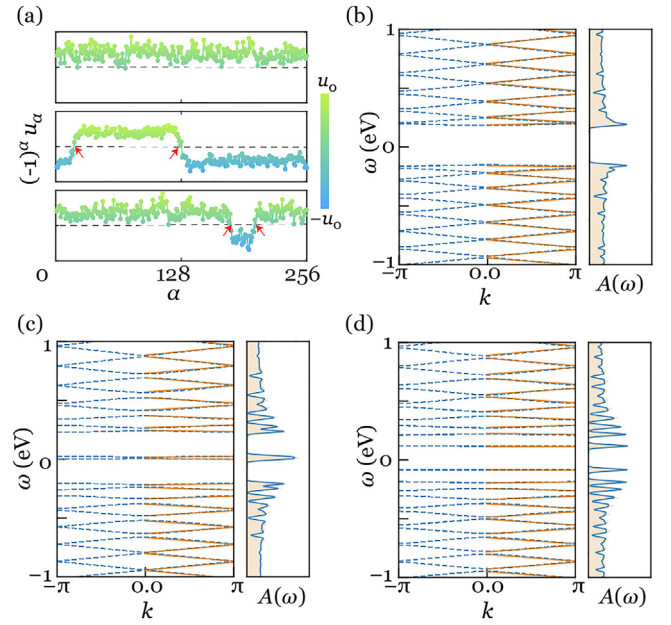
Fig. 3. (Color online) Comparison between TBworks predicted and *ab initio* calculated electronic structure, for three snapshots in a MD simulation of a 256-site carbyne chain. (a) Order parameters of the three configurations. Top: a polyyne configuration; Middle: an isolated soliton pair; And bottom: a coupled soliton pair. The red arrow point to the soliton locations. (b–d) The TBworks-predicted (blue line) vs. *ab initio* (yellow line and zone) band structures and density of states of the three configurations in (a), in the same order.
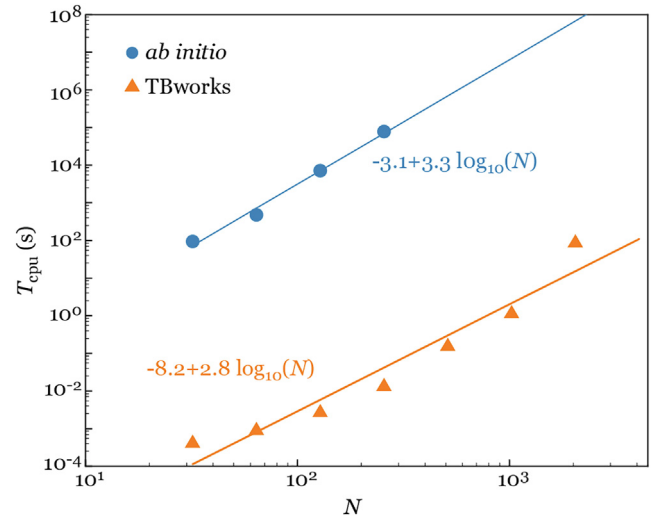
Fig. 4. (Color online) Compute time T_{cpu} for calculating the electronic spectra of carbyne vs. system size N in *ab initio* and TBworks calculations. The *ab initio* calculation is performed at Γ point [25], with planewave basis cutoff at 400 eV and total energy convergence to within 10^{-4} eV per atom. The dots (*ab initio*) and triangles (TBworks) mark the computation costs for different N. The solid lines are the least-square linear fittings between variable $\log_{10}(T_{cpu})$ and $\log_{10}(N)$.

[30,31]. The mutual information $I(a; b)$ is an asymmetric measure of the correlation between a pair of random variables a and b, with distributions $P(a)$ and $P(b)$, respectively,

$$I(a; b) = D_{KL}(P(a, b) || P(a) \otimes P(b)), \qquad (3)$$

where D_{KL} is the Kullback–Leibler divergence and $P(a, b)$ is the joint distribution [32]. We compute the mutual information between the input x (ionic coordinates inside a simulation box) and the output of the n-th layer y_n, and the mutual information between the training

label $\hat{\varepsilon}$ and the output y_n. Thus we can visualize the evolution of the neural network on the reduced information plane $I(x, y_n)$-$I(\hat{\varepsilon}, y_n)$, as depicted in Fig. 5. We see that $I(\hat{\varepsilon}; y_1)$ undergoes a continuous sharp increase throughout the training, which directly results from the optimization of the loss function, indicating continued improvement of model fitting. $I(x, y_1)$ increases slightly then decreases in the later stage. The decrease in $I(x, y_1)$ suggests a reduction in data redundancy, which is very important for the performance of the neural network. The dynamics of layers 2 and 3 are also bi-phasic, where redundancy elimination precedes the rapid model improvement. The redundant information on x arises from a large cut-off radius R_{cut} for the local chemical environment, which includes atoms having little impact on the hopping amplitude. Therefore, the information plane analysis uncovers critical stages in the learning process of neural network during the training process, and offers leads for systematically optimizing the choice of setup parameter for TBworks.

4. Applications to kinetic and equilibrium samplings

The efficient, accurate and transferable access to $H(\mathbf{r})$ offered by our TBworks provides a convenient computational tool for investigating the correlation between electronic and ionic degrees of freedom, which is particularly important for systems with strong electron–phonon coupling. As a first demonstration, TBworks is employed to evaluate correlation functions and corresponding physical properties of dynamical (both transient and equilibrium) systems. The detailed discussions centered around an application-oriented and task-specific example will help with an in-depth understanding of some of the issues in the implementation of the TBworks. In the first example, we compute the time-dependent spectral function of a kinetic process of soliton annihilation, whereupon non-adiabatic effects are revealed. As a second example, it is demonstrated TBworks can be combined with path-integral molecular dynamics to compute the current–current correlation function and therefore the optical conductivity. In this section, The evolution of ions is computed using ML force-field MD [8] within the BO approximation. TBworks is used to predict the electronic Hamiltonian along ion trajectories from MD, with which the electrodynamics is obtained by numerically integrating the Schrödinger equation. This approach allows us to capture non-BO effects on the electronic structure and responses of the electronic system.[1]

The soliton-antisoliton pair annihilation process is one such example, where the zero-energy electronic modes are gapped out dynamically. Non-adiabatic effects can be important but difficult to study in AIMD due to the high demand for computational resources. In order to study this kind of process, we interface our TBworks with the ML force field [8] to perform highly efficient classical molecular dynamics (CMD) and path-integral molecular dynamics (PIMD) on carbyne by the large-scale atomic/molecular massively parallel simulator (LAMMPS) [33]. The ML force field is generated by DeePMD-kit [34] using a $200 \times 200 \times 200$ deep neural network trained by 10000 snapshots of labeled AIMD data and tested by a validation set containing 1000 snapshots. The rms error is found to be within 0.01 eV/Å for predicted forces, and 10^{-5} eV per atom for predicted energies, which means the accuracy of the ML force field is at the same level of the *ab initio* calculation. Both the classical and path-integral MD simulations are performed in the canonical ensemble on a supercell $N = 1024$ with a time step of 0.2 femtoseconds (fs) at different temperatures. The Nose–Hoover thermostat [27,28,35] is employed to control the temperatures in the canonical ensemble. In PIMD, the paths are discretized
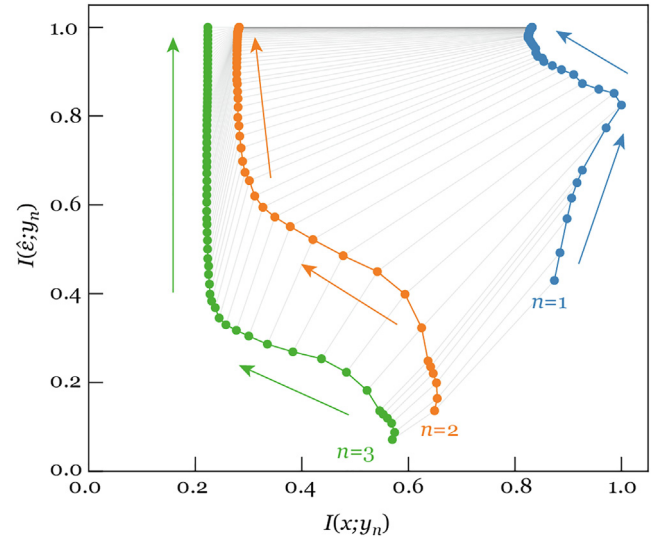
---

[1] Though no nonadiabatic effects on the ion dynamics is included in this approach.

Fig. 5. (Color online) Information plane loci of a 3-layer TBworks during training, showing the mutual information trajectories of a [200, 200, 200] neural network in 800 epoches. The colored solid arrows are guide-of-eye indications of the direction of progression for the three loci.

into 12 imaginary time slices. Fig. 6a shows the order parameters u_α of the ionic configurations in the dynamical process where a soliton-antisoliton pair is annihilated in the classical MD at $T = 1$ K.

To highlight the non-adiabatic effects in the soliton pair annihilation, we inspect the time-dependent electronic spectral functions. Because of the expedited access to TBworks Hamiltonian, we are able to let the electronic system evolve according to $\mathcal{T} \exp[-i \int_{t_0}^{t} H(r(t'))dt']$ $(\hbar = 1)$. Doing so explicitly takes ion dynamics into account and is manifestly beyond the BO approximation. As ion dynamics breaks time homogeneity for electrons, we define a time-dependent spectral function as

$$A(\omega, t) = \frac{i}{2\pi N} \hat{F} \text{Tr} \left[G^R(t + \tau, t) - G^A(t + \tau, t) \right], \qquad (4)$$

where $G_{\alpha\beta}^{A/R}(t + \tau, t) = \pm i \Theta(\mp \tau) \left\langle \left\{ c_\alpha(t + \tau), c_\beta^\dagger(t) \right\} \right\rangle$ are advanced/retarded Green's functions, and $\hat{F}$ means a Fourier transform w.r.t. the time difference τ [36]. At different t, the advanced and retarded Green's functions are evaluated by computing the expectation of the commutator $\left\{ c_\alpha(t + \tau), c_\beta^\dagger(t) \right\}$ for $\tau > 0$ and $\tau < 0$, respectively, with $|\tau| \leqslant 600$ for an energy resolution of ~ 3 meV ($\sim 1\%$ of the CDW gap). As the exact propagator from the TBworks Hamiltonian can be quickly computed, the $A(\omega, t)$ so obtained can be non-adiabatically calculated along a trajectory in the MD simulation. As a comparison, the static electronic spectral function within the BO approximation for configuration at t is calculated using instantaneous energy spectra, with $A_{BO}(\omega, t) = \frac{1}{N_k} \sum_{nk} \delta(\omega - \varepsilon_{nk}(t))$. We focus on the soliton-antisoliton pair annihilation event indicated by the dashed rectangle in Fig. 6a. During this event, two kinks meet and annihilate, and the polyyne-B phase completely disappears. At the same time, the degenerate zero-energy modes localized at the kinks become hybridized and a gap gradually develops. The question is whether the BO approximation is accurate enough to describe the electronic states for the dynamical annihilation process. As a comparison, we first examined the electronic spectral functions for structures well before (Fig. 6c) and after (Fig. 6d) the annihilation event. Prior to the annihilation event where the soliton and anti-soliton are well separated, both dynamical (beyond the BO
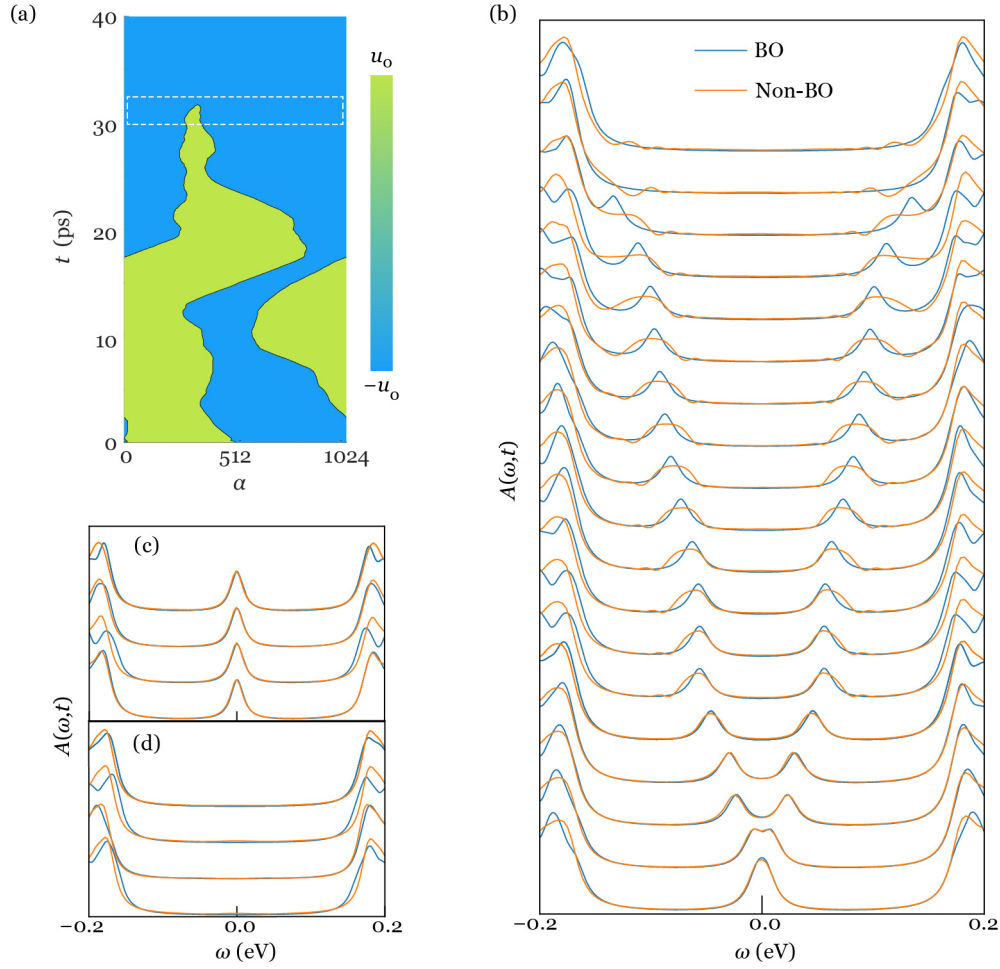
Fig. 6. (Color online) Non-adiabatic dynamical effects on the electronic spectral functions during the soliton-antisoliton pair annihilation process. (a) The order parameters along a polyyne chain as a function of time in a molecular dynamics trajectory. The dashed rectangle indicated the soliton-antisoliton pair annihilation event. (b) Computed time-dependent electronic spectral functions for structures at every 20 fs in the annihilation event. (c), (d) Computed electronic spectral functions for structures before (c) and after (d) annihilation event.

approximation) and static (within the BO approximation) spectral functions give the same topological protected zero-energy peaks, as shown in Fig. 6c, where little difference between the dynamical and static spectral functions is discernible. Post annihilation, the carbyne chain adopts a homogeneous polyyne-A structure, where both dynamical and static spectral functions present the CDW gap, as shown in Fig. 6d. Again, little difference between the dynamical and static spectral functions is uncovered, because the system is in a (transient) stationary state.

Interestingly, in the course of this annihilation event, the dynamical and static time-dependent spectral functions for the annihilation process display remarkable departures, as shown in Fig. 6b. Both dynamical and static spectral functions show the splitting of the zero-energy modes into two peaks, which subsequently merge into the conduction and valence bands, respectively, as the soliton pair approaching annihilation. However, the low-energy peaks in the dynamical spectral functions show significantly larger broadening than those in the static spectral functions. Thus, although the dynamical effects shows little renormalization on the spectral functions before and after the annihilation event, the non-adiabatic dynamical effects are clearly visible in the splitting of the topological in-gap modes in our calculations.

The non-adiabatic effects on the low-energy modes can be revealed in the spectral function projected on atomic sites,

$$A(\omega, \alpha) = \frac{i}{2\pi N}\left[G_{\alpha\alpha}^{R}(\omega) - G_{\alpha\alpha}^{A}(\omega)\right], \quad (5)$$

where $G_{\alpha\alpha}^{A/R}(\omega)$ are the on-site advanced/retarded Green's functions and α is the atomic index. For brevity, we leave out the index t shown in Eq. (4). For comparison, the static spectral function in the BO approximation is calculated from the instantaneous energy spectra and wavefunctions with $A_{BO}(\omega, \alpha) = \frac{1}{N_k}\sum_{nk}\delta(\omega - \varepsilon_{nk})|a_{n\alpha}|^2$. where the $a_{n\alpha}$ is the expansion coefficient of n-th wavefunction in the atomic orbital basis.

Both dynamical and static spectral functions are calculated for a carbyne configuration where the soliton and anti-soliton are moving toward each other before annihilation simulated in MD with $N = 1024$ and $T = 1$ K. As order parameter plotted in upper panel of Fig. 7b indicates, the soliton and antisoliton are about 60 sites apart and there is a significant coupling. As shown in Fig. 7a the low-energy modes are split because of the coupling, and the modes in the dynamical spectral function are clearly broadened compared to the static ones due to non-adiabatic effects.

The origin of the broadening of the low-energy peak is the retardation of electronic responses as ions move, which is by definition absent in the BO calculation. This can be demonstrated by inspecting the site-projected spectral weights. The contribution from the atomic orbitals to the low-energy modes is measured by
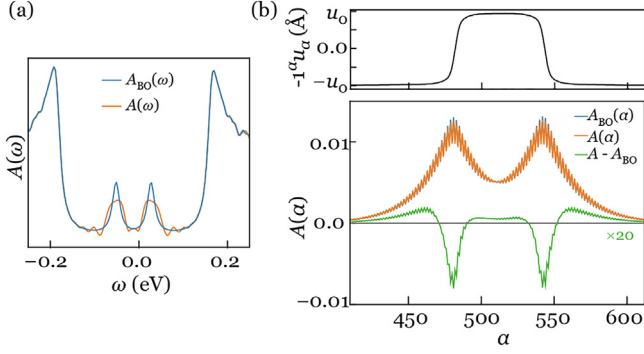
(a)

(b)

Fig. 7. (Color online) Spectral functions of the low-energy electronic modes. (a) Calculated static and dynamical spectral functions. (b) Upper: the order parameter of a carbyne chain with a soliton-antisoliton pair. Lower: static (BO) and dynamical spectral weights along the carbyne chain. Also shown (green line) is their difference.

$A(\alpha) = \int_{-0.1}^{0.1} d\omega A(\omega, \alpha)$. The lower panel of Fig. 7b shows that both the static and dynamical spectral functions peaked around the soliton and antisoliton, with significant overlap commensurate with the finite coupling. By taking the difference between the two spectral functions, we see that the dynamical spectral function is more delocalized, carrying especially a larger weight on the tails of the inward moving solitons. This in fact reflects the retardation effects that electrons do not immediately follow the moving solitons, which cannot be captured by the BO calculations. We emphasize that it is very difficult, if possible at all, to produce these results directly from *ab initio* calculations for two reasons. First, the pair annihilation process requires very long carbon chain in the simulation. Second, the non-adiabatic effects require a time-dependent electronic evolution concomitant to the MD, which is an added challenge on top of the already taxing requirement of large model system. But with TBworks, this type of calculations are performed with much reduced computational costs.

Now we move on to examine some of the equilibrium electronic correlation functions based on the TBworks, again with carbyne as the model material, using classical and path integral MD. In order to compute experimentally accessible electronic properties, retarded correlation functions of the form [37]

$$G_{AB}^{R}(t - t\prime) = -i \lambda \Theta(t) \mathrm{Tr}\{\varrho_{\mathrm{ion}}(\mathbf{r}\prime) \langle [A(t), B(t\prime)]_s \rangle\}, \tag{6}$$

usually need to be computed, where $s = \pm$ for Fermion/Boson and λ is a numerical constant. In an equilibrium ensemble, the sampling according to ionic density matrix $\varrho_{\mathrm{ion}}(\mathbf{r}(t\prime))$ is automatically offered by the trajectory from molecular dynamics. Anticipating little non-adiabatic effects, we use the approximate propagators: $U(t - t\prime) \approx e^{-iH(t\prime)(t - t\prime)}$. Then the electronic spectral function is computed with $A \rightarrow \hat{\psi}, B \rightarrow \hat{\psi}^\dagger$ and $\lambda = i/2\pi N$. To compute the electronic conductivity, $A, B \rightarrow j = -en\nu$ the current operator. Then the dissipative part of the conductivity is given by

$$\mathrm{Re}\sigma(\omega) = \mathrm{Re}\left[\frac{i}{\omega} G_{jj}^R(\omega)\right]. \tag{7}$$

For the Fourier transformation, time values $t - t\prime \in [-600, 600]$ fs are used with 0.1 fs increment. Further details of the calculation of these correlation functions can be found in the Supplementary materials Section S-2.

The computed spectral function and optical conductivity for carbyne at finite temperatures are shown in Fig. 8. We see that at $T = 1$ K, the spectral functions from both path-integral and classical simulations show a CDW gap $\Delta \sim 0.27$ eV, indicating

the equilibrium configuration of carbyne is in the polyyne structure. The optical conductivity also shows an onset at $\omega = \Delta$ at this temperature. At $T = 10$ K, the spectral function remains gapped in CMD, indicating absence of soliton. In contrast, PIMD shows a zero-energy peak in spectral function, which suggests quantum effects promote the generation of soliton-antisoliton pairs. As shown in Fig. S1 (online), the energy barrier for generation of a soliton-antisoliton pair from the polyyne structure is ~ 80 meV. Therefore, at low temperature 10 K, the soliton-antisoliton pair is hardly sampled in CMD. However, in PIMD due to quantum dynamical effects, the creation of soliton-antisoliton pair can be observed.

When the temperature is raised to moderately high value 100 K, in PIMD, the CDW gap and the zero-energy mode disappear and optical conductivity shows the single peak at $\omega = 0$, which is commensurate with the polycumulene phase. At the same temperature in CMD, the zero-energy peak in spectral function indicates presence of solitons. Above 300 K, in the CMD, the CDW melts and polycumulene is the equilibrium structure. Both spectral function and optical conductivity present the sequence of phase transitions in carbyne from polyyne to soliton and finally to polycumulene in CMD and PIMD. Due to the quantum dynamical effects, the soliton and polycumulene phase transitions happen at lower temperatures in PIMD than that in CMD, which could be detected by measuring the spectral function and/or optical conductivity experimentally. All these results exhibit the efficiency of our TBworks in calculating correlation functions that can be routinely measured in experiment to monitor the dynamics and phase transition but require large-scale and long-time simulations.

5. Summary

In this work, we have developed a neural network-based predictive representation method, which predicts tight-binding Hamiltonians for modeling electronic structures. This method provides an efficient approach to configuration-dependent Hamiltonians by directly supervised observation of the *ab initio* energy spectra. This enables an expedited access to electronic structures at dramatically reduced computational costs during the atomic configurations sampling, in subsequent MD or Monte Carlo. Using carbyne as an example, we have demonstrated that the accuracy of our method in predicting the electronic spectra is comparable to that of the *ab initio* calculations. The TBworks also shows a high degree of transferability, accurately predicting spectra for configurations drawn from unseen temperatures and system sizes.

By interfacing our TBworks with a ML force field it is demonstrated that we can efficiently sample the electronic structure of carbyne in MD to compute correlation functions. This enables the computation of Green's functions and the time-dependent spectral functions in the dynamical annihilation of a soliton-antisoliton pair. When comparing our results with those from the adiabatic approximation, it is found that the low-energy edge modes are significant renormalized due to the non-adiabatic dynamical effects. We have also applied the TBworks sampling of electronic structure in a PIMD, where the optical conductivity is computed by sampling the current–current correlation function.

As the carbyne example demonstrates, the TBworks offers an expedited access to electronic structure of a dynamical system with *ab initio* accuracy, based on an accurate and transferable representation of electronic structures from AIMD. This entails a dramatic speed-up and reduction of computational cost, when combined with ML force field MD, in computing electronic correlation functions and associated physical properties both at equilibrium and in a transient process. This will be especially useful for problems where large-scale MD and long sampling times are essential, as we have demonstrated in the non-adiabatic process
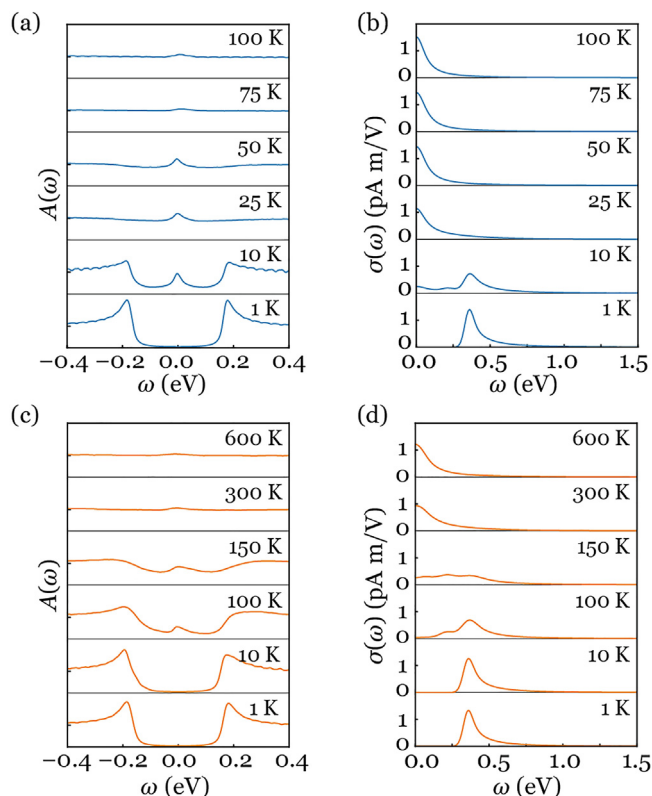
Fig. 8. (Color online) The averaged electronic spectral function $A(\omega)$ and optical conductivity $\sigma(\omega)$ for supercell $N = 1024$ at different temperatures. Spectral function (a, c) and optical conductivity (b, d) averaged in the canonical ensemble sampled by PIMD (blue lines) and CMD (orange lines).

of soliton annihilation. Evidently, this approach will significantly enhance our capability and broaden the scope in the modeling of electronic structure and processes in materials at finite temperatures or in kinetic phenomena.

Furthermore, we believe that the TBworks approach can have far-reaching potential applications in materials simulations. Because of the accurate and efficient access to electronic Hamiltonian as a function of ion configuration it offers, the approach will be especially useful to explore the properties emerging from dynamical processes and from electron–phonon coupling (e.g., phonon-assisted indirect optical absorption). Moreover, the high efficiency of TBworks in comparison with full *ab initio* electronic structure calculation will enable large-scale and long-time simulations, which can offer a direct comparison to experimentally relevant temporal and spatial scales. Clearly, further developments to deal with more complex systems are worthy goals to pursue in the future. Extension of the TBworks approach to higher dimensional materials is currently underway. It is also important to incorporate the full symmetry of the system and implement symmetry-adapted TBworks [8], which will further improve the training efficiency and model transferability.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

Author contributions

All authors made indispensable contributions to the research, wrote and critically revised the manuscript together.

Appendix A. Supplementary materials

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.scib.2021.09.010.

References

[1] Car R, Parrinello M. Unified approach for molecular dynamics and density-functional theory. Phys Rev Lett 1985;55:2471–4.
[2] Marx D, Hutter J. *Ab initio* molecular dynamics: basic theory and advanced methods. Cambridge University Press; 2009.
[3] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[4] Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett 2007;98:146401.
[5] Schütt KT, Arbabzadah F, Chmiela S, et al. Quantum-chemical insights from deep tensor neural networks. Nat Commun 2017;8:13890.
[6] Schütt KT, Sauceda HE, Kindermans P-J, et al. Schnet – a deep learning architecture for molecules and materials. J Chem Phys 2018;148:241722.
[7] Zhang L, Han J, Wang H, et al. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. Phys Rev Lett 2018;120:143001.
[8] Zhang L, Han J, Wang H, et al. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. Adv Neural Inf Process Syst 2018;32:4436–46.
[9] Ben Mahmoud C, Anelli A, Csányi G, et al. Learning the electronic density of states in condensed matter. Phys Rev B 2020;102:235130.
[10] Chandrasekaran A, Kamal D, Batra R, et al. Solving the electronic structure problem with machine learning. Npj Comput Mater 2019;5:22.
[11] Schütt KT, Gastegger M, Tkatchenko A, et al. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. Nat Commun 2019;10:5024.
[12] Brouder C, Panati G, Calandra M, et al. Exponential localization of Wannier functions in insulators. Phys Rev Lett 2007;98:046402.
[13] Cloizeaux JD. Analytical properties of n-dimensional energy bands and Wannier functions. Phys Rev 1964;135:A698–707.
[14] Nenciu G. Existence of the exponentially localised Wannier functions. Comm Math Phys 1983;91:81–5.
[15] Marzari N, Mostofi AA, Yates JR, et al. Maximally localized Wannier functions: Theory and applications. Rev Mod Phys 2012;84:1419–75.
[16] Kingma DP, Ba J. Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations (ICLR), 2015.
[17] Porezag D, Frauenheim T, Köhler T, et al. Construction of tight-binding-like potentials on the basis of density-functional theory: application to carbon. Phys Rev B 1995;51:12947–57.
[18] Elstner M, Porezag D, Jungnickel G, et al. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. Phys Rev B 1998;58:7260–8.
[19] Heimann RB, Kleiman J, Salansky NM. A unified structural approach to linear carbon polytypes. Nature 1983;306:164–7.
[20] Jin C, Lan H, Peng L, et al. Deriving carbon atomic chains from graphene. Phys Rev Lett 2009;102:205501.
[21] Cretu O, Botello-Mendez AR, Janowska I, et al. Electrical transport measured in atomic carbon chains. Nano Lett 2013;13:3487–93.
[22] Kertesz M, Koller J, Azman A. *Ab initio* Hartree-Fock crystal orbital studies. II. Energy bands of an infinite carbon chain. J Chem Phys 1978;68:2779–82.
[23] Peierls RE. Quantum theory of solids, international series of monographs on physics. Clarendon Press; 1996.
[24] Su WP, Schrieffer JR, Heeger AJ. Solitons in polyacetylene. Phys Rev Lett 1979;42:1698–701.
[25] Kresse G, Furthmüller J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. Phys Rev B 1996;54:11169–86.
[26] Perdew JP, Zunger A. Self-interaction correction to density-functional approximations for many-electron systems. Phys Rev B 1981;23:5048–79.
[27] Nosé S. A molecular dynamics method for simulations in the canonical ensemble. Mol Phys 1984;52:255–68.
[28] Hoover WG. Canonical dynamics: equilibrium phase-space distributions. Phys Rev A 1985;31:1695–7.
[29] Kohn W, Sham LJ. Self-consistent equations including exchange and correlation effects. Phys Rev 1965;140:A1133–8.

[30] Tishby N, Zaslavsky N. Deep learning and the information bottleneck principle. In: 2015 IEEE Information Theory Workshop (ITW). p. 1–5.
[31] Shwartz-Ziv R, Tishby N. Opening the black box of deep neural networks via information. Computing Research Repository (CoRR) abs/1703.00810.
[32] Kullback S, Leibler RA. On information and sufficiency. Ann Math Statist 1951;22:79–86.
[33] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 1995;117:1–19.
[34] Wang H, Zhang L, Han J, et al. A deep learning package for many-body potential energy representation and molecular dynamics. Comput Phys Commun 2018;228:178–84.
[35] Tuckerman ME, Berne BJ, Martyna GJ, et al. Efficient molecular dynamics and hybrid monte carlo algorithms for path integrals. J Chem Phys 1993;99:2796–808.
[36] Nghiem HTM, Costi TA. Time evolution of the kondo resonance in response to a quench. Phys Rev Lett 2017;119:156601.
[37] Mahan GD. Many-particle physics. Springer; 2013.
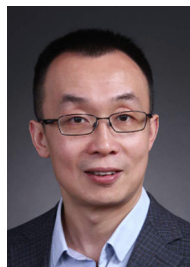
Ji Feng developed his interest in the electronic structure of materials during his training with Roald Hoffmann, Neil Ashcroft and Ju Li. As a professor of Physics in the International Center for Quantum Materials at Peking University, his research focuses on electronic structure theory and methods for computing physical properties.

Qiangqiang Gu currently holds a joint position of post-doctoral fellow at Peking University and research assistant at Artificial Intelligence for Science Institute (AISI) Beijing after graduating from Peking University in July 2021 while working at the group of Prof. Ji Feng. He has been involved in methodological developments related to models for simulation of physical properties of solids like electronic structure, molecular/lattice dynamics, electronic transport, phonon-assisted electronic responses, etc.