nature computational science

Brief Communication

https://doi.org/10.1038/s43588-024-00701-9

Generalizing deep learning electronic structure calculation to the plane-wave basis

Received: 19 March 2024

Accepted: 4 September 2024

Published online: 03 October 2024

Check for updates

Xiaoxun Gong[1,2,3], Steven G. Louie[2,3] ✉, Wenhui Duan[1,4,5] ✉ & Yong Xu[1,5,6] ✉

Deep neural networks capable of representing the density functional theory (DFT) Hamiltonian as a function of material structure hold great promise for revolutionizing future electronic structure calculations. However, a notable limitation of previous neural networks is their compatibility solely with the atomic-orbital (AO) basis, excluding the widely used plane-wave (PW) basis. Here we overcome this critical limitation by proposing an accurate and efficient real-space reconstruction method for directly computing AO Hamiltonian matrices from PW DFT results. The reconstruction method is orders of magnitude faster than traditional projection-based methods to convert PW results to the AO basis, and the reconstructed Hamiltonian matrices can faithfully reproduce the PW electronic structure, thus bridging the longstanding gap between the AO basis deep learning electronic structure approach and PW DFT. Advantages of the PW methods, such as high accuracy, high flexibility and wide applicability, thus can be all integrated into deep learning electronic structure methods without sacrificing these methods' inherent benefits. This allows for the construction of large-scale and high-fidelity training datasets with the help of PW DFT results towards the development of precise and broadly applicable deep learning electronic structure models.

Recent years have witnessed remarkable progress in the field of ab initio computation in combination with artificial intelligence[1–4]. For instance, neural-network force fields can facilitate ab initio molecular dynamics simulations at large length and time scales[5–9] and have become almost indispensable in molecular dynamics simulations nowadays. There are also numerous deep learning models for studying various material properties[10–13]. Recently, fruitful progress has been achieved in the generalization of deep learning methods from atomic structure calculations to electronic structure calculations. For instance, machine learning offers a pathway for designing accurate density functionals[14–18] as well as for predicting electronic properties, such as charge density and local density of states[19–31]. Deep learning methods have also been proposed to bypass the iterative solution of the Kohn–Sham equation of density functional theory (DFT) by directly predicting the converged DFT Hamiltonian under the atomic-orbital (AO) basis[32–43]. All these methods substantially expand the scope of theoretical and computational materials research towards unprecedented accuracy and efficiency.

When compared with other approaches, the deep learning DFT Hamiltonian method has several benefits[33,34]. First, eigenvalues and wavefunctions can be easily obtained from a one-shot diagonalization of the predicted sparse AO Hamiltonian matrix, from which all the DFT-based physical properties of materials can be derived. Furthermore, the method scales linearly with system size, and can be trained by DFT results for small-size structures and generalize to study unseen large-size structures with ab initio accuracy. The reason for these properties is that the AO Hamiltonian is a local and nearsighted

[1]State Key Laboratory of Low Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing, China. [2]Department of Physics, University of California at Berkeley, Berkeley, CA, USA. [3]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4]Institute for Advanced Study, Tsinghua University, Beijing, China. [5]Frontier Science Center for Quantum Information, Beijing, China. [6]RIKEN Center for Emergent Matter Science (CEMS), Wako, Japan. ✉e-mail: sglouie@berkeley.edu; duanw@tsinghua.edu.cn; yongxu@mail.tsinghua.edu.cn

https://doi.org/10.1038/s43588-024-00701-9

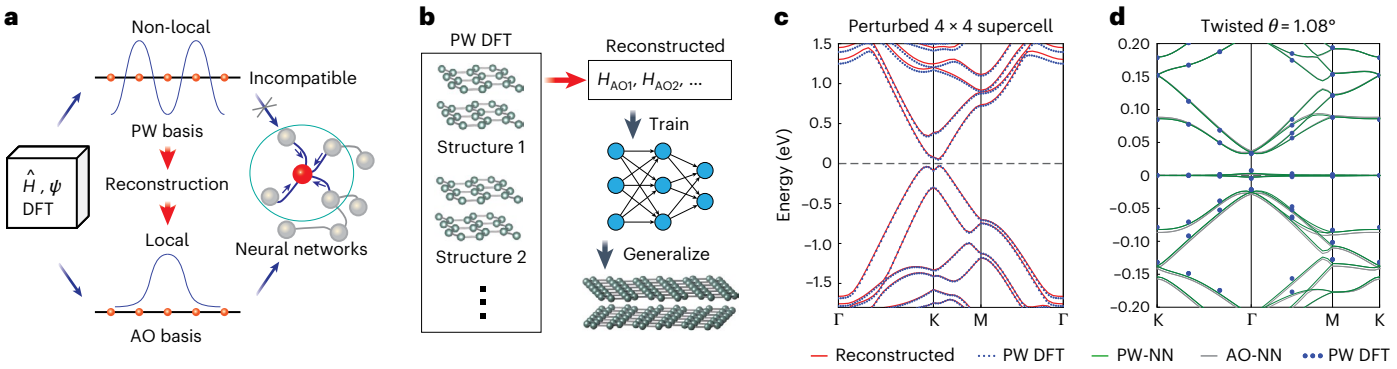

Fig. 1 | Idea of the deep learning DFT Hamiltonian under the PW basis and its applications to twisted bilayer graphene. a, The PW basis offers high accuracy and flexibility, but deep learning electronic structure calculation methods require local basis sets. With the reconstruction method, local AO Hamiltonians are reconstructed from PW DFT output, thus deep learning methods are now able to interface with PW DFT. **b**, Practical workflow of the method. PW DFT results on a set of small, non-twisted structures are used to reconstruct Hamiltonians under the AO basis. A neural network trained on these reconstructed Hamiltonians can then be generalized to predict the Hamiltonians of large, twisted structures. **c**, Band structures of a perturbed 4 × 4 graphene supercell in the training set. Those obtained from diagonalization of the reconstructed AO Hamiltonian are compared with the PW DFT results. **d**, Band structure of twisted bilayer graphene at the magic twist angle $\theta = 1.08°$ with 11,164 atoms in the Moiré supercell. The bands labeled PW-NN correspond to those predicted by the neural network trained on Hamiltonians reconstructed from PW DFT results, and those labeled AO-NN correspond to the prediction of the neural network trained on AO DFT results from ref. 34. PW DFT results are from ref. 67.

physical quantity, which can be determined only by its nearby atomic environment[44,45] (also see Supplementary Sections 1 and 2). Thus, methods of this kind can be designed to break the accuracy–efficiency dilemma in electronic structure simulations and are particularly useful for large-scale material simulations that would otherwise demand formidable computational resources. Similar to the role neural-network force fields are playing in today's molecular dynamics simulations, it is highly possible that future electronic structure simulations will also be primarily based on deep learning models of the DFT Hamiltonian.

However, the deep learning DFT Hamiltonian method faces a critical issue related to basis functions. The Hamiltonian is a quantum mechanical operator, which can be expressed as a matrix if a particular basis set is chosen. Commonly used basis sets in DFT are plane waves (PWs) and AOs. Up to now, all neural-network methods for DFT Hamiltonians only support the AO basis, because PWs are spread over the entire space (Fig. 1a) and will destroy the aforementioned locality property. Nevertheless, PW-based methods usually offer higher accuracy than those using AOs because the PW basis can usually achieve fuller completeness and is also easier to converge. It is also favorable over the AO basis in terms of its simplicity and flexibility. In fact, the majority of DFT calculations for solids are done using the PW basis. In this context, generalizing deep learning electronic structure calculation to the PW basis would be of critical importance to future development of the field.

In this work, we propose a real-space reconstruction method to reconstruct AO Hamiltonians based on PW DFT results. It is orders of magnitude faster than the traditional method of directly projecting the PW Hamiltonian or wavefunctions. Moreover, we show that the AO Hamiltonians generated using our method not only can faithfully reproduce the PW electronic structure but also can be very easily learned by neural-network models. Thus, the critical problem of the deep learning DFT Hamiltonian under the PW basis is solved (Fig. 1a). The high accuracy and efficiency of our method is beneficial for the construction of more versatile and accurate deep learning electronic structure calculation methods, which not only makes them accessible to a much broader scientific community, but also greatly enhances their suitability for general applications.

Results and discussion

Theory

The PW Hamiltonian and the AO Hamiltonian are actually the same physical quantity expressed under different basis sets. In principle, once we have the PW Hamiltonian $H_{\mathbf{k}}(\mathbf{G},\mathbf{G}') = \langle \mathbf{k}+\mathbf{G}|\hat{H}|\mathbf{k}+\mathbf{G}'\rangle$, we can always obtain the corresponding AO Hamiltonian $H_{i\alpha,j\beta} = \langle \phi_{i\alpha}|\hat{H}|\phi_{j\beta}\rangle$ by a change of basis, which can then be flexibly learned by current neural networks for AO Hamiltonians. However, there are several different ways of doing this. In Supplementary Section 3, we present a detailed review of the possible methods from literature as well as previous attempts at deep learning electronic structure calculation under the PW basis. Here we will briefly discuss three methods to convert PW Hamiltonians to the AO basis, and more details can be found in Methods.

We would naturally think of using the projection method, which is widely used to bridge the gap between PWs and AOs. The projection method was initially developed to evaluate the quality of the AO basis set[46–48] and has been adapted for various purposes, such as to analyze charge distribution and to interpret chemical bonding[49–55]. The projection method can also be modified to directly convert a Hamiltonian from the PW basis to the AO basis:

$$H_{i\alpha,j\beta} = \sum_{\mathbf{k}\mathbf{G}\mathbf{G}'} \langle \phi_{i\alpha}|\mathbf{k}+\mathbf{G}\rangle H_{\mathbf{k}}(\mathbf{G},\mathbf{G}')\langle \mathbf{k}+\mathbf{G}'|\phi_{j\beta}\rangle. \quad (1)$$

Here the PW basis is normalized in the Born–von Kármán (BvK) supercell: $\langle \mathbf{r}|\mathbf{k}+\mathbf{G}\rangle = \exp(i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r})/\sqrt{N\Omega}$, where $\mathbf{k}$ is a wavevector in the first Brillouin zone, $\mathbf{G}$ is a reciprocal lattice vector, N is the number of primitive unit cells forming the BvK supercell and Ω is the volume of the primitive unit cell. The AO basis function $|\phi_{i\alpha}\rangle$ is centered at atom i. There could be multiple basis functions (labeled by n) sharing the same angular momentum quantum number l and magnetic quantum number m. The index α is an abbreviation for n, l, m. Equation (1) will be referred to as the $H_{\mathbf{k}}(\mathbf{G}, \mathbf{G}')$ projection method in this Brief Communication.

If the eigenvalues $\varepsilon_{n\mathbf{k}}$ and wavefunctions $|\psi_{n\mathbf{k}}\rangle$ of the PW Hamiltonian are obtained, equation (1) can be further written as

$$H_{i\alpha,j\beta} = \sum_{n\mathbf{k}} \langle \phi_{i\alpha}|\psi_{n\mathbf{k}}\rangle \varepsilon_{n\mathbf{k}} \langle \psi_{n\mathbf{k}}|\phi_{j\beta}\rangle, \quad (2)$$

where

$$\langle \phi_{i\alpha}|\psi_{n\mathbf{k}}\rangle = \sum_{\mathbf{G}} \langle \phi_{i\alpha}|\mathbf{k}+\mathbf{G}\rangle\langle \mathbf{k}+\mathbf{G}|\psi_{n\mathbf{k}}\rangle. \quad (3)$$

In this Brief Communication equation (2) will be referred to as the $\psi_{n\mathbf{k}}(\mathbf{G})$ projection method.

Although equations (1) and (2) offer straightforward ways of converting PW Hamiltonians to the AO basis, they suffer from low computational efficiency, and the reasons are as follows. Equation (1) involves two multiplications and summations over **G**, and the number of **G** vectors is usually very large. Equation (2) requires a large number of Bloch wavefunctions to converge. Moreover, they all scale cubically with respect to the number of atoms in the system (Methods), which limits their scope of applications.

In fact, we can leverage locality in real space to considerably speed up the calculation. The Hamiltonian $H(\mathbf{r},\mathbf{r}') = \langle \mathbf{r}|\hat{H}|\mathbf{r}'\rangle$ in real space under atomic units is[56,57]

$$H(\mathbf{r},\mathbf{r}') = -\frac{1}{2}\nabla^2\delta(\mathbf{r}-\mathbf{r}') + [V_{\mathrm{Har}}(\mathbf{r}) + V_{\mathrm{xc}}(\mathbf{r}) + V_{\mathrm{loc}}(\mathbf{r})]\delta(\mathbf{r}-\mathbf{r}') + V_{\mathrm{nloc}}(\mathbf{r},\mathbf{r}'), \quad (4)$$

where the various terms correspond to the kinetic energy, the Hartree potential, the exchange–correlation potential and the local and non-local parts of the pseudopotential, respectively. In this Brief Communication we are only considering semilocal functionals for exchange and correlation. The three terms in the square bracket will be referred to as the total effective local potential: $V_{\mathrm{eff}}(\mathbf{r}) \equiv V_{\mathrm{Har}}(\mathbf{r}) + V_{\mathrm{xc}}(\mathbf{r}) + V_{\mathrm{loc}}(\mathbf{r})$, which is periodic over unit cells. Usually, PW DFT programs directly store $V_{\mathrm{eff}}(\mathbf{r})$ or its Fourier transform $V_{\mathrm{eff}}(\mathbf{G})$ in memory, and $V_{\mathrm{nloc}}(\mathbf{r},\mathbf{r}') = \langle \mathbf{r}|\hat{V}_{\mathrm{nloc}}|\mathbf{r}'\rangle$ can be read from pseudopotential files. Once we have $H(\mathbf{r},\mathbf{r}')$, we can calculate the AO Hamiltonian directly in real space as follows:

$$H_{i\alpha,j\beta} = \langle \phi_{i\alpha}|-\frac{1}{2}\nabla^2|\phi_{j\beta}\rangle + \int \mathrm{d}^3 r\, \phi^*_{i\alpha}(\mathbf{r})V_{\mathrm{eff}}(\mathbf{r})\phi_{j\beta}(\mathbf{r}) + \langle \phi_{i\alpha}|\hat{V}_{\mathrm{nloc}}|\phi_{j\beta}\rangle, \quad (5)$$

which will be referred to as the real-space reconstruction method in this Brief Communication. The first and last terms can be very efficiently calculated using two-center integral techniques[58] (also see Methods). The most time-consuming part is the evaluation of the second term in the equation, which is integrated directly on an evenly spaced real-space grid. Since the AOs are local in real space, the integration region can be chosen such that both of the orbitals $\phi_{i\alpha}(\mathbf{r})$ and $\phi_{j\beta}(\mathbf{r})$ are non-zero. The number of grid points involved in the integration thus does not depend on the overall system size. Therefore, the time required to evaluate the above formula is proportional to the number of non-zero AO Hamiltonian matrix elements, which scales linearly with the number of atoms in the system. Although they are theoretically equivalent and will yield the same results when converged, the proposed real-space reconstruction method is much more efficient than the first two projection-based methods.

It is worth noticing that none of the three methods described in this Brief Communication depend on the specific form of the AO basis functions. They only need to be separable into radial and angular parts, and the radial function needs to go to zero after a certain cutoff radius. This degree of freedom allows us to systematically improve the quality of the reconstructed AO Hamiltonian by customizing the AO basis using modern techniques such as a numerical AO basis[59–62]. The most important design principle of the AO basis is that it must be compatible with the pseudopotentials used in the PW calculation, otherwise it is difficult for the reconstructed Hamiltonian to give an accurate description of the band structure.

Application to twisted bilayer graphene

The real-space reconstruction method provides a very efficient way to calculate the Hamiltonian under an AO basis set from PW DFT results. The resulting AO Hamiltonian not only can accurately reproduce the PW electronic structure, but also can be learned by neural networks, thus enabling deep learning electronic structure calculations under the PW basis. The effectiveness of this workflow depends on two factors: the quality of the reconstructed AO Hamiltonian, and the compatibility of the reconstructed Hamiltonian with deep learning methods. In our tests, the first is measured by comparison of the band structures of the AO Hamiltonian with those from direct PW DFT calculations. The second can be evaluated through checking the quality of the band structure predicted by the neural networks that learn from reconstructed AO Hamiltonians. In all calculations reported in this Brief Communication, PW DFT calculations are performed with the Quantum ESPRESSO package[63] using norm-conserving pseudopotentials[64]. The results of these PW calculations are used to reconstruct AO Hamiltonians, where the AO basis is the numerical AOs generated using the SIESTA code[59] for the same set of pseudopotentials. Details of the convergence tests we have performed on the effect of the sizes of AO basis sets can be found in Supplementary Section 4.

The most remarkable capability of the deep learning DFT Hamiltonian method is that neural-network models can be trained on small structures and generalized to predict the Hamiltonians of much larger structures. In the study of bilayer graphene, the training set consists of 300 4 × 4 bilayer graphene supercells with different stackings and random perturbation of each atom site. After we train the neural-network model on the reconstructed AO Hamiltonians from PW DFT results (Fig. 1b), we can use the model to systematically study Moiré twisted superstructures with arbitrary twist angle. We first benchmark the reconstructed Hamiltonian on one of the structures in the training set by plotting its band structure alongside those calculated using PWs. As shown in Fig. 1c, the two band structures agree very well. After training the neural-network model, we use it to study the well-known 'magic-angle' twisted bilayer graphene at $\theta = 1.08°$ with 11,164 atoms in a Moiré supercell. This system is of substantial interest to researchers because of the discovery of a series of correlated phenomena[65,66], but is particularly challenging for electronic structure calculations because of its large system size and large-scale corrugation patterns. However, with the deep learning DFT Hamiltonian method available, the computational cost can be greatly reduced[34] (also see Supplementary Section 5). As illustrated in Fig. 1d, the neural network trained on the reconstructed AO Hamiltonian manages to give very accurate predictions when compared with the PW DFT benchmark[67], with an error of only a few millielectronvolts. Moreover, when the neural network is trained on reconstructed AO Hamiltonians from PW DFT output, the predicted band structure (PW-NN in Fig. 1d) has better agreement with the PW DFT results by Lucignano et al.[67] compared to the case where the neural network is trained on Hamiltonians calculated by AO DFT[34] (AO-NN in Fig. 1d). This shows that the deep learning Hamiltonian interfaced to PW DFT can indeed give results that have higher accuracy. This high accuracy, when combined with the flexibility and wide applicability of the PW method, will greatly enhance the capability of deep learning ab initio calculations and will be highly beneficial for future research.

Application to bilayer MoS2

The three previously discussed methods to obtain the AO Hamiltonian from PW DFT results are equivalent when converged, but the real-space reconstruction method (equation (5)) is the most efficient. Here we compare these three methods in the study of the bilayer MoS_2 system (Fig. 2a). First we tested the reconstruction method on the AB-stacked bilayer unit cell consisting of six atoms, and the band structures obtained from the reconstructed AO Hamiltonian agree well with PW DFT results (Fig. 2b). We then plot the band structures given by the three different methods, and the results are shown in Fig. 2c. They are almost the same, except that the band structure given by the $\psi_{n\mathbf{k}}(\mathbf{G})$ projection method is slightly different from the other two because we are only using a finite number of bands in evaluation of equation (2).

We further compared the computation times of the three methods. The systems we have studied here are bilayer MoS_2 structures with different numbers of atoms (unit cell with 6 atoms, 3 × 3 supercell with 54 atoms and Moiré twisted bilayer MoS_2 at $\theta = 13.17°$ and 9.43° with 114 and 222 atoms). The CPU times are shown in Fig. 2d along with the time

https://doi.org/10.1038/s43588-024-00701-9

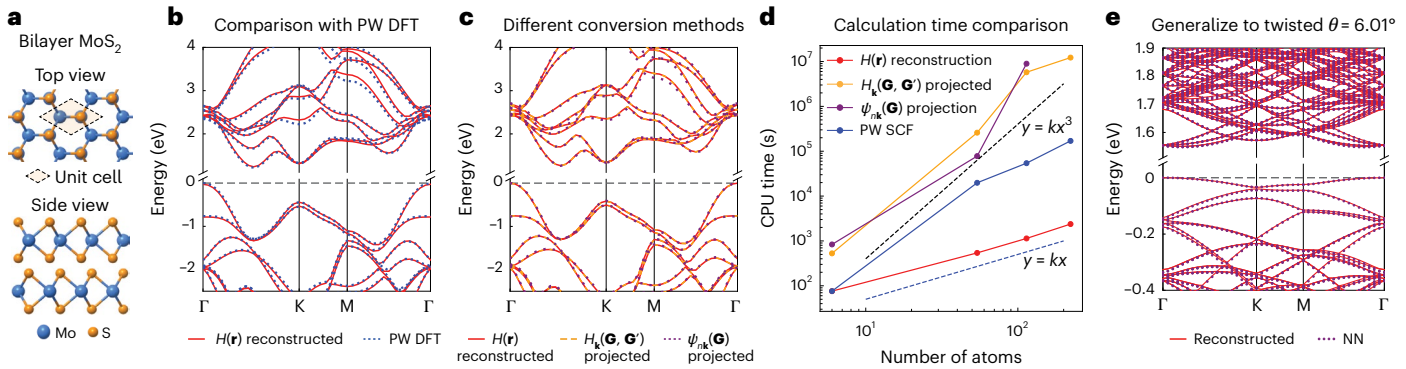

Fig. 2 | Application of the deep learning DFT Hamiltonian under the PW basis to study twisted bilayer MoS_2. a, Schematic illustration of bilayer MoS_2 with AB stacking. **b**, Band structures of the AB-stacked bilayer MoS_2 unit cell. Those obtained from diagonalization of the reconstructed AO Hamiltonian are compared with PW DFT results. **c**, The band structures obtained from the AO Hamiltonians constructed using three conversion methods are compared with each other, and the results are almost the same. **d**, Comparison of the computation times among different conversion methods when studying bilayer MoS_2 structures including varying numbers of atoms per supercell. The computation times of PW self-consistent field iterations until convergence (PW SCF) are also displayed as a reference. The $H(\mathbf{r})$ reconstruction method scales linearly with system size and is much faster than the other two methods. **e**, Band structure of the fully relaxed $\theta = 6.01°$ twisted bilayer MoS_2 with 546 atoms in the Moiré supercell. The neural-network-predicted band structure lies almost exactly on top of that diagonalized from the reconstructed Hamiltonian.

for PW self-consistent field calculation. Note that the time of diagonalization for PW wavefunctions for the $\psi_{n\mathbf{k}}(\mathbf{G})$ projection method is also included in the total CPU time for that method. As expected, the two projection-based methods show roughly cubic scaling. They are even more time consuming than the full self-consistent field calculation. Conversely, thanks to the locality of the AO basis, the real-space reconstruction method achieves linear scaling and can be several orders of magnitude faster than the projection methods. This acceleration will become more prominent when we investigate large-size materials. Therefore, our method will be essential when we want to construct large-scale training sets including various kinds of material and structures of different sizes to train accurate and versatile neural-network models, whereas the projection methods would not be affordable for this purpose in terms of computational cost.

Now, we follow the same workflow as illustrated in Fig. 1b and test the performance of the neural network on bilayer MoS_2, which is a more challenging material system than bilayer graphene. PW DFT calculations are performed on 256 non-twisted 3×3 supercells of bilayer MoS_2 with different stacking configurations and random perturbations of each atom site, and the neural-network model is trained on the reconstructed AO Hamiltonians. We test the generalizability of the neural network to a fully relaxed $\theta = 6.01°$ twisted bilayer MoS_2 with 546 atoms in the Moiré supercell. Since the neural network is trained on reconstructed AO Hamiltonians, the predicted bands are compared with the bands of the AO Hamiltonian, and results are shown in Fig. 2e. The absolute energy differences are as small as 0.30 and 2.22 meV for the highest valence band and the lowest conduction band, respectively, where errors are averaged along the high-symmetry $\mathbf{k}$ path Γ–K–M–Γ. This is remarkable considering that only small non-twisted structures are included in the training set.

Discussion

Our approach to reconstruct an AO Hamiltonian from PW DFT results facilitates deep learning electronic structure calculations based on PW DFT results and combines the advantages of the PW method and the deep learning approaches. One direct impact of our work is that it makes the deep learning electronic structure method applicable for those who are already familiar with the PW method but have less experience in AO DFT. Another promising future application of our method is to build universal deep learning models that can handle diverse families of materials and give accurate predictions of their electronic structure. The model can take advantage of the numerous materials databases that have already been set up. In fact, most of the materials databases of solids are built using the PW method, and they are thus made accessible through our reconstruction method. Moreover, the applicable scope of our method is not limited to PW DFT only. The spirit of the change of basis can also be generalized to apply to any kind of implementation of Kohn–Sham DFT and interface it with deep learning approaches. Further, the PW methods are even more widely used to implement advanced methods beyond the DFT level, such as the density functional perturbation theory to study electron–phonon interactions[68], the many-body perturbation theory (such as *GW* and *GW*-BSE methods) and time-dependent DFT for excited-state phenomena, and so on. Now, with our method to interface deep learning with PW methods, important generalizations of the deep learning approach to these advanced methods will become feasible in the foreseeable future.

Methods

Details of different ways to convert the PW Hamiltonian to the AO basis

The DFT Hamiltonian operator we are considering in this Brief Communication is given as equation (4) in the main text. Here, we will explain the non-local part of the pseudopotential $V_{\mathrm{nloc}}(\mathbf{r}, \mathbf{r}') = \langle \mathbf{r}|\hat{V}_{\mathrm{nloc}}|\mathbf{r}'\rangle$ in detail. It is constructed in a separable form known as the Kleinman and Bylander projectors[69]:

$$\hat{V}_{\mathrm{nloc}} = \sum_{i\alpha\beta} |p_{i\alpha}\rangle D_{i\alpha\beta} \langle p_{i\beta}|, \tag{6}$$

where the summation over atom sites i is carried out over all atoms in the whole BvK supercell, and this will apply to the remainder of this section unless otherwise stated. The projector function $|p_{i\alpha}\rangle$ is centered at atom i and can be separated into radial and angular parts:

$$\langle \mathbf{r}|p_{i\alpha}\rangle = R^{\mathrm{proj}}_{inl}(|\mathbf{r}_i|) Y_{lm}(\hat{\mathbf{r}}_i), \tag{7}$$

where $\mathbf{r}_i \equiv \mathbf{r} - \mathbf{R}_i$ and $\mathbf{R}_i$ is the position of the ith atom. There could be multiple projector functions (labeled by n) sharing the same l and m. The matrix $D_{i\alpha\beta}$ is non-zero only for $\alpha = \beta$ and is not system dependent (that is, unchanged in different atomic environments) if we only focus on norm-conserving pseudopotentials[70,71].

Here, we would like to discuss the locality of the Hamiltonian. The Hamiltonian in equation (4) is non-local (that is, it is non-zero when $|\mathbf{r} - \mathbf{r}'| \neq 0$) because of the presence of the non-local projectors of the pseudopotential. However, the pseudopotential projector functions

$|p_{i\alpha}\rangle$ are, by construction, highly localized around each nucleus. Therefore, we would expect that the non-local part $V_{\text{nloc}}(\mathbf{r}, \mathbf{r}')$ is non-zero only within the core region. The nearsightedness of the Hamiltonian is closely related to its locality but is a different concept, which is discussed in Supplementary Section 2.

In practical DFT calculations, two common choices of basis functions for expanding the Hamiltonians and wavefunctions are PWs and AOs. In the remainder of this section, we will first briefly review the forms of the Hamiltonian matrix under both kinds of basis set. Then, we will discuss in detail the three methods of transforming a PW Hamiltonian to the AO basis mentioned in the main text. Finally, we will discuss an extension of our method to the projector augmented-wave (PAW) formalism[72,73]. Details of the numerical techniques used to speed up the calculations are deferred to the next section.

PW Hamiltonian. The PW basis we are using here is normalized in the BvK supercell:

$$\langle \mathbf{r}|\mathbf{k}+\mathbf{G}\rangle = \frac{1}{\sqrt{N\Omega}} \exp(i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}). \tag{8}$$

Under the PW representation, the Kohn–Sham equation is written as

$$\sum_{\mathbf{G}'}(H_{\mathbf{k}}(\mathbf{G},\mathbf{G}') - \varepsilon_{n\mathbf{k}}\delta_{\mathbf{G}\mathbf{G}'})\psi_{n\mathbf{k}}(\mathbf{G}') = 0, \tag{9}$$

where $\varepsilon_{n\mathbf{k}}$ is the Kohn–Sham eigenvalue, $|\psi_{n\mathbf{k}}\rangle$ is the corresponding eigenstate, $H_{\mathbf{k}}(\mathbf{G},\mathbf{G}') = \langle \mathbf{k}+\mathbf{G}|\hat{H}|\mathbf{k}+\mathbf{G}'\rangle$ is the PW Hamiltonian matrix and $\psi_{n\mathbf{k}}(\mathbf{G}) = \langle \mathbf{k}+\mathbf{G}|\psi_{n\mathbf{k}}\rangle$ is the wavefunction.

The Hamiltonian can be written under the PW basis[56] as

$$H_{\mathbf{k}}(\mathbf{G},\mathbf{G}') = \frac{1}{2}|\mathbf{k}+\mathbf{G}|^2\delta_{\mathbf{G}\mathbf{G}'} + V_{\text{eff}}(\mathbf{G}-\mathbf{G}') + V_{\text{nloc},\mathbf{k}}(\mathbf{G},\mathbf{G}'), \tag{10}$$

where

$$V_{\text{eff}}(\mathbf{G}) = \frac{1}{\Omega}\int_{\Omega} \mathrm{d}^3 r\, V_{\text{eff}}(\mathbf{r}) \exp(-i\mathbf{G}\cdot\mathbf{r}), \tag{11}$$

and the integral is carried out within the primitive unit cell. Because $V_{\text{eff}}(\mathbf{r})$ is periodic over unit cells, it is convenient to discretize it on evenly spaced grid points in real space, and the Fourier transform can be efficiently calculated using the fast Fourier transform.

The last term in equation (10) is

$$V_{\text{nloc},\mathbf{k}}(\mathbf{G},\mathbf{G}') = \sum_{i\alpha\beta}\langle \mathbf{k}+\mathbf{G}|p_{i\alpha}\rangle D_{i\alpha\beta}\langle p_{i\beta}|\mathbf{k}+\mathbf{G}'\rangle, \tag{12}$$

where the Fourier transform $\langle \mathbf{k}+\mathbf{G}|p_{i\alpha}\rangle$ can be calculated efficiently using an algorithm described in the next subsection.

Finally, we would like to point out that the total effective local potential is the only term in equation (10) that is system dependent and needs to be obtained from self-consistent field iterations. The kinetic energy term is trivial, and the non-local pseudopotential term can be built from data read from the pseudopotential file. In practice, most of the PW DFT codes store the quantity $V_{\text{eff}}(\mathbf{G})$ or $V_{\text{eff}}(\mathbf{r})$ instead of the full Hamiltonian matrix, which substantially saves memory.

AO Hamiltonian. The AO basis functions are centered on atomic sites and are separated into radial and angular parts, similar to the projector function defined in equation (7):

$$\langle \mathbf{r}|\phi_{i\alpha}\rangle = \phi_{i\alpha}(\mathbf{r}) = R_{inl}(|\mathbf{r}_i|)Y_{lm}(\hat{\mathbf{r}}_i), \tag{13}$$

where the radial function $R_{inl}(r)$ can, in principle, take any arbitrary form and still be compatible with our reconstruction method, which will be described in a later section. It only needs to be local, which means that it goes to zero after a certain cutoff radius. This degree of freedom allows us to systematically improve the quality of the reconstructed AO Hamiltonian by customizing the AO basis using modern techniques such as a numerical AO basis[59–62]. The most important design principle of the AO basis is that it must be compatible with the pseudopotential used in the PW calculation, otherwise it is difficult for the reconstructed Hamiltonian to give an accurate description of the band structure.

The Kohn–Sham equation under the AO basis is written as

$$\sum_{j\beta}(H_{i\alpha,j\beta} - \varepsilon_{n\mathbf{k}}S_{i\alpha,j\beta})\psi_{n\mathbf{k},j\beta} = 0, \tag{14}$$

where $H_{i\alpha,j\beta} = \langle \phi_{i\alpha}|\hat{H}|\phi_{j\beta}\rangle$ is the Hamiltonian matrix, and $S_{i\alpha,j\beta} = \langle \phi_{i\alpha}|\phi_{j\beta}\rangle$ is the overlap matrix. Notice that we have to include the overlap matrix here because the AO basis functions are typically not orthonormal.

Reconstruction of AO Hamiltonian. In the main text, two projection-based methods are discussed: the $H_{\mathbf{k}}(\mathbf{G},\mathbf{G}')$ projection method (equation (1)) and the $\psi_{n\mathbf{k}}(\mathbf{G})$ projection method (equation (2)). Both methods scale cubically with system size, and here we will discuss the scaling of these two methods in detail. Equation (1) involves summations over two $\mathbf{G}$ vectors, and the summations are performed over all AO pairs ($i\alpha$,$j\beta$). The number of $\mathbf{G}$ vectors is usually very large, and is proportional to the system size. The number of orbital pairs ($i\alpha$,$j\beta$) within a certain cutoff radius is also proportional to the system size. Thus the projection has a scaling of $O(N^3)$, where N is the number of atoms in the unit cell. According to our tests, the typical calculation time for equation (1) is sometimes longer than that for a full self-consistent DFT calculation. The second method, using equation (2), involves the evaluation of equation (3), which also scales as N^3, because number of AOs, number of $\mathbf{G}$ vectors and number of wavefunctions are all proportional to system size. Because equation (3) only involves one summation over $\mathbf{G}$, this method is usually more efficient than the first one using equation (1). However, to converge the calculation of equation (2), we have to choose a relatively large n, which means we have to diagonalize the PW Hamiltonian for a large number of bands, including high-energy unoccupied bands that are typically not calculated by standard PW codes. Therefore, neither of the methods above is satisfactory in terms of efficiency.

In the main text, an efficient method is proposed to calculate the AO Hamiltonian directly in real space as equation (5). It can also be written as

$$\begin{aligned} H_{i\alpha,j\beta} = \langle \phi_{i\alpha}|-\frac{1}{2}\nabla^2|\phi_{j\beta}\rangle + \int \mathrm{d}^3 r\, \phi^*_{i\alpha}(\mathbf{r})V_{\text{eff}}(\mathbf{r})\phi_{j\beta}(\mathbf{r}) \\ + \sum_{a\gamma\delta}\langle \phi_{i\alpha}|p_{a\gamma}\rangle D_{a\gamma\delta}\langle p_{a\delta}|\phi_{j\beta}\rangle, \end{aligned} \tag{15}$$

where the terms $\langle \phi_{i\alpha}|-\frac{1}{2}\nabla^2|\phi_{j\beta}\rangle$ and $\langle \phi_{i\alpha}|p_{a\gamma}\rangle$ can be calculated very efficiently using the two-center integral technique[58], which will be described in the next subsection. The second term in the equation can be calculated directly on an evenly spaced real-space grid, and the integration region can be chosen such that both of the orbitals $\phi_{i\alpha}(\mathbf{r})$ and $\phi_{j\beta}(\mathbf{r})$ are non-zero in the integration region. The number of grid points involved in the integration thus does not depend on the overall system size. Therefore, the time required to evaluate the above formula scales linearly with the system size. Here, we would like to note that sometimes we can not directly obtain $V_{\text{eff}}(\mathbf{r})$ from a PW DFT code, but have to convert $V_{\text{eff}}(\mathbf{G})$ to real space using the inverse of equation (11). This involves the fast Fourier transform, which scales as $O(N \log N)$, but we only have to do it once, and the pre-factor is so small that the time is negligible when compared with those of other parts of the real-space reconstruction method, at least up to a few thousand atoms.

As pointed out before, $V_{\text{eff}}(\mathbf{r})$ or $V_{\text{eff}}(\mathbf{G})$ is the only term in the Hamiltonian that requires self-consistent field calculations. The quantities $|p_{a\gamma}\rangle$ and $D_{a\gamma\delta}$ can be directly read from pseudopotential files. Therefore,

for any material structure we are interested in, we only need to calculate $V_{\text{eff}}(\mathbf{r})$ with the PW DFT codes to use equation (5) to construct the AO Hamiltonian. Obtaining this quantity is also convenient because most PW DFT codes directly store $V_{\text{eff}}(\mathbf{r})$ or $V_{\text{eff}}(\mathbf{G})$ in memory.

Extension to the PAW method. Under the PAW formalism[72,73], the all-electron wavefunction $|\Psi\rangle$ is connected to the smooth pseudo-wavefunction $|\tilde{\Psi}\rangle$ by a linear transformation: $|\Psi\rangle = \mathcal{T}|\tilde{\Psi}\rangle$. Since the wavefunctions are changed by a linear transformation, any operator $\hat{A}$ in the PAW formalism is also changed according to the rule $\tilde{A} = \mathcal{T}^{\dagger}\hat{A}\mathcal{T}$. The Kohn–Sham equation still takes the same form:

$$(\tilde{H} - \varepsilon_{n\mathbf{k}}\tilde{S})|\tilde{\psi}_{n\mathbf{k}}\rangle = 0, \tag{16}$$

where the Hamiltonian operator is

$$\tilde{H}(\mathbf{r},\mathbf{r}') = \left(-\frac{1}{2}\nabla^2 + \tilde{V}_{\text{eff}}(\mathbf{r})\right)\delta(\mathbf{r}-\mathbf{r}') + \tilde{V}_{\text{nloc}}(\mathbf{r},\mathbf{r}'), \tag{17}$$

with

$$\tilde{V}_{\text{nloc}} = \sum_{i\alpha\beta}|\tilde{p}_{i\alpha}\rangle\tilde{D}_{i\alpha\beta}\langle\tilde{p}_{i\beta}|, \tag{18}$$

and the overlap operator is

$$\tilde{S} = 1 + \sum_{i\alpha\beta}|\tilde{p}_{i\alpha}\rangle\tilde{Q}_{i\alpha\beta}\langle\tilde{p}_{i\beta}|. \tag{19}$$

Comparing equations (4) and (6) with equations (17) and (18), we can see that the PAW Hamiltonian takes exactly the same form as in the case where the norm-conserving pseudopotential is used, so equation (15) can still be used. The process of evaluating the terms in equation (17) will be different for the DFT codes, but this is beyond the scope of this Brief Communication. The only thing we have to be careful of here is that the matrix $\tilde{D}_{i\alpha\beta}$ is now system dependent and needs to be obtained self-consistently. Thus, apart from obtaining $\tilde{V}_{\text{eff}}(\mathbf{r})$, we also need the matrix $\tilde{D}_{i\alpha\beta}$ from the DFT code when we convert the PW Hamiltonian to the AO basis.

Numerical techniques

Fourier transform of orbitals. Here we consider the Fourier transform of orbitals that can be separated into radial and angular parts: $\phi_{i\alpha}(\mathbf{r}) = R_{inl}(|\mathbf{r}|)Y_{lm}(\hat{\mathbf{r}})$. This will be useful when computing the Fourier transform of projector functions $\langle\mathbf{k}+\mathbf{G}|p_{i\alpha}\rangle$ or AOs $\langle\mathbf{k}+\mathbf{G}|\phi_{i\alpha}\rangle$. Using the identity

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l} i^l j_l(kr)Y_{lm}(\hat{\mathbf{k}})Y^*_{lm}(\hat{\mathbf{r}}), \tag{20}$$

where $r \equiv |\mathbf{r}|$, $k \equiv |\mathbf{k}|$, $\hat{\mathbf{r}} \equiv \mathbf{r}/r$, $\hat{\mathbf{k}} \equiv \mathbf{k}/k$ and j_l is the spherical Bessel function of order l, we can rewrite the Fourier transform as

$$\phi_{i\alpha}(\mathbf{k}) \equiv \int d^3r\, e^{-i\mathbf{k}\cdot\mathbf{r}}\phi_{i\alpha}(\mathbf{r}) = (-i)^l R_{inl}(|\mathbf{k}|)Y_{lm}(\hat{\mathbf{k}}), \tag{21}$$

where the radial part can be obtained using a spherical Bessel transformation

$$R_{inl}(k) = 4\pi\int_0^{\infty} dr\, r^2 j_l(kr)R_{inl}(r). \tag{22}$$

In practice, $R_{inl}(k)$ can be computed on a radial grid up to certain energy cutoff. Then all three-dimensional Fourier transforms of the orbital can be calculated easily and very efficiently using spline interpolation. If the orbital is not centered at the origin, we only need to add an additional phase factor.

It is also worth mentioning the inverse transformation here:

$$R_{inl}(r) = \frac{1}{2\pi^2}\int_0^{\infty} dk\, k^2 j_l(kr)R_{inl}(k). \tag{23}$$

Two-center integrals. Integrals of the product of two orbitals centered at two different positions are used frequently in AO calculations. Here we discuss a very efficient way to calculate this kind of integral, following Sankey and Niklewski[58]. Consider two orbitals $\phi_{i\alpha}(\mathbf{r})$ and $\phi_{j\beta}(\mathbf{r})$ with $\alpha = (n_1 l_1 m_1)$, $\beta = (n_2 l_2 m_2)$; their overlap integral is defined as

$$S(\mathbf{R}) \equiv \int d^3r\, \phi^*_{i\alpha}(\mathbf{r})\phi_{j\beta}(\mathbf{r}-\mathbf{R}). \tag{24}$$

This integral in real space can be converted to the integral in Fourier space:

$$S(\mathbf{R}) = \int \frac{d^3k}{(2\pi)^3}\phi^*_{i\alpha}(\mathbf{k})\phi_{j\beta}(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{R}}. \tag{25}$$

Plugging in equations (20) and (21), we have

$$S(\mathbf{R}) = \sum_{l=0}^{2l_{\max}}\sum_{m=-l}^{l} G_{l_1m_1,l_2m_2,lm}S_l(|\mathbf{R}|)Y_{lm}(\hat{\mathbf{R}}), \tag{26}$$

with $l_{\max} = \max\{l_1, l_2\}$, Gaunt coefficients $G_{l_1m_1,l_2m_2,lm}$ defined as

$$G_{l_1m_1,l_2m_2,lm} = \int_0^{\pi}\sin\theta\, d\theta\int_0^{2\pi} d\varphi\, Y^*_{l_1m_1}(\theta,\varphi)Y_{l_2m_2}(\theta,\varphi)Y^*_{lm}(\theta,\varphi), \tag{27}$$

and

$$S_l(R) = \frac{i^{l_1-l_2-l}}{2\pi^2}\int_0^{\infty} dk\, k^2 j_l(kR)R^*_{in_1l_1}(k)R_{jn_2l_2}(k). \tag{28}$$

In our calculations, $S_l(R)$ is computed on a radial grid, so that all overlap integrals $S(\mathbf{R})$ can be computed very efficiently using spline interpolation.

The above technique can be extended to calculate kinetic matrix elements

$$T(\mathbf{R}) \equiv \int d^3r\, \phi^*_{i\alpha}(\mathbf{r})(-\frac{1}{2}\nabla^2)\phi_{j\beta}(\mathbf{r}-\mathbf{R}), \tag{29}$$

with slight modifications to equation (28):

$$T_l(R) = \frac{i^{l_1-l_2-l}}{4\pi^2}\int_0^{\infty} dk\, k^4 j_l(kR)R^*_{in_1l_1}(k)R_{jn_2l_2}(k). \tag{30}$$

Preparation of datasets

Bilayer graphene dataset. The structures in the training set are the same as those used in ref. 33. There are 300 bilayer graphene 4 × 4 supercells with different interlayer stackings and random perturbations to atomic positions. The perturbations are uniformly distributed within ±0.1 Å along three Cartesian directions. The interlayer distance follows a normal distribution with mean 3.41 Å and s.d. 0.05 Å. The thickness of the unit cell along the non-periodic direction is chosen to be 20 Å. The PW DFT calculations are performed using the Perdew–Burke–Ernzerhof functional[74] with the Quantum ESPRESSO package[63] and norm-conserving Vanderbilt pseudopotential[64]. Energy cutoffs are 80 Ry for the wavefunctions and 320 Ry for the charge density. A 3 × 3 grid is used for the **k** sampling of the supercell. The double-zeta plus polarization (DZP) basis for the carbon atom with nodes is generated using SIESTA[59], which includes two orbitals in the 2*s* shell, two orbitals in the 2*p* shell and one orbital in the 3*d* shell polarized from the 2*p* orbital.

Bilayer MoS_2. PW DFT simulations are all performed with the Quantum ESPRESSO package using the Perdew–Burke–Ernzerhof functional with wavefunction cutoff 60 Ry and charge density cutoff 240 Ry. The unit cell is calculated using a 6×6 **k** sampling, and the mesh sizes are reduced for the supercells corresponding to the supercell sizes. The thickness of the unit cell along the non-periodic direction is chosen to be 20 Å. There are 256 3×3 supercells in the training set with different interlayer stackings and random perturbations to atomic positions. The perturbations are uniformly distributed within ±0.1 Å along three Cartesian directions. The average distance between the two Mo layers is 6.49 Å and the s.d. is 0.05 Å. The twisted structures are relaxed using the Perdew–Burke–Ernzerhof functional plus van der Waals interaction energy corrected using the DFT-D3 method[75]. The AOs are the standard split-norm DZP basis for Mo and S atoms generated by SIESTA. The orbitals for the Mo atom include one orbital in the 4*s* shell, one orbital in the 4*p* shell, two orbitals in the 4*d* shell, two orbitals in the 5*s* shell and one orbital in the 5*p* shell polarized from the 5*s* shell. The orbitals for the S atom include two orbitals in the 3*s* shell, two orbitals in the 3*p* shell and one orbital in the 3*d* shell polarized from the 3*p* orbital. One additional diffusion orbital in the 4*s* shell with cutoff distance 8.0 a.u. is added to the S atom to capture the interlayer hybridization.

Reconstruction from PW Hamiltonian to AO basis. The cutoffs for equation (5) of the two-center integrals are taken to be the same as the wavefunction cutoff used in the PW calculations. The sizes of the real-space grid for the real-space integrals in equation (5) are also taken to be the same as the size of the fast Fourier transform grid in the PW calculations. These are the same for all calculations in this Brief Communication.

Data availability

The data used in the current study are available at *Zenodo*[76]. PW DFT calculations in this study are performed with the Quantum ESPRESSO package (https://www.quantum-espresso.org/). AO basis functions are generated using the SIESTA package (https://siesta-project.org/siesta/index.html). Source Data for Figures 1 and 2 are available with this manuscript.

Code availability

The source code used in the current study is available at *Zenodo*[77], at *GitHub* (https://github.com/Xiaoxun-Gong/HPRO) and as Supplementary Software along with the manuscript. A demo of how to use it is also provided alongside the code.

References

1. Li, H., Xu, Y. & Duan, W. Ab initio artificial intelligence: future research of Materials Genome Initiative. *Mater. Genome Eng. Adv.* **1**, e16 (2023).
2. von Lilienfeld, O. A. & Burke, K. Retrospective on a decade of machine learning for chemical discovery. *Nat. Commun.* **11**, 4895 (2020).
3. Pederson, R., Kalita, B. & Burke, K. Machine learning and density functional theory. *Nat. Rev. Phys.* **4**, 357–358 (2022).
4. Fiedler, L., Shah, K., Bussmann, M. & Cangi, A. Deep dive into machine learning density functional theory for materials science and chemistry. *Phys. Rev. Mater.* **6**, 040301 (2022).
5. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
6. Zhang, L., Han, J., Wang, H., Car, R. & E, W. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
7. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet—a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
8. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
9. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
10. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. *Proc. Mach. Learn. Res.* **70**, 1263–1272 (2017).
11. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
12. Gasteiger, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations* Vol. 14, 10758 (Curran, 2020).
13. Eickenberg, M., Exarchakis, G., Hirn, M., Mallat, S. & Thiry, L. Solid harmonic wavelet scattering for predictions of molecule properties. *J. Chem. Phys.* **148**, 241732 (2018).
14. Lei, X. & Medford, A. J. Design and analysis of machine learning exchange–correlation functionals via rotationally invariant convolutional descriptors. *Phys. Rev. Mater.* **3**, 063801 (2019).
15. Dick, S. & Fernandez-Serra, M. Machine learning accurate exchange and correlation functionals of the electronic density. *Nat. Commun.* **11**, 3509 (2020).
16. Nagai, R., Akashi, R. & Sugino, O. Completing density functional theory by machine learning hidden messages from molecules. *npj Comput. Mater.* **6**, 43 (2020).
17. Fujinami, M., Kageyama, R., Seino, J., Ikabata, Y. & Nakai, H. Orbital-free density functional theory calculation applying semi-local machine-learned kinetic energy density functional and kinetic potential. *Chem. Phys. Lett.* **748**, 137358 (2020).
18. Bogojeski, M., Vogt-Maranto, L., Tuckerman, M. E., Müller, K.-R. & Burke, K. Quantum chemical accuracy from density functional approximations via machine learning. *Nat. Commun.* **11**, 5223 (2020).
19. Yeo, B. C., Kim, D., Kim, C. & Han, S. S. Pattern learning electronic density of states. *Sci. Rep.* **9**, 5879 (2019).
20. Ben Mahmoud, C., Anelli, A., Csányi, G. & Ceriotti, M. Learning the electronic density of states in condensed matter. *Phys. Rev. B* **102**, 235130 (2020).
21. Brockherde, F. et al. Bypassing the Kohn–Sham equations with machine learning. *Nat. Commun.* **8**, 872 (2017).
22. Tsubaki, M. & Mizoguchi, T. Quantum deep field: data-driven wave function, electron density generation, and atomization energy prediction and extrapolation with machine learning. *Phys. Rev. Lett.* **125**, 206401 (2020).
23. Schmidt, E., Fowler, A. T., Elliott, J. A. & Bristowe, P. D. Learning models for electron densities with Bayesian regression. *Comput. Mater. Sci.* **149**, 250–258 (2018).
24. Alred, J. M., Bets, K. V., Xie, Y. & Yakobson, B. I. Machine learning electron density in sulfur crosslinked carbon nanotubes. *Compos. Sci. Technol.* **166**, 3–9 (2018).
25. Grisafi, A. et al. Transferable machine-learning model of the electron density. *ACS Cent. Sci.* **5**, 57–64 (2018).
26. Fabrizio, A., Grisafi, A., Meyer, B., Ceriotti, M. & Corminboeuf, C. Electron density learning of non-covalent systems. *Chem. Sci.* **10**, 9424–9432 (2019).
27. Lewis, A. M., Grisafi, A., Ceriotti, M. & Rossi, M. Learning electron densities in the condensed phase. *J. Chem. Theory Comput.* **17**, 7203–7214 (2021).
28. Chandrasekaran, A. et al. Solving the electronic structure problem with machine learning. *npj Comput. Mater.* **5**, 22 (2019).
29. Ellis, J. A. et al. Accelerating finite-temperature Kohn–Sham density functional theory with deep neural networks. *Phys. Rev. B* **104**, 035120 (2021).

30. Dick, S. & Fernandez-Serra, M. Learning from the density to correct total energy and forces in first principle simulations. *J. Chem. Phys.* **151**, 144102 (2019).
31. Fiedler, L. et al. Predicting electronic structures at any length scale with machine learning. *npj Comput. Mater.* **9**, 115 (2023).
32. Schütt, K. T., Gastegger, M., Tkatchenko, A., Müller, K.-R. & Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **10**, 5024 (2019).
33. Li, H. et al. Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.* **2**, 367–377 (2022).
34. Gong, X. et al. General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian. *Nat. Commun.* **14**, 2848 (2023).
35. Li, H. et al. Deep-learning electronic-structure calculation of magnetic superstructures. *Nat. Comput. Sci.* **3**, 321–327 (2023).
36. Tang, Z. et al. Efficient hybrid density functional calculation by deep learning. Preprint at https://arxiv.org/abs/2302.08221 (2023).
37. Su, M., Yang, J.-H., Xiang, H.-J. & Gong, X.-G. Efficient determination of the Hamiltonian and electronic properties using graph neural network with complete local coordinates. *Mach. Learn. Sci. Technol.* **4**, 035010 (2023).
38. Zhong, Y., Yu, H., Su, M., Gong, X. & Xiang, H. Transferable equivariant graph neural networks for the Hamiltonians of molecules and solids. *npj Comput. Mater.* **9**, 182 (2023).
39. Yu, H., Xu, Z., Qian, X., Qian, X. & Ji, S. Efficient and equivariant graph networks for predicting quantum Hamiltonian. *Proc. Mach. Learn. Res.* **202**, 40412–40424 (2023).
40. Wang, Z. et al. Machine learning method for tight-binding Hamiltonian parameterization from ab-initio band structure. *npj Comput. Mater.* **7**, 11 (2021).
41. Gu, Q. et al. Deep learning tight-binding approach for large-scale electronic simulations at finite temperatures with ab initio accuracy. *Nat. Commun.* **15**, 6772 (2024).
42. Kim, R. & Son, Y.-W. Transferable empirical pseudopotenials from machine learning. *Phys. Rev. B* **109**, 045153 (2024).
43. Wang, Y. et al. Universal materials model of deep-learning density functional theory Hamiltonian. *Sci. Bull.* **69**, 2514–2521 (2024).
44. Kohn, W. Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.* **76**, 3168 (1996).
45. Prodan, E. & Kohn, W. Nearsightedness of electronic matter. *Proc. Natl Acad. Sci. USA* **102**, 11635–11638 (2005).
46. Chadi, D. J. Localized-orbital description of wave functions and energy bands in semiconductors. *Phys. Rev. B* **16**, 3572 (1977).
47. Sanchez-Portal, D., Artacho, E. & Soler, J. M. Projection of plane-wave calculations into atomic orbitals. *Solid State Commun.* **95**, 685–690 (1995).
48. Sánchez-Portal, D., Artacho, E. & Soler, J. M. Analysis of atomic orbital basis sets from the projection of plane-wave results. *J. Phys. Condens. Matter* **8**, 3859 (1996).
49. Segall, M. D., Pickard, C. J., Ahah, R. & Payne, M. C. Population analysis in plane wave electronic structure calculations. *Mol. Phys.* **89**, 571–577 (1996).
50. Dunnington, B. D. & Schmidt, J. R. Generalization of natural bond orbital analysis to periodic systems: applications to solids and surfaces via plane-wave density functional theory. *J. Chem. Theory Comput.* **8**, 1902–1911 (2012).
51. Dunnington, B. D. & Schmidt, J. R. A projection-free method for representing plane-wave DFT results in an atom-centered basis. *J. Chem. Phys.* **143**, 104109 (2015).
52. Maintz, S., Deringer, V. L., Tchougréeff, A. L. & Dronskowski, R. LOBSTER: a tool to extract chemical bonding from plane-wave based DFT. *J. Comput. Chem.* **37**, 1030–1035 (2016).
53. Nelson, R. et al. LOBSTER: local orbital projections, atomic charges, and chemical-bonding analysis from projector-augmented-wave-based density-functional theory. *J. Comput. Chem.* **41**, 1931–1940 (2020).
54. Aarons, J., Verga, L. G., Hine, N. D. M. & Skylaris, C.-K. Atom-projected and angular momentum resolved density of states in the ONETEP code. *Electron. Struct.* **1**, 035002 (2019).
55. Kundu, S., Bhattacharjee, S., Lee, S.-C. & Jain, M. Population analysis with Wannier orbitals. *J. Chem. Phys.* **154**, 104111 (2021).
56. Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge Univ. Press, 2004).
57. Sholl, D. S. & Steckel, J. A. *Density Functional Theory: a Practical Introduction* (Wiley, 2009).
58. Sankey, O. F. & Niklewski, D. J. Ab initio multicenter tight-binding model for molecular-dynamics simulations and other applications in covalent systems. *Phys. Rev. B* **40**, 3979 (1989).
59. Soler, J. M. et al. The SIESTA method for ab initio order-*N* materials simulation. *J. Phys. Condens. Matter* **14**, 2745 (2002).
60. Ozaki, T. Variationally optimized atomic orbitals for large-scale electronic structures. *Phys. Rev. B* **67**, 155108 (2003).
61. Larsen, A. H., Vanin, M., Mortensen, J. J., Thygesen, K. S. & Jacobsen, K. W. Localized atomic basis set in the projector augmented wave method. *Phys. Rev. B* **80**, 195112 (2009).
62. Bowler, D. R. et al. Highly accurate local basis sets for large-scale DFT calculations in CONQUEST. *Jpn. J. Appl. Phys.* **58**, 100503 (2019).
63. Giannozzi, P. et al. Quantum ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).
64. van Setten, M. et al. The PseudoDojo: training and grading a 85 element optimized norm-conserving pseudopotential table. *Comput. Phys. Commun.* **226**, 39–54 (2018).
65. Cao, Y. et al. Correlated insulator behaviour at half-filling in magic-angle graphene superlattices. *Nature* **556**, 80–84 (2018).
66. Cao, Y. et al. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018).
67. Lucignano, P., Alfè, D., Cataudella, V., Ninno, D. & Cantele, G. Crucial role of atomic corrugation on the flat bands and energy gaps of twisted bilayer graphene at the magic angle $\theta \sim 1.08°$. *Phys. Rev. B* **99**, 195419 (2019).
68. Li, H. et al. Deep-learning density functional perturbation theory. *Phys. Rev. Lett.* **132**, 096401 (2024).
69. Kleinman, L. & Bylander, D. M. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.* **48**, 1425 (1982).
70. Hamann, D. R., Schlüter, M. & Chiang, C. Norm-conserving pseudopotentials. *Phys. Rev. Lett.* **43**, 1494 (1979).
71. Hamann, D. R. Optimized norm-conserving Vanderbilt pseudopotentials. *Phys. Rev. B* **88**, 085117 (2013).
72. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
73. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).
74. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **78**, 1396 (1997).
75. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).
76. Gong, X., Louie, S. G., Duan, W. & Xu, Y. Dataset for 'Generalizing deep-learning electronic structure calculation to plane-wave basis'. *Zenodo* https://doi.org/10.5281/zenodo.13377497 (2024).
77. Gong, X., Louie, S. G., Duan, W. & Xu, Y. Code for 'Generalizing deep-learning electronic structure calculation to plane-wave basis'. *Zenodo* https://doi.org/10.5281/zenodo.13377785 (2024).

Acknowledgements
This work was supported by the Basic Science Center Project of NSFC (grant 52388201), the National Science Foundation of the United States (grant DMR-2325410), the Ministry of Science and Technology of China (grant 2023YFA1406400), the National Natural Science Foundation of China (grants 12334003, 12421004 and 12361141826), the National Science Fund for Distinguished Young Scholars (grant 12025405), the Beijing Advanced Innovation Center for Future Chip Technology (ICFC) and the Beijing Advanced Innovation Center for Materials Genome Engineering. The calculations were done at National Supercomputer Center in Tianjin using the Tianhe new generation supercomputer.
We thank W. Kim and Z. Tang for discussions.

Author contributions
Y.X., W.D. and S.G.L. proposed the project and supervised X.G. in carrying out the research. All authors discussed the results. Y.X. and X.G. prepared the manuscript with input from the other co-authors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-024-00701-9.

Correspondence and requests for materials should be addressed to Steven G. Louie, Wenhui Duan or Yong Xu.

Peer review information *Nature Computational Science* thanks Attila Cangi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Jie Pan, in collaboration with the *Nature Computational Science* team. Peer reviewer reports are available.