


# A deep equivariant neural network approach for efficient hybrid density functional calculations

Received: 4 March 2024

Accepted: 24 September 2024

Published online: 11 October 2024

 Check for updates

Zechen Tang<sup>1,11</sup>, He Li<sup>1,2,11</sup>, Peize Lin<sup>3,4,5,11</sup>, Xiaoxun Gong<sup>1,6</sup>, Gan Jin<sup>7</sup>, Lixin He<sup>5,7</sup>, Hong Jiang<sup>8</sup>, Xinguo Ren<sup>3,4</sup>✉, Wenhui Duan<sup>1,2,9</sup>✉ & Yong Xu<sup>1,9,10</sup>✉

Hybrid density functional calculations are essential for accurate description of electronic structure, yet their widespread use is restricted by the substantial computational cost. Here we develop DeepH-hybrid, a deep equivariant neural network method for learning the hybrid-functional Hamiltonian as a function of material structure, which circumvents the time-consuming self-consistent field iterations and enables the study of large-scale materials with hybrid-functional accuracy. Our extensive experiments demonstrate good reliability as well as effective transferability and efficiency of the method. As a notable application, DeepH-hybrid is applied to study large-supercell Moiré-twisted materials, offering the first case study on how the inclusion of exact exchange affects flat bands in magic-angle twisted bilayer graphene. The work generalizes deep-learning electronic structure methods to beyond conventional density functional theory, facilitating the development of deep-learning-based ab initio methods.

A milestone development of density functional theory (DFT) is the invention of hybrid functionals<sup>1</sup>, developed first as an ad hoc correction to local density or generalized gradient approximations (LDA/GGA)<sup>2</sup>, and later formulated more rigorously in the generalized Kohn–Sham framework<sup>3</sup>. Superior to conventional density functionals, hybrid functionals provide a viable route to solve the critical “band-gap problem” of DFT<sup>4,5</sup>, thus indispensable for reliable material prediction and particularly useful for computational studies in (opto-)electronics, spintronics, topological electronics, etc. The practical use of hybrid functionals, however, is limited for large-scale materials simulations, because their computational cost is considerably higher than local and semi-local DFT methods. Great efforts have been devoted to improving the numerical

algorithms<sup>6–14</sup>. This helps reduce the computational overhead and facilitates linear-scaling hybrid-functional calculations, but cannot fundamentally change the landscape of ab initio computation.

Deep learning methods shed light on revolutionizing ab initio materials simulation<sup>15–40</sup>. For instance, the use of artificial neural networks to represent DFT Hamiltonian enables efficient electronic-structure calculations with ab initio accuracy, whose computational cost is as low as that of empirical tight-binding calculations<sup>26–30</sup>. The so-called deep-learning DFT Hamiltonian (DeepH) approach has been demonstrated powerful in large-scale materials simulation for both non-magnetic and magnetic systems<sup>26–30</sup>. However, the method was originally designed within the Kohn–Sham (KS) framework. Therein

<sup>1</sup>State Key Laboratory of Low Dimensional Quantum Physics and Department of Physics, Tsinghua University, 100084 Beijing, China. <sup>2</sup>Institute for Advanced Study, Tsinghua University, 100084 Beijing, China. <sup>3</sup>Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, 100190 Beijing, China. <sup>4</sup>Songshan Lake Materials Laboratory, 523808 Dongguan, Guangdong, China. <sup>5</sup>Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, 230026 Hefei, Anhui, China. <sup>6</sup>School of Physics, Peking University, 100871 Beijing, China. <sup>7</sup>Key Laboratory of Quantum Information, University of Science and Technology of China, 230026 Hefei, Anhui, China. <sup>8</sup>College of Chemistry and Molecular Engineering, Peking University, 100871 Beijing, China. <sup>9</sup>Frontier Science Center for Quantum Information, Beijing, China. <sup>10</sup>RIKEN Center for Emergent Matter Science (CEMS), Wako, Saitama 351-0198, Japan. <sup>11</sup>These authors contributed equally: Zechen Tang, He Li, Peize Lin. ✉e-mail: [renxg@iphy.ac.cn](mailto:renxg@iphy.ac.cn); [duanw@tsinghua.edu.cn](mailto:duanw@tsinghua.edu.cn); [yongxu@mail.tsinghua.edu.cn](mailto:yongxu@mail.tsinghua.edu.cn)

the deep-learning problem is simplified by the local nature of the exchange-correlation potential. In contrast, hybrid functionals are usually done within the generalized Kohn–Sham (gKS) scheme<sup>3</sup>, giving rise to non-local exchange potentials. Considering that the deep-learning approach relies critically on the locality property<sup>41–43</sup>, whether the same strategy is applicable to the gKS scheme or not is an important open question.

In this work, we find that the gKS-DFT Hamiltonian of hybrid functionals  $H_{\text{DFT}}^{\text{hyb}}$  can be represented by neural networks similar to conventional DFT, benefiting from the preservation of the nearsightedness principle on a localized basis. We apply deep E(3)-equivariant neural networks to model  $H_{\text{DFT}}^{\text{hyb}}$  as a function of material structure (Fig. 1a), from which the electronic structure and physical properties of materials can be predicted without invoking ab initio codes. The method is tested to show good performance by systematic numerical experiments and further applied to study Moiré-twisted superstructures, such as magic-angle twisted bilayer graphene, demonstrating the capability for large-scale electronic-structure calculations with hybrid-functional accuracy. Our work paves the way for accurate, efficient materials simulation, and also opens a door for developing deep-learning electronic structure methods beyond DFT.

## Results

### Consideration of nearsightedness principle

In the KS-DFT<sup>44</sup>, the challenging interacting-electron problem is mapped to an auxiliary non-interacting problem whereby the complicated many-body effects are incorporated in an exchange-correlation function. Within the conventional approximations of KS-DFT, the exchange-correlation energy is expressed as an *explicit functional of density* and a *local* form of exchange-correlation potential  $V_{\text{xc}}(\mathbf{r})$  is assumed. Such approximations greatly simplify the problem and are widely used in ab initio calculations. Unfortunately, the delocalization error<sup>45</sup> is prevalent in such density-based functionals, which could result in systematic failures of DFT, including the band gap problem<sup>4,46</sup>. In fact, the fundamental band gap will be underestimated within the KS-DFT framework if the functional derivative discontinuity is not taken into account<sup>47–49</sup>. This critical issue needs to be addressed to make reliable property predictions on electronic materials.

The gKS scheme allows the use of orbital-dependent exchange-correlation potential, which helps relieve the band-gap problem. As a typical example, hybrid-functional methods replace a portion of semi-local exchange with the (screened) Hartree-Fock exact exchange, by

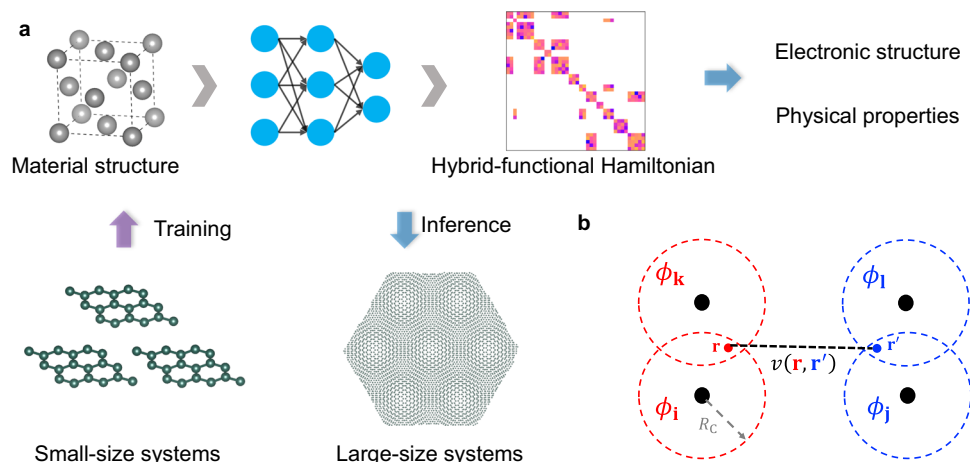
which the band gap problem could be largely resolved. However, a non-local, exact-exchange potential  $V_{\text{Ex}}(\mathbf{r}, \mathbf{r}')$  will be introduced into the effective Hamiltonian, which significantly complicates the calculation. Let us illustrate this with the localized orbital basis functions  $\phi_i(\mathbf{r}) = R_{ipl}(r)Y_{lm}(\hat{r})$ , where  $R_{ipl}$  is the radial function centered at the  $i$ th atom labeled by the multiplicity  $p$  and angular momentum quantum number  $l$ ,  $Y_{lm}$  is the real spherical harmonics of degree  $l$  and order  $m$ , and  $\mathbf{i} \equiv (iplm)$  is used for simplicity. In the localized basis, the  $n$ th KS eigenstate is  $\psi_n(\mathbf{r}) = \sum_i c_{ni} \phi_i(\mathbf{r})$ , and the (screened) exact-exchange potential is written as

$$V_{ij}^{\text{Ex}} = - \sum_n^{\text{occ}} \sum_{\mathbf{k}, \mathbf{l}} c_{nk} c_{nl}^* (\mathbf{i} \mathbf{k} | \mathbf{l} \mathbf{j}), \quad (1)$$

$$(\mathbf{i} \mathbf{k} | \mathbf{l} \mathbf{j}) = \iint d\mathbf{r} d\mathbf{r}' \phi_i^*(\mathbf{r}) \phi_k(\mathbf{r}) v(\mathbf{r} - \mathbf{r}') \phi_l^*(\mathbf{r}') \phi_j(\mathbf{r}'),$$

where  $v(\mathbf{r} - \mathbf{r}')$  denotes the Coulomb potential  $1/|\mathbf{r} - \mathbf{r}'|$  or its screened version. Note that the two-electron Coulomb repulsion integral  $(\mathbf{i} \mathbf{k} | \mathbf{l} \mathbf{j})$  involves a four-center integration over six spatial coordinates (Fig. 1b), and the number of integrals to be calculated is enormous, growing quickly with the system size. Hence the computation becomes much more expensive than local or semi-local DFT. The situation is alleviated with significant algorithm improvements (e.g., resolution of identity and linear-scaling techniques)<sup>13,14,50</sup>, but the significant increase of the computational cost from semi-local to hybrid DFT methods is not fundamentally changed. This is the major drawback of hybrid functionals, which restricts broad applications of the methods.

The exchange-correlation potentials of hybrid functionals share the form  $V_{\text{xc}}^{\text{hyb}}(\mathbf{r}, \mathbf{r}') = V'_{\text{xc}}(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}') + \alpha V_{\text{Ex}}(\mathbf{r}, \mathbf{r}')$ , where a fraction  $\alpha$  of the semi-local exchange potential is replaced by  $V_{\text{Ex}}$  and the remaining part  $V'_{\text{xc}}$  is the same as that of the semi-local DFT. For example, the Heyd–Scuseria–Ernzerhof (HSE) hybrid functional uses an error-function-screened Coulomb potential in  $V_{\text{Ex}}$  with  $\alpha = 25\%$ <sup>51</sup>. According to the Hohenberg–Kohn theorem<sup>52</sup>, the auxiliary non-interacting Hamiltonian is uniquely determined by the external potential  $V_{\text{ext}}$  that is defined by the material structure  $\{\mathcal{R}\}$ . Thus  $H_{\text{DFT}}^{\text{hyb}}$  and  $V_{\text{Ex}}$  can be expressed as a function of  $\{\mathcal{R}\}$ . It has been established that the KS-DFT Hamiltonian  $H_{\text{DFT}}^{\text{KS}}(\{\mathcal{R}\})$  can be well represented by deep neural networks<sup>26–28</sup>. We will attempt to use neural networks to model  $H_{\text{DFT}}^{\text{hyb}}(\{\mathcal{R}\})$  to generalize the deep-learning approach to achieve hybrid-functional accuracy (Fig. 1a). Compared with the KS case, a special non-



**Fig. 1 | Schematic workflow and illustration of non-local nature in hybrid functionals. a** Schematic workflow of the DeepH-hybrid method. Neural networks are used to learn the relationship between hybrid-functional Hamiltonian and material structure. The method applies ab initio calculations of small-size systems for training and makes inferences on large-size systems for predicting electronic

structure and physical properties. **b** Non-local nature of the exact-exchange potential. Two non-overlapping atomic orbitals  $\phi_i$  and  $\phi_j$  with a finite cutoff radius  $R_C$  can contribute non-zero hybrid-functional Hamiltonian matrix elements due to the presence of non-local potential  $v(\mathbf{r}, \mathbf{r}')$ . Other atomic orbitals  $\phi_k$  and  $\phi_l$  are involved in the non-local coupling.

local component  $V_{\text{Ex}}(\{\mathcal{R}\})$  is introduced here, whose neural-network representation has not been considered before.

Satisfying the nearsightedness principle is essential to simplify the deep-learning Hamiltonian problem, as learned from the study of KS-DFT. Multiple existing studies have investigated the nearsightedness of various quantities, such as total energy<sup>53,54</sup>, yet discussions regarding the nearsightedness of gKS-DFT Hamiltonians remain exclusive to date. On a localized basis, the KS-DFT Hamiltonian can be viewed as an *ab initio* tight-binding Hamiltonian. The hopping between atoms  $i$  and  $j$ , namely the Hamiltonian matrix block  $H_{ij}$ , is nonzero only when the atomic distance  $r_{ij}$  is smaller than a cutoff radius  $R_C$ . Moreover,  $H_{ij}$  is predominately determined by the neighboring environment, whose value is insensitive to distant variations of atomic structure. Thus  $H_{ij}$  can be simplified to be a function of  $\{\mathcal{R}\}_N$ , which includes structural information of neighboring atoms  $\{k\}$  with  $r_{ik}, r_{jk} < R_N$ , where  $R_N$  denotes a nearsightedness length<sup>26</sup>.

For hybrid functionals, whether the non-local exact exchange is compatible with the nearsightedness principle or not should be checked. In Eq. (1), the KS eigenvector  $c_{nk}$  or  $c_{nl}$  can be influenced by distant changes in boundary conditions, which breaks the nearsightedness principle. The two-electron Coulomb repulsion integral  $(\mathbf{ik}|\mathbf{lj})$  displays long-distance (short-distance) decay between atoms  $i$  and  $j$  when the bare (screened) Coulomb potential is considered. Thus the product  $c_{nk}c_{nl}(\mathbf{ik}|\mathbf{lj})$  is a non-local quantity, whose dependence on material structure is expected to be complicated. However, the summation over occupied states  $\sum_n^{\text{occ}} c_{nk}c_{nl}^*$  yields the density matrix element  $\rho_{k,l}$  that is a local quantity<sup>55,56</sup>. Moreover, the Coulomb integral  $(\mathbf{ik}|\mathbf{lj})$  is nonzero only when both atom pairs  $i-k$  and  $l-j$  have finite orbital overlaps (Fig. 1b). Hence the local property gets preserved for  $V_{ij}^{\text{Ex}}$ . This is the reminiscent of W. Kohn's principle, which states that the nearsightedness of electronic matter is originated from wave-mechanical destructive interference in many-particle systems<sup>55,56</sup>. Benefiting from the nearsightedness property,  $V_{ij}^{\text{Ex}}$  can be determined by local structural information of neighborhood, similar to local exchange-correlation potentials. *The merit enables us to treat the conventional and generalized KS-DFT within a unified deep-learning framework.* The inclusion of exact exchange, however, could significantly weaken the sparseness and nearsightedness properties of the DFT Hamiltonian. This will be accounted for by changing the important length scales  $R_C$  and  $R_N$  in the design of deep neural networks.

The DFT Hamiltonian matrix of local or semi-local functionals is sparse, whose elements are nonzero only when the distance between atom pairs  $i$  and  $j$  is smaller than a cutoff length  $R_C = R_i + R_j$ , where  $R_i$  and  $R_j$  denote the basis cutoff radii of atoms  $i$  and  $j$ , respectively. In contrast, the Hamiltonian matrix of hybrid functionals is denser. To describe this feature, we set a larger cutoff length  $R_C^{\text{hyb}} = \gamma R_C$ , with  $\gamma$  being an adjustable parameter. Choosing a larger  $\gamma$  generally results in a more accurate DeepH-hybrid model capable of capturing more non-local features, at the expense of increased time and memory costs. Numerical tests on this parameter are presented in Supplementary Section 5, indicating that a  $\gamma$  value of 2.0 is appropriate. This value is used for studies presented in the main text. This selection is a compromise between accuracy and efficiency, which reflects the non-local nature of hybrid functionals.

### Equivariance neural networks

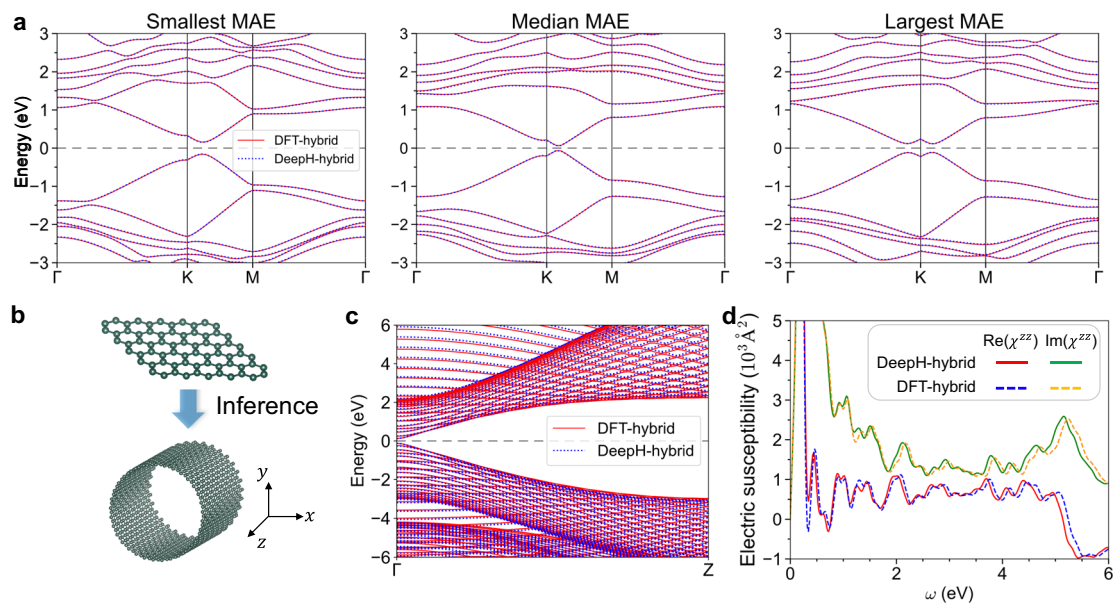
In our work, we use the method of E(3)-equivariant deep learning DFT Hamiltonian (DeepH-E3)<sup>28</sup> to model the mapping from the material structure  $\{\mathcal{R}\}$  to the corresponding hybrid-functional DFT Hamiltonian  $H_{\text{DFT}}^{\text{hyb}}$  under numerical atomic orbital (NAO) basis. A graph is associated with each material structure, with each vertex representing an atom, and edges are connected between atoms within a certain cutoff. The feature vectors associated with vertices and edges are iteratively updated with neural networks, and the final edge features are used to construct the output hopping matrices. Updating a vertex or edge only uses information within its neighborhood, and so the nearsightedness

property is utilized for the prediction of Hamiltonians. Moreover, since the Hamiltonian transforms covariantly between coordinate frames, it is most natural and advantageous to construct neural networks that explicitly handle the covariant property of the Hamiltonian. To achieve this, all the input, output and internal vectors of the neural networks transfer according to irreducible representations of the O(3) group under coordinate rotations and inversion. The incorporation of the requirements of locality and symmetry as *a priori* knowledge has greatly enhanced the performance of DeepH-E3 and has led to its sub-meV level accuracy and excellent generalization ability<sup>28</sup>.

Utilizing the equivariant neural network (ENN)<sup>57,58</sup>, the mapping  $\{\mathcal{R}\} \rightarrow \hat{H}_{\text{DFT}}$  in the DeepH-E3 approach is equivariant with respect to the Euclidean group in three-dimensional space, the E(3) group, which is composed of translations, rotations and spatial inversion in three-dimensional space, thus preserving the fundamental symmetries. To realize equivariance in neural networks, DeepH-E3 labels each network feature with angular quantum number  $l$ . Upon spatial rotation  $\mathbf{R}$  to the input structure, all features will transform as  $\mathbf{x}_m^l \rightarrow \sum_{m'} D_{mm'}^l(\mathbf{R}) \mathbf{x}_{m'}^l$ , in which  $D_{mm'}^l(\mathbf{R})$  is the Wigner-D matrix. DFT Hamiltonians block  $H_{ij}$  can be divided into sub-blocks  $\mathbf{h} \equiv [H_{ij}]^{p_1 p_2}$  by grouping orbitals with the same  $p$  together. The resulting sub-blocks are equivariant tensor whose elements transform as  $\mathbf{h}_{m_1 m_2}^{l_1 l_2} \rightarrow \sum_{m'_1 m'_2} D_{m_1 m'_1}^{l_1}(\mathbf{R}) D_{m_2 m'_2}^{l_2}(\mathbf{R}) \mathbf{h}_{m'_1 m'_2}^{l_1 l_2}$  upon rotation  $\mathbf{R}$ . Equivariant tensors and equivariant vectors can be associated by Wigner-Eckart theorem  $l_1 \otimes l_2 = |l_1 - l_2| \oplus \dots \oplus (l_1 + l_2)$ . Hamiltonian sub-blocks are regarded as equivariant tensors with representation  $l_1 \otimes l_2$  and constructed accordingly. DeepH-E3 starts with  $l=0$  (scalar) features by embedding atomic numbers ( $Z_i$ ) and distance between atom pairs ( $|r_{ij}|$ ). Relative direction of atom pairs are also taken as input features with  $l=1, 2, \dots$  by enforcing spherical harmonics on  $\hat{\mathbf{r}}_{ij}$ . Regarding equivariance with respect to spatial inversion, feature vectors are additionally labeled by their parity upon spatial inversion, either even (e) or odd (o). All intermediate ENN operations are designed to preserve features' parity characteristics. In all training, our neural networks are composed of three message-passing blocks, with  $64 \times 0e + 32 \times 1o + 16 \times 2e + 8 \times 3o + 8 \times 4e$  equivariant layer for each intermediate layer. Here,  $64 \times 0e$  stands for 64 even-parity equivariant vectors with  $l=0$ ,  $32 \times 1o$  stands for 32 odd-parity equivariant vectors with  $l=1$ . Atomic configuration information is embedded into 64-dimensional equivariant vectors as initial vertex and edge features. Mean squared error of Hamiltonian matrix elements is selected as the loss function for training neural network models. Datasets are randomly split into training, validation, and test sets with a ratio of 6:2:2.

### Case studies

To demonstrate the capability of DeepH-hybrid, we carry out example studies on various material systems, including monolayers and bilayers of graphene and MoS<sub>2</sub>. Figure 2 shows the performance of DeepH-hybrid on studying monolayer graphene and related systems. Neural network models are trained by a dataset containing randomly perturbed structures of graphene supercells, which will be generalized to investigate new structures of graphene supercells as well as carbon nanotubes (CNTs). The mean absolute errors (MAEs) of gKS-DFT Hamiltonian matrix elements are 0.207, 0.208, and 0.208 meV for training, validation, and test sets, respectively. The MAEs are even smaller than that for the DeepH study using the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional (MAE = 0.40 meV)<sup>28</sup>. The test set consisting of 100 perturbed graphene supercells is sorted in terms of MAE. The band structures corresponding to the best, median and worst MAEs are shown in Fig. 2a. All of them agree well with the benchmark calculations, demonstrating the good accuracy of DeepH-hybrid. In addition, we use the trained neural network model to study CNTs that have a curved geometry unseen in the training set for testing the generalization ability of the method (Fig. 2b). As shown in Fig. 2c for (49, 0) CNT, a good agreement



**Fig. 2 | Example studies on monolayer graphene and carbon nanotube (CNT).**

**a** Band structures of supercells of monolayer graphene computed by hybrid density functional theory (DFT-hybrid) and DeepH-hybrid. Representative test structures with the smallest, median and largest mean absolute error (MAE) in the prediction of hybrid-functional Hamiltonian are displayed. **b** Schematic workflow of the DeepH-hybrid method, which learns from training datasets of nearly flat structures

and then generalizes to study curved nanotube structures. **c** Band structures and **d** real and imaginary parts of electric susceptibility  $\chi^{zz}$  as a function of frequency  $\omega$  of (49, 0) CNT computed by DFT-hybrid and DeepH-hybrid. The periodic direction of the nanotube is defined as the z-axis. Source data are provided as a Source Data file.

between DeepH-hybrid and DFT-hybrid (i.e., benchmark calculations using the HSE hybrid functional) is achieved for the study of band structure. One major improvement of hybrid function over LDA/GGA is the improved description of the band gap, which is closely related to optical properties. We further calculate the electric susceptibility by the method developed in ref. 59 using the gKS-DFT Hamiltonians obtained from DFT-hybrid and DeepH-hybrid, respectively. The real and imaginary parts of electric susceptibility as a function of light frequency  $\omega$  are presented in Fig. 2d. The calculated results of DeepH-hybrid are in good consistency with the DFT-hybrid benchmark data. All of these prove the reliability of our neural network method.

Next, we apply the developed method to study material systems of twisted bilayer graphene (TBG), which belongs to a general class of Moiré-twisted materials, gaining increasing interest in recent years<sup>60</sup>. Various kinds of intriguing correlated phases, such as correlated insulators, ferromagnetism, and superconductivity, have been discovered in the TBG system<sup>61–63</sup>. In particular, the TBG with twist angle  $\theta \approx 1.08^\circ$  is theoretically proposed to have ultra flat bands near the Fermi energy, which is thus named magic-angle TBG. The flat band structure of magic-angle TBG has been reproduced by DFT calculations using the PBE exchange-correlation functional in GGA. The DFT-PBE study is inherently challenging, considering that the magic-angle TBG contains 11,164 atoms per Moiré cell. In principle, the more advanced hybrid functional methods might improve the description of electronic structure but are much more expensive than DFT-PBE. Whether the flat-band feature is preserved or not in the description of hybrid functionals is a fundamentally important problem, but has not been investigated before due to the computational challenge.

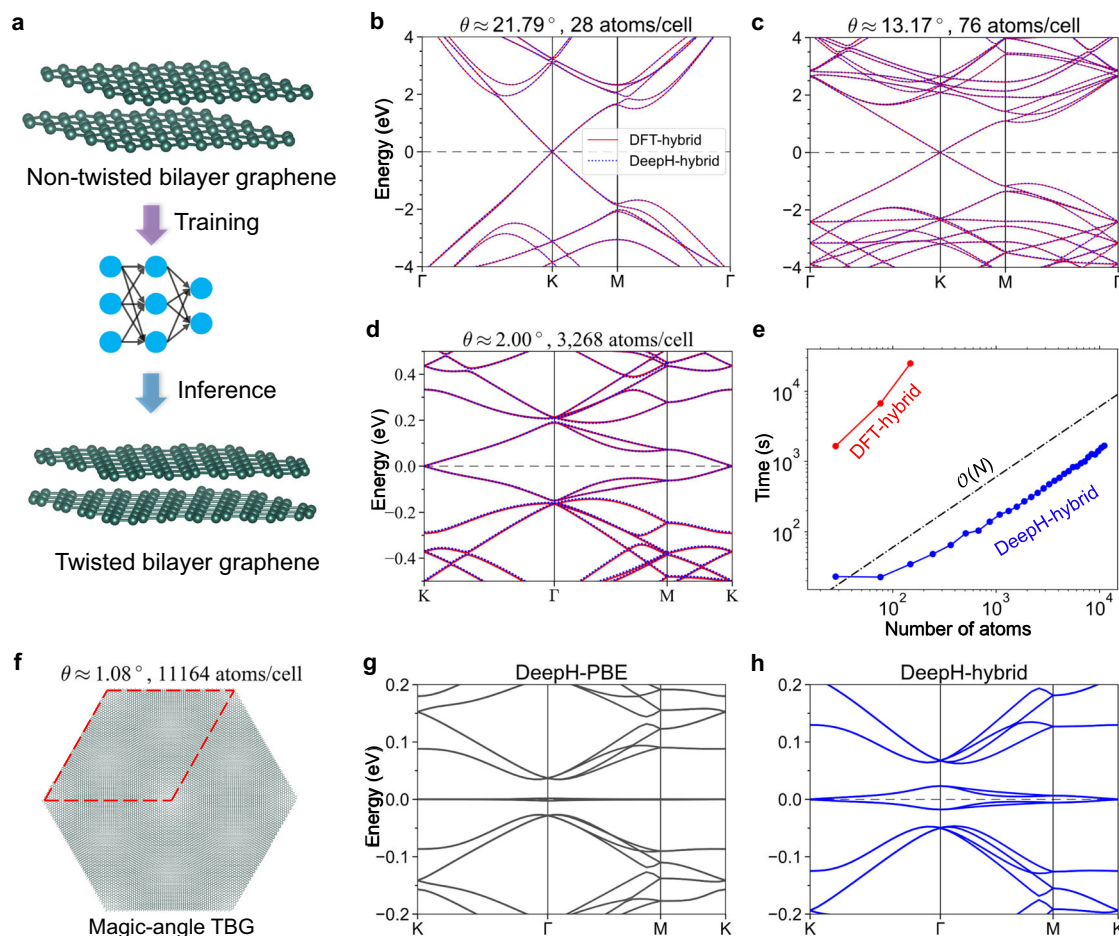
DeepH-hybrid is able to overcome the computational challenge. The strategy is as follows (Fig. 3a)<sup>26,28</sup>. Use datasets containing small-size supercells of non-twisted bilayer graphene to train neural networks, and then apply the trained neural network models to study TBGs with varying twist angles, including those with large-supercell Moiré structures. Training on bilayer graphene datasets creates a neural network model of DeepH-hybrid, whose MAEs of gKS-DFT Hamiltonian matrix elements are 0.146, 0.147, and 0.147 meV for

training, validation, and test sets, respectively. Such low MAEs ensure accurate prediction of band structures, which is confirmed by studying representative test structures as summarized in Supplementary Fig. 1. We further check the reliability of DeepH-hybrid on calculating TBGs with varying twist angles, focusing on systems with small Moiré cells for facilitating benchmark calculations. Figure 3b–d displays comparisons of the band structure of (2, 1) TBG (twist angle  $\theta \approx 21.79^\circ$ , 28 atoms/cell), (3, 2) TBG ( $\theta \approx 13.17^\circ$ , 76 atoms/cell), and (17, 16) TBG ( $\theta \approx 2.00^\circ$ , 3,268 atoms/cell). Note that we applied low-scaling algorithms at large computational cost by DFT-hybrid to compute the benchmark data for the final TBG system. Notably, the MAE of DeepH-hybrid and low-scaling DFT-hybrid Hamiltonian is 0.179 meV for (17, 16) TBG, which is comparable to the training loss, thus exemplifying DeepH-hybrid's ability to generalize to exascale systems. For all the case studies, band structures predicted by DeepH-hybrid can precisely match the benchmark results obtained by DFT-hybrid calculations.

The CPU times of calculating Hamiltonians by DFT-hybrid and DeepH-hybrid as a function of system size (i.e. number of atoms per supercell) are compared in Fig. 3e. Here linear-scaling algorithms as implemented in the ABACUS package<sup>14,64,65</sup> are applied in the DFT-hybrid calculations. In contrast, DeepH-hybrid can still reduce the computational cost by orders of magnitude, and its computational time roughly grows linearly with the system size, which demonstrates the superior efficiency of the neural network method. A more comprehensive analysis of DeepH-hybrid's time cost, taking dataset preparation and neural network optimization time into consideration, is presented in Supplementary Section 6.

Benefitting from the good accuracy and high efficiency of DeepH-hybrid, the hybrid-functional electronic structure of magic-angle TBG (Fig. 3f) can be predicted. Figure 3g,h demonstrates our results of PBE and HSE band structures of magic-angle TBG with structure relaxed by ref. 63. Both PBE and HSE Hamiltonians bear four flat bands near the Fermi surface. Compared with the PBE band structure, the bandwidth of the flat bands of the HSE band structure is increased to 41.1 meV from 4.1 meV. In addition, perturbation theory is applied to calculate





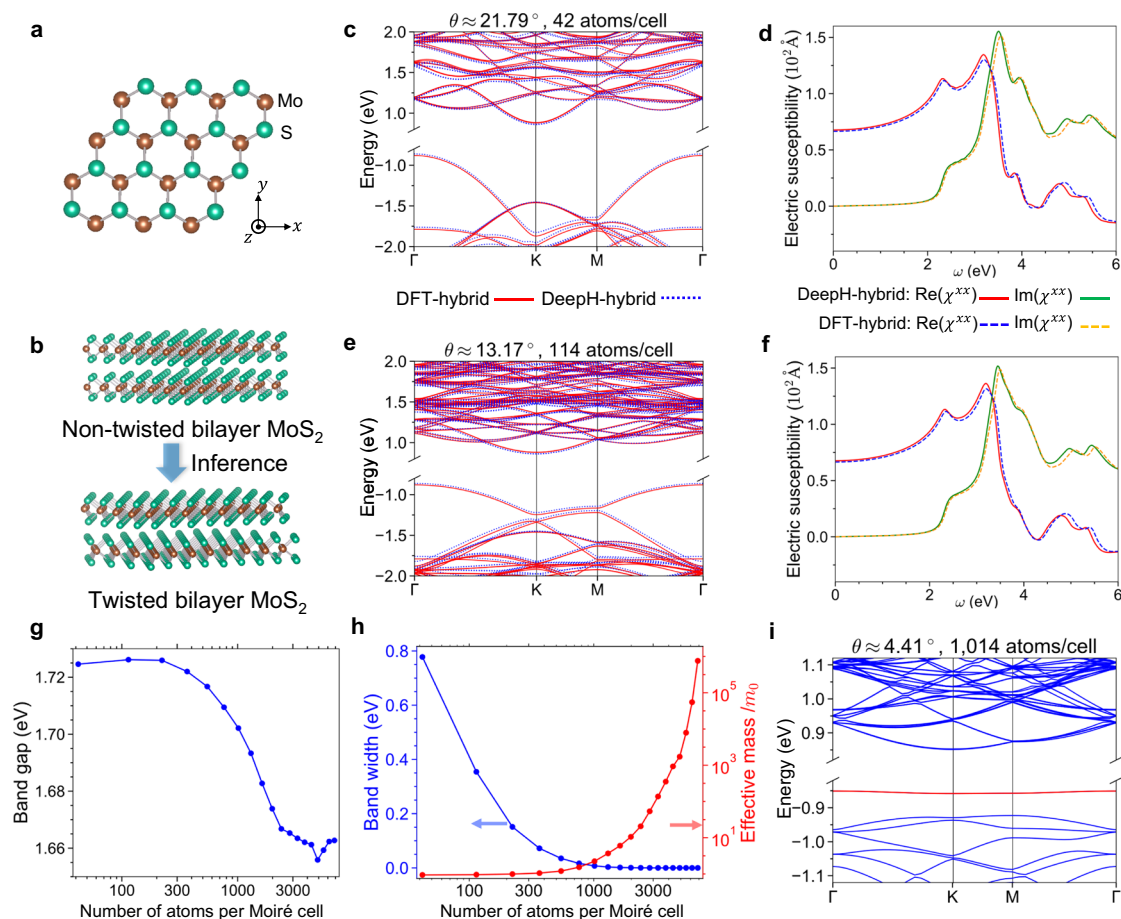
**Fig. 3 | Example studies on twisted bilayer graphene (TBG).** **a** Schematic workflow of the DeepH-hybrid method, which applies deep neural networks to learn from training datasets of non-twisted bilayer graphene and then generalizes to study TBG systems with varying twist angle  $\theta$ . **b–d** Band structures of **b** (2, 1) TBG ( $\theta \approx 21.79^\circ$ , 28 atoms/cell), **c** (3, 2) TBG ( $\theta \approx 13.17^\circ$ , 76 atoms/cell), and **d** (17, 16) TBG ( $\theta \approx 2.00^\circ$ , 3,268 atoms/cell) computed by hybrid density functional theory (DFT-hybrid) and DeepH-hybrid. **e** Computation time for predicting hybrid-functional Hamiltonians of TBGs by DFT-hybrid using linear-scaling algorithms versus by

DeepH-hybrid using neural-network inference. The dashed black line represents the time cost trend as a function of system size for an algorithm with  $O(N)$  scaling. **f** Top-view atomic structure of magic-angle TBG ( $\theta \approx 1.08^\circ$ , 11,164 atoms/cell). The Moiré supercell is highlighted by a red dashed frame. **g, h** Band structures of magic-angle TBG computed by **g** DeepH-PBE and **h** DeepH-hybrid. DeepH-PBE denotes the deep-learning DFT Hamiltonian of the PBE functional. The relaxed atomic structure of magic-angle TBG adapted from ref. 63 is used. The results of DeepH-PBE are adapted from ref. 28. Source data are provided as a Source Data file.

the Fermi velocity of HSE and PBE Hamiltonians, yielding  $v_F = 36.2$  and  $0.9 \text{ m s}^{-1}$ , respectively. From our calculation, introduction of exact exchange dramatically weakens the flatness of magic-angle TBG's flat bands, and thus could have a qualitative impact on the flat band physics of magic-angle TBG.

Regarding hybrid functionals' capability to solve the “band-gap problem”, we carried out further example studies on monolayer and bilayer H-MoS<sub>2</sub> (Fig. 4a, b). H-MoS<sub>2</sub> is a representative material of transition metal dichalcogenides, a family of materials gathering interest for its emerging new physics<sup>66,67</sup>. HSE calculations are anticipated to yield a heightened level of accuracy in the electronic band gap. Consequently, these calculations are poised to establish a more robust foundation for predicting optical properties, as well as other quasiparticle calculations. Analogous to graphene, two datasets containing monolayer and bilayer H-MoS<sub>2</sub> are constructed and used to train DeepH-hybrid models. For monolayer MoS<sub>2</sub>, the final MAEs of the predicted hybrid-functional Hamiltonian are 0.259, 0.259, and 0.258 meV for training, validation, and test sets, respectively. Band structures of representative test data predicted by DeepH-hybrid are summarized in Supplementary Fig. 2, showing good matching with DFT-hybrid band structures. For bilayer MoS<sub>2</sub>, the MAEs of the predicted hybrid-functional Hamiltonian are 0.266,

0.266, and 0.265 meV for training, validation, and test sets, respectively. Band structures of representative test data predicted by DeepH-hybrid are summarized in Supplementary Fig. 3. Comparisons of band gap of test structures between DFT-hybrid and DeepH-hybrid are summarized in Supplementary Fig. 4. Figure 4 demonstrates DeepH-hybrid's robustness in case studies of MoS<sub>2</sub>. Mean average error of the band gap on test sets of the two models is 15.1 and 16.0 meV, which are the order of magnitude smaller than the gap difference between PBE and HSE functionals. Figure 4c–f examines the capability of generalization from untwisted bilayer MoS<sub>2</sub> to twisted structures. Band structures as well as electric susceptibility of (2, 1) and (3, 2) twisted bilayer MoS<sub>2</sub> predicted by DeepH-hybrid, match well with the results of DFT-hybrid. DeepH-hybrid's efficiency makes it applicable to Moiré-twisted MoS<sub>2</sub> superstructures. Figure 4g shows the band gap of a series of ( $n$ ,  $n - 1$ ) twisted bilayer MoS<sub>2</sub>. The band gap shifts for up to 70 meV in the tested twist angles. Flat bands can be observed in twisted MoS<sub>2</sub> with large Moiré cells. The band width of the topmost occupied band and its effective mass at the  $\Gamma$  point are summarized in Fig. 4h. Band structures of these materials predicted by DeepH-hybrid are presented in Supplementary Fig. 5. A representative band structure with a flat valence band is shown in Fig. 4i.



**Fig. 4 | Example studies on twisted bilayer MoS<sub>2</sub>.** **a** Top-view atomic structure of monolayer MoS<sub>2</sub>. **b** Schematic workflow of the DeepH-hybrid method, which learns from training datasets of non-twisted bilayer MoS<sub>2</sub> and then generalizes to study twisted bilayer MoS<sub>2</sub> with varying twist angle  $\theta$ . **c, e** Band structures and **d, f** real and imaginary parts of electric susceptibility  $\chi^{xx}$  as a function of frequency  $\omega$  for **c, d** (2,1) twisted bilayer MoS<sub>2</sub> ( $\theta \approx 21.79^\circ$ , 42 atoms/cell) and **e, f** (8,7) twisted bilayer MoS<sub>2</sub>

( $\theta \approx 13.17^\circ$ , 114 atoms/cell) computed by hybrid density functional theory (DFT-hybrid) and DeepH-hybrid. The x-axis is defined in (a). **g** Band gap, **h** bandwidth and effective mass of the top valence band as a function of system size for Moiré-twisted bilayer MoS<sub>2</sub> predicted by DeepH-hybrid. **i** Band structure of (8,7) twisted bilayer MoS<sub>2</sub> ( $\theta \approx 4.41^\circ$ , 1,014 atoms/cell), which displays a flat valence band (highlighted in red). Source data are provided as a Source Data file.

## Discussions

In conclusion, we demonstrate that the conventional and generalized KS DFT can be treated within a unified deep-learning framework. Owing to the preservation of the nearsightedness principle, DeepH-hybrid's capability is proved in predicting the hybrid-functional Hamiltonian as a function of the material structure by multiple case studies. While an increased cutoff radius compared with semi-local DFT Hamiltonians is essential for handling the non-local exact exchange in DeepH-hybrid, the Hamiltonians in conventional and generalized KS frameworks share similar physical priors. This enables the deep-learning modeling of the gKS Hamiltonians with high accuracy without requiring extensive modifications to the neural network architecture, thus facilitating the simultaneous development of the neural network frameworks for both tasks in the future. Regarding the improved accuracy of band gaps with hybrid functionals, our work enables efficient study of optical properties, non-adiabatic molecular dynamics, etc., in which unoccupied conduction bands play an important role. By bypassing the time-consuming self-consistent field iterations and four-center Coulomb repulsion integrals, DeepH-hybrid has the potential to study electronic properties of superstructures at the hybrid functional level, which was previously bottlenecked due to the timely cost. Owing to the increased time cost in hybrid functionals, the relative time-saving of DeepH-hybrid will be even more significant than that in (semi)local density functionals. Test study on magic-angle TBG reveals DeepH-hybrid's ability to apply hybrid-level functionals to

Moiré-twisted superstructures with over  $10^4$  atoms. Application of DeepH-hybrid to magic-angle TBG shows a dramatic change of flat band properties when the HSE functional is applied, implying that the exact exchange could have a qualitative impact on the flat band physics of magic-angle TBG. DeepH-hybrid's success on the HSE functional can be a starting point for the generalization of DeepH to Hamiltonians from higher-level electronic structure theory, and this methodology may overcome the accuracy-efficiency dilemma of ab initio methods in general.

While the DeepH-hybrid method applies successfully to the HSE functional with a screened exact exchange, its applicability to other hybrid functionals, particularly those with unscreened exchange, is also a significant concern. There is no fundamental barrier for DeepH-hybrid being applied to unscreened hybrid functionals, since the "nearsightedness principle" is not violated. This is demonstrated by a numerical experiment on the graphene dataset using the PBE0 functional, a well-known unscreened hybrid functional<sup>68</sup>. A test-set mean absolute error (MAE) of 0.458 meV is achieved, indicating that DeepH-hybrid can also predict on PBE0 datasets with sub-meV level accuracy. Supplementary Fig. 6 shows a comparison of DFT-calculated and DeepH-predicted band structures for the test-set structure with the largest MAE, displaying a reasonable match.

Regarding the deep-learning modeling of higher-level electronic structure theories beyond hybrid DFT, the *GW* approximation may be of great concern, as the *GW* approximation offers significantly

improved description in quasiparticle excitations<sup>69</sup>. Previous research has proposed several pathways to machine learning *GW*, including modeling *GW* self-energy at imaginary frequencies<sup>70</sup>, quasiparticle spectra from the *GW* approximation<sup>71</sup>, etc. As a common practice, quantities in the *GW* approximation are projected under localized bases for machine learning, allowing them to fit within the DeepH-hybrid framework. Nevertheless, previous methods either predominantly focus on molecular systems or have relatively limited accuracy, arguably due to the non-local nature of the *GW* approximation. More careful design will be beneficial for deep-learning *GW* using DeepH-based frameworks in the future.

## Methods

### Datasets

We use the ABACUS package<sup>14,64,65</sup> to carry out hybrid functional calculations with norm-conserving pseudopotentials<sup>72</sup> using the NAO basis. The HSE06 functional with Hartree–Fock mixing constant  $\alpha = 0.25$  and screening parameter  $\omega = 0.11 \text{ Bohr}^{-1}$  is applied in all calculations<sup>51,73</sup>. While full-range hybrid functionals yield problematic results in metals due to the long-range part of the Hartree–Fock exchange<sup>74</sup>, the HSE functional screens out such contributions, facilitating its use in general solid systems. The energy cutoff for real-space grid is 400 Ry. C6.0-2s2p1d, Mo8.0-3s2p2d and S7.0-2s2p1d NAOs are applied for carbon, molybdenum, and sulfur atoms, respectively, including 13 basis functions with a cutoff radius of 6.0 Bohr for carbon, 19 basis functions with a cutoff radius of 8.0 Bohr for molybdenum and 13 basis functions with a cutoff radius of 7.0 Bohr for sulfur. For monolayer graphene and CNT, dataset is composed of 500 random structures with  $5 \times 5$  graphene supercell. Random structures are generated by introducing random offsets up to 0.1 Å, on each atom about the equilibrium configuration. For bilayer graphene, dataset is composed of 1000 random structures with  $4 \times 4$  supercell of bilayer graphene. In addition to random offsets on each atom, an overall in-plane shift is randomly assigned to each structure of bilayer graphene, and the interlayer distance is randomly sampled with normal distribution with mean 3.408 Å, and standard deviation 0.047 Å. A  $9 \times 9 \times 1$  Monkhorst–Pack *k*-mesh<sup>75</sup> is applied for supercells of monolayer and bilayer graphene. The dataset for monolayer MoS<sub>2</sub> consists of 500 randomly generated structures, while the dataset for bilayer MoS<sub>2</sub> comprises 1000 randomly generated structures. Both datasets include random structures with random offsets up to 0.1 Å on each atom in the  $4 \times 4$  supercell. A random overall interlayer shift is introduced to bilayer MoS<sub>2</sub> structures, analogous to the dataset of bilayer graphene. The interlayer distance of bilayer MoS<sub>2</sub> is fixed at 2.931 Å. A  $5 \times 5 \times 1$  Monkhorst–Pack *k*-mesh is employed for supercells of monolayer and bilayer MoS<sub>2</sub>. For the (17, 16) TBG including 3268 atoms/cell studied in Fig. 3d, a Gamma-only calculation is performed by using low-scaling techniques of resolution of identity and prescreening<sup>14</sup>. The tolerance for constructing auxiliary basis functions was set at  $10^{-4}$ . Screening tolerances for expansion coefficients, Coulomb matrix, and density matrix were set to be  $10^{-4}$ , 1.0, and  $10^{-3}$ , respectively.

### Calculation of electric susceptibility

To demonstrate the ability of DeepH-hybrid in computing physical properties, electric susceptibility is computed by using the DFT-hybrid or DeepH-hybrid Hamiltonian together with the HopTB package<sup>59</sup> via the formula:

$$\chi^{ab} = \frac{e^2}{\epsilon_0 \hbar} \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \sum_{n,m} f_{nm} \frac{r_{nm}^a r_{mn}^b}{\omega_{mn}(\mathbf{k}) - \omega - i\eta}, \quad (2)$$

where  $a, b$ , and  $c$  are Cartesian directions, while  $\epsilon_0$ ,  $\hbar$ , and  $e$  represent the vacuum permittivity, the reduced Planck's constant, and the charge of the electron, respectively.  $\omega_{mn}(\mathbf{k}) = \frac{E_{mk} - E_{nk}}{\hbar}$  and  $f_{nm} = f_n(\mathbf{k}) - f_m(\mathbf{k})$  are abbreviations for the difference of band energy and

Fermi–Dirac occupations of bands  $n$  and  $m$  at wave vector  $\mathbf{k}$ , respectively.  $r_{nm}^a$  is the Berry connection, which is defined to be zero when  $n = m$ . To ensure the integration over the Brillouin zone remains invariant with respect to the vacuum layer in low-dimensional systems, the displayed  $\chi$  are multiplied by the cross-sectional area  $A$  for quasi-one-dimensional systems or the thickness  $d$  of the supercell along the non-periodic direction for quasi-two-dimensional systems. For the CNT studied in Fig. 2d, the periodic direction is defined as the  $z$ -axis. A  $1 \times 1 \times 40$  *k*-grid was employed for the Brillouin zone integration. For the twisted bilayer MoS<sub>2</sub> studied in Fig. 4, the out-of-plane direction is defined as the  $z$ -axis. A  $20 \times 20 \times 1$  *k*-grid is employed for the Brillouin zone integration.

### Data availability

The datasets used in the current study are available via Zenodo<sup>76</sup>. Source data are provided with this paper.

### Code availability

The additional codes for DeepH-hybrid are available at GitHub (<https://github.com/aaaashanghai/DeepH-hybrid>) and Zenodo<sup>76</sup>. The code may be interfaced with the DeepH-E3 model, which is available at GitHub (<https://github.com/Xiaoxun-Gong/DeepH-E3>) and Zenodo<sup>77</sup>.

### References

1. Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372 (1993).
2. Becke, A. D. Density-functional thermochemistry. iii. The role of exact exchange. *J. Chem. Phys.* **98**, 5648 (1993).
3. Seidl, A., Görling, A., Vogl, P., Majewski, J. A. & Levy, M. Generalized Kohn–Sham schemes and the band-gap problem. *Phys. Rev. B* **53**, 3764 (1996).
4. Perdew, J. P. Density functional theory and the band gap problem. *Int. J. Quantum Chem.* **28**, 497 (1985).
5. Perdew, J. P. et al. Understanding band gaps of solids in generalized Kohn–Sham theory. *Proc. Natl Acad. Sci. USA* **114**, 2801 (2017).
6. Almlöf, J., Faegri Jr, K. & Korsell, K. Principles for a direct SCF approach to LCAO-MO ab-initio calculations. *J. Comput. Chem.* **3**, 385 (1982).
7. Häser, M. & Ahlrichs, R. Improvements on the direct SCF method. *J. Comput. Chem.* **10**, 104 (1989).
8. Burant, J. C., Scuseria, G. E. & Frisch, M. J. A linear scaling method for Hartree–Fock exchange calculations of large molecules. *J. Chem. Phys.* **105**, 8969 (1996).
9. Wu, X., Selloni, A. & Car, R. Order- $N$  implementation of exact exchange in extended insulating systems. *Phys. Rev. B* **79**, 085102 (2009).
10. Shang, H., Li, Z. & Yang, J. Implementation of exact exchange with numerical atomic orbitals. *J. Phys. Chem. A* **114**, 1039 (2010).
11. Ren, X. et al. Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *N. J. Phys.* **14**, 053020 (2012).
12. Ihrig, A. C. et al. Accurate localized resolution of identity approach for linear-scaling hybrid density functionals and for many-body perturbation theory. *N. J. Phys.* **17**, 093020 (2015).
13. Lin, P., Ren, X. & He, L. Accuracy of localized resolution of the identity in periodic hybrid functional calculations with numerical atomic orbitals. *J. Phys. Chem. Lett.* **11**, 3082 (2020).
14. Lin, P., Ren, X. & He, L. Efficient hybrid density functional calculations for large periodic systems using numerical atomic orbitals. *J. Chem. Theory Comput.* **17**, 222 (2021).
15. Lorenz, S., Groß, A. & Scheffler, M. Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks. *Chem. Phys. Lett.* **395**, 210 (2004).
16. Carleo, G. et al. Machine learning and the physical sciences. *Rev. Mod. Phys.* **91**, 045002 (2019).



17. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
18. 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).
19. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. In *Proc. 34th International Conference on Machine Learning (ICML)*, (eds Precup, D. & Teh, Y. W.) Vol. 70, 1263–1272 (JMLR.org, 2017).
20. 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).
21. Jørgensen, P. B., Jacobsen, K. W. & Schmidt, M. N. Neural message passing with edge updates for predicting properties of molecules and materials. Preprint at <https://arxiv.org/abs/arXiv:1806.03146> (2018).
22. 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).
23. 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).
24. Anderson, B., Hy, T. S. & Kondor, R. Cormorant: covariant molecular neural networks. In *Advances in Neural Information Processing Systems*, Vol. 32 (eds Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. & Garnett, R.) (Curran Associates, Inc., 2019).
25. Unke, O. T. et al. Spookynet: learning force fields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.* **12**, 7273 (2021).
26. Li, H. et al. Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.* **2**, 367 (2022).
27. Li, H. & Xu, Y. Improving the efficiency of ab initio electronic-structure calculations by deep learning. *Nat. Comput. Sci.* **2**, 418 (2022).
28. Gong, X. et al. General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian. *Nat. Commun.* **14**, 2848 (2023).
29. Li, H. et al. Deep-learning electronic-structure calculation of magnetic superstructures. *Nat. Comput. Sci.* **3**, 321–327 (2023).
30. Li, H. & Xu, Y. A deep-learning method for studying magnetic superstructures. *Nat. Comput. Sci.* **3**, 287 (2023).
31. Klicpera, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations (ICLR)* (ICLR, 2020).
32. Unke, O. T. et al. SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. In *Advances in Neural Information Processing Systems*, (eds Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S. & Wortman Vaughan, J.) 14434–14447 (Curran Associates, Inc., 2021).
33. Gu, Q., Zhang, L. & Feng, J. Neural network representation of electronic structure from ab initio molecular dynamics. *Sci. Bull.* **67**, 29 (2022).
34. 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).
35. 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).
36. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
37. Musaelian, A. et al. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **14**, 579 (2023).
38. Qiao, Z. et al. Informing geometric deep learning with electronic interactions to accelerate quantum chemistry. *Proc. Natl Acad. Sci. USA* **119**, e2205221119 (2022).
39. Nigam, J., Willatt, M. J. & Ceriotti, M. Equivariant representations for molecular Hamiltonians and *N*-center atomic-scale properties. *J. Chem. Phys.* **156**, 014115 (2022).
40. Zhang, L. et al. Equivariant analytical mapping of first principles Hamiltonians to accurate and transferable materials models. *npj Comput. Mater.* **8**, 158 (2022).
41. Unke, O. T. et al. Machine learning force fields. *Chem. Rev.* **121**, 10142 (2021).
42. Chen, Y., Zhang, L., Wang, H. & E, W. Deepks: a comprehensive data-driven approach toward chemically accurate density functional theory. *J. Chem. Theory Comput.* **17**, 170 (2021).
43. Zepeda-Núñez, L. et al. Deep density: circumventing the Kohn–Sham equations via symmetry preserving neural networks. *J. Comput. Phys.* **443**, 110523 (2021).
44. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133 (1965).
45. Mori-Sánchez, P., Cohen, A. J. & Yang, W. Localization and delocalization errors in density functional theory and implications for band-gap prediction. *Phys. Rev. Lett.* **100**, 146401 (2008).
46. Cohen, A. J., Mori-Sánchez, P. & Yang, W. Insights into current limitations of density functional theory. *Science* **321**, 792 (2008).
47. Perdew, J. P., Parr, R. G., Levy, M. & Balduz, J. L. Density-functional theory for fractional particle number: derivative discontinuities of the energy. *Phys. Rev. Lett.* **49**, 1691 (1982).
48. Perdew, J. P. & Levy, M. Physical content of the exact Kohn–Sham orbital energies: band gaps and derivative discontinuities. *Phys. Rev. Lett.* **51**, 1884 (1983).
49. Yang, W., Cohen, A. J. & Mori-Sánchez, P. Derivative discontinuity, bandgap and lowest unoccupied molecular orbital in density functional theory. *J. Chem. Phys.* **136**, 204111 (2012).
50. Levchenko, S. V. et al. Hybrid functionals for large periodic systems in an all-electron, numeric atom-centered basis framework. *Comput. Phys. Commun.* **192**, 60 (2015).
51. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **118**, 8207 (2003).
52. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864 (1964).
53. Chmiela, S. et al. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).
54. Wilkins, D. M. et al. Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proc. Natl Acad. Sci. USA* **116**, 3401 (2019).
55. Kohn, W. Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.* **76**, 3168 (1996).
56. Prodan, E. & Kohn, W. Nearsightedness of electronic matter. *Proc. Natl Acad. Sci. USA* **102**, 11635 (2005).
57. Geiger, M. et al. e3nn/e3nn: 2022-04-13. zenodo <https://doi.org/10.5281/zenodo.6459381> (2022).
58. Geiger, M. & Smidt, T. e3nn: Euclidean neural networks. Preprint at <https://arxiv.org/abs/arXiv:2207.09453> (2022).
59. Wang, C. et al. First-principles calculation of optical responses based on nonorthogonal localized orbitals. *N. J. Phys.* **21**, 093001 (2019).
60. Andrei, E. Y. & MacDonald, A. H. Graphene bilayers with a twist. *Nat. Mater.* **19**, 1265 (2020).
61. Bistritzer, R. & MacDonald, A. H. Moiré bands in twisted double-layer graphene. *Proc. Natl Acad. Sci. USA* **108**, 12233 (2011).
62. Tarnopolsky, G., Kruchkov, A. J. & Vishwanath, A. Origin of magic angles in twisted bilayer graphene. *Phys. Rev. Lett.* **122**, 106405 (2019).



63. 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  $\theta \sim 1.08^\circ$ . *Phys. Rev. B* **99**, 195419 (2019).
64. Li, P. et al. Large-scale ab initio simulations based on systematically improvable atomic basis. *Comput. Mater. Sci.* **112**, 503 (2016).
65. Chen, M., Guo, G. & He, L. Systematically improvable optimized atomic basis sets for ab initio calculations. *J. Phys. Condens. Matter* **22**, 445501 (2010).
66. Manzeli, S., Ovchinnikov, D., Pasquier, D., Yazyev, O. V. & Kis, A. 2D transition metal dichalcogenides. *Nat. Rev. Mater.* **2**, 1 (2017).
67. Cai, J. et al. Signatures of fractional quantum anomalous hall states in twisted MoTe<sub>2</sub>. *Nature* **622**, 63 (2023).
68. Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* **110**, 6158 (1999).
69. Hybertsen, M. S. & Louie, S. G. Electron correlation in semi-conductors and insulators: band gaps and quasiparticle energies. *Phys. Rev. B* **34**, 5390 (1986).
70. Dong, X., Gull, E. & Wang, L. Equivariant neural network for Green's functions of molecules and materials. *Phys. Rev. B* **109**, 075112 (2024).
71. Westermayr, J. & Maurer, R. J. Physically inspired deep learning of molecular excitations and photoemission spectra. *Chem. Sci.* **12**, 10755 (2021).
72. Morrison, I., Bylander, D. M. & Kleinman, L. Nonlocal hermitian norm-conserving Vanderbilt pseudopotential. *Phys. Rev. B* **47**, 6728 (1993).
73. Krukau, A. V., Vydrov, O. A., Izmaylov, A. F. & Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **125**, 224106 (2006).
74. Monkhorst, H. J. Hartree–Fock density of states for extended systems. *Phys. Rev. B* **20**, 1504 (1979).
75. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188 (1976).
76. Tang, Z. et al. Dataset for the article “A deep equivariant neural network approach for efficient hybrid density functional calculations” <https://doi.org/10.5281/zenodo.13444159> (2023).
77. Gong, X. et al. Code for “General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian” <https://doi.org/10.5281/zenodo.7554314> (2023).

## Acknowledgements

We thank Wenfei Li and Xinyang Dong (from AI for Science Institute, Beijing) for helping with this project. This work was supported by the Basic Science Center Project of NSFC (Grant No. 52388201), the National Natural Science Foundation of China (Grant Nos. 12334003, 12421004, 12361141826, 12134012, 12188101 and 12204332), the National Science Fund for Distinguished Young Scholars (Grant No. 12025405), the National Key Basic Research and Development Program of China (Grant No. 2023YFA1406400), the Beijing Advanced Innovation Center for

Future Chip (ICFC), and the Beijing Advanced Innovation Center for Materials Genome Engineering. The work was carried out at the National Supercomputer Center in Tianjin using the Tianhe new generation supercomputer.

## Author contributions

Y.X., W.D. and X.R. proposed the project and supervised Z.T., H.L. and P.L. in carrying out the research, with the help of X.G., G.J., L.H. and H.J. All authors discussed the results. Y.X. and Z.T. 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/s41467-024-53028-4>.

**Correspondence** and requests for materials should be addressed to Xinguo Ren, Wenhui Duan or Yong Xu.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024