nature computational science

Article

https://doi.org/10.1038/s43588-024-00747-9

Approaching coupled-cluster accuracy for molecular electronic structures with multi-task learning

Received: 16 May 2024

Accepted: 21 November 2024

Published online: 27 December 2024

Check for updates

Hao Tang[1], Brian Xiao[2], Wenhao He[3], Pero Subasic[4], Avetik R. Harutyunyan[4], Yao Wang[5], Fang Liu[5], Haowei Xu[6] ✉ & Ju Li[1,6] ✉

Machine learning plays an important role in quantum chemistry, providing fast-to-evaluate predictive models for various properties of molecules; however, most existing machine learning models for molecular electronic properties use density functional theory (DFT) databases as ground truth in training, and their prediction accuracy cannot surpass that of DFT. In this work we developed a unified machine learning method for electronic structures of organic molecules using the gold-standard CCSD(T) calculations as training data. Tested on hydrocarbon molecules, our model outperforms DFT with several widely used hybrid and double-hybrid functionals in terms of both computational cost and prediction accuracy of various quantum chemical properties. We apply the model to aromatic compounds and semiconducting polymers, evaluating both ground- and excited-state properties. The results demonstrate the model's accuracy and generalization capability to complex systems that cannot be calculated using CCSD(T)-level methods due to scaling.

Computational methods for molecular and condensed matter systems play essential roles in physics, chemistry and materials science, which can reveal underlying mechanisms of diverse physical phenomena and accelerate materials design[1]. Among various types of computational methods, quantum chemistry calculations of electronic structures are usually the bottleneck, limiting the computational speed and scalability[2]. In recent years, machine learning methods have been successfully applied to accelerate molecular dynamics simulations and improve their accuracy in many application scenarios[3]. Particularly, machine-learned inter-atomic potentials can predict energy and force of molecular systems with much lower computational costs than quantum chemistry methods[4–7]. Indeed, recent advances in universal machine-learned potentials enable large-scale molecular dynamics simulation with the complexity of realistic physical systems[8–11]. In addition to machine-learned inter-atomic potentials, rapid advances also appear in another promising direction, namely, the machine learning density functional, which focuses on further improving the energy prediction towards chemical accuracy (1 kcal mol^{-1})[12,13].

Aside from energy and force, other electronic properties that explicitly involve the electron degrees of freedom are also essential in molecular simulations[14]. In the past few years, machine learning methods have also been extended to electronic structure of molecules, predicting various electronic properties such as electric multipole moments[15–17], electron population[18], excited-state properties[19,20], and the electronic band structure of condensed matter[21,22]. Most of these methods take the density functional theory (DFT) results as the training data, and use neural networks to fit the single-configurational representation (either the Kohn–Sham Hamiltonian or molecular orbitals) of the DFT calculations[15,19,21,23]. Along with the rapid advances of machine learning techniques, the neural network predictions match the DFT

[1]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]The Center for Computational Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Honda Research Institute USA, San Jose, CA, USA. [5]Department of Chemistry, Emory University, Atlanta, GA, USA. [6]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: haoweixu@mit.edu; liju@mit.edu

results increasingly well, approaching chemical accuracy[8,16]. However, as a mean-field-level theory, DFT calculations themselves induce a systematic error that is usually several times larger than the chemical accuracy[24], limiting the overall accuracy of the machine learning model trained on DFT datasets.

By comparison, the correlated wavefunction method CCSD(T) is considered the gold-standard in quantum chemistry[25]. It provides high-accuracy predictions on various molecular properties. Unfortunately, the computational cost of CCSD(T) calculations scales unfavorably with system size. It can therefore only handle small molecules with up to hundreds of electrons. This urges the combination of CCSD(T) with machine learning methods, which, together, can potentially have both high accuracy and low computational cost. However, the above-mentioned machine learning methods that directly fit the single-configurational representation of the DFT calculations cannot be directly applied to the CCSD(T) training data. This is because CCSD(T) does not provide either Kohn–Sham Hamiltonians or single-body electronic wavefunctions due to the many-body quantum entanglement nature of its representation.

In this work we develop a unified multi-task machine learning method for molecular electronic structures. Instead of focusing solely on energy, our method also provides accurate predictions for various electronic properties. By contrast to machine learning models trained on DFT datasets, our method learns from CCSD(T)-accuracy training data. The method incorporates the E3-equivariant neural network (E3NN)[4,26], in which vectors and tensors are involved in the message-passing step. For brevity we refer to our method as multi-task electronic Hamiltonian network (MEHnet). Using hydrocarbon organic molecules as a testbed, our method predicts molecular energy within chemical accuracy as compared with both CCSD(T) calculations and experiments. It also predicts various properties such as electric dipole and quadrupole moments, atomic charge, bond order, energy gap and electric polarizability with better accuracy than B3LYP, one of the most widely used hybrid DFT functionals[27]. Our trained model shows robust generalization capability from small molecules in the training dataset (molecular weight < 100 unified atomic mass unit) to larger molecules such as naphthalene and even semiconducting polymers (molecular weight up to several thousands). Systematically predicting multiple electronic properties using a single model with local DFT computational speed, the method provides a high-performance tool for computational chemistry and a promising framework for machine learning electronic structure calculations.

Results

Computational workflow

In this section we briefly describe the theoretical background and model architecture of the MEHnet method (see Methods for details). We basically use a neural network to simulate the non-local exchange-correlation interactions of a many-body system. A physics-informed approach is then used to predict multiple properties from the output of a single neural network.

Given an input atomic configuration, our goal is to acquire an effective single-body Hamiltonian matrix that is then used to predict quantum chemical properties from physics principles (Fig. 1a). First, a fast-to-evaluate single-configurational method such as DFT or Hartree–Fock is used to obtain a mean-field effective Hamiltonian, $\mathbf{F}'$. Note that $\mathbf{F}'$ is easy and fast to compute, but its accuracy is relatively low. We will use $\mathbf{F}'$ as the starting point of our machine learning model, and the total effective Hamiltonian of the system $\mathbf{H}^{\text{eff}} = \mathbf{F}' + \mathbf{V}^{\theta}$ is obtained by adding the machine learning correction term $\mathbf{V}^{\theta}$. In the current formalism, $\mathbf{F}'$ is obtained from a local DFT calculation, and it contains only a local-exchange-correlation contribution, and the correction term $\mathbf{V}^{\theta}$ would account for the non-local exchange-correlation effects. Generally, the non-local exchange-correlation effects can be captured in CCSD(T) calculations. However, as mentioned before, the computational costs of CCSD(T) methods are formidably high for large systems. The essence of our machine learning method is to obtain the non-local exchange-correlation effects from a neural network, whose computational cost scales only linearly with system size.

To obtain the machine learning correction term, we build a neural network model to predict $\mathbf{V}^{\theta}$. The workflow consists of the input, convolutional and output layers. The input layer takes atomic configurations as input, encoding them into the node features $\mathbf{x}_{I,\text{in}}$ for atom information, and edge features $\mathbf{f}_{IJ,\text{in}}$ for bond information (I,J are indices of atoms). The E3NN framework is employed for the convolutional layer (Fig. 1b; see Methods for details) due to its good performance in predicting molecular properties[4]. The convolutional layer outputs $\mathbf{x}_{I,\text{out}}$ and $\mathbf{f}_{IJ,\text{out}}$, which encode E3-equivariant features of atoms and bonds, as well as their atomic environment. The equivariant machine learning correction Hamiltonian $\mathbf{V}^{\theta}$ is then constructed using $\mathbf{x}_{I,\text{out}}$ and $\mathbf{f}_{IJ,\text{out}}$ in the output layer. The effective electronic structure of a molecule is obtained by solving the eigenvalue equations of the total Hamiltonian $\mathbf{H}^{\text{eff}} = \mathbf{F}' + \mathbf{V}^{\theta}$, giving ϵ_i, the ith energy level, and $\mathbf{c}^i$, the corresponding molecular orbital represented on atomic orbital basis set.

Multiple learning tasks

Our scheme aims to predict multiple observable molecular properties (more than just energy). To achieve reduced computational costs, we do not include information about the entire electronic Hilbert space as learning targets. MEHnet is instead trained on a series of molecular properties to capture their shared underlying representation, that is, the effective single-body Hamiltonian $\mathbf{H}^{\text{eff}}$. The corresponding single-body energy levels and molecular orbitals are used to evaluate a series of ground-state properties O_{g} according to the rules of quantum mechanics:

$$O_{\text{g}}^{\text{MEHnet}} = f_{O_{\text{g}}}(\{\epsilon_i\},\{\mathbf{c}^i\}), \quad O_{\text{g}} = E,\vec{p},\mathbf{Q},C_I,B_{IJ}, \tag{1}$$

where O_{g} goes through the ground-state energy (E), the electric dipole ($\vec{p}$) and quadrupole ($\mathbf{Q}$) moments, the Mulliken atomic charge[28] of each atom C_I, and Mayer bond order[29] of each pair of atoms B_{IJ}. We also evaluate the energy gap (first excitation energy, E_{g}) and static electric polarizability α:

$$\begin{aligned} E_{\text{g}}^{\text{MEHnet}} &= f_{E_{\text{g}}}(\{\epsilon_i\},\{\mathbf{c}^i\},\mathbf{G}),\\ \alpha^{\text{MEHnet}} &= f_{\alpha}(\{\epsilon_i\},\{\mathbf{c}^i\},\mathbf{T}). \end{aligned} \tag{2}$$

In principle, the ground-state electronic structure does not contain the information on the energy gap and electric polarizability. We therefore use the model-output correction terms $\mathbf{G}$ (energy gap correction) and $\mathbf{T}$ (dielectric screening matrix) to account for the information on excited states and the linear response, respectively. We provide more details on the function forms of $f_{O_{\text{g}}}$, $f_{E_{\text{g}}}$ and f_{α} in the Methods. Note that these properties are all derived from the underlying electronic structure, so they are internally related. Multi-task learning methods can therefore utilize these relations to mutually enhance the model's generalization capability.

The goal of our multi-task learning is to predict the properties listed above with coupled-cluster accuracy. Hence, the total loss function L_{Total} for each molecule is constructed as follows:

$$\begin{aligned} L_{\text{Total}} &= l_V + \sum_{O\in O_{\text{g}}\cup\{E_{\text{g}},\alpha\}} l_O,\\ l_O &= w_O \times \text{MSEloss}(O^{\text{MEHnet}}, O^{\text{label}}),\\ l_V &= \frac{w_V}{N_{\text{basis}}^2}\sum_{I\mu,J\nu} |V^{\theta}_{I\mu,J\nu}|^2. \end{aligned} \tag{3}$$

Here, for each property O, l_O is the the mean-square error loss between O^{MEHnet} and O^{label}, the MEHnet predictions (equations (1) and (2)) and

https://doi.org/10.1038/s43588-024-00747-9

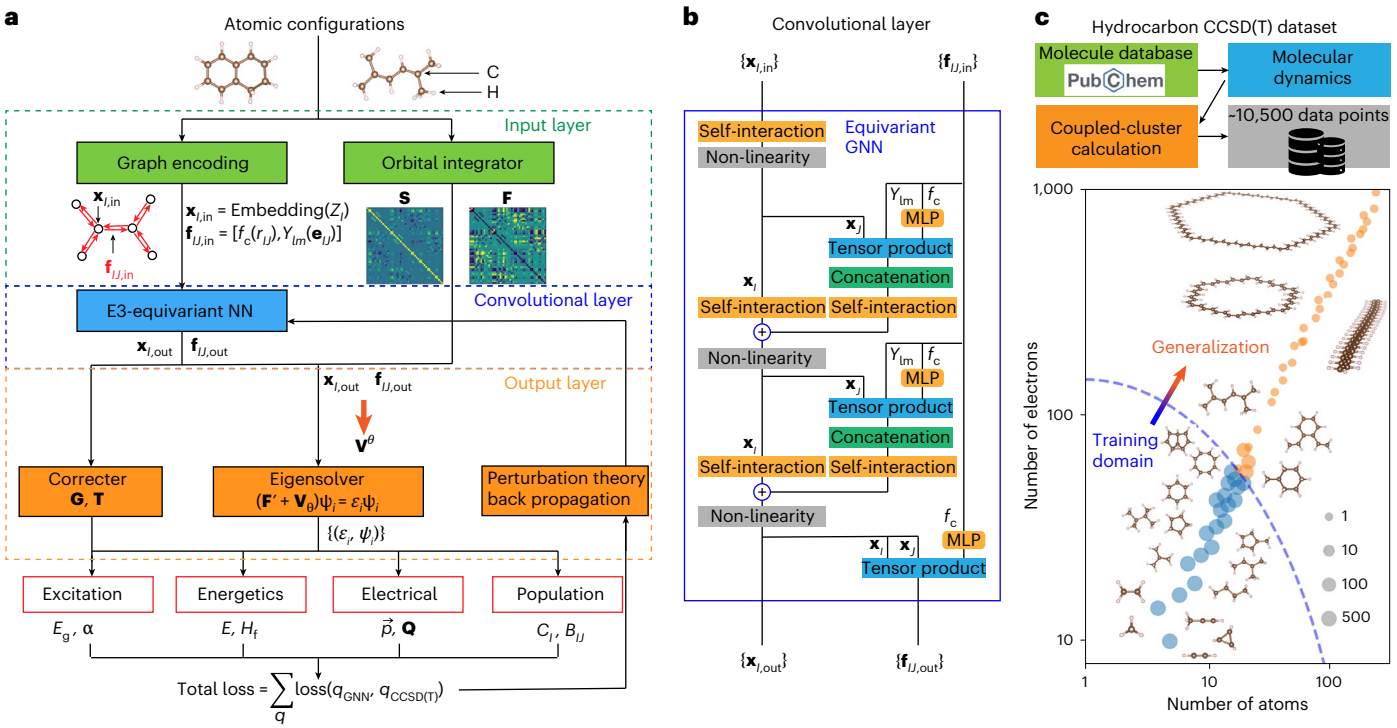


Fig. 1 | Schematic of the MEHnet electronic structure workflow. **a**, Computation graph of the MEHnet method that predicts multiple quantum chemical properties from atomic configurations inputs. The computational graph consists of an input layer (green blocks), convolutional layer (blue block) and output layer (orange blocks). **b**, Model architecture of the E3-equivariant NN with two layers of graph convolution. The output contains both node feature $\mathbf{x}_{I,\text{out}}$ and edge feature $\mathbf{f}_{IJ,\text{out}}$. **c**, Training and testing dataset generation. Each dot represents molecules with the same chemical formula, and is plotted to show the number of electrons and atoms. The blue and orange colors correspond to molecules in the training and generalization domains, respectively. The model is trained with small molecules, and is subsequently tested with large molecules. The dot size reflects the number of conformers and/or vibrational configurations with the same chemical formula in the dataset (from 1 to 500, in log scale).

coupled-cluster labels in the training dataset, respectively. Meanwhile, l_V is a regularization that penalizes large correction matrix $\mathbf{V}^\theta$, whereas N_{basis} is the total number of basis functions in the molecule. The weights w_V and w_O are hyperparameters whose values are listed in the Methods. The weights are chosen to balance the training tasks so that the training errors of all tasks decrease to satisfactory levels. Minimizing L_{Total} requires the back-propagation through the diagonalization of $\mathbf{H}^{\text{eff}}$ (that is, calculating $\partial\epsilon_i/\partial\mathbf{H}^{\text{eff}}$ and $\partial\mathbf{c}^i/\partial\mathbf{H}^{\text{eff}}$), which is numerically unstable with direct numerical differentiation. To overcome this issue, we derive customized back-propagation schemes for each property using perturbation theory in quantum mechanics (see Methods for details), giving

$$\begin{aligned}\nabla_\theta\epsilon_i &= (\mathbf{c}^i)^\dagger(\nabla_\theta\mathbf{V}^\theta)\mathbf{c}^i\\ \nabla_\theta\mathbf{c}^i &= \sum_{p\neq i}\frac{(\mathbf{c}^p)^\dagger(\nabla_\theta\mathbf{V}^\theta)\mathbf{c}^i}{\epsilon_i-\epsilon_p}\mathbf{c}^p.\end{aligned} \tag{4}$$

When evaluating the gradients of properties in equations (1) and (2) using the chain rule, terms that analytically cancel each other are removed in the numerical evaluation, making the scheme numerically stable.

Atomic configurations of molecules in our training dataset are generated by the workflow shown in Fig. 1c. Our dataset covers various classes of hydrocarbons (saturated, unsaturated, alicyclic and aromatic) and molecular structures (linear, branched and cyclic), containing both stable and metastable conformers with diverse types of carbon–carbon bonds (single, double, triple and conjugated π-bonds; see Supplementary Section 1). Coupled-cluster calculations are implemented for various hydrocarbon molecules. The MEHnet model is trained on small-molecules training dataset (training domain; Fig. 1c). The model is then tested on both small molecules in the training domain but outside the training dataset (in-domain validation) and larger molecules outside the training domain (out-of-domain validation).

Benchmark of model performance

We then benchmark the performance of the MEHnet model and display potential applications of the model in systems of practical importance. The following discussions focus on close-shell hydrocarbon molecules (except for the QM9 version of MEHnet that we will describe later).

The model's generalization capability from small to large molecules is essential for its usefulness on complex systems for which coupled-cluster calculations cannot be implemented on current computational platforms due to their formidable computational costs. To test the generalization capability and data efficiency of our model, we train the model with a varying training dataset size, N_{train}, which ranges from 10 to 7,440 atomic configurations of hydrocarbon molecules. The testing root-mean-square error (RMSE, absolute error in atomic units) of different trained properties exhibits a decreasing trend when the training dataset size increases (Fig. 2a), indicating effective model generalization. Notably, the energy error has the fastest drops with a slope of −0.38 (meaning that the testing error $\propto N_{\text{train}}^{-0.38}$). In comparison, some of the recently developed advanced machine learning potentials (that directly learn energies and their derivatives, such as the potentials in refs. 4,8) exhibit lower slopes of about −0.25. This implies a potential advantage of the multi-task method: as a multi-task method learns different molecular properties through a shared representation (the electronic structure), the domain information learned from one property can help the model's generalization on predicting other properties[30], providing improved data efficiency.

https://doi.org/10.1038/s43588-024-00747-9

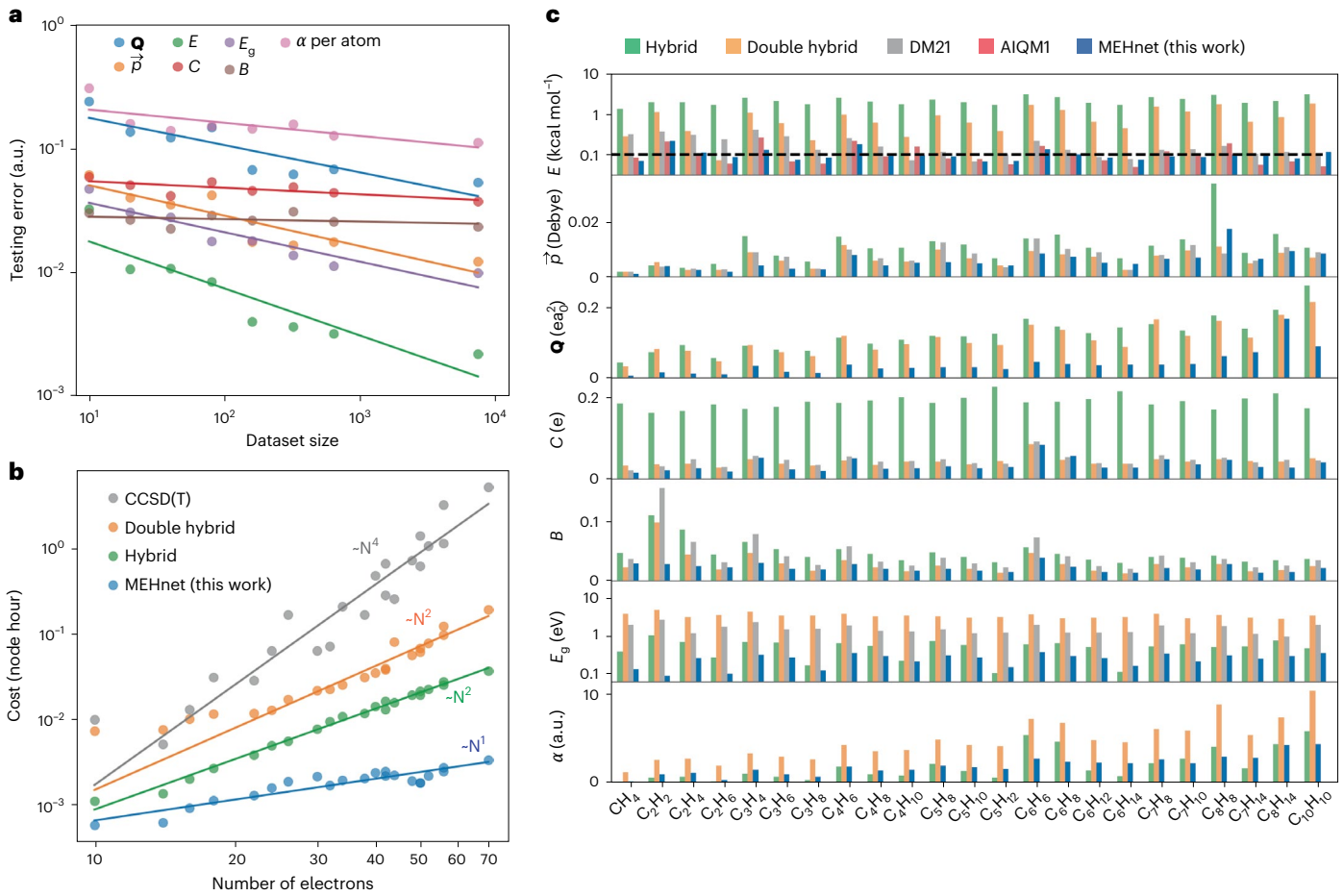


Fig. 2 | Benchmark of the model performance on the testing dataset. **a**, Testing RMSE of different quantities as a function of training dataset size. **b**, Computational costs of different methods plotted against the number of electrons. The computational cost is measured as the calculation time (node hour) on a single Intel Xeon Platinum 8260 CPU node with 48 cores with sufficient memory for all calculations. The scaling deviates from the theoretical asymptotic scaling N^7 for CCSD(T), because the parallelization efficiency is higher for larger molecules. In principle, the N^7 scaling for CCSD(T) would appear in the large N limit. **c**, Prediction RMSE of the energy (E per atom, reference to separate atoms), electric dipole moment ($\vec{p}$), electric quadrupole moment (**Q**), Mulliken atomic charge (C), Mayer bond order (B), energy gap (first excitation energy, E_g) and static electric polarizability (α, a.u. means atomic unit) with respect to the coupled-cluster results. The MEHnet method is compared with the B3LYP hybrid functional, DSD-PBEP86 double-hybrid functional[31], DM21 machine learning functional[12] and AIQM1 machine learning potential[11].

We then benchmark the computational costs and prediction accuracy of our model trained on 7,440 atomic configurations with 70 different molecules, which will be used in the rest of this paper. The MEHnet method exhibits smaller computational cost and slower scaling with system size, as compared with the hybrid functional, double-hybrid functional[31] and the CCSD(T) method (Fig. 2b). Compared to the hybrid functional, our method avoids the expensive calculation of the exact exchange, thus substantially reducing computational cost[27]. Using the gold-standard CCSD(T) calculation as a reference, the prediction accuracy of the MEHnet method on various molecular properties is compared with that of several popular functionals and existing machine learning methods (Fig. 2c and Table 1). The comparison is implemented on both the in-domain (ID) and out-of-domain (OOD) testing dataset of hydrocarbon molecules. Note that although the B3LYP hybrid functional is widely used, it is known to exhibit certain failure modes in hydrocarbon molecules[32], we therefore include several other high-performance hybrid and double-hybrid functionals[32,33] with DFT-D3 correction[34] in the comparison (Supplementary Section 2).

The MEHnet predictions consistently exhibit smaller RMSEs than the hybrid (B3LYP and B3PW91[35]), double-hybrid (DSD-PBEP86[31] and PWPB95[36]) and DM21[12] functionals on most molecular properties (Table 1, with the exception of the electric dipole moment on the OOD dataset, for which DSD-PBEP86 gives the smallest RMSE). Remarkably, the RMSE of the combination energy predicted by MEHnet is about 0.1 kcal mol^{-1} (~4 meV) per atom in both the ID and OOD datasets. Our method exhibits a similar combination energy RMSE to the AIQM1 machine learning potential, which features energy predictions within chemical accuracy. These results confirm that MEHnet's predictions on reaction energies can approach quantum chemical accuracy (assuming that on average 1 mole of molecules in reactants contain ~10 moles of atoms). Note that the B3LYP functional (with the def2-SVP basis set) exhibits large RMSEs for Mulliken charge mainly because of the basis set error[37]. Although using a large basis set for the B3LYP Mulliken charge gives a much smaller error, the MEHnet model still gives better overall accuracy (Supplementary Fig. 3).

Aside from the ground-state properties, MEHnet also provides the excited-state property E_g and linear response property α with better overall accuracy than other methods (Table 1). For intensive quantities (E per atom, C, B and E_g), the errors are on a similar level for molecules with different sizes; for extensive quantities (**p**, **Q**, α), there is a trend of increasing error with increasing system size, because the absolute values of these quantities themselves increase with system size. Furthermore, the MEHnet model gives similar prediction accuracy among different classes of hydrocarbons such as alkanes, alkenes, alkynes and

https://doi.org/10.1038/s43588-024-00747-9

Table 1 | Benchmark of MEHnet model's RMSE in predicting different quantum chemical properties on the ID testing dataset and OOD testing dataset with respect to the coupled-cluster calculations

RMSE	(ID/OOD)	Hybrid		Double	Hybrid		ML	
	Unit	B3LYP	B3PW91	DSD-PBEP86	PWPB95	DM21	AIQM1	MEHnet (ours)
Energy (per atom)	kcal mol^{-1}	2.20/2.41	2.03/2.73	0.94/1.20	1.64/1.98	0.22/0.11	0.13/0.06	0.11/0.10
Dipole	Debye	0.06/0.06	0.06/0.04	0.03/0.03	0.07/0.05	0.04/0.04	–	0.03/0.04
Quadrupole	ea_0^2	0.12/0.21	0.32/0.51	0.11/0.18	0.10/0.14	–	–	0.03/0.12
Atomic charge	e	0.19/0.20	0.16/0.16	0.04/0.05	0.05/0.05	0.05/0.04	–	0.04/0.03
Bond order	–	0.05/0.03	0.06/0.04	0.04/0.02	0.06/0.03	0.06/0.03	–	0.02/0.02
Bandgap	eV	0.59/0.63	0.65/0.54	3.71/3.26	2.19/1.98	1.71/1.47	–	0.26/0.31
Polarizability	a.u.	2.22/4.32	2.53/4.72	4.74/8.05	–	–	–	1.85/3.91

The numbers in the table are ID/OOD RMSE. Other DFT and machine learning methods are compared. We leave some of the spaces blank when the method does not directly output the quantity for fair comparison.

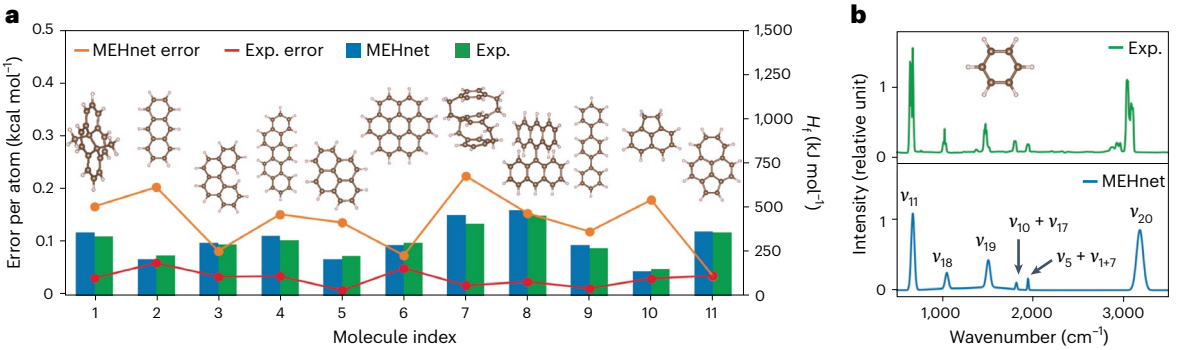


Fig. 3 | Validation of the MEHnet's predictions on gas phase aromatic hydrocarbon molecules, as compared with experimental results. a, Standard enthalpy of formation. The MEHnet predictions and experimental values from ref. 38 (right axis) are compared for 11 molecules (see Supplementary Table 2 for details). The difference between the MEHnet method and experimental values are shown by the orange line, and the experimental uncertainty is shown by the red line (left axis). **b**, Infrared spectrum of benzene. The experimental data is from the NIST Chemistry WebBook[54]. Vibration modes corresponding to the peaks are labeled following the convention in ref. 55.

arenes (Supplementary Fig. 1), suggesting consistent generalizability in the hydrocarbon chemical space.

Aromatic molecules

Hydrocarbon molecules have a vast structural space, including various types of local atomic environments. To further examine the model's generalization capability in more complex structures, we apply MEHnet to a series of aromatic hydrocarbon molecules synthesized in experiments[38]. The gas phase standard enthalpy of formation H_f is an essential thermochemical property of molecules that can be accurately measured in experiments. In this regard, we use the MEHnet model to predict H_f of various aromatic molecules in a comprehensive experimental review paper (ref. 38). The MEHnet predictions on H_f are well consistent with experiments on all molecules, and their difference is only around ~0.1–0.2 kcal mol^{-1} per atom (Fig. 3a). Note that the MEHnet prediction error is on the same order of magnitude as the experimental error bar (though numerically larger), indicating high prediction accuracy.

In addition to thermochemical properties, MEHnet can also predict spectral properties (Fig. 3b and Supplementary Fig. 4). Infrared spectra, especially, reflect essential information on molecular vibrational modes and their interaction with light. In a past work on machine learning electronic structure[16], the predicted peak intensity is usually inconsistent with the experiment. In comparison, the MEHnet predictions on both the peak positions and intensity agree well with experimental results in several common hydrocarbon molecules, and it also provides both the fundamental bands and combination bands known as benzene fingers in the infrared spectrum. The good consistency of peak intensity is attributed to accurate predictions on the transition dipole moments that determine the intensity of light–matter interaction. See Supplementary Section 3 for details on calculating the infrared spectrum, as well as the calculated infrared spectra for several other molecules.

Large-scale semiconducting polymers

Aside from small molecules, we also apply MEHnet to semiconducting polymers comprising hundreds of atoms, which are difficult to calculate by rigorous correlated methods such as CCSD(T). The essential electronic properties of semiconducting polymers originate from the conjugated π-bonds with delocalized molecular orbitals. As the delocalized molecular orbitals extend through the whole molecule (Fig. 4a), the polymers' electronic properties also involve long-range correlation, making it challenging for machine learning methods. It is therefore important to examine whether MEHnet can capture semiconducting polymers' electronic properties involving delocalized molecular orbitals.

Three kinds of semiconducting polymers: *trans*-polyacetylene (t-PA), cyclic polyacetylene (c-PA) and polyphenylene (PPP) are studied using MEHnet. The model correctly captures the delocalized π-bond feature of frontier orbitals, that is, the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) (Fig. 4a). Various important electronic properties of semiconducting polymers depend on the chain length, including the energy gap E_g and polarizability α. We calculate such chain-length dependence (up to more than 400 atoms) using the MEHnet model (Fig. 4b,c). One can see that E_g is larger for shorter oligomers and converges to a smaller value for long chains. This is in analogy to the size effect on the energy

https://doi.org/10.1038/s43588-024-00747-9

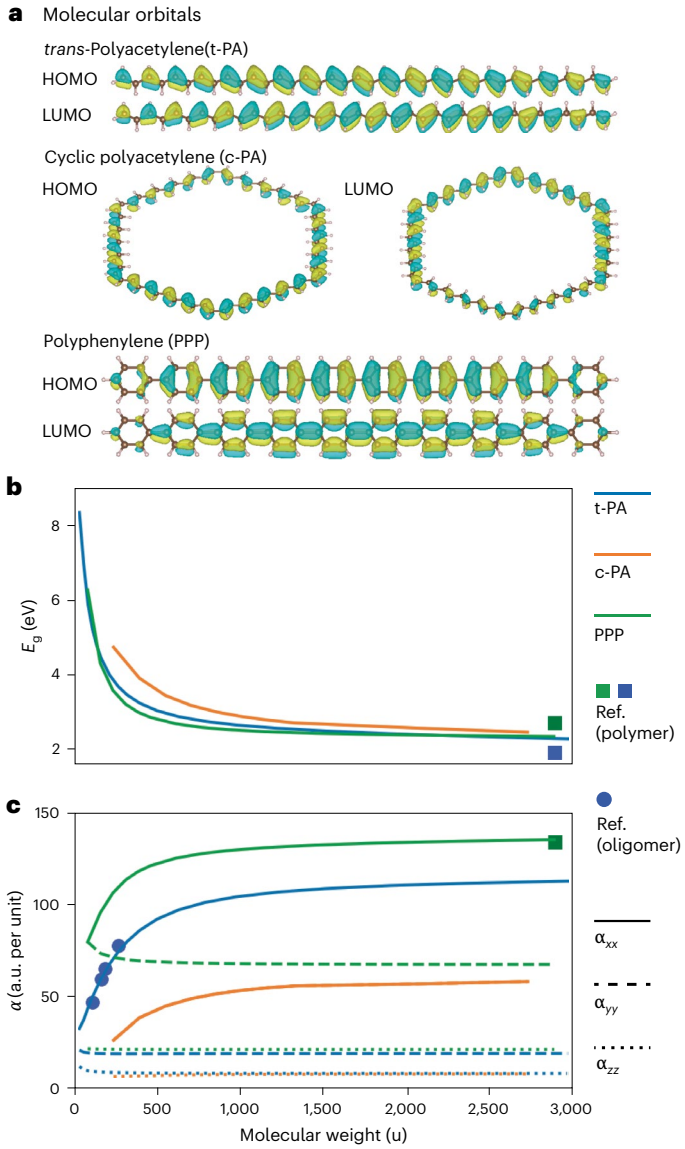


Fig. 4 | MEHnet predictions for the electronic properties of semiconducting polymers. a, Atomic structure and HOMO and LUMO wavefunctions of t-PA, polyphenylene PPP and c-PA. The wavefunctions are visualized by isosurfaces at the level of ±0.01 $Å^{-2/3}$ (positive isosurface colored blue and negative isosurface colored yellow). **b**,**c**, Energy gap (**b**) and static electric polarizability (**c**) of t-PA (blue lines), PPP (green lines) and c-PA (orange lines) with different polymer chain length. Longitudinal polarizability α_{xx}, horizontal polarizability α_{yy} and vertical polarizability α_{zz} are shown as solid, dashed and dotted lines, respectively. Squares (blue for t-PA and green for PPP) represent literature values for polymers in experiments[39,40] and correlated calculations[41], whereas blue dots represent literature values for t-PA oligomers from the MP2 correlated calculations[42].

gap of quantum dots and quantum wells. The converged energy gap for long t-PA and PPP polymers calculated by MEHnet are in reasonable agreement with the experimental values (relative errors within 10%)[39,40], which are shown as squares in Fig. 4b. The longitudinal static electric polarizability α_{xx} (per monomer) is positively related to the polymer chain length. This is because in longer chains, more delocalized electron distributions can have larger displacements under an external electric field. The predicted α_{xx} for t-PA oligomers and PPP polymers are in perfect agreement with previous correlated calculations using the high-accuracy MP2 method[41,42] (Fig. 4c). The chain-length-dependent E_g and α of c-PA, to the best of our knowledge, have not been reported. We provide their values as a prediction to be examined by future work.

QM9 version of MEHnet

Although the results in this paper mainly focus on hydrocarbons, our method is readily applicable to systems with different elements. To examine the generality of our method to the chemical space beyond hydrocarbons, we trained an MEHnet model on 10,000 molecules randomly sampled from the QM9 dataset[43]—a common quantum chemistry database including molecules comprising H, C, N, O and F atoms. The model is then tested on 4,000 other molecules randomly sampled from the QM9 dataset (see Table 2 and Supplementary Fig. 5). The prediction accuracy on the QM9 testing dataset is even better than that in the case of hydrocarbons (Table 1), suggesting that our method can be applied to more general cases with various types of elements (refer to Supplementary Section 4 for details on the benchmark of the QM9 version).

Discussion

The current MEHnet scheme has several limitations: it is not readily applicable to periodic crystals, open shell molecules or molecules with strong multi-reference character.

In principle, our approach can also be generalized to extended systems, where the periodic boundary condition (PBC) is applied. The band structure and Bloch wavefunction can then be obtained by solving the eigenvalue problem for each wave vector $\vec{k}$ after a Fourier transformation from the real space to the reciprocal space. Although the CCSD(T) method for training data generation does not directly support PBC, one can use CCSD(T) calculations for finite atom clusters (that is, a truncated and possibly passivated supercell) to train the model and subsequently use the model to predict the properties of extended systems. Alternatively, the training data of extended systems can be generated by high-accuracy methods other than CCSD(T), such as double-hybrid DFT, which allows for PBC. Aside from CCSD(T), our scheme can also use other high-level quantum chemistry methods to generate the training labels of molecule properties. Quantum chemistry methods can be selected according to the desired accuracy and the character of systems under consideration.

Note that the results of CCSD(T) calculations may not be consistently accurate for all molecules. Some of the polyaromatic hydrocarbons are more multi-reference in nature (although it is rare for the molecules studied in this paper; see Supplementary Section 5), so that the CCSD(T) calculations themselves exhibit larger errors for these molecules than for other molecules. As all of our training and testing data take CCSD(T) as the ground truth, our model cannot capture the strong multi-reference effects that are not captured by CCSD(T). One possible way to adapt the workflow to systems with strong multi-reference nature is by using the multi-reference configuration interaction method[44] to generate the training dataset. It is also possible to include one-particle reduced density matrix as an output descriptor of the MEHnet model to better describe electronic structure with strong multi-reference nature, as demonstrated in refs. 16. The one-particle reduced density matrix contains complete information on ground-state single-body properties for both single- and multi-reference systems. Adapting the MEHnet method to more comprehensive datasets can produce a general-purpose electronic structure predictor, which is left for future work.

Methods

Graph encoding of atomic configuration

The input layer takes atomic configurations as input, including the information on atomic numbers $(Z_1, Z_2, \ldots, Z_n)$ and atomic coordinates $(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n)$ of an n-atom system. A molecular graph is constructed, in which atoms are mapped to graph nodes, whereas bonds between atoms (neighboring atoms within a cut-off radius $r_{cut} = 2$ Å) are mapped to graph edges. The atomic numbers Z_I of input elements are encoded as node features $\mathbf{x}_{I,\text{in}}$ by one-hot embedding. The atomic coordinates are encoded as edge features $\mathbf{f}_{IJ,\text{in}} \equiv [f_c(r_{IJ}), Y_{lm}(\vec{e}_{IJ})]$, where $f_c(r) \equiv \frac{1}{2}[\cos(\pi \frac{r}{r_{cut}}) + 1]$ is a smooth cut-off function reflecting the

Table 2 | RMSE of the QM9 version of MEHnet model on the testing dataset of 4,000 randomly sampled configurations in the QM9 dataset

Property	*E* per atom	*p*	Q	*C*	*B*	E_g	α
Unit	kcal mol^{-1}	Debye	a.u.	e	–	eV	a.u.
RMSE	0.07	0.03	0.04	0.03	0.04	0.25	1.19

The dash indicates that the bond order *B* has no unit.

bond length $r_{IJ} \equiv |\vec{r}_I - \vec{r}_J|$, and $Y_{lm}(\vec{e}_{IJ})$ is the spherical harmonic functions acting on the unit vector $\vec{e}_{IJ} \equiv \frac{\vec{r}_I - \vec{r}_J}{|\vec{r}_I - \vec{r}_J|}$ representing the bond orientation[26]. We include Y_{lm} tensors up to $l = 2$. The electron wavefunction is represented using an atomic orbital basis set $\{|\phi_{I,\mu}\rangle\}$ (ref. 45), where I is the index of atom and μ is the index of atomic orbital basis function.

BP86 single-body effective Hamiltonian

A quantum chemistry calculation[46] (the orbital integrator block in Fig. 1a) is then used to evaluate the single-body effective Hamiltonian $F_{I\mu,J\nu}$ and overlap matrix $S_{I\mu,J\nu} \equiv \langle\phi_{I,\mu}|\phi_{J,\nu}\rangle$ in the non-orthogonal atomic orbital representation, where $I\mu$ is the row index and $J\nu$ is the column index. The **S** and **F** matrices are evaluated by the ORCA quantum chemistry program package[46] (v.5.0.4) with the quick-to-evaluate BP86 local density functional[47] and the medium-sized cc-pVDZ basis set[45]. As the hydrocarbon molecules we study are all close-shell molecules, we use spin-restricted DFT calculations to obtain **F**. We also assume the neural network correction term $\mathbf{V}^\theta$ is spin-independent as well. Namely, the spin-up and -down molecular orbitals and energy levels are the same, and all molecular orbitals are either doubly occupied or vacant.

The total BP86 energy E_{BP86} equals the molecular orbital energy $2\sum_{i=1}^{n_e/2}\epsilon_i$ (where ϵ_i is the ith molecular orbital energy level and n_e is the number of electrons) plus a many-body energy E_{MB}:

$$E_{\mathrm{BP86}} = 2\sum_{i=1}^{n_e/2}\epsilon_i + E_{\mathrm{MB}} \tag{5}$$

E_{MB} originates from the double-counting of the electron–electron interaction in the band structure energy and can be obtained from the output of the ORCA BP86 DFT calculation. The Lowdin-symmetrized Kohn–Sham Hamiltonian[48] is then obtained as

$$\mathbf{F}' \equiv \mathbf{S}^{-1/2}\mathbf{F}\mathbf{S}^{-1/2} + \frac{E_{\mathrm{MB}}}{n_e}\mathbf{I}, \tag{6}$$

where the last term is an identity shift to account for the many-body energy term. In this case, the direct summation of molecular orbital energies given by $\mathbf{F}'$ equals:

$$\begin{aligned} 2\sum_{i=1}^{n_e/2}\mathrm{eig}_i(\mathbf{F}') &= 2\sum_{i=1}^{n_e/2}\mathrm{eig}_i\left(\mathbf{S}^{-1/2}\mathbf{F}\mathbf{S}^{-1/2} + \tfrac{E_{\mathrm{MB}}}{n_e}\mathbf{I}\right) \\ &= 2\sum_{i=1}^{n_e/2}\left[\mathrm{eig}_i(\mathbf{S}^{-1/2}\mathbf{F}\mathbf{S}^{-1/2}) + \tfrac{E_{\mathrm{MB}}}{n_e}\right] \\ &= 2\sum_{i=1}^{n_e/2}\epsilon_i + E_{\mathrm{MB}}, \end{aligned} \tag{7}$$

where eig_i is a function that returns the ith lowest eigenvalue of a matrix, and we use the fact that the energy level ϵ_i is the eigenvalue of the Lowdin-symmetrized Hamiltonian $\mathrm{eig}_i(S^{-1/2}FS^{-1/2})$ (ref. 48). After this transformation, the Kohn–Sham effective Hamiltonian $\mathbf{F}'$ already includes the many-body energy E_{MB}, and the total electronic energy is just the summation of molecular orbital energies. Adding the E_{MB} term does not change the eigenfunction and relative energy levels, and thus all other properties are unchanged.

Architecture of the convolutional layer

In the convolutional layer, the input feature first goes through a $N_{\mathrm{species}} \times N_{\mathrm{species}}$ linear transformation (the first self-interaction block, N_{species} is the number of different elements in the system) and an activation layer (the first non-linearity block). All activation layers in MEHnet are tanh functions applied to scalar features. The input features then go through the first-step convolution, where the fully connected tensor product (the first tensor product block) of node feature $\mathbf{x}_J$ and the spherical Harmonic components of all connected edge features $\mathbf{f}_{IJ}$ are mapped to an irreducible representation '8 × 0e + 8 × 1o + 8 × 2e' (denoted as Irreps1), meaning there are eight even scalars, eight odd vectors and eight even rank-2 tensors. Weights in the fully connected tensor product are from a multilayer perceptron (the first multilayer perceptron (MLP) block), taking $f_c(r_{IJ})$ as input. All MLP blocks in Fig. 1b have a $1 \times 16 \times 16 \times 16 \times N_w$ structure and tanh activation function, where N_w is the number of weights in the tensor product. Then, in the concatenation block, tensor products from different edges $\mathbf{f}_{IJ}$ connected to the node I are summed to a new node feature on I. The new node features then go through a linear transformation (self-interaction block) that maps to Irreps1. In all of the self-interaction layers, linear combinations are only applied to features with the same tensor order. The new node features are added to the original node features, before undergoing the linear transformation to complete the first-step of the convolution process. The second-step convolution has the same architecture. The only difference is that the second tensor product block takes input node features of Irreps1 and output features of '8 × 0e + 8 × 0o + 8 × 1e + 8 × 1o + 8 × 2e + 8 × 2o' (denoted as Irreps2). The output of the self-interaction blocks is also Irreps2. After another activation function, the node features are output as $\mathbf{x}_{I,\mathrm{out}}$. Another tensor product is applied to the node features of the two endpoints of each edge to attain the output bond feature $\mathbf{f}_{IJ,\mathrm{out}}$; this output also has a dimension of Irreps2, with weight parameters from the MLP taking $f_c(r_{IJ})$ as input.

Finally, the output features are used to construct the correction matrix $\mathbf{V}^\theta$. The neural network correction matrix is as follows:

$$V^\theta_{I\mu,J\nu} = \begin{cases} \left[V_{\mathrm{node}}(\mathbf{x}_{I,\mathrm{out}})\right]_{\mu,\nu} & \text{if } I = J \\ \frac{1}{2}\left[V_{\mathrm{edge}}(\mathbf{f}_{IJ,\mathrm{out}})\right]_{\mu,\nu} + \frac{1}{2}\left[V_{\mathrm{edge}}(\mathbf{f}_{JI,\mathrm{out}})\right]_{\nu,\mu} & \text{if } I \neq J \end{cases} \tag{8}$$

where $V_{\mathrm{node}}(\mathbf{x}_{I,\mathrm{out}})$ is a $N_I \times N_I$ symmetric matrix rearranged from node features $\mathbf{x}_{I,\mathrm{out}}$, whereas $V_{\mathrm{edge}}(\mathbf{f}_{IJ,\mathrm{out}})$ is a $N_I \times N_J$ matrix obtained from edge features $\mathbf{f}_{IJ,\mathrm{out}}$. Here N_I and N_J are the numbers of basis functions of the atom I, J. Note that the output matrices $\mathbf{V}^\theta$ are Hermitian and equivariant under rotation according to the transformation rule of the basis set $\{|\phi_{I,\mu}\rangle\}$. $V_{\mathrm{node}}(\mathbf{x}_{I,\mathrm{out}})$ first applies a linear layer from the input dimension of Irreps2 to the output dimension Irreps$(I)^{\otimes 2}$, where:

$$\mathrm{Irreps}(I) = \begin{cases} (2 \times 0\mathrm{e} + 1 \times 1\mathrm{o}) & \text{if } I \text{ is H} \\ (3 \times 0\mathrm{e} + 2 \times 1\mathrm{o} + 1 \times 2\mathrm{e}) & \text{if } I \text{ is C} \end{cases} \tag{9}$$

The output dimension corresponds to the irreducible representation of the block diagonal terms of the Hamiltonian of the cc-pVDZ basis set. The output is then arranged into the $N_I \times N_I$ matrix form, $V_{I,\mathrm{out}}$, according to the Wigner–Eckart theorem[22], and symmetrized to obtain $V_{\mathrm{node}}(\mathbf{x}_{I,\mathrm{out}}) = \frac{\lambda_V}{2}(V_{I,\mathrm{out}} + V^T_{I,\mathrm{out}})$; λ_V is a constant hyperparameter and is set to 0.2 for our model. Similarly, the off-diagonal term $V_{\mathrm{edge}}(\mathbf{f}_{IJ,\mathrm{out}})$ in equation (8) applies a linear layer from the input dimension of Irreps2 to the output the dimension Irreps(I,J), which equals the direct product of Irreps(I) and Irreps(J). The outputs are then arranged into the $N_I \times N_J$ matrix and multiplied by λ_V, giving $V_{\mathrm{edge}}(\mathbf{f}_{IJ,\mathrm{out}})$.

Furthermore, the energy gap correction term **G** is obtained from a $8 \times 32 \times 3$ MLP that takes the even scalars of $\mathbf{x}_{I,\mathrm{out}}$ as input and outputs a three-component scalar array, $\mathbf{g}_{I;0,1,2}$, with tanh activation. The first component is for attention pooling:

$$\mathbf{G}_K = \sum_I \frac{e^{g_{I,0}}}{\sum_J e^{g_{J,0}}} g_{I,K}, \qquad K = 1, 2, \tag{10}$$

giving the two-component bandgap correction term **G**. The polarizability correction term, the screening matrix **T** is obtained from the edge features $\mathbf{f}_{IJ,\text{out}}$ going through a Irreps2 to 32 × 0e + 1 × 2e linear layer, an tanh activation layer, and a 32 × 0e + 1 × 2e to 1 × 0e + 1 × 2e linear layer. The 1 × 0e + 1 × 2e array is then multiplied by a factor λ_T (set to 0.01 in our case) and rearranged into the six independent components of the symmetric matrix, **T**.

Evaluating molecular properties

Using $\mathbf{H}^{\text{eff}}$, the electronic structure is evaluated by Schordinger equation $\mathbf{H}^{\text{eff}}\mathbf{c}^i = \epsilon_i \mathbf{c}^i$, and the molecular orbitals are $|\psi_i\rangle = \sum_{I,\mu} \tilde{c}^i_{I,\mu} |\phi_{I,\mu}\rangle$, $\tilde{\mathbf{c}}^i = \mathbf{S}^{-1/2}\mathbf{c}^i$. The ground-state properties in equation (1) are evaluated from the electronic structure from physics principles, that is, refs. 28,29:

$$\begin{aligned}
E^{\text{MEHnet}} &= E_{\text{NN}} + 2\sum_{i=1}^{n_e/2} \epsilon_i \\
\vec{p}^{\,\text{MEHnet}} &= -2e \sum_{i=1}^{n_e/2} \sum_{I\mu,J\nu} (\tilde{c}^i_{I,\mu})^* \tilde{c}^i_{J,\nu} \langle \phi_{I,\mu} | \hat{\vec{r}} | \phi_{J,\nu} \rangle \\
\mathbf{Q}^{\text{MEHnet}} &= -2e \sum_{i=1}^{n_e/2} \sum_{I\mu,J\nu} (\tilde{c}^i_{I,\mu})^* \tilde{c}^i_{J,\nu} \langle \phi_{I,\mu} | \hat{\vec{r}}\hat{\vec{r}} | \phi_{J,\nu} \rangle \\
C_I^{\text{MEHnet}} &= e\left[Z_I - 2\sum_{i=1}^{n_e/2} \sum_{J\mu\nu} (\tilde{c}^i_{I,\mu})^* \tilde{c}^i_{J,\nu} S_{I\mu,J\nu} \right] \\
B_{IJ}^{\text{MEHnet}} &= 4 \sum_{i,j=1}^{n_e/2} \sum_{KL\mu\nu\lambda\sigma} (\tilde{c}^i_{K,\lambda})^* \tilde{c}^i_{I,\mu} S_{K\lambda,J\nu} (\tilde{c}^j_{L,\sigma})^* \tilde{c}^j_{J,\nu} S_{L\sigma,I\mu}
\end{aligned} \tag{11}$$

where E_{NN} is the Coulomb repulsion energy between nuclei, and e and $\hat{\vec{r}}$ are the electron charge and position operator, respectively.

Besides, using the ground-state electronic structure, E_g can be roughly estimated as $\epsilon_{n_e/2+1} - \epsilon_{n_e/2}$, the HOMO–LUMO gap. However, in principle, the ground-state electronic structure (ϵ_n, $\mathbf{c}^n$) does not contain the information on excited states (once a electron is excited, ϵ_n and $\mathbf{c}^n$ undergo relaxation and become different). We therefore use MEHnet to output two correction terms G_1 and G_2; E_g is then evaluated as a linear transformation of the HOMO–LUMO gap using G_1 and G_2 as the coefficients:

$$E_{\text{g}}^{\text{MEHnet}} = (1 + G_1)(\epsilon_{n_e/2+1} - \epsilon_{n_e/2}) + G_2 \tag{12}$$

Evaluation of the static electric polarizability is done in two steps. First, we evaluate the single-particle polarizability α_0 using perturbation theory:

$$\alpha_0 = 2e^2 \sum_{a=n_e/2+1}^{N_{\text{basis}}} \sum_{i=1}^{n_e/2} \frac{\vec{r}_{ai}\vec{r}_{ia}}{\epsilon_a - \epsilon_i} \tag{13}$$

where N_{basis} is the number of basis functions of the molecule, and $\vec{r}_{ai} \equiv \sum_{I\mu,J\nu} (\tilde{c}^a_{I,\mu})^* \tilde{c}^i_{J,\nu} \langle \phi_{I,\mu} | \hat{\vec{r}} | \phi_{J,\nu} \rangle$. However, the single-particle approximation used in equation (13) does not consider the electric screening effect from electron–electron interaction. We use MEHnet to output a screening matrix T and evaluate the corrected polarizability α as follow:

$$\alpha^{\text{MEHnet}} = (\mathbf{I} + \alpha_0 \mathbf{T})^{-1} \alpha_0. \tag{14}$$

We evaluate the gas phase standard enthalpy of formation of molecules in Fig. 3 using atomic configurations relaxed by the BP86 functional with cc-pVDZ basis set. The total energy at the relaxed atomic configuration is then calculated by the MEHnet. The zero-point energy (ZPE) and thermal vibration, rotation, and translation energy at T = 298.15 K are also calculated by the BP86 functional with cc-pVDZ basis set implemented in ORCA. The ZPE is corrected by the optimal scaling factor of 1.0393 according to Ref. 49. Summing all energy terms give the inner energy U, and the enthalpy is evaluated as $H \simeq U + k_B T$ (k_B is the Boltzmann constant), where we use the ideal gas law. To obtain the standard enthalpy of formation, we subtract the reference state enthalpy of graphite and hydrogen gas at standard condition. The reference enthalpy for each carbon and hydrogen atom are determined as −38.04639 a.u. and −0.57550 a.u., respectively, using CCSD(T) calculation with cc-pVTZ basis set combined with measured standard enthalpy of formation of atomic carbon, atomic hydrogen, and benzene. Atomic configurations of semiconducting polymers in Fig. 4 are relaxed using the PreFerred Potential v.5.0.0 (ref. 6,7).

Perturbation theory-based back-propagation

In MEHnet training, gradient of the loss function to the model parameters needs to be calculated. Gradient back-propagation schemes are well-developed for all computation steps, with the exception of solving the Schrodinger equation. The gradients are numerically unstable when there are near-degenerate energy levels, which is usually the case in molecules. Here we first use quantum perturbation theory to obtain the first-order change of energy levels and molecular orbitals:

$$\begin{aligned}
\delta\epsilon_i &= (\mathbf{c}^i)^\dagger \delta H^{\text{eff}} \mathbf{c}^i \\
\delta\mathbf{c}^i &= \sum_{p\neq i} \frac{(\mathbf{c}^p)^\dagger \delta H^{\text{eff}} \mathbf{c}^i}{\epsilon_i - \epsilon_p} \mathbf{c}^p
\end{aligned} \tag{15}$$

We then have the gradients to model parameters as equation (4). Using these equations, we derive the gradients of each molecule properties in equation (11), as follows:

$$\begin{aligned}
\nabla_\theta f_E &= 2\sum_{i=1}^{n_e/2} \nabla V_{ii} \\
\nabla_\theta f_p &= -4e \sum_{i=1}^{n_e/2} \sum_{a=n_e/2+1}^{N_{\text{basis}}} \text{Re}\, \frac{\nabla V_{ai}}{\epsilon_i - \epsilon_a} \langle \psi_i | \hat{\vec{r}} | \psi_a \rangle \\
\nabla_\theta f_{\mathbf{Q}} &= -4e \sum_{i=1}^{n_e/2} \sum_{a=n_e/2+1}^{N_{\text{basis}}} \text{Re}\, \frac{\nabla V_{ai}}{\epsilon_i - \epsilon_a} \langle \psi_i | \hat{\vec{r}}\hat{\vec{r}} | \psi_a \rangle \\
\nabla_\theta f_{C_I} &= -4e \sum_{i=1}^{n_e/2} \sum_{a=n_e/2+1}^{N_{\text{basis}}} \text{Re}\, \frac{\nabla V_{ai} (\tilde{\mathbf{c}}^i)^\dagger (\mathbf{I}_I \mathbf{S}) \tilde{\mathbf{c}}^a}{\epsilon_i - \epsilon_a} \\
\nabla_\theta f_{B_{IJ}} &= 4 \sum_{i=1}^{n_e/2} \sum_{a=n_e/2+1}^{N_{\text{basis}}} \text{Re}\, \frac{\nabla V_{ai}}{\epsilon_i - \epsilon_a} \\
&\quad \times (\tilde{\mathbf{c}}^i)^\dagger (\mathbf{S}\mathbf{I}_J \mathbf{P} \mathbf{S}\mathbf{I}_I + \mathbf{S}\mathbf{I}_I \mathbf{P} \mathbf{S}\mathbf{I}_J) \tilde{\mathbf{c}}^a
\end{aligned} \tag{16}$$

where $\nabla V_{ai} \equiv (\mathbf{c}^a)^\dagger (\nabla_\theta \mathbf{V}^\theta) \mathbf{c}^i$, the $N_{\text{basis}} \times N_{\text{basis}}$ matrix $\mathbf{I}_J$ is identity in the block diagonal part of atom J and zero elsewhere. Meanwhile, we define $\mathbf{P} \equiv 2\sum_{i=1}^{n_e/2} \tilde{\mathbf{c}}^i (\tilde{\mathbf{c}}^i)^\dagger$. The essential method to avoid numerical instability is to remove terms that can be proved to cancel each other out. Taking $\nabla_\theta f_p$ as an example: in equation (4), the summation over p goes through all states except i. But as the summed formula in equation (16) is antisymmetric to i and a, the terms for which a ranges from 1 to $n_e/2$ cancel each other out. Only terms for which a ranges from $n_e/2 + 1$ to N_{basis} have a non-zero contribution to the final gradient. Therefore, i is always occupied, and a is always unoccupied in the summation. As close-shell molecules have a finite bandgap, ϵ_i and ϵ_a are not close to each other in any term of the summation, so evaluating equation (16) is numerically stable.

Similarly, the gradients of E_g and α are as follow:

$$\begin{aligned}
\nabla_\theta f_{E_{\text{g}}} &= (1 + G_1)\left[\nabla V_{n_e/2+1,n_e/2+1} - \nabla V_{n_e/2,n_e/2}\right] \\
&\quad + (\epsilon_{n_e/2+1} - \epsilon_{n_e/2}) \nabla_\theta G_1 + \nabla_\theta G_2
\end{aligned} \tag{17}$$

To calculate the gradient of α, we first evaluate the gradient of α_0, and then derive $\nabla_\theta f_\alpha$ using the chain rule:

$$\begin{aligned}\nabla_\theta \alpha_0 &= 2e^2 \sum_{a=n_e/2+1}^{N_{\text{basis}}} \sum_{i=1}^{n_e/2} \text{Re}\left\{\frac{\vec{r}_{ai}\vec{r}_{ia}(\nabla V_{ii}-\nabla V_{aa})}{(\epsilon_a-\epsilon_i)^2}\right. \\ &\quad \left. - 2\sum_{p\neq a,i}\frac{\vec{r}_{ai}}{(\epsilon_a-\epsilon_i)}\left[\frac{\vec{r}_{ip}\nabla V_{pa}}{(\epsilon_p-\epsilon_a)}+\frac{\vec{r}_{pa}\nabla V_{ai}}{(\epsilon_p-\epsilon_i)}\right]\right\} \qquad (18)\\ \nabla_\theta f_\alpha &= (\mathbf{I}+\alpha_0\mathbf{T})^{-1}(\nabla_\theta\alpha_0)(\mathbf{I}-\mathbf{T}\alpha) \\ &\quad - \alpha_0(\nabla_\theta\mathbf{T})(\mathbf{I}+\alpha_0\mathbf{T})^{-1}\alpha\end{aligned}$$

The above equations give gradients of all terms in the loss function expressed by gradients to the direct outputs of the MEHnet, $\nabla_\theta \mathbf{V}^\theta$, $\nabla_\theta \mathbf{G}$ and $\nabla_\theta \mathbf{T}$.

Dataset generation

First, 85 small hydrocarbon molecule structures are collected from the PubChem database[50]. The training domain (out-of-domain testing dataset) includes 20 (3) different chemical formula correspond to the horizontal axis labels of the first 20 (last 3) columns in Fig. 2c. Each chemical formula includes up to five different molecules (conformers) taken from the PubChem database. The total number of molecules (conformers) in the training domain and out-of-domain testing dataset is 70 and 15, respectively. The full list of molecules and the number of atomic configurations for each molecule are listed in Supplementary Table 1. Refer to Supplementary Section 1 for a discussion on the principles of selecting these molecules, and their diversity.

Molecular dynamics simulation with TeaNet interatomic potential[6,7] is then performed for each molecule structure to sample an ensemble of atomic configurations. The molecular dynamics simulation uses PreFerred Potential v.4.0.0 (ref. 10) at a temperature of 2,000 K, which enables large bond distortion but does not break the bonds. The initial velocity is set as a Maxwell Boltzmann distribution with the same temperature. Langevin *NVT* dynamics is used with the friction factor of 0.001 fs^{-1} and timestep of 2 fs. The TeaNet potential run for 100,000 steps for each chemical formula, and one atomic configuration is sampled every 200 timesteps in the molecular dynamics trajectory; 500 configurations (including the initial equilibrium configuration) are sampled for each chemical formula in the training domain; three-quarters of the 10,000 configurations are sampled to form the training dataset, and the remaining one-quarter forms the in-domain testing dataset. The out-of-domain testing dataset contains 500 configurations. Note that as we aim to include structures out of equilibrium positions, geometric relaxation is not needed before CCSD(T) calculation (otherwise all structures will relax back to the equilibrium positions).

A CCSD(T) calculation with the cc-pVTZ basis set is then implemented in ORCA[46] for each selected configuration, giving the training labels of total energy, electric dipole and quadrupole moment, Mulliken atomic charge, and Mayer bond order. An EOM-CCSD calculation[51] with the cc-pVDZ basis set is then implemented to obtain the first excitation energy (energy gap). Finally, we conduct a polarizability calculation with the CCSD and cc-pVDZ basis set. The overlap matrix S and starting-point effective Hamiltonian F is obtained from a BP86 DFT calculations with the cc-pVDZ basis set.

Model training

The weight parameters in the loss function is listed as follow: $w_V = 0.1$, $w_E = 1$, $w_p = 0.2$, $w_\mathbf{Q} = 0.01$, $w_C = 0.01$, $w_B = 0.02$, $w_{E_g} = 0.2$, $w_\alpha = 3\times10^{-5}$. All quantities are in atomic unit. The model training is implemented by full gradient descend (FGD) with Adam optimizer. For the finally deployed model (7,440 training data points), it is first trained on 1,240 data points sampled from the whole training dataset for 5,000 FGD steps with initial learning rate of 0.01. The learning rate is decayed by a constant factor $\gamma_1 = 10^{-1/10}$ per 500 steps. The model is then trained on the whole dataset with 7,440 data points for 6,000 FGD steps with a learning rate of 0.001. For other models trained on smaller dataset in Fig. 2a in the main text, the model is trained for 3,000 FGD steps with initial learning rate of 0.01 decayed by $\gamma_2 = 10^{-1/6}$ per 500 steps. As the model trained on 640 data points do not converge in the 3,000-step training, we implement a 10,000-step training, with an initial learning rate of 0.01 that decays by γ_1 every 500 steps in the first 5,000 steps and keeps constant at the last 5,000 steps.

Data availability

Raw computational data files and the training and testing datasets are available with this manuscript through FigShare at https://doi.org/10.6084/m9.figshare.25762212 (ref. 52). Source Data are provided with this paper.

Code availability

The source code to generate the training dataset, train the MEHnet model, and apply the trained MEHnet model to hydrocarbon molecules has been deposited into a publicly available GitHub repository at https://github.com/htang113/Multi-task-electronic (ref. 53), and is also available in the Supplementary Software. The repository contains two branches: the branch v.1.6 is for all results of hydrocarbon molecules in this paper, and the branch v.2.0 is for the benchmark on the QM9 dataset.

References

1. Carter, E. A. Challenges in modeling materials properties without experimental input. *Science* **321**, 800–803 (2008).
2. Kulik, H. J. et al. Roadmap on machine learning in electronic structure. *Electron. Struct.* **4**, 023004 (2022).
3. Dral, P. O. *Quantum Chemistry in the Age of Machine Learning* (Elsevier, 2022).
4. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
5. Zhang, L., Han, J., Wang, H., Car, R. & Weinan, E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
6. Takamoto, S., Izumi, S. & Li, J. Teanet: universal neural network interatomic potential inspired by iterative electronic relaxations. *Comput. Mater. Sci.* **207**, 111280 (2022).
7. Takamoto, S., Okanohara, D., Li, Q. & Li, J. Towards universal neural network interatomic potential. *J. Materiomics* **9**, 447–454 (2023).
8. Merchant, A. et al. Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023).
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. Takamoto, S. et al. Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements. *Nat. Commun.* **13**, 2991 (2022).
11. Zheng, P., Zubatyuk, R., Wu, W., Isayev, O. & Dral, P. O. Artificial intelligence-enhanced quantum chemical method with broad applicability. *Nat. Commun.* **12**, 7022 (2021).
12. Kirkpatrick, J. et al. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science* **374**, 1385–1389 (2021).
13. 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).
14. Helgaker, T., Jorgensen, P. & Olsen, J. *Molecular Electronic-Structure Theory* (John Wiley & Sons, 2013).

15. 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).
16. Shao, X., Paetow, L., Tuckerman, M. E. & Pavanello, M. Machine learning electronic structure methods based on the one-electron reduced density matrix. *Nat. Commun.* **14**, 6281 (2023).
17. Feng, C., Xi, J., Zhang, Y., Jiang, B. & Zhou, Y. Accurate and interpretable dipole interaction model-based machine learning for molecular polarizability. *J. Chem. Theory Comput.* **19**, 1207–1217 (2023).
18. Fan, G., McSloy, A., Aradi, B., Yam, C.-Y. & Frauenheim, T. Obtaining electronic properties of molecules through combining density functional tight binding with machine learning. *J. Phys. Chem. Lett.* **13**, 10132–10139 (2022).
19. Cignoni, E. et al. Electronic excited states from physically constrained machine learning. *ACS Central Sci.* **10**, 637–648 (2023).
20. Dral, P. O. & Barbatti, M. Molecular excited states through a machine learning lens. *Nat. Rev. Chem.* **5**, 388–405 (2021).
21. Li, H. et al. Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.* **2**, 367–377 (2022).
22. Gong, X. et al. General framework for E(3)-equivariant neural network representation of density functional theory hamiltonian. *Nat. Commun.* **14**, 2848 (2023).
23. Unke, O. et al. SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. *Advances in Neural Information Processing Systems* **34**, 14434–14447 (2021).
24. Mardirossian, N. & Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **115**, 2315–2372 (2017).
25. Bartlett, R. J. & Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Modern Phys.* **79**, 291 (2007).
26. Geiger, M. & Smidt, T. e3nn: Euclidean neural networks. Preprint at https://arxiv.org/abs/2207.09453 (2022).
27. Tirado-Rives, J. & Jorgensen, W. L. Performance of B3LYP density functional methods for a large set of organic molecules. *J. Chem. Theory Comput.* **4**, 297–306 (2008).
28. Mulliken, R. S. Electronic population analysis on LCAO–MO molecular wave functions. I. *J. Chem. Phys.* **23**, 1833–1840 (1955).
29. Mayer, I. Bond order and valence indices: a personal account. *J. Comput. Chem.* **28**, 204–221 (2007).
30. Caruana, R. Multitask learning. *Mach. Learn.* **28**, 41–75 (1997).
31. Kozuch, S. & Martin, J. M. Spin-component-scaled double hybrids: an extensive search for the best fifth-rung functionals blending DFT and perturbation theory. *J. Comput. Chem.* **34**, 2327–2344 (2013).
32. Karton, A. How reliable is DFT in predicting relative energies of polycyclic aromatic hydrocarbon isomers? Comparison of functionals from different rungs of jacob's ladder. *J. Comput. Chem.* **38**, 370–382 (2017).
33. Goerigk, L. & Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **13**, 6670–6688 (2011).
34. 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).
35. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
36. Goerigk, L. & Grimme, S. Efficient and accurate double-hybrid-meta-GGA density functionals evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **7**, 291–309 (2011).
37. Jablonnski, M. & Palusiak, M. Basis set and method dependence in atoms in molecules calculations. *J. Phys. Chem. A* **114**, 2240–2244 (2010).
38. Slayden, S. W. & Liebman, J. F. The energetics of aromatic hydrocarbons: an experimental thermochemical perspective. *Chem. Rev.* **101**, 1541–1566 (2001).
39. Heeger, A. J. Nobel lecture: Semiconducting and metallic polymers: the fourth generation of polymeric materials. *Rev. Modern Phys.* **73**, 681 (2001).
40. Grem, G., Leditzky, G., Ullrich, B. & Leising, G. Realization of a blue-light-emitting device using poly (*p*-phenylene). *Adv. Mater.* **4**, 36–37 (1992).
41. Otto, P., Piris, M., Martinez, A. & Ladik, J. Dynamic (hyper) polarizability calculations for polymers with linear and cyclic π-conjugated elementary cells. *Synth. Metals* **141**, 277–280 (2004).
42. Champagne, B. et al. Assessment of conventional density functional schemes for computing the polarizabilities and hyperpolarizabilities of conjugated oligomers: an ab initio investigation of polyacetylene chains. *J. Chem. Phys.* **109**, 10489–10498 (1998).
43. Ramakrishnan, R., Dral, P. O., Rupp, M. & Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022 (2014).
44. Szalay, P. G., Muller, T., Gidofalvi, G., Lischka, H. & Shepard, R. Multiconfiguration self-consistent field and multireference configuration interaction methods and applications. *Chem. Rev.* **112**, 108–181 (2012).
45. Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. i. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).
46. Neese, F., Wennmohs, F., Becker, U. & Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **152**, 224108 (2020).
47. Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **33**, 8822 (1986).
48. Löwdin, P.-O. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *J. Chem. Phys.* **18**, 365–375 (1950).
49. Kesharwani, M. K., Brauer, B. & Martin, J. M. Frequency and zero-point vibrational energy scale factors for double-hybrid density functionals (and other selected methods): can anharmonic force fields be avoided? *J. Phys. Chem. A* **119**, 1701–1714 (2015).
50. Kim, S. et al. PubChem 2023 update. *Nucl. Acids Res.* **51**, D1373–D1380 (2022).
51. Krylov, A. I. Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: the hitchhiker's guide to Fock space. *Annu. Rev. Phys. Chem.* **59**, 433–462 (2008).
52. Tang, H. et al. Training and testing datasets in 'Multi-task learning for molecular electronic structure approaching coupled-cluster accuracy'. (FigShare, 2024); https://doi.org/10.6084/m9.figshare.25762212
53. *htang113/Multi-task-electronic* (GitHub, 2024); https://github.com/htang113/Multi-task-electronic

54. Linstrom, P. & W.G. Mallard, E. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69* (National Institute of Standards and Technology, 2024).
55. Wilson Jr, E. B. The normal modes and frequencies of vibration of the regular plane hexagon model of the benzene molecule. *Phys. Rev.* **45**, 706 (1934).

Acknowledgements

This work was supported by Honda Research Institute (HRI-USA). H.T. acknowledges support from the Mathworks Engineering Fellowship. The calculations in this work were performed in part on the Matlantis high-speed universal atomistic simulator, the Texas Advanced Computing Center (TACC), the MIT SuperCloud, and the National Energy Research Scientific Computing (NERSC).

Author contributions

All authors contributed to the discussions of theory and results and to writing the manuscript. J.L. designed and guided the project and formulated the research goals. H.T., H.X., B.X. and W.H. designed and developed the computational method and code package, generated the quantum chemistry dataset, implemented the machine learningtraining and applications, and did data analysis and visualization. A.H. initiated the theme and formulated the research goals. Y.W., F.L. and P.S. provided important comments for the paper.

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/s43588-024-00747-9.

Correspondence and requests for materials should be addressed to Haowei Xu or Ju Li.

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

nature computational science

Supplementary information

https://doi.org/10.1038/s43588-024-00747-9

Approaching coupled-cluster accuracy for molecular electronic structures with multi-task learning

In the format provided by the authors and unedited

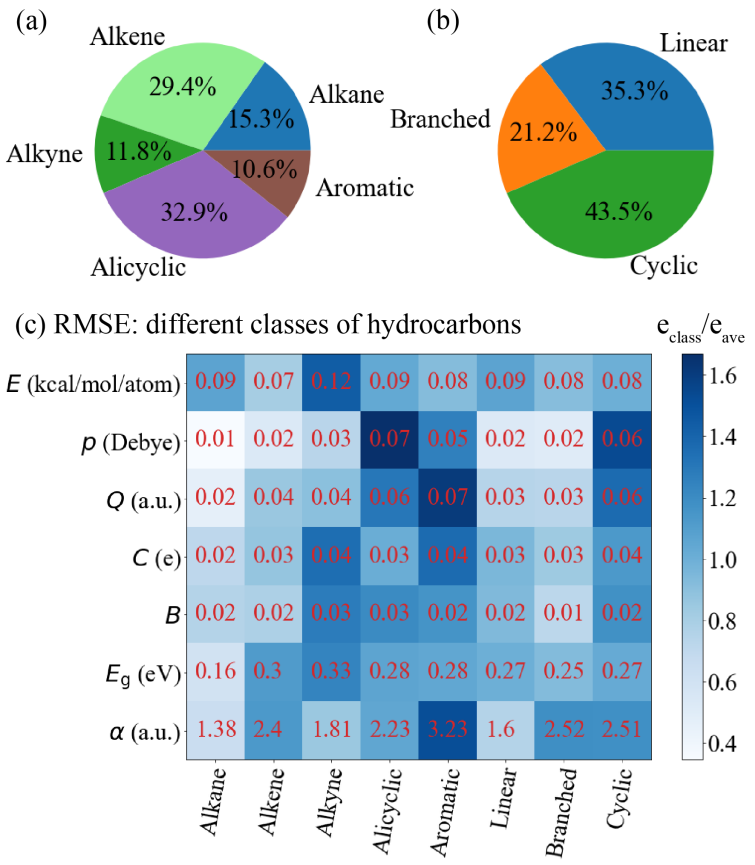
CONTENTS

Supplementary Section 1. DETAILS ON DATASET

The molecules are selected based on the principle of structural diversity, covering various different classes of hydrocarbons. Details about the structural diversity of our collected hydrocarbons are summarized in Supplementary Figure 1. Hydrocarbon molecules can be classified into 4 classes [Supplementary Figure 1a]: saturated hydrocarbon (alkane), unsaturated hydrocarbon (alkene and alkyne), alicyclic hydrocarbon, and aromatic hydrocarbon. On the other hand, the molecule structure can be categorized into 3 classes [Supplementary Figure 1(b)]: linear structure, branched structure, and cyclic structure. Our training dataset covers a number of molecules in each class. We further examine the testing errors of our model in different classes of molecules in the dataset, which are shown as numbers in Supplementary Figure 1(c). We can see for each quantity (labelled on y-axis), the errors are generally close among all classes of molecules (labelled on x-axis). For each quantity, we also calculated how the error for each class of molecules deviates from the average value among all classes, which are marked with the color. One can see that the deviation is no more than 60%, showing that our trained model has consistently good prediction accuracy for various classes of hydrocarbons. This further validates that the molecules we selected for training provides sufficient and balanced training data so that the model learns the electronic structure of different classes of hydrocarbons.

Supplementary Table 1. Composition of the hydrocarbon dataset. The table contains a list of molecule names and number of atomic configurations (labeled in the superscript) for each molecule.

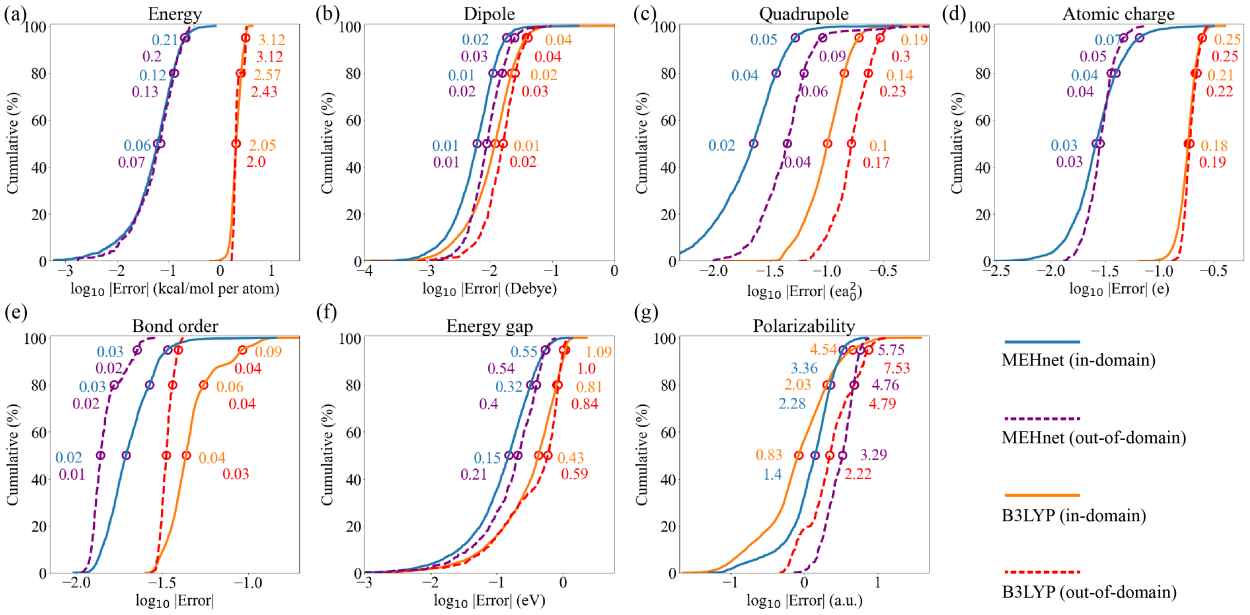
Chemical formula	Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5
CH_4	Methane^{500}	–	–	–	–
C_2H_2	Acetylene^{500}	–	–	–	–
C_2H_4	Ethylene^{500}	–	–	–	–
C_2H_6	Ethane^{500}	–	–	–	–
C_3H_4	Propyne^{300}	Allene^{100}	Cyclopropene^{100}	–	–
C_3H_6	Propylene^{250}	Cyclopropane^{250}	–	–	–
C_3H_8	Propane^{500}	–	–	–	–
C_4H_6	1,2-Butadiene^{100}	1,3-Butadiene^{100}	1-Butyne^{100}	2-Butyne^{100}	1-Methylcyclopropene^{100}
C_4H_8	Isobutylene^{100}	Cyclobutane^{100}	1-Butene^{100}	2-Butene^{100}	Methylcyclopropane^{100}
C_4H_{10}	Butane^{250}	Isobutane^{250}	–	–	–
C_5H_8	Isoprene^{100}	Cyclopentene^{100}	1-Pentyne^{100}	Methylene-cyclobutane^{100}	1,3-Pentadiene^{100}
C_5H_{10}	Cyclopentane^{100}	1-Pentene^{100}	2-Methyl-1-Butene^{100}	2-Methyl-2-Butene^{100}	3-Methyl-1-Butene^{100}
C_5H_{12}	Neopentane^{200}	Isopentane^{200}	Pentane^{100}	–	–
C_6H_6	Benzene^{100}	1,5-Hexadiyne^{100}	2,4-Hexadiyne^{100}	Divinylacetylene^{100}	3,4-Dimethylene-cyclobut-1-ene^{100}
C_6H_8	1,3-Cyclohexadiene^{100}	1,4-Cyclohexadiene^{100}	Hexa-1,3,5-triene^{100}	Methyl-cyclopentadiene^{100}	Divinylethylene^{100}
C_6H_{12}	Methyl-cyclopentane^{100}	Cyclohexane^{100}	1-Hexene^{100}	cis-4-Methyl-2-pentene^{100}	2-Methyl-1-Pentene^{100}
C_6H_{14}	2,2-Dimethylbutane^{100}	2,3-Dimethylbutane^{100}	3-Methylpentane^{100}	2-Methylpentane^{100}	Hexane^{100}
C_7H_8	Toluene^{100}	2,5-Norbornadiene^{100}	Quadricyclane^{100}	1,6-Heptadiyne^{100}	Cycloheptatriene^{100}
C_7H_{10}	Norbornene^{100}	1,3-Cycloheptadiene^{100}	1-Methyl-1,3-cyclohexadiene^{100}	2-Methyl-1,3-cyclohexadiene^{100}	3-Methylenecyclohexene^{100}
C_7H_{14}	Methyl-cyclohexane^{50}	Cycloheptane^{50}	1-Heptene^{50}	(E)-4,4-Dimethyl-2-pentene^{50}	trans-3-Heptene^{50}
C_8H_8	Styrene^{100}	Benzocyclobutene^{100}	Cubane^{100}	Semibullvalene^{100}	Cyclooctatetraene^{100}
C_8H_{14}	Bimethallyl^{25}	Diisocrotyl^{50}	1,7-Octadiene^{25}	CYCLOOCTENE^{25}	(4E)-2,3-dimethylhexa-1,4-diene^{25}
$C_{10}H_{10}$	1,3-Divinylbenzene^{20}	Naphthalene^{20}	1,4-Divinylbenzene^{20}	Divinylbenzene^{20}	4-Phenyl-1-butyne^{20}



Supplementary Figure 1. Composition of the dataset in different (a) hydrocarbon molecule classes (alkane, alkene, alkyne, cyclic, and aromatic) and (b) structural classes (linear, branched, and alicyclic). Percentage of the number of molecules in the dataset is shown in the plot. (c) RMSE of the MEHnet model predictions on different classes of molecules in the testing dataset. The red numbers denote absolute values of the RMSE, and the color reflects the ratio of the RMSE on a specific class of molecule to the average RMSE on all molecules in the testing dataset.

Supplementary Section 2. DETAILS ON BENCHMARKS

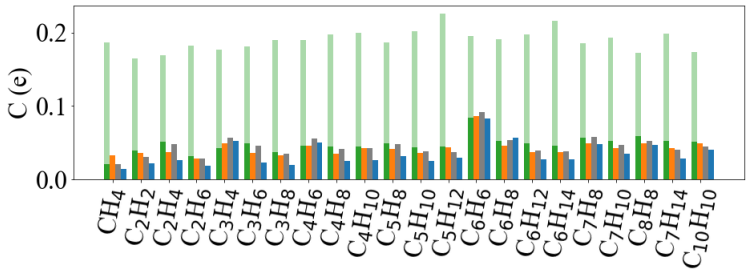
For comparison, B3LYP and B3PW91 hybrid DFT calculations are implemented with def2-SVP basis set in ORCA. DSD-PBEP86 and PWPB95 double-hybrid DFT calculations are implemented with the def2-TZVP basis set in ORCA. The DM21 machine learning DFT calculations are implemented with the def2-TZVP basis set in PySCF program package [1]. The AIQM1 calculations are implemented in the MLatom platform [2]. For DM21 and AIQM1, the isolated-atom energies of carbon and hydrogen are re-calibrated according to the least-mean-square criterion to give optimal combination energies in our dataset. Besides



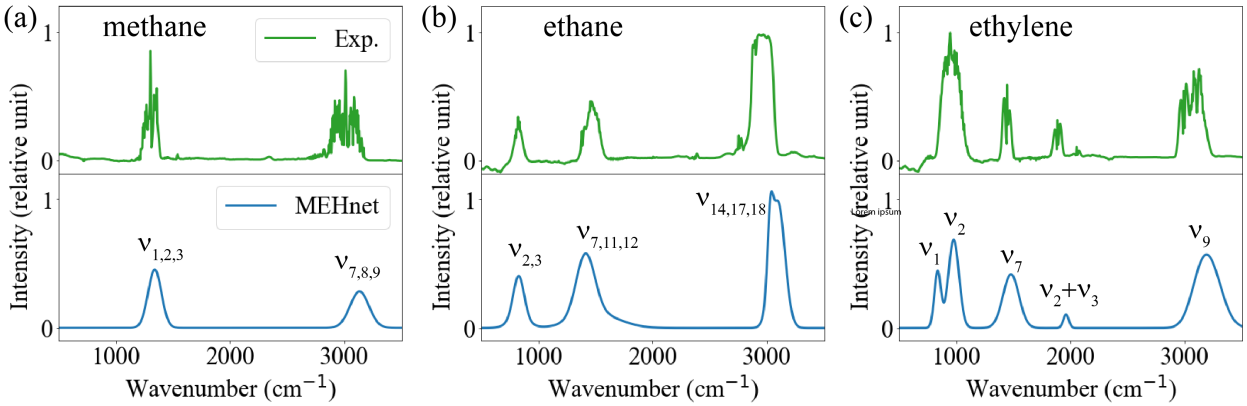
Supplementary Figure 2. The distribution of model prediction accuracy on the testing dataset compared to the B3LYP DFT calculations using the CCSD(T) results as the ground truth. (a-g) Cumulative distribution of prediction errors for the (a) energy, (b) electric dipole moment, (c) electric quadrupole moment, (d) Mulliken atomic charge, (e) Mayer bond order, (f) energy gap (1^{st} excitation energy), and (g) static electric polarizability (a.u. means atomic unit). The blue and orange solid lines represent MEHnet and B3LYP results on the in-domain testing dataset, and the purple and red dashed lines represent MEHnet and B3LYP results on the out-of-domain testing dataset, respectively. We denote the model errors at 50%, 80%, and 95% percentile from the bottom to the top by hollow circles.

the RMSE shown in Fig. 2 in the main text, the error distribution of our MEHnet model and the B3LYP calculations is also shown in Supplementary Figure 2. When calculating the RMSE in Fig. 2c and Table I in the main text, for some of the DFT functionals and other ML methods in comparison, we average part of the data points (rather than all) to save computational costs. Specifically, we use all data points in our dataset for B3LYP and AIQM1; 100 data points per chemical formula for B3PW91, DSD-PBEP86, and PWPB95; and 50 data points per chemical formula for DM21.

Note that the Mulliken charges are sensitive to basis set and the B3LYP RMSE shown in Fig. 2c mainly comes from the basis set error. Therefore, we further calculate the Mulliken atomic charge using the B3LYP/def2-TZVP, a larger basis set that usually gives fairly



Supplementary Figure 3. Calculated Mulliken atomic charge using the B3LYP/def2-TZVP compared with other methods. The green transparent bars are the RMSE of the B3LYP/def2-SVP results shown in the main text Fig. 2c, and the green solid bars are that of the large basis set B3LYP/def2-TZVP results. All other bars have the same meaning as Fig. 2c.



Supplementary Figure 4. Calculated IR spectra of (a) methane, (b) ethane, and (c) ethylene using the MEHnet model. The blue lines are our model predictions and the green lines are exper-imental results from the NIST Chemistry WebBook [3]. Peaks in the IR spectra are assigned to vibrational modes $\{\nu_i\}$, where ν_i means the mode with the ith-lowest vibrational frequency.

converged results for comparison. The calculated RMSE is shown in Supplementary Figure 3. The results confirm that the large RMSE of B3LYP in the main text Fig. 2c mainly comes from the basis set error. Although the B3LYP/def2-TZVP calculations give much smaller RMSEs than the smaller basis B3LYP/def-SVP calculations in Fig. 2c, our model still exhibits a better overall accuracy even if compared with the B3LYP/def2-TZVP results.

Aromatic molecules in the main text Fig. 3 include all molecules with serial numbers dividable by 4 and enthalpy of formation provided in Ref. [4]. Details of these aromatic molecules are listed in Table Supplementary Table 2. We also provided the calculated

Supplementary Table 2. List of serial numbers (Sr. No., defined in Ref. [4]) and thermochemical properties of aromatic molecules in the main text Fig. 3. The name of each molecule is 1: trans-10b,10c-dimethyl-10b,10c-dihydropyrene; 2: anthracene; 3: benzo[c]phenanthrene; 4: 5-ring phenacene, picene; 5: Pyrene; 6: Coronene; 7: 1,4:2,5-[2.2.2.2]cyclophane; 8: 9,9'-bianthryl; 9: *p*-terphenyls; 10: acenaphthene; 11: Aceplaidylene. U and G are the inner energy and Gibbs free energy reference to separated atoms; H_f^{MEHnet} and $H_f^{\mathrm{Exp.}}$ are the enthalpy of formation by the MEHnet predictions and experiments in Ref. [4], respectively; and T1 is the result of T1-diagnostic of coupled cluster calculations.

Mol. index	Sr. No.	U (Hartree)	G (Hartree)	H_f^{MEHnet} (kJ/mol)	$H_f^{\mathrm{Exp.}}$ (kJ/mol)	T1
1	12	-5.9022	-5.5028	362.9	338.8	0.0097
2	20	-4.4302	-4.151	208.5	229.1	0.0102
3	24	-5.617	-5.2615	301.9	291.2	0.01
4	28	-6.8246	-6.3928	340.5	317.3	0.01
5	32	-4.9635	-4.6556	210.6	225.7	0.01
6	36	-7.2236	-6.7812	290.4	302.0	0.0102
7	40	-8.773	-8.1545	458.8	409.5	0.0088
8	44	-8.6805	-8.1234	484.4	454.3	0.0102
9	48	-5.776	-5.4015	289.2	272.9	0.0098
10	64	-3.9225	-3.6712	139.3	156	0.0098
11	68	-4.9042	-4.5976	366.1	362	0.0109

inner energies U and Gibbs free energies G of these molecules as additional thermochemical quantities as predictions of our model.

Supplementary Section 3. INFRARED SPECTRUM

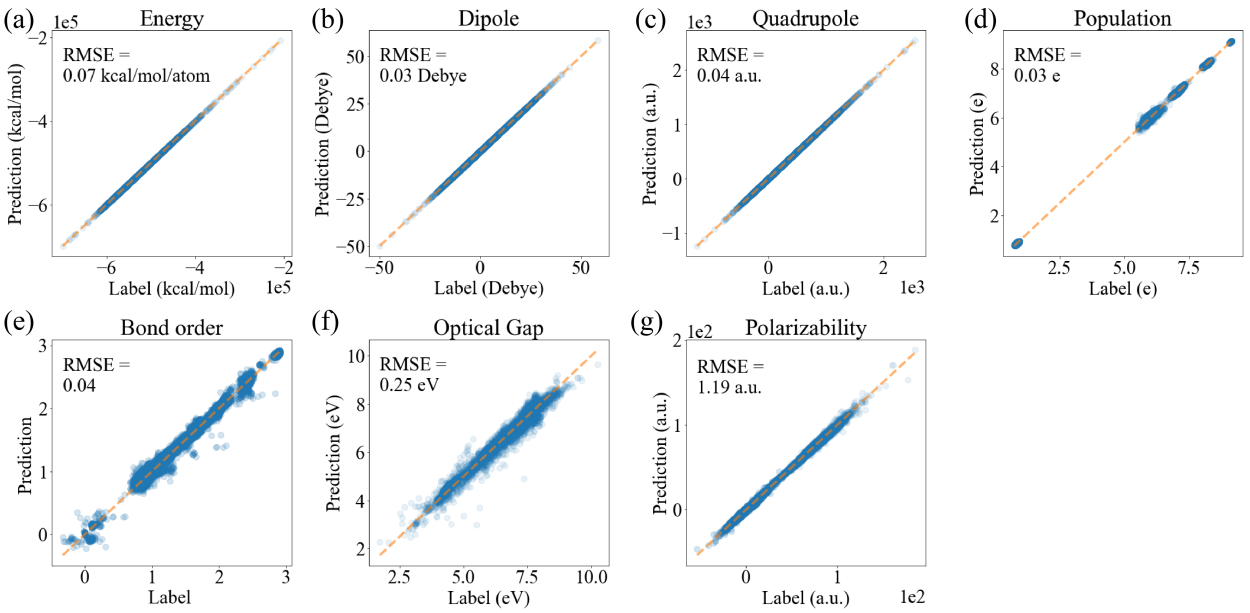
In order to evaluate the infrared spectrum, we first implement a B3LYP hybrid DFT calculation with def2-TZVP basis set to obtain the vibrational modes and frequency of a benzene molecule. Then, we generate atomic configurations displaced from the equilibrium configuration along each vibrational mode by a displacement of -0.1, -0.05, 0.05, and 0.1 Å. Our MEHnet model is used to evaluate the electric dipole moment at each configuration, and the dipole-moment derivative with respect to each normal coordinate is evaluated by linear regression. The infrared band intensity of fundamental bands are then evaluated following the method in Ref. [5].

As the two combination bands at 1800 - 2000 cm^{-1} are mainly contributed by $\nu_{10}+\nu_{17}$ and $\nu_5+\nu_{17}$ [6], we generate atomic configurations displaced from the equilibrium configuration by displacement vectors of $0.1(\vec{e}_i+\vec{e}_j), 0.1(\vec{e}_i-\vec{e}_j), 0.1(-\vec{e}_i+\vec{e}_j)$, and $0.1(-\vec{e}_i-\vec{e}_j)$ Å, where $(\vec{e}_i, \vec{e}_j)$ are the pair of vibrational modes contributing to each combination band. The second-order dipole-moment derivatives with respect to each pair of normal coordinates $\frac{\partial^2 \vec{p}}{\partial Q_i \partial Q_j}$ are then obtained by finite difference method. The leading-order anharmonic constants are also evaluated by finite difference method. These parameters are then used to calculate intensity of the combination bands by Fermi's golden rule. Using the calculated infrared spectrum peak positions and intensity, we add Gaussian broadening to each peak and fit their bandwidth to the experimental spectrum.

The same calculation procedure is applied to three other common molecules, methane, ethane, and ethylene, as shown in Supplementary Figure 4. In ethylene, the combination bind at around 2000 cm^{-1} is mainly contributed by $\nu_2+\nu_3$. Similar to the results of benzene, the MEHnet model predicts both the peak positions and peak intensity well consistent with experimental results.

Supplementary Section 4. QM9 VERSION OF MEHNET

About 10% of molecules (14,000 molecules) are randomly selected from the QM9 dataset and separated into the 10,000 molecules training set and 4,000 molecules testing set. In order to reduce computational costs, we implement CCSD(T) calculations following the idea described in Ref. [7]. Ground-state properties q (energy, dipole, quadrupole, charge, bond

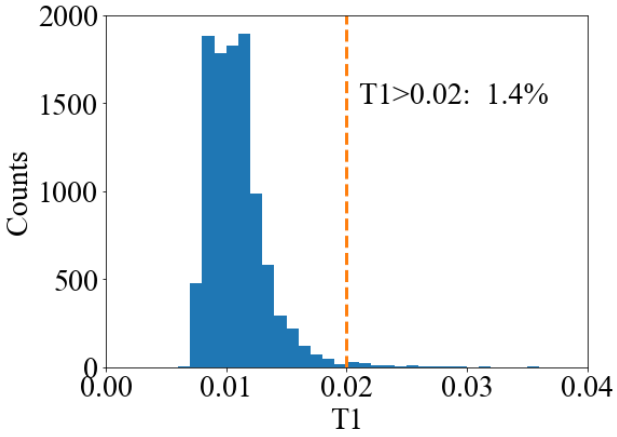


Supplementary Figure 5. Benchmark of the QM9 version of MEHnet on 4,000 molecules randomly sampled from the QM9 dataset for (a) energy, (b) electric dipole moment, (c) electric quadrupole moment, (d) Mulliken population, (e) Mayer bond order, (f) optical gap, and (g) static electric polarizability. The horizontal axis and vertical axis represent the coupled-cluster labels and MEHnet model predictions, respectively.

order) are evaluated by CCSD(T)/cc-pVDZ, DLPNO-CCSD(T)/cc-pVDZ, and DLPNO-CCSD(T)/cc-pVTZ, respectively. The training label is then evaluated as:

$$q_{\text{label}} = q^{\text{CCSD(T)}}_{\text{cc-pVDZ}} - q^{\text{DLPNO-CCSD(T)}}_{\text{cc-pVDZ}} + q^{\text{DLPNO-CCSD(T)}}_{\text{cc-pVTZ}} \tag{S1}$$

The Optical gap and polarizability are evaluated in the same way as the hydrocarbon dataset as described in the main text Methods D. The weight parameters in the training loss function are: $w_V = 0.01$, $w_{\vec{p}} = 0.1$, and other weight parameters are the same as the hydrocarbon version of MEHnet. The dataset is evenly divided into 25 minibatches with 400 data points in each of them, and 150,000 gradient descend steps are implemented on the minibatches alternatively (6,000 steps for each minibatch). The prediction accuracy of the QM9 version of MEHnet is shown in Supplementary Figure 5. We can see that the predictions on the testing dataset are generally consistent with the coupled-cluster labels, confirming that the MEHnet model successfully learns the molecular electronic structures in the QM9 dataset.



Supplementary Figure 6. Histogram of T1 diagnostic results of our coupled-cluster dataset. Most of the data points have T1 values below 0.02, indicating that the dataset does not have strong multireference character.

Supplementary Section 5. ABOUT MULTI-REFERENCE ISSUE

In order to study the multireference issue, we implemented T1 diagnostic to both our CCSD(T) dataset and aromatic molecules in Fig. 3a to get more insight into their multireference character [8]. The evaluated T1 values are shown in Supplementary Figure 6 and Table Supplementary Table 2. In most cases, the T1 values are below 0.02, suggesting that the studied system does not exhibit strong multireference character.

As our work aims to generalize from small to large molecules, including 1-rdm as the output descriptor also involves certain challenges. As some molecules we studied have delocalized orbitals (as shown in the main text Fig. 4a), the 1-rdm, unlike the Fock matrix, can have non-zero off-diagonal terms between atomic orbitals far from each other. Therefore, predicting the 1-rdm requires the neural network architecture to directly capture delocalized features of the whole molecules. In previous works [9, 10], the machine learning models involve all-to-all connection of atoms in the molecules in order to capture the delocalized features of 1-rdm. However, these model architectures also have their limitation: they only applies to molecules with the same number of atoms as the training data. Therefore, in order to generalize from small to large molecules, different model architecture is needed. Constructing appropriate model architectures for systems with strong multireference char-

acters is an intriguing direction to explore, which is left to future work.

---

[1] Sun, Q. *et al.* Recent developments in the PySCF program package. *The Journal of Chemical Physics* **153**, 024109 (2020). URL `https://doi.org/10.1063/5.0006074`.

[2] Dral, P. O. *et al.* Mlatom 3: A platform for machine learning-enhanced computational chemistry simulations and workflows. *Journal of Chemical Theory and Computation* (2024).

[3] Linstrom, P. & W.G. Mallard, E. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, Ch. Infrared Spectra (National Institute of Standards and Technology, Gaithersburg MD, 20899, 2024).

[4] Slayden, S. W. & Liebman, J. F. The energetics of aromatic hydrocarbons: an experimental thermochemical perspective. *Chemical reviews* **101**, 1541–1566 (2001).

[5] Hess Jr, B. A., Schaad, L. J., Carsky, P. & Zahradnik, R. Ab initio calculations of vibrational spectra and their use in the identification of unusual molecules. *Chemical Reviews* **86**, 709–730 (1986).

[6] Maslen, P. E., Handy, N. C., Amos, R. D. & Jayatilaka, D. Higher analytic derivatives. iv. anharmonic effects in the benzene spectrum. *The Journal of chemical physics* **97**, 4233–4254 (1992).

[7] Smith, J. S. *et al.* Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature communications* **10**, 2903 (2019).

[8] Lee, T. J. & Taylor, P. R. A diagnostic for determining the quality of single-reference electron correlation methods. *International Journal of Quantum Chemistry* **36**, 199–207 (1989).

[9] Hazra, S., Patil, U. & Sanvito, S. Predicting the one-particle density matrix with machine learning. *Journal of Chemical Theory and Computation* (2024).

[10] Shao, X., Paetow, L., Tuckerman, M. E. & Pavanello, M. Machine learning electronic structure methods based on the one-electron reduced density matrix. *Nature communications* **14**, 6281 (2023).

nature portfolio

Peer Review File

Approaching coupled-cluster accuracy for molecular electronic structures with multi-task learning

Corresponding Author: Professor Ju Li

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

A version of this paper was originally rejected for publication by Nature Computational Science, however that decision was reconsidered after appeal by the authors.

Version 0:

Decision Letter:

Dear Professor Li,

Thank you for submitting "Multi-task learning for molecular electronic structure approaching coupled-cluster accuracy" to Nature Computational Science. Regretfully, we cannot offer to publish it in its current form.

Among the considerations that arise at this stage are the manuscript's likely interest to a broad range of researchers in computational science, the pressure on space for the various disciplines covered by Nature Computational Science, and the likelihood that a manuscript would seem of great topical interest to those working in the same or related areas of computational science. We do not doubt the technical quality of your work or that it will be of interest to others working in this area of research. However, I regret that we are unable to conclude that the paper provides the sort of substantial practical or conceptual advance that would be of immediate interest to a broad readership of researchers in computational science.

Should future experimental data allow you to:

1) Perform quantitative comparisons against existing methods in the field, such as https://www.nature.com/articles/s41467-023-41953-9 and https://www.nature.com/articles/s41524-023-01070-z

2) Provide quantitative comparisons to higher level of theory functionals (in addition to the comparisons to B3LYP)

then we would be happy to look at a revised manuscript and reassess the work to decide whether or not we will send it out to peer review (unless, of course, something similar has by then been accepted at Nature Computational Science or appeared elsewhere). This includes submission or publication of a portion of this work somewhere else. In the case of eventual publication, the received date would be that of the revised paper.

If you are interested in submitting a suitably revised manuscript in the future or if you have any questions, please contact me. In case you are not interested in submitting a revised version, you may transfer your manuscript to another journal in the Nature Portfolio using the link I provide at the end of this email.

Thank you for your interest in Nature Computational Science. I am sorry that on this occasion we cannot be more positive.

Best regards,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

Although we cannot publish your paper, it may be appropriate for another journal in the Nature Portfolio. If you wish to explore the journals and transfer your manuscript please use our manuscript transfer portal.

You will not have to re-supply manuscript metadata and files, unless you wish to make modifications. For more information, please see our manuscript transfer FAQ page.

For Nature Portfolio general information and news for authors, see https://www.nature.com/nature-research/for-authors.

Version 1:

Decision Letter:

** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors. **

Dear Professor Li,

Your manuscript "Multi-task learning for molecular electronic structure approaching coupled-cluster accuracy" has now been seen by 3 referees, whose comments are appended below. You will see that while they find your work of interest, they have raised points that need to be addressed before we can make a decision on publication.

The referees' reports seem to be quite clear. Naturally, we will need you to address *all* of the points raised.

While we ask you to address all of the points raised, the following points need to be substantially worked on:

1) Please clarify the generalizability of the proposed approach as requested by reviewers
2) Please add additional quantitative benchmarks as requested by reviewers
3) Please add additional methodological details as requested by reviewers

Please use the following link to submit your revised manuscript and a point-by-point response to the referees' comments (which should be in a separate document to any cover letter):

Link Redacted

** This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first. **

To aid in the review process, we would appreciate it if you could also provide a copy of your manuscript files that indicates your revisions by making use of Track Changes or similar mark-up tools. Please also ensure that all correspondence is marked with your Nature Computational Science reference number in the subject line.

In addition, please make sure to upload a Word Document or LaTeX version of your text, to assist us in the editorial stage.

To improve transparency in authorship, we request that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We hope to receive your revised paper within three weeks. If you cannot send it within this time, please let us know.

We look forward to hearing from you soon.

Best regards,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

Reviewers comments:

Reviewer #1 (Remarks to the Author):

The authors use multi-task learning to learn the non-local exchange interaction after calculating the Lowdin-symmetrized KS Hamiltonian (local DFT Hamiltonian). They predict a variety of ground state properties: energy, electric dipole, quadrupole moments, Mulliken charge, Mayer bond order, first excitation energy, and static electric polarizability, successfully demonstrating linear scaling up to 70 electrons, outperforming CCSD(T), hybrid, and double hybrid DFT for more than 400 atoms.

While the effort seeks to address a critical challenge of great interest to the AI for Science community, it is my opinion that the developed workflow and demonstrated benchmarks are too limited in scope to be suitable for Nature Computational Science.

On workflow: No justification for which 85 hydrocarbons are collected from PubChem; Furthermore, there is comparison against only one hybrid and one double-hybrid to represent those classes of DFT. However, B3LYP has known failure modes for hydrocarbons [1-2], and there have been several updates to DSD-PBE86 in the last 10 years or so. If the authors would like to use these functionals, more discussion with other density functionals is needed.

Additionally, the prediction of Mulliken charges are known to be sensitive to basis set choice [4]. Can the authors comment on how the choice of B3LYP/cc-pVDZ affects their results?

On benchmarks: Limited to hydrocarbons only, and it is not clear how diverse the carbon bonds really are. Additionally, the enthalpy of formation is the only type of thermochemical data computed, and the IR spectra for only benzene is computed.

Other comments:
1. How many steps was the TeaNet potential run for? How were structures selected? Was additional geometric relaxation needed before the CCSD(T) calculation?
2. What is the reason for lack of data in Table 1 for AIQM1?
3. Since all systems are closed-shell and calculated with spin-restricted DFT, can the statement, "One can see that Eg is larger for oligomers...due to the quantum confinement effect... " still be made?
4. I know this is not a classical simulation, but it would be helpful to show the generality to hydrocarbons in the context of what has been done for classical force fields like OPLS-AA.
5. "Relatively small NN with only 511,589 parameters". Please cite some sources to provide context.

If the authors can broaden the workflow and benchmarks, perhaps via the directions suggested above, the manuscript will garner much greater excitement for the community.

[1] A. Karton. Journal of Computational Chemistry, 2017, 38, 370–382.
[2] J. Tirado-Rives and W. L. Jorgensen, Journal of Chemical Theory and Computation 2008 4 (2), 297-306
[3] J. M. L. Martin and G. Santra. Israel Journal of Chemistry, 2020, 60, 787.
[4] M. Jablonski. J. Phys. Chem. A 2010, 114, 5, 2240–2244.

Reviewer #2 (Remarks to the Author):

This is a very interesting paper. Authors applied machine learning technique to build DFT like Hamiltonian using data from coupled-cluster calculations as training set. They went further and computed physical quantities in addition to the total energy. Their effort allows low computational cost and high accuracy prediction/calculation for molecular systems. I recommend this paper for publication in Nature Computational Science after the following questions being addressed --

1. Can this approach be applied to extended systems? If so, please explain if there is a plan to do so. If not, what is bottle neck?
2. Is this scheme general enough such that one can also use other high-level quantum chemistry methods (e.g. full-CI) to train the network?

3. A specific question about the supplement materials: Towards end of the left column, it says "in Eq. (S2), the summation over m goes through all states except n. But in equation S2, there is no summation over m and n. Therefore the subsequent argument about symmetry does not work. Also in S3, there is no epsilon_m or n.

Reviewer #2 (Remarks on code availability):

I only downloaded the code, no time to run or test.

Reviewer #3 (Remarks to the Author):

The manuscript "Multi-task learning for molecular electronic structure approaching coupled cluster accuracy" by Tang et al is an interesting and much needed approach for machine learning approaches in electronic structure theory. Instead of targeting energy as the output descriptor which is a low dimensional and often degenerate observable, the authors achieve multi-task learning by targeting the Fock operator. As a result one can derive any observable that is well defined within the single reference mean field framework. Furthermore, the high accuracy learning is achieved for smaller molecules with less than ~100 electrons while the prediction space is much larger and can ideally be as large as required within the hydrocarbon framework of the chemical space. Therefore, I find the goal and achievement of the approach to be quite laudable.

However, there are significant improvement that is required in the manuscript to improve its readability.
1. The major problem I have with the paper is that due to condensed format in which it is written, it is very difficult to understand the workflow in detail. The methodology section needs to be more descriptive for the readability and general

reproducibility of the work that is mentioned.
2. I am guessing the workflow is applicable for only hydrocarbons. This is what I surmise from the figures. However, I do not see where that is explicitly written. Page 5 (Model Performance and Applicability) should mention which class of small to large molecules are considered. If it is hydrocarbons, are they all possible hydrocarbons? How is even this chemical space created? Are all metastable hydrocarbons included? This is crucial to describe because the nature of molecules are quite diverse in this chemical space and the exact degree of transferability of learning should be understood.
3. Fig.1c also describes some of the data set creation. However, it is a bit unclear what the orange and blue dots refer to. Are these MD and CCSD(T) calculations? But that doesn't quite make sense. What is the meaning behind the different size of these dots? No of calculations? These are totally not mentioned in the paper.
4. In Fig. 2c, the RMSE of each property is shown. I am guessing this RMSE is with respect to CCSD(T) numbers. In such a case, it will be interesting to see the prediction efficiency (RMSE) in the training domain versus the generalization domain. Furthermore, it can be easily observed that different properties show different error ranges. The reason for this should be discussed. Also, static electric polarizability shows a trend of increasing error with increasing system size. Why?
5. Also a question that constantly bothers me is that the intended accuracy bar is set for CCSD(T) which may or may not be relevant for different classes of molecules (even within hydrocarbons). There are polyaromatic hydrocarbons which are more multireference in nature. Are they even included in the data set? Maybe some of this aspect can be seen in Fig. 3a. The experimental versus EGNN uncertainty is quite different for certain molecules.

In short, I find the idea of the paper quite intriguing but due to brevity and condensed writing style, much of the details is quite obscure. One further aspect that needs some discussion is why use Fock operator as the output descriptor and not include something like a one particle density matrix? It would still have the same order of difficulty in the problem while resolving the issue of multireference nature of the problem. An interesting example (albeit in much smaller class of systems) can be seen in J. Chem. Theory Comput. 2024, 20, 11, 4569–4578.

Reviewer #3 (Remarks on figshare data availability):

I have already included that in the report.

Version 2:

Decision Letter:

Our ref: NATCOMPUTSCI-24-0907B

6th November 2024

Dear Dr. Li,

Thank you for submitting your revised manuscript "Multi-task learning for molecular electronic structure approaching coupled-cluster accuracy" (NATCOMPUTSCI-24-0907B). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Computational Science, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

TRANSPARENT PEER REVIEW
Nature Computational Science offers a transparent peer review option for original research manuscripts. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please remember to choose, using the manuscript system, whether or not you want to participate in transparent peer review.**
Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

Thank you again for your interest in Nature Computational Science. Please do not hesitate to contact me if you have any questions.

Sincerely,

Kaitlin McCardle, PhD

Senior Editor
Nature Computational Science

ORCID
IMPORTANT: Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance: https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

Reviewer #1 (Remarks to the Author):

The authors have addressed all comments and substantially revised the manuscript and SI to a satisfactory level.

The updated evaluation on QM9, comparisons to B3PW91 and PWPB95, tests for Mulliken charges with B3LYP/def2-TZP, calculated IR spectra, have significantly expanded the generality of the work. Furthermore, the updates enabled by Reviewer 2's question about extended systems adds some very interesting new perspectives to the work.

All of my concerns have been comprehensively addressed, so I can now recommend this article for publication.

Reviewer #1 (Remarks on code availability):

The code is well-documented and the demo is clear and user-friendly. Note that the large dataset can and should also be uploaded to open databases like Materials Cloud.

Reviewer #3 (Remarks to the Author):

The authors have answered all the questions and have corrected the paper satisfactorily. Therefore, I believe the paper can be accepted.

Reviewer #3 (Remarks on code availability):

The authors have answered all the questions and have corrected the paper satisfactorily. Therefore, I believe the paper can be accepted.

Reviewer #3 (Remarks on figshare data availability):

All the data is available and satisfactory.

Version 3:

Decision Letter:

21st November 2024

Dear Dr. Li,

I am delighted to tell you that your manuscript NATCOMPUTSCI-24-0907C has been accepted for publication in Nature Computational Science.

We will be publishing your paper on an accelerated schedule. **Please carefully review the details below and contact us immediately at computationalscience@nature.com if you have any travel plans or other conflicts that may make you unable to respond to us for the next 5-7 days.**

In approximately 2 business days you will receive a link to choose the appropriate publishing options for your paper and complete the appropriate grant of rights necessary to publish your work. As it is vital that this process not be delayed, we strongly encourage you to whitelist the email address do-not-reply@springernature.com to ensure that this message is received.

You will receive a link to your electronic proof via email with a request to make any necessary corrections as soon as possible. You will find that we have made minor changes to enhance the clarity of the text and to ensure that your paper conforms to the journal's style so we ask that you review these proofs carefully to ensure that we have not inadvertently introduced errors or altered the sense of your text in any way.

Please return your proof within 24 hours of receiving it. If you have any questions about your proofs or anticipate any delays please contact rjsproduction@springernature.com immediately.

Once a publication date is set for your paper, the Springer Nature press office will be in touch with the full embargo details. We request that you do not send out your own publicity or contact any journalists until you hear from us that the paper has a confirmed publication date.

If you would like to inform your Public Relations or Press Office about your paper, we suggest that you do so immediately to allow them as much time as possible to prepare an appropriate press release and organize publicity if they choose to do so. Please include your manuscript tracking number NATCOMPUTSCI-24-0907C and the name of the journal, which they will need if they contact our press office.

Please note that Nature Computational Science is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com.

If you have not already done so, we strongly recommend that you upload the step-by-step protocols used in this manuscript to the Protocol Exchange. Protocol Exchange is an open online resource that allows researchers to share their detailed experimental know-how. All uploaded protocols are made freely available, assigned DOIs for ease of citation and fully searchable through nature.com. Protocols can be linked to any publications in which they are used and will be linked to from your article. You can also establish a dedicated page to collect all your lab Protocols. By uploading your Protocols to Protocol Exchange, you are enabling researchers to more readily reproduce or adapt the methodology you use, as well as increasing the visibility of your protocols and papers. Upload your Protocols at www.nature.com/protocolexchange/. Further information can be found at www.nature.com/protocolexchange/about .

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

Sincerely,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

P.S. Click here if you would like to recommend Nature Computational Science to your librarian - this will link directly to the Recommend page.

http://www.nature.com/subscriptions/recommend.html#forms

** Visit the Springer Nature Editorial and Publishing website at www.springernature.com/editorial-and-publishing-jobs for more information about our career opportunities. If you have any questions please click here.**

Response letter to reviewers

Reviewer #1 (Remarks to the Author):

The authors use multi-task learning to learn the non-local exchange interaction after calculating the Lowdin-symmetrized KS Hamiltonian (local DFT Hamiltonian). They predict a variety of ground state properties: energy, electric dipole, quadrupole moments, Mulliken charge, Mayer bond order, first excitation energy, and static electric polarizability, successfully demonstrating linear scaling up to 70 electrons, outperforming CCSD(T), hybrid, and double hybrid DFT for more than 400 atoms. While the effort seeks to address a critical challenge of great interest to the AI for Science community, it is my opinion that the developed workflow and demonstrated benchmarks are too limited in scope to be suitable for Nature Computational Science.

We thank the reviewer for the critical review and for pointing out the limitation in the data generation workflow and benchmarks. Based on the insightful comments below, we made significant efforts in generalizing the workflow and expanding the scope of demonstrated benchmarks. We now have an *updated version* of the model that supports a diverse range of organic molecules consisted of not only H and C as in the previous version, but also N, O, and F. We benchmarked the model on randomly sampled structures from a main-stream quantum chemistry dataset (QM9), giving consistently high prediction accuracy. As suggested by the reviewer, we also provided more comprehensive benchmark in terms of functionals, basis set, and predicted quantities. We added a section in the main text about the QM9 version of our model, and leave other extensive discussions in the supporting information (SI) to avoid making the main text too crowded.. We believe these updates can resolve the concerns of the reviewer, particularly about the scope and generality of our work. Please see the point-to-point response below:

On workflow: No justification for which 85 hydrocarbons are collected from PubChem;

Thank you for this insightful comment. In the revised manuscript, we include justifications on the selection of the 85 hydrocarbons from the PubChem database on paragraph 2, page 5 of the main text and SI section S2. The molecules are selected based on the principle of structural diversity, covering different classes of hydrocarbons. Details about the structural diversity of our collected hydrocarbons are summarized in the revised supporting information Fig. S1, which is reproduced below:

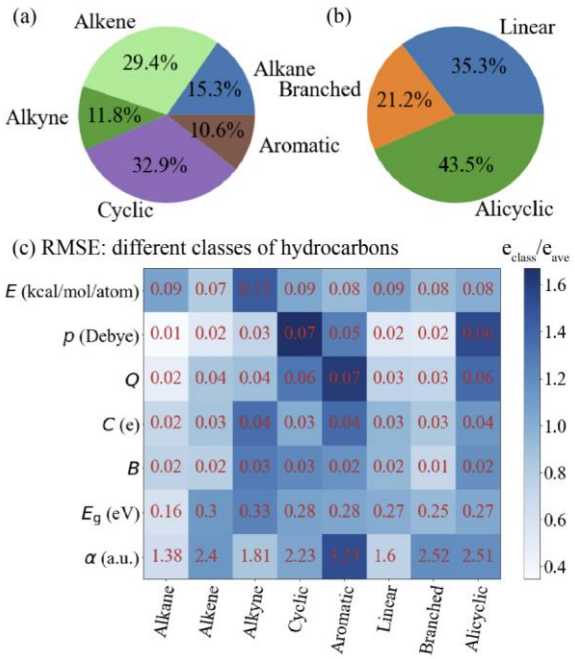


FIG. S1. Composition of the dataset in different (a) hydrocarbon molecule classes (alkane, alkene, alkyne, cyclic, and aromatic) and (b) structural classes (linear, branched, and alicyclic). Percentage of the number of molecules in the dataset is shown in the plot. (c) RMSE of the MteHnet model predictions on different classes of molecules in the testing dataset. The red numbers denote absolute values of the RMSE, and the color reflects the ratio of the RMSE on a specific class of molecule to the average RMSE on all molecules in the testing dataset.

Hydrocarbon molecules can be classified into 4 classes [Fig. S1(a)]: saturated hydrocarbon (Alkane), unsaturated hydrocarbon (Alkene and Alkyne), Alicyclic hydrocarbon, and Aromatic hydrocarbon. On the other hand, the molecule structure can be categorized into 3 classes [Fig. S1(b)]: linear structure, branched structure, and cyclic (either alicyclic or aromatic) structure. Our training dataset is diverse as it covers a significant number of molecules in each class. We further examine the testing errors of our model in different classes of molecules in the dataset, which are shown as numbers in Fig. S1(c). We can see for each quantity (labelled on y-axis), the errors are generally close among all classes of molecules (labelled on x-axis).. For each quantity, we also calculated how the error for each class of molecules deviates from the average value among all classes, which are marked with the color. One can see that the deviation is no more than 60%, showing that our trained model has consistently good prediction accuracy for various classes of hydrocarbons. This further validates that the molecules we selected for training provides sufficient and balanced training data so that the model learns the electronic structure of different classes of hydrocarbons.

Besides, we would like to mention that we intentionally selected 85 relatively *small* molecules from the PubChem database, which can reduce the cost for model training. The model, however, can be applied to large molecules beyond the training dataset (Figure 1c in the main text), showing the significant generalizability of our model.

Furthermore, there is comparison against only one hybrid and one double-hybrid to represent those classes of DFT. However, B3LYP has known failure modes for hydrocarbons [1-2], and there have been several updates to DSD-PBE86 in the last 10 years or so. If the authors would like to use these functionals, more discussion with other density functionals is needed.

Thanks. We agree that using only the B3LYP and DSD-PBEP86 functionals may not be sufficiently representative for hybrid and double hybrid functionals. In order to more effectively represent the hybrid and double hybrid class of functionals, we choose another two representative hybrid and double hybrid functionals to benchmark on the hydrocarbon dataset. Specifically, we choose the B3PW91 hybrid functional and PWPB95 double hybrid functional, which are referred as the best performed functionals in the Ref. [1] provided by the reviewer, [*Phys. Chem. Chem. Phys.* 13.14 (2011): 6670-6688], and orca manual [ORCA Input Library - DFT calculations (google.com)]. The comparison results are listed in Table I (reproduced below) in the revised main text with discussions on paragraph 1-2, page 6:

TABLE I. Benchmark of MteHnet model's RMSE in predicting different quantum chemical properties on the in-domain (ID) testing dataset and out-of-domain (OOD) validation dataset with respect to the coupled cluster calculations. The numbers in the table are ID/OOD RMSD. Other DFT and machine learning methods are compared. We leave some of the spaces blank when the method does not directly output the quantity for fair comparison.

RMSE	(ID/OOD)	Hybrid		Double Hybrid		ML		
	Unit	B3LYP	B3PW91	D3D-PBEP86	PWPB95	DM21	AIQM1	MteHnet (ours)
Energy (/atom)	kcal/mol	2.20/2.41	2.03/2.73	0.94/1.20	1.64/1.98	0.22/0.11	0.13/0.06	**0.11/0.10**
Dipole	Debye	0.06/0.06	0.06/0.04	0.03/0.03	0.07/0.05	0.04/0.04	–	**0.03/0.04**
Quadrupole	ea_0^2	0.12/0.21	0.32/0.51	0.11/0.18	0.10/0.14	–	–	**0.03/0.12**
Atomic Charge	e	0.19/0.20	0.16/0.16	0.04/0.05	0.05/0.05	0.05/0.04	–	**0.04/0.03**
Bond Order	–	0.05/0.03	0.06/0.04	0.04/0.02	0.06/0.03	0.06/0.03	–	**0.02/0.02**
Bandgap	eV	0.59/0.63	0.65/0.54	3.71/3.26	2.19/1.98	1.71/1.47	–	**0.26/0.31**
Polarizability	a.u.	2.22/4.32	2.53/4.72	4.74/8.05	–	–	–	**1.85/3.91**

From the comparison, we can see that although the newly added hybrid and double hybrid functionals exhibit better accuracy in certain properties than the B3LYP and D3D-PBEP86 functional, our multitask machine learning method still shows better overall accuracy. Therefore, we consider the main conclusion we made on this comparison remains valid after adding these comparisons.

As suggested by the reviewer, we also include discussions on the limitation of the B3LYP functional in the manuscript on paragraph 1, page 6.

Additionally, the prediction of Mulliken charges are known to be sensitive to basis set choice [4]. Can the authors comment on how the choice of B3LYP/cc-pVDZ affects their results?

We agree that the Mulliken charges are sensitive to basis set, and that the B3LYP RMSE shown in Fig. 2c of the main text can (partially) come from the basis set error. Therefore, we further calculate the Mulliken atomic charge using the B3LYP/def2-TZVP, a larger basis set that usually gives fairly converged results. The calculated RMSE is shown below:

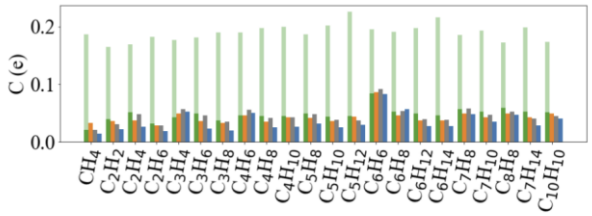


FIG. S3. Calculated Mulliken atomic charge using the B3LYP/def2-TZVP compared with other methods. The green transparent bars are the RMSE of the B3LYP/def2-SVP results shown in the main text Fig. 2c, and the green solid bars are that of the large-basis B3LYP/def2-TZVP results. All other bars has the same meaning as Fig. 2c.

where the green transparent bars are the RMSE of the small-basis B3LYP results and the green solid bars (overlapped on the green transparent bars) are that of the large-basis B3LYP results. The results confirm that, as the reviewer suggested, the large RMSE of B3LYP in Fig. 2c mainly comes from the basis set error. Despite this, our model still exhibits a better overall accuracy even if we compare with the B3LYP/def2-TZVP results, so the main conclusion we made is still valid. We can also see that our ML model can achieve better overall accuracy with a smaller cc-pVDZ basis set ($N^H_{\text{basis}} = 5, N^C_{\text{basis}} = 14$) compared to B3LYP with the larger def2-TZVP basis set (($N^H_{\text{basis}} = 14, N^C_{\text{basis}} = 30$)).

In the revised manuscript, we clarify the basis set error of the B3LYP functional in paragraph 2, page 6, and include the above plot as Fig. S3 with discussions in SI section S3.

On benchmarks: Limited to hydrocarbons only, and it is not clear how diverse the carbon bonds really are. Additionally, the enthalpy of formation is the only type of thermochemical data computed, and the IR spectra for only benzene is computed.

Thanks. In order to make the benchmarks more general, we include a series of benchmarks according to the suggestions of the reviewer, as shown below:

1. To include molecules beyond hydrocarbons, we applied our MtElect workflow to the QM9 molecule database [https://paperswithcode.com/dataset/qm9], which includes various molecules consisted of 5 elements: H, C, N, O, F. The model is trained on 10,000 molecules randomly sampled from the QM9 database and tested on 4,000 other randomly sampled molecules. The model performance is shown in the main text Table II and SI Fig. S5 (reproduced below):

TABLE II. RMSE of the QM9 version of MtElect model on the testing dataset of 4,000 randomly sampled configurations in the QM9 dataset.

Property	E/atom	p	Q	C	B	E_g	α
Unit	kcal/mol	Debye	a.u.	e	–	eV	a.u.
RMSE	0.07	0.03	0.04	0.03	0.04	0.25	1.19

where we can see that out MtElect model gives consistently high prediction accuracy for all properties for molecules in the QM9 dataset. We believe these results suggest that although we mainly focus on the hydrocarbons in this paper, our machine learning method is applicable to a much broader chemical space. On paragraph 2, page 5 of the main text, we also included clarification on the diversity of the carbon-carbon bonds (Fig. S1, see also our replies to the first comment).

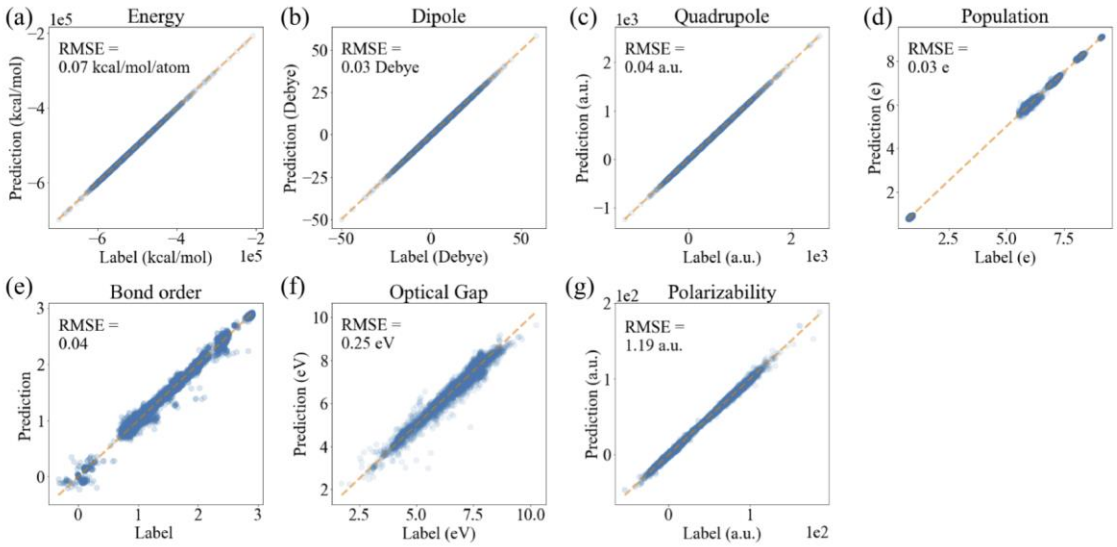


FIG. S5. Benchmark of the QM9 version of MteHnet on 4,000 molecules randomly sampled from the QM9 dataset for (a) energy, (b) electric dipole moment, (c) electric quadrupole moment, (d) Mulliken population, (e) Mayer bond order, (f) optical gap, and (g) static electric polarizability. The horizontal axis and vertical axis represent the coupled-cluster labels and MteHnet model predictions, respectively.

2. Besides Benzene, we calculated IR spectrum for another three common hydrocarbon molecules, as shown in Fig. S4:

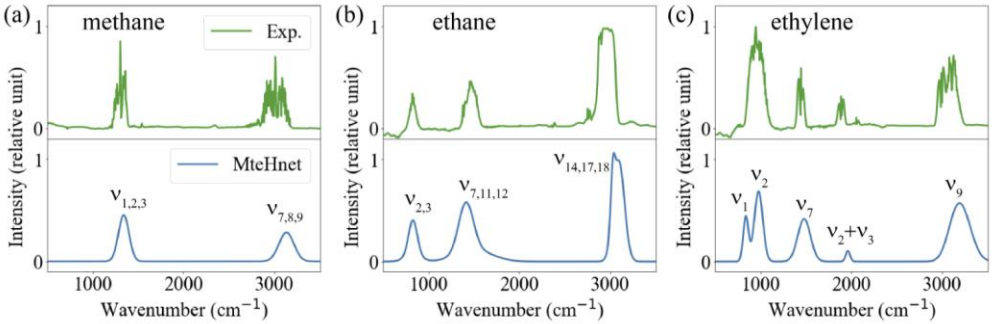


FIG. S4. Calculated IR spectra of (a) methane, (b) ethane, and (c) ethylene using the MteHnet model. The blue lines are our model predictions and the green lines are experimental results from the NIST Chemistry WebBook [3]. Peaks in the IR spectra are assigned to vibrational modes ν_i, the ith-lowest vibrational frequency.

The results confirms that in methane, ethane, and ethylene, the MtElect calculations also give reasonably good predictions in both the IR peak positions and peak intensity compared with experimental results. The consistency supplements to the result of benzene shown in Fig. 3b of the main text, further validating the prediction accuracy of our method on IR spectra. We added relevant discussions to the new results to paragraph 2, page 7 of the main text, and SI section 4.

3. Besides enthalpy of formation, we calculated two more thermochemical quantities, the inner energy (U) and Gibbs free energy (G) for the same set of molecules as those in in Fig. 3a of the main text. The results are added into SI table II. Discussions on the calculated results are also included in SI section S3.

In summary, we included several more benchmarks, and we believe the included results make the benchmark in the manuscript more thorough and general.

TABLE II. List of serial numbers (Sr. No., defined in Ref. [4]) and thermochemical properties of aromatic molecules in the main text Fig. 3. The name of each molecule is 1: trans-10b,10c-dimethyl-10b,10c-dihydropyrene; 2: anthracene; 3: benzo[c]phenanthrene; 4: 5-ring phenacene, picene; 5: Pyrene; 6: Coronene; 7: 1,4:2,5-[2.2.2.2]cyclophane; 8: 9,9'-bianthryl; 9: *p*-terphenyls; 10: acenaphthene; 11: Aceplaidylene.

Mol. index	Sr. No.	U (Eh)	G (Eh)	H_f^{MteHnet} (kJ/mol)	$H_f^{\mathrm{Exp.}}$ (kJ/mol)	T1
1	12	-5.9022	-5.5028	362.9	338.8	0.0097
2	20	-4.4302	-4.151	208.5	229.1	0.0102
3	24	-5.617	-5.2615	301.9	291.2	0.01
4	28	-6.8246	-6.3928	340.5	317.3	0.01
5	32	-4.9635	-4.6556	210.6	225.7	0.01
6	36	-7.2236	-6.7812	290.4	302.0	0.0102
7	40	-8.773	-8.1545	458.8	409.5	0.0088
8	44	-8.6805	-8.1234	484.4	454.3	0.0102
9	48	-5.776	-5.4015	289.2	272.9	0.0098
10	64	-3.9225	-3.6712	139.3	156	0.0098
11	68	-4.9042	-4.5976	366.1	362	0.0109

Other comments:

1. How many steps was the TeaNet potential run for? How were structures selected? Was additional geometric relaxation needed before the CCSD(T) calculation?

The TeaNet potential runs for about 20,000 steps for each molecule/conformer and 100,000 steps for each chemical formula (C_xH_y, in our dataset, each chemical formula contains 5 different molecules/conformers in most cases). Structures are selected by taking 1 structure per 200 molecular dynamics steps. As we aim to include out-of-equilibrium configurations, we do not perform geometric relaxation. In other words, our model applies to structures with and without geometric relaxation. We added clarifications of these points in Methods D.

2. What is the reason for lack of data in Table 1 for AIQM1?

Thanks. This is because AIQM1 model applies machine learning corrections only to the ground-state energy. Other electron-related properties, such as electric dipole, are not modified by AIQM1, so they are left the same as the ODM2 semiempirical method [J. Chem. Theory Comput.15, 1743 (2019)]. Therefore, we think it is not a fair comparison for these semi-classical quantities without ML correction to other methods with ML corrections. Clarification is added to the caption of Table 1.

3. Since all systems are closed-shell and calculated with spin-restricted DFT, can the statement, "One can see that Eg is larger for oligomers...due to the quantum confinement effect... " still be made?

We agree that referring the increasing bandgap for shorter chain length as quantum confinement effect can lead to confusion, as we are not actually calculating the open-shell system with confined electron/hole. Therefore, we revised the sentence as:

“One can see that E_g is larger for oligomers with shorter chain length and converges to a smaller value for long polymer. This is in analogy to the size effect on e.g., the energy gap of quantum dots”

4. I know this is not a classical simulation, but it would be helpful to show the generality to hydrocarbons in the context of what has been done for classical force fields like OPLS-AA.

Thanks. We are not very sure about what "what has been done for classical force fields like OPLS-AA" refers to. We guess it refers to more benchmarks of the model on different class and/or size of hydrocarbons. If this is the case, we think our responses to the previous comments (Fig. S1, S3, S4, S5 in the SI, and Table I and Table II in the main text) provide more comprehensive benchmarks that validate the generality of our method. If the reviewer believes more benchmark is necessary, we are open to include more benchmark calculations.

We notice that the OPLS-AA force field was also used for molecular dynamics (MD) simulation of hydrocarbon liquid and liquid-gas phase transformations [JPCB 105.28 (2001): 6474-6487; JCTC 8.4 (2012): 1459-1470]. Such large-scale MD simulation is, however, beyond the scope of this work. Although our method is much faster than CCSD(T) and hybrid functional DFT, it is certainly more computationally involved than a classical force field. In principle, it is possible to calculate hydrocarbon liquids with a large supercell using our method, but this is not a trivial work. At current stage, our work focuses more on electronic structure. One can also use OPLS-AA or machine learning potentials to do the MD simulations and then use our model to evaluate electronic properties on structures sampled from the obtained MD trajectories.

5. "Relatively small NN with only 511,589 parameters". Please cite some sources to provide context.

Thanks. We added two representative references that use ML methods to fit the electronic properties of molecules. One uses 93,000,000 parameters [Nat. Commun. 10.1 (2019): 5024] and the other uses 17,000,000 parameters [NeurIPS 34 (2021): 14434-14447].

If the authors can broaden the workflow and benchmarks, perhaps via the directions suggested above, the manuscript will garner much greater excitement for the community.

[1] A. Karton. Journal of Computational Chemistry, 2017, 38, 370–382.

[2] J. Tirado-Rives and W. L. Jorgensen, Journal of Chemical Theory and Computation 2008 4 (2), 297-306

[3] J. M. L. Martin and G. Santra. Israel Journal of Chemistry, 2020, 60, 787.

[4] M. Jablonski. J. Phys. Chem. A 2010, 114, 5, 2240–2244.

Reviewer #2 (Remarks to the Author):

This is a very interesting paper. Authors applied machine learning technique to build DFT like Hamiltonian using data from coupled-cluster calculations as training set. They went further and computed physical quantities in addition to the total energy. Their effort allows low computational cost and high accuracy prediction/calculation for molecular systems. I recommend this paper for publication in Nature Computational Science after the following questions being addressed –

We thank the reviewer for the encouraging comments. Please see our point-to-point response below:

1. Can this approach be applied to extended systems? If so, please explain if there is a plan to do so. If not, what is bottle neck?

We thank the reviewer for the helpful suggestion. In principle, the approach can be applied to extended systems, and we indeed plan to work along this direction. We consider the main challenge in this direction is that our training data generation method CCSD(T) is not directly applicable to extended systems. Despite this, it is still possible to apply our model to extended systems. We included the following discussion on paragraph 6, page 8 of the main text:

"In principle, our approach can also be applied to extended systems, where the periodic boundary condition (PBC) is applied to Eq. (3). The band structure and Bloch wavefunctions can then be obtained by solving the eigenvalue problem for each wave vector $\vec{k}$ after a Fourier transformation from the real space to the reciprocal space. Although the CCSD(T) method for training data generation does not directly support PBC, one can use CCSD(T) calculations for finite atom clusters (i.e., a truncated and possibly passivated supercell) to train the model and subsequently use the model to predict the properties of extended systems. Alternatively, the training data of extended systems can be generated by high-accuracy methods other than CCSD(T), such as double-hybrid DFT, that allows periodic boundary conditions."

2. Is this scheme general enough such that one can also use other high-level quantum chemistry methods (e.g. full-CI) to train the network?

Yes, it is possible to use other high-level quantum chemistry methods different from CCSD(T) to generate training labels and train the network. In our machine learning training scheme, only the calculated properties (energy, electric dipole, electric quadrupole, …) are used as training labels. Therefore, other high-level quantum chemistry methods, including full-CI, can be used to train the network as long as they can provide the target properties. We included the following discussion on paragraph 1, page 9:

"Besides CCSD(T), our scheme can also use other high-level quantum chemistry methods, such as full-CI, CASSCF, MP2, double-hybrid functional, and multireference electronic structure methods, to generate the training labels of molecule properties. Quantum chemistry methods can be selected according to the desired accuracy and the nature of systems under consideration."

3. A specific question about the supplement materials: Towards end of the left column, it says "in Eq. (S2), the summation over m goes through all states except n. But in equation S2, there is no summation over m and n. Therefore the subsequent argument about symmetry does not work. Also in S3, there is no epsilon_m or n.

We thank the reviewer for the careful examination and we apologize for the typo in summation index. We have corrected the paragraph as follow to make the index consistent:

“in Eq. (S2), the summation over p goes through all states except i. But as the summed formula in Eq. (S3) is antisymmetric to a and i, the terms that a goes from 1 to ne/2 cancel each other. Only terms that a goes from ne/2 + 1 to N_{basis} have a non-zero contribution to the final gradient. Therefore, i is always occupied, and a is always unoccupied in the summation. As close-shell molecules have a finite bandgap, ϵ_i and ϵ_a are not close to each other in any term of the summation, so evaluating Eq. (S3) is numerically stable.”

At the meantime, we double checked all equations in the manuscript and SI and corrected another typo of pre-factor in f_Q (Eq. 13) from $2e^2$ to $-2e$.

Reviewer #3 (Remarks to the Author):

The manuscript "Multi-task learning for molecular electronic structure approaching coupled cluster accuracy" by Tang et al is an interesting and much needed approach for machine learning approaches in electronic structure theory. Instead of targeting energy as the output descriptor which is a low dimensional and often degenerate observable, the authors achieve multi-task learning by targeting the Fock operator. As a result one can derive any observable that is well defined within the single reference mean field framework. Furthermore, the high accuracy learning is achieved for smaller molecules with less than ~100 electrons while the prediction space is much larger and can ideally be as large as required within the hydrocarbon framework of the chemical space. Therefore, I find the goal and achievement of the approach to be quite laudable.
However, there are significant improvement that is required in the manuscript to improve its readability.

We thank the reviewer for the positive comments and helpful suggestions on the readability of the manuscript. We made careful revision to the manuscript accordingly. Please see our point-to-point response below:

1. The major problem I have with the paper is that due to condensed format in which it is written, it is very difficult to understand the workflow in detail. The methodology section needs to be more descriptive for the readability and general reproducibility of the work that is mentioned.

To make the methodology section more descriptive, we included more descriptions on the detailed workflow of dataset generation (as section IV.D on page 11) and model training (as section IV.E on page 11) in the methodology section. In the dataset generation part, we added details on how we obtained the molecular structures and implemented coupled cluster calculations to build the datasets. In the model training part, details on the training hyperparameters are discussed. We also revised the model architecture part (section IV.C and IV.D on page 10) of the methodology section to improve the readability and reproducibility. The newly included and revised contents are highlighted in the resubmitted manuscript. We believe after the revision, the methodology section becomes more descriptive.

Finally, we recently notice that the word “EGNN” we used to call our model can lead to confusion with the model architecture in E(n) Equivariant Graph Neural Networks (mlr.press). Therefore, we refer to our model as "Multi-task electronic Hamiltonian network" (MEHnet) for brevity.

2. I am guessing the workflow is applicable for only hydrocarbons. This is what I surmise from the figures. However, I do not see where that is explicitly written. Page 5 (Model Performance and Applicability) should mention which class of small to large molecules are considered. If it is hydrocarbons, are they all possible hydrocarbons? How is even this chemical space created? Are all metastable hydrocarbons included? This is crucial to describe because the nature of molecules are quite diverse in this chemical space and the exact degree of transferability of learning should be understood.

We included more clarification about the class of molecules considered and tested our model performance in different class of hydrocarbons to understand the model transferability. Please see our detailed response and revisions as follow:

1. We added clarifications in paragraph 3, page 5 to explicitly state that our primary results focus on close-shell hydrocarbons. While, some additional results in the newly added section titled "QM9 version of MEHnet" cover a broader range of molecules.
2. We added descriptions about the diversity of our hydrocarbon dataset in paragraph 2, page 5 and SI section S2. We selected 85 hydrocarbon molecules from the PubChem database that covers diverse types of carbon-carbon bonds (single, double, triple bonds, conjugated π bonds) and connections (linear, branched, and cyclic). Metastable hydrocarbons are included, and out-of-equilibrium atomic configurations are introduced via molecular dynamics simulations. Please see also our replies to the first comment of reviewer 1 (page 1,2 of this document).
3. As the structure space of hydrocarbons is in principle infinite, it is impractical to test the prediction accuracy of our model for all possible hydrocarbons or all metastable structures. Nevertheless, we can classify the molecules based on structural features, such as saturate/unsaturated/cyclic/aromatic, linear/branched/ring-containing molecules. The performance of our model is tested on different types of molecules, which helps understand the transferability and generalizability of our model. The results show that our model gives consistently good prediction accuracy in various classes of hydrocarbon, including metastable and our-of-equilibrium structures. The results are added in SI section S2 with brief discussions in the main text paragraph 3, page 6.

3. Fig.1c also describes some of the data set creation. However, it is a bit unclear what the orange and blue dots refer to. Are these MD and CCSD(T) calculations? But that doesn't quite make sense. What is the meaning behind the different size of these dots? No of calculations? These are totally not mentioned in the paper.

We apologize for the confusion in the meaning of dots. The blue dots refer to chemical formula in the training domain, while the orange dots refer to chemical formula out of the training domain, which are typically larger than those in the training domain. All blue dots and some orange dots contain CCSD(T) calculations, while some other orange dots have experimental values to check with. The size of these dots means how many conformers and vibrational configurations are included for each chemical formula. For example, the chemical formula C_6H_6 contains 5 conformers: Benzene, 1,5-Hexadiyne, 2,4-Hexadiyne, Divinylacetylene, and 3,4-Dimethylene-cyclobut-1-ene. 100 vibrational configurations are sampled for each conformer. Then, there are 500 overall configurations in the training dataset, so a large dot corresponding to 500 is assigned.

We add detailed clarification in the caption of Fig. 1c.

4. In Fig. 2c, the RMSE of each property is shown. I am guessing this RMSE is with respect to CCSD(T) numbers. In such a case, it will be interesting to see the prediction efficiency (RMSE) in the training domain versus the generalization domain. Furthermore, it can be easily observed that different properties show different error ranges. The reason for this should be discussed. Also, static electric polarizability shows a trend of increasing error with increasing system size. Why?

Yes, the RMSE is with respect to CCSD(T) numbers, and we added clarifications to the caption of Fig. 2c and Table 1 of the main text. We also show the numerical comparison of the RMSE in the training domain versus the generalization domain in Table I as $RMSE_{train}/RMSE_{generalization}$ for all compared

quantities, where we can see that despite some numerical difference, $RMSE_{train}$ and $RMSE_{generalization}$ are generally close to each other for intensive quantities (E per atom, C, B, and E_g). The extensive quantities (p, Q, α) exhibit larger RMSE in generalization domain because the quantities themselves are larger for larger molecules in the generalization domain. The relatively error, however, still remains similar for training and generalization domain.

Discussions on the error range are added to paragraph 3, page 6 in the main text. For intensive quantities, the errors are in a similar level for molecules with different sizes; for extensive quantities, as mentioned by the reviewer, there is a trend of increasing error with increasing system size. This is because the absolute values of these quantities increase with system size.

5. Also a question that constantly bothers me is that the intended accuracy bar is set for CCSD(T) which may or may not be relevant for different classes of molecules (even within hydrocarbons). There are polyaromatic hydrocarbons which are more multireference in nature. Are they even included in the data set? Maybe some of this aspect can be seen in Fig. 3a. The experimental versus EGNN uncertainty is quite different for certain molecules.

Thanks. We agree that the results of CCSD(T) calculations may not be consistently accurate for all molecules we studied. Some of the polyaromatic hydrocarbons are more multireference in nature, so that the CCSD(T) calculations themselves may exhibit larger errors for these molecules than for other molecules. As all our training and testing data takes CCSD(T) as the ground truth, our model cannot capture the strong multireference effects that are not captured by CCSD(T).

As suggested by the reviewer, molecules exhibiting larger errors in Fig. 3a of the main text may involves stronger multireference character. In order to study this issue, we implemented T1 diagnostic to both our CCSD(T) dataset and aromatic molecules in Fig. 3a to get more insight into their multireference character. The evaluated T1 values are shown in SI Fig. S6 and Table II:

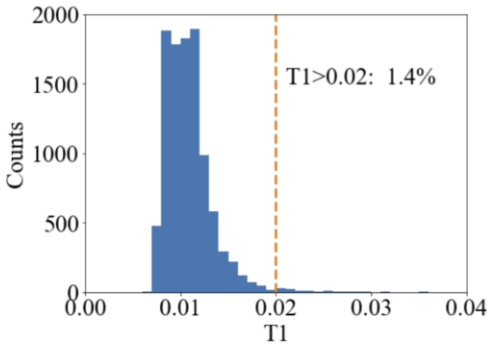


FIG. S6. Histogram of T1 diagnostic results of our coupled-cluster dataset. Most of the data points have T1 values below 0.02, indicating that the dataset does not have strong multireference character.

TABLE II. List of serial numbers (Sr. No., defined in Ref. [4]) and thermochemical properties of aromatic molecules in the main text Fig. 3. The name of each molecule is 1: trans-10b,10c-dimethyl-10b,10c-dihydropyrene; 2: anthracene; 3: benzo[c]phenanthrene; 4: 5-ring phenacene, picene; 5: Pyrene; 6: Coronene; 7: 1,4:2,5-[2.2.2.2]cyclophane; 8: 9,9'-bianthryl; 9: *p*-terphenyls; 10: acenaphthene; 11: Aceplaidylene.

Mol. index	Sr. No.	U (Eh)	G (Eh)	H_f^{MteHnet} (kJ/mol)	$H_f^{\mathrm{Exp.}}$ (kJ/mol)	T1
1	12	-5.9022	-5.5028	362.9	338.8	0.0097
2	20	-4.4302	-4.151	208.5	229.1	0.0102
3	24	-5.617	-5.2615	301.9	291.2	0.01
4	28	-6.8246	-6.3928	340.5	317.3	0.01
5	32	-4.9635	-4.6556	210.6	225.7	0.01
6	36	-7.2236	-6.7812	290.4	302.0	0.0102
7	40	-8.773	-8.1545	458.8	409.5	0.0088
8	44	-8.6805	-8.1234	484.4	454.3	0.0102
9	48	-5.776	-5.4015	289.2	272.9	0.0098
10	64	-3.9225	-3.6712	139.3	156	0.0098
11	68	-4.9042	-4.5976	366.1	362	0.0109

We see in most cases, the T1 values are below 0.02, suggesting that the studied system does not exhibit strong multireference character. Hence, it is generally reasonable to use CCSD(T) as the ground truth for molecules considered in the current work. We included discussion in paragraph 1, page 9 of the main text and SI section S6 about the potential error source from the multireference nature of certain molecules. One potential way to further improve the prediction accuracy is to conduct multireference configuration interaction (MRCI) calculations for the molecules with large T1 values. This is, however, beyond the scope of this paper, so we leave the implementation of the MRCI dataset generation to future work.

In short, I find the idea of the paper quite intriguing but due to brevity and condensed writing style, much of the details is quite obscure. One further aspect that needs some discussion is why use Fock operator as the output descriptor and not include something like a one particle density matrix? It would still have the same order of difficulty in the problem while resolving the issue of multireference nature of the problem. An interesting example (albeit in much smaller class of systems) can be seen in J. Chem. Theory Comput. 2024, 20, 11, 4569–4578.

We agree that using one particle density matrix (1-rdm) as the output descriptor is an insightful idea to resolve the multireference issue. The method was explored in the paper mentioned by the reviewer as well as [Nat. Commun. 14.1 (2023): 6281].

In our case, as our work aims to generalize from small to large molecules, including 1-rdm as the output descriptor also involves certain challenges. As some molecules we studied have delocalized orbitals (as shown in Fig. 4a), the 1-rdm, unlike the Fock matrix, can have non-zero off-diagonal terms between atomic orbitals far from each other. Therefore, predicting the 1-rdm requires the neural network architecture to directly capture delocalized features of the whole molecules.

In both reference papers mentioned above, the machine learning models involve all-to-all connection of atoms in the molecules in order to capture the delocalized features of 1-rdm. However, these model architectures have their limitation: it only applies to molecules with the same number of atoms as the training data. Therefore, as we intend to generalize from small to large molecules, different model architecture is needed. We think constructing new model architectures for systems with strong multireference characters is an intriguing direction to explore.

We include discussions about outputting density matrix in paragraph 1, page 9 of the main text and SI section S6.