Editors' Suggestion

# Neural-network Density Functional Theory Based on Variational Energy Minimization

Yang Li[1,*], Zechen Tang[1,*], Zezhou Chen[1,*], Minghui Sun[1], Boheng Zhao[1], He Li[1,2], Honggeng Tao[1], Zilong Yuan[1], Wenhui Duan[1,2,3,†] and Yong Xu[1,3,4,‡]

[1]*State Key Laboratory of Low Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing 100084, China*
[2]*Institute for Advanced Study, Tsinghua University, Beijing 100084, China*
[3]*Frontier Science Center for Quantum Information, Beijing, China*
[4]*RIKEN Center for Emergent Matter Science (CEMS), Wako, Saitama 351-0198, Japan*

Deep-learning density functional theory (DFT) shows great promise to significantly accelerate material discovery and potentially revolutionize materials research. However, current research in this field primarily relies on data-driven supervised learning, making the developments of neural networks and DFT isolated from each other. In this work, we present a theoretical framework of neural-network DFT, which unifies the optimization of neural networks with the variational computation of DFT, enabling physics-informed unsupervised learning. Moreover, we develop a differential DFT code incorporated with deep-learning DFT Hamiltonian, and introduce algorithms of automatic differentiation and backpropagation into DFT, demonstrating the capability of neural-network DFT. The physics-informed neural-network architecture not only surpasses conventional approaches in accuracy and efficiency, but also offers a new paradigm for developing deep-learning DFT methods.

Deep-learning *ab initio* calculation is an emerging interdisciplinary field, which aims to greatly enhance the capability of *ab initio* methods by using state-of-the-art neural-network approaches [1–13]. For instance, the use of neural-network quantum states significantly improves the accuracy of quantum Monte Carlo calculations [14,15]; the integration of deep learning and density functional theory (DFT) can speed up material simulations by several orders of magnitude [6–11]. In particular, deep-learning DFT can potentially have a revolutionary impact on future research due to the indispensable role of DFT in various fields ranging from physics and chemistry to materials science. Stimulated by the Materials Genome Initiative launched in 2011, great efforts have been devoted to building computational material databases via DFT. Deep-learning DFT will act as a game changer in the field, since neural-network algorithms can considerably accelerate the construction of bigger material databases, and the bigger data would in turn allow for training more powerful neural-network models. In this context, combining neural networks with DFT database construction holds great promise for advancing materials discovery.

Current research on deep-learning DFT, however, treats the tasks of DFT and neural networks separately: People first compute material datasets by DFT and then train neural-network models by data-driven approaches. With this strategy, individuals can focus on the methodological development of neural networks without needing to delve into the intricacies of DFT algorithms. This results in the development of several valuable neural-network representations of DFT [6–11]. In contrast, a more intriguing strategy is to achieve a synergistic combination of neural networks and DFT, termed neural-network DFT, which enables their methodological developments to benefit each other mutually. This objective is theoretically feasible due to the resemblance between the variational principle in physics and the loss minimization rule in deep learning. In analogy to neural-network quantum Monte Carlo [14,15], one may express the total energy as a functional of DFT quantities, such as the Hamiltonian, wave function, and charge density [16–18], and define the energy functional as a loss function for training neural networks. While algorithms for neural networks and DFT are well developed individually, the coherent integration of the two for creating an advanced deep-learning architecture remains elusive.

In this work, we introduce a theoretical framework of neural-network DFT, which unifies the minimization of loss in neural networks with the optimization of the energy functional in DFT. The central idea is to express the total energy as a functional of DFT quantities while simultaneously representing these quantities using neural networks, allowing one to define the energy functional as the loss function of neural networks. We suggest selecting the DFT Hamiltonian as the target quantity to acquire transferable neural-network models and variational energy

---

functionals. To illustrate this concept, we develop a computational code called AI2DFT for implementing neural-network DFT. In AI2DFT, equivariant neural networks are employed to learn the mapping from material structure to the DFT Hamiltonian, and DFT algorithms are adapted to be differentiable, introducing modern techniques of automatic differentiation and backpropagation into DFT. Remarkably, neural-network DFT enables physics-informed unsupervised learning, offering superior accuracy and efficiency compared to conventional supervised learning methods. The work establishes a new pathway for developing deep-learning DFT methods.

The Kohn-Sham DFT is the most widely used *ab initio* approach in material simulations. The method maps the complicated problem of interacting electrons to a simplified problem of noninteracting electrons described by an effective single-particle Kohn-Sham Hamiltonian, which takes the intricate many-body effects into account by employing approximated exchange-correlation functionals [16]. Typically, DFT calculations are performed by solving the Kohn-Sham equation via self-consistent field (SCF) iterations. A more fundamental approach involves the variational principle, which computes ground-state properties of materials by minimizing an energy functional of DFT. In fact, the Kohn-Sham equation is formally derived from the variational principle. Although the variational approach is more fundamental and favored by theoretical physics, it is typically not employed for DFT computation. One possible reason is that the variational method requires searching through high-dimensional parameter spaces, which may not be as efficient as solving differential equations iteratively. The situation, however, could potentially change as advanced algorithms and hardware developed for deep learning become available.

The integration of deep learning and DFT has revolutionized the paradigm of method development [1–12]. Nevertheless, previous research in deep-learning DFT has primarily relied on data-driven supervised learning techniques. As illustrated in Fig. 1(a), DFT SCF calculations are first performed to generate training data for varying material structures; neural networks are designed and trained for predicting data resembling DFT results. During this process, the DFT computation and neural-network optimization are separated. We call this scenario "neural networks and DFT".

An intriguing strategy is to intimately integrate neural networks and DFT together, termed neural-network DFT [Fig. 1(b)]. This advanced architecture allows us to pursue the synergistic effects between the two. For instance, algorithms for neural networks and DFT may be shared with each other, and their developments can be mutually stimulated. More importantly, by explicitly introducing the knowledge of DFT into deep learning, neural network models might be trained to better emulate real physics than the previous data-training approach.
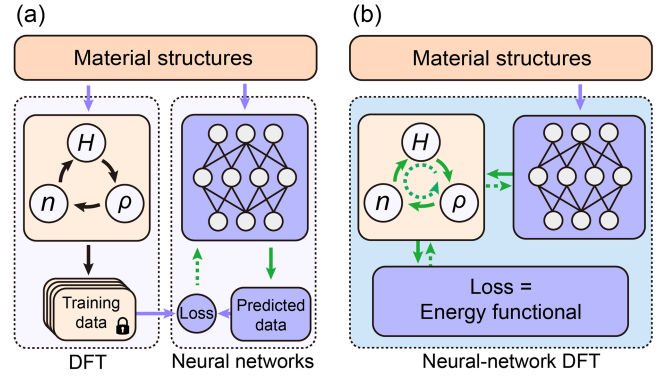


FIG. 1. Two different approaches for deep-learning DFT. (a) Conventional data-driven supervised learning, where DFT calculations on varying material structures are performed to generate training data, and the data are used to train neural networks. The DFT Hamiltonian $H$, density matrix $\rho$, and charge density $n$ are iteratively computed. (b) Physics-informed unsupervised learning based on neural-network DFT, which integrates the minimization of the energy functional in DFT with the optimization of the loss function in neural networks. Black arrows depict normal computation, and green arrows denote differential programming computation, with solid and dashed lines representing forward pass and gradient backpropagation, respectively.

Variational DFT is preferred for implementing neural-network DFT, because the minimization of the energy functional in DFT is similar in spirit to the optimization of loss function in neural networks. The total energy of DFT can be implicitly written as a functional of charge density $n$:

$$E_{\mathrm{DFT}}[n] = T_s + E_{\mathrm{ext}}[n] + E_{\mathrm{Hartree}}[n] + E_{\mathrm{XC}}[n] + E_{II},$$

which includes the single-particle kinetic energy $T_s$, the external potential energy $E_{\mathrm{ext}}$, the Hartree energy $E_{\mathrm{Hartree}}$, the exchange correlation energy $E_{\mathrm{XC}}$, and the classic interaction between nuclei $E_{II}$ [16]. Note that $n$ can be derived from other DFT quantities, such as the Kohn-Sham eigenstates ($\{\psi_i\}$), density matrix ($\rho$), and DFT Hamiltonian ($H_{\mathrm{DFT}}$). The subscript "DFT" will be omitted without leading to confusion. Hence different kinds of energy functionals can be defined: $E[Q]$ [16], where the target quantity $Q = n, \{\psi_i\}, \rho$, or $H$. Inspired by neural-network quantum Monte Carlo methods [14], neural networks are utilized to represent the target DFT quantity, denoted as $Q_\theta$, where $\theta$ represents parameters or weights of neural networks. Consequently, the total energy becomes a function of neural-network parameters, denoted as $E[Q_\theta]$. In principle, one may use $E[Q_\theta]$ as the loss function, and the optimization of neural networks naturally completes the computation of variational DFT, as depicted in Fig. 1(b).

The selection of target DFT quantity $Q$ merits careful consideration. Here, we present three selection criteria. Firstly, the target of deep learning should closely adhere to the "nearsightedness" or locality principle proposed by

Walter Kohn [19,20], thereby enhancing the transferability of neural network models. Local physical quantities such as $\rho$, $n$, and $H$ meet the criterion [6,16], whereas nonlocal quantities such as Kohn-Sham eigenstates do not. Secondly, neural network representations that satisfy the fundamental requirements for the target quantity should be readily available. Here are some essential requirements: $n$ is positive in value and normalized to a constant; $H$ is Hermitian; $\rho$ at zero temperature is idempotent [16,19]. While the former two conditions are relatively straightforward to achieve in deep learning, meeting the strict idempotency requirement of $\rho$ remains challenging. Thus, we may further consider $E[H]$ or $E[n]$. Thirdly, the energy functional should be variational. $E[H]$ satisfies the criterion [16], whereas $E[n]$ does not [21], meaning that minimizing $E[n]$ may not yield the correct ground state. Based on the above considerations, we choose the energy functional $E[H]$ for our subsequent study.

Neural-network representations of the DFT Hamiltonian have been developed and applied for accelerating large-scale material simulations [6–8]. In the approach of deep-learning DFT Hamiltonian (DEEPH) [6], neural networks are employed to represent $H$ as a function of the material structure $\{\mathcal{R}\}$. Message-passing graph neural networks [22] are utilized, with vertices representing atoms and edges denoting atom pairs. Features of vertices and edges are updated through message passing from neighboring atoms, facilitating the aggregation of information from distant chemical environments. In line with the locality principle, $H$ is expressed under localized atomic basis sets. We will employ the DEEPH-E3 framework [7] to ensure that the mapping $\{\mathcal{R}\} \mapsto H$ is equivariant under the Euclidean group in three-dimensional space. In DEEPH-E3, only equivariant vectors carrying irreducible representations of the three-dimensional rotation group are permitted, and their tensor product is determined by the Wigner-Eckart theorem. The principle of equivariance is thus maintained. As illustrated in Fig. 2(a), equivariant neural networks take the embedding of material structure information as input and represent the Hamiltonian matrix through their output. In this way, $H_\theta$ parametrized by neural-network weights is obtained, and thus the energy functional $E[H]$ can be viewed as the loss function $E[H_\theta]$ for neural networks.

In neural-network DFT, the DFT program must supply $\nabla_H E$ for optimizing neural network parameters. This poses a significant challenge in terms of DFT programming, as $\nabla_H E$ is not mandatory for standard SCF calculations. Fulfilling this requirement calls for a DFT program that is end-to-end differentiable. Automatic differentiation (AD) offers a suite of methods for numerically computing the derivatives of functions embedded in computer programs [23]. It systematically applies the chain rule and calculus principles, eliminating the need for manual derivation, thus making it well suited for computing differential quantities like $\nabla_H E$. Currently, most DFT codes do not fully support
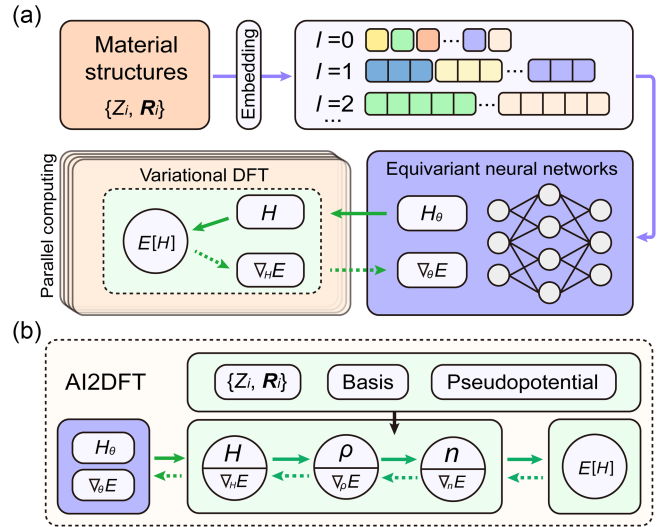


FIG. 2. Architecture and implementation of neural-network DFT. (a) Overall architecture: variational DFT, which minimizes the energy functional $E[H]$, is linked to equivariant neural networks representing the DFT Hamiltonian $H_\theta$. $\theta$ denotes the parameters of neural networks. The structural information of materials, including the element type $Z_i$ and atomic coordinate $\boldsymbol{R}_i$ for each atom $i$, is embedded into equivariant vectors labeled by the angular momentum quantum number $l$, serving as inputs to neural networks. (b) Implementation by the AI2DFT code employing differentiable programming. The forward pass is from $H$ to density matrix $\rho$ and charge density $n$, and finally to $E[H]$. Automatic differentiation is applied to compute the gradients $\nabla_n E$, $\nabla_\rho E$, $\nabla_H E$, and $\nabla_\theta E$, and backpropagation is utilized for optimization.

the function of AD. Our neural-network DFT necessitates a differentiable implementation of DFT capable of accommodating periodic boundary conditions as well as localized atomiclike bases, a feature that, to our knowledge, has not been developed. Consequently, we have developed our own autodifferentiable DFT program named "AI2DFT" using the JULIA language with the AD capability realized by the ZYGOTE package [24]. Method details are described in the Supplemental Material [25].

Remarkably, AD is available for both DFT computation and neural-network training in AI2DFT [Fig. 2(b)]. In DFT, one first derives $\rho$ and $n$ from $H$ and then uses the two quantities to compute the total energy. Based on the chain rule, AI2DFT uses the reverse-mode AD to compute $\nabla_n E$, $\nabla_\rho E$, and $\nabla_H E$ in turn. The gradient information $\nabla_\theta E$ is used in neural networks for optimization. AI2DFT supports three autodifferentiable working modes: the conventional SCF DFT and new functionalities of variational DFT and neural-network DFT.

By comparing with SCF DFT, we noticed a critical problem emerging in variational DFT: The variation of $E[H]$ can correctly predict certain ground-state properties, such as the total energy, charge density, and density matrix, but not the DFT Hamiltonian. Upon analysis of the

calculated electronic structure, we observed that while the occupied manifold matches the result of SCF DFT, the unoccupied part does not. The unoccupied states have zero contribution to the ground-state energy at zero temperature, which cannot be uniquely determined via energy minimization. For computing DFT total energy, $H$ contains complete but redundant information; $\rho$ is just sufficient, which forms the foundation of density matrix functional theory [19,33]. Since $\rho$ can be reliably predicted by energy minimization, one may use $\rho$ to reconstruct a new Hamiltonian, denoted as $\tilde{H}$. The Hamiltonian reconstruction may be used for postprocessing in variational DFT, giving the ground-state Hamiltonian the same as SCF DFT.

To be compatible with neural-network DFT, we propose to solve the above problem of the Hamiltonian in a more elegant way by introducing a generalized energy functional: $\tilde{E}[H] = E[H] + f[H - \tilde{H}]$, with the extra term $f[H - \tilde{H}]$ to emulate the spirit of Hamiltonian reconstruction. The functional $f$ should attain its minimum value for $H = \tilde{H}$, ensuring that $\tilde{E}[H]$ shares the same ground-state energy with $E[H]$. The generalized energy gradient $\nabla_H \tilde{E}$ is relevant for optimization. According to our preliminary tests, we selected a gradient of the form $\nabla_H E + \lambda(H - \tilde{H})$ in our study, where $\lambda$ is an adjustable coefficient as discussed in the Supplemental Material [25]. Note that one may replace the energy functional $E[H]$ with a free energy functional $F[H]$ for introducing temperature into DFT [34]. In neural-network DFT, the second gradient term trains the Hamiltonian predicted by neural networks to resemble the reconstructed Hamiltonian. The energy gradient term $\nabla_H E$, which is typically not included in deep-learning DFT, drives neural networks to evolve toward lower energy configurations and enables neural networks to learn the underlying physics in an unsupervised manner. Therefore, we refer to our proposed neural-network DFT as physics-informed unsupervised learning.

We have established a theoretical framework of neural-network DFT and numerically implemented it by the AI2DFT code employing differentiable programming. Next, we comprehensively test AI2DFT by studying various types of materials, including the $H_2O$ molecule, graphene, monolayer $MoS_2$, and bulk body-centered-cubic Na. These examples of materials span from molecules to periodic crystals and from metals to insulators. Firstly, we checked that SCF iterations of AI2DFT can well reproduce the benchmark results of the SIESTA code [26] (Fig. S3). Then, we applied variational DFT to study the same materials. The total energy can converge below the μeV scale after tens of variation steps [Figs. 3(a) and 3(b), and Fig. S1]. Other physical quantities such as the energy gradient, Hamiltonian, density matrix, and charge density also exhibit exponential convergence behavior, validating the reliability and robustness of variational DFT.

Moreover, we assess the performance of neural-network DFT, which combines variational DFT with the DEEPH-E3
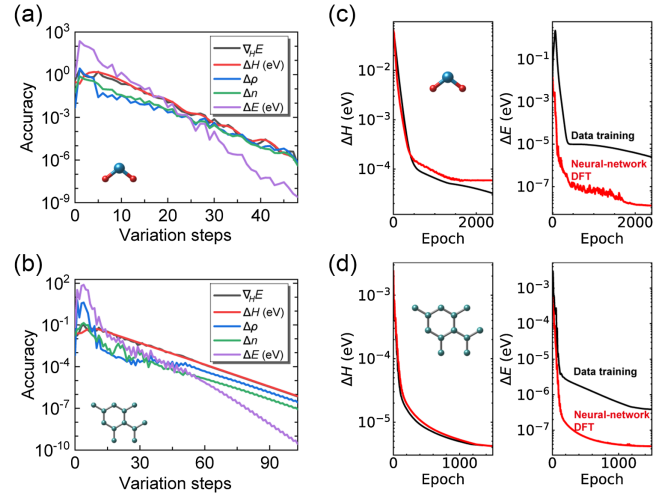


FIG. 3. Validation of AI2DFT. (a),(b) Variational DFT versus SCF DFT: the mean average values of $\nabla_H E$, $\Delta H$, $\Delta \rho$, $\Delta n$, and $\Delta E$ display exponential convergence behavior as increasing variation steps. (c),(d) Neural-network DFT versus data training: the mean average values of $\Delta H$ and $\Delta E$ steadily decrease with increasing training epochs. The upper and lower panels present example calculations on the $H_2O$ molecule and graphene. Results obtained by SCF DFT calculations are used as a reference.

[7] neural networks, through comparison with conventional data-driven supervised learning approaches [Figs. 3(c) and 3(d)]. For the study of the $H_2O$ molecule, the DFT Hamiltonian can indeed be optimized to achieve a high level of accuracy: 0.06 meV by neural-network DFT and 0.02 meV by data training. Even higher accuracies of 0.004 meV are achieved for graphene using both approaches. The reliability of neural network approaches is thus confirmed. Meanwhile, we computed physical quantities using the $H$ predicted by neural networks and monitored their accuracies throughout the training process. An intriguing trend is observed: Neural-network DFT shows significantly improved performance over data training in predicting the derived physical quantities. For instance, in the study of the $H_2O$ molecule, the prediction accuracy of energy reaches 0.013 μeV by neural-network DFT, over 60 times better than data training (0.83 μeV) [Fig. 3(c)]. A similar trend is noticed for graphene [Fig. 3(d)] as well as for other quantities such as $\rho$ and $n$ (Fig. S12). Furthermore, we devised an artificial experiment by introducing Gaussian noise to the benchmark DFT Hamiltonian to simulate numerical inaccuracies [25]. For Hamiltonians with comparable levels of accuracy, the physical quantities obtained from neural-network DFT exhibit orders of magnitude better precision than those computed using Hamiltonians generated with Gaussian noise (Fig. S11). The underlying reason is that neural-network DFT introduces the energy gradient $\nabla_H E$ into optimization, enabling effective filtering of unphysical high-energy noises in the learned Hamiltonian. In this
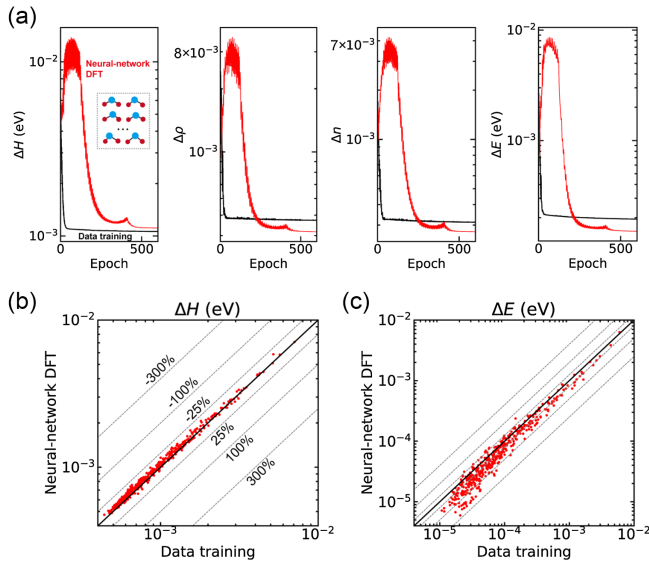
FIG. 4. Unsupervised learning on multiple $H_2O$ structures by neural-network DFT. (a) Learning curves of neural-network DFT compared with data-driven supervised learning. Both are initiated from a pretrained neural network model. The mean absolute errors of Hamiltonian ($\Delta H$), energy ($\Delta E$), density matrix ($\Delta\rho$), and charge density ($\Delta n$) are computed across all 300 training structures. (b),(c) Mean average errors of (b) Hamiltonian and (c) energy across 435 test structures: Neural-network DFT versus data training. The gray dashed lines indicate performance improvement ratios between the two methods.

this method is elaborated upon in the Supplemental Material [25]. Moreover, we numerically implemented neural-network DFT using deep-learning DFT Hamiltonian and differentiable programming, bringing modern techniques of automatic differentiation and backpropagation into DFT. In this context, the developments of neural networks and DFT computation are no longer isolated but will get mutual benefits and stimulate each other. The work introduces new avenues for the collaborative development of artificial intelligence and *ab initio* methods, significantly enriching the scope of deep-learning *ab initio* research.

regard, the physics-informed approach surpasses conventional data-driven methods.

Finally, we apply neural-network DFT to compute multiple material structures and demonstrate its capability of unsupervised learning. Using $H_2O$ molecules as an example study, we initially obtained a pretrained neural network model through data-driven supervised learning using the DEEPH-E3 approach, fine-tuned the model across 300 training structures using neural-network DFT, and achieved high prediction accuracies for Hamiltonian as well as other quantities [Fig. 4(a)]. The trained neural-network model was further employed to make predictions on 435 test structures that were unseen during training, showing good generalization ability [Figs. 4(b) and 4(c), Figs. S10 and S13] as discussed in the Supplemental Material [25]. Compared with data-driven supervised learning, neural-network DFT yields Hamiltonians with a slightly larger mean absolute error. Nevertheless, it exhibits superior performance in predicting derived physical quantities, as confirmed by both training and test results. This is consistent with the above results, highlighting the advantage of physics-informed learning.

In summary, we proposed a theoretical framework of neural-network DFT that coherently combines variational DFT and equivariant neural networks together, enabling physics-informed unsupervised learning. The advantage of

[1] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

[2] L. Zhang, J. Han, H. Wang, R. Car, and W. E, Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics, Phys. Rev. Lett. **120**, 143001 (2018).

[3] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer, Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions, Nat. Commun. **10**, 5024 (2019).

[4] O. T. Unke, M. Bogojeski, M. Gastegger, M. Geiger, T. Smidt, and K.-R. Müller, SE(3)-equivariant prediction of molecular wavefunctions and electronic densities, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., Red Hook, New York, 2021), p. 14434.

[5] Q. Gu, L. Zhang, and J. Feng, Neural network representation of electronic structure from *ab initio* molecular dynamics, Sci. Bull. **67**, 29 (2022).

[6] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, and Y. Xu, Deep-learning density functional theory Hamiltonian for efficient *ab initio* electronic-structure calculation, NATO ASI series Series F, Computer and system sciences **2**, 367 (2022).

[7] X. Gong, H. Li, N. Zou, R. Xu, W. Duan, and Y. Xu, General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian, Nat. Commun. **14**, 2848 (2023).

[8] H. Li, Z. Tang, X. Gong, N. Zou, W. Duan, and Y. Xu, Deep-learning electronic-structure calculation of magnetic superstructures, NATO ASI series Series F, Computer and system sciences **3**, 321 (2023).

[9] H. Li, Z. Tang, J. Fu, W.-H. Dong, N. Zou, X. Gong, W. Duan, and Y. Xu, Deep-learning density functional perturbation theory, Phys. Rev. Lett. **132**, 096401 (2024).

[10] Z. Tang, H. Li, P. Lin, X. Gong, G. Jin, L. He, H. Jiang, X. Ren, W. Duan, and Y. Xu, Efficient hybrid density functional calculation by deep learning, arXiv:2302.08221.

[11] Y. Wang, H. Li, Z. Tang, H. Tao, Y. Wang, Z. Yuan, Z. Chen, W. Duan, and Y. Xu, DEEPH-2: Enhancing deep-learning electronic structure via an equivariant local-coordinate transformer, arXiv:2401.17015.

[12] H. Yu, Z. Xu, X. Qian, X. Qian, and S. Ji, Efficient and equivariant graph networks for predicting quantum Hamiltonian, in *International Conference on Machine Learning* (PMLR, 2023), pp. 40412–40424, https://dl.acm.org/doi/10.5555/3618408.3620100.

[13] Z. Yuan, Z. Xu, H. Li, X. Cheng, H. Tao, Z. Tang, Z. Zhou, W. Duan, and Y. Xu, Equivariant neural network force fields for magnetic materials, Quantum Front **3**, 8 (2024).

[14] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[15] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, Phys. Rev. Res. **2**, 033429 (2020).

[16] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, England, 2004), 10.1017/CBO9780511805769.

[17] T. Li, M. Lin, Z. Hu, K. Zheng, G. Vignale, K. Kawaguchi, A. C. Neto, K. S. Novoselov, and S. YAN, D4FT: A deep learning approach to Kohn-Sham density functional theory, in *International Conference on Learning Representations (ICLR)* (ICLR, Appleton, Wisconsin, 2023).

[18] A. Mathiasen, H. Helal, P. Balanca, A. Krzywaniak, A. Parviz, F. Hvilshøj, B. Banaszewski, C. Luschi, and A. W. Fitzgibbon, Reducing the cost of quantum chemical data by backpropagating through density functional theory, arXiv:2402.04030.

[19] W. Kohn, Density functional and density matrix method scaling linearly with the number of atoms, Phys. Rev. Lett. **76**, 3168 (1996).

[20] E. Prodan and W. Kohn, Nearsightedness of electronic matter, Proc. Natl. Acad. Sci. U.S.A. **102**, 11635 (2005).

[21] J. Harris, Simplified method for calculating the energy of weakly interacting fragments, Phys. Rev. B **31**, 1770 (1985).

[22] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, in *Proceedings of the 34th International Conference on Machine Learning (ICML)* (Microtome Publishing located in Brookline, Massachusetts, 2017), Vol. 70, p. 1263, JMLR.org.

[23] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, Automatic differentiation in machine learning: A survey, J. Mach. Learn. Res. **18**, 1 (2018) [arXiv:1502.05767].

[24] M. Innes, Don't unroll adjoint: Differentiating SSA-form programs, arXiv:1810.07951.

[25] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.133.076401 for calculation methods as well as other detailed information, which includes references [6,7,15,16,24,26–32].

[26] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, The SIESTA method for *ab initio* order-N materials simulation, J. Phys. Condens. Matter **14**, 2745 (2002).

[27] D. R. Hamann, M. Schlüter, and C. Chiang, Norm-conserving pseudopotentials, Phys. Rev. Lett. **43**, 1494 (1979).

[28] L. Kleinman and D. M. Bylander, Efficacious form for model pseudopotentials, Phys. Rev. Lett. **48**, 1425 (1982).

[29] N. Troullier and J. L. Martins, Efficient pseudopotentials for plane-wave calculations, Phys. Rev. B **43**, 1993 (1991).

[30] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).

[31] J. M. Bowman, C. Qu, R. Conte, A. Nandi, P. L. Houston, and Q. Yu, The MD17 datasets from the perspective of datasets for gas-phase "small" molecule potentials, J. Chem. Phys. **156**, 240901 (2022).

[32] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, JULIA: A fresh approach to numerical computing, SIAM Rev. **59**, 65 (2017).

[33] X.-P. Li, R. W. Nunes, and D. Vanderbilt, Density-matrix electronic-structure method with linear system-size scaling, Phys. Rev. B **47**, 10891 (1993).

[34] N. D. Mermin, Thermal properties of the inhomogeneous electron gas, Phys. Rev. **137**, A1441 (1965).