Order-N orbital-free density-functional calculations with machine learning of functional derivatives for semiconductors and metals

Fumihiro Imoto [1,2] Masatoshi Imada [2,3] and Atsushi Oshiyama [1]

[1]*Institute of Materials and Systems for Sustainability, Nagoya University, Nagoya 464-8603, Japan*
[2]*Waseda Research Institute for Science and Engineering, Waseda University, Tokyo 169-8555, Japan*
[3]*Toyota Physical and Chemical Research Institute, Nagakute 480-1192, Japan*

Orbital-free density-functional theory (OFDFT) offers a challenging way of electronic-structure calculations scaled as $\mathcal{O}(N)$ computation for system size N. We here develop a scheme of the OFDFT calculations based on the accurate and transferrable kinetic-energy density-functional (KEDF), which is created in an unprecedented way using appropriately constructed neural network (NN). We show that our OFDFT scheme reproduces the electron density obtained in the state-of-the-art DFT calculations and then provides accurate structural properties of 24 different systems, ranging from atoms, molecules, metals, semiconductors, and an ionic material. The accuracy and the transferability of our KEDF is achieved by our NN training system in which the kinetic-energy functional derivative (KEFD) at each real-space grid point is used. The choice of the KEFD as a set of training data is essentially important, because first it appears directly in the Euler equation, which one should solve, and second, its learning assists in reproducing the physical quantity expressed as the first derivative of the total energy. More generally, the present development of KEDF $T[\rho]$ is in the line of systematic expansion in terms of the functional derivatives $\delta^{\ell_1} T/\delta \rho^{\ell_1}$ through progressive increase of ℓ_1. The present numerical success demonstrates the validity of this approach. The computational cost of the present OFDFT scheme indeed shows the $\mathcal{O}(N)$ scaling, as is evidenced by the computations of the semiconductor SiC used in power electronics.

I. INTRODUCTION

Density functional theory (DFT) [1] proves that the ground-state total energy E of an interacting electron system is a unique universal functional $G[\rho]$ of the electron density $\rho(\mathbf{r})$ plus the electrostatic energy $V_{\text{ext}}[\rho]$ under the external potential $v_{\text{ext}}(\mathbf{r})$, opening a possibility to compute physical properties of real materials by solving an Euler equation $\delta E[\rho]/\delta \rho(\mathbf{r}) = \mu$, where μ is the Lagrange multiplier that enforces density normalization. Various attempts have been made by introducing virtual noninteracting electron systems, in which the electron densities are identical to those in corresponding real materials, and then decomposing G to the kinetic energy of the noninteracting system $T_s[\rho]$, the classical electron-electron interaction energy $E_{\text{H}}[\rho]$, and the remaining exchange-correlation energy $E_{\text{xc}}[\rho]$ [2]. In this scheme, T_s is expressed as a sum of the kinetic-energy contribution from each orbital $\phi_i(\mathbf{r})$ (Kohn-Sham orbital) as

$$T_s[\rho] = \frac{1}{2} \sum_i \int |\nabla \phi_i(\mathbf{r})|^2 d\mathbf{r}, \tag{1}$$

and thus the original Euler equation in DFT

$$\frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} + v_{\text{ext}}(\mathbf{r}) + \frac{\delta E_{\text{H}}[\rho]}{\delta \rho(\mathbf{r})} + \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})} = \mu \tag{2}$$

becomes a set of Schrödinger-like equations (Kohn-Sham equations), which in turn determine ϕ_i self-consistently.

A number of works adopting this Kohn-Sham (KS) scheme (KSDFT) has been applied to a various materials and achieved unprecedented success [3,4], depending on the level of the approximation to the exchange-correlation functional (Jacob's ladder) [5]. However, solving the Kohn-Sham equations for all the occupied orbitals in the system is a computational burden scaling with the system size N as $\mathcal{O}(N^3)$, thus restricting the applicability of DFT. The scheme with lower-order scaling is highly demanded in materials science and also in advancing DFT. One of the solutions in a legitimate way is the orbital-free density-functional theory (OFDFT) in which $T_s[\rho]$ is expressed as a functional of ρ, the kinetic-energy density-functional (KEDF), and the Euler equation Eq. (2) remains as a single equation for ρ, thus OFDFT being expected to be $\mathcal{O}(N)$ scheme.

Such OFDFT approach free from the orbitals is in the heart of the original DFT [1] and was initiated much before DFT, being known as Thomas-Fermi (TF) theory for the homogeneous electron gas [6,7] and von Weizsäcker (vW) gradient expression [8]. Based on these approaches, the kinetic-energy functional T_s is generally written as

$$T_s[\rho] = \int \tau^{\text{TF}}(\mathbf{r}) F[\rho] d\mathbf{r}, \tag{3}$$

where $\tau^{TF}(\boldsymbol{r})$ is the kinetic energy density in the TF approximation, $\tau^{TF} = (3/10)(3\pi^2)^{2/3}\rho^{5/3}$, and F is so called the enhancement factor.

There have been a lot of efforts to develop the enhancement factor [9,10] either in a semilocal [11–14] or a nonlocal form [15–22] to reproduce the KS kinetic energy T^{KS} as accurate as possible. Among the various semilocal forms, recent two functionals, PGSLβ (Pauli-Gaussian second order and Laplacian with a parameter β) [12,13] and LKT (Luo-Karasiev-Trickey) [14], reproduce experimental structural properties successfully within the error of less than a few percent for the lattice constants and of about ten percent for the bulk moduli. Semilocal KEDFs have been usually developed in the form of the generalized gradient approximation (GGA) and satisfy some of the exact conditions for (a) the small limit of the density gradient in the gradient expansion (GE), (b) the large limit of the density gradient in GE, (c) the positivity of Pauli potential [23], and (d) the linear response function in the homogeneous density limit, namely being equal to the Lindhard function [24]. PGSLβ satisfies the first three conditions (a)–(c). The parameter β is fitted so as to minimize the mean absolute relative errors of physical quantities such as the cell volume, the bulk modulus, the total energy at the equilibrium volume, and the electron density, with respect to the corresponding values obtained in the KS scheme, leading to the value of $\beta = 0.25$ [12,13]. LKT satisfies the conditions (b) and (c), and is left with a single parameter "a". The allowable range of "a" is determined by the condition (c) for atomic densities generated by pseudopotentials. The obtained range is $0.0 \leqslant a \leqslant 1.3$ and the value of $a = 1.3$ is typically used [14]. Karasiev *et al.* [25] proposed another semilocal parameter-free KEDF that satisfies the conditions (a)–(c). However, The bulk moduli from this functional are around 50% higher than the reference KS values. Hence, we will not include the results by this functional in our benchmark in the present paper.

Inclusion of the nonlocal effects in the enhancement factor indeed improves the performance of KEDF [18,19,22,26,27]. The nonlocal functional typically reproduces the structural properties within the error of less than 1% for the lattice constants and 10% for the bulk moduli [27]. Nonlocal KEDFs such as WGC (Wang-Govind-Carter) [19] and related nonlocal KEDFs [15–18,20,21] were constructed to satisfy the conditions (a), (b), and (d). HC (Huang-Carter) [22] functional improves the performance of WGC and its relatives in semiconductors by using two parameters adjusted to reproduce bulk moduli, equilibrium volumes, and equilibrium energies by KSDFT. EvW-WGC (enhanced von Weizsäcker Wang-Govind-Carter) [26] functional is an another extension of WGC, which is a linear combination of vW, TF, and WGC, where a system-dependent parameter determines the portions of each KEDFs. EvW-WGC is more accurate than HC if an optimally adjusted parameter is used. MGP (Mi-Genova-Pavanello) [27] functional uses a unique way of imposing the condition (d), namely the functional integration of the inverse of the Lindhard function in homogeneous density limit. In spite of the better performance generally observed, the nonlocal scheme requires the heavier computational cost, which scales as $\mathcal{O}(N \log N)$ and, if possible,

the true $\mathcal{O}(N)$ scheme enabled by the semilocal scheme is desired.

In this paper, to establish an alternative scheme that follows strictly the $\mathcal{O}(N)$ scaling, we propose a new enhancement factor in KEDF by neural-network (NN) machine learning. Without resorting to the reproducibility of the structural properties of target materials, we focus on reproducing the electron density, which is the quantity of basics in DFT. We show that our NN, which is trained only with the electron density in diamond generates a KEDF and successfully reproduces structural properties of a variety of materials, thus demonstrating the potential of the machine learning for further developments of the density functional.

Coming back to the DFT itself, the energy as a functional of the electron density, KEDF in the present case, is primary. Its functional derivative defines the Euler equation and, on the other hand, provides the physical properties related to the first-order derivative of the total energy. The higher derivatives are obviously important to describe the linear and nonlinear responses of materials. It is thus desirable to construct KEDF as well as its functional derivatives by term-by-term conformation of the kinetic-energy functional-derivative (KEFD) $\delta^{\ell_1} T / \delta \rho^{\ell_1}$ to the KS derivative through progressive increase of ℓ_1. This is combined with the variational optimization of KEDF as a functional of both ρ and its spatial derivatives $d^{\ell_2} \rho / d\boldsymbol{r}^{\ell_2}$ with increasing ℓ_2. This systematic approach was formidable in the past but now may be practical using the machine learning. We here demonstrate a success of our first attempt with $\ell_1 = 1$ and $\ell_2 = 2$.

The organization of the present paper is as follows. In Sec. II, we explain the way of constructing our NN KEDFs. Computational details are also explained. The obtained NN-KEDFs are applied to 24 systems, which include 7 atoms, 6 diatomic molecules, and 11 solids ranging from metals, semiconductors to an ionic solid in Sec. III. The accuracy of our NN KEDFs is assessed in detail and discussed. In Sec. III, we also demonstrate the $\mathcal{O}(N)$ computational time scaling of our orbital-free implementation achieved for the system with thousands of atoms. We summarize our findings in Sec. IV.

II. METHODS

A. Neural network for developing kinetic-energy density functional

The neural network (NN) in general consists of an input layer, multiple hidden layers, and an output layer. Each of those layers with an index l ($l = 0, \ldots, N$) is composed of neurons with indices k ($k = 0, \ldots, D_l$), where $N - 1$ is the number of hidden layers and D_l is the number of neurons in the lth layer (Fig. 1). The output from the jth neuron in the lth layer ($1 \leqslant l \leqslant N - 1$) is generally written as

$$z_j^{(l)} = \sigma^{(l)}(a_j^{(l)}) = \sigma^{(l)}\left(\sum_{k=0}^{D_{l-1}} W_{jk}^{(l)} z_k^{(l-1)}\right), \qquad (4)$$

where $\sigma^{(l)}$ is the activation function with the variable $a_j^{(l)}$ being $\sum_{k=0}^{D_{l-1}} W_{jk}^{(l)} z_k^{(l-1)}$. In our notation, the inputs are $z_j^{(0)}$ ($j = 1, 2$) and the output is $z_1^{(N)}$. The parameters trained in the NN, $W_{jk}^{(l)}$ ($1 \leqslant l \leqslant N$, $1 \leqslant j \leqslant D_l$, $0 \leqslant k \leqslant D_{l-1}$), are
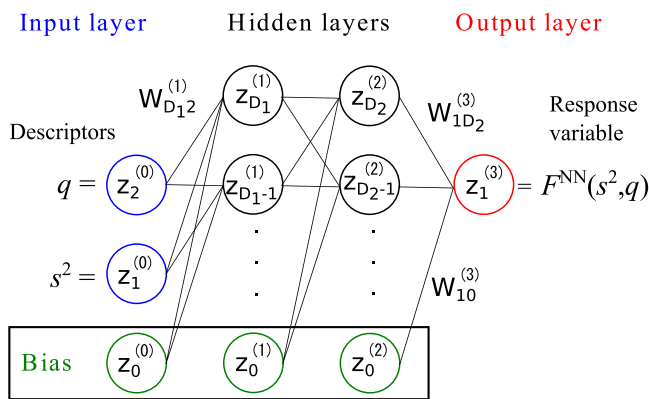
FIG. 1. Schematic structure of the NN with two hidden layers ($N = 3$).

called weights. For each layer, by using additional inputs $z_0^{(l)} \equiv 1$, $W_{j0}^{(l)}$ works as the bias parameter.

In order to develop our NN KEDF, we seek for the enhancement factor F^{NN} as a functional of dimensionless quantities derived from the gradient and Laplacian of the electron density, i.e., s^2 and q, (namely, up to $\ell_2 = 2$) where $s = |\nabla \rho|/[2(3\pi^2)^{1/3}\rho^{4/3}]$ and $q = \nabla^2 \rho/[4(3\pi^2)^{2/3}\rho^{5/3}]$. The KEDF with this enhancement factor

$$T^{\mathrm{NN}}[\rho] = \int \tau^{\mathrm{TF}}(\boldsymbol{r}) F^{\mathrm{NN}}(s^2, q; \boldsymbol{W}) d\boldsymbol{r} \qquad (5)$$

satisfies the uniform scaling condition [28]. Here, the generic notation $\boldsymbol{W} \equiv \{W^i; i = 1 \ldots n_W\}$ represents the set of weight parameters in the NN, $\{ W_{jk}^{(l)} \}$ in Eq. (4). Figure 1 shows the schematic structure of the NN with two inputs s^2 and q. The output is the enhancement factor $F^{\mathrm{NN}}(s^2, q)$ introduced in Eq. (5). The inputs and outputs are of course spatial dependent and thus the functions of the real-space position $\boldsymbol{r}$.

The explicit formula of the enhancement factor in the present NN KEDF is, e.g., for $D_l = D$ and $N = 4$ (see subsection II C for determination of the structure of our present NN),

$$F^{\mathrm{NN}}(s^2, q)$$
$$= \sum_{i=0}^{D} W_{1i}^{(4)} \sigma \left(\sum_{j=0}^{D} W_{ij}^{(3)} \sigma \left(\sum_{k=0}^{D} W_{jk}^{(2)} \sigma \left(\sum_{l=0}^{2} W_{kl}^{(1)} z_l^{(0)} \right) \right) \right), \qquad (6)$$

where $(z_1^{(0)}, z_2^{(0)}) = (s^2, q)$. In the present paper we adopt the activation function defined as the exponential linear unit (ELU) [29]

$$\sigma\left(a_j^{(l)}\right) = \begin{cases} a_j^{(l)}, & \text{if } a_j^{(l)} > 0 \\ \exp\left(a_j^{(l)}\right) - 1, & \text{if } a_j^{(l)} \leqslant 0 \end{cases} \qquad (7)$$

for each layer ($1 \leqslant l \leqslant N - 1$) and an identity function for an output layer ($l = N$).

We adopt the enhancement factor $F^{\mathrm{NN}}(s^2, q)$ as the output variable $z_1^{(N)}$. However, $F^{\mathrm{NN}}(s^2, q)$ is not directly learned as the training data. Instead, the KEFD with $\ell_1 = 1$ from the KS scheme, $\delta T^{\mathrm{KS}}/\delta\rho$, is used as the training data since we adopt the cost function defined in terms of the KEFD [Eq. (8) below]

is used as a criterion of the performance of our NN in the present paper.

B. Density-functional computation scheme

Actual density-functional (DFT) computations in the construction of our NN and its validation have been performed by our real-space code RSDFT [30–32], which is highly optimized for the massively parallel architecture. In RS-DFT scheme, all the quantities are computed on grid points in real space and the converged results are obtained by reducing the grid spacing systematically. Thanks to the real-space treatments, the scheme is essentially free from Fourier Transform, which releases heavy communication burden in the massively-parallel architecture of computers. For the exchange-correlation functional, we use generalized gradient approximation by Perdew, Burke, and Ernzerhof (PBE) [33].

In usual first-principles DFT calculations, orbital-dependent nonlocal pseudopotentials (NLPSs) are used to simulate nuclei and core electrons in real materials [34,35]. Yet in the OFDFT scheme, the orbitals are unavailable so that we need to construct *ab initio local ionic* pseudopotentials (LIPSs) to simulate nuclei and core electrons. We have newly generated such LIPSs following the scheme by Carter and collaborators [22,36–38]. Our scheme to generate LIPS consists of (i) the generation of the electron density of the target element in KSDFT scheme, (ii) obtaining the effective potential from the above density by solving the inverse problem with the Kadantsev-Stott method [39], and (iii) proper skeletonizing of the effective potential (Appendix A for details). It should be emphasized that our construction scheme is totally free from adjustable parameters. We have generated LIPSs for 7 elements, Li, C, Na, Al, Si, Cl, and Cu. The validity and transferability of the obtained LIPSs are evidenced by the electron densities and the structural properties of 8 various materials, bcc-Li, fcc-Al, fcc-Cu, bcc-Na, NaCl, ds-Si, ds-C, and zincblende(3C)-SiC (Appendix A).

C. Training of neural network toward kinetic-energy functional derivative

In order to develop the KEDF by our NN that best reproduces the KS-KEDF $T^{\mathrm{KS}}[\rho]$ within the framework of $\ell_1 = 1$, we minimize during the training process the cost function for KEFD,

$$L = \frac{1}{N_t} \sum_{n=1}^{N_t} \frac{1}{2} \left[\frac{\delta T^{\mathrm{NN}}(\boldsymbol{r}_n)}{\delta\rho} - \frac{\delta T^{\mathrm{KS}}(\boldsymbol{r}_n)}{\delta\rho} \right]^2, \qquad (8)$$

which is the mean-squared error between the kinetic-energy functional derivative (KEFD) $\delta T^{\mathrm{NN}}/\delta\rho$ obtained by our NN and KEFD $\delta T^{\mathrm{KS}}/\delta\rho$ from the KS orbitals $\{\phi_i(\boldsymbol{r})\}$. Here, N_t is the total number of the training data at the real space position $\boldsymbol{r}_n$. The analytical expressions of KEFDs are given by (see Appendix B for the derivation)

$$\frac{\delta T^{\mathrm{NN}}}{\delta\rho} = c_0 \rho^{2/3} \left[\frac{5}{3} F^{\mathrm{NN}} - \frac{8}{3} s^2 \frac{\partial F^{\mathrm{NN}}}{\partial(s^2)} - \frac{5}{3} q \frac{\partial F^{\mathrm{NN}}}{\partial q} \right]$$
$$- \frac{3}{20} \nabla \cdot \left[\frac{\partial F^{\mathrm{NN}}}{\partial(s^2)} \frac{\nabla\rho}{\rho} \right] + \frac{3}{40} \nabla^2 \left(\frac{\partial F^{\mathrm{NN}}}{\partial q} \right), \qquad (9)$$

with $c_0 = 3(3\pi^2)^{2/3}/10$, and

$$\frac{\delta T^{\mathrm{KS}}[\rho^{\mathrm{KS}}]}{\delta\rho}$$

$$= \frac{1}{\rho^{\mathrm{KS}}} \sum_i f_i \left[-\frac{1}{2}\phi_i^* \nabla^2 \phi_i + (\varepsilon^{\mathrm{HOKS}} - \varepsilon_i)|\phi_i(\boldsymbol{r})|^2 \right],$$

(10)

where $\varepsilon^{\mathrm{HOKS}}$ is the highest-occupied KS eigenvalue, and ε_i and f_i are the KS eigenvalue and occupation number of the ith KS orbital, and $\rho^{\mathrm{KS}} = \sum_i f_i |\phi_i|^2$, respectively [23]. Instead of directly learning $F^{\mathrm{NN}}(s^2, q)$ from the training data, we minimize the cost function Eq. (8) between KEFDs because we observe that the functional derivative is poorly reproduced when KEDF is trained [40] and this training fails to optimize the electron density in some cases. This problem in the functional derivative leading to the erroneous solution of Euler equation is also reported in the previous efforts to develop KEDF with machine learning [41–47], where the training set is the KEDF itself. As the training and test sets in the deep learning, we adopt the *kinetic-energy functional derivative* (KEFD) at each real-space grid point. It is noteworthy that the KEFD, $[\delta T[\rho]/\delta\rho](\boldsymbol{r})$, appears directly in the Euler equation Eq. (2) and thus assures accuracy of the solution of the Euler equation, which is crucial to reproduce the target electron density $\rho(\boldsymbol{r})$.

The training set consists of the KEFD at 13 824 real-space grid points obtained for diamond. For efficient training, we adopt the stochastic natural gradient descent (SNGD), which is a mini-batch training based on natural gradient method [48]. In SNGD, we randomly select N_b training data, then update the ith weight W^i at tth epoch as

$$W^i(t+1) = W^i(t) - \eta \sum_{k=1}^{n_W} [G_{ik} + \nu(t)\mathrm{tr}(\boldsymbol{G})\delta_{ik}]^{-1} \frac{\partial L}{\partial W^k},$$

(11)

where $\boldsymbol{G}$ is a metric tensor defined by

$$G_{ik} = \frac{1}{N_b} \sum_{p=1}^{N_b} \frac{\partial}{\partial W^i}\left[\frac{\delta T^{\mathrm{NN}}(\boldsymbol{r}_p)}{\delta\rho}\right] \frac{\partial}{\partial W^k}\left[\frac{\delta T^{\mathrm{NN}}(\boldsymbol{r}_p)}{\delta\rho}\right]. \quad (12)$$

Here the learning rate η is fixed at $\eta = 0.1$, and n_W is the total number of the NN weight parameters, $\{W^k\}$. To achieve efficient convergence in Eq. (11), we blend the natural gradient and the ordinary gradient by introducing a weighted diagonal matrix $\nu(t)\mathrm{tr}(\boldsymbol{G})\delta_{jk}$, where the scheduling $\nu(t) = \nu_0/(1 + bt)$ with $b = 0.01$ is employed during the epochs being decreased from $\nu_0 = 10^{-5}$. The determination of those training parameters is explained in Appendix C.

The calculations of $\partial L/\partial\boldsymbol{W}$ and Eq. (12) are performed by the backpropagation [49,50]. The algorithm of the backpropagation in the present paper is briefly explained in Appendix D. The training data set was divided into 90% for the training and 10% for the validation.

We have examined the dependence of the accuracy of the obtained $\delta T^{\mathrm{NN}}/\delta\rho$ on the structure of our NN, namely the number of hidden layers and the number of neurons in each layer. The accuracy is assessed by the deviation from the

TABLE I. Accuracy of our KEDF $\delta T^{\mathrm{NN}}/\delta\rho$ depending on the NN structures, i.e., the number of hidden layers and the number of neurons per layer. The deviation from $\delta T^{\mathrm{KS}}/\delta\rho$ is represented in terms of the root mean square error (RMSE) in Ha (atomic unit of energy).

# of hidden layers	# of neurons of per layer	# of weight parameters	RMSE
1	5	21	0.356
1	10	41	0.310
1	15	61	0.302
1	20	81	0.306
2	5	51	0.304
2	10	151	0.311
3	5	81	0.296

training set $\{\delta T^{\mathrm{KS}}/\delta\rho\}$ at grid points in the real space. The 13 824 grid points in a unit cell of diamond are chosen to construct the training set. Table I shows the root mean square error (RMSE) of $\delta T^{\mathrm{NN}}/\delta\rho$ from $\delta T^{\mathrm{KS}}/\delta\rho$ obtained in several NNs. Hereafter we abbreviate the atomic unit of energy (Hartree) as Ha. From this examination, we have adopted the NN with three hidden layers and five neurons in each layer in the present paper. The actual weight parameters $\boldsymbol{W}$ of our NN is available upon request. The trend that the RMSE decrease with increasing the number of weight parameters observed in Table I is indicative of further improvement of our NN for better KEDF in future.

D. Augmentation of the enhancement factor

The enhancement factor in the present paper is further augmented by requiring rigorous limits for $s \to 0$ and $s \to \infty$: When $s \to 0$, it should be the second-order gradient expansion [51] as $1 + (5/27)s^2$, whereas when $s \to \infty$ it should be
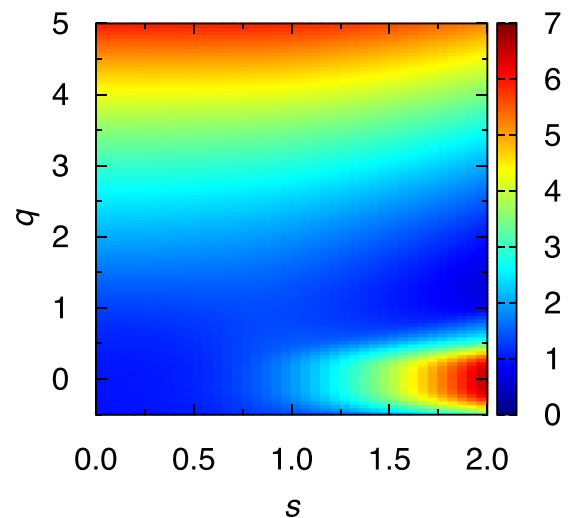

FIG. 2. Contour color map of the enhancement factor NN KEDF $\tilde{F}^{\mathrm{NN}}(s^2, q)$ plotted as a function of s and q within a typical range of s and q ($0 \leqslant s \leqslant 2.0$, $-0.5 \leqslant q \leqslant 5.0$). The abscissa is the reduced density gradient s and the ordinate is the reduced Laplacian of density q.
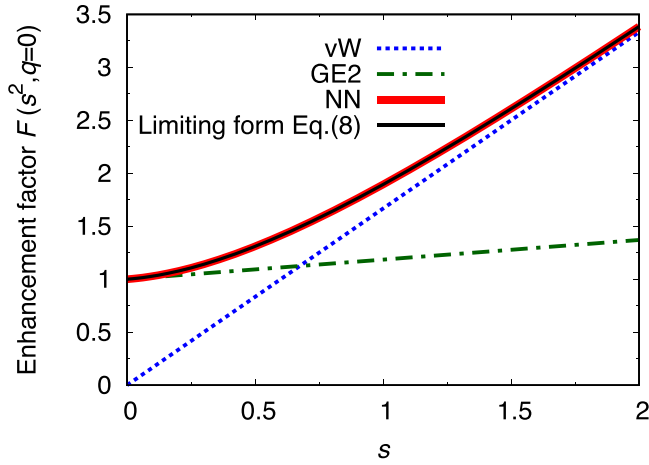
FIG. 3. s-Dependence of the enhancement factors at $q = 0$ within a typical range of s ($0 \leqslant s \leqslant 2.0$). The abscissa is the reduced density gradient s and the enhancement factor $F(s^2, q = 0)$. vW, GE2, and NN denote the enhancement factors of the vW KEDF [8], the second-order gradient expansion [51], and $\tilde{F}^{\mathrm{NN}}$ [the present paper Eq. (14)], respectively.

equal to the vW KEDF [8], $F \rightarrow (5/3)s^2$. A form that satisfies those limits within the first order of q^2 is

$$F^{(0)}(s^2, q) = \frac{5}{3}s^2 + e^{-\alpha s^2} + \beta q^2, \quad (13)$$

with α being 40/27.

Here, β was previously adjusted to reproduce structural properties of target materials [12,13]. Instead, we here determine β so that the second functional derivative of KEDF derived from Eq. (13) in the small-s limit reproduces the Lindhard function [24]. We have found that $\beta = 0.382$ well satisfies this homogeneous-limit condition (for the fitting procedure, see Appendix E).
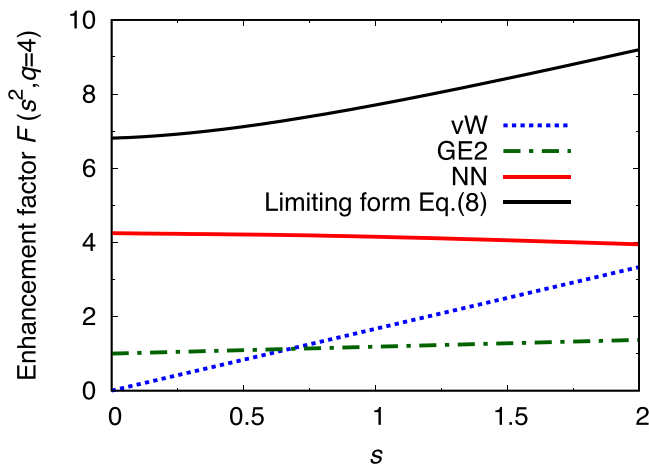

FIG. 4. s-Dependence of the enhancement factors at $q = 4$ within a typical range of s ($0 \leqslant s \leqslant 2.0$). The abscissa is the reduced density gradient s and the enhancement factor $F(s^2, q = 4)$. vW, GE2, and NN denote the enhancement factors of the vW KEDF [8], the second-order gradient expansion [51], and $\tilde{F}^{\mathrm{NN}}$ [the present paper Eq. (14)], respectively.
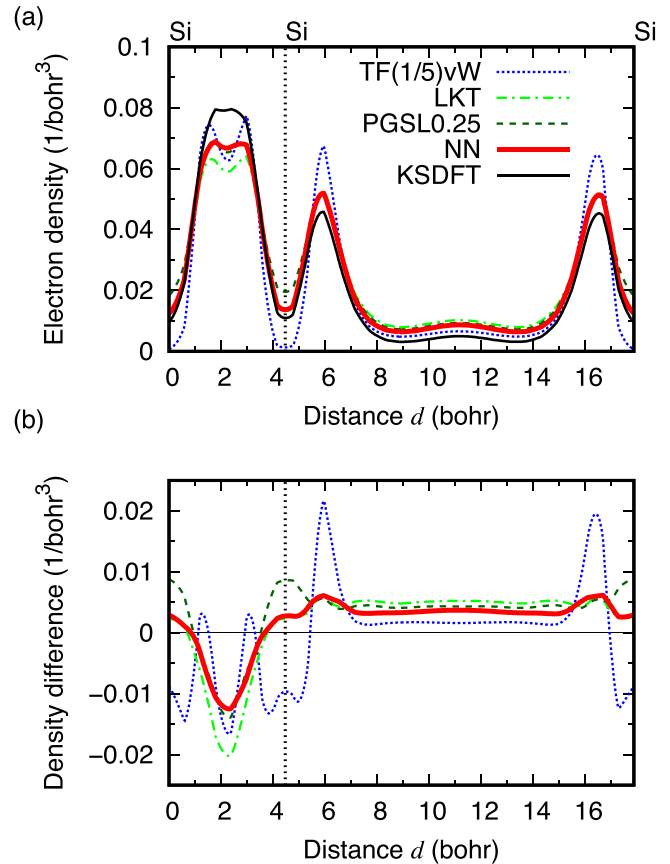

FIG. 5. (a) SCF electron density in ds-Si along [111] direction obtained by different approximations to KEDF. The horizontal axis is the distance d from a Si atom along [111]. Si atoms are located at the positions depicted by the vertical dashed line, and the left and right ends. (b) Difference of densities with respect to the KS density [black-solid line in (a)]. The 0 value on the ordinate is indicated by a horizontal black solid line.

Then we finally propose our enhancement factor,

$$\tilde{F}^{\mathrm{NN}}(s^2, q) = X(q)F^{(0)}(s^2, q) + [1 - X(q)]F^{\mathrm{NN}}(s^2, q), \quad (14)$$

where $X(q) = \exp(-Aq^4)$ is an interpolation function between the small-q and large-q subsystems. This is to enhance the accuracy especially for large q obtained by the flexibility of the NN, following the concept of subsystem functionals in DFT [52–55]. Our final KEDF $\tilde{T}^{\mathrm{NN}}$ is given by Eq. (5) with F^{NN} being replaced by $\tilde{F}^{\mathrm{NN}}$.

The parameter A can be optimized by minimizing the cost function L by expanding the metric tensor G_{jk} to $(n_W + 1) \times (n_W + 1)$ dimensions by adding the components related to $\partial[\delta\tilde{T}^{\mathrm{NN}}(\mathbf{r}_p)/\delta\rho]/\partial A$. The optimized A, however, shows multi-minima structure with nearly flat dependence of L on A in the range $10 < A < 32$, causing uncertainty in the optimization. Then we additionally minimize the difference between the electron density from our KEDF and the KS electron density, which more reliably settles A at 31.62 (see Appendix F for details).

TABLE II. RMSE of SCF density in $10^{-2} \times \text{bohr}^{-3}$ with respect to the KS density in the periodic systems obtained by 5 different KEDFs. The right end column ("ratio") indicates the ratio of RMSE averaged over all the systems to the averaged RMSE obtained by NN KEDF.

	diamond	graphene	ds-Si	fcc-Si	β-tin Si	3C-SiC	bcc-Li	fcc-Al	fcc-Cu	bcc-Na	NaCl	ratio
NN	1.1450	0.6467	0.3580	0.1242	0.2370	0.6180	4.1915	0.0971	8.4183	2.8791	2.1998	1
NN[bare]	0.9370	0.4807	0.3050	0.0889	0.2120	0.6110	2.1484	0.0800	8.7775	1.6449	1.2761	0.777
PGSL0.25	1.2850	0.7097	0.4350	0.1583	0.2580	0.7140	5.3014	0.1205	11.193	4.1229	3.4720	1.255
LKT	1.6760	0.6665	0.5110	0.1113	0.3120	0.9630	5.5690	0.1130	12.866	4.3014	3.7826	1.357
TF(1/5)vW	2.0560	0.3151	0.5500	0.5755	0.4850	1.2010	5.7562	0.3488	25.770	5.0822	3.1791	2.153

III. RESULTS AND DISCUSSION

A. Behavior of NN kinetic-energy density functional

Figure 2 shows our NN enhancement factor $\tilde{F}^{\text{NN}}(s^2, q)$, Eq. (14), for a typical range of s and q. It is clear that the NN enhancement factor is continuous in (s^2, q)-plane, indicating that $\tilde{F}^{\text{NN}}$ smoothly connects F^{NN}, Eq. (6), and the limiting form $F^{(0)}$, Eq. (13). To compare $\tilde{F}^{\text{NN}}$, $F^{(0)}$, $F^{\text{vW}} = (5/3)s^2$ (the vW KEDF [8]), and also the second-order gradient expansion $F^{\text{GE2}} = 1 + (5/27)s^2$ [51]), we plotted the s-dependence of these enhancement factors for two fixed values of q, namely $q = 0$ and $q = 4$ as the typical large q value in most systems, in Figs. 3 and 4. At $q = 0$, $\tilde{F}^{\text{NN}}$ recovers the limiting form $F^{(0)}$ as expected. The behavior of $\tilde{F}^{\text{NN}}$ is substantially different from the limiting form $F^{(0)}$ at $q = 4$, which is ascribed to the form of NN KEDF, Eq. (6), and relevant to its accuracy.

B. Accuracy of NN KEDF

1. Electron densities

We assess the accuracy of KEDFs with the self-consistent-field (SCF) density obtained by minimizing the total energy with the developed KEDF. One way is the direct minimization with iterative techniques [56–59]. The other way is to solve the Euler equation Eq. (2) by the matrix diagonalization, which we adopt in this paper. By introducing vW KEDF [8], $T_{\text{vW}}[\rho]$, which satisfies $\delta T_{\text{vW}}/\delta \rho = -(\nabla^2 \sqrt{\rho})/(2\sqrt{\rho})$, the Euler equation becomes a Schrödinger-type equation for $\sqrt{\rho}$ [60]. However, it is recognized that the diagonalization of such Schrödinger-type equation suffers from ill convergence [61]. We have found that this difficulty can be circumvented in all KEDFs we considered by introducing a parameter λ and by performing the rearranged diagonalization scheme. Previously, this sort of rearrangement has been applied to the TF(λ)vW functional with $\lambda = 1/5$ or $1/9$ [62]. We first express KEDF as the sum of scaled vW functional

and the rest part:

$$T_s[\rho] = \lambda T_{\text{vW}}[\rho] + T_r[\rho]. \tag{15}$$

Then Eq. (2) becomes

$$\left[-\frac{\nabla^2}{2} + \frac{1}{\lambda}\left(\frac{\delta T_r[\rho]}{\delta \rho}(\boldsymbol{r}) + v_{\text{KS}}([\rho];\boldsymbol{r})\right)\right]\sqrt{\rho(\boldsymbol{r})} = \frac{\mu}{\lambda}\sqrt{\rho(\boldsymbol{r})}, \tag{16}$$

where $v_{\text{KS}} = v_{\text{ext}} + \delta E_{\text{H}}/\delta \rho + \delta E_{\text{xc}}/\delta \rho$ is the usual KS potential. In our scheme, we have found that $\lambda = 10$ leads to a satisfactory convergence in the SCF calculations. Note that any choice of λ offers the same correct SCF solution, if it converges.

Figure 5 shows calculated SCF electron density obtained by our NN KEDF. For comparison, the computed densities using the KEDFs in the past, i.e., PGSL0.25 [12] defined by Eq. (13) with $\beta = 0.25$, LKT [14], and the conventional TF(1/5)vW (TF(λ)vW KEDF with $\lambda = 1/5$) [63] are also shown. Figure 5(a) is the electron density along [111] direction in diamond-structured (ds)-Si and Fig. 5(b) shows its difference from the KS density. The obtained electron density shows overall superiority of our NN KEDFs against PGSL0.25 and TF(1/5)vW most clearly visible at the nuclear site ($d = 4.47$ bohr). When compared with LKT, NN tends to be more accurate in the bonding region ($0 < d < 4.47$ bohr). We have also calculated the electron densities with our NN KEDF of other 23 systems (10 solids, 6 diatomic molecules, and 7 atoms), and found that the obtained densities satisfactorily reproduce the corresponding KS densities (see Supplemental Material [64] for details). It is noteworthy that our NN KEDF obtained by the training data of only ds-C shows high transferability for various systems.

The superiority of our NN KEDF in reproducing the KS densities is quantified by the RMSE of the SCF density ρ_{scf} with respect to the KS density ρ^{KS} for all the 24 systems examined in the present paper. Table II shows

TABLE III. RMSE of SCF density in $10^{-2} \times \text{bohr}^{-3}$ with respect to the KS density in atoms. The right column ("ratio") indicates the ratio of RMSE to NN averaged over all systems.

	Li	C	Na	Al	Si	Cl	Cu	ratio
NN	1.1155	0.1469	1.2508	0.0271	0.0481	0.2943	2.2284	1
NN[bare]	0.7003	0.1253	0.9795	0.2322	0.0415	0.2250	2.7214	1.9544
PGSL0.25	1.6069	0.3482	1.7854	0.2686	0.1527	0.4664	4.4330	3.2613
LKT	2.1197	0.2540	1.6340	0.0532	0.0776	0.4145	5.0251	1.7393
TF(1/5)vW	1.3733	0.1315	1.2050	0.0329	0.0635	0.2832	1.3952	1.0302

TABLE IV. RMSE of SCF density in $10^{-2} \times$ bohr^{-3} with respect to the KS density in diatomic molecules. The right column ("ratio") indicates the ratio of RMSE to NN averaged over all systems.

	Li$_2$	C$_2$	Na$_2$	Al$_2$	Si$_2$	Cl$_2$	ratio
NN	0.7036	0.3545	1.4726	0.2066	0.1000	0.5534	1
NN[bare]	0.4036	0.2258	1.3239	0.2137	0.1047	0.4204	0.8250
PGSL0.25	0.6932	0.3919	1.6260	0.1882	0.2661	0.5130	1.2823
LKT	0.7616	0.6422	2.4138	0.1548	0.2108	0.7267	1.4505
TF($\frac{1}{5}$)vW	0.4342	0.7200	3.8073	0.1618	0.2019	0.4080	1.4620

the RMSE for periodic systems, including semiconductors, metals, and an ionic material with some of their polytypes: diamond, graphene, ds-Si, face-centered-cubic(fcc)-Si, β-tin Si, zincblende(3C)-SiC, body-centered-cubic (bcc)-Li, fcc-Al, fcc-Cu, bcc-Na, and NaCl. For comparison, the results obtained with the KEDFs in the past are also shown. We also present the results by the uninterpolated NN enhancement factor F^{NN} [i.e., $A \rightarrow \infty$ in Eq. (14)] (labelled as NN[bare]) trained by the same scheme as NN. The overall good performance of NN is demonstrated by the RMSE for each material. The RMSE averaged over all 11 materials (the right end column "ratio" in Table II) clearly shows the superiority of the present NN KEDF to other KEDFs in the past.

The RMSE of the SCF density has been also computed for 13 isolated systems, Li, C, Na, Al, Si, Cl, and Cu atoms, and Li$_2$, C$_2$, Na$_2$, Al$_2$, Si$_2$, and Cl$_2$ molecules, and then averaged over all the atoms and molecules, as shown in Table III and IV. The same RMSE "ratio" indicates that NN and TF(1/5)vW are most accurate in atoms, whereas NN and NN[bare] are most accurate in molecules.

These results indicate that we have succeeded to construct NN KEDFs that outperform previous ones without resorting to system-dependent parameters. The electron density calculated with NN[bare] shows better performance than NN in some cases. However, physical quantities shown below indicate the limitation of NN[bare]. The present NN up to $\ell_1 = 1$ makes the agreement with KS at the first-order derivative of

T, while the density is a quantity related at the first-order level $\rho(\mathbf{r}) = -\delta E[\rho]/\delta\mu(\mathbf{r})$ where $\mu(\mathbf{r})$ is the local chemical potential. Hence the success in reproducing the KS density is intrinsic to the present scheme.

The NN adopted in the present paper produces the minimum value of the cost function $L_{\min} = 3.519 \times 10^{-2}$ Ha^2. We have examined other NNs, which produce the cost function of $L_{\min} < L \leqslant 1.8 L_{\min}$ Ha^2. We have confirmed that the obtained RMSE of the electron density with such NN increases less than 1% at most.

2. Structural properties

To further examine the validity of our OFDFT scheme with the NN KEDF, we have calculated structural properties and energetics for test systems: the lattice constants a_0, bulk moduli B_0, and cohesive energies E_{coh} of 11 solids, and the bond lengths r_0 of 6 molecules.

Tables V and VI show the lattice constants and bulk moduli for 11 solids along with the corresponding values obtained using other KEDFs in the past. We compare the relative errors with respect to the KS values (numbers in the parentheses in Tables V and VI). Our NN KEDF provides the smallest relative errors in 10 cases (a_0 for β-tin Si, fcc-Al, fcc-Cu, graphene; B_0 for diamond, fcc-Si, β-tin Si, fcc-Al, fcc-Cu, NaCl), whereas the number of the cases with the smallest errors are 6 for LKT and 5 for PGSL0.25. The overall quantitative index of the superiority is evaluated as the mean absolute relative errors (MAREs) with respect to the KS values for those quantities. The NN KEDF clearly shows the minimum MAREs for both a_0 and B_0 (Table VI), indicating its superior performance. For all structural properties, NN[bare] produces larger MAREs than the NN functional. This indicates the importance of augmenting the enhancement factor and the validity of concept of the subsystem DFT [52]. It is noteworthy that our NN KEDF functional is trained by a part of the data in diamond and reproduces KSDFT results reasonably for the structural properties of a variety of materials including metals and ionic crystals with no adjustable parameters, where the machine learning parameters are

TABLE V. Comparison of equilibrium lattice constant (a_0) in Å and bulk moduli (B_0) in GPa obtained by NN and NN[bare] with those obtained by different approximations to KEDF. Numbers in the parentheses are the relative errors in % with respect to the KS values. In β-tin Si, the ratio of another lattice constant c_0 to a_0 (c_0/a_0) is also listed. Some values are left as blanks because the total energy monotonically decreases with respect to the volume expansion.

	diamond		ds-Si		fcc-Si		β-tin Si			3C-SiC		bcc-Li		fcc-Al	
	a_0	B_0	a_0	B_0	a_0	B_0	a_0	c_0/a_0	B_0	a_0	B_0	a_0	B_0	a_0	B_0
NN	3.428	411	5.367	85.4	3.642	143	4.673	0.529	165	3.092	157	3.479	15.5	4.118	79.1
	(−2.53)	(25.7)	(−0.46)	(−14.6)	(−0.22)	(5.93)	(0.17)	(−1.67)	(5.10)	(2.28)	(−36.2)	(−0.83)	(4.03)	(1.68)	(2.06)
NN[bare]	3.130	867	5.187	83.5	3.529	149	–	–	–	2.904	253	3.329	18.9	3.841	129
	(−11.0)	(165)	(−3.80)	(−16.5)	(−3.32)	(10.4)				(−3.94)	(2.85)	(−5.10)	(26.8)	(−5.16)	(66.5)
PGSL0.25	3.430	433	5.384	93.4	3.702	118	4.744	0.529	137	3.073	217	3.496	15.5	4.197	67.4
	(−2.47)	(32.4)	(−0.15)	(−6.6)	(1.42)	(−13.0)	(1.69)	(−1.67)	(−12.7)	(1.65)	(−11.8)	(−0.33)	(3.78)	(3.62)	(−13.0)
LKT	3.343	578	5.336	104	3.644	161	4.643	0.532	186	3.066	227	3.492	15.3	4.144	86.0
	(−4.95)	(76.8)	(−1.04)	(4.00)	(−0.16)	(19.5)	(−0.47)	(−1.12)	(18.5)	(1.42)	(−7.72)	(−0.46)	(2.52)	(2.32)	(10.9)
TF(1/5)vW			5.708	39.6	3.847	54.0		–		3.353	63.2	3.384	16.4	4.243	44.0
			(5.86)	(−60.4)	(5.40)	(−60.0)				(10.9)	(−74.3)	(−3.54)	(10.1)	(4.76)	(−43.2)
KSDFT	3.517	327	5.392	100	3.650	135	4.665	0.538	157	3.023	246	3.508	14.9	4.050	77.5

TABLE VI. Comparison of equilibrium lattice constant (a_0) in Å and bulk moduli (B_0) in GPa obtained by NN and NN[bare] with different approximations to KEDF. Numbers in the parentheses are the relative errors in % with respect to the KS values. Mean absolute relative errors (%) with respect to the KS values averaged over the systems including the ones in Table V are also listed as MARE.

	fcc-Cu		bcc-Na		NaCl		graphene	MARE	
	a_0	B_0	a_0	B_0	a_0	B_0	a_0	a_0	B_0
NN	3.730	169	4.227	7.98	5.678	27.7	2.448		
	(2.33)	(6.96)	(−0.91)	(3.50)	(3.61)	(6.94)	(0.29)	1.39	11.1
NN[bare]	3.887	228	4.045	9.73	5.187	32.4	2.377		
	(6.64)	(44.3)	(−5.18)	(26.2)	(−5.35)	(25.1)	(−2.62)	4.74	38.4
PGSL0.25	3.795	138	4.250	8.00	5.595	22.9	2.433		
	(4.12)	(−12.4)	(−0.38)	(3.75)	(2.10)	(−11.5)	(−0.32)	1.66	12.1
LKT	3.762	175	4.245	7.91	5.596	23.8	2.402		
	(3.21)	(10.7)	(−0.50)	(2.61)	(2.12)	(−8.20)	(−1.59)	1.66	14.5
TF(1/5)vW	3.799	88.4	4.116	8.49	6.056	6.75	2.593		
	(4.21)	(−44.1)	(−3.51)	(10.1)	(10.5)	(−73.9)	(6.24)	6.10	47.0
KSDFT	3.645	158	4.266	7.71	5.480	25.9	2.441		

determined uniquely in an *ab initio* fashion after minimizing the cost function.

Table VII shows the bond lengths of 6 molecules. The MAREs with respect to the KS values show that NN KEDF performs overall better than other KEDFs. However, the obtained MAREs of the bond lengths of the molecules are certainly larger than the MAREs of the lattice constants of the solids. This is presumably due to the fact that we have trained our NN using the data of the solid, diamond-structured carbon. This demonstrates the difficulty to develop the KEDF quantitatively valid for both localized and delocalized systems. Table VIII shows the calculated cohesive energies (E_{coh}) for 11 solids. The chemical trend obtained in the KS scheme is reproduced by our NN KEDF. However, the quantitative reproduction of the KS values is not satisfactory. In fact, all the KEDFs including those in the past provide MAREs of the cohesive energy larger than 20% (the right-end column). Among them, the NN KEDF keeps the smallest MAREs.

The lattice constant and the bulk modulus are obtained by the behavior of the total energy E as a function of the volume V around its minimum point: a_0 is obtained as the zero point of $\delta E/\delta V$ and B_0 requires the second derivative $\delta^2 E/\delta V^2$. In our OFDFT with the NN KEDF with $\ell_1 = 1$,

the first derivative of E with respect to the density is trained. Hence the high reproducibility of a_0 may be intrinsic in the present scheme, while it is natural that the MAREs for B_0 are worse than those for a_0. Along this line, the machine learning up to $\ell_1 = 2$ is desired for physical quantities associated with the second derivative of E such as B_0. Since B_0 calculated with $\ell_1 = 1$ already shows fair agreements with that by KS-DFT and experiments, the higher-order machine learning has potential to provide us with systematically more accurate KEDF.

C. Order-N DFT computations

Finally, we have analyzed the computational time of both KSDFT and OFDFT in order to demonstrate our $\mathcal{O}(N)$ scheme. We decompose the computational time for a single SCF iteration (t_{SCF}) as $t_{SCF} = t_1 + t_2 + t_3 + t_{others}$, where t_1 is the time for the subspace diagonalization (SD), conjugate-gradient minimization (CG) and Gram-Schmidt (GS) orthonormalization, t_2 for updating density and calculating the total energy, t_3 for the mixing procedure to obtain the new input potential, and t_{others} for other procedures such as MPI gathering and broadcasting eigenvectors. The test systems are 4H-SiC supercells with sizes of 576, 1024, 1600,

TABLE VII. Comparison of the equilibrium bond length (r_0) in Å for molecules. Numbers in the bottom parentheses are the relative errors in % with respect to the KS values. Mean absolute relative errors in % with respect to the KS values are also listed as MARE.

	Li$_2$	C$_2$	Na$_2$	Al$_2$	Si$_2$	Cl$_2$	MARE
NN	3.118	1.425	2.913	2.608	2.325	1.966	
	(8.04)	(18.5)	(−5.56)	(−4.32)	(5.26)	(5.20)	7.81
NN[bare]	2.332	1.414	2.995	2.886	2.243	1.964	
	(−19.2)	(17.6)	(−2.90)	(5.90)	(1.54)	(3.54)	8.45
PGSL0.25	2.970	1.388	3.369	2.815	2.497	1.966	
	(2.91)	(15.4)	(9.23)	(3.31)	(13.0)	(3.62)	7.92
LKT	2.846	1.391	3.211	3.153	1.974	1.777	
	(−1.37)	(15.7)	(4.12)	(15.7)	(−10.6)	(−6.32)	8.96
TF(1/5)vW	3.011	1.489	3.042	2.505	2.582	2.020	
	(4.33)	(23.9)	(−1.37)	(−8.09)	(16.9)	(6.49)	10.2
KSDFT	2.886	1.203	3.084	2.725	2.209	1.897	

TABLE VIII. Comparison of cohesive energies (E_{coh}) for solids in Ha. Numbers in the parentheses are the relative errors in % with respect to the KS values. The mean absolute relative errors (MAREs) in % with respect to the KS values are also listed.

	diamond	graphene	ds-Si	fcc-Si	β-tin Si	3C-SiC	bcc-Li	fcc-Al	fcc-Cu	bcc-Na	NaCl	MARE
NN	−0.3544	−0.3038	−0.3714	−0.2036	−0.3710	−0.3445	−0.0668	−0.1258	−0.1073	−0.0443	−0.3068	
	(22.1)	(40.2)	(14.8)	(8.55)	(15.2)	(32.2)	(7.1)	(19.1)	(16.6)	(21.0)	(22.9)	20.0
NN[bare]	−0.3517	−0.2332	−0.3731	−0.2416	−0.4186	−0.3341	−0.0926	−0.5550	−0.1786	−0.0692	−0.4944	
	(22.7)	(54.1)	(14.4)	(8.50)	(4.38)	(34.3)	(48.4)	(425)	(94.0)	(89.0)	(98.1)	81.2
PGSL0.25	−1.2546	−0.7595	−0.7839	−0.4581	−0.8854	−1.0818	−0.2079	−0.6102	−0.2471	−0.0992	−0.6907	
	(176)	(49.6)	(79.8)	(106)	(102)	(113)	(233)	(477)	(168)	(171)	(177)	168
LKT	−1.1272	−0.5836	−0.6768	−0.4422	−0.8340	−0.9147	−0.2134	−0.2770	−0.1856	−0.0771	−0.5235	
	(148)	(14.9)	(55.2)	(98.6)	(90.5)	(79.9)	(242)	(162)	(102)	(111)	(110)	110
TF(1/5)vW	−0.3860	−0.1877	−0.3634	−0.2187	−0.4148	−0.4257	−0.0664	−0.1369	−0.1158	−0.0499	−0.3484	
	(15.2)	(137)	(16.7)	(1.77)	(5.25)	(16.3)	(6.51)	(29.5)	(25.8)	(36.3)	(39.6)	30.0
KSDFT	−0.4551	−0.5077	−0.4361	−0.2226	−0.4377	−0.5084	−0.0624	−0.1057	−0.0921	−0.0366	−0.2496	

2400, and 4704 atoms. Only gamma point is sampled for the Brillouin zone integration in KSDFT. The grid-spacing is chosen as 0.39 Å. We use 36 eigenvectors for the diagonalization of Eq. (16).

Figure 6(a) shows the results. The predominant computational time during a single SCF iteration in KSDFT is t_1, whereas it is t_2 or t_3 in OFDFT. The latter scales with $\mathcal{O}(N^{1.2})$ in our numerical confirmation. The overall computational time in our OFDFT scheme and in our KS-scheme using RSDFT code [30–32], which may be the fastest available code, is shown in Fig. 6(b), indicating that the present OFDFT scheme has now achieved essentially the $\mathcal{O}(N)$ scaling.

Several $\mathcal{O}(N)$ density-functional calculations have been proposed and developed in the past. One primitive way is to introduce localized-orbital basis sets to express Kohn-Sham equations and then truncate the overlap of localized orbitals in the actual computations [65,66]. This is obviously not based on a legitimate principle but relies on the incompleteness of the basis set practically. Another scheme is based on the "near-sightedness principle" [67] of many-electron systems, which states that principal quantities to describe physical properties are essentially local. One of such quantities may be the density matrix and the $\mathcal{O}(N)$ scheme with the density matrix has been developed [68]. Yet in the actual computations, one need to truncate the density matrix in real space, which is actually a system dependent procedure. On the other hand, OFDFT does not require such system-dependent procedure once the suitable kinetic-energy functional is developed. In this sense, OFDFT has a potential to become the most legitimate and practical $\mathcal{O}(N)$ scheme, which is applicable to a broad range of materials on the equal footing.
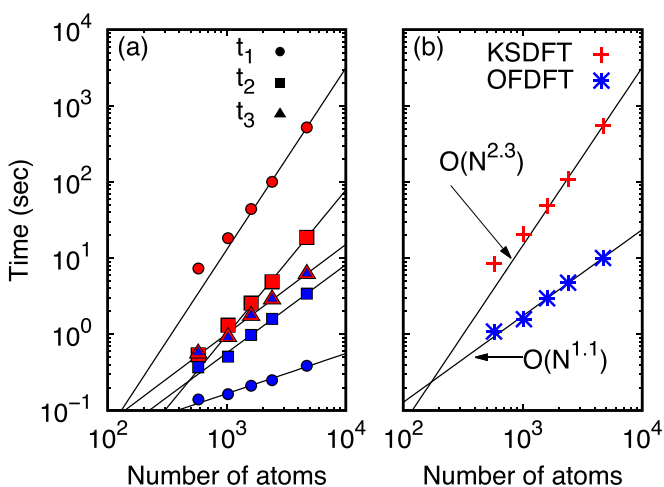

FIG. 6. Computational time of KSDFT and OFDFT. Comparison of (a) t_1, t_2, and t_3; and (b) t_{SCF}. The test system is 4H-SiC supercell with the number of atoms of 576–4704. In both (a) and (b), data points obtained by KSDFT and OFDFT are plotted in red and blue colors, respectively. Each set of data points has been fitted by AN^β with parameters A and β, where N is the number of atoms, and the fitted results are plotted as black solid lines. In (a), these fittings scale as $\mathcal{O}(N^\beta)$ where $\beta = 2.4, 1.9, 1.2, 1.2, 0.5$ from the top line to the bottom line. t_1, t_2, and t_3 are plotted as circles, squares, and triangles, respectively.

IV. CONCLUSION

We have developed a scheme of the orbital-free density-functional-theory (OFDFT) calculations based on the accurate and transferrable kinetic-energy density functional (KEDF), which is created in an unprecedented way using appropriately constructed neural network (NN). Our OFDFT scheme has reproduced the electron density obtained in the state-of-the-art DFT calculations and then provided accurate structural properties of 24 different systems, ranging from atoms, molecules, metals, semiconductors, and an ionic material. The accuracy and the transferability of our KEDF have been achieved by our NN training system in which the kinetic-energy functional derivative (KEFD) at each real-space grid point in diamond-structured carbon is used as the training data. The choice of the KEFD as the training data is essential in the following sense: First, it appears directly in the Euler equation, which one should solve and it allows us a transparent and intuitive understanding, where the density and its derivatives are the primary and fundamental quantities in the spirit of DFT. Second, its leaning assists in reproducing the physical quantity expressed as the first derivative of the total energy. More generally, the present development of KEDF $T[\rho]$ is in the line of systematic expansion in terms of the functional derivatives $\delta^{\ell_1} T/\delta\rho^{\ell_1}$ through progressive increase of ℓ_1. The present numerical success has demonstrated the validity of

this approach with the detailed results for case of $\ell_1 = 1$. The computational cost of the present OFDFT scheme for the system size N has indeed shown the scaling of $\mathcal{O}(N)$ inherent to OFDFT, as is evidenced by the computations of SiC consisting of thousands of Si and C atoms, which are important in developing power-electronics devices.

APPENDIX A: CONSTRUCTION OF *AB INITIO* LOCAL IONIC PSEUDOPOTENTIALS

In this Appendix, we explain our scheme to generate *ab initio* local ionic pseudopotentials (LIPSs), its application to 7 elements, lithium, carbon, sodium, aluminum, silicon, chlorine, and copper, and the validity and the transferability of the generated NLPSs.

Local pseudopotentials combined with OFDFT are developed in the pioneering paper by Carter and her collaborators [22,36–38]. They are called "bulk-derived local pseudopotential" (BLPS) since the potentials contain parameters, which are optimized to reproduce structural properties of target bulk materials: The parameters are (1) the value of the non-Coulombic part of the BLPS in reciprocal space at $G = 0$ and (2) the cutoff radius r_c beyond which the Coulombic tail is imposed on the BLPS in real space. In the construction of BLPS, the KSDFT calculations with NLPS are first performed to obtain the bulk modulus B_0, equilibrium volumes V_0, and also the energy ordering of various phases. Then the above parameters in BLPS are adjusted to reproduce those structural characteristics obtained by the KSDFT calculations. Such fitting is also improved by minimizing the difference between forces obtained by the BLPS with OFDFT and those by an NLPS with KSDFT [38]. Currently, these BLPSs are available for Al, As, Ga, In, Li, Mg, P, Sb, Si [69]. After the appropriate fitting, the bulk properties calculated from the BLPSs are in good agreement with those by NLPSs [37].

In spite of relatively good performance of BLPSs, the above mentioned fitting process is manual and cumbersome. The iterative improvement of the BLPS is unavoidable until required target properties can be obtained within acceptably small errors. Furthermore, the previous procedure

to solve KS equation inversely and obtain the effective potential [22,36,37] using the Wang-Parr method [70] or the Wu-Yang method [71] require good initial guess for the effective potential.

In our paper, we aim to eliminate the fitting parameters and to circumvent the iterative fitting procedure. In our approach, the LIPS can be treated solely in reciprocal space and thus there is no need to perform Fourier-Bessel transform, which requires elaborate interpolation in reciprocal space.

The first step is to find an effective potential $V_{\mathrm{eff}}(\boldsymbol{r})$ that reproduces a target electron density generated by the KSDFT calculations using NLPSs. Among many ways to invert the KS equation, we adopt the Kadantsev-Stott method [39] based on the Haydock-Foulkes variational principle [72], which does not require a particularly good initial guess for the local pseudopotential. In this method, a functional $\Upsilon[V_{\mathrm{eff}}(\boldsymbol{r})]$ is defined as

$$\Upsilon[V_{\mathrm{eff}}(\boldsymbol{r})] = -\sum_{i:\mathrm{occ}} \epsilon_i[V_{\mathrm{eff}}(\boldsymbol{r})] + \int V_{\mathrm{eff}}(\boldsymbol{r})\rho_{\mathrm{target}}(\boldsymbol{r})d\boldsymbol{r}, \quad (A1)$$

where ϵ_i are eigenvalues corresponding to the KS orbitals $\{\phi_i\}$, and $\rho_{\mathrm{target}}(\boldsymbol{r})$ is the target electron density obtained by the KSDFT calculation with NLPSs, which is to be reproduced here. The Haydock-Foulkes variational principle ensures that the true effective potential corresponding to the target density satisfies the stationary condition $\delta\Upsilon[V_{\mathrm{eff}}]/\delta V_{\mathrm{eff}} = 0$.

In the kth iteration of the minimization of $\Upsilon[V]$, the KS equation is solved inversely using the $V_{\mathrm{eff}}^{(k-1)}(\boldsymbol{r})$ in the previous iteration to obtain the density $\rho^{(k)}(\boldsymbol{r})$. We note that the functional derivative of Υ satisfies an identity $\delta\Upsilon[V]/\delta V = -\rho(\boldsymbol{r}) + \rho_{\mathrm{target}}(\boldsymbol{r})$. Hence in the kth iteration, $V_{\mathrm{eff}}^{(k)}(\boldsymbol{r})$ is explored along a line defined by

$$V_{\mathrm{eff}}^{(k)}(\boldsymbol{r}) = V_{\mathrm{eff}}^{(k-1)}(\boldsymbol{r}) - \alpha_1 \left.\frac{\delta\Upsilon[V]}{\delta V}\right|_{\rho=\rho^{(k)}}, \quad (A2)$$

and α_1 is determined by the line search. We iteratively continue this process until σ,

$$\sigma = \sqrt{\frac{1}{\Omega} \int_\Omega |\rho^{(k)}(\boldsymbol{r}) - \rho_{\mathrm{target}}(\boldsymbol{r})|^2 d\boldsymbol{r}}, \quad (A3)$$

becomes less than the preset value. We have implemented this variational minimization method in the RSDFT code [30–32].

In actual computations, we first generate the target bulk electron densities by solving the KS equations using NLPSs. We then invert the KS equation using these target bulk electron densities to obtain the local effective potentials $V_{\mathrm{eff}}(\boldsymbol{r})$ for these targets using the Kadantsev-Stott method explained above. The settings of k-point grids and real-space grid spacing used in the inversion are the same as those used in generating the target bulk electron densities.

Next, the Hartree potential $V_{\mathrm{H}}(\boldsymbol{r})$ and exchange-correlation potential $V_{\mathrm{xc}}(\boldsymbol{r})$ among valence electrons are subtracted from $V_{\mathrm{eff}}(\boldsymbol{r})$ to extract the local ionic potential for the bulk:

$$V_{\mathrm{bulk}}^{\mathrm{ion}}(\boldsymbol{r}) = V_{\mathrm{eff}}(\boldsymbol{r}) - V_{\mathrm{H}}(\boldsymbol{r}) - V_{\mathrm{xc}}(\boldsymbol{r}). \quad (A4)$$

We then convert $V_{\mathrm{bulk}}^{\mathrm{ion}}(\boldsymbol{r})$ into the atom-centered local ionic pseudopotential in reciprocal space $\tilde{V}_{\mathrm{loc}}^{\mathrm{ion}}(G)$ as follows. First,

TABLE IX. Parameters of local ionic pseudopotentials for Li, C, Na, Al, Si, Cl, and Cu.

	c_1^{core}	α_1^{core}	α_2^{core}	r_c	C_1	C_2	C_3	C_4	C_5
Li	1.0000	3.1250		0.4000	-3.12247	-5.29585	1.29259	-0.0299128	0
C	1.0000	1.4635		0.5845	8.4107	-13.007	4.8809	-0.6743	0.02793
Na	1.0000	1.1619		0.6560	-3.85643	-8.09377	2.90894	-0.190011	0
Al	1.0000	0.5732		0.9340	3.22841	-1.4132	0.147102	-0.00494713	0
Si	1.6054	2.1600	0.8600	0.7999	9.0231	-3.7692	0.5453	-0.02952	0
Cl	1.0000	1.3171		0.616128	5.29287	-2.12203	0.169072	-0.014369	0
Cu	1.0000	1.5097		0.5755	-2.58512	-17.0765	5.29496	-0.31904	0

$V_{bulk}^{ion}(\boldsymbol{r})$ is Fourier-transformed into reciprocal space as

$$V_{bulk}^{ion}(\boldsymbol{G}) = \frac{1}{\Omega} \int_\Omega V_{bulk}^{ion}(\boldsymbol{r}) e^{i\boldsymbol{G}\cdot\boldsymbol{r}} d\boldsymbol{r}, \tag{A5}$$

where Ω is the unit cell volume. Then using the structure factor $S(\boldsymbol{G})$ of the target bulk material, we obtain the Fourier transformed atom-centered ionic pseudopotential:

$$V^{ion}(\boldsymbol{G}) = \frac{V_{bulk}^{ion}(\boldsymbol{G})}{S(\boldsymbol{G})}. \tag{A6}$$

Finally, by spherically averaging $V^{ion}(\boldsymbol{G})$, we obtain the atom-centered local ionic pseudopotential in reciprocal space, $\tilde{V}_{loc}^{ion}(G)$:

$$\tilde{V}_{loc}^{ion}(G) = \frac{1}{N_G} \sum_{|\boldsymbol{G}_i|=G} V^{ion}(\boldsymbol{G}_i), \tag{A7}$$

where N_G is the number of G vectors with the length being G. The procedure to construct $\tilde{V}_{loc}^{ion}(G)$ explained above ensures that the resulting $\tilde{V}_{loc}^{ion}$ has the proper Coulombic tail due to the nucleus and core electrons in real space. Hence we fit $\tilde{V}_{loc}^{ion}(G)$ numerically obtained above to the form,

$$
\begin{aligned}
\tilde{V}_{loc}^{ion}(G) &= \frac{4\pi}{\Omega} \int_0^\infty r^2 j_0(Gr) \tilde{V}_{loc}^{ion}(r) dr \\
&= -\frac{4\pi Z}{\Omega G^2} \sum_{i=1}^2 c_i^{core} \exp\left[-\frac{G^2}{4\alpha_i^{core}}\right] + \frac{\sqrt{(2\pi)^3}}{\Omega} r_c^3 \exp\left[-\frac{(Gr_c)^2}{2}\right]\{C_1 + C_2[3 - (Gr_c)^2] + C_3[15 - 10(Gr_c)^2 + (Gr_c)^4] \\
&\quad + C_4[105 - 105(Gr_c)^2 + 21(Gr_c)^4 - (Gr_c)^6] + C_5[945 - 1260(Gr_c)^2 + 378(Gr_c)^4 - 36(Gr_c)^6 + (Gr_c)^8]\},
\end{aligned} \tag{A8}
$$

which corresponds to the Fourier-inversed-transform,

$$
\begin{aligned}
\tilde{V}_{loc}^{ion}(r) &= -\frac{Z}{r} \sum_{i=1}^2 c_i^{core} \mathrm{erf}\left[(\alpha_i^{core})^{1/2} r\right] \\
&\quad + \exp\left[-\frac{1}{2}\left(\frac{r}{r_c}\right)^2\right] \sum_{n=1}^5 C_n \left(\frac{r}{r_c}\right)^{2n-2}, \quad (A9)
\end{aligned}
$$

where $c_1^{core} + c_2^{core} = 1$ [35,73].

The KSDFT calculations to obtain the target bulk densities were performed using the RSDFT code [30–32]. We adopted the generalized gradient approximation of Perdew, Burke, and Ernzerhof [33] for the exchange-correlation functional. We used Troullier-Martins (TM) NLPSs [74,75]. As target densities, we chose the densities of ds-Si, ds-C, bcc-Li, bcc-Na, fcc-Al, fcc-Cu, and NaCl with the lattice constants of 5.466 Å, 3.560 Å, 3.430 Å, 4.212 Å, 4.051 Å, 3.634 Å, and 5.694 Å, respectively, all of which are equilibrium values within the KSDFT scheme using NLPSs. The k-point sampling grids were taken to be $4 \times 4 \times 4$ in all calculations. The real-space grid spacing was 0.190 Å for Si, 0.178 Å for C, 0.172 Å for Li, 0.211 Å for Na, 0.169 Å for Al, 0.151 Å for Cu, and 0.190 Å for NaCl.

The convergence criteria of the KS inversion was set to be $\sigma < 3.5 \times 10^{-4}$ bohr^{-3} for Li, $\sigma < 1.0 \times 10^{-3}$ bohr^{-3} for C, $\sigma < 2.0 \times 10^{-5}$ bohr^{-3} for Na, $\sigma < 1.0 \times 10^{-5}$ bohr^{-3} for Al, $\sigma < 1.0 \times 10^{-4}$ bohr^{-3} for Si, $\sigma < 3.0 \times 10^{-5}$ bohr^{-3} for NaCl, and $\sigma < 8.5 \times 10^{-3}$ bohr^{-3} for Cu. We minimized the functional $\Upsilon[V_{eff}(\boldsymbol{r})]$ via the conjugate gradient (CG) method. The obtained parameters for Li, C, Na, Al, Si, Cl, and Cu LIPSs are shown in Table IX.

We now assess the quality of our LIPSs. We compare the electron densities of bcc-Li, fcc-Al, fcc-Cu, bcc-Na, NaCl, ds-Si, ds-C, and zincblende(3C)-SiC calculated with the NLPS and our LIPS. We denote the density obtained by the NLPS as ρ_{NLPS} and that by our LIPS as ρ_{LIPS}. Figures 7, 8, and 9 show the electron densities of the target materials along selected directions: bcc-Li along [100], fcc-Al along [100], fcc-Cu along [100], bcc-Na along [100], NaCl along [111], ds-Si along [111], ds-C along [111], and SiC along [111]. For Al and Si, the data from BLPS is also available. Hence we plot the density obtained by the BLPS, labeled as ρ_{BLPS} in Figs. 7 and 9. Figures 7, 8, and 9 clearly show that the electron density obtained by our LIPS well reproduces the density by NLPS. The density by BLPS also seems to be good enough. For Al [Fig. 7(b)] and Si [Fig. 9(a)], it appears that $|\rho_{LIPS} - \rho_{NLPS}|$ is generally smaller than $|\rho_{BLPS} - \rho_{NLPS}|$. As for diamond C and SiC, the calculated ρ_{LIPS} satisfactorily reproduce the characteristic features of ρ_{NLPS}: The peculiar double peaks between the C-C bonds in diamond C and the substantial iconicity
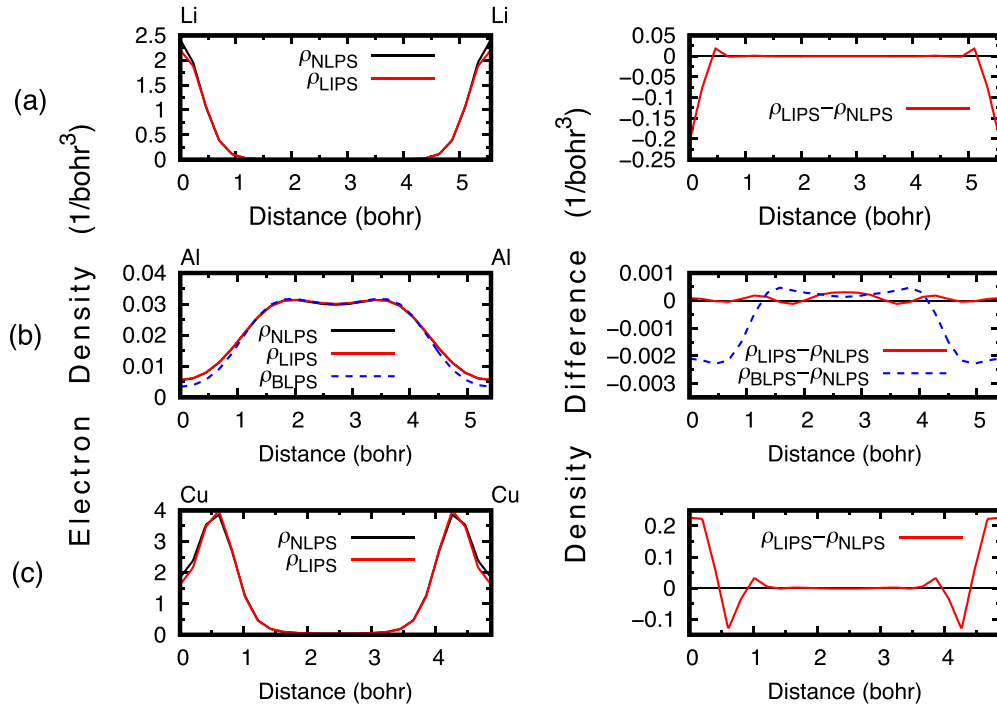
FIG. 7. Electron densities (left panels) obtained by NLPS and LIPS and the electron-density differences (right panels) for (a) body-centered-cubic(bcc)-Li, (b) face-centered-cubic(fcc)-Al, and (c) fcc-Cu along the [100] direction. The position of each atom along the [100] is marked on the top lines of each figure. For Al, the data obtained by using BLPS is also shown.

in SiC (Fig. 9). The maximum deviation of $\rho_{LIPS} - \rho_{NLPS}$ for diamond is 4.38×10^{-2} bohr^{-3} at the C nuclear site for diamond C. For SiC it is 4.49×10^{-2} bohr^{-3} again at the C nuclear site.

We also evaluate the quality of our LIPSs by calculating the total energy as a function of the volume $E(V)$ for ds-Si, ds-

C, bcc-Li, bcc-Na, fcc-Al, fcc-Cu, NaCl, and zincblende(3C)-SiC. The obtained $E(V)$ is fitted to Murnaghan's equation of state [76] to deduce the equilibrium lattice constant a_0 and the bulk modulus B_0. The results are shown in Table X along with those obtained by the NLPSs. The results from the BLPS for Li, Al, and Si are also tabulated.
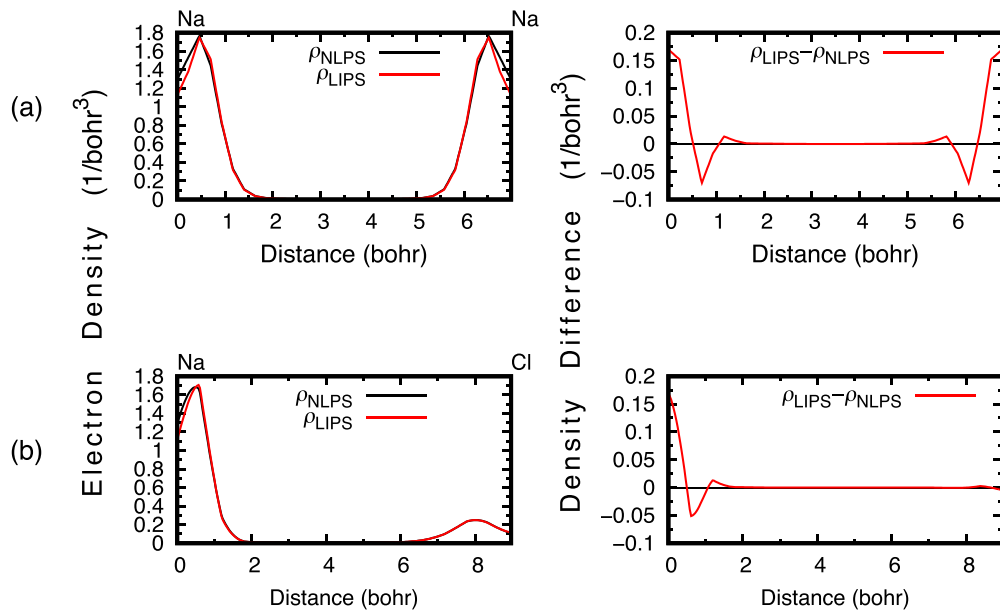

FIG. 8. Electron densities (left panels) obtained by NLPS and LIPS and the electron-density differences (right panels) for (a) body-centered-cubic(bcc)-Na along the [100] direction and (b) NaCl along the [111] direction on the (110) plane. The position of each atom along each direction is marked on the top lines of each figure.
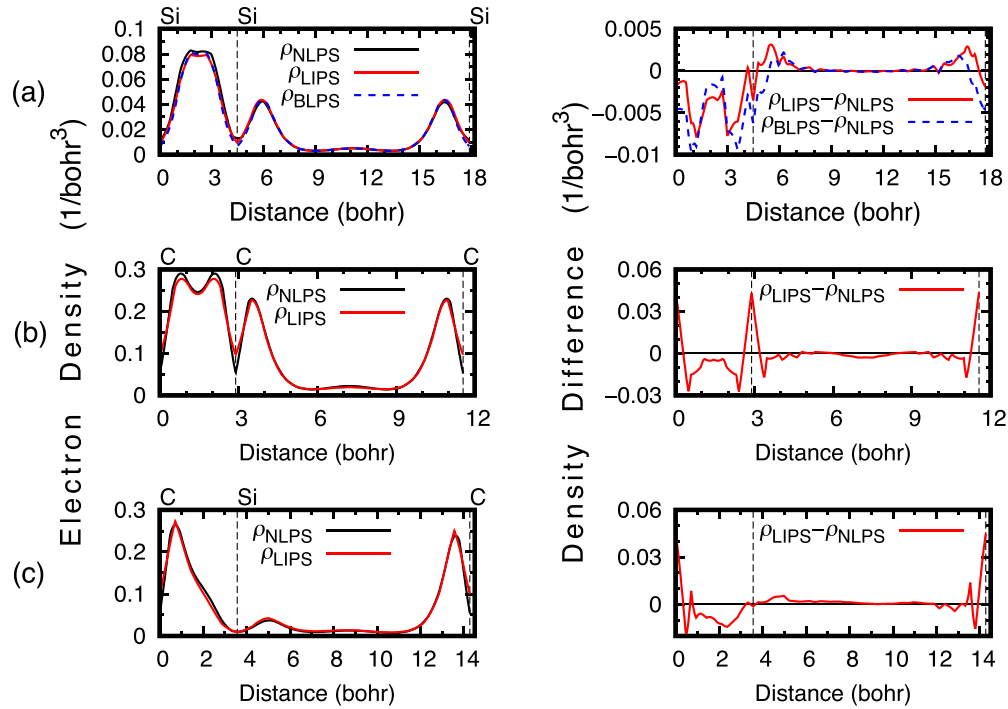
FIG. 9. Electron densities (left panels) obtained by NLPS and LIPS and the electron-density differences (right panels) for (a) diamond-structured Si, (b) diamond-structured C, and (c) SiC along the [111] direction on the (110) plane. The position of each atom along the [111] on the (110) is marked on the top lines of each figure. For Si, the data obtained by using BLPS is also shown.

It is clear that the structural properties of the 8 different materials produced by our LIPSs are as accurate as those from the NLPSs. The electron densities and the structural properties obtained above certainly assures the reliability and the transferability of the frozen-core approximation (pseudopotential scheme) using our LIPSs.

APPENDIX B: FUNCTIONAL DERIVATIVE OF LAPLACIAN-LEVEL KINETIC-ENERGY FUNCTIONAL

For any Laplacian level KEDF of the form

$$T[\rho] = \int \tau(\rho, |\nabla\rho|^2, \nabla^2\rho)d\boldsymbol{r}, \qquad (B1)$$

the variation with respect to the electron density is

$$\delta T[\rho] = \int \left[\frac{\partial\tau}{\partial\rho}\delta\rho + \frac{\partial\tau}{\partial(|\nabla\rho|^2)}\delta(|\nabla\rho|^2) \right. $$
$$\left. + \frac{\partial\tau}{\partial(\nabla^2\rho)}\delta(\nabla^2\rho) \right]d\boldsymbol{r}$$
$$= \int \left[\frac{\partial\tau}{\partial\rho} - 2\nabla\cdot\left(\frac{\partial\tau}{\partial(|\nabla\rho|^2)}\nabla\rho \right) \right. $$
$$\left. + \nabla^2\left(\frac{\partial\tau}{\partial(\nabla^2\rho)} \right) \right]\delta\rho d\boldsymbol{r}, \qquad (B2)$$

where we used $\delta(|\nabla\rho|^2) = 2\nabla\rho\cdot\nabla(\delta\rho)$, $\delta(\nabla^2\rho) = \nabla^2(\delta\rho)$, and the first $(\int \boldsymbol{v}(\boldsymbol{r})\cdot\nabla f(\boldsymbol{r})d\boldsymbol{r} = -\int f(\boldsymbol{r})\nabla\cdot v(\boldsymbol{r})d\boldsymbol{r})$

TABLE X. Equilibrium lattice constants (a_0) in Å and bulk moduli (B_0) in GPa obtained by using the NLPS and our LIPS of 8 different solids. For bcc-Li, fcc-Al, and ds-Si, the obtained values using the BLPS are also listed. Numbers in parentheses are the relative errors with respect to the experimental values. The mean absolute relative errors (MAREs) in % with respect to the experimental values are 0.59 (NLPS), 1.11 (BLPS), and 1.13 (LIPS) for a_0, and 6.21 (NLPS), 6.78 (BLPS), and 11.0 (LIPS) for B_0.

|	bcc-Li		diamond		bcc-Na		fcc-Al		ds-Si		NaCl		fcc-Cu		SiC	
	a_0	B_0	a_0	B_0	a_0	B_0	a_0	B_0	a_0	B_0	a_0	B_0	a_0	B_0	a_0	B_0
NLPS	3.430	13.3	3.560	434	4.212	7.68	4.051	77.8	5.466	86.9	5.694	23.9	3.634	146	3.077	226
	(−1.75)	(−4.32)	(−0.20)	(−1.81)	(−0.31)	(21.9)	(0.02)	(2.37)	(0.63)	(−12.0)	(0.96)	(−2.05)	(0.66)	(4.29)	(−0.19)	(0.89)
BLPS	3.481	14.8					3.968	84.0	5.377	95.5						
	(−0.29)	(6.47)					(−2.02)	(−9.08)	(−1.01)	(−17.1)						
LIPS	3.508	14.9	3.517	327	4.266	7.71	4.050	77.5	5.392	100	5.480	25.9	3.645	158	3.023	246
	(0.49)	(7.19)	(−1.40)	(−26.0)	(0.97)	(22.4)	(0.00)	(1.97)	(−0.74)	(1.21)	(−2.84)	(6.15)	(0.97)	(12.9)	(−1.62)	(9.82)
Exp.	3.491	13.9	3.567	442	4.225	6.3	4.05	76	5.432	98.8	5.64	24.4	3.61	140	3.083	224
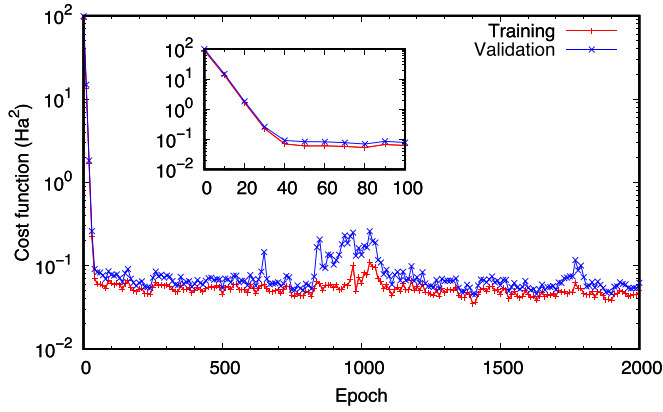
FIG. 10. Evolution of the cost function during the optimization process. The inset shows a zoom-in for the epochs smaller than 100.

and the second Green's identities $(\int f(\mathbf{r})\nabla^2 g(\mathbf{r})d\mathbf{r} = \int \nabla^2 f(\mathbf{r})g(\mathbf{r})d\mathbf{r})$. Specifically, when the kinetic energy is expressed in meta-GGA form [Eq. (5)], substituting the expressions

$$\frac{\partial \tau}{\partial \rho} = \tau^{\mathrm{TF}}\left(\frac{5}{3\rho}F + \frac{\partial s^2}{\partial \rho}\frac{\partial F}{\partial s^2} + \frac{\partial q}{\partial \rho}\frac{\partial F}{\partial q}\right) \quad (B3)$$

$$\frac{\partial \tau}{\partial(|\nabla \rho|^2)} = \tau^{\mathrm{TF}}\frac{\partial F}{\partial s^2}\frac{\partial s^2}{\partial(|\nabla \rho|^2)} \quad (B4)$$

$$\frac{\partial \tau}{\partial(\nabla^2 \rho)} = \tau^{\mathrm{TF}}\frac{\partial F}{\partial q}\frac{\partial q}{\partial(\nabla^2 \rho)} \quad (B5)$$

in Eq. (B2) leads to Eq. (9) with the use of $\partial s^2/\partial \rho = -8s^2/(3\rho)$, $\partial s^2/\partial(|\nabla \rho|^2) = \rho^{-8/3}/(4(3\pi^2)^{2/3})$, $\partial q/\partial \rho = -5q/(3\rho)$, and $\partial q/\partial(\nabla^2 \rho) = \rho^{-5/3}/(4(3\pi^2)^{2/3})$.

APPENDIX C: TRAINING DETAILS

We here determine the training hyperparameters, η and ν, appearing in the NN weight updating formula Eq. (11). We have first performed a grid search for optimum constant values of η and ν in the range of $10^{-4} \leqslant \eta \leqslant 10$ and $10^{-8} \leqslant \nu \leqslant 1$, and found that a pair of constants, $(\eta = 0.1, \nu = 10^{-5})$, provides the smallest RMSE $\sqrt{2L}$ of 0.228. Since the ordinary gradient scheme represented by ν is helpful only at early stages of training, we introduce proper time scheduling of ν, keeping $\eta = 0.1$ constant. We consider three options for $\nu(t)$, $\nu(t) = \nu_0/(1 + bt)$, $\nu_0/\sqrt{1 + bt}$, and $\nu_0 \exp(-bt)$, where t is the number of the training epoch and $\nu_0 = 10^{-5}$. We have found that $\nu = \nu_0/(1 + bt)$ with $b = 10^{-2}$ provides the smallest RMSE and adopted this scheduling in the present work.

Figure 10 shows variation of the cost function L during the optimization for the case of the fixed $A = 10^{1.5}$ (see Appendix F below). It is noteworthy that the training of only 100 epochs makes the cost function small by three-orders-of-magnitude, being less than 0.1 Ha^2. It is shown that the cost function becomes smallest at the 1400-th epoch. We thus adopt the NN weight at the 1400th epoch as the final trained weight. The total computational time for the optimization of the NN weights is typically less than 30 minutes on usual house computers, which is the orders-of-magnitudes shorter than the computational time for the SCF calculations. We note

that 3000 training points are sufficient to minimize the cost function to the value less than 0.06 Ha^2.

APPENDIX D: FUNCTIONAL DERIVATIVE TRAINING

In order to calculate $\partial L/\partial \mathbf{W}$, we need to obtain several derivatives, $\partial F^{\mathrm{NN}}/\partial \mathbf{W}$, $\partial(\partial F^{\mathrm{NN}}/\partial(s^2))/\partial \mathbf{W}$ and $\partial(\partial F^{\mathrm{NN}}/\partial q)\partial \mathbf{W}$. In this paper, the first derivative is obtained by the conventional backpropagation, and the second and the third derivatives are obtained by the backpropagation for the derivatives [49,50]. Here we briefly overview the algorithm of the backpropagation in order to explain how to compute the partial derivative of NN outputs with respect to the NN weights $\mathbf{W}$. The number of the neuron in the final layer is $D_N = 1$ in our case. In the conventional backpropagation, our aim is to compute $\partial z_1^{(N)}/\partial W_{ji}^{(l)}$ from $\partial z_1^{(N)}/\partial a_1^{(N)} = \sigma^{(N)'}(a_1^{(N)})$. Using the quantity $\delta_j^{(l)}$ that satisfies the recursive formula

$$\delta_j^{(l)} \equiv \frac{\partial z_1^{(N)}}{\partial a_j^{(l)}} = \sum_k \frac{\partial z_1^{(N)}}{\partial a_k^{(l+1)}}\frac{\partial a_k^{(l+1)}}{\partial a_j^{(l)}}$$

$$= \sum_k \delta_k^{(l+1)}\frac{\partial}{\partial a_j^{(l)}}\sum_i W_{ki}^{(l+1)}\sigma^{(l)}(a_i^{(l)})$$

$$= \sum_k \delta_k^{(l+1)}W_{kj}^{(l+1)}\sigma^{(l)'}(a_j^{(l)}), \quad (D1)$$

we can obtain $\partial z_1^{(N)}/\partial W_{ji}^{(l)}$ as

$$\frac{\partial z_1^{(N)}}{\partial W_{ji}^{(l)}} = \delta_j^{(l)}z_i^{(l-1)}. \quad (D2)$$

This formula allows us to recursively compute $\partial z_1^{(N)}/\partial W_{ji}^{(l)}$ from $\delta_1^{(N)}$. Next, we derive a similar recursive method to obtain $\partial \gamma_{1r}^{(N)}/\partial W_{ji}^{(l)}$, where $\gamma_{1r}^{(N)} = \partial z_1^{(N)}/\partial z_r^{(1)}$ is the derivative of the output with respect to the inputs. Here $\gamma_{1r}^{(l)}$ has been computed recursively as

$$\gamma_{1r}^{(l)} = \sigma^{(l)'}(a_j^{(l)})\sum_k W_{jk}^{(l)}\gamma_{kr}^{(l-1)} \quad (D3)$$

in advance. Using Eq. (D2), we have

$$\frac{\partial \gamma_{1r}^{(N)}}{\partial W_{ji}^{(l)}} = z_i^{(l-1)}\frac{\partial \delta_j^{(l)}}{\partial z_r^{(1)}} + \frac{\partial z_i^{(l-1)}}{\partial z_r^{(1)}}\delta_j^{(l)}$$

$$= z_i^{(l-1)}\frac{\partial \delta_j^{(l)}}{\partial z_r^{(1)}} + \gamma_{ir}^{(l-1)}\delta_j^{(l)}. \quad (D4)$$

By introducing a quantity $\zeta_{jr}^{(l)}$ that satisfies the recursive formula

$$\zeta_{jr}^{(l)} \equiv \frac{\partial \delta_j^{(l)}}{\partial z_r^{(1)}} = \frac{\partial}{\partial z_r^{(1)}}\left[\sigma^{(l)'}(a_j^{(l)})\sum_k \frac{\partial z_1^{(N)}}{\partial a_k^{(l+1)}}W_{kj}^{(l+1)}\right]$$

$$= \frac{\partial \sigma^{(l)'}(a_j^{(l)})}{\partial z_j^{(l)}}\gamma_{jr}^{(l)}\sum_k W_{kj}^{(l+1)}\delta_k^{(l+1)}$$

$$+ \sigma^{(l)'}(a_j^{(l)})\sum_k W_{kj}^{(l+1)}\zeta_{kr}^{(l+1)}, \quad (D5)$$
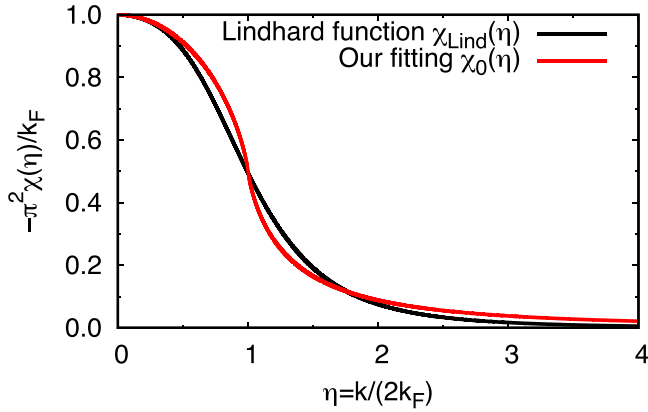
FIG. 11. Comparison of our response function $\chi_0(\eta)$ with the Lindhard response function $\chi_{\text{Lind}}(\eta)$ after the optimization of the parameter β.

we can compute $\partial \gamma_{1r}^{(N)}/\partial W_{ji}^{(l)}$ starting from $\zeta_{jr}^{(N)} = 0$ since we take the activation function in the output layer as an identity function, $z_1^{(N)} = \sigma^{(N)}(a_1^{(N)}) = a_1^{(N)}$.

APPENDIX E: DETERMINATION OF THE ENHANCEMENT FACTOR $F^{(0)}(s^2, q)$

We have optimized β in the enhancement factor $F^{(0)}(s^2, q)$, Eq. (13), so that the inverse of the response function derived from this enhancement factor in the homogeneous-gas limit,

$$-\frac{1}{\chi_0(\eta)} = \frac{\pi^2}{k_F}\left(1 + \frac{\eta^2}{3} + \frac{9}{5}\beta\eta^4\right) \quad \text{(E1)}$$

reproduces the Lindhard function

$$-\frac{1}{\chi_{\text{Lind}}(\eta)} = \frac{\pi^2}{k_F}\left(\frac{1}{2} + \frac{1-\eta^2}{4\eta}\ln\left|\frac{1+\eta}{1-\eta}\right|\right)^{-1}. \quad \text{(E2)}$$

Here $\eta = k/(2k_F)$ with k_F being $(3\pi^2\rho)^{1/3}$.

To this end, we have discretized η and defined 10^4 points $\eta_i \in [10^{-5}, 4]$. Then, we have performed the least square fitting that minimizes the cost function C

$$C = \frac{1}{2}\sum_i\left[\frac{1}{\chi_0(\eta_i)} - \frac{1}{\chi_{\text{Lind}}(\eta_i)}\right]^2, \quad \text{(E3)}$$

which results in the optimized value, $\beta = 0.382$. The obtained $\chi_0(\eta)$ is compared with $\chi_{\text{Lind}}(\eta)$ in Fig. 11.

APPENDIX F: DETERMINATION OF SUBSPACE DECOMPOSITION PARAMETER A

The parameter A, which defines our subsystem DFT functional [Eq. (14)] can be determined by minimizing the cost function L [Eq. (8)]. In the case, the metric tensor G_{ik} [Eq. (12)] is expanded to $(n_W + 1) \times (n_W + 1)$ dimension by adding the components related to $\partial[\delta\tilde{T}^{\text{NN}}(\boldsymbol{r}_p)/\delta\rho]/\partial A$. We have indeed minimized the cost function with the expanded metric tensor and determined the parameter A. In the actual minimization, we have performed grid search for the initial values of $A_m = 10^{0.5m}$ with m being integers from 0 to 9 (a typical logarithmic grid) and the subsequent neural-network

TABLE XI. The subsystem-DFT parameter A and the corresponding RMSE between our NN KEFD and KS KEFD (the column "$\sqrt{2L}$" in the unit of Ha). KS density ρ^{KS} is used to evaluate $\sqrt{2L}$. The RMSE of the SCF density ρ_{SCF} obtained by our OFDFT scheme with each value of A with respect to ρ^{KS} is also shown (the column "RMSE" in the unit of $10^{-2} \times \text{bohr}^{-3}$). $[(n_W + 1) \times (n_W + 1)]\boldsymbol{G}$ and $[n_W \times n_W]\boldsymbol{G}$ denote the results with the expanded metric tensor and the original metric tensor without the $\partial[\delta\tilde{T}^{\text{NN}}(\boldsymbol{r}_p)/\delta\rho]/\partial A$-related components, respectively.

m	$[(n_W + 1) \times (n_W + 1)]\boldsymbol{G}$			$[n_W \times n_W]\boldsymbol{G}$	
$(A_m = 10^{0.5m})$	A^{opt}	$\sqrt{2L}$	RMSE	$\sqrt{2L}$	RMSE
0	2.751020	0.433	1.904	0.289	1.271
1	4.220258	0.331	1.534	0.273	1.262
2	10.54878	0.269	1.225	0.264	1.201
3	31.76219	0.284	1.232	0.274	1.186
4	100.0558	0.328	1.204	0.326	1.194
5	316.2413	0.404	1.300	0.387	1.246
6	1000.005	0.408	1.274	0.400	1.246
7	3162.279	0.337	1.231	0.346	1.266
8	10000.00	0.297	1.243	0.297	1.243
9	31622.78	0.279	1.238	0.279	1.238

minimization of the cost function L. We have found that the optimized value A^{opt} depends on the initial choice A_m as shown in Table XI, indicating that the cost function L shows multistability as a function of A. This is presumably due to a fact that the NN weights $\{W^i\}$ rather than A are decisive to determine the cost function. The computed RMSE of our KEFD with respect to the KS-KEFD, i.e., $\sqrt{2L}$ itself, for each value of A^{opt} are tabulated in Table XI. For $A = 10.549$, $\sqrt{2L}$ has the minimum value of 0.269 Ha although it increases only by 7% for $A = 31.76$.

Observing the insensitivity of the cost function to the parameter A, we have also performed the minimization of the cost function with the original metric tensor with the dimension of n_W with the fixed value of $A = A_m$. The obtained $\sqrt{2L}$ is shown in Table XI. The $\sqrt{2L}$ values are comparable with those obtained for A^{opt} and the minimum value 0.264 Ha for $A_m = 10$ is even smaller than the value 0.269 Ha for $A = 10.549$ above. This is presumably because we have succeeded to optimize $\{W^i\}$ better with the fixed choice of A. This may indicate the limitation of the present SNGD optimization scheme for the two parameter sets with different mathematical structures.

Having observed that the $\sqrt{2L}$ values are insensitive to the value of A, we introduce another criterion to determine the A value. That is to minimize the difference of the electron density ρ_{SCF} obtained by the SCF calculations in our OFDFT scheme with each A and $\{W^i\}$ values from that ρ^{KS} obtained by the KSDFT scheme. The minimum RMSE 1.186 $\times 10^{-2}$ bohr^{-3} of those two densities are obtained for the A value of $10^{1.5}$. This set of parameters, A and $\{W^i\}$ in turn produces value of 0.274 Ha for $\sqrt{2L}$, which is just 3.8% larger than the minimum $\sqrt{2L}$ explained above. Recalling that the electron density is the fundamental quantity in DFT, we adopt the value of $A = 10^{1.5}$ in the present paper. Further examination of the A value remains in future.

[1] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, Phys. Rev. **136**, B864 (1964).

[2] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. **140**, A1133 (1965).

[3] R. O. Jones, Density functional theory: Its origins, rise to prominence, and future, Rev. Mod. Phys. **87**, 897 (2015).

[4] N. Mardirossian and M. Head-Gordon, Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals, Mol. Phys. **115**, 2315 (2017).

[5] J. P. Perdew and K. Schmidt, *Density Functional Theory and Its Application to Materials* (AIP Press, Melville, NY, 2001).

[6] L. H. Thomas, The calculation of atomic fields, Proc. Cambridge Philos. Soc. **23**, 542 (1927).

[7] E. Fermi, Eine statistische methode zur bestimmung einiger eigenschaften des atoms und ihre anwendung auf die theorie des periodischen systems der elemente, Z. Phys. **48**, 73 (1928).

[8] C. F. von Weizsäcker, Zur theorie der kernmassen, Z. Phys. **96**, 431 (1935).

[9] S. D. Schwartz (ed.), *Theoretical Methods in Condensed Phase Chemistry*, Progress in Theoretical Chemistry and Physics, Vol. 5 (Kluwer Academic, Dordrecht, 2002).

[10] V. Bach and L. D. Site, eds., *Many-Electron Approaches in Physics, Chemistry, and Mathematics* (Springer International, New York, 2014).

[11] D. Garcia-Aldea and J. E. Alvarellos, Kinetic energy density study of some representative semilocal kinetic energy functionals, J. Chem. Phys. **127**, 144109 (2007), and references therein.

[12] L. A. Constantin, E. Fabiano, and F. D. Sala, Semilocal Pauli–Gaussian kinetic functionals for orbital-free density functional theory calculations of solids, J. Phys. Chem. Lett. **9**, 4385 (2018).

[13] L. A. Constantin, E. Fabiano, and F. D. Sala, Performance of semilocal kinetic energy functionals for orbital-free density functional theory, J. Chem. Theory Comput. **15**, 3044 (2019).

[14] K. Luo, V. V. Karasiev, and S. B. Trickey, A simple generalized gradient approximation for the noninteracting kinetic energy density functional, Phys. Rev. B **98**, 041111(R) (2018).

[15] L. W. Wang and M. P. Teter, Kinetic-energy functional of the electron-density, Phys. Rev. B **45**, 13196 (1992).

[16] D. Garcia-Aldea and J. E. Alvarellos, Kinetic-energy density functionals with nonlocal terms with the structure of the Thomas-Fermi functional, Phys. Rev. A **76**, 052504 (2007).

[17] P. García-González, J. E. Alvarellos, and E. Chacón, Nonlocal kinetic-energy-density functionals, Phys. Rev. B **53**, 9509 (1996).

[18] Y. A. Wang, N. Govind, and E. A. Carter, Orbital-free kinetic-energy density functionals for the nearly free electron gas, Phys. Rev. B **58**, 13465 (1998).

[19] Y. A. Wang, N. Govind, and E. A. Carter, Orbital-free kinetic-energy density functionals with a density-dependent kernel, Phys. Rev. B **60**, 16350 (1999).

[20] B. J. Zhou, V. L. Ligneres, and E. A. Carter, Improving the orbital-free density functional theory description of covalent materials, J. Chem. Phys. **122**, 044103 (2005).

[21] D. Garcia-Aldea and J. E. Alvarellos, Approach to kinetic energy density functionals: Nonlocal terms with the structure of the von Weizsäcker functional, Phys. Rev. A **77**, 022502 (2008).

[22] C. Huang and E. A. Carter, Nonlocal orbital-free kinetic energy density functional for semiconductors, Phys. Rev. B **81**, 045206 (2010).

[23] M. Levy and Hui Ou-Yang, Exact properties of the Pauli potential for the square root of the electron density and the kinetic energy functional, Phys. Rev. A **38**, 625 (1988).

[24] J. Lindhard, On the properties of a gas of charged particles, Dan. Mat. Fys. Medd. **28**, 8 (1954).

[25] V. V. Karasiev, D. Chakraborty, O. A. Shukruto, and S. B. Trickey, Nonempirical generalized gradient approximation free-energy functional for orbital-free simulations, Phys. Rev. B **88**, 161108(R) (2013).

[26] I. Shin and E. A. Carter, Enhanced von Weizsäcker Wang-Govind-Carter kinetic energy density functional for semiconductors, J. Chem. Phys. **140**, 18A531 (2014).

[27] W. Mi, A. Genova, and M. Pavanello, Nonlocal kinetic energy functionals by functional integration, J. Chem. Phys. **148**, 184107 (2018).

[28] M. Levy and J. P. Perdew, Hellmann-Feynman, virial, and scaling requisites for the exact universal density functionals. shape of the correlation potential and diamagnetic susceptibility for atoms, Phys. Rev. A **32**, 2010 (1985).

[29] D. A. Clevert, T. Unterthiner, and S. Hochreiter, Fast and accurate deep network learning by exponential linear units (ELUs), in *Proceedings of ICLR 2016*, Vol. 1 (2016) pp. 1–14.

[30] J.-I. Iwata, D. Takahashi, A. Oshiyama, T. Boku, K. Shiraishi, S. Okada, and K. Yabana, A massively-parallel electronic-structure calculations based on real-space density functional theory, J. Comput. Phys. **229**, 2339 (2010).

[31] Y. Hasegawa, J.-I. Iwata, M. Tsuji, D. Takahashi, A. Oshiyama, K. Minami, T. Boku, H. Inoue, Y. Kitazawa, I. Miyoshi, and M. Yokokawa, Performance evaluation of ultra-large-scale first-principles electronic structure calculation code on the K computer, Int. J. High Perform. Comput. Appl. **28**, 335 (2014).

[32] J.-I. Iwata, https://github.com/j-iwata/RSDFT.

[33] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).

[34] D. R. Hamann, M. Schlüter, and C. Chiang, Norm-Conserving Pseudopotentials, Phys. Rev. Lett. **43**, 1494 (1979).

[35] G. B. Bachelet, D. R. Hamann, and M. Schlüter, Pseudopotentials that work: From H to Pu, Phys. Rev. B **26**, 4199 (1982).

[36] B. Zhou, Y. A. Wang, and E. A. Carter, Transferable local pseudopotentials derived via inversion of the Kohn-Sham equations in a bulk environment, Phys. Rev. B **69**, 125109 (2004).

[37] C. Huang and E. A. Carter, Transferable local pseudopotentials for magnesium, aluminum and silicon, Phys. Chem. Chem. Phys. **10**, 7109 (2008).

[38] B. G. del Rio, J. M. Dieterich, and E. A. Carter, Globally-optimized local pseudopotentials for (orbital-free) density functional theory simulations of liquids and solids, J. Chem. Theory Comput. **13**, 3684 (2017).

[39] E. S. Kadantsev and M. J. Stott, Variational method for inverting the Kohn-Sham procedure, Phys. Rev. A **69**, 012502 (2004).

[40] F. Imoto, Development of orbital-free density functional theory with machine learning and its applications, Ph.D. thesis, University of Tokyo, 2019.

[41] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, Finding Density Functionals With Machine Learning, Phys. Rev. Lett. **108**, 253002 (2012).

[42] J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K.-R. Müller, and K. Burke, Orbital-free bond breaking via machine learning, J. Chem. Phys. **139**, 224104 (2013).

[43] K. Yao and J. Parkhill, Kinetic energy of hydrocarbons as a function of electron density and convolutional neural networks, J. Chem. Theory Comput. **12**, 1139 (2016).

[44] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, Bypassing the Kohn-Sham equations with machine learnings, Nat. Commun. **8**, 872 (2017).

[45] J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, and H. Nakai, Semi-local machine-learned kinetic energy density functional with third-order gradients of electron density, J. Chem. Phys. **148**, 241705 (2018).

[46] J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, and H. Nakai, Semi-local machine-learned kinetic energy density functional demonstrating smooth potential energy curves, Chem. Phys. Lett. **734**, 136732 (2019).

[47] M. Fujinami, R. Kageyama, J. Seino, Y. Ikabata, and H. Nakai, Orbital-free density functional theory calculation applying semi-local machine-learned kinetic energy density functional and kinetic potential, Chem. Phys. Lett. **748**, 137358 (2020).

[48] S. Amari, Natural gradient works efficiently in learning, Neural Comput. **10**, 251 (1998).

[49] A. Pukrittayakamee, M. Hagan, R. Raff, S. Bukkapatnam, and R. Komanduri, Practical training framework for fitting a function and its derivatives, IEEE Trans. Neural. Netw. **22**, 936 (2011).

[50] A. Pukrittayakamee, M. Malshe, M. Hagan, L. M. Raff, R. Narulkar, S. Bukkapatnum, and R. Komanduri, Simultaneous fitting of a potential-energy surface and its corresponding force fields using feedforward neural networks, J. Chem. Phys. **130**, 134101 (2009).

[51] M. Brack, B. K. Jennings, and Y. H. Chu, On the extended Thomas-Fermi approximation to the kinetic energy density, Phys. Lett. B **65**, 1 (1976).

[52] W. Kohn and A. E. Mattsson, Edge Electron Gas, Phys. Rev. Lett. **81**, 3487 (1998).

[53] A. E. Mattson and W. Kohn, An energy functional and surfaces, J. Chem. Phys. **115**, 3441 (2001).

[54] R. Armiento and A. E. Mattson, Subsystem functionals in density-functional theory: Investigating the exchange energy per particle, Phys. Rev. B **66**, 165117 (2002).

[55] A. E. Mattson and R. Armiento, The subsystem functional scheme: The Armiento-Mattsson 2005 (AM05) functional and beyond, Int. J. Quantum Chem. **110**, 2274 (2010).

[56] G. S. Ho, V. L. Lignères, and E. A. Carter, Introducing PROFESS: A new program for orbital-free density functional theory calculations, Comput. Phys. Comm. **179**, 839 (2008).

[57] L. Hung, C. Huang, I. Shin, G. S. Ho, V. L. Lignères, and E. A. Carter, Introducing PROFESS 2.0: A parallelized, fully linear scaling program for orbital-free density functional theory calculations, Comput. Phys. Comm. **181**, 2208 (2010).

[58] M. Chen, J. Xia, C. Huang, J. M. Dieterich, L. Hung, I. Shin, and E. A. Carter, Introducing PROFESS 3.0: An advanced program for orbital-free density functional theory molecular dynamics simulations, Comput. Phys. Comm. **190**, 228 (2015).

[59] W. Mi, X. Shao, C. Su, Y. Zhou, S. Zhang, Q. Li, H. Wang, L. Zhang, M. Miao, Y. Wang, and Y. Ma, ATLAS: A real-space finite-difference implementation of orbital-free density functional theory, Comput. Phys. Comm. **200**, 87 (2016).

[60] M. Levy, J. P. Perdew, and V. Sahni, Exact differential equation for the density and ionization energy of a many-particle system, Phys. Rev. A **30**, 2745 (1984).

[61] V. V. Karasiev and S. B. Trickey, Issues and challenges in orbital-free density functional calculations, Comput. Phys. Comm. **183**, 2519 (2012).

[62] J. Lehtomäki, I. Makkonen, M. A. Caro, A. Harju, and O. Lopez-Acevedo, Orbital-free density functional theory implementation with the projector augmented-wave method, J. Chem. Phys. **141**, 234102 (2014).

[63] A. Berk, Lower-bound energy functionals and their application to diatomic systems, Phys. Rev. A **28**, 1908 (1983).

[64] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.3.033198 for the calculated Self-Consistent-Field (SCF) electron densities obtained by our NN KEDF along with those obtained by using the KEDFs in the past, i.e., PGSL0.25, LKT, and the conventional TF(1/5)vW KEDFs.

[65] 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).

[66] T. Ozaki and H. Kino, Efficient projector expansion for the *ab initio* LCAO method, Phys. Rev. B **72**, 045121 (2005).

[67] W. Kohn, Density Functional And Density Matrix Method Scaling Linearly With The Number Of Atoms, Phys. Rev. Lett. **76**, 3168 (1996).

[68] M. J. Gillan, D. R. Bowler, A. S. Torralba, and T. Miyazaki, Order-*N* first-principles calculations with the CONQUEST code, Comput. Phys. Comm. **177**, 14 (2007).

[69] E. A. Carter, https://github.com/PrincetonUniversity/BLPSLibrary.

[70] Y. Wang and R. G. Parr, Construction of exact Kohn-Sham orbitals from a given electron density, Phys. Rev. A **47**, R1591(R) (1993).

[71] Q. Wu and W. Yang, A direct optimization method for calculating density functionals and exchange-correlation potentials from electron densities, J. Chem. Phys. **118**, 2498 (2003).

[72] W. M. C. Foulkes, and R. Haydock, Tight-binding models and density-functional theory, Phys. Rev. B **39**, 12520 (1989).

[73] S. Goedecker, M. Teter, and J. Hutter, Separable dual-space Gaussian pseudopotentials, Phys. Rev. B **54**, 1703 (1996).

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

[75] N. Troullier and J. L. Martins, Efficient pseudopotentials for plane-wave calculations. II. Operators for fast iterative diagonalization, Phys. Rev. B **43**, 8861 (1991).

[76] F. D. Murnaghan, The compressibility of media under extreme pressures, Proc. Nat. Acad. Sci. USA **30**, 244 (1944).