Toward the Exact Exchange-Correlation Potential: a 3D Convolutional Neural Network Construct

Yi Zhou, Jiang Wu, Shuguang Chen, GuanHua Chen[*]

Department of Chemistry, The University of Hong Kong, Hong Kong S.A.R, China

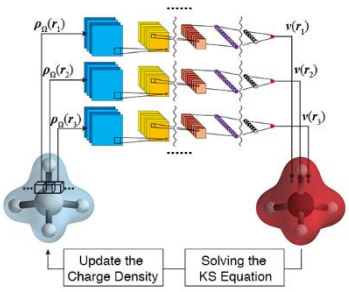
AUTHOR INFORMATION

Corresponding Author:

*Email: ghc@everest.hku.hk

ABSTRACT A deep neural network is constructed to yield in-principle exact exchange-correlation potential. It requires merely the electron densities of small molecules and ions, and yet, can determine the exchange-correlation potentials of large molecules. We train and validate the neural network based on the data for H_2 and HeH^+, and subsequently determine the ground state electron density of stretched HeH^+, linear H_3^+ and H–He–He–H^{2+}. Moreover, the deep neural network is proven to model the van der Waals interaction by being trained and validated on dataset containing He_2. Comparisons to B3LYP are given to illustrate the accuracy and transferability of the neural network.

TOC GRAPHIC



Density-functional theory (DFT)[1, 2] has become the most widely used quantum mechanical simulation method. Despite its success, DFT suffers from the chronic deficiencies such as the self-interaction error,[3] chemical accuracy issue for thermochemistry,[4] poor treatment to van der Waals interaction,[5, 6] inaccurate reaction barrier,[7] hydrogen bond[8], charge-transfer,[9] delocalization error[10, 11], static correlation and strong correlation error.[12, 13] Medvedev *et al.* showed that although the accuracy in energy of DFT methods had been improving over the past decade or two, the calculated electron density distribution had become less accurate.[14] Despite the continuing efforts to find more accurate exchange-correlation (xc) functional, the improvement in calculated accuracy has slowed down. Alternative beyond the traditional approaches[15-18] must be sought.

Efforts have been made to construct high quality xc potentials for given electron densities via optimization procedures.[19, 20] Modern machine learning techniques have also been incorporated in the DFT method aiming at boosting calculations,[21-24] calibrating results,[4] or improving the xc functional.[25] In 2003 we proposed a neural network based B3LYP xc functional.[26] The three hybrid parameters in B3LYP xc functional, a_0, a_x, and a_c, are in principle the functionals of ground state electron density function and thus system dependent. Their values were evaluated via a neural network which had been trained and tested against the experimental data. The input descriptors of the neural network were the number of electrons, dipole moment, quadrupole moment, kinetic energy and spin multiplicity of the system, which are the functionals of the ground state electron density function of the molecule. In 2012, Burke and coworkers proposed a machine learning based density functional for the kinetic energy of one-dimensional electron gas,[25] which was later extended for real molecules.[27-29] All these machine learning based functionals require the knowledge of the *global electron density function*. For small molecules or ions, it is straight forward to obtain their electron density functions and the corresponding energies using highly

accurate quantum chemistry methods such as the coupled-cluster method, quantum Monte Carlo,[30] and the density matrix renormalization group technique.[31] However, electron density functions and energies of larger molecules or ions are difficult to determine, if not impossible. As a result, no or little data are available for the large molecular systems. This problem is difficult to circumvent, and prevents all the existing machine learning based methods for DFT xc functional applicable for large molecular systems.

The xc potential $v_{\mathrm{xc}}(\boldsymbol{r})$ of the density-functional theory (DFT)[2]

$$v_{\mathrm{xc}}(\boldsymbol{r}) = \frac{\delta E_{\mathrm{xc}}[\rho(\boldsymbol{r})]}{\delta \rho(\boldsymbol{r})} \tag{1}$$

at any point $\boldsymbol{r}_i$ is a functional of electron density function $\rho(\boldsymbol{r})$. In principle, the determination the potential at any $\boldsymbol{r}_i$, $v_{\mathrm{xc}}(\boldsymbol{r}_i)$ requires the knowledge of global electron density function $\rho(\boldsymbol{r})$. Indeed, Nagai *et al.*[32] proposed a neural network that maps $\rho(\boldsymbol{r})$ to $v_{\mathrm{xc}}(\boldsymbol{r})$, and as a proof of concept, determined the $v_{\mathrm{xc}}(\boldsymbol{r})$ of a one-dimensional two-particle spinless fermion model via a neural network that one-to-one maps its $\rho(\boldsymbol{r})$ to its $v_{\mathrm{xc}}(\boldsymbol{r})$. Just as other existing machine-learnt xc functional,[26, 27] their approach requires the knowledge of the global electron density function $\rho(\boldsymbol{r})$ to determine the $v_{\mathrm{xc}}(\boldsymbol{r})$,[32] and thus cannot be applicable for large molecules as their electron density functions are not accurately known. Handy and co-workers[33] attempted to map one-to-one the electron density $\rho(\boldsymbol{r}_i)$ and the corresponding $v_{\mathrm{xc}}(\boldsymbol{r}_i)$ at one spatial point $\boldsymbol{r}_i$; however, the electron density at single spatial point $\boldsymbol{r}_i$ does not contain enough information such as the environment of the region around this point, and thus yields improved but not very accurate $v_{\mathrm{xc}}(\boldsymbol{r})$.

The holographic electron density theorem states that the electron density function of any finite volume determines in principle the entire electron density function of the real atomic or molecular system.[34, 35] As the electron density distribution $\rho(\boldsymbol{r})$ of a finite volume contains already the knowledge of the global electron density distribution, we do not need the entire electron density

function to determine the xc potential. Instead, the local xc potential $v_{\mathrm{xc}}(\boldsymbol{r}_i)$ at a point $\boldsymbol{r}_i$ can be in principle determined by the electron density distribution $\rho(\boldsymbol{r})$ over a finite volume around $\boldsymbol{r}_i$, *i.e.*, the quasi-local electron density around $\boldsymbol{r}_i$. This is useful as any large molecular system is made of small fragments whose electron density functions are similar to those of the corresponding small molecules, and thus, the neural network that has been trained for the xc potentials of small molecules can also be used to determine the xc potential $v_{\mathrm{xc}}(\boldsymbol{r}_i)$ of the larger molecule, simply by scanning $\boldsymbol{r}_i$ over its entire spatial region.

In this work, we propose to use the quasi-local electron density function around a spatial point to determine the corresponding xc potential at $\boldsymbol{r}_i$, $v_{\mathrm{xc}}(\boldsymbol{r}_i)$, via a deep-learning based scheme. A large amount of highly accurate electron density functions of small molecules and ions and their corresponding xc potentials are required to train the neural network. CCSD/aug-cc-pVQZ[36] is employed to calculate the electron density functions. To determine the corresponding effective xc potentials that reproduce the calculated CCSD electron density functions via Kohn-Sham (KS) self-consistent-field solution,[2] we employ a procedure for the optimized effective potential (OEP) developed by Wu and Yang (Wu-Yang method).[20] In current study, we chose H_2, He_2, and HeH^+ to train and validate the neural network, and in particular, He_2 is chosen because few xc functionals account properly for van der Waals interaction.

We employ a deep neural network to map the quasi-local electron density around $\boldsymbol{r}_i$ onto the local xc potential $v_{\mathrm{xc}}(\boldsymbol{r}_i)$ at $\boldsymbol{r}_i$. It is important to note that the resulting neural network is transferable from smaller molecules to larger molecules. **Figure 1a** illustrates the mapping from the quasi-local electron density onto the local xc potential. The distribution inside the grey cube is the quasi-local electron density and the corresponding potential $v_{\mathrm{xc}}(\boldsymbol{r}_i)$ is the xc potential at $\boldsymbol{r}_i$. We employ a three-dimensional convolutional neural network (3D-CNN, shown in the middle of

Figure 1a) to "learn" the relationship between the quasi-local density and xc potential from the exact or high precision data of electron density functions and xc potentials. The CCSD method is used to compute highly accurate electron density, and the corresponding xc potential of KS equation is obtained via the Wu-Yang method.[20] The 3D-CNN uses the discretized quasi-local densities and their gradients (along *x*-, *y*- and *z*-directions) centered at a spatial point $\boldsymbol{r}_i$ as the inputs or descriptors, and outputs the value of the xc potential at $\boldsymbol{r}_i$, $v_{\mathrm{xc}}(\boldsymbol{r}_i)$. It is worth mentioning that a non-negative penalty term is added in the loss function of the 3D-CNN (see the last term of Eq. S3) to ensure the invariance of the density–potential mapping with respect to the reflection symmetry. By scanning the grey cube $\rho_{\Omega}(\boldsymbol{r}_i)$ over the molecule, we obtain the xc potential of the entire physical space. The details of the Wu-Yang method, 3D-CNN architecture and training process are provided in the Supporting Information.

Self-consistent-field Kohn-Sham/Neural Network scheme (KS-DFT/NN) The 3D-CNN maps the CCSD-precision electron density to its effective KS potential. However, accurate electron density is unknown prior to the calculation, and thus, a self-consistent-field (SCF) procedure is necessary. In our work, the electron density computed with B3LYP is used as the initial guess for the density and is inputted into the 3D-CNN (see the lower part of **Figure 1a**). The output along with the fixed reference potential is thus the initial effective potential which is further subjected to the zero-force condition.[37] The corresponding Kohn-Sham equation is then solved self-consistently. The convergence is examined at each iteration; and if not converged, the resulting electron density is submitted back to the 3D-CNN to yield the updated potential. The above iterative steps are repeated until the convergence is reached.

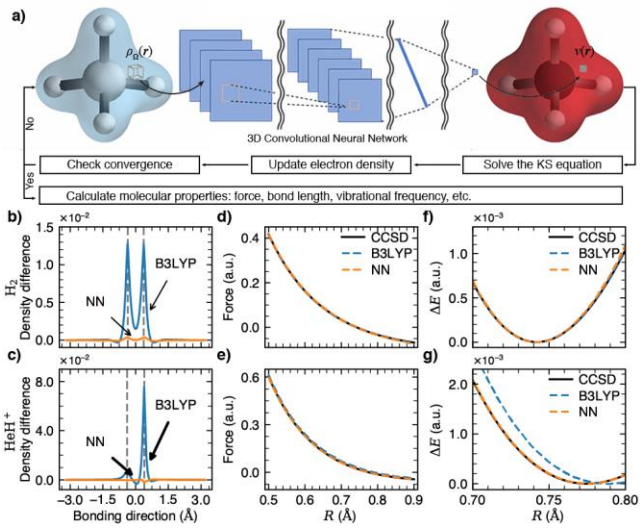


Figure 1. Schematic illustration of the KS-DFT/NN procedure and comparison of calculated results for H_2 and HeH^+. (a) The blue color at the left represents the electron density of the molecule, and the red at the right the xc potential. The quasi-local density $\rho_\Omega(\boldsymbol{r}_i)$ around $\boldsymbol{r}_i$ is represented by a grey cube, and is mapped to the xc potential $v_{\mathrm{xc}}(\boldsymbol{r}_i)$ at the center of the cube $\boldsymbol{r}_i$ (the grey dot) through a 3D convolutional neural network. Given a trained 3D-CNN model and B3LYP density, the potential and electron density can be updated iteratively until converging to a more accurate result. The differences between the calculated electron densities and CCSD benchmarks for equilibrium structures are shown in (b) for H_2 (0.7420 Å) and (c) for HeH^+ (0.7748 Å). Blue and orange lines are for B3LYP and KS-DFT/NN results, respectively. Atomic forces for varieties of structures are plotted in (d) for H_2 and (e) for HeH^+. The relative energies near the equilibrium distances are shown in (f) for H_2 and (g) for HeH^+. For comparison, the lowest energy is shifted to zero.

H_2/HeH^+ system We first trained a 3D-CNN based on H_2 and HeH^+ to demonstrate its capability to learn the density-potential mapping. The dataset is constructed using 50 H_2 and HeH^+ structures with interatomic distance evenly distributed from 0.504 Å to 0.896 Å. As both systems have cylindrical symmetry, the center of quasi-local density $\boldsymbol{r}_i$ is placed on the *xz* plane (parallel to the

internuclear axis), and selected according to a level 3 Lebedev-Laikov quadrature[38] with the cube size a = 0.9 Bohr. Typically, a H_2 molecule has 3382 unique grids and a HeH^+ ion has 3108 ones. In total, the H_2/HeH^+ dataset contains 324500 data points and is randomly divided into a training set of size 243400 (75%) and a validation set of size 81000 (25%).

After training, the ground state electron densities of H_2 and HeH^+ are obtained via KS-DFT/NN SCF calculation using B3LYP density as the initial electron density function and the 3D-CNN to construct effective potential. The use of B3LYP result as initial guess is for the convenience of following comparison and we have shown that KS-DFT/NN calculation can converge to a consistent density function from different starting points (see **Table S1** for details). We use the normalized squared difference I to measure the overlap between two densities ρ_1 and ρ_2:[39]

$$I_{1,2} = \frac{\int |\rho_1(\boldsymbol{r}) - \rho_2(\boldsymbol{r})|^2 \mathrm{d}\boldsymbol{r}}{\int |\rho_1(\boldsymbol{r})|^2 \mathrm{d}\boldsymbol{r} + \int |\rho_2(\boldsymbol{r})|^2 \mathrm{d}\boldsymbol{r}} \tag{2}$$

where the subscripts 1 can be B3LYP, NN (for KS-DFT/NN), and in this work 2 is chosen to be CCSD as the benchmark density. The I value lies between 0 and 1, and smaller $I_{1,2}$ implies better agreement between ρ_1 and ρ_2. We have calculated the I values for 201 H_2 and HeH^+ structures with the internuclear distances distributed evenly from 0.5 Å to 0.9 Å (per 0.002 Å), among which the majority are different from the 50 H_2 and HeH^+ structures in the training and validation set. The average $I_{\mathrm{B3LYP,CCSD}}$ values are 5.9×10^{-5} and 4.7×10^{-5} for H_2 and HeH^+, respectively, while the corresponding $I_{\mathrm{NN,CCSD}}$ are 2.8×10^{-6} and 5.1×10^{-7} for, H_2 and HeH^+, respectively. More detailed $I_{\mathrm{B3LYP,CCSD}}$ and $I_{\mathrm{NN,CCSD}}$ values are summarized in **Table S2**. It is clear that the $I_{\mathrm{NN,CCSD}}$ are at least one order of magnitude smaller than $I_{\mathrm{B3LYP,CCSD}}$, confirming that the electron densities obtained via KS-DFT/NN calculations are closer to the CCSD values than the B3LYP densities. Besides, the average $I_{\mathrm{B3LYP,CCSD}}$ values for those 50 structures within the dataset is 2.7×10^{-6} for H_2 and 5.1×10^{-7} for HeH^+. The similarity between the mean I values for overall

201 structures and the minority 50 ones used for training indicates little over-fitting in the trained 3D-CNN model. To illustrate further, we plot the calculated electron density difference (the CCSD density ρ_{CCSD} as the benchmark) in **Figure 1b** and **1c**. For both H_2 and HeH^+, KS-DFT/NN calculation yields significantly less error than B3LYP. Both *I* values and density differences show convincingly that KS-DFT/NN yields highly accurate electron density (compared to the CCSD benchmark). This is expected as KS-DFT/NN is designed to reproduce the CCSD electron density.

With the highly accurate electron density function, the force on atom can be computed with Hellman-Feynman theorem plus the basis set correction.[40, 41] The resulting internal forces of H_2 and HeH^+ structures are plotted in **Figure 1d** and **1e**, respectively. The relative energy is then calculated by integrating the force (**Figure 1f** and **1g)**. For H_2, three methods, CCSD, B3LYP, and KS-DFT/NN, yield consistent values; and for HeH^+, KS-DFT/NN yields better result than B3LYP. It is important to note that KS-DFT/NN leads high precision electron density and energy simultaneously.

The KS eigenvalues are important properties as in exact DFT calculation, the negative value of HOMO energy (ϵ_{HOMO}) represents the ionization potential E_{IP}. We have examined the HOMO energy of the equilibrium H_2 and HeH^+ structure. For H_2, $-\epsilon_{\mathrm{HOMO}}$ obtained via the KS-DFT/NN calculation is 0.611 Ha, which is much closer to the CCSD's E_{IP} = 0.603 Ha. As a comparison, the $-\epsilon_{\mathrm{HOMO}}$ of B3LYP is 0.434 Ha. For HeH^+, E_{IP} computed using CCSD is 1.630 Ha. KS-DFT/NN leads to a precise $-\epsilon_{\mathrm{HOMO}}$ of 1.631 Ha, while the B3LYP one is 1.376 Ha. We have also obtained the LUMO energy (ϵ_{LUMO}) to see whether the KS-DFT/NN result can be used to predict the electron affinity E_{EA} (although DFT is known to produce unreliable E_{EA} from ϵ_{LUMO}). The ϵ_{LUMO} obtained from KS-DFT/NN for H_2 and HeH^+ is 0.0472 Ha and -0.2866 Ha, respectively, which is quite close to the E_{EA} of 0.0454 Ha (H_2) and -0.2864 Ha (HeH^+) computed using CCSD. It is clear

that KS-DFT/NN yields much better $\epsilon_{\mathrm{HOMO/LUMO}}$ that is useful to measure chemical hardness, since the 3D-CNN provides accurate xc potentials.

To investigate whether the 3D-CNN can "learn" the van der Waals (vdW) interaction from highly accurate dataset, we applied the model to He_2, a vdW system and performed KS-DFT/NN calculation. The smaller $I_{\mathrm{NN,CCSD}}$ values show that the density distribution obtained via the KS-DFT/NN calculation is more accurate than the B3LYP result (see **Table S3** for detailed *I* values). The improvement, however, is not significant, indicating more data and training process is necessary for the 3D-CNN to fully capture the vdW interaction.

Transferability Adoption of quasi-local density (instead of total density function) as the input ensures the transferability of the 3D-CNN from small to large molecular systems. We performed KS-DFT/NN calculation on another 1050 HeH^+ structures with internuclear distance ranging from 0.9 Å to 3.0 Å. Indeed, the average value of $I_{\mathrm{NN,CCSD}}$ for the 1050 structures, 8.5×10^{-6} is less than $I_{\mathrm{B3LYP,CCSD}}$, 5.7×10^{-5}. The typical result of density difference plotted in **Figure 2a** confirms that KS-DFT/NN calculation yields more accurate electron density even for ions that are out of the dataset. **Figure 2b** exhibits the relative energy for HeH^+ ions with internuclear distance up to 3.0 Å. Compared to the B3LYP result, the energy curve obtained through KS-DFT/NN calculation is closer to the CCSD benchmark. Note that the results for the internuclear distance $R_{\mathrm{He–H}} > 0.9$ Å are the extrapolation, as the training dataset contains only the structures with $R_{\mathrm{He–H}} < 0.9$ Å.

To demonstrate the transferability of the 3D-CNN to larger molecule, we choose H_3^+, a system containing three nuclei and two electrons, and He–H–H–He^{2+}, a system containing four nuclei and four electrons, while H_2 and HeH^+ that were used to constructed the 3D-CNN have two nuclei and two electrons. We compute the ground state electron densities and forces of 441 linear H_3^+ and He–H–H–He^{2+} structures with He–H and H–H distances varying equally between 0.60 Å and

0.80 Å (0.01 Å per bond length variation) using CCSD, B3LYP and KS-DFT/NN. In **Figure 2c** and **2d**, we examine the accuracy of B3LYP and KS-DFT/NN densities of typical H_3^+ and He–H–H–He^{2+} structure by plotting their differences compared to the CCSD results along the internuclear axis (more results are provided in **Figure S2** and **S4**). As anticipated, the KS-DFT/NN density exhibits smaller and narrower error distribution than B3LYP. $I_{B3LYP,CCSD}$ and $I_{NN,CCSD}$ values further show that the KS-DFT/NN leads to the electron densities that are much closer to CCSD densities than B3LYP for the 441 structures (see **Figure S3** and **S5**). All these show that KS-DFT/NN yields much better results than B3LYP for H_3^+ and He–H–H–He^{2+}, and thus confirm that the 3D-CNN which is trained and verified on H_2/HeH^+ dataset is indeed transferable to the larger molecules.

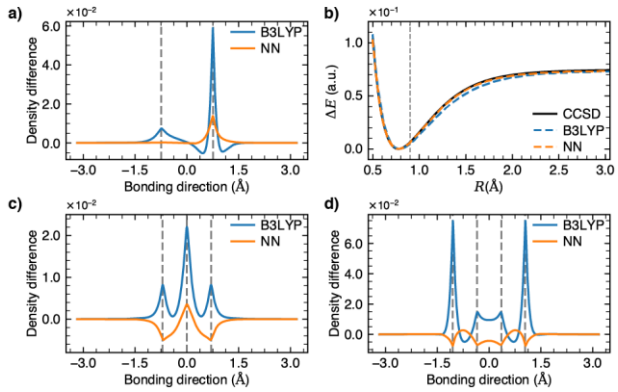


Figure 2. Comparison of B3LYP and KS-DFT/NN results on out-of-dataset HeH^+ and He–H–H–He^{2+}. (a) Density difference of HeH^+ along the internuclear axis with $R_{He–H}$ = 1.50 Å. (b) Relative energy of HeH^+. The lowest energy is shifted to zero for comparison. The vertical dashed line at $R_{He–H}$ = 0.9 Å indicates that the training dataset contains only the data for $R_{He–H}$ < 0.9 Å. (c) Density difference of linear H_3^+ with $R_{H–H}$ = 0.7 Å. (d) Density difference of linear He–H–H–He^{2+} along the internuclear axis with $R_{He–H}$ = $R_{H–H}$ = 0.70 Å.

Van der Waals interaction: He_2 We choose He_2 to examine the ability of the 3D-CNN to model van der Waals (vdW) systems. Since the distance between two He atoms is much larger than the

typical bonding length, a larger cube with size of a = 3.0 Bohr is used. Similar to the H_2 and HeH^+, a level 3 quadrature is employed to select the grids for He_2, and the centers of cubes are placed at the *xz*-plane. We compute 61 He_2 structures with internuclear distance varying from 2.00 Å to 3.20 Å. In total, the dataset consisting of 86437 data points is divided into the training and validation set according to the ratio of 4:1.

Employing KS-DFT/NN, we compute the ground state electron densities of 121 He_2 structures with internuclear distance evenly distributed between 2.00 Å and 3.20 Å; and B3LYP density is taken as the initial guess. The average value is 4.5×10^{-5} for $I_{\mathrm{B3LYP,CCSD}}$ and 8.3×10^{-9} for $I_{\mathrm{NN,CCSD}}$. Other representative values are available in **Table S4**. Compared to B3LYP density, KS-DFT/NN density is about 4 order closer to the CCSD values. This confirms that the slight density perturbation can be captured by the 3D-CNN which yields the correct xc potential that accounts for vdW interaction. In **Figure 3a**, we plot the density deviations (compared to CCSD density) for B3LYP and KS-DFT/NN in *xz*-plane for the equilibrium He_2 structure, respectively. KS-DFT/NN density is far closer to CCSD values. The relative energy of He_2 versus the internuclear distance obtained by integrating the force is shown in **Figure 3b** and **3c**. The KS-DFT/NN potential energy curve is closer to CCSD values, leading to far better equilibrium geometry than B3LYP.

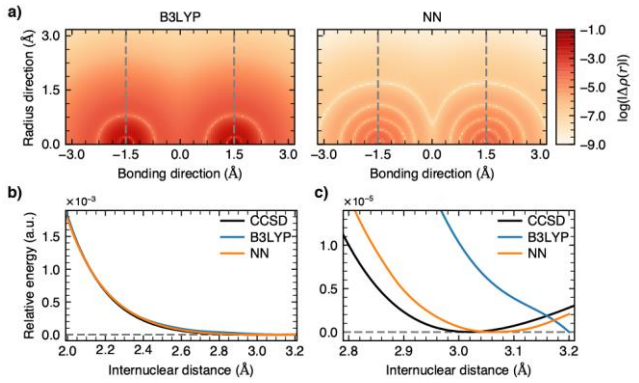


Figure 3. Comparison among CCSD, B3LYP and KS-DFT/NN results for He_2. (a) Absolute density deviations compared to the CCSD benchmark for B3LYP (left) and KS-DFT/NN (right)

result in *xz*-plane (logarithmic scale). The He–He distance is 3.00 Å. (b) and (c) Relative energy versus the internuclear distance ranging from (b) 2.00 Å to 3.20 Å and (c) 2.80 Å to 3.20 Å.

H_2/HeH^+/He_2 system Now we combine the datasets of H_2, HeH^+, and He_2 as the H_2/HeH^+/He_2 dataset, and train/validate the 3D-CNN again with a = 3.0 Bohr. The resulting model is then used to calculate various properties of H_2, HeH^+, He_2, and He–H–H–He^{2+}. As shown in **Figure 4a** to **4c**, KS-DFT/NN densities of H_2, HeH^+, and He_2 coincide far better with CCSD benchmark than B3LYP results, confirming again the accuracy and reliability of our KS-DFT/NN method. KS-DFT/NN calculation is also carried out on the 441 He–H–H–He^{2+} structures. Compared to CCSD benchmarks, the density deviations of KS-DFT/NN are much less than that of B3LYP, as shown in **Figure 4d** and **4e**, implying that the new 3D-CNN constructed on the H_2/HeH^+/He_2 dataset yields the accurate xc potential for He–H–H–He^{2+}, and thus, confirming again its transferability. More comparison of electron density of individual structures and I values are provided in **Figure S6** and **S7**.

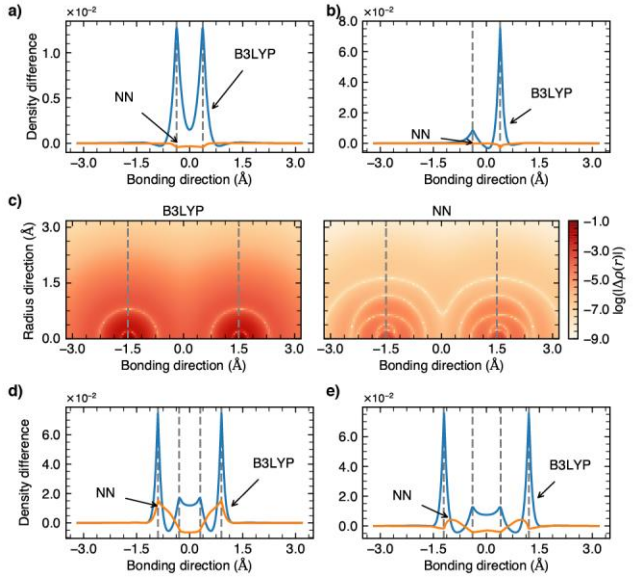


Figure 4. Comparison of B3LYP and KS-DFT/NN densities against CCSD benchmarks. Upper panel: density deviations for (a) H_2 with the internuclear distance at 0.7420 Å and (b) HeH^+ at

0.7748 Å. The nuclear positions are marked with the vertical dashed lines. (c) Density deviation for equilibrium He_2 structures with the internuclear distance at 3.00 Å in *xz*-plane calculated using B3LYP and KS-DFT/NN (logarithmic scale). Lower panel: density deviation for $He–H–H–He^{2+}$ structures with $R_{He–H}$ and $R_{H–H}$ equal to (d) 0.60 Å and (e) 0.80 Å, respectively.

With precision KS-DFT/NN densities we further calculate the forces for H_2, HeH^+ and He_2, and determine their equilibrium structures and the corresponding vibrational frequencies numerically. Results are summarized and compared to CCSD and B3LYP in **Table 1**. KS-DFT/NN results are very close to CCSD values, and overall more precise than B3LYP ones. In particular, KS-DFT/NN results on He_2 are far better than B3LYP: KS-DFT/NN equilibrium distance is 3.02 Å, comparable with 3.01 Å of CCSD, and the vibrational frequency is 32.7 cm^{-1}, comparable with 31.4 cm^{-1} of CCSD. In contrast, B3LYP yields an equilibrium distance of 4.04 Å and a vibrational frequency of 18.2 cm^{-1}.

Table 1. Summary of calculated equilibrium properties of H_2, HeH^+, and He_2

Mol.	Method*	Eq. dis. (Å)	Freq. (cm^{-1})
H_2	CCSD	0.742	4398
	B3LYP	0.742	4414
	KS-DFT/NN	0.744	4377
HeH^+	CCSD	0.775	3219
	B3LYP	0.788	3072
	KS-DFT/NN	0.774	3195
He_2	CCSD	3.01	31.4†
	B3LYP	4.04	18.2†
	KS-DFT/NN	3.02	32.7†

*CCSD and B3LYP results are obtained via analytical geometry optimization

[†]vdW potential of He_2 is anharmonic, and thus, the calculated vibration frequency is theoretical and for the assessment of calculation accuracy only.

Our 3D-CNN aims to yield the xc potential that reproduces high precision electron density, and as a result, the corresponding KS-DFT/NN scheme results in highly accurate electron density; with the accurate electron density, precision force is obtained, and further the relative energy and other properties such as the vibrational frequency. Our KS-DFT/NN scheme results in both precision electron density and also accurate energetics. One stone two birds!

We have addressed a chronic problem of DFT: van der Waals interaction, and demonstrated convincingly that the 3D-CNN for xc potential is sensitive enough to capture the slight variance of electron density caused by weak vdW interaction and thus model the vdW systems such as He_2. Machine learning requires data, and yet unfortunately, data for large molecular systems are rarely available. This poses a serious obstacle for constructing machine learning based xc functional for large molecules. We circumvent the problem by mapping quasi-local electron density to local xc potential, and thus, the neural network trained with data of small molecules can yield the xc potentials for larger systems. In this work, the 3D-CNN constructed on the data sets of H_2 and HeH^+ can be used to yield accurately the xc potential of H_3^+ and $He–H–H–He^{2+}$, and the KS-DFT/NN calculation yields much better electron density of H_3^+ and $He–H–H–He^{2+}$ than B3LYP. Transferability from small to larger systems is another essential step for developing the universal machine learning based xc potential. Going forward, our approach can be extended to more complex systems like organic molecules using molecular fragments containing several atoms as quasi-local density inputs for neural networks. Of course, more comprehensive investigation will be necessary to construct better dataset and neural network architectures.

ASSOCIATED CONTENT

Supporting Information Available: Details of methodology and comparison of electron densities for more structures (PDF).

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We thank Prof. Eberhard K. U. Gross (Max Planck Institute of Microstructure Physics) and Prof. Garnet K. Chan (California Institute of Technology) for the helpful discussion. The financial support from Research Grant Council of HKSAR (AoE/P-04/08 and 17316016) is acknowledged.

REFERENCE

(1) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964,** *136*, B864-B871.

(2) Kohn, W.; Sham, L. J. Self-consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965,** *140*, A1133-A1138.

(3) Zhang, Y.; Yang, W. A Challenge for Density Functionals: Self-interaction Error Increases for Systems with a Noninteger Number of Electrons. *J. Chem. Phys.* **1998,** *109*, 2604-2608.

(4) Hu, L.; Wang, X.; Wong, L.; Chen, G. Combined First-Principles Calculation and Neural-Network Correction Approach for Heat of Formation. *J. Chem. Phys.* **2003,** *119*, 11501-11507.

(5) Kristyán, S.; Pulay, P. Can (Semi)Local Density Functional Theory Account for the London Dispersion Forces? *Chem. Phys. Lett.* **1994,** *229*, 175-180.

(6) Pérez-Jordá, J.; Becke, A. D. A Density-Functional Study of van der Waals Forces: Rare Gas Diatomics. *Chem. Phys. Lett.* **1995,** *233*, 134-137.

(7) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Development and Assessment of a New Hybrid Density Functional Model for Thermochemical Kinetics. *J. Phys. Chem. A* **2004,** *108*, 2715-2719.

(8) Zhao, Y.; Truhlar, D. G. Benchmark Databases for Nonbonded Interactions and Their Use to Test Density Functional Theory. *J. Chem. Theory Comput.* **2005,** *1*, 415-432.

(9) Cai, Z.-L.; Sendt, K.; Reimers, J. R. Failure of Density-Functional Theory and Time-Dependent Density-Functional Theory for Large Extended π-Systems. *J. Chem. Phys.* **2002,** *117*, 5543-5549.

(10) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Fractional Charge Perspective on the Band Gap in Density-Functional Theory. *Phys. Rev. B* **2008,** *77*, 115123.

(11) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Localization and Delocalization Errors in Density Functional Theory and Implications for Band-Gap Prediction. *Phys. Rev. Lett.* **2008,** *100*, 146401.

(12) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Fractional Spins and Static Correlation Error in Density Functional Theory. *J. Chem. Phys.* **2008,** *129*, 121104.

(13) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012,** *112*, 289-320.

(14) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. Density Functional Theory is Straying from the Path toward the Exact Functional. *Science* **2017,** *355*, 49-52.

(15) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996,** *77*, 3865-3868.

(16) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003,** *91*, 146401.

(17) Becke, A. D. Density-Functional Thermochemistry. V. Systematic Optimization of Exchange-Correlation Functionals. *J. Chem. Phys.* **1997,** *107*, 8554-8560.

(18) Perdew, J. P.; Schmidt, K. Jacob's Ladder of Density Functional Approximations for the Exchange–Correlation Energy. *AIP Conf. Proc.* **2001,** *577*, 1-20.

(19) Kanungo, B.; Zimmerman, P. M.; Gavini, V. Exact Exchange-Correlation Potentials from Ground-State Electron Densities. *Nat. Commun.* **2019,** *10*, 4497.

(20) Wu, Q.; Yang, W. A Direct Optimization Method for Calculating Density Functionals and Exchange–Correlation Potentials from Electron Densities. *J. Chem. Phys.* **2003,** *118*, 2498-2509.

(21) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modelling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012,** *108*, 058301.

(22) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham Equations with Machine Learning. *Nat. Commun.* **2017,** *8*, 872.

(23) Chandrasekaran, A.; Kamal, D.; Batra, R.; Kim, C.; Chen, L.; Ramprasad, R. Solving the Electronic Structure Problem with Machine Learning. *npj Comput. Mater.* **2019,** *5*, 22.

(24) Ryczko, K.; Strubbe, D. A.; Tamblyn, I. Deep Learning and Density-Functional Theory. *Phys. Rev. A* **2019,** *100*, 022512.

(25) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012,** *108*, 253002.

(26) Zheng, X.; Hu, L.; Wang, X.; Chen, G. A Generalized Exchange-Correlation Functional: The Neural-Networks Approach. *Chem. Phys. Lett.* **2004,** *390*, 186-192.

(27) Snyder, J. C.; Rupp, M.; Hansen, K.; Blooston, L.; Müller, K.-R.; Burke, K. Orbital-Free Bond Breaking via Machine Learning. *J. Chem. Phys.* **2013,** *139*, 224104.

(28) Yao, K.; Parkhill, J. Kinetic Energy of Hydrocarbons as a Function of Electron Density and Convolutional Neural Networks. *J. Chem. Theory Comput.* **2016,** *12*, 1139-1147.

(29) Seino, J.; Kageyama, R.; Fujinami, M.; Ikabata, Y.; Nakai, H. Semi-Local Machine-Learned Kinetic Energy Density Functional with Third-Order Gradients of Electron Density. *J. Chem. Phys.* **2018,** *148*, 241705.

(30) Ceperley, D.; Alder, B. Quantum Monte Carlo. *Science* **1986,** *231*, 555-560.

(31) Chan, G. K.-L.; Head-Gordon, M. Highly Correlated Calculations with a Polynomial Cost Algorithm: A Study of the Density Matrix Renormalization Group. *J. Chem. Phys.* **2002,** *116*, 4462-4476.

(32) Nagai, R.; Akashi, R.; Sasaki, S.; Tsuneyuki, S. Neural-Network Kohn-Sham Exchange-Correlation Potential and Its Out-Of-Training Transferability. *J. Chem. Phys.* **2018,** *148*, 241737.

(33) Tozer, D. J.; Victoria, I. E.; Handy, N. C. Exchange-Correlation Potentials. *J. Chem. Phys.* **1996,** *105*, 9200-9213.

(34) Riess, J.; Münch, W. The Theorem of Hohenberg and Kohn for Subdomains of a Quantum System. *Theor. Chim. Acta* **1981,** *58*, 295-300.

(35) Zheng, X.; Wang, F.; Yam, C. Y.; Mo, Y.; Chen, G. Time-Dependent Density-Functional Theory for Open Systems. *Phys. Rev. B* **2007,** *75*, 195127.

(36) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of Static Electrical Response Properties. *J. Chem. Phys.* **1994,** *100*, 2975-2988.

(37) Kurzweil, Y.; Head-Gordon, M. Improving Approximate-Optimized Effective Potentials by Imposing Exact Conditions: Theory and Applications to Electronic Statics and Dynamics. *Phys. Rev. A* **2009,** *80*, 012509.

(38) Lebedev, V. I.; Laikov, D. N. A Quadrature Formula for the Sphere of the 131st Algebraic Order of Accuracy, *Dokl. Math.* **1999,** *59*, 477-481.

(39) Bochevarov, A. D.; Friesner, R. A. The Densities Produced by the Density Functional Theory: Comparison to Full Configuration Interaction. *J. Chem. Phys.* **2008,** *128*, 034102.

(40) Feynman, R. P. Forces in Molecules. *Phys. Rev.* **1939,** *56*, 340-343.

(41) Pulay, P. Ab Initio Calculation of Force Constants and Equilibrium Geometries in Polyatomic Molecules. *Mol. Phys.* **1969,** *17*, 197-204.