Deep Charge: Deep learning model of electron density from a one-shot density functional theory calculation

Taoyuze Lv [1], Zhicheng Zhong [2,*], Yuhang Liang [3], Feng Li [1], Jun Huang [3], and Rongkun Zheng [1,†]

[1] *School of Physics, The University of Sydney, NSW 2006, Australia*

[2] *Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Science, Ningbo, Zhejiang 315201, China*

[3] *School of Chemical and Biomolecular Engineering, The University of Sydney, NSW 2006, Australia*

(Received 25 September 2023; revised 14 November 2023; accepted 30 November 2023; published 20 December 2023)

Electron charge density is a fundamental physical quantity, determining various properties of matter. In this study, we have proposed a deep learning model for accurate charge-density prediction. Our model naturally preserves physical symmetries and can be effectively trained from one-shot density functional theory calculation toward high accuracy. It captures detailed atomic environment information, ensuring accurate predictions of charge density across bulk, surface, molecules, and amorphous structures. This implementation exhibits excellent scalability and provides efficient analyses of material properties in large-scale condensed matter systems.

DOI: 10.1103/PhysRevB.108.235159

I. INTRODUCTION

In the vast domain of materials science and condensed matter physics, electron charge density is a fundamental property that describes the spatial distribution of electrons in a material. It provides invaluable insights into overall material behaviors and is an essential quantity in various quantum-mechanical frameworks, especially in density functional theory (DFT) [1], where the ground-state properties of a many-body system can be described as functionals of the electron charge density. Commonly, DFT solves the Kohn-Sham (KS) equations [2] via the self-consistent loop [3] to adopt the ground-state charge density $\varrho(\boldsymbol{r}, \{\boldsymbol{R}_i\}_{i=1}^n)$, where charge density ϱ is a function of the space position $\boldsymbol{r}$ and the positions of all n atoms in the system $\{\boldsymbol{R}_i\}_{i=1}^n$. A significant challenge of this approach is the computational complexity of $O(N^3)$ with respect to the number of electrons. The harshness of density functional optimization impedes the application of DFT in mesoscopic systems with thousands or even millions of atoms. This limitation has spurred the search for alternative approaches that can efficiently represent charge density without compromising accuracy.

The electron density generally obtains "nearsightedness" and screening effect in many-atom systems [4,5]. These effects make it possible to introduce a radius r_c around a space point, beyond which any perturbation on the atomic environment only provides negligible impacts on the electron density at that point. This is the locality property of electron density, which does not apply to the KS wave functions. In that case, the original time-consuming KS self-consistent calculation can be turned into mapping local atomic environment information to the electron density. Such a mapping can naturally be obtained via techniques such as machine learning. A well-trained machine learning model can be used for predicting charge density in different atomic systems, and even be scalable for determining large-scale material systems. Currently, several machine learning models have been proposed to predict electron charge density as alternatives to first-principles calculations, as listed in Table I.

One commonly used strategy is to express the charge density as the summation of a set of basis functions such as spherical harmonics, with the coefficients fitted via machine learning [6–12]. The rotational invariance of charge density can be automatically protected with a reasonable basis set. These methods are efficient but mainly used in charge-localized systems, and the accuracy is highly dependent on the quality and complexity of basis functions and usually requires DFT calculations of plenty of different atomic structures to train. On the other hand, the grid-based methods can be adopted to obtain more flexibility [13–18]. These methods focus on mapping every unique local environment around real space points to nondegenerate values in the latent space, which thus contains enough information for charge-density representation. The mapping should generally be invariant under three physical symmetry operations: translation, rotation, and permutation, as will be discussed in detail below. Achieving these without losing structural detail information is a nontrivial task. Moreover, grid-point methods usually require large numbers of data points from various structures to obtain satisfactory accuracy.

Herein, to overcome the above challenges for the grid-based method, we introduce a deep learning charge model (Deep Charge) to represent electron charge density. We have adopted the embedding framework from Deep Potential [19] to extract the local atomic environment, which naturally preserves the system's symmetry. This implementation requires very few structures for training. Remarkably, a one-shot DFT calculation is sufficient for simple systems like crystalline silicon and aluminum to reach a mean absolute error (MAE) below 7×10^{-4} e/Å^3. The model provides a scalable

*zhong@nimte.ac.cn

†rongkun.zheng@sydney.edu.au

TABLE I. The published works on machine learning charge density. The top seven methods belong to basis-function methods, while the bottom six are grid-point methods. The unit for MAE and RMSE values is e/Å^3.

Method	Systems	Accuracy	Data size	Ref.
SA-GPR	Small molecules	MAE: 1.0%∼1.2%	∼1000 structures	[6]
SA-GPR	Organic dimers	rMAE[b]: 0.3%∼1.8%	∼2300 structures	[7]
SALTED	Al, Si, I_h−Ice	%RMSE[c]:∼1.5%	100 structures	[8]
EGNN[a]	DNA	rMAE ∼1%	922 structures	[9]
SchNOrb	Small molecules		25 000 structures	[10]
EGNN	Water clusters	rMAE: ∼1.4%	6000 structures	[11]
ML-HK map	Small molecules		100 ∼ 350 structures	[12]
Gaussians	PE and Al thin films	RMSE[d]: $\sim 5 \times 10^{-4}$	$>10^7$ points	[13]
CGCNN	Polymer, zeolite	RMSE: 0.04 ∼ 0.18	$>3 \times 10^5$ points	[14]
Deep Density	Molecules, Al	rRMSE[e]: 0.5%∼2.2%	$10^5 \sim 10^6$ points	[15]
Gaussians	Organics	RMSE: 0.03 ∼ 1.52	$\sim 1.5 \times 10^9$ points	[16]
EGNN	Molecules, cathode	rMAE: 0.06%∼0.27%	$10^3 \sim 10^5$ structures	[17]
JLCDMs	Benzene, Al, MoS_2	RMSE: 0.006 ∼ 0.008	$\sim 1.6 \times 10^6$ points	[18]
This work	Si, Al, Al-Mg, water	MAE: 0.05%∼0.4%	$>10^5$ points	

[a]Equivariant graph neural network.
[b]Relative mean absolute error. $\sum_i^N |\rho_i - \hat{\rho}_i| / \sum_i^N |\rho_i|$, in which ρ_i is the real charge density of the ith grid point; $\hat{\rho}_i$ is the represented value of that point.
[c]Percentage root-mean-square error, defined in the original paper.
[d]Root-mean-square error. $\sqrt{\sum_i^N (\rho_i - \hat{\rho}_i)^2}/N$, in which N is the total number of grid points.
[e]Relative root-mean-square error. $\sqrt{\sum_i^N (\rho_i - \hat{\rho}_i)^2}/\sqrt{\sum_i^N \rho_i^2}$.

representation of charge density in bulk, surface, and complex systems such as alloy and amorphous, involving semiconducting and metallic systems. Our Deep Charge model can thus efficiently assist in the prediction and analysis of charge density in large-scale condensed matter systems with DFT accuracy.

II. DEEP CHARGE MODEL

The most important reasons why the Deep Charge model can effectively capture the physics consistent with KS-DFT are the nearsightedness [5] and the screening effect [20,21] of materials. When a perturbation in the atomic structure is introduced within a local environment of a material, the ground-state distribution of electrons adjusts accordingly. The magnitude of the charge-density change diminishes with increasing distance from the perturbation due to the destructive interference of the density amplitude concerning single-particle eigenstates. In gapped systems or disordered systems, the decay is exponential, causing a rapid decrease in influence. Meanwhile, in ordered gapless systems such as metals, the decay follows power laws rather than exponential, breaking the nearsightedness to some extent. On the other hand, the screening effect of metallic electrons shields the long-range Coulomb potential from the perturbed atoms, forcing the system to be nearsighted. These results ensure that almost all the information about local charge density is encoded in the atomic structures within a suitable cutoff radius. The Deep Charge model can thus be trained to establish the relationship between the charge density and atomic structures approaching the DFT accuracy.

The continuous charge density is discretized into a space grid in grid-point methods. Suppose there are n atoms in the system; the charge-density value at position $\mathbf{r}$ can be written as a function of the coordinates of atoms $\{\boldsymbol{R}_i\}_{i=1}^n$,

$$\rho = \varrho\left(\boldsymbol{r}, \{\boldsymbol{R}_i\}_{i=1}^n\right), \tag{1}$$

where all the information of the system is required to determine the density at a single point. Instead, we can use only the coordinates of neighboring atoms $\{\boldsymbol{R}_i | i \in n_{r_c}(\boldsymbol{r})\}$, where $n_{r_c}(\boldsymbol{r})$ denotes the set of atoms within the cutoff radius r_c surrounding $\mathbf{r}$ with a size of n_c.

$$\varrho\left(\boldsymbol{r}, \{\boldsymbol{R}_i\}_{i=1}^n\right) \cong \varrho(\boldsymbol{r}, \{\boldsymbol{R}_i | i \in n_{r_c}(\boldsymbol{r})\}). \tag{2}$$

Ideally, this charge-density representation should be designed carefully to fulfill the invariance under three symmetry operations:

1. Any translation $\hat{\mathcal{T}}_{\boldsymbol{d}}$ of the whole system $\hat{\mathcal{T}}_{\boldsymbol{d}}\varrho(\boldsymbol{r}, \{\boldsymbol{R}_i\}_{i=1}^n) = \varrho(\boldsymbol{r}+\boldsymbol{d}, \{\boldsymbol{R}_i+\boldsymbol{d}\}_{i=1}^n)$ should result in the same value as $\varrho(\boldsymbol{r}, \{\boldsymbol{R}_i\}_{i=1}^n)$;
2. Any operation $\hat{O} \in \mathrm{O}(3)$ centered at $\mathbf{r}$ should obey $\hat{O}\varrho(\boldsymbol{r}, \{\boldsymbol{R}_i\}_{i=1}^n) = \varrho(\boldsymbol{r}, \{\boldsymbol{R}_i\}_{i=1}^n)$;
3. Any permutation $\hat{\pi}$ on labels of the same atomic species fulfills $\varrho(\boldsymbol{r}, \{\boldsymbol{R}_{\hat{\pi}(i)}\}_{i=1}^n) = \varrho(\boldsymbol{r}, \{\boldsymbol{R}_i\}_{i=1}^n)$.

To preserve the translational invariance of the density, the relative coordinates $\boldsymbol{r}_i = \boldsymbol{r} - \boldsymbol{R}_i = (x_i, y_i, z_i)$ are adopted:

$$\varrho(\boldsymbol{r}, \{\boldsymbol{R}_i | i \in n_{r_c}(\boldsymbol{r})\}) = \varrho(\{\boldsymbol{r}_i | i \in n_{r_c}(\boldsymbol{r})\}). \tag{3}$$

The density can be represented via neural networks as

$$\varrho(\{\boldsymbol{r}_i | i \in n_{r_c}(\boldsymbol{r})\}) = \mathcal{F}(\mathcal{D}(\{\boldsymbol{r}_i | i \in n_{r_c}(\boldsymbol{r})\}; \boldsymbol{\theta}_d); \boldsymbol{\theta}_f), \tag{4}$$

where $\mathcal{D}$ represents the descriptor that extracts all the local environment information, and $\mathcal{F}$ represents the fitting net for charge density. The trainable neural network parameters for them are $\boldsymbol{\theta}_d$ and $\boldsymbol{\theta}_f$, respectively. In the implementation

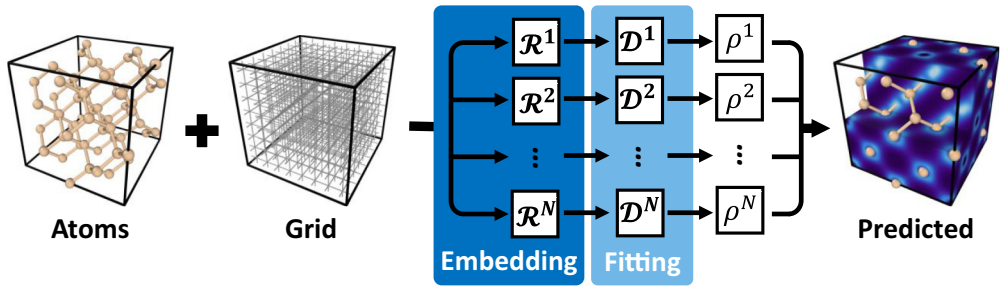

FIG. 1. Workflow of Deep Charge prediction. Suppose there are N grid points in the system; each grid point goes through one network separately.

of Deep Charge, the descriptor is constructed to conserve the rotation and permutation symmetry. Herein, we use the two-body embedding DEEPPOT-SE descriptor with radial and angular information [22]:

$$\mathcal{D} = \frac{1}{n_c^2}\mathcal{G}^T\mathcal{R}\mathcal{R}^T\mathcal{G}_<, \tag{5}$$

in which $\mathcal{R} \in \mathbb{R}^{n_c\times\{1,\,4\}}$ is the coordinate matrix with the form of

$$\mathcal{R}_i = s(r_i)\left(1, \frac{x_i}{r_i}, \frac{y_i}{r_i}, \frac{z_i}{r_i}\right), \tag{6}$$

in which $r_i = \|\boldsymbol{r}_i\|$ is the norm of relative coordinates. $s(r)$ is the switching function defined with a fifth-order polynomial:

$$s(r) = \begin{cases} \frac{1}{r}, & r < r_s \\ \frac{[u^3(-6u^2+15u-10)+1]}{r}, & r_s \leqslant r < r_c,\ u = \dfrac{r-r_s}{r_c-r_s} \\ 0, & r \geqslant r_c \end{cases} \tag{7}$$

where r_s is the radius where smoothing starts. The construction of $\mathcal{R}\mathcal{R}^T$ is naturally invariant under O(3) symmetry operation. The embedding matrix $\mathcal{G}$ is for obtaining the permutation invariance according to Ref. [23], and $\mathcal{G}_<$ is the first few columns of $\mathcal{G}$ (ten columns in our case). The form of $\mathcal{G}$ is trained with neural networks $\mathcal{N}$:

$$\mathcal{G}_i = \mathcal{N}(s(r_i)). \tag{8}$$

The fitting net $\mathcal{F}$ obtains a simple fully connected feed-forward network structure. The detailed charge-density prediction workflow is shown in Fig. 1. The continuous space is discretized into grids. The model will take the atomic configuration and grid points as input and find the local environment for each grid point in parallel, which will be used in the embedding networks to construct the descriptor and then input to the fitting networks. The final charge density can thus be collected from the grid points.

Other similar grid-point methods with high accuracy, such as those that use Gaussian fingerprints [13] and Jacobi-Legendre charge density models (JLCDMs) [18], usually require a predefined set of filter functions to construct fingerprints. These functions are applied to the relative distances between the grid points and each surrounding atom and are then summed over the results from all the atoms to obtain the fingerprint. Consequently, the number of filter functions corresponds to the number of consistent fingerprints. This construction method has the advantage of reducing the complexity of training, as environmental information is extracted via manually designed functions. Meanwhile, the quality of extracted information can highly depend on the quality and number of the filters and may lack flexibility in very complex circumstances. In the case of Deep Charge, on the other hand, the descriptor is trainable, providing additional degrees of freedom to map local environmental information into latent space. This implementation endowed the model with more adaptability to complex systems such as amorphous structures, as demonstrated and discussed later in this paper.

III. RESULTS AND DISCUSSION

A. Performance in bulk systems

In this section, we examine the convergence relation between the number of data points for training and the accuracy in bulks. We chose crystalline silicon with 64 atoms for this study. Silicon is an extensively studied semiconductor with localized valence electrons at bond positions. The training sets were based on the charge density obtained from only one structure, while the testing was performed on all the grid points from an unseen structure. The results are shown in Fig. 2. In addition, we compared the convergence with models trained on five different structures, 10^5 data points for each structure selected, displayed with blue dashed lines in Fig. 2(a). The model can reach an MAE of 6.7×10^{-4} e/Å^3 trained with 10^5 data points. It is noticeable that the one-structure models exhibit similar accuracy to the five-structure models, indicating that a little charge-density information of one single structure of structure is enough to learn the patterns in crystals. The parity plot of Fig. 2(b) has a coefficient of determination (R^2) of 0.999 957, also indicating a nearly linear relationship. Meanwhile, as shown in Figs. 2(c)–2(f), the absolute charge-density errors are larger around the center of atoms, where the fluctuation and charge density are also higher. The absolute error distribution is also shown in Fig. S1(a) in Supplemental Material (SM) [24] (see also Refs. [3,25–31] therein).

B. Performance in surface systems

To test the capability of the model under different atomic environments, we further evaluated the charge-density representation in surface structures. The aluminum surface with 144 atoms was chosen, because it is a benchmark material for testing charge-density representation with highly delocalized valence electrons [8,13,15,18]. As shown in

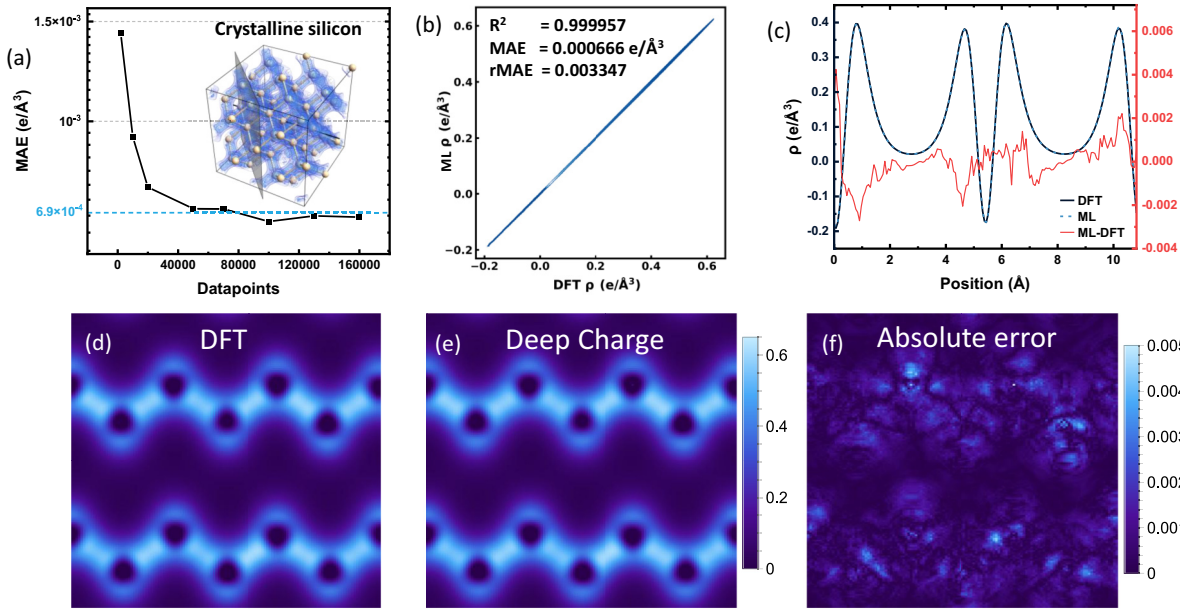

FIG. 2. The performance of the Deep Charge model on bulk silicon. (a) The convergence relation between the number of data points used for training and the MAE values for the testing set. (b) The scatter plot of predicted and test densities from a whole testing structure. Notice that the negative values of charge density are induced by the on-site correction terms in projector augmented-wave (PAW) pseudopotentials. (c) The charge density along the line is indicated in the inset of (a). (d), (e) The charge density computed from DFT and machine learning, respectively. The slice is denoted in the inset of (a). (f) The absolute error. Color bars are in the unit of $e/Å^3$.

Figs. 3(a) and 3(b), the accuracy of our model approached below $6 \times 10^{-4}\,e/Å^3$ after training with 10^5 data points, which is quite similar to the five-structure model. We have extracted the charge density along the line across the surface and bond positions, as shown in Fig. 3(c). The charge density near the surface is lower than that in the bulk. This is owing to the surface relaxation; the atoms near the surface are loosely bonded, reducing the charge density between atoms. Moreover, the charge density inside the film also varies considerably with the disturbance of atomic structures than that in silicon, which is also clearly shown in Fig. 4. The results from the Deep Charge model greatly overlap that from the DFT calculation, meaning that the embedding approach can accurately capture the subtle changes in the environment, with the error fluctuating under $0.004\,e/Å^3$ along the line.

In addition, many other works have tested their model in aluminum with different system sizes and structures. Thus, we herein briefly compare other models with ours. The SALTED [8], Deep Density [15], and JLCDM [18] models have been examined in small bulk aluminum systems, in which the errors are ∼1% [%RMSE (root-mean-square error), 4 atoms], $2 \times 10^{-2}\,eV/Å^3$ (MAE, 32 atoms), $4.81 \times 10^{-4}\,eV/Å^3$ (MAE, 32 atoms), respectively. For comparison, we trained and tested our model in the same systems and obtained an MAE of $3.37 \times 10^{-4}\,eV/Å^3$, lower than their values. Meanwhile, Gaussian fingerprints [13] can reach an RMSE of about $10^{-3}\,eV/Å^3$ in six-layer thin-film systems (144 atoms) with about 10^5 datapoints, while the error can be further decreased to $6 \times 10^{-4}\,eV/Å^3$ but requires two orders of magnitude more data than us. Overall, the Deep Charge model can provide an

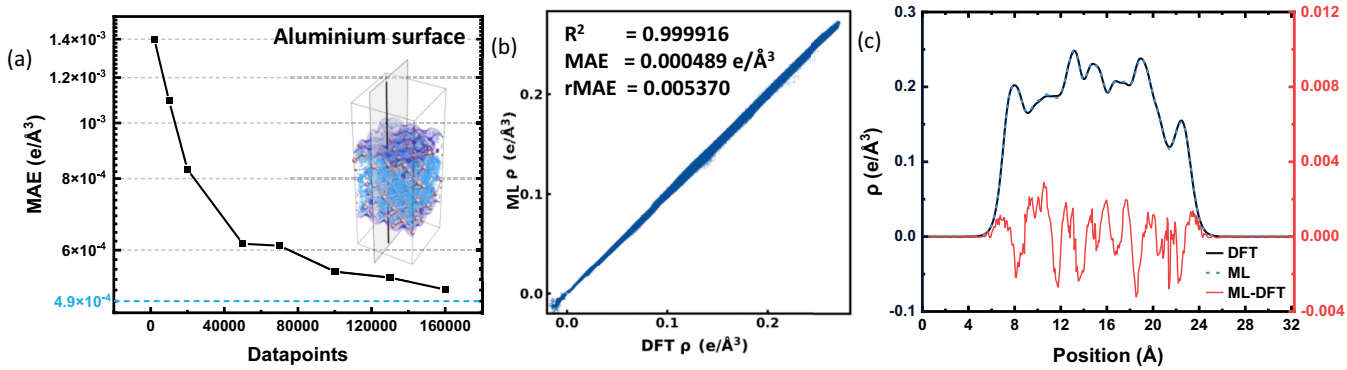

FIG. 3. (a) The convergence relation of machine learning model in aluminum surface. (b) The scatter plot of predicted and test densities. (c) The charge density along the line is indicated in the inset of (a).

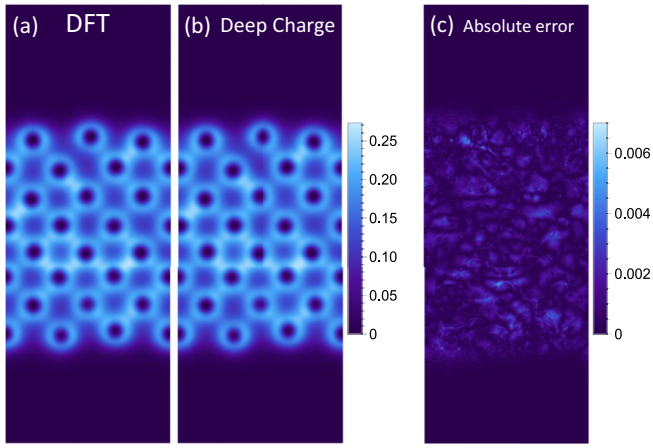

FIG. 4. (a), (b) The computed charge density from DFT and machine learning, respectively. The slice is denoted in Fig. 3(a). (c) The absolute error.

efficient and accurate way of representing charge density in aluminum.

C. Performance and scalability in complex structures

In this section, we examine the scalability of the machine learning model in complex structures. Scalability is one of the most valuable aims of machine learning, since complex structures normally require a large system size in practical applications, while the ordinary DFT calculations for large supercells are extremely challenging. Therefore, we have studied the model in alloy and amorphous materials. Specifically, aluminum-magnesium alloy (108 atoms), amorphous silicon (64 atoms), and water (48 atoms) were tested.

First, a face-centered cubic structure with randomly placed aluminum and magnesium was used for the alloy study. The training convergence towards the number of structures trained on was tested, and 10^5 grid points from each structure were chosen for training, as shown in Fig. 5(a). The MAE approaches $8.0 \times 10^{-4}\,\mathrm{e}/\mathrm{\AA}^3$ with more than five structures of training data. Figures 5(b) and 5(c) indicate that the increase in system size barely influences the predictability of the machine learning model. The MAE in the system with 256 atoms is $8.14 \times 10^{-4}\,\mathrm{e}/\mathrm{\AA}^3$, slightly smaller than $7.66 \times 10^{-4}\,\mathrm{e}/\mathrm{\AA}^3$ from the system with 108 atoms. The detailed charge density is shown in the first row of Fig. 6. The absolute errors are mostly contributed from the center area of magnesium atoms, with a value of around $0.06\,\mathrm{e}/\mathrm{\AA}^3$.

Next, amorphous silicon and water reveal similar results, as shown in Figs. 5(d)–5(i) and Fig. 6. For amorphous silicon, the MAE reached below $0.004\,\mathrm{e}/\mathrm{\AA}^3$ after training with seven structures of data. The max absolute error for the test structure is $0.037\,\mathrm{e}/\mathrm{\AA}^3$, nearing the silicon atoms. Moreover, we have also examined the performance of our model in the silicon melting process, which obtains MAEs of about $0.005\,\mathrm{e}/\mathrm{\AA}^3$. The details are discussed in the SM. In the case of water, the MAE decreases below $0.0023\,\mathrm{e}/\mathrm{\AA}^3$ after training with six structures. The predictions in silicon with 216 atoms and water with 384 atoms are as good as that in smaller systems. However, with the increase in system size, the accuracy for water decreases relatively largely, which may be induced by the rather small training structures. These results show that the accuracy of Deep Charge in amorphous structures is relatively lower than that of crystalline but can still be a good representation. Besides, the larger error around atom centers could be related to the cutoff function we used, which increases the weight of adjacent atoms but probably not be the best description of the electron distributions. A reasonable design of the cutoff function by reference to the near-core charge density should improve the predictability of machine learning models. The extremely high scalability will pave the way for predicting charge density in large-scale simulations.

From all the results above, we can see that the Deep Charge model can effectively capture the radial distribution of the charge near atoms as well as the angular distribution around bond positions, due to two factors. First, the construction of the local environmental matrix $\mathcal{R}$ of each grid point has included the generalized coordinates with angular information for every neighboring atom. Second, the embedding net $\mathcal{G}$ of each grid point works as filters that add weights to each neighboring atom according to the distance and chemical species. From a physical point of view, this can be regarded as the contribution of each atom to the charge-density representation in latent space. Different from the above-mentioned filters such as Gaussian, the parameters of embedding nets are adaptive and can be learned to align with the DFT results via training. Together with the weight of atomic contribution and detailed generalized coordinates, the model obtains its ability to represent the charge density of arbitrary material systems.

IV. CONCLUSION

In summary, we have realized a charge-density representation with machine learning techniques. Our Deep Charge model can accurately represent charge density in semiconductor, metal, and molecular systems. The symmetry-preserving descriptor enhances the training efficiency. Only one single DFT calculation is enough for training extremely high-accuracy models for simple crystal structures. It can also capture the fluctuations of the environment and represent complex systems by training with fewer data. Meanwhile, further improvement in the accuracy and efficiency of this model is still possible. A reasonable design of the cutoff function and loss function may increase the charge-density accuracy near the atoms. Moreover, implementing a concurrent learning strategy can automate the sampling process of data points used for training, leading to more efficient learning and less human intervention.

In addition, reasonable modifications to the model can extend its application. Apart from the descriptor employed in our study, the attention-based model [32] offers the potential to create a universal charge-density representation covering more elements. Furthermore, the model can be adapted to predict the spatial distribution of other fields, such as spin charge density and the local density of states on grids. Achieving this requires suitable modifications to the network architecture

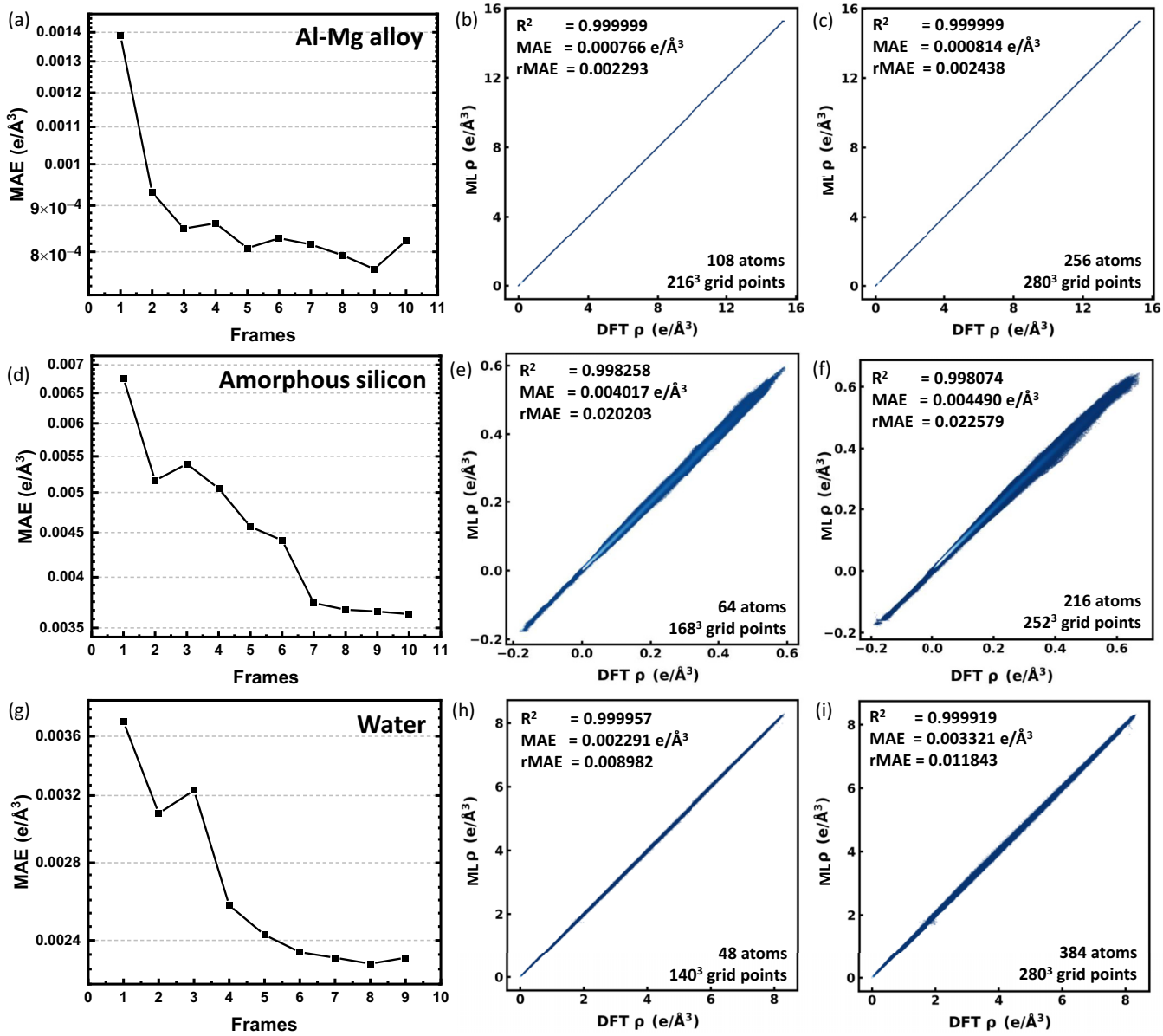

FIG. 5. (a), (d), (g) The convergence relation between the number of structures used for training and the MAE values for the testing set of aluminum-magnesium alloy, amorphous silicon, and water, respectively. (b), (e), (h) The corresponding scatter plots for charge density calculated by DFT and Deep Charge for a whole grid of data from a testing structure. (c), (f), (i) The corresponding parity plots in larger supercells.

and transitioning from scalar output values to two- or multidimensional vectors. The feasibility and implementation will be carried out in future works.

Our model is easy to implement and can provide an approach for fast charge-density prediction avoiding the usage of KS quasiparticle wave functions. Looking forward, with the predicted charge density, many other physical quantities like ground-state energy and force can be derived theoretically based on the orbital-free DFT frameworks [33], which thus assist the electronic structure analysis, stability prediction, elastic behavior, etc. in large-scale condensed matter systems with the first-principle accuracy. Additionally, the predicted charge density can be integrated into physics-informed neural networks to efficiently derive these quantities. This approach further enhances the versatility and efficiency of our model in capturing and analyzing the complex physics of condensed matter systems.

ACKNOWLEDGMENTS

We appreciate the discussion and coding instruction with Mr. Xuejian Qin and Mr. Yifan Shan. We greatly thank Dr. Linfeng Zhang for the information and suggestions provided. This work is partially funded by the Grand Challenge scheme of the Physics Foundation at the University of Sydney. The research undertakes resources provided by National Computational Infrastructure (NCI Australia) and the Pawsey supercomputer center. We acknowledge the financial

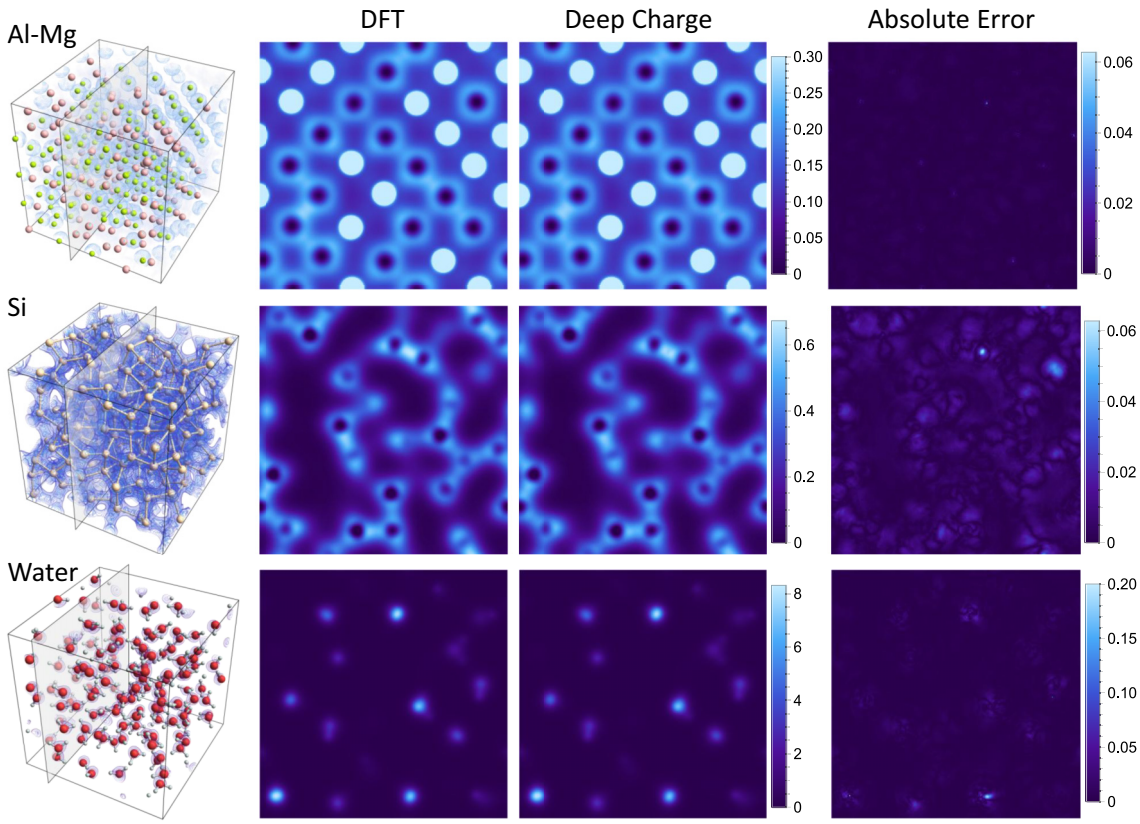

FIG. 6. Charge density in large aluminum-magnesium alloy, amorphous silicon, and water supercells. Notice that the maximum value of charge density for aluminum-magnesium alloy is above 15.3 $e/Å^3$. We only display the charge-density values in the interval of $0 \sim 0.3\, e/Å^3$ to show the details clearly.

support from the National Key R&D Program of China (Grants No. 2022YFA1403000 and No. 2021YFA0718900), and the Australian Research Council (Grants No. DP200100940 and No. DE180100167). We express our gratitude for the academic exchange program provided by AI for Science Institute, Beijing.

[1] R. O. Jones, Density functional theory: Its origins, rise to prominence, and future, Rev. Mod. Phys. **87**, 897 (2015).
[2] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. **140**, A1133 (1965).
[3] G. Kresse and J. Furthmüller, Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).
[4] E. H. Lieb and B. Simon, The Thomas-Fermi theory of atoms, molecules and solids, Adv. Math. **23**, 22 (1977).
[5] E. Prodan and W. Kohn, Nearsightedness of electronic matter, Proc. Natl. Acad. Sci. USA **102**, 11635 (2005).
[6] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, Transferable machine-learning model of the electron density, ACS Cent. Sci. **5**, 57 (2019).
[7] A. Fabrizio, A. Grisafi, B. Meyer, M. Ceriotti, and C. Corminboeuf, Electron density learning of non-covalent systems, Chem. Sci. **10**, 9424 (2019).
[8] A. M. Lewis, A. Grisafi, M. Ceriotti, and M. Rossi, Learning electron densities in the condensed phase, J. Chem. Theory Comput. **17**, 7203 (2021).
[9] A. J. Lee, J. A. Rackers, and W. P. Bricker, Predicting accurate ab initio DNA electron densities with equivariant neural networks, Biophys. J. **121**, 3883 (2022).
[10] K. T. Schütt, M. Gastegger, A. Tkatchenko, K. R. Müller, and R. J. Maurer, Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions, Nat. Commun. **10**, 5024 (2019).
[11] J. A. Rackers, L. Tecot, M. Geiger, and T. E. Smidt, A recipe for cracking the quantum scaling limit with machine learned electron densities, Mach. Learn.: Sci. Technol. **4**, 015027 (2023).
[12] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, Bypassing the Kohn-Sham equations with machine learning, Nat. Commun. **8**, 872 (2017).
[13] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, Solving the electronic structure problem with machine learning, npj Comput. Mater. **5**, 22 (2019).
[14] S. Gong, T. Xie, T. Zhu, S. Wang, E. R. Fadel, Y. Li, and J. C. Grossman, Predicting charge density distribution of materials using a local-environment-based graph convolutional network, Phys. Rev. B **100**, 184103 (2019).

[15] L. Zepeda-Núñez, Y. Chen, J. Zhang, W. Jia, L. Zhang, and L. Lin, Deep Density: Circumventing the Kohn-Sham equations via symmetry preserving neural networks, J. Comput. Phys. **443**, 110523 (2021).

[16] D. Kamal, A. Chandrasekaran, R. Batra, and R. Ramprasad, A charge density prediction model for hydrocarbons using deep neural networks, Mach. Learn.: Sci. Technol. **1**, 025003 (2020).

[17] P. B. Jørgensen and A. Bhowmik, Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids, npj Comput. Mater. **8**, 183 (2022).

[18] B. Focassio, M. Domina, U. Patil, A. Fazzio, and S. Sanvito, Linear Jacobi-Legendre expansion of the charge density for machine learning-accelerated electronic structure calculations, npj Comput. Mater. **9**, 87 (2023).

[19] J. Zeng, D. Zhang, D. Lu, P. Mo, Z. Li, Y. Chen, M. Rynik, L. Huang, Z. Li, S. Shi *et al.*, DeePMD-kit v2: A software package for deep potential models, J. Chem. Phys. **159**, 054801 (2023).

[20] A. G. Eguiluz, Self-consistent static-density-response function of a metal surface in density-functional theory, Phys. Rev. B **31**, 3303 (1985).

[21] T. Rice, The effects of electron-electron interaction on the properties of metals, Ann. Phys. **31**, 100 (1965).

[22] L. Zhang, J. Han, H. Wang, W. Saidi, and R. Car, End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems, Adv. Neural Inf. Process. Syst. **31**, 4441 (2018).

[23] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, Deep sets, in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017), Vol. 30.

[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.108.235159 for additional figures used in our work.

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

[26] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).

[27] J. Sun, A. Ruzsinszky, and J. P. Perdew, Strongly constrained and appropriately normed semilocal density functional, Phys. Rev. Lett. **115**, 036402 (2015).

[28] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen *et al.*, LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, Comput. Phys. Commun. **271**, 108171 (2022).

[29] W. Jiang, Y. Zhang, L. Zhang, and H. Wang, Accurate Deep Potential model for the Al–Cu–Mg alloy in the full concentration space, Chin. Phys. B **30**, 050706 (2021).

[30] C. Zhang, F. Tang, M. Chen, J. Xu, L. Zhang, D. Y. Qiu, J. P. Perdew, M. L. Klein, and X. Wu, Modeling liquid water by climbing up Jacob's Ladder in density functional theory facilitated by using deep neural network potentials, J. Phys. Chem. B **125**, 11444 (2021).

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

[32] H. B. Duo Zhang, F.-Z. Dai, W. Jiang, L. Zhang, and H. Wang, DPA-1: Pretraining of attention-based deep potential model for molecular simulation, arXiv:2208.08236.

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