

Predicting the Electron Density of Charged Systems Using Machine Learning

Sherif Abdulkader Tawfik,* Sunil Gupta, and Svetha Venkatesh



Cite This: *J. Phys. Chem. A* 2025, 129, 2117–2122



Read Online

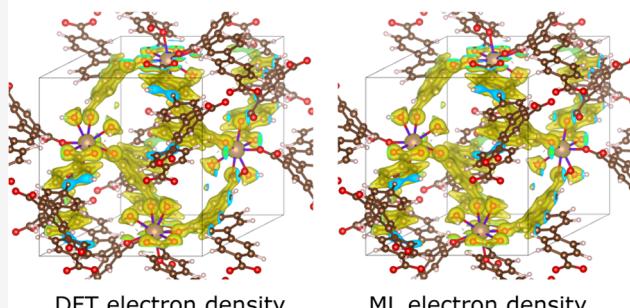
ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: The prediction of the electron density in molecules and crystals is a key pillar in the first-principles computation of their properties. Using machine learning to predict the electron density by using the atomic structure alone can save the computational cost of performing first-principles computations. While various machine learning approaches have been introduced for predicting the electron density, none of them predict the electron density for charged systems. This work extends a recent machine learning charge density model, DeepDFT, by including the charge of the structure as an input parameter into the model. We establish an input charge representation approach that successfully predicts the charged electron densities for several test cases, including charged defective perovskites, LiCoO₂ supercells with multiple Li vacancies, diamond-based defects, metal-organic frameworks, and molecular crystals.

APAYOH MOF with an excess electron: error 2.6%



1. INTRODUCTION

Molecules and crystals are seldom electrically neutral. In crystals, excess or depletion of charge compensates for the presence of defects, which are present in significant percentages, even in the purest samples of materials such as diamond (the negatively charged nitrogen vacancy). Charges on molecular systems drive a number of mechanisms. The opposing charges on the components of an ionic liquid result in their solubility and drive their electrostatic interactions. Proteins are most often charged due to the charged amino acids,^{1,2} and charges in proteins are the origin of the biochemical functions of proteins.¹ While charges on small molecules are generally static, they are far from static on large molecules such as proteins;¹ predicting where an extra positive or negative charge will be for a given structure is a challenge. Likewise, while one can expect the position of an extra positive or negative charge in a defective crystal structure, this becomes far from intuitive when the crystal structure is larger and hosts multiple vacancies.

The standard workhorse method for obtaining the charge distribution for any given system is density functional theory (DFT). The charge distribution is obtained from the electron density function, $\rho(\mathbf{r})$, which is a scalar spatial function of position \mathbf{r} that determines the probability of finding an electron at \mathbf{r} . DFT can routinely calculate $\rho(\mathbf{r})$ for neutral and charged structures, but it struggles with system scalability owing to its $O(N^3)$ complexity, where N is the number of valence electrons in the system. To solve this problem, various machine learning (ML) models have been proposed as surrogates to DFT, where

the model can accurately generate $\rho(\mathbf{r})$ for a given atomic structure.^{3–12} Of particular interest to the present work is DeepDFT,⁵ which enables accurate prediction of the electron density of molecular and crystal systems by embedding two sets of geometric structures that are derived from input electron density: atom–atom distances and atom-probe distances. The “probe points” are imaginary points that are arbitrarily specified within the simulation cell. DeepDFT achieves high prediction accuracy for electron density in neutral systems. However, neither DeepDFT nor any of the other mentioned approaches have considered the prediction of the influence of charge addition or removal on the structure of $\rho(\mathbf{r})$.

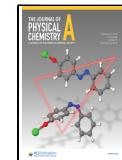
Based on the success of the DeepDFT ML model in predicting $\rho(\mathbf{r})$ for neutral systems, this work modifies DeepDFT to incorporate the effective charge of charged supercells as an input parameter. We introduce two alternative approaches to represent the input charge, and we call our modified code CDeepDFT. We create a data set of 12k charged unit cells based on the materials in the TinyUnitCells data set in the <https://github.com/>

Received: December 19, 2024

Revised: January 29, 2025

Accepted: February 7, 2025

Published: February 13, 2025



[sherifawfikabbas/materialsalchemist](#) repository, and train the CDeepDFT model on this data set, and subsequently test our trained models on several data sets of charged defect supercells: charged defective perovskites, LiCoO₂ supercells with multiple Li vacancies, diamond-based defects, metal–organic frameworks (MOFs), and molecular crystals. The CDeepDFT models achieve high prediction accuracy for inorganic structures but do not perform well with diamond-based structures.

2. COMPUTATIONAL DETAILS

2.1. Incorporating the Charge as an Input. For a given structure with atomic positions \mathbf{r}_i , where $i = 1, \dots, N$, N is the number of atoms, DeepDFT describes the structure by the \mathbf{r}_i vectors, the lattice vectors (if it is a crystal structure), and “imaginary atoms”⁴ or “probe points”,¹⁰ which are simply a discretization of the three-dimensional (3D) grid into grid points. Then, each atom-probe distance is calculated $p_{ij} = |\mathbf{p}_i - \mathbf{r}_j|$ within a cutoff distance r_{cut} , where j is the probe index. The quantities p_{ij} are expanded using the sinc function $\phi(p_{ij}) = \sin(n\pi p_{ij}/r_{\text{cut}})/p_{ij}$, n is between 1 and 20, and are subsequently embedded in the neural network as described in ref 10.

We incorporate the effective supercell charge using two alternative approaches:

1. The charge is directly incorporated into the structure embedding by modifying the sinc function as follows

$$\phi'(p_{ij}) = \sin(2^{-c}n\pi p_{ij}/r_{\text{cut}})/p_{ij} \quad (1)$$

where c is the charge. The presence of excess electrons is represented by augmenting the value of $\phi'(p_{ij})$. The system is negatively charged (that is, has excess electrons) with $c < 0$, and hence, the impact of negative charge is to magnify the magnitude and reduce the period of the sinc function. That coefficient becomes 1 in neutral systems ($c = 0$). The total number of trainable parameters in the present setup is 98,671, which is the same number of parameters in the vanilla DeepDFT model. eq 1 was proposed because the coefficient 2^{-c} is positive definite and magnifies the change in charge because of the exponential dependence. However, a key shortcoming in this method is that, while it has a distinguished impact on the values of the expansion for positive c , it does not have significantly different change on the expansion for negative values of c .

2. The charge is included as an additional input node to the neural network. The total number of trainable parameters in the present setup is 104,874.

Hereafter, we label the modified DeepDFT models Charged-DeepDFT, or CDeepDFT in short.

2.2. Training and Validation Sets. We pick 4900 structures from the TinyUnitCells data set and filter them by removing any structure where the cell height $< 3 \text{ \AA}$. The number of structures remaining is 4140. Then, for each of these structures, we assign an effective supercell charge by randomly picking one of the values $\pm 1, \pm 2, \pm 3$, charging the cell with that value and subsequently optimizing both the lattice and the atomic positions. The total number charged an neutral structures in the training set then becomes 11,980. We display the distribution of the unit cells according to charge in Figure 1. We split the training and validation sets in 95:5. The validation error is quantified in terms of the mean absolute error (MAE) %, defined in ref 10, which is given by

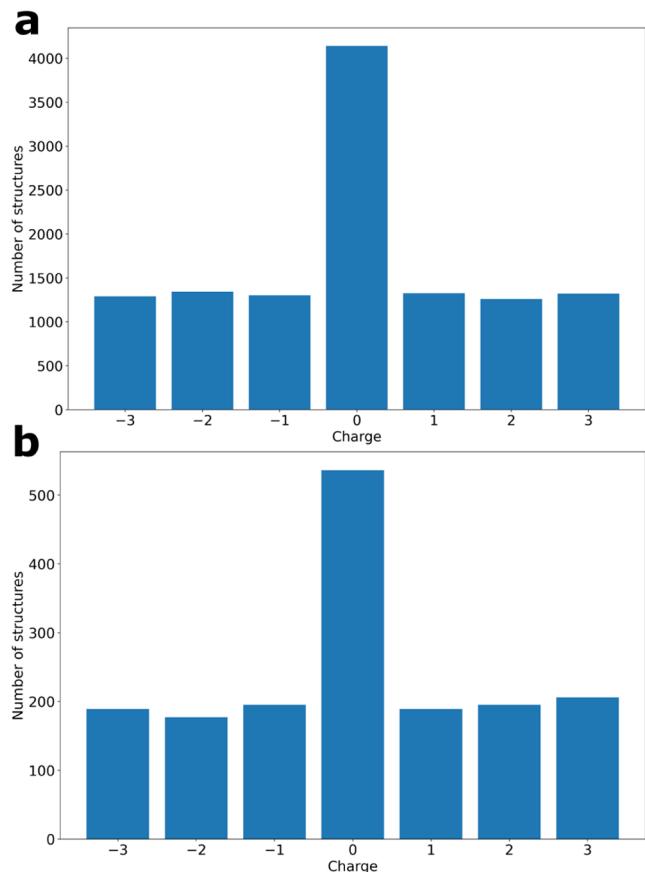


Figure 1. Number of structures with the charge values displayed along the x -axes in the (a) training set and (b) test set.

$$\text{MAE\%} = \frac{\int d\mathbf{r}|\rho_t(\mathbf{r}) - \rho_p(\mathbf{r})|}{\int \rho_t(\mathbf{r})} \times 100 \quad (2)$$

where $\rho_t(\mathbf{r})$ is the target electron density function and $\rho_p(\mathbf{r})$ is the predicted electron density function.

2.3. Test Set. We construct 810 defective structures by building $2 \times 2 \times 1$ supercells from a selection of the structures in the TinyUnitCells data set. Then for each supercell, we create a substitutional or vacancy defect for a randomly chosen atom from the supercell and then optimize both the lattice and the atomic positions. For each defect structure, we assign a charge to the supercell by randomly picking one of the values $\pm 1, \pm 2, \pm 3$, charging the cell with that value and then optimizing the electronic structure without optimizing the atomic or lattice structures. The total number of neutral and charged defects in the test set is 1688. We display the distribution of the defects according to charge in Figure 1.

2.4. DFT Calculations. To generate the charge densities for the training and test sets, we performed DFT calculations using the VASP 5.4.4 code within the generalized gradient approximation (GGA) of Perdew, Burke, and Ernzerhof (PBE).¹³ We set the energy cutoff for the plane wave basis set at 520 eV and the energy tolerance at 10^{-6} eV to ensure the accuracy of the calculations. In the structural energy minimization, the internal coordinates are allowed to relax until all of the forces are less than 0.01 eV/ \AA .

3. RESULTS AND DISCUSSION

We display in **Table 1** the MAE % values computed by using eq 2 for training the CDeepDFT models according to the two

Table 1. Average Mean Absolute Error (MAE) % for the Validation and Test Sets Predictions of the CDeepDFT Model Using the two Charge Representation Methods 1 and 2. Also Shown Is the MAE % for the Validation Set of the Neutral Training Set

method	average validation set MAE %	average test set MAE %
1	1.40	2.83
2	2.09	1.73
uncharged	0.65	

charge representation methods described above. We also display in the table the MAE % for the validation set of the neutral TinyUnitCells training set in order to benchmark the accuracy of DeepDFT in comparison with the prediction accuracies reported in refs 10,14 for neutral systems. The errors are generally higher than the MAE % calculated in for the uncharged materials in **Table 1** (0.65%) as well as the errors reported in ref 10 for the LiCoO₂ supercells and the errors reported in ref 14 for the Materials Project (MP) data set. Such a difference is due to several factors: the inclusion of the charge as an input parameter; the method of representing the charge into the input; and the inadequacy of the number of charged cells for each structure within the training set (for each neutral structure, 2 charged structures are included in the training set). According to the table, the two charge representation methods have different prediction performances: method 1 performs better in predicting the validation set, whereas method 2 performs better in predicting the test set.

Next, we examine the ability of the trained CDeepDFT models in predicting the $\rho(\mathbf{r})$ of charged defective supercells with larger sizes than those in the training and test sets in **Table 1**. Similarly to the above test set, the electron densities of the following set of structures have been computed using single-point SCF. We examine the impact of the following four scenarios on the accuracy of CDeepDFT:

1. Charged single-site defect: we examine the charged diamond defects that were examined in our work ref 15, namely: nitrogen vacancy (N-V, charges),¹⁶ oxygen vacancy (O-V),¹⁷ phosphorus vacancy (P-V), germanium vacancy (Ge-V),¹⁸ tin vacancy (Sn-V),¹⁹ silicon vacancy (Si-V),²⁰ sulfur vacancy (S-V),²¹ substitutional boron (B_C),²² substitutional nitrogen (N_C),^{22,23} substitutional oxygen (O_C),¹⁷ interstitial hydrogen (H_i),²⁴ and interstitial carbon (C_i).²⁵ For each of these defects, we assigned an input charge by randomly choosing from the charges $\pm 1, \pm 2, \pm 3$. The size of the diamond supercell is $2 \times 2 \times 2$. For these 11 defects, we generated $3 \times 11 = 33$ charged supercells.
2. Defect dilution: we examine the effect of diluting the C vacancy defect on the accuracy of the predictions of CDeepDFT for this defect. The supercells examined here are $3 \times 2 \times 2$, $3 \times 3 \times 2$, and $3 \times 3 \times 3$, and we created $3 \times 3 = 9$ charged supercells.
3. Charged multisite diamond defects: we examine the situation where the defect supercell with multiple defect sites is charged with a randomly picked charge from the

set $\{\pm 1, \pm 2, \pm 3\}$. The size of the diamond supercell is $3 \times 3 \times 3$, and the defect structures were generated by the combination of two groups of defects: an antisite N, O, S, Si, P, Sn vacancy and a substitutional B, O, and N defect. For 6×3 defective supercells, we generated $3 \times 6 \times 3 = 54$ charged supercells.

4. Similar to the tests conducted in refs 10,14 where the accuracy of density prediction models is assessed in predicting the charge density of the lithium cathode material LiCoO₂, we examine a $3 \times 3 \times 2$ supercell of LiCoO₂ and remove 3 Li atoms from it. Using the `bsym` library, we generate 54 structures, which represent all the possible permutations of the three Li vacancies, and then optimize the neutral structures using DFT. We then apply a charge by randomly picking a charge from the set $\{-1, -2, -3, -4\}$, creating 2 charged supercells for each structure.
5. Defective perovskites: we randomly pick 50 perovskite unit cells and create a vacancy defect in the $2 \times 2 \times 2$ supercell. Then, the neutral defect is optimized with DFT, and then we perform single-point SCF calculations on the charged supercells by charging them with the four charges: $\pm 1, \pm 2$.
6. Charged MOFs: we selected 50 MOFs from the QMOFs database²⁶ and randomly assigned charges from the set $\{\pm 1, \pm 2\}$. The number of samples in the data set is 120.
7. Charged molecular crystals: we selected three organic molecular crystals from the MP database which have known experimental structures: guanidine (mp-1196017, ICSD: 237237), ammonium dicyanamide (mp-30094, ICSD: 281051), and 1,2-bis(*N,N*-Dimethylamino)-1,2-bis((*p*-tolyl)imino)ethane (mp-605048, ICSD: 163106). For each of these, we applied each of the 4 charges in the set $\{\pm 1, \pm 2\}$.

The results of the tests are displayed in **Table 2**. The table demonstrates the general advantages of using Method 2 for

Table 2. Average Mean Absolute Error (MAE) % for the Validation and Test Sets

charged defects	method 1 (MAE %)	method 2 (MAE %)
LiCO ₂ multiple Li vacancies	4.17	2.73
perovskites with single vacancies	2.20	1.60
single-site diamond defects	149.66	6.67
effect of diluting single-site diamond defects	237.24	6.46
multisite diamond defects	116.25	6.75
MOFs	~50k	3.10
organic crystals	8.90	2.95

density representation. The MAEs for the predictions of the LiCO₂ multiple Li vacancies and the perovskite vacancies are much smaller than those obtained for the prediction of diamond-based defects, particularly the Method 1 MAEs. For the case of the defective LiCO₂ supercells, the MAE (2.73%) is higher than the error of the prediction of Method 2 for the test set in **Table 1**, while for the perovskites, the MAE (1.6%) is slightly smaller than the MAE of the test set in **Table 1**. The errors of the diamond defects are close (6.67, 6.46, and 6.75%) and are much higher than those obtained for the LiCO₂ multiple Li vacancies and perovskites, and this is due to the

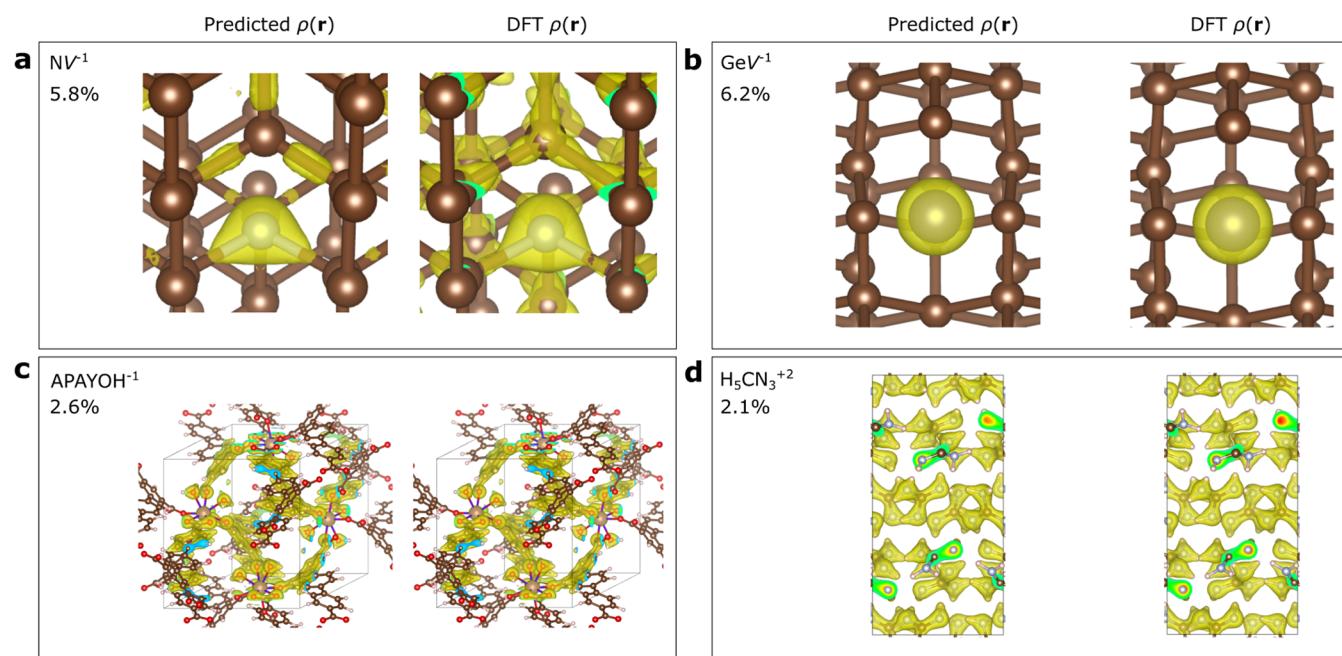


Figure 2. Comparison between the predicted and target electron densities for the (a) N_V^{-1} defect (isosurface at $0.25e$), the (b) Ge_V^{-1} defect (isosurface at $0.29e$), the (c) APAYOH $^{-1}$ MOF (isosurface at $0.2e$) and the (d) guanidine ($\text{H}_5\text{CN}_3^{+2}$) crystal with a supercell charge of +2 (isosurface at $0.2e$). The predictions are computed using Method 2.

low representation of carbon-based covalent systems in the training set: 8 out of the 199 C-containing materials in the training set are pure carbon polymorphs, and 7 are p-block covalent carbon compounds (1 CS (mp-1147643), 1 SiC (mp-11714), 1 $\text{C}_2\text{O}_2\text{H}$ (mp-675784), 1 CN (mp-1009818), and 3 CB_2N (mp-1008525, mp-1008526, mp-1008527)).

Even though the errors are higher in diamond defects on average using Method 2, the critical features of the defect density are well reproduced by the representation method, as shown in Figure 2: the charge accumulation at the N site in the N_V^{-1} defect and around the C–C bonds surrounding the defect (Figure 2(a)), and the charge accumulation at the Ge site in the Ge_V^{-1} defect and the lack of charge accumulation around the defect site (Figure 2(b)). Even though the MAEs of these two defects are very low using Method 1 (3.2% for Ge_V^{-1} and 2.9% for N_V^{-1}), the average MAE using Method 1 is high owing to the large errors in some of the defects, such as the N substitutional defect with charge +2 (N_C^{+2}), where the MAE is excessively high ($>2\%$). We display a comparison between the predicted density of the N_C^{+2} defect using the two methods in Figure 3. Method 2 is able to distribute the charge density around each C–C and C–N bond almost evenly as is the case in the target electron density plot, whereas Method 1 overlocalizes the charge density far away from the location of the defect. The possible avenues to reduce the errors of Method 1 would include: better representation of the charge within the embedding function in eq 1 and representing more C-containing materials in the training set.

For the MOF test set, Method 1 fails in predicting the electron densities of the structures, while the average error of Method 2 is 3.1%, which is half of the errors obtained in carbon-based systems. We display an example prediction for the charge density of one of the MOFs in the test set: the APAYOH structure with one excess electron charge (APAYOH $^{-1}$). Method 2 is nearly as accurate in predicting

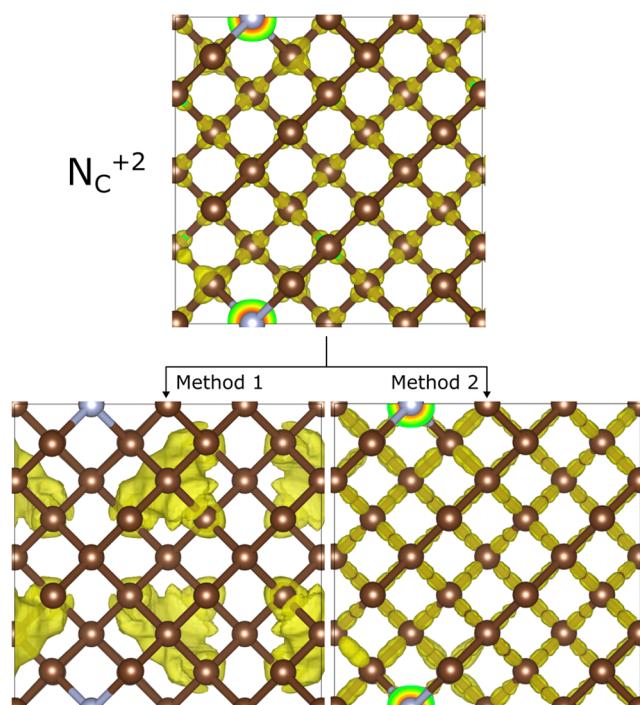


Figure 3. Comparison between the predicted and target electron densities N substitutional defect N_C^{+2} using Method 1 and Method 2.

the charged organic molecular crystals, with an average error of 2.95%. For this set of materials, Method 1 achieves an error of 8.9%, which is nearly 3 times the error obtained using Method 2. This analysis shows that Method 1, even though it can be useful for some classes of materials, is unreliable in general. As was pointed out earlier, the method is insensitive to variations in the positive charge and maps negative charges exponentially. This, however, is not a statement against representation

methods that are similar to Method 1, where the excess charge is embedded into the DeepDFT or any other electron density ML model by direct mathematical manipulation within the edge expansion function. A more careful choice of the edge expansion function would be required to improve the accuracy of such methods.

4. CONCLUSIONS

To summarize, we presented a proof-of-concept for the accurate inclusion of the charge of a structure to predict the electron charge distribution of charged structures. We proposed CDeepDFT, which is based on the DeepDFT method, where we included the charge density as a machine learning input by using two alternative methods: Method 1, where charge is included as a factor in the embedding function, and Method 2, where the charge is included as an additional input into the neural network. We found that, generally, Method 2 surpassed the accuracy of Method 1 for a range of test cases: multiple Li vacancies in a LiCO₂ supercell, perovskites with single vacancies, single-site diamond defects, diluted single defects in diamond and multisite diamond defects. Carbon-based systems generally had higher prediction errors, owing to the low representation of carbon in the training set. The CDeepDFT is the first proposal for the prediction of the impact of charge on the electron charge distribution and can be applied for predicting the charge density in charged systems such as proteins, solvated electrons, and charged defects in solid structures. A carefully generated polymer data set can be used to train CDeepDFT on predicting polaronic states in polymer systems.

■ AUTHOR INFORMATION

Corresponding Author

Sherif Abdulkader Tawfik – *Applied Artificial Intelligence Institute, Deakin University, Geelong, Victoria 3216, Australia; ARC Centre of Excellence in Exciton Science, Deakin University, Geelong, Victoria 3216, Australia;*  [0000-0003-3592-1419](https://orcid.org/0000-0003-3592-1419); Email: s.abbas@deakin.edu.au

Authors

Sunil Gupta – *Applied Artificial Intelligence Institute, Deakin University, Geelong, Victoria 3216, Australia*
Svetna Venkatesh – *Applied Artificial Intelligence Institute, Deakin University, Geelong, Victoria 3216, Australia*

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpca.4c08583>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The density functional theory calculations in this work were supported by computational resources provided by the Australian Government through NCI and Pawsey under the National Computational Merit Allocation Scheme.

■ REFERENCES

- (1) Gitlin, I.; Carbeck, J. D.; Whitesides, G. M. Why Are Proteins Charged? Networks of Charge–Charge Interactions in Proteins Measured by Charge Ladders and Capillary Electrophoresis. *Angew. Chem., Int. Ed.* **2006**, *45*, 3022–3060.
- (2) Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118*, 1691–1741.
- (3) Sunshine, E. M.; Shuaibi, M.; Ulissi, Z. W.; Kitchin, J. R. Chemical Properties from Graph Neural Network-Predicted Electron Densities. *J. Phys. Chem. C* **2023**, *127*, 23459–23466.
- (4) Gong, S.; Xie, T.; Zhu, T.; Wang, S.; Fadel, E. R.; Li, Y.; Grossman, J. C. Predicting charge density distribution of materials using a local-environment-based graph convolutional network. *Phys. Rev. B* **2019**, *100*, No. 184103.
- (5) Jørgensen, P. B.; Bhowmik, A. DeepDFT: Neural Message Passing Network for Accurate Charge Density Prediction 2020 <http://arxiv.org/abs/2011.03346>.
- (6) Willatt, M. J.; Musil, F.; Ceriotti, M. Atom-density representations for machine learning. *J. Chem. Phys.* **2019**, *150*, No. 154110.
- (7) Tsubaki, M.; Mizoguchi, T. Quantum Deep Field: Data-Driven Wave Function, Electron Density Generation, and Atomization Energy Prediction and Extrapolation with Machine Learning. *Phys. Rev. Lett.* **2020**, *125*, No. 206401.
- (8) Lewis, A. M.; Grisafi, A.; Ceriotti, M.; Rossi, M. Learning Electron Densities in the Condensed Phase. *J. Chem. Theory Comput.* **2021**, *17*, 7203–7214.
- (9) Lv, T.; Zhong, Z.; Liang, Y.; Li, F.; Huang, J.; Zheng, R. Deep Charge: Deep learning model of electron density from a one-shot density functional theory calculation. *Phys. Rev. B* **2023**, *108*, No. 235159.
- (10) Jørgensen, P. B.; Bhowmik, A. Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. *npj Comput. Mater.* **2022**, *8*, No. 183.
- (11) Fabrizio, A.; Grisafi, A.; Meyer, B.; Ceriotti, M.; Corminboeuf, C. Electron density learning of non-covalent systems. *Chem. Sci.* **2019**, *10*, 9424–9432.
- (12) Grisafi, A.; Fabrizio, A.; Meyer, B.; Wilkins, D. M.; Corminboeuf, C.; Ceriotti, M. Transferable Machine-Learning Model of the Electron Density. *ACS Cent. Sci.* **2019**, *5*, 57–64.
- (13) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, No. 3865.
- (14) Koker, T.; Quigley, K.; Taw, E.; Tibbetts, K.; Li, L. Higher-order equivariant neural networks for charge density prediction in materials. *npj Comput. Mater.* **2024**, *10*, No. 161.
- (15) Tawfik, S. A.; Nguyen, T. M.; Russo, S. P.; Tran, T.; Gupta, S.; Venkatesh, S. Embedding material graphs using the electron-ion potential: application to material fracture. *Digital Discovery* **2024**, *3*, 2618–2627.
- (16) Rondin, L.; Tetienne, J.-P.; Hingant, T.; Roch, J.-F.; Maletinsky, P.; Jacques, V. Magnetometry with nitrogen-vacancy defects in diamond. *Rep. Prog. Phys.* **2014**, *77*, No. 056503.
- (17) Thiering, G.; Gali, A. Characterization of oxygen defects in diamond by means of density functional theory calculations. *Phys. Rev. B* **2016**, *94*, No. 125202.
- (18) Iwasaki, T.; Ishibashi, F.; Miyamoto, Y.; et al. Germanium-Vacancy Single Color Centers in Diamond. *Sci. Rep.* **2015**, *5*, No. 12882.
- (19) Iwasaki, T.; Miyamoto, Y.; Taniguchi, T.; Siyushev, P.; Metsch, M. H.; Jelezko, F.; Hatano, M. Tin-Vacancy Quantum Emitters in Diamond. *Phys. Rev. Lett.* **2017**, *119*, No. 253601.
- (20) Flatae, A. M.; Lagomarsino, S.; Sledz, F.; Soltani, N.; Nicley, S. S.; Haenen, K.; Rechenberg, R.; Becker, M. F.; Sciortino, S.; Gelli, N.; et al. Silicon-vacancy color centers in phosphorus-doped diamond. *Diamond Relat. Mater.* **2020**, *105*, No. 107797.
- (21) Baker, J. M.; Van Wyk, J. A.; Goss, J. P.; Briddon, P. R. Electron paramagnetic resonance of sulfur at a split-vacancy site in diamond. *Phys. Rev. B* **2008**, *78*, No. 235203.
- (22) Mainwood, A. Substitutional impurities in diamond. *J. Phys. C: Solid State Phys.* **1979**, *12*, 2543–2549.
- (23) Kalish, R. The search for donors in diamond. *Diamond Relat. Mater.* **2001**, *10*, 1749–1755.

- (24) Goss, J. P.; Jones, R.; Heggie, M. I.; Ewels, C. P.; Briddon, P. R.; Öberg, S. Theory of hydrogen in diamond. *Phys. Rev. B* **2002**, *65*, No. 115207.
- (25) Weigel, C.; Peak, D.; Corbett, J. W.; Watkins, G. D.; Messmer, R. P. Carbon Interstitial in the Diamond Lattice. *Phys. Rev. B* **1973**, *8*, No. 2906.
- (26) Rosen, A. S.; Iyer, S. M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J. M.; Snurr, R. Q. Machine learning the quantum-chemical properties of metal-organic frameworks for accelerated materials discovery. *Matter* **2021**, *4*, 1578–1597.

CAS BIOFINDER DISCOVERY PLATFORM™

ELIMINATE DATA SILOS. FIND WHAT YOU NEED, WHEN YOU NEED IT.

A single platform for relevant, high-quality biological and toxicology research

Streamline your R&D

CAS A division of the American Chemical Society