An Efficient Deep Learning Scheme to Predict the Electronic Structure of Materials and Molecules: The Example of Graphene-Derived Allotropes

Beatriz G. del Rio, Christopher Kuenneth, Huan Doan Tran, and Rampi Ramprasad*

School of Materials Science and Engineering, Georgia Institute of Technology, 771 Ferst Drive NW, Atlanta, GA 30332, United States

E-mail: rampi.ramprasad@mse.gatech.edu

Abstract

Computations based on density functional theory (DFT) are transforming various aspects of materials research and discovery. However, the effort required to solve the central equation of DFT, namely the Kohn-Sham equation, remains a major obstacle for studying large systems with hundreds of atoms in a practical amount of time with routine computational resources. Here, we propose a deep learning architecture that systematically learns the input-output behavior of the Kohn-Sham equation and predicts the electronic density of states, a primary output of DFT calculations, with unprecedented speed and chemical accuracy. The algorithm also adapts and progressively improves in predictive power and versatility as it is exposed to new diverse atomic configurations. We demonstrate this capability for a diverse set of carbon allotropes spanning a large configurational and phase space. The electronic density of states,

along with the electronic charge density, may be used downstream to predict a variety of materials properties, bypassing the Kohn-Sham equation, leading to an ultrafast and high-fidelity DFT emulator.

Introduction

Density functional theory (DFT)[1,2] has become an invaluable computational workhorse for materials development and design. It has impacted a variety of fields ranging from energy storage,[3–5] catalysis,[6,7] fuel production and chemical transformations,[8,9] design of advanced electronic and functional materials,[10,11] and the understanding of materials behavior under a variety of extreme conditions,[12,13] to just name a few. DFT addresses the many-electron many-nuclear problem of quantum mechanics through a series of approximations and leaps of imagination and ingenuity, and ultimately involves solving the effective one-electron Kohn-Sham equation.[2] For a given configuration of atoms, the solutions of the Kohn-Sham equation include the one-electron wavefunctions (or the electronic charge density), one-electron energy spectrum (or the electronic density of states), atomic forces, potential energy, as well as a variety of application-relevant equilibrium materials properties.

Despite its versatility and reach, DFT remains a laborious computational enterprise. It requires high-performance computing hardware, robust and specialized software, and fairly in-depth knowledge and expertise to execute the calculations in a credible manner. Even with the availability of such resources, modern DFT ecosystems only allow the practical or routine treatment of systems involving not more than a few hundreds of atoms per repeating unit cell.

In order to accelerate the speed with which one may reliably predict application-relevant properties of new materials, the community is beginning to focus attention on the creation of "surrogate" models that can be much faster than a fresh and direct DFT calculation, but mimics it in the accuracy. Such surrogate models are trained, using machine learning algorithms, on a set of reference data produced by prior DFT calculations. The last decade

has seen several successful examples of such predictive machine learning models applied to a variety of materials properties and application spaces.[14–25] In these efforts, a mapping is established between atomic configurations and appropriate quantities of interest, such as atomic forces, potential energies and a variety of materials or molecular properties of interest.

The primary bottleneck in DFT-based workflows is the computation of the electronic density of states (DOS) and charge density. Once computed, the DOS and charge density, owing to their fundamental nature, may be used to determine the above listed other quantities of interest at negligible cost. Thus, creation of a capability that can significantly speed up the prediction of DOS and charge density will impart unprecedented efficiency to the overall DFT workflow, and can lead to an ultrafast DFT emulator that can produce DFT-like output with a high degree of verisimilitude. This manner of solving the electronic structure problem will be a radical departure from attempting to directly solve the Kohn-Sham equation.

In this contribution, we mainly focus on the creation of an efficient deep learning capability for the instantaneous prediction of the electronic DOS for a given configuration of atoms. A neural network (NN) architecture is trained on a database of prior reference DFT computations, and learns the relationship between the atomic configuration and the electronic DOS. Specifically, the NN is designed to take as input the environment around an atom, i.e., the distribution of its neighbouring atoms, producing as output the corresponding atomic DOS spectrum. The NN is trained such that the sum of the thus-predicted atomic DOS of all atoms in the system is required to be equal to the correct total electronic DOS calculated by DFT. As we will show here, this deep learning capability proves to be several orders of magnitude faster than the parent DFT calculation. Moreover, the present development is also a significant advance, both in terms of conceptual aspects and in terms of efficiency, compared to a recipe we proposed recently.[26] While this past work also utilized deep neural networks to predict the DOS (and the electronic charge density), the training data consisted of the projected DOS at each spatial grid point. As the typical ratio of the number of grid points to the number of atoms in any system is *about a million*, the method of Ref. 26 leads

to an enormous memory requirement (for the storage of the training data), and an enormous amount of training and prediction time, hindering its use with large datasets.

(a)

Dataset 1

Structures	# atoms
Graphene	128
Graphite	108
C20	20
C40	40
C60	60
C(6,4)	152
C(9,9)	144
C(8,0)	128

Dataset 2

Structures		# atoms
Defects	SW	128
	SV	127-124
	DV	126-120
2000 K	Graphene	128
	C60	60
	C(6,4)	152
	C(9,9)	144
	C(8,0)	128

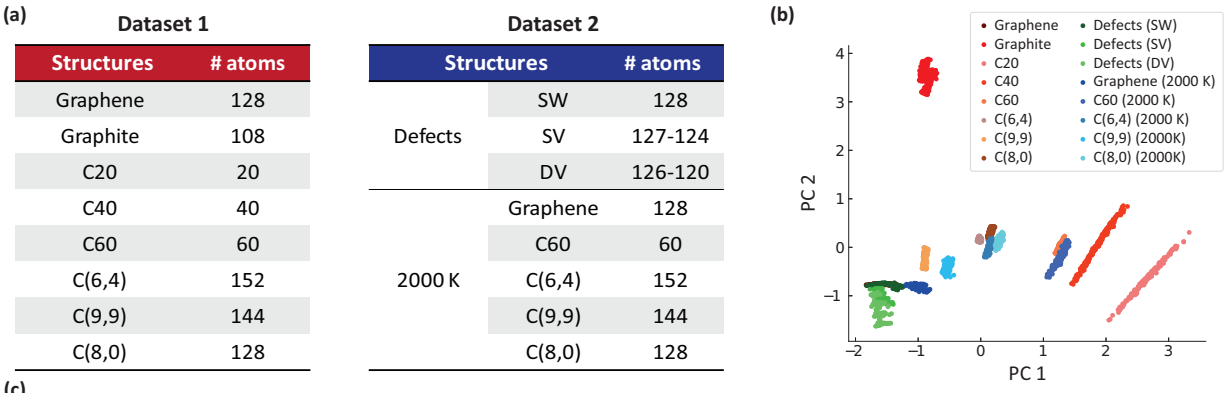

Figure 1: (a) Datasets used to train the models in this work. The number of atoms refers to those in the repeating unit cell of each structure. (b) Main two principal components (PC) of the mean atomic fingerprints of each configuration. (c) Snapshots of carbon allotropes included in dataset 1 and in dataset 2, with red-colored atoms to identify the location of the defects.

As a demonstration of the present development, we train our NN to predict the total electronic DOS of a variety of graphene-derived allotropes, including carbon nanotubes of various types, fullerene molecules, as well as graphene and graphite. Specifically, as listed in the tables of Figure 1(a), we created two different datasets, for which accurate reference DFT calculations were done. We first trained the model using dataset 1, comprised of graphene, graphite, C20, C40 and C60 fullerene molecules, and C(6,4), C(9,9) and C(8,0) single-walled carbon nanotubes (SWCNT) with different chirality. To provide the NN with sufficient examples of configurational diversity within the space of the above list of structures, 200 random snapshots of each of these structures from DFT-based molecular dynamics (MD) runs at 300 K and 600 K were procured. From this set, 80 % of the configurations were used for training and 20 % for validation. An additional separate test set of 20 configurations of each structure was created to select the best performing model after cross-validation. Figure

1(c) (left) shows some representative structures contained in dataset 1.

In order to test the generality of the model built out of dataset 1, and also to unambiguosly demonstrate the ability to systematically improve the model through exposure to newer environments, dataset 2 was considered. Dataset 2 is comprised of defects in graphene and highly disordered structures. The considered graphene defects are Stone-Wales (SW), single-vacancies (SV), and double-vacancies (DV). A representative set of such defected structures are shown in Figure 1(c) (right). The highly disordered structures were procured from DFT-MD runs at 2000 K. The train and test sets of dataset 2 were composed of a total of 830 and 83 configurations, respectively. Overall, the structural and topological diversity of the cases included in the training data considered here is enormous relative to past studies. We choose the vacuum energy as the global energy reference and the DOS of every atomic configuration was aligned with respect to it. The DFT DOS curve is partitioned into 310 windows of 0.1 eV, from -30 eV to 1 eV.

To describe the atomic environment surrounding each atom in a machine-readable form, we used the same set of permutation, translation, and rotation invariant fingerprints introduced in our previous NN DOS protocol,[26] but centered at each atom instead of at grid points. The fingerprints consist of a hierarchy of scalar, vector, and tensor expressions which capture the radial (scalar) and angular (vector and tensor) features of the surrounding atomic environment. The fingerprints are based on a predefined set of Gaussian functions with varying widths centered at every atom. Figure 1(b) shows the variation of the two principal components (PC) of the fingerprint features for each type of the aforementioned structures, spanning a large region of configurational space.

The atomic fingerprint vectors are provided as input layer for the NN, resulting in a DOS per atom as the NN output. Addition of all the atomic DOS for a given configuration results in the predicted total DOS. To ensure an accurate prediction of the Fermi level, the cumulative sum is concatenated to the predicted total DOS. Owing to the variability in size of the structures in the datasets, the prediction is normalized by the number of atoms in

the configuration. This normalization ensures an equal contribution to the error metric from each structure. Figure 2 provides a schematic view of the entire protocol.

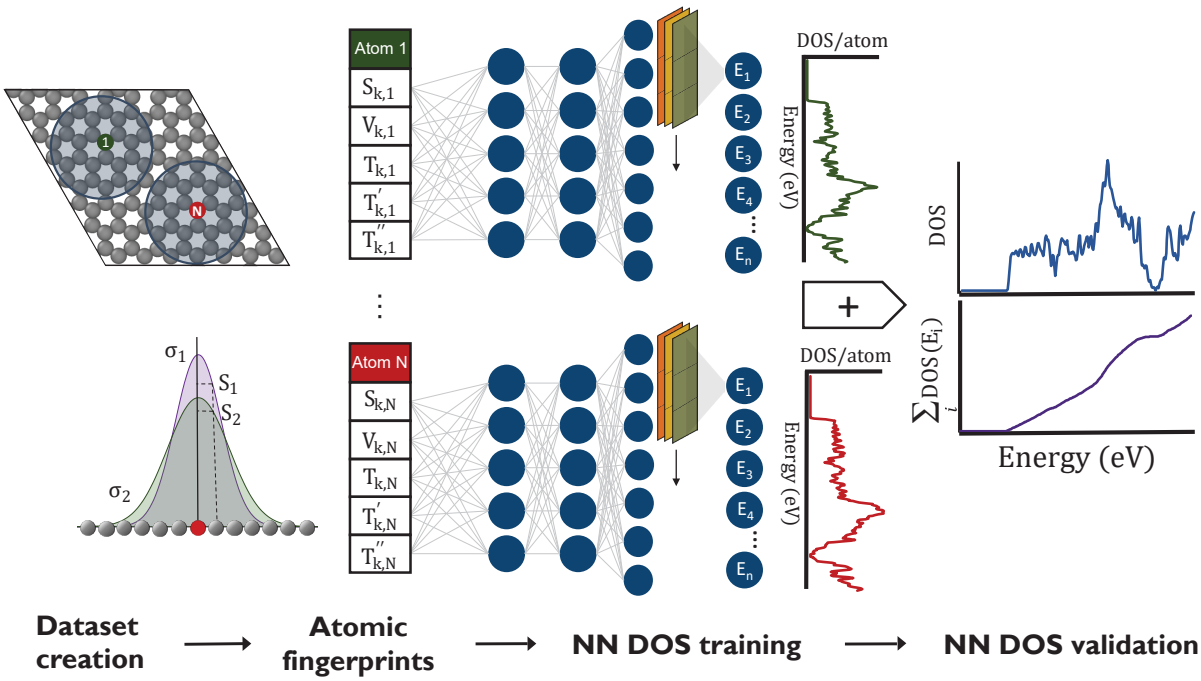

Figure 2: Overall scheme and workflow designed to compute the DOS of a given atomic structure using deep NN with atom centered fingerprints. After the dataset is created, each atom in an atomic structure is replaced by a fingerprint vector representing the surrounding environment. These fingerprints are provided as input layers to the NN, resulting in a DOS per atom as the output layer. All atomic DOS from the same atomic configuration are added to obtain the total DOS. Once normalized, the DOS is concatenated to its cumulative sum and validated against the DFT reference.

As we will demonstrate below, the NN DOS model predicts the electronic structure of sp^2-type carbon allotropes (metallic or semiconducting, and with or without a variety of defects and significant disorder) with unprecedented accuracy and speed (relative to conventional DFT computations). Owing to the flexibility of training afforded by the NN architecture, model prediction performance can be systematically and continuously improved via persistent exposure to newer varieties of configurational diversity. Further, the predicted DOS allows for a precise evaluation of the contribution of the occupied energy levels to the total energy of the system; a necessary step to achieving a machine learned DFT emulator.

Methods

DFT Details

All the reference data calculations were performed using DFT-MD simulations using the Vienna Ab Initio Simulation Package (VASP).[27,28] The exchange-correlation functional was modelled using the Perdew-Burke-Ernzerhof approximation[29] and the ion-electron interaction was modelled using projector-augmented wave (PAW) potentials.[30] We employed a Monkhorst-Pack grid[31] with a density of 0.03 Å^{-1} to sample the Brillouin zone. A plane wave basis set with kinetic energy cutoff of 800 eV was used. The chosen kinetic energy cutoff and k-point sampling converged the total energy to less than 1 meV per atom. Grimme's D2 vdW correction was included.[32] A Gaussian smearing of 0.2 eV was used. The MD simulations were performed in the NVT ensemble, with a time step of 1 fs. All structures were thermalized for 500 time steps at the desired temperature (300 K, 600 K, and 2000 K) and the snapshots were taken from the subsequent thermalized simulations spanning 2 ps.

Fingerprint Details

The scalar fingerprint for a given atom, i, is expressed as the sum over the number of Gaussian functions (k) of width σ_k,

$$S_k = c_k \sum_{j=1}^{N} \exp\left(\frac{-R_{ij}^2}{2\sigma_k^2}\right) f_c(R_{ij}) \tag{1}$$

where c_k is the normalization constant defined as $\left(\frac{1}{\sqrt{2\pi}\sigma_k}\right)^3$, R_{ij} the distance between atom j and the center atom i, and $f_c(R_{ij})$ a cutoff function defined as $0.5\left[\cos\left(\frac{\pi R_{ij}}{d_c}\right)+1\right]$ for $R_{ij} \leq d_c$, and equal to 0 for $R_{ij} > d_c$. In this work, we employed 18 different Gaussian widths, on a logarithmic scale (base 10) from 0.25 to 6.0 Å, with a cutoff distance of $d_c = 7$Å.

The vector and tensor components are defined as

$$V_k^{\alpha} = c_k \sum_{j=1}^{N} \frac{r_{ij}^{\alpha}}{R_{ij}} \exp\left(\frac{-R_{ij}^2}{2\sigma_k^2}\right) f_c(R_{ij}) \tag{2}$$

$$T_k^{\alpha\beta} = c_k \sum_{j=1}^{N} \frac{r_{ij}^{\alpha} r_{ij}^{\beta}}{R_{ij}^2} \exp\left(\frac{-R_{ij}^2}{2\sigma_k^2}\right) f_c(R_{ij}) \tag{3}$$

where α and β represent the x, y or z components of the radial vector between atoms i and j. While S_k is rotational invariant, V_k^{α} and $T_k^{\alpha\beta}$ are variant, but can be combined into four rotational invariant expressions,

$$V_k = \sqrt{(V_k^x)^2 + (V_k^y)^2 + (V_k^z)^2} \tag{4}$$

$$T_k = T_k^{xx} + T_k^{yy} + T_k^{zz} \tag{5}$$

$$T_k' = T_k^{xx} T_k^{yy} + T_k^{yy} T_k^{zz} + T_k^{xx} T_k^{zz} - (T_k^{xy})^2 - (T_k^{yz})^2 - (T_k^{xz})^2 \tag{6}$$

$$T_k'' = \det\left(T_k^{\alpha\beta}\right) \tag{7}$$

Therefore, for each width there are five features. We employed 18 different widths, providing a feature vector for each atom with 90 components.

Neural Network Architecture and Performance

The number of hidden dense layers and nodes per layer were optimized to five with the first four with 300 neurons each and the last one with 312. After the last dense layer, there is a 1D convolution layer with three filters of size 3, resulting in 1D vectors of size 310, equal to the number of energy windows used to discretize the reference DOS curve. Finally, the

average value of the three 1D vectors for each bin is selected as the output for the value of each energy window in the DOS. The final 1D Convolution layer is included to introduce the correlation between adjacent points and ensure a smooth shape in the predicted atomic DOS. Details on the performance of the NN with different number of dense layers can be found on the Supporting Information (SI).

The activation function used for each dense layer as well as for the final 1D convolution layer is rectified linear unit (ReLU). To prevent over-fitting, a $L2$ regularizer with 0.1 was used in each hidden layer. A dropout rate of 0.1 was also included for the hidden layers, meaning that for every pass, each node in the layer has a 10% probability of not being active. The benefit of including the dropout is twofold. First, during training, it acts as a regularization technique to reduce over fitting and second, during prediction, it allows for an evaluation of the uncertainty in such prediction. The latter technique of activating the dropout during prediction is known as Monte Carlo dropout.[33,34]

We used Keras[35] with Tensorflow backend to implement the NN DOS model. A mini-batch training of 30 with random sampling was employed along with Adam optimizer with a learning rate of 0.0001 and momentum vectors $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The RMSE was employed as the objective function.

We compare the computational performance of DFT and our NN DOS model, for a given graphene configuration of 128 atoms. DFT employs 3615 s to solve the Kohn-Sham equation and calculate the DOS on a Broadwell node with 28 cores and 128 GB of RAM. On the other hand, the NN requires only 5 s on a Tesla P100-PCIe GPU with 16 GB of RAM. Out of that time, the fingerprinting process requires 3 s, while the DOS prediction only takes 2 s. Albeit the comparison limitations due to the different architectures used, the achieved speed up is several orders of magnitude. Furthermore, it is worth noting the quadratic scaling (at best) of modern DFT codes with system size as opposed to the linear scaling of NN.

As a final note, we performed a baseline comparison with the model in Ref. 26 by training and testing our model on the same aluminum configurations from the study. The accuracy

of the prediction is similar for both models: the R^2 reported in Ref. 26 is 0.9992, whereas in with our model is 0.9996. The major improvement is on the training time: Ref. 26 reported 5-6 hours (also on a GPU), whereas our NN DOS model only required 10 minutes. Additionally, the prediction time is further reduced with our protocol due to the atomic fingerprint (3 s) as opposed to the grid-point fingerprint in Ref. 26, requiring 20 s.

Results and Discussion

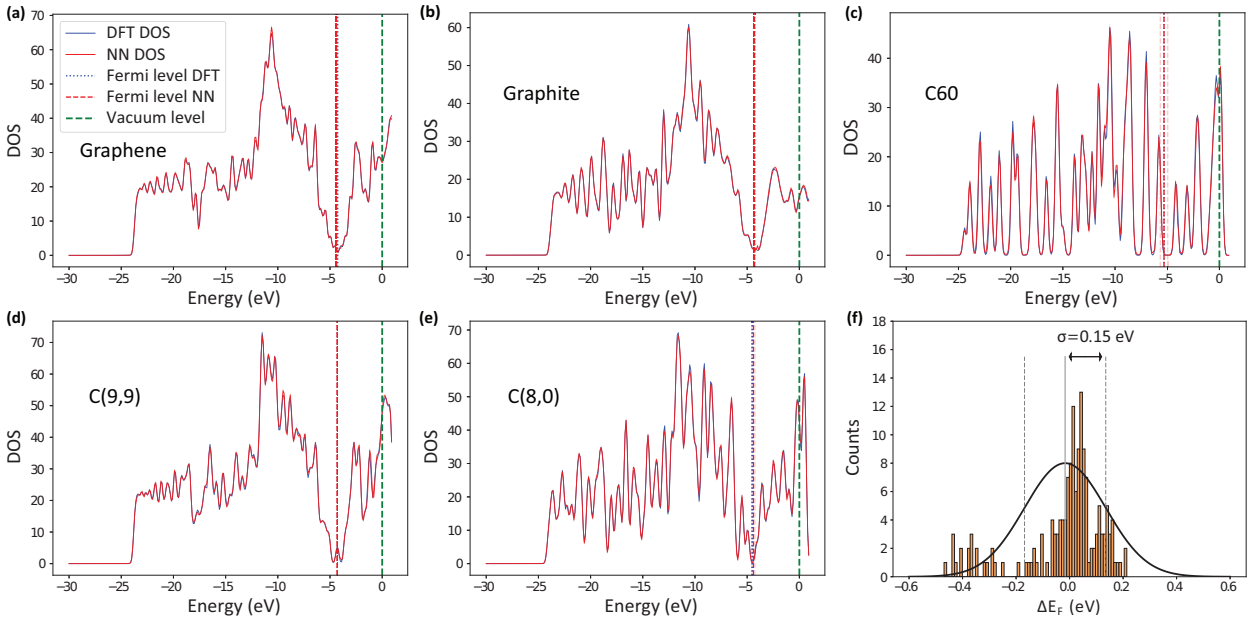

Figure 3: DFT DOS (blue) and predicted NN DOS (red) for test configurations of (a) graphene, (b) graphite, (c) C60, and the SWCNTs of (d) C(9,9) and (e) C(8,0). The DFT and NN Fermi levels calculated as the cumulative integral of the DOS curves are included as vertical dashed blue and red lines, respectively. The vertical dashed green line indicates the vacuum energy used as the global energy reference. The uncertainty in the Fermi level prediction is marked by the dashed pink vertical lines. (f) Histogram of the Fermi level difference between DFT and the NN prediction.

Figure 3 summarizes the results of the model trained and tested on dataset 1. From a 5-fold cross-validation, the predicted DOS curves have a mean root-mean-square-error (RMSE) per atom of 0.0192 states/eV with a standard deviation of 0.0004 states/eV, and a mean $R^2 = 0.9756$ with a standard deviation of 0.0012. Using the separate test set from

dataset 1, we selected the best performing NN model, with $R^2 = 0.977$, RMSE= 0.0188 states eV^{-1}/atom, and 1% highest error (HE) of 0.0716 states eV^{-1}/atom. These accurate metrics are reflected in figure 3 where the NN DOS (red curve) follows very closely the reference DFT DOS (blue curve). Likewise, the calculated Fermi level from the cumulative integral of NN DOS (dashed red line) coincides with the DFT Fermi level (dashed blue line). The dashed pink vertical lines represent the uncertainty in the predicted Fermi level. More information on the evaluation of the uncertainty can be found in the SI.

Besides an accurate DOS prediction, quantities such as bandgap requires a precise calculation of the Fermi level from the predicted DOS. Figure 3 (f) displays the histogram of the error in the calculated Fermi level between the DFT DOS and the NN DOS, for dataset 1. The results follow a Gaussian distribution with a standard deviation of 0.15 eV. The number of instances with a higher error drastically decays after ± 0.2 eV, with maximum values up to -0.47 eV. Albeit such good results, it is worth mentioning that the Fermi level is a very sensitive quantity, especially in cases with bandgaps, where very small deviations from the total number of electrons can shift the Fermi level to the other side of the bandgap.

Systematic Improvement with New Cases

Transferability to new environments along with a capability for systematic improvement are essential for a NN model in a field of ever growing datasets and need to explore newer configurational environments. As such, first we decided to test the NN trained on dataset 1 (Model 1) on the test configurations of dataset 2, and afterwards evaluate the improvement in the predictions once the model is trained on both datasets. Figure 4(a)(b) (grey) shows the results for the mean RMSE of predicted DOS and mean absolute error in the Fermi level calculation. The results on dataset 1 are included as a comparison baseline. As expected when using machine learning models on unseen cases, the performance of Model 1 on dataset 2 is worse than on dataset 1, for both the graphene defects, DS 2 (Defects), and for the highly disordered structures, DS 2 (2000 K). However, given the considerable difference between

the atomic environments and types of carbon hybridization between dataset 1 and dataset 2 (see SI), the results are still surprisingly good. Nevertheless, the model can be extended and improved by training on both datasets 1 and 2 resulting in Model 2, figure 4(a)(b) (dark red). The drastic error decrease on dataset 2 along with the slight error reduction on dataset 1 illustrates the capability of the model for systematic improvement as the dataset size is expanded with entirely new information. In addition, the specific atom-based NN allows to study the atomic DOS of specific atoms, and learns the chemical changes introduced to the electronic structure of the system by these defects (see SI for an example).

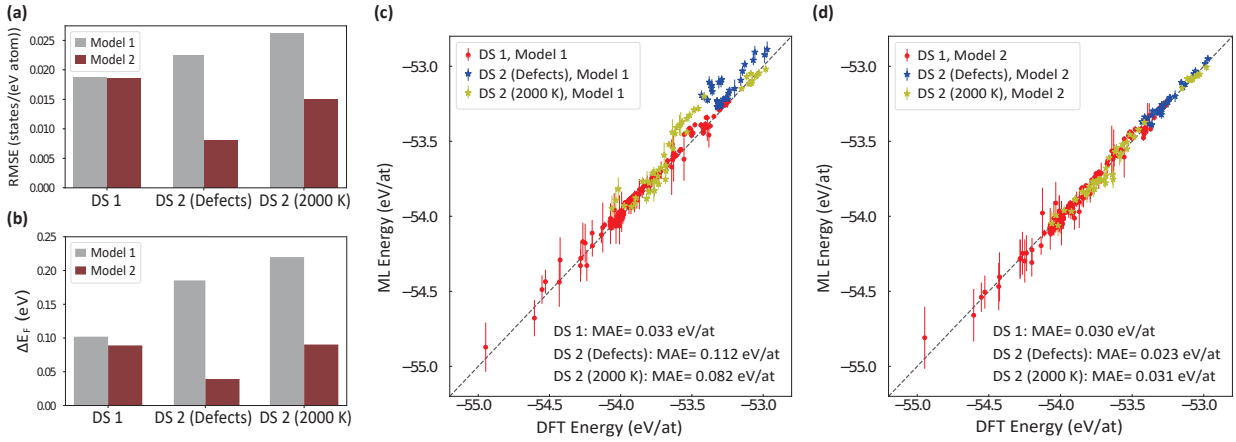

Figure 4: (a)(b) Transferability and systematic improvement of the NN DOS on dataset 2 (DS 2): defects on graphene and highly disordered structures at 2000 K. Mean RMSE (a) and mean absolute error in the Fermi level calculation (b) for Model 1 trained on dataset 1 (DS 1) (grey) and for Model 2 trained on both datasets (dark red). Parity plots for the contribution to the total energy from the NN DOS using Model 1 (c) and Model 2 (d). The error bars represent the standard deviation obtained from the uncertainty in the predicted NN DOS.

Total Energy Contribution

The culminating goal of utilizing machine learning to emulate *and* dramatically accelerate DFT is to bypass the computationally expensive Kohn-Sham equation by directly predicting the electronic structure. To realize this vision, the electronic structure prediction requires highly accurate results of both the eigenvalues and the charge density in order to compute

the total energy as,[2,26]

$$E = 2\sum_{i}^{N_e/2} \epsilon_i - E_H[\rho(\mathbf{r})] + E_{\mathrm{xc}}[\rho(\mathbf{r})] - \int \frac{\delta E_{\mathrm{xc}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} d\mathbf{r} + E_{n-n} \tag{8}$$

where, ρ, N_e, E_H, E_{xc}, E_{n-n} are the charge density, number of electrons, Hartree energy, exchange-correlation energy and nuclear-nuclear interaction energy, respectively. ϵ_i is the eigenvalue of the i-th Kohn-Sham orbital. In Eq. (8), the first term $2\sum_i^{N_e/2} \epsilon_i$ can be written in terms of the DOS as,

$$2\sum_{i}^{N_e/2} \epsilon_i = \int_{-\infty}^{E_{\mathrm{F}}} \mathrm{DOS}(\epsilon)\epsilon d\epsilon \tag{9}$$

while the remaining terms are known functions of the charge density (for a given level of theory).

As a final assessment of the DOS prediction model, we evaluate the accuracy of the contribution to the total energy from the predicted DOS and Fermi level using Equation 9. Figures 4 (c)(d) display the parity plots of the predicted energy contribution compared to the reference energy contribution calculated from the DFT DOS and Fermi level. The performance of Model 1 on both datasets is displayed in figure 4(c). Model 1 successfully predicts the total energy contribution with a mean absolute error (MAE) of 0.033 eV/atom, below the chemical accuracy threshold of 0.043 eV/atom (1 kcal/mol). Nonetheless, the results on graphene with defects and highly disordered structures at 2000 K display a decay in accuracy with a MAE of 0.112 eV/atom and 0.082 eV/atom, respectively. However, this lower accuracy can be mainly ascribed to some specific structure types which present a more significant challenge to Model 1 due to different carbon hybridizations or to more significant disorder in the geometry of the system (see SI). Nevertheless, once the NN is trained on both datasets, the resulting Model 2 outperforms Model 1 for all the datasets, all of them considerably below the chemical accuracy threshold (see figure 4(d)). More detailed results can be found in the SI. These successful results outline the promising capability of our NN

DOS model within the envisaged DFT emulator.

Conclusions

In summary, we have developed a NN DOS predictor which outperforms DFT in computational time by several orders of magnitude while preserving chemical accuracy. Despite the myriad of diverse structures and topologies of carbon considered (albeit within generic sp^2-type environments), the model is flexible enough to perfectly adapt to every atomic environment and to systematically improve the predictions as the dataset is expanded to new chemical spaces. A very promising outcome, owing to the linear scaling of NN with the system size, is the deployment of the DOS prediction model on extremely large systems, impractical with conventional DFT or any available electronic structure code.

Going forward, we plan on exploiting the capabilities of the NN DOS model in two parallel but interconnected pathways. First, to profit from the performance and scaling of the model for large systems, we will develop a NN DOS predictor to provide 'immediate' access to the electronic structure of complex polymeric structures composed of sp, sp^2 and sp^3 hybridizations, *and* multiple elements. To achieve such a goal we will extend the model to include multi-elemental systems starting with hydrocarbons, and progressively expand to polymers with increasing chemical complexity. Second, to continue work towards DFT emulation we will couple our NN DOS predictor along with a charge density predictor to calculate the total energy of the system following Eq. 8. The former pathway will subsequently feed off of the latter, allowing for molecular dynamics simulations of polymers which preserve DFT accuracy and provide information on the electronic structure at each step, all within short computational times.

Despite the outstanding results of the present model, a promising avenue for improvement is the development of alternate fingerprint representations in the form of NNs instead of hand-crafted features. While the fingerprints employed in the present work provide very

good results, such hand-crafted features may impose a bias, limiting the mapping between the structure and the DOS. By eliminating some of those constraints and allowing the NN to find the best mapping, a further increase in accuracy, versatility, and transferability is expected. Promising representations to be considered will employ spherical and icosahedral convolutional NNs within an approach that still preserves the permutation, translation, and rotation invariance of the atomic structure.

Acknowledgement

This work is funded by the National Science Foundation under Award Number 1900017. C. K. is supported by the Alexander von Humboldt Foundation.

Supporting Information Available

Neural network optimization, uncertainty evaluation of the predictions, performance details of Model 1 and Model 2, and atomic NN DOS predictions for graphene with four double vacancies.

References

(1) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Physical Review* **1964**, B864.

(2) Kohn, W.; Sham, L. Self-consistent equations including exchange and correlation effects. *Physical Review* **1965**, A1133–8.

(3) Ceder, G.; Chiang, Y.-M.; Sadoway, D. R.; Aydinol, M. K.; Jang, Y.-I.; Huang, B. Identification of cathode materials for lithium batteries guided by first-principles calculations. *Nature* **1998**, 694–696.

(4) Schlapbach, L.; Züttel, A. Hydrogen-storage materials for mobile applications. *Nature* **2001**, 353–358.

(5) Kang, K.; Meng, Y. S.; Bréger, J.; Grey, C. P.; Ceder, G. Electrodes with high power and high capacity for rechargeable lithium batteries. *Science* **2006**, 977–980.

(6) Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I. B.; Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nat. Mater.* **2006**, 909–913.

(7) Nørskov, J. K.; Abild-Pedersen, F.; Studt, F.; Bligaard, T. Density functional theory in surface chemistry and catalysis. *PNAS* **2011**, 937–943.

(8) Yan, J.; Gorai, P.; Ortiz, B.; Miller, S.; Barnett, S.; Mason, T.; Stevanovićab, V.; Toberer, E. Materials descriptors for predicting thermoelectric performance. *Energy Environ. Sci* **2015**, 983–994.

(9) Seh, Z.; Kibsgaard, J.; Dickens, C.; Chorkendorff, I.; Nørskov, J.; Jaramillo, T. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **2017**, eaad4998.

(10) Körzdörfer, T.; Brédas, J.-L. Organic electronic materials: recent advances in the DFT description of the ground and excited states using tuned range-separated hybrid functionals. *Acc. Chem. Res.* **2014**, 3284–3291.

(11) Yan, F.; Zhang, X.; Yu, Y.; Yu, L.; Nagaraja, A.; Mason, T.; Zunger, A. Design and discovery of a novel half-Heusler transparent hole conductor made of all-metallic heavy elements. *Nat. Commun.* **2015**, 7308.

(12) Christensen, N. E.; Novikov, D. L. Predicted superconductive properties of lithium under pressure. *Phys. Rev. Lett.* **2001**, 1861–1864.

(13) Kolmogorov, A.; Shah, S.; Margine, E.; Bialon, A.; Hammerschmidt, T.; Drautz, R. New superconducting and semiconducting Fe-B compounds predicted with an ab initio evolutionary search. *Phys. Rev. Lett.* **2010**,

(14) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, 146401.

(15) Bartók, A.; Payne, M.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, 136403.

(16) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, 074106.

(17) Botu, V.; Ramprasad, R. Learning scheme to predict atomic forces and accelerate materials simulations. *Phys. Rev. B* **2015**, 094306.

(18) Mueller, T.; Kusne, A.; Ramprasad, R. Machine learning in materials science: Recent progress and emerging applications. *Rev. Comp. Ch.* **2016**, 186–273.

(19) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2017**, 511–522.

(20) Huan, T.; Batra, R.; Chapman, J.; Krishnan, S.; Chen, L.; Ramprasad, R. A universal strategy for the creation of machine learning-based atomistic force fields. *Npj Comput. Mater.* **2017**, 37.

(21) Imbalzano, G.; Anelli, A.; Giofré, D.; Klees, S.; Behler, J.; Ceriotti, M. Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. *J. Chem. Phys* **2018**, 241730.

(22) Mannodi-Kanakkithodi, A.; Chandrasekaran, A.; Kim, C.; Huan, T.; Pilania, G.;

Botu, V.; Ramprasad, R. Scoping the polymer genome: A roadmap for rational polymer dielectrics design and beyond. *Mater. Today* **2018**, 785–96.

(23) Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platformfor Property Predictions. *J. Phys. Chem. C* **2018**, 17575–17585.

(24) Batra, R.; Tran, H.; Kim, C.; Chapman, J.; Chen, L.; Chandrasekaran, A.; Ramprasad, R. General Atomic Neighborhood Fingerprint for Machine Learning-Based Methods. *J. Phys. Chem. C* **2019**, 15859–15866.

(25) Kim, C.; Chandrasekaran, A.; Jha, A.; Ramprasad, R. Active-learning and materials design: the example of high glass transition temperature polymers. *MRS Commun.* **2019**, 860–6.

(26) Chandrasekaran, A.; Kamal, D.; Batra, R.; Kim, C.; Chen, L.; Ramprasad, R. Solving the electronic structure problem with machine learning. *Npj Comput. Mater.* **2019**, 22.

(27) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. . *Phys. Rev. B* **1996**, 11169–86.

(28) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *J. Comput. Mater. Sci.* **1996**, 15–50.

(29) Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. . *Phys. Rev. Lett.* **1996**, 3865–8.

(30) Kresse, G. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, 1758–75.

(31) Monkhorst, H.; Pack, J. Special points for Brillouin-zone integrations. *Phys. Rev. B* **1976**, 5188–92.

(32) Grimme, S. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comp. Chem.* **2006**, 1787.

(33) Gal, Y.; Ghahramani, Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *Proceedings of the 33rd International Conference on Machine Learning (ICML-16)* **2016**,

(34) Gal, Y. Uncertainty in Deep Learning. *Ph.D. thesis, University of Cambridge* **2016**,

(35) Chollet, F., et al. Keras. `https://keras.io`, 2015.

Graphical TOC Entry