

# Can a deep-learning model make fast predictions of vacancy formation in diverse materials?

Cite as: AIP Advances 13, 095109 (2023); doi: 10.1063/5.0135382

Submitted: 17 July 2023 • Accepted: 14 August 2023 •

Published Online: 6 September 2023



View Online



Export Citation



CrossMark

Kamal Choudhary<sup>1,a)</sup> and Bobby G. Sumpter<sup>2</sup>

## AFFILIATIONS

<sup>1</sup> Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

<sup>2</sup> Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

<sup>a)</sup> Author to whom correspondence should be addressed: [kamal.choudhary@nist.gov](mailto:kamal.choudhary@nist.gov)

## ABSTRACT

The presence of point defects, such as vacancies, plays an important role in materials design. Here, we explore the extrapolative power of a graph neural network (GNN) to predict vacancy formation energies. We show that a model trained only on perfect materials can also be used to predict vacancy formation energies ( $E_{vac}$ ) of defect structures without the need for additional training data. Such GNN-based predictions are considerably faster than density functional theory (DFT) calculations and show potential as a quick pre-screening tool for defect systems. To test this strategy, we developed a DFT dataset of 530  $E_{vac}$  consisting of 3D elemental solids, alloys, oxides, semiconductors, and 2D monolayer materials. We analyzed and discussed the applicability of such direct and fast predictions. We applied the model to predict 192 494  $E_{vac}$  for 55 723 materials in the JARVIS-DFT database. Our work demonstrates how a GNN-model performs on unseen data.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0135382>

Defects play an important role in our pursuit to engineer the performance of a material. Vacancies are a type of defects that are ubiquitous, and their presence can significantly alter catalytic, electronic, optoelectronic, electrochemical, diffusion, and neuromorphic properties.<sup>1–10</sup> Experimentally, vacancy formation energies can be determined using positron annihilation experiments.<sup>11</sup> Theoretically, they can be computed using classical force-field (FF)<sup>12,13</sup> and density functional theory (DFT)<sup>2,14–17</sup> calculations. However, such a computation can be very computationally expensive and non-generalizable for DFT and FF based calculations, respectively.

Recently, machine learning techniques have been proposed as a faster method for predicting defect energetics, but they still require time-consuming defect data generation for model training and limit the applicability and generalizability of the defect energetic predictions.<sup>18–27</sup> In particular, graph neural network based deep-learning models<sup>28–32</sup> have become very popular for predicting properties of materials and have been used for several bulk property predictions and their applicability needs to be tested for defect property predictions. Two key ingredients needed for accomplishing this task are (1) a pretrained deep-learning model that can directly predict the total energy of perfect and defect structures and (2) a test

DFT dataset of vacancy formation energies on which the DL model could be applied.

In this work, we demonstrate that the unified atomistic line graph neural network force-field (ALIGNN-FF)<sup>33,34</sup> based total energy prediction model (trained on the JARVIS-DFT OptB88vdW data for perfect bulk materials<sup>35</sup>) can be directly used to predict vacancy formation energy of an arbitrary material without requiring additional training data. The performance in terms of mean absolute error for the energy per atom model was reported as 0.034 eV in Ref. 33. Note that we do not train any machine learning/deep learning model in this work for defects and just use the model parameters for energy prediction that was developed and shared publicly in Ref. 33 showing the extrapolative nature of the model.

Developing a vacancy formation energy dataset can be extremely time-consuming and depends on several computational setup parameters, such as supercell-size, choice of k-points, considering neutral vs charged defects, and selection of appropriate chemical potentials. For testing the strategy adopted in this work, we generated a DFT dataset of 530 vacancy formation energies ( $E_{vac}$ ) in 295 materials with charge-neutral defects using a high-throughput approach. The dataset consists of elemental solids, oxides, alloys,

and 2D materials. We plan to expand the dataset in the future. In addition to predicting the vacancy formation energies, we analyze the trends, strengths, and limitations of such predictions. Finally, we used this strategy to develop a database of vacancy formation energies for all the materials in the JARVIS-DFT database. The deep-learning model, the DFT dataset, and the workflow are made publicly available through the JARVIS (Joint Automated Repository for Various Integrated Simulations) infrastructure.<sup>35</sup>

First, we discuss the generation of our vacancy formation energy dataset that is used for testing the deep-learning model. We obtained stable elemental solids, binary alloys, binary semiconductors, oxides, and 2D materials from the JARVIS-DFT dataset. Some of the alloy systems are Al–Ni, Al–Co, Al–Cu, Al–Ti, Nb–Al, Zr–Ni, and Zr–Cu, while some of the oxides are Cu<sub>2</sub>O, MnO<sub>2</sub>, NiO<sub>2</sub>, TiO<sub>2</sub>, Cu<sub>2</sub>O, GeO<sub>2</sub>, CoO<sub>2</sub>, PbO, and CaO<sub>2</sub>. Similarly, some of the 2D materials are WSe<sub>2</sub>, MoS<sub>2</sub>, SnS, RuCl<sub>3</sub>, FeTe, ZrS<sub>3</sub>, and WTe<sub>2</sub>, while there are 65 elemental solids from the Periodic Table, which were found to be on the convex hull. We used at least 8 Å lattice parameter constraints in x, y, z directions to build the supercell. We removed an atom with a unique Wyckoff position to generate the vacancy structure using the JARVIS-Tools package (<https://github.com/usnistgov/jarvis>). The defect structures were then subjected to energy minimization using the OptB88vdW functional<sup>36</sup> and projected augmented wave formalism<sup>37</sup> in the Vienna *Ab initio* Simulation Package (VASP) package.<sup>38,39</sup> Please note that the commercial software is identified to specify procedures. Such identification does not imply recommendation by the National Institute of Standards and Technology (NIST). We used the converged k-point and cutoff from the JARVIS-DFT dataset based on total energy convergence.<sup>40</sup> We used an energy convergence of 10<sup>-6</sup> eV during the self-consistent cycle. Currently, we have 530 entries for the vacancies and the dataset is still growing.

For the deep-learning predictions, we used the recently developed atomistic line graph neural network force-field (ALIGNN-FF),<sup>33</sup> which is publicly available at <https://github.com/usnistgov/alignn>. ALIGNN was used to train fast and accurate models for more than 70 properties of solids and molecules with high accuracy,<sup>29,30,34,41</sup> and ALIGNN-FF was developed using both energies and forces for applications such as quick structure optimization.

In ALIGNN, a crystal structure is represented as a graph using atomic elements as nodes and atomic bonds as edges. Each node in the atomistic graph is assigned nine input node features based on its atomic species: electronegativity, group number, covalent radius, valence electrons, first ionization energy, electron affinity, block, and atomic volume. The inter-atomic bond distances are used as edge features with radial basis function up to 8 Å cutoff and a 12-nearest-neighbor ( $N$ ). This atomistic graph is then used for constructing the corresponding line graph using interatomic bond-distances as nodes and bond-angles as edge features. ALIGNN uses edge-gated graph convolution for updating nodes and edge features using a propagation function ( $f$ ) for layer ( $l$ ), atom features ( $h$ ), and node ( $i$ ), of which the details can be found in Refs. 33 and 34,

$$h_i^{(l+1)} = f(h_i^l \{h_j^l\}_i). \quad (1)$$

Unlike many other conventional GNNs, ALIGNN uses bond-distances and bond-angles to distinguish atomic structures. The

ALIGNN model is implemented in PyTorch<sup>42</sup> and deep graph library (DGL).<sup>43</sup> A model to predict energy per atom was developed in Ref. 33, which will be used as an energy predictor for both perfect and defect structures in this work.

In Fig. 1, we show the results of the analysis of the DFT database for vacancy formation energies developed in this work. We used this dataset for testing purposes only. Although there have been several studies in generating a vacancy formation energy dataset, a fully atomistic dataset consistent with bulk and vacancy energetics information is not available to our knowledge. Hence, we generated a DFT dataset for vacancies consisting of a wide variety of material classes, such as elemental solids, 2D materials, oxides, and metallic alloys. We visualize the defect formation energies of materials in Fig. 1. As mentioned above, we only considered the charge-neutral vacancies within a finite 8 Å cell size with the OptB88vdW functional.

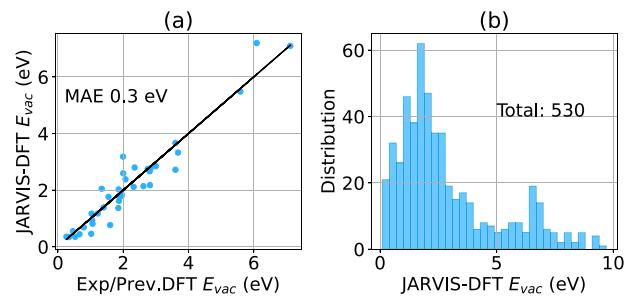
The vacancy formation energy was calculated as

$$E_{vacancy} = E_{defect} - E_{perfect} + \mu, \quad (2)$$

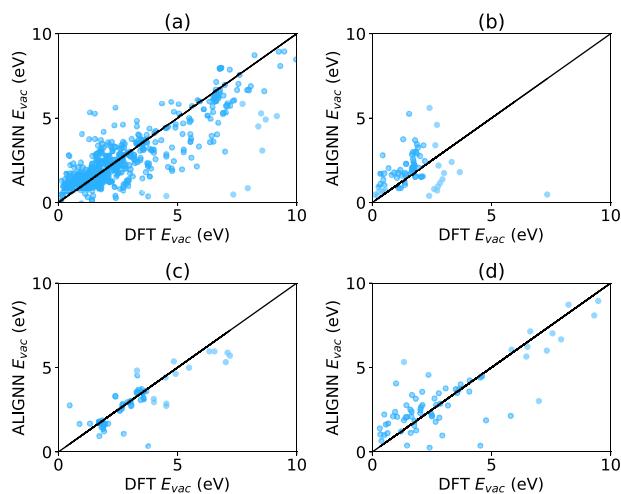
where  $E_{vacancy}$  is the vacancy formation energy,  $E_{defect}$  is the energy of the defect structure with an atom missing,  $E_{perfect}$  is the energy of the perfect structure,  $\mu$  is the chemical potential used as energy per atom of the most stable structure of an element.

We compare a subset of this dataset with data from previous experimental and DFT-studies<sup>14,44–48</sup> in Fig. 1(a). These comparison points were chosen based on the results that we could find in the literature from previous DFT and experimental studies on elemental solids, 2D materials, oxides, semiconductors, etc. We find an excellent agreement between our dataset and that from the literature with a mean absolute error (MAE) of 0.3 eV. In Fig. 1(b), we show the histogram of all the vacancy formation energy data. We find that most of the vacancy formation energy data lie below 3 eV.

Next, we used an ALIGNN-FF based pretrained total energy per atom model trained on the JARVIS-DFT dataset for predicting defect energy and perfect energy required for vacancy formation energy following Eq. (2). This model was trained using energies and forces for 307 113 entries.<sup>33</sup> The defect structures were generated by deleting an atom with a unique Wyckoff position without optimizing the atomic positions of other atoms in the defect structures. We used the same chemical potential for elements from JARVIS-DFT. The



**FIG. 1.** Analysis of vacancy formation energy dataset generated in this work. (a) Comparison of a subset of vacancy formation energies with respect to available experimental and previous DFT calculations from the literature. Our dataset agrees very well with previously reported values. (b) Data distribution of all the vacancy formation energy values. Most of the values are below 3 eV.



**FIG. 2.** Comparison of DFT and ALIGNN-FF (with scissor-shift) based vacancy formation energy predictions for (a) the entire DFT dataset generated in this work, (b) elemental system subset, (c) oxide subset, and (d) 2D monolayer subset.

comparison of ALIGNN-FF based predictions with respect to DFT data is shown with blue dots in Fig. 2(a). Interestingly, we observed that there is a noticeable correlation between the ALIGNN-FF direct predictions and DFT data. In addition, we apply a scissor-correction of 0.8 eV on all data points, resulting in an overall mean absolute error of 0.87 eV compared to DFT. The value of 0.8 eV was chosen as a fitting parameter such that the overall MAE is minimized. The usage of scissor-shifts is not completely uncommon for materials application, for example in Ref. 49. Although we do not have a concrete justification for the fitting parameter right now, we believe that this is due to the dilute limit of the coordination/bonding environments that are missing in the dataset of dense materials. Here, the dilute limit implies that the ALIGNN model has seen dense materials data but very small amount of defect-like environment data. Large errors in vacancy formation energy predictions can also be attributed to relaxation of atomic positions around the defect, which the model might have not seen. We noted that the previous reports on machine learning for vacancy formation energies resulted in mean absolute error values of 0.40 and 0.67 eV in Refs. 18 and 21, respectively. These models were trained on specific material classes, such as the GeTe system and 2D materials, while the approach used here acts as a generalized application of the model with MAE closer to specific case studies. Moreover, comparing these models with ours is not fair because (1) those models did not use a chemically and structurally diverse dataset as ours, (2) we trained on total energy data only, and (3) the datasets and code are not readily available/implementable for those studies and beyond the scope of this work to re-implement/validate.

To further analyze the predictions for different types of materials, we compared the DFT and ALIGNN-FF predictions for elemental solids, oxides, and 2D monolayers in Figs. 2(b)–2(d), respectively. The corresponding ALIGNN-FF scissor-shifted values are shown in Table I. As shown in Table I, the MAEs for the subsets follow the trend of Oxides < All < 2D < Elements, while the Spearman correlation follows the trend of Oxides > All > 2D > Elements. The

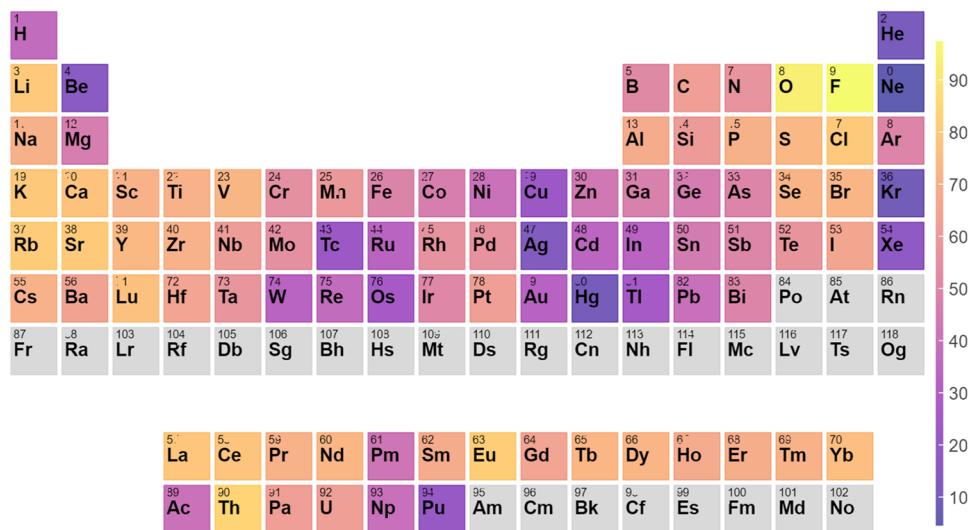
**TABLE I.** Analysis and comparison of ALIGNN based vacancy formation energy with scissor-correction. The mean absolute errors (MAEs) in eV are calculated for all the values as well as oxide and 2D-monolayer subsets.

Sets	Count	Spearman corr.	MAE (eV)
All	530	0.76	0.87
Oxides—subset	65	0.78	0.61
2D materials—subset	72	0.70	1.00
Elements—subset	74	0.30	1.19

Spearman correlation helps understand the trend. Hence, we found that the ALIGNN-FF based models perform well especially for oxide materials. This behavior can be explained based on the fact that GNN architectures usually perform message passing locally and may work well for insulating materials with fewer bonds rather than for elemental solids and other systems that are usually metallic and have delocalized electronic structures leading to a higher number of interatomic bonds. One possible way to further improve the model would be to include more perfect structures that can have a varied number of bonds for various systems. As the JARVIS-DFT dataset for perfect structures is still growing, we believe that the defect property prediction models should further improve in the future.

Now, we applied the above strategy to predict the vacancy formation energies of all the 55 723 materials in the JARVIS-DFT, leading to 192 494 vacancy formation energies. In Fig. 3, we visualize the probability that compounds with vacancy of a given element have a vacancy formation energy of more than 2 eV as a threshold value. Interestingly, we found that C, N, O, and F are some of the common elements with high vacancy formation energies, which is similar to the elemental solid  $E_{vac}$ . These elements usually make ionic/covalent bonded compounds that can have higher bond-breaking enthalpies than compounds with metallic bonds, leading to high  $E_{vac}$  values. Furthermore, we plot the ALIGNN-FF based vacancy formation energy dataset against the OptB88vdW based formation energy available in the JARVIS-DFT database and color code with the corresponding OptB88vdW bandgaps in Fig. 4. We also provided the histogram distribution of the formation energy per atom and vacancy formation energy per atom. Interestingly, we found that high vacancy formation energies were favored by lower formation energies per atom and high electronic bandgaps. We observe that the majority of the  $E_{vac}$  values are around 1 to 3 eV and there are only a very few data points for very high values. We note that very high values of  $E_{vac}$  could be nonphysical.

In summary, we have developed a diverse dataset of vacancy formation energies using density functional theory and demonstrate that the total energy per atom model using ALIGNN-FF can be directly used as a fast pre-screening tool for predicting vacancy formation energies without the need for additional training data. We have discussed the assumptions used in this work, such as excluding charge defects and using finite cell sizes. Most importantly, our work has explored how a deep-learning model performs on unseen data without additional data. A detailed analysis for physical insights into such deep-learning based models and their performance behavior is challenging and is a subject of future work. We have applied a heuristic scissor-shift of energy that further improves the accuracy.



**FIG. 3.** Periodic Table trend of 192 494 vacancy formation energies predicted from scissor-shifted ALIGNN based predictions. We visualize the probability that compounds with vacancy of a given element have a vacancy formation energy of more than 2 eV.

Using the current strategy, we have predicted vacancy formation energies of around 55 723 compounds with 192 494 entries, leading to the largest dataset of defect properties, which could have been very expensive from DFT calculations. We provide data from this work and machine learning models to help accelerate the design of new materials.<sup>50</sup> We anticipate that similar extrapolation studies could be carried out in other branches of materials design, such as catalysts and interfaces and grain-boundaries, where DFT and other methods are highly computationally expensive.

K.C. thanks the National Institute of Standards and Technology for funding, computational, and data-management resources. DFT calculations were also conducted at the Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, a U.S. Department of Energy Office of Science User Facility.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Kamal Choudhary:** Conceptualization (lead); Writing – original draft (equal). **Bobby G. Sumpter:** Conceptualization (supporting); Investigation (supporting); Writing – original draft (equal); Writing – review & editing (equal).

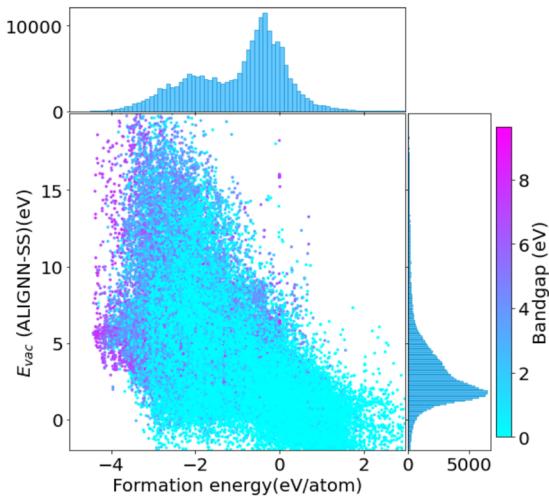
### DATA AVAILABILITY

The data that support the findings of this study are openly available in FigShare at <https://doi.org/10.6084/m9.figshare.23000573>.<sup>50</sup>

The JARVIS-DFT dataset can be obtained from <https://jarvis.nist.gov/> and <https://jarvis-tools.readthedocs.io/en/master/databases.html>. The machine learning model is available at <https://github.com/usnistgov/alignn>. The results from this work are also available on the JARVIS-Leaderboard benchmark [https://pages.nist.gov/jarvis\\_leaderboard/](https://pages.nist.gov/jarvis_leaderboard/).

### REFERENCES

- Y. Hinuma, T. Toyao, T. Kamachi, Z. Maeno, S. Takakusagi, S. Furukawa, I. Takigawa, and K.-I. Shimizu, “Density functional theory calculations of oxygen vacancy formation and subsequent molecular adsorption on oxide surfaces,” *J. Phys. Chem. C* **122**, 29435–29444 (2018).



**FIG. 4.** ALIGNN vacancy formation energy with scissor-shift (SS) dataset (192 494 entries) against OptB88vdW based formation energy available in the JARVIS-DFT database and color coded with corresponding OptB88vdW bandgaps.

- <sup>2</sup>A. A. Emery, J. E. Saal, S. Kirklin, V. I. Hegde, and C. Wolverton, "High-throughput computational screening of perovskites for thermochemical water splitting applications," *Chem. Mater.* **28**, 5621–5634 (2016).
- <sup>3</sup>M. Watkins, D. Pan, E. G. Wang, A. Michaelides, J. VandeVondele, and B. Slater, "Large variation of vacancy formation energies in the surface of crystalline ice," *Nat. Mater.* **10**, 794–798 (2011).
- <sup>4</sup>A. A. Emery and C. Wolverton, "High-throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of  $\text{ABO}_3$  perovskites," *Sci. Data* **4**, 170153 (2017).
- <sup>5</sup>X. Yu, Z. Zhan, J. Rong, Z. Liu, L. Li, and J. Liu, "Vacancy formation energy and size effects," *Chem. Phys. Lett.* **600**, 43–45 (2014).
- <sup>6</sup>S.-L. Shang, B.-C. Zhou, W. Y. Wang, A. J. Ross, X. L. Liu, Y.-J. Hu, H.-Z. Fang, Y. Wang, and Z.-K. Liu, "A comprehensive first-principles study of pure elements: Vacancy formation and migration energies and self-diffusion coefficients," *Acta Mater.* **109**, 128–141 (2016).
- <sup>7</sup>X. Yan, Q. Zhao, A. P. Chen, J. Zhao, Z. Zhou, J. Wang, H. Wang, L. Zhang, X. Li, Z. Xiao *et al.*, "Vacancy-induced synaptic behavior in 2D  $\text{WS}_2$  nanosheet-based memristor for low-power neuromorphic computing," *Small* **15**, 1901423 (2019).
- <sup>8</sup>N. S. Parmar, L. A. Boatner, K. G. Lynn, and J.-W. Choi, "Zn vacancy formation energy and diffusion coefficient of CVT  $\text{ZnO}$  crystals in the sub-surface micron region," *Sci. Rep.* **8**, 13446 (2018).
- <sup>9</sup>J. Ita and R. E. Cohen, "Effects of pressure on diffusion and vacancy formation in MGO from nonempirical free-energy integrations," *Phys. Rev. Lett.* **79**, 3198 (1997).
- <sup>10</sup>W. Huh, D. Lee, and C.-H. Lee, "Memristors based on 2D materials as an artificial synapse for neuromorphic electronics," *Adv. Mater.* **32**, 2002092 (2020).
- <sup>11</sup>B. McKee, W. Triftshäuser, and A. Stewart, "Vacancy-formation energies in metals from positron annihilation," *Phys. Rev. Lett.* **28**, 358 (1972).
- <sup>12</sup>K. Choudhary, A. J. Biacchi, S. Ghosh, L. Hale, A. R. H. Walker, and F. Tavazza, "High-throughput assessment of vacancy formation and surface energies of materials using classical force-fields," *J. Phys.: Condens. Matter* **30**, 395901 (2018).
- <sup>13</sup>S. Huygh, A. Bogaerts, A. C. Van Duin, and E. C. Neyts, "Development of a ReaxFF reactive force field for intrinsic point defects in titanium dioxide," *Comput. Mater. Sci.* **95**, 579–591 (2014).
- <sup>14</sup>B. Medasani, M. Haranczyk, A. Canning, and M. Asta, "Vacancy formation energies in metals: A comparison of MetaGGA with LDA and GGA exchange-correlation functionals," *Comput. Mater. Sci.* **101**, 96–107 (2015).
- <sup>15</sup>H. Ding, B. Medasani, W. Chen, K. A. Persson, M. Haranczyk, and M. Asta, "PyDII: A python framework for computing equilibrium intrinsic point defect concentrations and extrinsic solute site preferences in intermetallic compounds," *Comput. Phys. Commun.* **193**, 118–123 (2015).
- <sup>16</sup>A. Goyal, P. Gorai, S. Anand, E. S. Toberer, G. J. Snyder, and V. Stevanovic, "On the dopability of semiconductors and governing material properties," *Chem. Mater.* **32**, 4467–4480 (2020).
- <sup>17</sup>A. Goyal, P. Gorai, H. Peng, S. Lany, and V. Stevanović, "A computational framework for automation of point defect calculations," *Comput. Mater. Sci.* **130**, 1–9 (2017).
- <sup>18</sup>Y. Cheng, L. Zhu, G. Wang, J. Zhou, S. R. Elliott, and Z. Sun, "Vacancy formation energy and its connection with bonding environment in solid: A high-throughput calculation and machine learning study," *Comput. Mater. Sci.* **183**, 109803 (2020).
- <sup>19</sup>M. Arrigoni and G. K. Madsen, "Evolutionary computing and machine learning for discovering of low-energy defect configurations," *npj Comput. Mater.* **7**, 71 (2021).
- <sup>20</sup>A. Manzoor, G. Arora, B. Jerome, N. Linton, B. Norman, and D. S. Aidhy, "Machine learning based methodology to predict point defect energies in multi-principal element alloys," *Front. Mater.* **8**, 129 (2021).
- <sup>21</sup>N. C. Frey, D. Akinwande, D. Jariwala, and V. B. Shenoy, "Machine learning-enabled design of point defects in 2D materials for quantum and neuromorphic information processing," *ACS Nano* **14**, 13406–13417 (2020).
- <sup>22</sup>A. Mannodi-Kanakkithodi, X. Xiang, L. Jacoby, R. Biegaj, S. T. Dunham, D. R. Gamelin, and M. K. Chan, "Universal machine learning framework for defect predictions in zinc blende semiconductors," *Patterns* **3**, 100450 (2022).
- <sup>23</sup>V. Sharma, P. Kumar, P. Dev, and G. Pilania, "Machine learning substitutional defect formation energies in  $\text{ABO}_3$  perovskites," *J. Appl. Phys.* **128**, 034902 (2020).
- <sup>24</sup>R. B. Wexler, G. S. Gautam, E. B. Stechel, and E. A. Carter, "Factors governing oxygen vacancy formation in oxide perovskites," *J. Am. Chem. Soc.* **143**, 13212–13227 (2021).
- <sup>25</sup>A. M. Deml, A. M. Holder, R. P. OHayre, C. B. Musgrave, and V. Stevanovic, "Intrinsic material properties dictating oxygen vacancy formation energetics in metal oxides," *J. Phys. Chem. Lett.* **6**, 1948–1953 (2015).
- <sup>26</sup>J. B. Varley, A. Samanta, and V. Lordi, "Descriptor-based approach for the prediction of cation vacancy formation energies and transition levels," *J. Phys. Chem. Lett.* **8**, 5059–5063 (2017).
- <sup>27</sup>M. Witman, A. Goyal, T. Ogitsu, A. McDaniel, and S. Lany, "Defect graph neural networks for materials discovery in high-temperature clean-energy applications," *Nat. Comput. Sci.* **3**, 675–686 (2023).
- <sup>28</sup>K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. W. Park, A. Choudhary, A. Agrawal, S. J. Billinge *et al.*, "Recent advances and applications of deep learning methods in materials science," *npj Comput. Mater.* **8**, 59 (2022).
- <sup>29</sup>K. Choudhary, T. Yildirim, D. W. Siderius, A. G. Kusne, A. McDannald, and D. L. Ortiz-Montalvo, "Graph neural network predictions of metal organic framework  $\text{CO}_2$  adsorption properties," *Comput. Mater. Sci.* **210**, 111388 (2022).
- <sup>30</sup>K. Choudhary and K. Garrity, "Designing high- $T_c$  superconductors with BCS-inspired screening, density functional theory, and deep-learning," *npj Comput. Mater.* **8**, 244 (2022).
- <sup>31</sup>V. Fung, J. Zhang, E. Juarez, and B. G. Sumpter, "Benchmarking graph neural networks for materials chemistry," *npj Comput. Mater.* **7**, 84 (2021).
- <sup>32</sup>V. Fung, G. Hu, P. Ganesh, and B. G. Sumpter, "Machine learned features from density of states for accurate adsorption energy prediction," *Nat. Commun.* **12**, 88 (2021).
- <sup>33</sup>K. Choudhary, B. DeCost, L. Major, K. Butler, J. Thiyyagalingam, and F. Tavazza, "Unified graph neural network force-field for the periodic table: Solid state applications," *Digital Discovery* **2**, 346–355 (2023).
- <sup>34</sup>K. Choudhary and B. DeCost, "Atomistic line graph neural network for improved materials property predictions," *npj Comput. Mater.* **7**, 185 (2021).
- <sup>35</sup>K. Choudhary, K. F. Garrity, A. C. Reid, B. DeCost, A. J. Biacchi, A. R. Hight Walker, Z. Trautt, J. Hattrick-Simpers, A. G. Kusne, A. Centrone *et al.*, "The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design," *npj Comput. Mater.* **6**, 173 (2020).
- <sup>36</sup>J. Klimeš, D. R. Bowler, and A. Michaelides, "Chemical accuracy for the van der Waals density functional," *J. Phys.: Condens. Matter* **22**, 022201 (2009).
- <sup>37</sup>P. E. Blöchl, "Projector augmented-wave method," *Phys. Rev. B* **50**, 17953 (1994).
- <sup>38</sup>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).
- <sup>39</sup>G. Kresse and J. Furthmüller, "Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set," *Comput. Mater. Sci.* **6**, 15–50 (1996).
- <sup>40</sup>K. Choudhary and F. Tavazza, "Convergence and machine learning predictions of Monkhorst-Pack k-points and plane-wave cut-off in high-throughput DFT calculations," *Comput. Mater. Sci.* **161**, 300–308 (2019).
- <sup>41</sup>P. R. Kaundinya, K. Choudhary, and S. R. Kalidindi, "Prediction of the electron density of states for crystalline compounds with Atomistic Line Graph Neural Networks (ALIGNN)," *JOM* **74**(4), 1395–1405 (2022).
- <sup>42</sup>A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, "PyTorch: An imperative style, high-performance deep learning library," preprint arXiv:1912.01703 (2019).
- <sup>43</sup>M. Wang, D. Zheng, Z. Ye, Q. Gan, M. Li, X. Song, J. Zhou, C. Ma, L. Yu, Y. Gai *et al.*, "Deep graph library: A graph-centric, highly-performant package for graph neural networks," arXiv:1909.01315 (2019).
- <sup>44</sup>H. H. Pham and L.-W. Wang, "Oxygen vacancy and hole conduction in amorphous  $\text{TiO}_2$ ," *Phys. Chem. Chem. Phys.* **17**, 541–550 (2015).
- <sup>45</sup>A. Kumar, A. Chernatyuk, T. Liang, K. Choudhary, M. J. Noordhoek, Y.-T. Cheng, S. R. Phillpot, and S. B. Sinnott, "Charge optimized many-body (COMB) potential for dynamical simulation of Ni-Al phases," *J. Phys.: Condens. Matter* **27**, 336302 (2015).

<sup>46</sup>K. Choudhary, T. Liang, A. Chernatynskiy, S. R. Phillpot, and S. B. Sinnott, “Charge optimized many-body (COMB) potential for Al<sub>2</sub>O<sub>3</sub> materials, interfaces, and nanostructures,” *J. Phys.: Condens. Matter* **27**, 305004 (2015).

<sup>47</sup>Y. Guo, D. Liu, and J. Robertson, “Chalcogen vacancies in monolayer transition metal dichalcogenides and fermi level pinning at contacts,” *Appl. Phys. Lett.* **106**, 173106 (2015).

<sup>48</sup>K. Wu, T. Liu, R. Sun, J. Song, and C. Shi, “First-principles calculations of oxygen vacancy in CaO crystal,” *Eur. Phys. J. D* **74**, 209 (2020).

<sup>49</sup>C. Wang, P. Elliott, S. Sharma, and J. Dewhurst, “Real time scissor correction in TD-DFT,” *J. Phys.: Condens. Matter* **31**, 214002 (2019).

<sup>50</sup>K. Choudhary (2023). “vacancydb,” Figshare <https://doi.org/10.6084/m9.figshare.23000573.v3>.