Navigating the Materials Space with Machine-Learning-Generated Electronic Fingerprints

I. Neporozhnii,[1,2] Z. Wang,[1,2] R. Bajpai,[1] C. Gomez,[1] N. Chakraborty,[1]
T. Dong,[1] I. Tamblyn,[2,3] S. Hoogland,[2,4] and O. Voznyy[1,2,*]

[1]*Department of Physical and Environmental Sciences, University of Toronto, Scarborough, 1065 Military Trail, Toronto, Ontario M1C 1A4, Canada*

[2]*The Alliance for AI-Accelerated Materials Discovery, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada*

[3]*Department of Physics, University of Ottawa, 150 Louis-Pasteur Pvt, Ottawa, Ontario K1N 6N5, Canada*

[4]*Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada*

(Received 7 July 2024; accepted 15 July 2026; published 25 August 2026)

Identifying materials with strong performance in a specific application, particularly when the origin of that performance is unclear or difficult to compute, remains a central challenge in materials science. Trial-and-error exploration is prohibitively expensive due to the vastness of the materials space. A more practical strategy is to search for new materials within the proximity of known compounds exhibiting the desired property. This requires defining a meaningful representation to assess materials similarity. Structural fingerprints are commonly used, yet structural similarity often fails to translate into similarity in properties. Electronic fingerprints, such as density of states or band structure, were proposed as a better alternative, but computing them for >100 000 materials is still too costly for rapid screening. Here, we present ProDosNet, a graph convolutional network trained on atom- and orbital-resolved projected density of states (PDOS) data, capable of predicting electronic structure at extremely low computational cost. Using this model, we generated PDOS fingerprints for all compounds in the Materials Project database and clustered them by orbital-resolved PDOS similarity. We demonstrate that these electronic fingerprints allow finding materials with similar electronic properties but drastically different structures for applications in photovoltaics, catalysis, and batteries.

DOI: 10.1103/g86f-dn26

I. INTRODUCTION

The global transition to renewable energy sources requires advancements in photovoltaic (PV) devices and battery technology to meet the increasing demand for clean energy generation and storage. The performance of modern PV devices and batteries is mostly limited by the properties of materials used for their production. Furthermore, state-of-the-art materials usually contain expensive and toxic elements that increase device prices and complicate their disposal, limiting accessibility and applications of clean energy technology [1,2]. Therefore, the discovery of new materials, especially lower cost and free from heavy elements, is necessary to facilitate successful transition from fossil-based fuels.

The screening of inorganic materials for specific applications is complicated by the number of possible structures in the materials space, which is estimated to be on the order of 10^{12} [3]. The efficient exploration of such vast space requires methods beyond random sampling and heuristics. A much higher probability of finding promising materials can be achieved by creating a structured materials space where compounds that possess similar properties are clustered together. If such structured materials space can be achieved, new materials can be sampled from the space around known state-of-the-art structures, increasing the probability of successful discovery.

In the field of organic chemistry, molecular fingerprints have been widely used to identify similar molecules and conduct virtual screening of chemical space [4–6]. Molecular fingerprints are fixed-size vectors that encode information about molecular structure. After fingerprints are generated for all molecules in the dataset, similar structures can be found by comparing their fingerprints using metrics such as Tanimoto coefficient [7]. The main disadvantage of the fingerprinting method is that it relies exclusively on the material's structure to generate fingerprints, which often does not translate into similarity in properties [8].

Electronic structure fingerprints that use properties such as density of states (DOS) and band structure were proposed as an alternative to crystal structure fingerprints [9–11]. Incorporation of DOS and band structure information into fingerprint vector enables structurization of materials space by electronic and chemical properties. However, the DOS fingerprinting approach relies on available electronic structure calculations to generate fingerprints. One of the largest publicly available databases of computed material properties—Materials Project—contains information about 154 000 materials; however, only for 60% of these materials density of states

*Contact author: o.voznyy@utoronto.ca

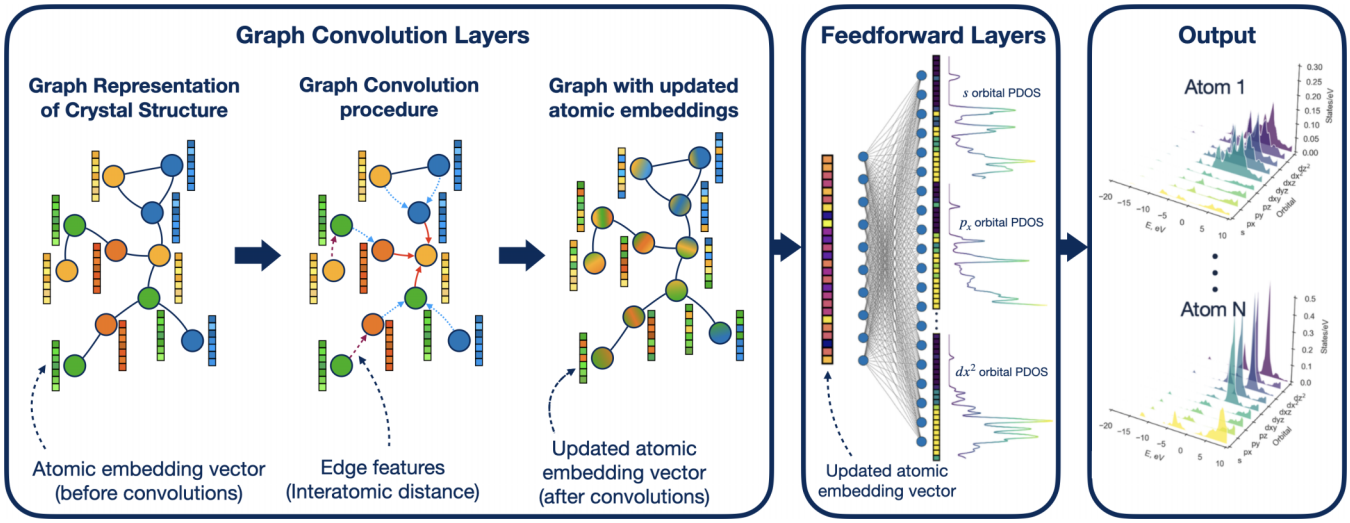

FIG. 1. Schematic representation of ProDosNet model architecture. The first stage contains graph convolution layers that are responsible for the exchange of information between nodes and generation of updated atomic embedding vectors that contain information about atom's local environment. After the graph convolution layers, feedforward layers are used to obtain single orbital densities from updated atomic embedding vectors. The final output of the model is nine orbital densities for each atom of the structure.

is calculated, which severely limits the electronic structure fingerprints approach [12,13]. It would be too slow to compute DOS for more than 60 000 materials using classical methods such as density functional theory (DFT) and even more expensive to apply this approach to a larger space of compounds.

Machine-learning (ML) models demonstrated prominent results in predicting material properties such as formation energy, band gap, etc. Once trained, machine-learning models can produce predictions with extremely low computational cost. Previous studies demonstrated that machine-learning models can predict the DOS of inorganic materials with respectable accuracy [14–19]. However, these studies did not attempt to predict orbital-resolved densities that are important for assessing materials' electronic similarity, nor did they attempt to use this approach for materials screening. In this work, we propose to use the available projected density of states (PDOS) data to train a machine-learning model capable of predicting orbital-resolved DOS and use these predictions to generate electronic fingerprints for all compounds in the materials space.

II. METHODS

A. Graph neural network

Graph convolutional networks (GCNs) are a subclass of machine-learning algorithms designed for operating with graph representation of data. GCNs are widely used in the field of computational chemistry and materials science due to the convenience of graph representation of material structures and leading performance in property prediction tasks. In this work, we developed a ProDosNet–GCN trained to predict PDOS of inorganic materials.

We use a modified version of crystal graph material representation that was first introduced along with the CGCNN model [20]. In crystal graph representation, the material is described by a set of nodes and edges. Nodes represent atoms in the unit cell and contain atomic descriptors, while edges represent bonds and provide information about interatomic distances and crystal geometry. In addition to the original crystal graph atomic descriptors, we also utilize *s*, *p*, and *d* atomic orbital energies and van der Waals, ionic, and covalent radii as input features for each element. The interatomic distances were provided as bond features to enable the model to accurately capture the crystal geometry.

The model receives a graph representation of a crystal structure as input and produces output for each node of the graph (Fig. 1). Node-level predictions correspond to orbital PDOS for each atom in the unit cell of the material and eliminate the need for pooling, a step required to aggregate information from all atoms into one output value. Such an approach allows us to predict and train on orbital densities—the most fundamental components of DOS. Using orbitally resolved data increases the number of training examples for each material by factor $\sigma = N_O \times N_A$, where N_O is the number of orbitals per atom and N_A is the number of atoms in the unit cell. The average factor $\langle\sigma\rangle$ across the dataset is 227, resulting in more than 2 orders of magnitude increase in training information when using orbitally resolved data, compared to training on the total DOS.

B. Training the model

To train the model, we curated a dataset of 23 140 materials from the Materials Project, preprocessing the non-spin-polarized orbital-projected densities onto a standardized 256-point energy grid (see the Supplemental Material [21] for data curation details, which includes Refs. [22–32]).

One of the main challenges that arise when training a model to predict the density of states is a choice of metric to measure model performance. A minor misalignment in sharp density peaks can lead to high loss, even when the overall DOS prediction is accurate. Furthermore, standard loss functions such as root mean square error (RMSE) or mean absolute error (MAE) do not capture improvement, i.e.,

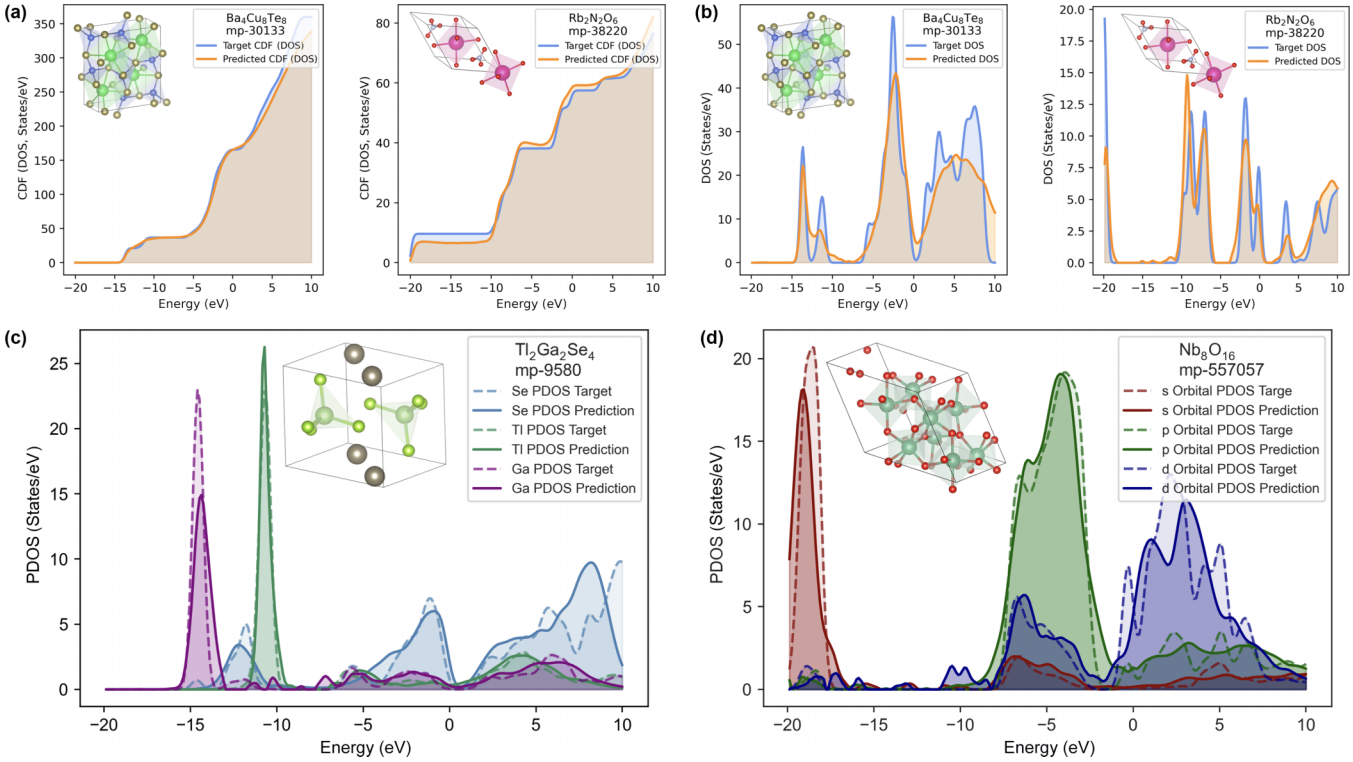

FIG. 2. Median predictions on the test set. (a) The CDF representation of the total DOS predictions, (b) total DOS predictions, (c) element-projected PDOS predictions, and (d) orbital-projected PDOS predictions.

the gradient, in PDOS prediction until target and predicted densities start to overlap, which complicates model training for sharper density peaks. To overcome these issues, we used Wasserstein distance (also known as earth mover's distance) as a metric to measure PDOS similarity and thus the accuracy of PDOS predictions. The Wasserstein distance between two distributions can be defined as the area between their cumulative distribution functions (CDFs) [Eq. (1)]. Unlike RMSE, the Wasserstein metric accounts for energy shifts in density peaks, preventing gradient vanishing problems when peaks are misaligned. The trained model achieved average training loss of 4.29×10^{-3}, validation loss of 6.84×10^{-3}, and test loss of 6.27×10^{-3} (see Fig. S4 in the Supplemental Material [21]).

$$W_p(\alpha, \beta) = \left(\int_{\mathbb{R}} |C_\alpha(x) - C_\beta(x)|^p dx \right)^{\frac{1}{p}}. \quad (1)$$

In this study, factor $p = 2$ is used.

C. PDOS fingerprints

The density of states provides sufficient (and arguably full) insight into materials' electronic and chemical properties and can serve as a fingerprint to measure materials similarity and, consequently, materials screening. Furthermore, explicitly including s, p, and d orbital densities in the fingerprint vectors enables their use for applications where orbital-specific projections can play a crucial role. For example, in optoelectronic applications, transition dipole moments depend on the parity of the participating wavefunctions and only optical transitions between different orbitals are allowed, requiring knowledge of not only the total DOS but also the PDOS. In catalysts, the d-band center approximation, which is often used to predict catalytic activity, relies on the projected d states of transition metals and therefore also requires orbitally resolved PDOS.

To construct the fingerprint vector for each material, we first predict its s, p, and d orbital densities. Each orbital density is represented as an amplitude on a 256-element energy grid that is defined by the ProDosNet model architecture. The total 768-element PDOS fingerprint vector is obtained by concatenation of s, p, and d orbital density vectors. To accommodate materials with varying numbers of atoms, the orbital-resolved PDOS vectors are summed across all atoms in the unit cell. Finally, the obtained PDOS fingerprint vector is normalized by the number of atoms in the unit cell. To measure the similarity of PDOS vectors, we again use Wasserstein distance. Smaller Wasserstein distance indicates more similar PDOS, and therefore more similar electronic and chemical properties of materials.

III. RESULTS

A. Predicting projected density of states

The ProDosNet architecture was designed to enable prediction of PDOS for every orbital of every atom in the material. The representative predictions illustrate the model's ability to predict various projections of DOS, including element- and orbital-specific components (Fig. 2).

Training with Wasserstein distance as a metric resulted in improved accuracy of predictions compared to RMSE or MAE (a detailed comparison of model's performance with different accuracy metrics is provided in the Supplemental

Material [21]). One of the most common visual features encountered throughout the test set predictions is the over-broadening of sharp density peaks by the model. This kind of error, however, does not produce any meaningful deviations in the Wasserstein distance, and is arguably an intended and better outcome of training. In other words, starting from the same DOS and applying different values of broadening to it would produce significantly different RMSE but nearly identical Wasserstein loss. While the model performs well generally, it struggles with environments like molecular crystals (e.g., HCN) and is prone to amplitude normalization errors, as detailed in the worst-case prediction examples shown in Figs. S11–S13 of the Supplemental Material [21].

B. Exploring structured materials space

The trained ProDosNet model was used to predict PDOS for all materials in the Materials Project database, excluding compounds that contain radioactive elements, resulting in a total of 147 472 predictions. The predicted densities were used to generate PDOS fingerprints. To evaluate the applicability of a similarity-based search for accelerated screening of materials, we evaluated four different applications: oxygen evolution catalysis, optoelectronics, solid-state ionic conductors, and Li-ion battery cathodes, using the known state-of-the-art materials as a starting point in each category: IrO_2, GaAs, $Li_{10}Ge(PS_6)_2$, and $LiFePO_4$, respectively. The neighbors were selected from the list of the top 25 most similar stable materials by their predicted PDOS (Fig. 3).

IrO_2 is one of the best-known catalysts for oxygen evolution reaction in acidic environment [33]. Before the advanced DFT methods were developed for measuring the thermodynamic energy barriers for each proton-electron transfer step as well as for surface Pourbaix diagrams calculations, a d-band center was used as an empirical metric of the catalyst activity, which supports the idea that electronic structure fingerprints, and specifically the one that distinguishes s, p, and d orbitals, are potentially a comprehensive descriptor not only for catalytic activity but also for materials stability in acid [34].

The nearest neighbors identified based on PDOS similarity are VCo_3AsO_8, $NiMoO_4$, and RuO_2 [Fig. 3(a)]. An additional imposed requirement was that only metallic materials (i.e., zero band gap) were selected to ensure better electrical conductivity. Because computational validation of oxygen evolution reaction (OER) catalysts using surface descriptors remains unreliable, we instead assess our approach by examining the neighborhood of IrO_2 in the predicted PDOS space [35]. We find that this neighborhood is enriched with experimentally validated, high-performing OER catalysts while also containing previously unexplored compounds.

More than a third of materials in the nearest neighbors list have been experimentally investigated and shown promising OER activity, including RuO_2, $NiTiO_3$, $NiMoO_4$, $CoWO_4$, Co_2GeO_4, PdO_2, $CoSeO_3$, $NiSeO_3$, Fe_2O_3, $MnCoO_3$, $NbMoO_4$, and MoO_2 (the full list of neighbors is shown in Fig. S15 of the Supplemental Material [21]) [36–47]. Importantly, most materials identified as PDOS-similar to IrO_2 have different crystal structure and chemical composition. This highlights the value of electronic-structure-based descriptors, such as PDOS fingerprints, for navigating materials space and identifying candidates for targeted applications.

Gallium arsenide is a well-studied semiconductor material commonly used in the production of photovoltaic devices, light-emitting diodes, and photodetectors. The primary properties of interest, optical band gap and transition dipole moment, can be well captured by the PDOS. Moreover, it is important to distinguish s, p, and d orbitals, as the valence and conduction bands should have different symmetry of the orbitals in order to allow for angular momentum change that enables optically allowed electronic transitions. Some of the nearest neighbors of GaAs by predicted PDOS are InP, GaP, InAs, and their alloys, which are also technologically popular semiconductors with similar properties. The other two close neighbors are $Ca(GaP)_2$ and In_5AgTe_8, both of which are experimentally known. While not necessarily the most desirable compositions in terms of simplicity or cost, they serve to demonstrate the possibility of using a significantly different crystal structure and composition to achieve the same electronic structure. A longer list of neighbors is provided in the Supplemental Material [21].

Solid-state inorganic electrolytes facilitate the emergence of new technologies including solid-state batteries and solid oxide fuel cells [48,49]. While the ionic conductivity seems to be purely defined by the crystal symmetry, such as the availability of empty sites for ion diffusion and Li-anion bond lengths, we hypothesize and argue that PDOS can potentially capture these properties indirectly. Specifically, the coordination number of Li ions and their coupling strength to the anionic lattice will inevitably affect the broadening of the original Li orbitals, which can be captured in PDOS. To test this, we explored the neighbors of $Li_{10}Ge(PS_6)_2$—an electrolyte with one of the highest reported Li ionic conductivities [50]. We identified 25 stable materials from the Materials Project database with the highest predicted PDOS similarity to $Li_{10}Ge(PS_6)_2$, constructing a list of nearest-neighbor candidates. From this list, we manually selected Li_3PS_4 (a well-known decently conducting electrolyte), Li_3BS_3, and Li_3SbS_3 [50,51]. To estimate the ionic conductivity of these materials, we performed DFT-based molecular dynamics simulations and calculated Li-ion diffusivity for each compound. While further investigation is needed to comprehensively evaluate the proposed materials, all compounds demonstrate high ionic conductivity. These results highlight the potential of the machine-learning-generated PDOS fingerprinting approach as a screening tool for identifying promising solid-state inorganic electrolytes. Even more interestingly, alloying all the above compounds can provide a wide compositional space with an opportunity to fine-tune the solid electrolyte conductivity. More details about molecular dynamics simulations and diffusivity analysis are presented in the Supplemental Material [21].

Lithium-ion batteries (LIBs) are dominant in the market of consumer electronics and begin to power electrification of the transportation sector, but their price and performance require further advances to be competitive and surpass the combustion engines [52]. The chemical stability of LIB cathodes is commonly represented through the PDOS, where metal and oxygen orbitals determine the maximum voltage achievable before oxygen evolution happens. Therefore, PDOS can serve

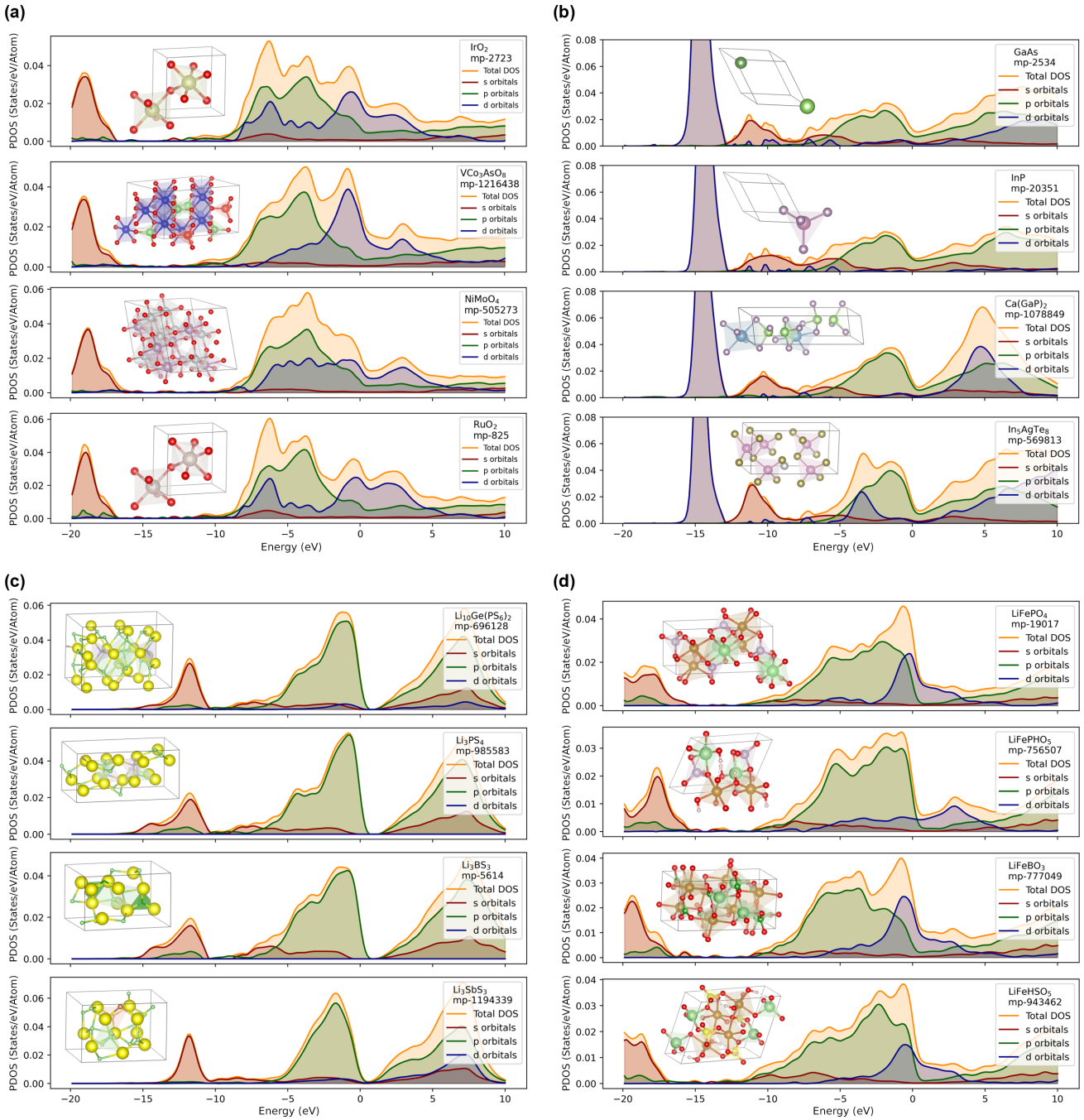

FIG. 3. Similar materials found by model using PDOS fingerprints. Each column shows a target material and three neighbors suggested by the model based on the PDOS similarity. Neighbors are sorted from most similar (first) to least similar (last).

as a proxy to evaluate the suitability of materials as cathodes. As a starting structure, we use $LiFePO_4$ (LFP)—a widely used low-cost cathode material with good stability albeit poor electronic conductivity. Among the nearest neighbors of LFP by PDOS similarity are $LiFePHO_5$, $LiFeBO_3$, and $LiFeHSO_5$, all of which are actively studied as promising cathode materials for LIBs [53–55]. This demonstrates that the utilization of PDOS fingerprints has led to the clustering of perspective cathode materials and successfully created a subspace of potentially valid compounds, which could also be alloyed for further exploration.

Defect-tolerant materials are important for many applications, including solar cells. Although discovering a novel defect-tolerant material is a significant challenge, in this work we would like to demonstrate the potential usefulness of PDOS fingerprints for this task. Previous studies have shown that the antibonding character of the valence band maximum (VBM) orbitals is often a sign of a high defect-tolerant material. While the bonding or antibonding character of the VBM orbitals cannot be determined directly from the predicted PDOS, similar electronic structures may also lead to similar characteristics of VBM orbitals in the neighboring materials.

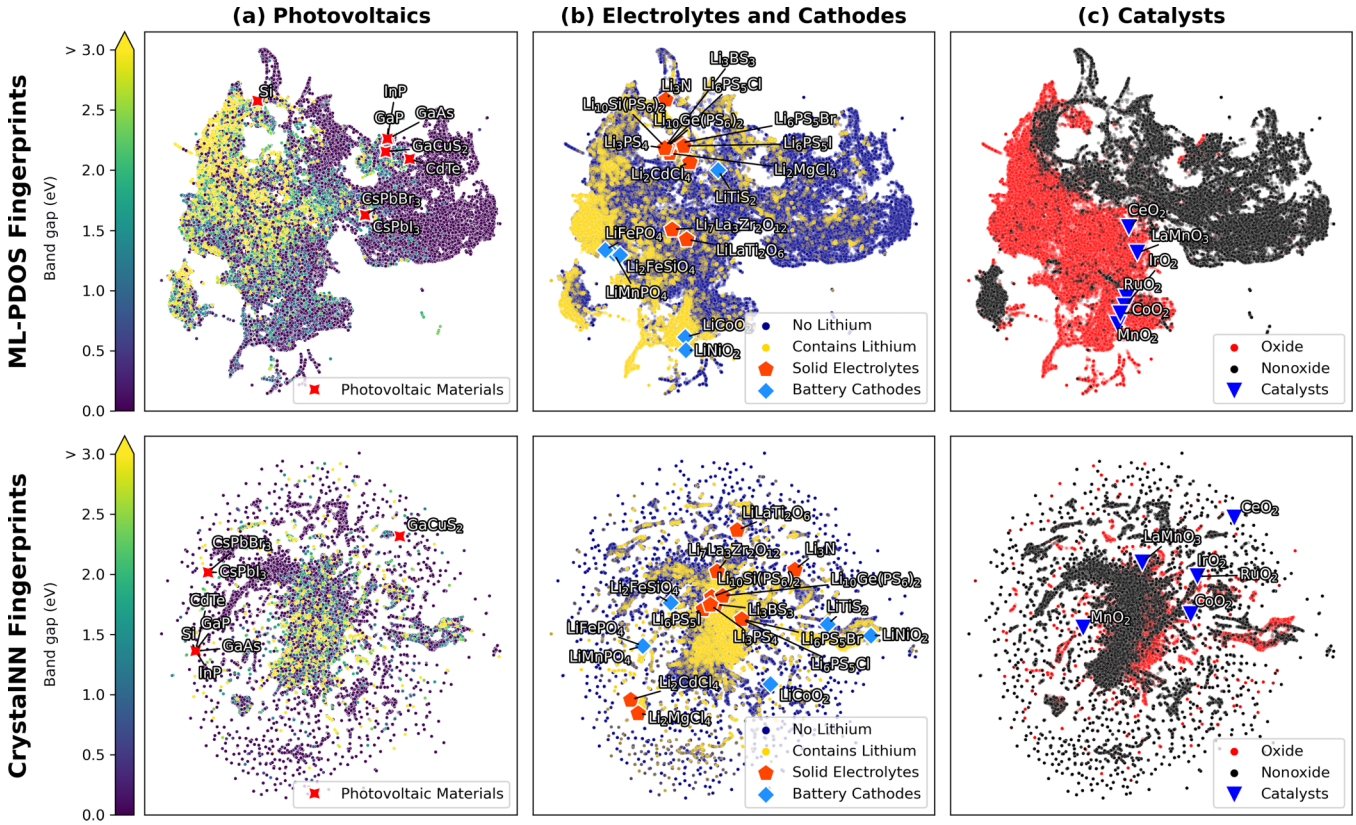

FIG. 4. UMAPs of materials space obtained using ML-generated PDOS fingerprints and CrystalNN fingerprints. The first column shows materials colored according to the band gap energy obtained from Materials Project database. The selected photovoltaic materials are highlighted in red. The second column shows distribution of materials that contain lithium (yellow) and do not contain lithium (blue). The selected solid electrolyte materials are highlighted in orange and battery cathode materials are highlighted in blue. The third column demonstrates distribution of oxide (red) and nonoxide (black) materials. The selected catalysts are highlighted in blue. The top row displays UMAP projections obtained with PDOS fingerprints predicted by ProDosNet, while the bottom row shows projections based on CrystalNN fingerprints.

We selected nearest PDOS neighbors of a $CsPbI_3$ perovskite material known for its defect tolerance and performed DFT calculations of their electronic structures followed by crystal orbital Hamilton population (COHP) analysis to determine the character of VBM orbitals. Our results show that, similarly to $CsPbI_3$, its neighboring compounds $CaPbI_6$, Rb_2SnI_6, and Rb_3BiI_9 also have antibonding character of the VBM orbitals suggesting high defect tolerance. The details of this analysis are presented in the Supplemental Material [21]. Although further research is required to reliably validate the defect tolerance of suggested materials, this example shows another possible application of the ML-generated PDOS fingerprinting approach.

IV. DISCUSSION

We used a trained ProDosNet model to predict orbitally resolved PDOS for 147 472 materials in the Materials Project, which covers more than 95% of the database and only excludes materials that contain elements with atomic numbers above 83. The predicted electronic densities were then used to calculate PDOS fingerprints and create a map of materials space organized by the similarity of electronic densities. We visualized this space in two dimensions using the uniform manifold approximation and projection (UMAP) algorithm (Fig. 4) [56]. We highlighted the state-of-the-art materials for specific applications such as photovoltaics, solid electrolytes, battery cathodes, and catalysts. Finally, we compared our PDOS-based UMAP with those generated using standard inorganic descriptors, including CrystalNN fingerprints (Fig. 4), as well as classical force-field inspired descriptors (CFIDs) and Magpie descriptors (see Fig. S14 in the Supplemental Material [21]) [57–59].

While not trained and not assessed explicitly in the presence of certain elements (e.g., lithium or oxygen), these maps nevertheless group together the oxygen-containing and lithium-containing materials. Similarly, materials with similar band gaps are also grouped. This highlights the predictive ability of the PDOS fingerprints. The maps also show that while some materials targeting the same application group well together, highlighting their electronic similarities, other materials for the same application might be quite far apart. Indeed, e.g., lead halide perovskites possess a very different electronic structure than III–V materials, and this difference isresponsible for defect tolerance of perovskites, which is not present in III–V materials. Battery materials and catalysts also form distinct subgroups. We can also see that their similarity is not purely due to composition; e.g., the space of oxides is

significantly more vast than the space of catalysts. Furthermore, we demonstrate that ML-generated PDOS fingerprints are more versatile and accurate in measuring materials similarity for specific applications compared to CrystalNN fingerprints, CFIDs, and Magpie descriptors (more details in the Supplemental Material [21]).

One limitation of the proposed approach is its reliance on relaxed material geometries to accurately predict the PDOS and generate reliable electronic fingerprints. Although geometry optimization using DFT provides high-quality structures, it remains prohibitively expensive for large-scale materials screening. In contrast, recent machine-learned force fields (MLFFs), such as universal model of atoms (UMA) and message passing atomic cluster expansion (MACE), achieve comparable structural accuracy at a fraction of the computational cost, making them well suited for high-throughput virtual screening applications [60,61]. These MLFFs can predict total energies and atomic forces, enabling efficient structural relaxation of materials. However, unlike DFT, they do not provide direct access to electronic properties of the relaxed structures. We therefore believe that combining MLFF-based structural relaxation with the proposed PDOS fingerprinting methodology offers a promising strategy for prescreening large materials spaces at substantially lower computational cost than conventional DFT-based workflows.

V. CONCLUSIONS

We demonstrate that ML-generated PDOS fingerprints can be used to create a structured materials space where compounds that possess similar electronic properties, and not just structural or compositional features, are clustered together. The presented examples illustrate the versatility of ML-generated PDOS fingerprinting approach across a wide array of applications. These include, but are not limited to, accelerated screening of novel catalysts, battery cathode materials, solid electrolytes, and photovoltaic materials. We also show a potential application of PDOS fingerprints to the identification of defect-tolerant materials. Furthermore, the proposed approach can be extended to a much larger materials space with low additional computational cost. The PDOS fingerprinting approach has the potential to become a valuable tool for high-throughput material exploration by providing application-specific pool of promising materials to validate with more expensive computational or experimental techniques.

ACKNOWLEDGMENTS

The authors acknowledge support from the Alliance for AI-Accelerated Materials Discovery (A3MD). This research was enabled in part by support provided by Compute Ontario [62] and the Digital Research Alliance of Canada [alliancecan.ca]. We thank Ted Sargent, Jason Hattrick-Simpers, Alexander Davis, and Felix Therrien for fruitful discussions.

Conceptualization: I.N. and O.V.; methodology: I.N. and O.V.; software: I.N; data acquisition: I.N. and Z.W.; model development: I.N. and R.B.; analysis of predictions: I.N., C.G., and N.C.; investigation: I.N.; writing original draft, review, and editing: I.N. and O.V.; supervision: O.V., I.T., and S.H.; funding acquisition: O.V.

The authors declare no competing interests.

DATA AVAILABILITY

The code and data for this study are available in Ref. [63].

[1] Editorial. Lithium-ion batteries need to be greener and more ethical, Nature (London) **595**, 7 (2021).

[2] E. A. Olivetti, G. Ceder, G. G. Gaustad, and X. Fu, Lithium-ion battery supply chain considerations: Analysis of potential bottlenecks in critical metals, Joule **1**, 229 (2017).

[3] A. Walsh, The quest for new functionality, Nat. Chem. **7**, 274 (2015).

[4] D. Rogers and M. Hahn, Extended-connectivity fingerprints, J. Chem. Inf. Model. **50**, 742 (2010).

[5] A. Cereto-Massagué, M. J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallvé, and G. Pujadas, Molecular fingerprint similarity search in virtual screening, Methods **71**, 58 (2015).

[6] P. Willett, Similarity-based virtual screening using 2D fingerprints, Drug Discovery Today **11**, 1046 (2006).

[7] M. A. Fligner, J. S. Verducci, and P. E. Blower, A modification of the Jaccard–Tanimoto similarity index for diverse selection of chemical compounds using binary strings, Technometrics **44**, 110 (2002).

[8] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, Automatic chemical design using a data-driven continuous representation of molecules, ACS Cent. Sci. **4**, 268 (2018).

[9] O. Isayev, D. Fourches, E. N. Muratov, C. Oses, K. Rasch, A. Tropsha, and S. Curtarolo, Materials cartography: Representing and mining materials space using structural and electronic fingerprints, Chem. Mater. **27**, 735 (2015).

[10] M. Kuban, S. Rigamonti, M. Scheidgen, and C. Draxl, Density-of-states similarity descriptor for unsupervised learning from materials data, Sci. Data **9**, 646 (2022).

[11] N. R. Knøsgaard and K. S. Thygesen, Representing individual electronic states for machine learning GW band structures of 2D materials, Nat. Commun. **13**, 468 (2022).

[12] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Mater. **1**, 011002 (2013).

[13] J. M. Munro, K. Latimer, M. K. Horton, S. Dwaraknath, and K. A. Persson, An improved symmetry-based approach to reciprocal space path selection in band structure calculations, npj Comput. Mater. **6**, 1 (2020).

[14] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, Solving the electronic structure problem with machine learning, npj Comput. Mater. **5**, 22 (2019).

[15] 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).

[16] K. Bang, B. C. Yeo, D. Kim, S. S. Han, and H. M. Lee, Accelerated mapping of electronic density of states patterns of metallic nanoparticles via machine-learning, Sci. Rep. **11**, 11604 (2021).

[17] S. Kong, F. Ricci, D. Guevarra, J. B. Neaton, C. P. Gomes, and J. M. Gregoire, Density of states prediction for materials discovery via contrastive learning from probabilistic embeddings, Nat. Commun. **13**, 949 (2022).

[18] V. Fung, P. Ganesh, and B. G. Sumpter, Physically informed machine learning prediction of electronic density of states, Chem. Mater. **34**, 4848 (2022).

[19] W. B. How, S. Chong, F. Grasselli, K. K. Huguenin-Dumittan, and M. Ceriotti, Adaptive energy reference for machine-learning models of the electronic density of states, Phys. Rev. Mater. **9**, 013802 (2025).

[20] T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. **120**, 145301 (2018).

[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/g86f-dn26 for details on data curation, model training, model predictions, and results analysis.

[22] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, Automatic differentiation in PyTorch, 2017, https://docs.pytorch.org/tutorials/beginner/basics/autogradqs_tutorial.html.

[23] M. Fey and J. E. Lenssen, Fast graph representation learning with PyTorch geometric, arXiv:1903.02428.

[24] G. Kresse and J. Hafner, *Ab initio* molecular dynamics for liquid metals, Phys. Rev. B **47**, 558 (1993).

[25] 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 (1996).

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

[27] R. Nelson, C. Ertural, J. George, V. L. Deringer, G. Hautier, and R. Dronskowski, LOBSTER: Local orbital projections, atomic charges, and chemical-bonding analysis from projector-augmented-wave-based density-functional theory, J. Comput. Chem. **41**, 1931 (2020).

[28] V. L. Deringer, A. L. Tchougréeff, and R. Dronskowski, Crystal orbital Hamilton population (COHP) analysis as projected from plane-wave basis sets, J. Phys. Chem. A **115**, 5461 (2011).

[29] S. Maintz, V. L. Deringer, A. L. Tchougréeff, and R. Dronskowski, Analytic projection from plane-wave and PAW wavefunctions and application to chemical-bonding analysis in solids, J. Comput. Chem. **34**, 2557 (2013).

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

[31] J. P. Perdew, M. Ernzerhof, and K. Burke, Rationale for mixing exact exchange with density functional approximations, J. Chem. Phys. **105**, 9982 (1996).

[32] A. R. McCluskey, A. G. Squires, J. Dunn, S. W. Coles, and B. J. Morgan, Kinisi: Bayesian analysis of mass transport from molecular dynamics simulations, J. Open Source Software **9**, 5984 (2024).

[33] Z. W. Seh, J. Kibsgaard, C. F. Dickens, I. Chorkendorff, J. K. Nørskov, and T. F. Jaramillo, Combining theory and experiment in electrocatalysis: Insights into materials design, Science **355**, eaad4998 (2017).

[34] G. T. Kasun Kalhara Gunasooriya and J. K. Nørskov, Analysis of acid-stable and active oxides for the oxygen evolution reaction, ACS Energy Lett. **5**, 3778 (2020).

[35] S. Chatterjee, A. Davis, L. Podina, D. Sharma, Y. Bengio, A. Duval, O. Voznyy, A. Hernández-Garcia, D. Rolnick, and F. Therrien, Adsorption energies are necessary but not sufficient to identify good catalysts, arXiv:2512.05938.

[36] C. Roy, R. R. Rao, K. A. Stoerzinger, J. Hwang, J. Rossmeisl, I. Chorkendorff, Y. Shao-Horn, and I. E. L. Stephens, Trends in activity and dissolution on RuO_2 under oxygen evolution conditions: Particles versus well-defined extended surfaces, ACS Energy Lett. **3**, 2045 (2018).

[37] A. Guiet, T. N. Huan, C. Payen, F. Porcher, V. Mougel, M. Fontecave, and G. Corbel, Copper-substituted $NiTiO_3$ ilmenite-type materials for oxygen evolution reaction, ACS Appl. Mater. Interfaces **11**, 31038 (2019).

[38] A. Rajput, M. K. Adak, and B. Chakraborty, Intrinsic lability of $NiMoO_4$ to excel the oxygen evolution reaction, Inorg. Chem. **61**, 11189 (2022).

[39] B. Zhang, X. Zheng, O. Voznyy, R. Comin, M. Bajdich, M. García-Melchor, L. Han, J. Xu, M. Liu, L. Zheng, *et al.*, Homogeneously dispersed multimetal oxygen-evolving catalysts, Science **352**, 333 (2016).

[40] Y. Wang, Y. Guo, J. Wu, H. Liu, and X. Tang, Bifunctional electrocatalyst spinel oxide Co_2GeO_4 and Ni_2GeO_4 with high oxygen evolution reaction and oxygen reduction reaction activity, J. Electron. Mater. **53**, 6563 (2024).

[41] H. He, J. Chen, D. Zhang, F. Li, X. Chen, Y. Chen, L. Bian, Q. Wang, P. Duan, Z. Wen, and X. Lv, Modulating the electrocatalytic performance of palladium with the electronic metal–support interaction: A case study on oxygen evolution reaction, ACS Catal. **8**, 6617 (2018).

[42] S. Anantharaj, H. Sugime, B. Chen, N. Akagi, and S. Noda, Achieving increased electrochemical accessibility and lowered oxygen evolution reaction activation energy for Co^{2+} sites with a simple anion preoxidation, J. Phys. Chem. C **124**, 9673 (2020).

[43] T. Wang, Z. Pei, D. Chen, S. Yang, X. Liu, J. Zhao, X.-P. Zhang, R. Cao, and W. Zhang, Single crystalline CoNi selenite for investigating a metal synergic effect in electrocatalytic water oxidation, ACS Catal. **15**, 11519 (2025).

[44] R. Speelman, E. J. Marker, M. D. Boamah, J. Kupferberg, J. Z. Bye, M. Engelhard, Y. Zhao, A. B. F. Martinson, K. M. Rosso, and F. M. Geiger, Water flipping and the oxygen evolution reaction on Fe_2O_3 nanolayers, Nat. Commun. **16**, 3585 (2025).

[45] J. Villalobos, D. M. Morales, D. Antipin, G. Schuck, R. Golnak, J. Xiao, and M. Risch, Stabilization of a Mn–Co oxide during oxygen evolution in alkaline media, ChemElectroChem **9**, e202200482 (2022).

[46] M. Ram, A. Roy, and K. K. Haldar, Exploring the synergistic electrochemical benefits of a niobium-doped MoO_3/Ta_2O_5 catalyst for an enhanced oxygen evolution reaction, Eur. J. Inorg. Chem. **28**, e202400681 (2025).

[47] P. Guha, B. Mohanty, R. Thapa, R. M. Kadam, P. V. Satyam, and B. K. Jena, Defect-engineered MoO_2 nanostructures as an efficient electrocatalyst for oxygen evolution reaction, ACS Appl. Energy Mater. **3**, 5208 (2020).

[48] J.-M. Tarascon and M. Armand, Issues and challenges facing rechargeable lithium batteries, Nature (London) **414**, 359 (2001).

[49] N. Q. Minh, Ceramic fuel cells, J. Am. Ceram. Soc. **76**, 563 (1993).

[50] J. C. Bachman, S. Muy, A. Grimaud, H.-H. Chang, N. Pour, S. F. Lux, O. Paschos, F. Maglia, S. Lupart, P. Lamp, L. Giordano, and Y. Shao-Horn, Inorganic solid-state electrolytes for lithium batteries: Mechanisms and properties governing ion conduction, Chem. Rev. **116**, 140 (2016).

[51] S. P. Ong, Y. Mo, W. D. Richards, L. Miara, H. S. Lee, and G. Ceder, Phase stability, electrochemical stability and ionic conductivity of the $Li_{10\pm1}MP_2X_{12}$ (M = Ge, Si, Sn, Al or P, and X = O, S or Se) family of superionic conductors, Energy Environ. Sci. **6**, 148 (2012).

[52] A. Masias, J. Marcicki, and W. A. Paxton, Opportunities and challenges of lithium ion batteries in automotive applications, ACS Energy Lett. **6**, 621 (2021).

[53] H. Y. Asl and A. Choudhury, Phosphorous acid route synthesis of iron tavorite phases, $LiFePO_4(OH)_xF_{1-x}[0 \leqslant x \leqslant 1]$ and comparative study of their electrochemical activities, RSC Adv. **4**, 37691 (2014).

[54] A. H. Reshak and W. Khan, Electronic structure, optical and thermoelectric transport properties of layered polyanionic hydrosulfate $LiFeSO_4OH$: Electrode for Li-ion batteries, J. Alloys Compd. **591**, 362 (2014).

[55] L. Tao, G. Rousse, J. N. Chotard, L. Dupont, S. Bruyère, D. Hanžel, G. Mali, R. Dominko, S. Levasseur, and C. Masquelier, Preparation, structure and electrochemistry of $LiFeBO_3$: A cathode material for Li-ion batteries, J. Mater. Chem. A **2**, 2060 (2014).

[56] L. McInnes, J. Healy, N. Saul, and L. Großberger, UMAP: Uniform manifold approximation and projection, J. Open Source Software **3**, 861 (2018).

[57] N. E. R. Zimmermann and A. Jain, Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity, RSC Adv. **10**, 6063 (2020).

[58] K. Choudhary, B. DeCost, and F. Tavazza, Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape, Phys. Rev. Mater. **2**, 083801 (2018).

[59] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, npj Comput. Mater. **2**, 1 (2016).

[60] B. M. Wood, M. Dzamba, X. Fu, M. Gao, M. Shuaibi, L. Barroso-Luque, K. Abdelmaqsoud, V. Gharakhanyan, J. R. Kitchin, D. S. Levine, K. Michel, A. Sriram, T. Cohen, A. Das, A. Rizvi, S. J. Sahoo, Z. W. Ulissi, and C. L. Zitnick, UMA: A family of universal models for atoms, arXiv:2506.23971.

[61] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, *et al.*, A foundation model for atomistic materials chemistry, J. Chem. Phys. **163**, 184110 (2025).

[62] https://www.computeontario.ca.

[63] https://github.com/ineporozhnii/pdos_gnn.