EVALUATING MACHINE LEARNING POTENTIALS ON BULK STRUCTURES WITH NEUTRAL SUBSTITUTIONAL DEFECTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Substitutional defects, either intentionally introduced as dopants or unintentionally as contaminants, are often primary determiners of the performance, efficiency, and versatility of semiconductors. However, the high computational cost of density functional theory (DFT) calculations limits the efficiency of large-scale screening. Machine learning interatomic potentials (MLIP) offer a promising alternative, as they can achieve high accuracy when trained on computational datasets, while being significantly faster than DFT calculations. In this work, we assess the generalization of the MACE-MP potential on a newly developed dataset (Perovs-Dopants) for perovskites with neutral substitutional defects. To disentangle the impact of computational settings from compositional novelty, we performed single-point DFT calculations on a subset of the MPtrj dataset using CP2K, and analyzed the distribution of force discrepancies between different DFT softwares. Our results indicate that both differences in DFT settings and out-of-distribution chemical compositions contribute to the prediction error when using MACE-MP. We then systematically compared standard finetuning and multihead finetuning approaches, demonstrating that multihead finetuning better preserves knowledge from the original training dataset while adapting to the new defect dataset.

1 INTRODUCTION

Among the various families of semiconductors, perovskites have garnered significant attention due to their diverse chemical compositions that create a large design space to work in (Zhang et al., 2017; Green et al., 2014; Fergus, 2007). Perovskites are a class of material with a general chemical formula of ABX_3, and which share the same crystal structure as $CaTiO_3$. Perovskites can exhibit remarkable properties, such as high carrier mobility, tunable band gaps, and strong light absorption, making them ideal for a wide range of applications, including photovoltaics, light-emitting diodes, and sensors (Abdelhady et al., 2016; Colella et al., 2013; Pan et al., 2017; Zhou et al., 2016).

The development of dopants is pivotal for advancing semiconductor technologies, as they enable precise tuning of electronic, optical, and thermal properties to meet the needs of devices. By introducing carefully chosen dopants, materials can be engineered with functionalities tailored for a wide variety of applications. Doping in perovskite materials has been extensively studied. For instance, dopants have been used to improve the operational stability of perovskites in photovoltaic applications (Chan et al., 2017; Ding et al., 2024; Li & Zhang, 2022), reduce nonradiative recombination losses in light-emitting diodes (Luo et al., 2020), and enhance carrier mobility to enable faster operation (Lee et al., 2017; Reo et al., 2022). Dopants are one type of substitutional point defect – specifically, one intentionally engineered to introduce additional free carriers (holes or electrons), enhancing electronic conductivity and related properties of materials. At the atomistic scale, defects can be studied through computational chemistry methods like Density Functional Theory (DFT) calculations, which provide insights into effects on systems' stability, carrier mobility, to name a few (Wang et al., 2012; Hu et al., 2019; Liu et al., 2019). Such calculations are crucial in understanding the mechanisms by which defects alter material behavior, and therefore in designing engineered defects such as dopants. However, these simulations are computationally intensive, especially when evaluating a large number of defect and host material combinations. The vast compositional landscape of perovskites and substituents remains largely unexplored, presenting significant opportunities for discovering materials

with unprecedented performance. Previous work explored how dopants have been difficult to model for current machine learning interatomic potentials (MLIPs) due to distribution shifts in forces and energies (Wang et al., 2024). In this work, we provide further analysis into the challenges of machine learning methods to model substitutional defects.

2 METHOD

For this study, we employed the MACE-MP0B3 model, which was pretrained on the MPtrj dataset that contains 1.5M atomic configurations and DFT calculated properties (Deng et al., 2023; Batatia et al., 2023). To test the model performance on the Perovs-Dopants dataset that includes diverse perovskite relaxation trajectories. (Wang et al., 2024). First, we split it into an 8:1:1 ratio for training, validation and testing. The training set also included 89 isolated atoms to adjust the atomic energies when training the MACE models and account for the variation in the atomic energies between CP2K (Kühne et al., 2020) and VASP (Kresse & Furthmüller, 1996a; Kresse, 1995; Kresse & Hafner, 1994; Kresse & Furthmüller, 1996b; Kresse & Joubert, 1999) DFT codes (Bosoni et al., 2023). To evaluate the impact of using different DFT software on prediction accuracy, we performed single-point calculations using CP2K on a randomly selected subset of the MPtrj test set. The computed forces were then compared to those provided in the original MPtrj dataset. This comparison enabled us to quantify the distribution shift in force predictions arising from the differences in DFT settings, including the pseudopotential, basis sets, and other numerical settings, including the plane-wave cutoffs, k-point density, and smearing. Understanding this shift is crucial for evaluating the generalization of machine learning models trained on data generated with one DFT code but applied to systems evaluated with another.

The pretrained model serves as a starting point for testing on the Perovs-Dopants dataset. Given the differences between these structures and the majority of chemical systems in the MPtrj dataset, we expect some level of fine-tuning to be necessary to adapt the pretrained model to the new data. To test the effectiveness of different finetuning strategies for the MACE-MP model, we compared vanilla finetuning and multihead finetuning approaches. In the vanilla finetuning strategy, the defective perovskite dataset was used to retrain the model, with a single output head predicting energy and forces. The multihead finetuning strategy involves a pretrained output head for retaining knowledge from the original MPtrj training dataset and a finetuned output head tailored to the Perovs-Dopants dataset. Additionally, 10,000 data points were selected from the original MPtrj training set using the farthest point selection algorithm (Li et al., 2022; Han et al., 2023), and these points were combined with the defective perovskite training data for finetuning. Both finetuning strategies were trained with identical hyperparameters, a learning rate of 10^{-5}, and 20 epochs of training. Additionally, we randomly selected 2,000 DFT data points from the MPtrj test set to create a small evaluation dataset to measure the finetuned model's performance on the pretraining dataset.

3 RESULTS

3.1 DISTRIBUTION SHIFT FROM VASP TO CP2K

We describe the details of the Perovs-Dopants dataset in Appendix A.1. First, we aim to understand the distribution shift between different DFT codes. From the CP2K single-point calculations performed on the MPtrj test set, we observed a substantial distribution shift in the calculated forces compared to the original MPtrj dataset, which was generated using VASP. This discrepancy, illustrated in Figure 1, highlights the impact of differences in DFT software, including the use of distinct basis sets and other numerical settings, even though all calculations were performed with the PBE exchange-correlation functional (Perdew et al., 1996). Additionally, we evaluated the performance of the MACE-MP model by comparing its predictions on the MPtrj test set against CP2K-calculated forces for the same structures. Interestingly, the difference between MACE-MP predictions and CP2K forces was comparable to the observed discrepancy between VASP and CP2K forces. These results suggest that both the out-of-domain nature of defective structures and the methodological differences between DFT codes (e.g., plane-wave basis sets in VASP versus Gaussian-type orbitals in CP2K) contribute significantly to the distribution shift. Several studies have specifically examined the impact of different DFT software, pseudopotentials, and basis sets, demonstrating that these factors can lead to significant variations in calculated results (Lejaeghere et al., 2016; Bosoni et al., 2023). This

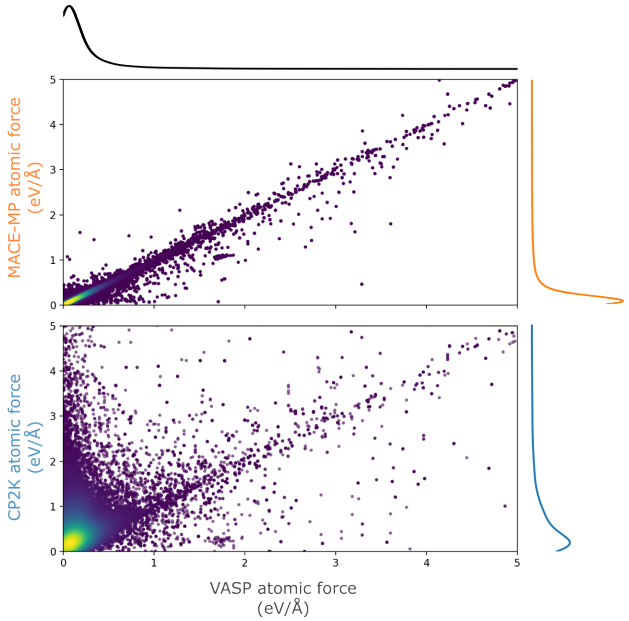

Figure 1: Density plot for the comparison of forces calculated using CP2K and VASP for the MPtrj test set. The distributions include: MACE-MP model predictions on the MPtrj test set, and CP2K-calculated forces for the MPtrj test set, compared to the original VASP calculated forces. The VASP-calculated force distribution is shown in the top panel. The forces predicted with MACE-MP show similar distribution with VASP, while the forces calculated with CP2K show significant shift in distribution.

underscores the dual challenge of adapting ML models to out-of-distribution data and reconciling inconsistencies introduced by computational methods when integrating datasets from diverse sources.

3.2 Finetuning experiments

We extracted descriptors for the MPtrj dataset (Deng et al., 2023), the OMat24 dataset (Barroso-Luque et al., 2024), and the Perovs-Dopants dataset (including all initial structures with different perovskite/substituent element combinations) using three different MACE-MP model checkpoints: the pretrained MACE-MP0B3, MACE-MP0B3 finetuned on Perovs-Dopants, and MACE-MP0B3 multihead-finetuned on Perovs-Dopants. We then performed both t-SNE (Van der Maaten & Hinton, 2008) and PHATE (Moon et al., 2019) analyses to visualize the resulting embeddings in a lower-dimensional space. Both t-SNE and PHATE are dimensionality reduction techniques used to reveal structure in high-dimensional data. While t-SNE primarily preserves local structure, PHATE preserves both local (forming clusters) and global (meaningful intercluster distance) information. From the results visualized in Figure 2, we observe that the embeddings generated from the pretrained MACE-MP (left column) and the multihead-finetuned model (right column) are qualitatively more similar, suggesting that multihead finetuning better preserves the original learned representation from the MPtrj dataset. Additionally, in the PHATE plot for the multihead-finetuned model, the Perovs-Dopants data shifts further away from the MPtrj distribution, indicating that the model recognizes defective perovskites as a distinct class of materials. In contrast, the vanilla finetuning approach (middle column) results in greater overlap between the Perovs-Dopants and MPtrj embeddings, indicating that the model struggles to differentiate defective perovskite structures from the original dataset, potentially leading to catastrophic forgetting.

In the PHATE plots, we used the color map to indicate the energy error from the finetuned and multihead finetuned MACE models on the Perovs-Dopants data. Even after finetuning, we observe that certain structures exhibit much higher energy errors, as indicated by the darker points in the figure. Interestingly, these high-error structures correspond to the systems where Nd is the substituent. As an f-block element, Nd is challenging to model accurately due to its complex electronic structure and

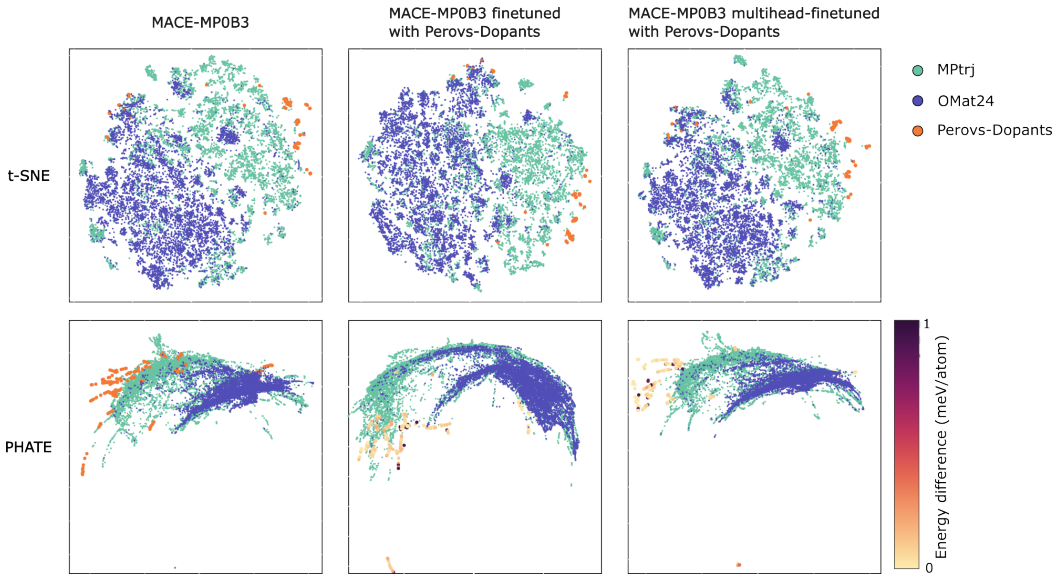

Figure 2: t-SNE (top row) and PHATE (bottom row) visualizations of the learned embeddings for the MPtrj dataset, OMat24 dataset, and Perovs-Dopants dataset using different MACE model checkpoints. The first column represents the pretrained MACE-MP0B3 model, the middle column shows the model finetuned on the Perovs-Dopants dataset, and the right column corresponds to the multihead-finetuned model. The color map for the middle and right PHATE plots indicates the predicted energy error on the defective perovskite structures.

localized 4f electrons. The difficulty in capturing the behavior of f-block elements accurately with the standardized DFT settings in our high-throughput workflow could be a potential contributing factor to the higher prediction errors in these cases. However, as can be seen in Figure A1, other f-block elements do not exhibit significantly higher force errors. Some other factors might be contributing to the high energy errors observed in Nd-containing structures such details of its oxidation state or spin configuration.

Table 1 summarizes the performance of the MACE-MP model under different finetuning strategies across various test datasets. The pretrained MACE-MP model performance on the MPtrj subset is consistent with previously published results, with a force mean absolute error (MAE) of 0.03 eV/Å(Batatia et al., 2023). When the pretrained model was directly applied to the Perovs-Dopants dataset, we observed a decline in performance as the model tends to overestimate the atomic forces. This result is expected and consistent with our earlier analysis: as indicated by Figure 3b, the Perovs-Dopants dataset represents an out-of-domain challenge. While standard finetuning significantly improves accuracy on the Perovs-Dopants dataset, it comes at the cost of catastrophic forgetting, leading to degraded performance on the original MPtrj dataset. In contrast, multihead finetuning, combined with replaying a subset of MPtrj dataset, enables the model to maintain high accuracy on both the MPtrj and Perovs-Dopants datasets.

Another approach we explored was using CP2K to re-evaluate 2,000 randomly selected points from the MPtrj dataset, and use the CP2K evaluated points as a finetuning set for the MACE-mp model. By exposing the model to CP2K-calculated points, we aimed to provide it with some knowledge of the distribution shift between CP2K and VASP. However, when tested on the Perovs-Dopants dataset, the finetuned model still resulted in high force errors. This result provides some evidence that the difference in DFT software may not be the primary factor contributing to the out-of-distribution challenge compared to the novelty in the defective perovskite materials.

In addition to evaluating the MACE-MP model on the Perovs-Dopants dataset, we also tested its performance on the OMat24 dataset, which, like MPtrj, is computed using VASP with similar

Table 1: Performance comparison of the MACE-MP model.

Model	Finetune		Force MAE (eV/Å)		
	Dataset	Method	MPtrj	Perovs-Dopants	OMat24
MACE-MP0B3	N/A		0.027	0.119	0.418
	Perovs-Dopants	Standard	0.246	0.037	
	Perovs-Dopants	Multihead	0.086	0.033	
	MPtrj (CP2K)	Standard	0.096	0.118	
	OMat24	Multihead	0.063	0.118	0.271

DFT settings, with the only difference being the version of pseudopotentials and the electronic minimization algorithm (Barroso-Luque et al., 2024). However, the MACE-MP performance on the OMat24 dataset is worse than that on the Perovs-Dopants datasets (0.4 eV/Åv.s. 0.1 eV/Å). Given the overlap in computational settings, we can attribute any performance issues primarily to compositional differences between the two datasets, rather than discrepancies in the DFT software. The OMat24 dataset contains non-equilibrium structures, despite these structural differences, the embedding analysis revealed significant overlap between MPtrj and OMat24. This suggests that the MACE-MP model's learned representation may not fully capture the atomic interactions required for generalization to other datasets.

4 Conclusion

In this work, we demonstrate our work towards the development of a neutral substitutional defect dataset for perovskite materials. The current benchmark dataset consists of over 20,000 DFT data points from the relaxation trajectories of 438 defective perovskite systems. We investigated the challenges of applying foundation models to defective perovskite structures and tested different finetuning strategies to improve their generalization. We demonstrated the potential of ML models, specifically the MACE-MP0B3 model, to predict the properties of these defective systems efficiently. The t-SNE analysis with the MACE-MP embeddings illustrates that the defective perovskite structures are outside of the distribution of its training set. This emphasizes the importance of building new computational datasets for advancing the development of foundation models and accelerating material discovery. Our results highlight that while the pretrained MACE-MP model shows extraordinary performance on its original MPtrj dataset, it struggles with the defects in perovskites. Finetuning the model improved its accuracy on the Perovs-Dopants dataset, but also resulted in catastrophic forgetting on the MPtrj dataset. Multihead finetuning, combined with the reintroduction of a subset of MPtrj training data, better preserved the original model's learned representations while adapting to the new dataset. Embedding analysis using PHATE and tSNE further confirmed that multihead finetuning maintains the distinction between different datasets.

This study aims to provide a valuable perovskite defect dataset for the materials science community to fill a critical gap in the field of semiconductor research. In this study, all DFT calculations for defective perovskites were performed with neutral systems. We note that to truly evaluate an element as a dopant, it is necessary to perform charged defect calculations to assess its behavior as an effective electron donor or acceptor. However, neutral calculations dramatically simplify the workflow, and stability as a neutral substitutional defect is certainly a necessary (albeit not sufficient) criterion for a dopant. Future extensions of the Perovs-Dopants dataset could incorporate charged systems to explore these important properties in more detail, though we note that appropriate machine learning interatomic potential (MLIP) architectures for such systems are less mature at present and are an area of active development. Future efforts will focus on expanding the dataset to cover more chemical spaces, and exploring other pretrained models' performance on the Perovs-Dopants dataset based on training frameworks for MLIPs training (Miret et al., 2023; Lee et al., 2023). Moving forward, there is a need to improve fine-tuning strategies to maintain the generalization ability of MLIP models. This can include broader integration of multi-head architectures to address the heterogeneity of DFT settings across different datasets, as well as utilizing different pretraining strategies (Lee et al., 2023; Barroso-Luque et al., 2024; Neumann et al., 2024).

REFERENCES

Ahmed L. Abdelhady, Makhsud I. Saidaminov, Banavoth Murali, Valerio Adinolfi, Oleksandr Voznyy, Khabiboulakh Katsiev, Erkki Alarousu, Riccardo Comin, Ibrahim Dursun, Lutfan Sinatra, Edward H. Sargent, Omar F. Mohammed, and Osman M. Bakr. Heterovalent dopant incorporation for bandgap and type engineering of perovskite crystals. *Journal of Physical Chemistry Letters*, 7: 295–301, 1 2016. ISSN 19487185. doi: 10.1021/acs.jpclett.5b02681.

Ahmed H. Al-Naggar, Nanasaheb M. Shinde, Jeom Soo Kim, and Rajaram S. Mane. Water splitting performance of metal and non-metal-doped transition metal oxide electrocatalysts. *Coordination Chemistry Reviews*, 474:214864, 2023. ISSN 00108545. doi: 10.1016/j.ccr.2022.214864. URL https://doi.org/10.1016/j.ccr.2022.214864.

Luis Barroso-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M. Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, C. Lawrence Zitnick, and Zachary W. Ulissi. Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771*, 10 2024.

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, Dávid P. Kovács, Janosh Riebesell, Xavier R. Advincula, Mark Asta, Matthew Avaylon, William J. Baldwin, Fabian Berger, Noam Bernstein, Arghya Bhowmik, Samuel M. Blau, Vlad Cărare, James P. Darby, Sandip De, Flaviano Della Pia, Volker L. Deringer, Rokas Elijošius, Zakariya El-Machachi, Fabio Falcioni, Edvin Fako, Andrea C. Ferrari, Annalena Genreith-Schriever, Janine George, Rhys E. A. Goodall, Clare P. Grey, Petr Grigorev, Shuang Han, Will Handley, Hendrik H. Heenen, Kersti Hermansson, Christian Holm, Jad Jaafar, Stephan Hofmann, Konstantin S. Jakob, Hyunwook Jung, Venkat Kapil, Aaron D. Kaplan, Nima Karimitari, James R. Kermode, Namu Kroupa, Jolla Kullgren, Matthew C. Kuner, Domantas Kuryla, Guoda Liepuoniute, Johannes T. Margraf, Ioan-Bogdan Magdău, Angelos Michaelides, J. Harry Moore, Aakash A. Naik, Samuel P. Niblett, Sam Walton Norwood, Niamh O'Neill, Christoph Ortner, Kristin A. Persson, Karsten Reuter, Andrew S. Rosen, Lars L. Schaaf, Christoph Schran, Benjamin X. Shi, Eric Sivonxay, Tamás K. Stenczel, Viktor Svahn, Christopher Sutton, Thomas D. Swinburne, Jules Tilly, Cas van der Oord, Eszter Varga-Umbrich, Tejs Vegge, Martin Vondrák, Yangshuai Wang, William C. Witt, Fabian Zills, and Gábor Csányi. A foundation model for atomistic materials chemistry. *arXiv:2401.00096*, 12 2023. URL http://arxiv.org/abs/2401.00096.

Emanuele Bosoni, Louis Beal, Marnik Bercx, Peter Blaha, Stefan Blügel, Jens Bröder, Martin Callsen, Stefaan Cottenier, Augustin Degomme, Vladimir Dikan, Kristjan Eimre, Espen Flage-Larsen, Marco Fornari, Alberto Garcia, Luigi Genovese, Matteo Giantomassi, Sebastiaan P. Huber, Henning Janssen, Georg Kastlunger, Matthias Krack, Georg Kresse, Thomas D. Kühne, Kurt Lejaeghere, Georg K.H. Madsen, Martijn Marsman, Nicola Marzari, Gregor Michalicek, Hossein Mirhosseini, Tiziano M.A. Müller, Guido Petretto, Chris J. Pickard, Samuel Poncé, Gian Marco Rignanese, Oleg Rubel, Thomas Ruh, Michael Sluydts, Danny E.P. Vanpoucke, Sudarshan Vijay, Michael Wolloch, Daniel Wortmann, Aliaksandr V. Yakutovich, Jusong Yu, Austin Zadoks, Bonan Zhu, and Giovanni Pizzi. How to verify the precision of density-functional-theory implementations via reproducible and universal workflows. *Nature Reviews Physics*, 2023. ISSN 25225820. doi: 10.1038/s42254-023-00655-3.

Ivano E. Castelli, David D. Landis, Kristian S. Thygesen, Soren Dahl, Ib Chorkendorff, Thomas F. Jaramillo, and Karsten W. Jacobsen. New cubic perovskites for one- and two-photon water splitting using the computational materials repository. *Energy and Environmental Science*, 5:9034–9043, 10 2012. ISSN 17545692. doi: 10.1039/c2ee22341d.

Shun Hsiang Chan, Ming Chung Wu, Kun Mu Lee, Wei Cheng Chen, Tzu Hao Lin, and Wei Fang Su. Enhancing perovskite solar cell performance and stability by doping barium in methylammonium lead halide. *Journal of Materials Chemistry A*, 5(34):18044–18052, 2017. ISSN 20507496. doi: 10.1039/c7ta05720b.

Silvia Colella, Edoardo Mosconi, Paolo Fedeli, Andrea Listorti, Francesco Gazza, Fabio Orlandi, Patrizia Ferro, Tullo Besagni, Aurora Rizzo, Gianluca Calestani, Giuseppe Gigli, Filippo De Angelis, and Roberto Mosca. Mapbi3-xclx mixed halide perovskite for hybrid solar cells: The role of chloride as dopant on the transport and structural properties. *Chemistry of Materials*, 25: 4613–4618, 11 2013. ISSN 0897-4756. doi: 10.1021/cm402919x.

Bowen Deng, Peichen Zhong, Kyu Jung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5:1031–1041, 9 2023. ISSN 25225839. doi: 10.1038/s42256-023-00716-3.

Bin Ding, Yong Ding, Jun Peng, Jan Romano-deGea, Lindsey E.K. Frederiksen, Hiroyuki Kanda, Olga A. Syzgantseva, Maria A. Syzgantseva, Jean Nicolas Audinot, Jerome Bour, Song Zhang, Tom Wirtz, Zhaofu Fei, Patrick Dörflinger, Naoyuki Shibayama, Yunjuan Niu, Sixia Hu, Shunlin Zhang, Farzaneh Fadaei Tirani, Yan Liu, Guan Jun Yang, Keith Brooks, Linhua Hu, Sachin Kinge, Vladimir Dyakonov, Xiaohong Zhang, Songyuan Dai, Paul J. Dyson, and Mohammad Khaja Nazeeruddin. Dopant-additive synergism enhances perovskite solar modules. *Nature*, 628(8007): 299–305, 2024. ISSN 14764687. doi: 10.1038/s41586-024-07228-z.

Jeffrey W. Fergus. Perovskite oxides for semiconductor-based gas sensors, 5 2007. ISSN 09254005.

Alex Ganose, Hrushikesh Sahasrabuddhe, Mark Asta, Kevin Beck, Tathagata Biswas, Alexander Bonkowski, Joana Bustamante, Xin Chen, Yuan Chiang, Daryl Chrzan, Jacob Clary, Orion Cohen, Christina Ertural, Max Gallant, Janine George, Sophie Gerits, Rhys Goodall, Rishabh Guha, Geoffroy Hautier, Matthew Horton, Aaron Kaplan, Ryan Kingsbury, Matthew Kuner, Bryant Li, Xavier Linn, Matthew McDermott, Rohith Srinivaas Mohanakrishnan, Aakash Naik, Jeffrey Neaton, Kristin Persson, Guido Petretto, Thomas Purcell, Francesco Ricci, Benjamin Rich, Janosh Riebesell, Gian-Marco Rignanese, Andrew Rosen, Matthias Scheffler, Jonathan Schmidt, Jimmy-Xuan Shen, Andrei Sobolev, Ravishankar Sundararaman, Cooper Tezak, Victor Trinquet, Joel Varley, Derek Vigil-Fowler, Duo Wang, David Waroquiers, Mingjian Wen, Han Yang, Hui Zheng, Jiongzhi Zheng, Zhuoying Zhu, and Anubhav Jain. Atomate2: Modular workflows for materials science, jan 2025. URL https://chemrxiv.org/engage/chemrxiv/article-details/678e76a16dde43c9085c75e9.

Martin A. Green, Anita Ho-Baillie, and Henry J. Snaith. The emergence of perovskite solar cells, 2014. ISSN 17494893.

Meng Han, Liang Wang, Limin Xiao, Hao Zhang, Chenhao Zhang, Xiangrong Xu, and Jianfeng Zhu. Quickfps: Architecture and algorithm co-design for farthest point sampling in large-scale point clouds. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2023.

Shuwei Hu, Bin Liu, Zhen Li, Jian Zhou, and Zhimei Sun. Identifying optimal dopants for Sb2Te3 phase-change material by high-throughput ab initio calculations with experiments. *Computational Materials Science*, 165(March):51–58, 2019. ISSN 09270256. doi: 10.1016/j.commatsci.2019.04.028. URL https://doi.org/10.1016/j.commatsci.2019.04.028.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation, 2013. ISSN 2166532X.

G. Kresse. Ab initio molecular dynamics for liquid metals. *Journal of Non-Crystalline Solids*, 192-193:222–229, 1995. ISSN 00223093. doi: 10.1016/0022-3093(95)00355-X.

G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, jul 1996a. ISSN 09270256. doi: 10.1016/0927-0256(96)00008-0.

G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B - Condensed Matter and Materials Physics*, 54(16): 11169–11186, 1996b. ISSN 1550235X. doi: 10.1103/PhysRevB.54.11169.

G. Kresse and J. Hafner. Ab initio molecular-dynamics simulation of the liquid-metalamorphous-semiconductor transition in germanium. *Physical Review B*, 49(20):14251–14269, 1994. ISSN 01631829. doi: 10.1103/PhysRevB.49.14251.

G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758–1775, jan 1999. ISSN 0163-1829. doi: 10.1103/PhysRevB.59.1758. URL https://link.aps.org/doi/10.1103/PhysRevB.59.1758.

Thomas D. Kühne, Marcella Iannuzzi, Mauro Del Ben, Vladimir V. Rybkin, Patrick Seewald, Frederick Stein, Teodoro Laino, Rustam Z. Khaliullin, Ole Schütt, Florian Schiffmann, Dorothea Golze, Jan Wilhelm, Sergey Chulkov, Mohammad Hossein Bani-Hashemian, Valéry Weber, Urban Borštnik, Mathieu Taillefumier, Alice Shoshana Jakobovits, Alfio Lazzaro, Hans Pabst, Tiziano Müller, Robert Schade, Manuel Guidon, Samuel Andermatt, Nico Holmberg, Gregory K. Schenter, Anna Hehn, Augustin Bussy, Fabian Belleflamme, Gloria Tabacchi, Andreas Glöß, Michael Lass, Iain Bethune, Christopher J. Mundy, Christian Plessl, Matt Watkins, Joost VandeVondele, Matthias Krack, and Jürg Hutter. Cp2k: An electronic structure and molecular dynamics software package -quickstep: Efficient and accurate electronic structure calculations. *Journal of Chemical Physics*, 152, 5 2020. ISSN 10897690. doi: 10.1063/5.0007045.

Kin Long Kelvin Lee, Carmelo Gonzales, Marcel Nassar, Matthew Spellings, Mikhail Galkin, and Santiago Miret. Matsciml: A broad, multi-task benchmark for solid-state materials modeling. *arXiv preprint arXiv:2309.05934*, 2023.

Woong Jhae Lee, Hyung Joon Kim, Jeonghun Kang, Dong Hyun Jang, Tai Hoon Kim, Jeong Hyuk Lee, and Kee Hoon Kim. Transparent Perovskite Barium Stannate with High Electron Mobility and Thermal Stability. *Annual Review of Materials Research*, 47:391–423, 2017. ISSN 15317331. doi: 10.1146/annurev-matsci-070616-124109.

Kurt Lejaeghere, Gustav Bihlmayer, Torbjörn Björkman, Peter Blaha, Stefan Blügel, Volker Blum, Damien Caliste, Ivano E. Castelli, Stewart J. Clark, Andrea Dal Corso, Stefano De Gironcoli, Thierry Deutsch, John Kay Dewhurst, Igor Di Marco, Claudia Draxl, Marcin Dułak, Olle Eriksson, José A. Flores-Livas, Kevin F. Garrity, Luigi Genovese, Paolo Giannozzi, Matteo Giantomassi, Stefan Goedecker, Xavier Gonze, Oscar Grånäs, E. K.U. Gross, Andris Gulans, François Gygi, D. R. Hamann, Phil J. Hasnip, N. A.W. Holzwarth, Diana Iuşan, Dominik B. Jochym, François Jollet, Daniel Jones, Georg Kresse, Klaus Koepernik, Emine Küçükbenli, Yaroslav O. Kvashnin, Inka L.M. Locht, Sven Lubeck, Martijn Marsman, Nicola Marzari, Ulrike Nitzsche, Lars Nordström, Taisuke Ozaki, Lorenzo Paulatto, Chris J. Pickard, Ward Poelmans, Matt I.J. Probert, Keith Refson, Manuel Richter, Gian Marco Rignanese, Santanu Saha, Matthias Scheffler, Martin Schlipf, Karlheinz Schwarz, Sangeeta Sharma, Francesca Tavazza, Patrik Thunström, Alexandre Tkatchenko, Marc Torrent, David Vanderbilt, Michiel J. Van Setten, Veronique Van Speybroeck, John M. Wills, Jonathan R. Yates, Guo Xu Zhang, and Stefaan Cottenier. Reproducibility in density functional theory calculations of solids. *Science*, 351(6280), 2016. ISSN 10959203. doi: 10.1126/science.aad3000.

Bowei Li and Wei Zhang. Improving the stability of inverted perovskite solar cells towards commercialization. *Communications Materials*, 3(1):1–13, 2022. ISSN 26624443. doi: 10.1038/s43246-022-00291-x.

Jingtao Li, Jian Zhou, Yan Xiong, Xing Chen, and Chaitali Chakrabarti. An adjustable farthest point sampling method for approximately-sorted point cloud data. In *2022 IEEE Workshop on Signal Processing Systems (SiPS)*, pp. 1–6, 2022. doi: 10.1109/SiPS55645.2022.9919246.

Tianjun Liu, Xiaoming Zhao, Jianwei Li, Zilu Liu, Fabiola Liscio, Silvia Milita, Bob C. Schroeder, and Oliver Fenwick. Enhanced control of self-doping in halide perovskites for improved thermoelectric performance. *Nature Communications*, 10(1):1–9, 2019. ISSN 20411723. doi: 10.1038/s41467-019-13773-3. URL http://dx.doi.org/10.1038/s41467-019-13773-3.

Deying Luo, Rui Su, Wei Zhang, Qihuang Gong, and Rui Zhu. Minimizing non-radiative recombination losses in perovskite solar cells. *Nature Reviews Materials*, 5(1):44–60, 2020. ISSN 20588437. doi: 10.1038/s41578-019-0151-y. URL http://dx.doi.org/10.1038/s41578-019-0151-y.

Santiago Miret, Kin Long Kelvin Lee, Carmelo Gonzales, Marcel Nassar, and Matthew Spellings. The open matsci ML toolkit: A flexible framework for machine learning in materials science. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=QBMyDZsPMd.

Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Natalia B.

Ivanova, Guy Wolf, and Smita Krishnaswamy. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37:1482–1492, 12 2019. ISSN 15461696. doi: 10.1038/s41587-019-0336-3.

Christopher P. Muzzillo, Cristian V. Ciobanu, and David T. Moore. High-entropy alloy screening for halide perovskites. *Materials Horizons*, 5 2024. ISSN 20516355. doi: 10.1039/d4mh00464g.

Mark Neumann, James Gin, Benjamin Rhodes, Steven Bennett, Zhiyi Li, Hitarth Choubisa, Arthur Hussey, and Jonathan Godwin. Orb: A Fast, Scalable Neural Network Potential. *arXiv preprint arXiv:2410.22570*, pp. 1–26, 2024. URL http://arxiv.org/abs/2410.22570.

Gencai Pan, Xue Bai, Dongwen Yang, Xu Chen, Pengtao Jing, Songnan Qu, Lijun Zhang, Donglei Zhou, Jinyang Zhu, Wen Xu, Biao Dong, and Hongwei Song. Doping lanthanide into perovskite nanocrystals: Highly improved and expanded optical properties. *Nano Letters*, 17:8005–8011, 12 2017. ISSN 1530-6984. doi: 10.1021/acs.nanolett.7b04575.

John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple, 1996.

Youjin Reo, Huihui Zhu, Ao Liu, and Yong Young Noh. Molecular Doping Enabling Mobility Boosting of 2D Sn2+-Based Perovskites. *Advanced Functional Materials*, 32(38):1–9, 2022. ISSN 16163028. doi: 10.1002/adfm.202204870.

R. Rohib, Saeed Ur Rehman, Eunjik Lee, Changki Kim, Hyunjoon Lee, Seung Bok Lee, and Gu Gon Park. Synergistic effect of perovskites and nitrogen-doped carbon hybrid materials for improving oxygen reduction reaction. *Scientific Reports*, 13, 12 2023. ISSN 20452322. doi: 10.1038/s41598-023-47304-4.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Joost VandeVondele and Jürg Hutter. An efficient orbital transformation method for electronic structure calculations. *The Journal of Chemical Physics*, 118(10):4365–4369, 03 2003. ISSN 0021-9606. doi: 10.1063/1.1543154. URL https://doi.org/10.1063/1.1543154.

Fenggong Wang, Cristiana Di Valentin, and Gianfranco Pacchioni. Doping of WO 3 for photocatalytic water splitting: Hints from density functional theory. *Journal of Physical Chemistry C*, 116(16): 8901–8909, 2012. ISSN 19327447. doi: 10.1021/jp300867j.

Xiaoxiao Wang, Suehyun Park, and Santiago Miret. Perovs-dopants: Machine learning potentials for doped bulk structures. In *AI for Accelerated Materials Design - NeurIPS 2024*, 2024. URL https://openreview.net/forum?id=sEpHuS8CWQ.

Xu Zhang, Xiaodong Ren, Bin Liu, Rahim Munir, Xuejie Zhu, Dong Yang, Jianbo Li, Yucheng Liu, Detlef M. Smilgies, Ruipeng Li, Zhou Yang, Tianqi Niu, Xiuli Wang, Aram Amassian, Kui Zhao, and Shengzhong Liu. Stable high efficiency two-dimensional perovskite solar cells via cesium doping. *Energy and Environmental Science*, 10:2095–2102, 10 2017. ISSN 17545706. doi: 10.1039/c7ee01145h.

Yunqin Zhang, Datao Tu, Luping Wang, Chenliang Li, Yuhan Liu, and Xueyuan Chen. Transition metal ion-doped cesium lead halide perovskite nanocrystals: doping strategies and luminescence design. *Materials Chemistry Frontiers*, 8(1):192–209, 2023. ISSN 20521537. doi: 10.1039/d3qm00691c.

Yuanyuan Zhou, Zhongmin Zhou, Min Chen, Yingxia Zong, Jinsong Huang, Shuping Pang, and Nitin P. Padture. Doping and alloying for improved perovskite solar cells. *Journal of Materials Chemistry A*, 4(45):17623–17635, 2016. ISSN 20507496. doi: 10.1039/C6TA08699C.

A Appendix

A.1 Perovs-Dopant dataset

To construct a broad dataset on substitutional defects in perovskites, we begin by selecting a diverse set of perovskite materials and substituents to ensure extensive coverage of different chemical environments and structural variations. The base perovskite materials were queried from the Materials Project database (Jain et al., 2013), an oxide perovskite dataset (Castelli et al., 2012), and a halide perovskite dataset (Muzzillo et al., 2024). Based on the structures contained in these datasets, we randomly selected perovskites that are stable and have a DFT-calculated band gap ranging from 1 to 3 eV. For the substituent elements, we focused primarily on transition metals and nitrogen based on prior work related to doped semiconductor materials (Zhang et al., 2023; Al-Naggar et al., 2023; Rohib et al., 2023). These elements were chosen due to their diverse electronic configurations and their potential to introduce significant property modifications to the host perovskite material. The defective perovskite structures were then generated by substituting one atom in either the A or B site of a perovskite supercell with the selected substituent. We also included cases with vacancies on the A or B site.

The workflow for constructing the defect dataset was developed using Atomate2 (Ganose et al., 2025). CP2K was employed as the DFT code for all calculations (Kühne et al., 2020), and we used the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional (Perdew et al., 1996). The calculations were conducted using the Orbital Transformations method from the Quickstep code in CP2K (VandeVondele & Hutter, 2003), with the TZVP basis set and GTH pseudopotentials. The number of multi-grids was set to 5, the planewave cutoff for the finest level of multi-grid was 500 Ry, and the plane-wave cutoff of a reference grid was 80 Ry. Geometric relaxation was performed to obtain the optimized structures for the defect systems. During these simulations, the atomic positions were optimized with the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) to minimize the forces acting on the atoms to ensure the structure reached a stable state. The relaxation process was iterated until the forces on all atoms were reduced to below 0.02 eV/Å.

The Perovs-Dopants dataset contains 438 defective perovskite structures, and the element distribution is shown in Figure 3a. We performed a t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten & Hinton, 2008) analysis to help qualitatively analyze the difference in the chemical space coverage between the MPtrj dataset and the defect dataset from the model's perspective. The node features for 10,000 randomly selected systems from MPtrj training dataset and the entire Perovs-Dopants test set were extracted from the pretrained MACE-MP model. These 256-dimensional vector features represent the atomic neighborhood of each atom in a chemical system. We averaged the per-atom vectors within each system to obtain a system-level descriptor. t-SNE was then applied to reduce the dimensionality and visualize the distribution of the systems. As shown in Figure 3b, while there is some overlap suggesting shared features between the datasets, a significant portion of the defect dataset lies in the areas that are not covered by MPtrj. This observation confirms that the Perovs-Dopants dataset explores new chemical spaces, emphasizing the need for fine-tuning the pretrained MACE model to better adapt to these out-of-distribution data points.

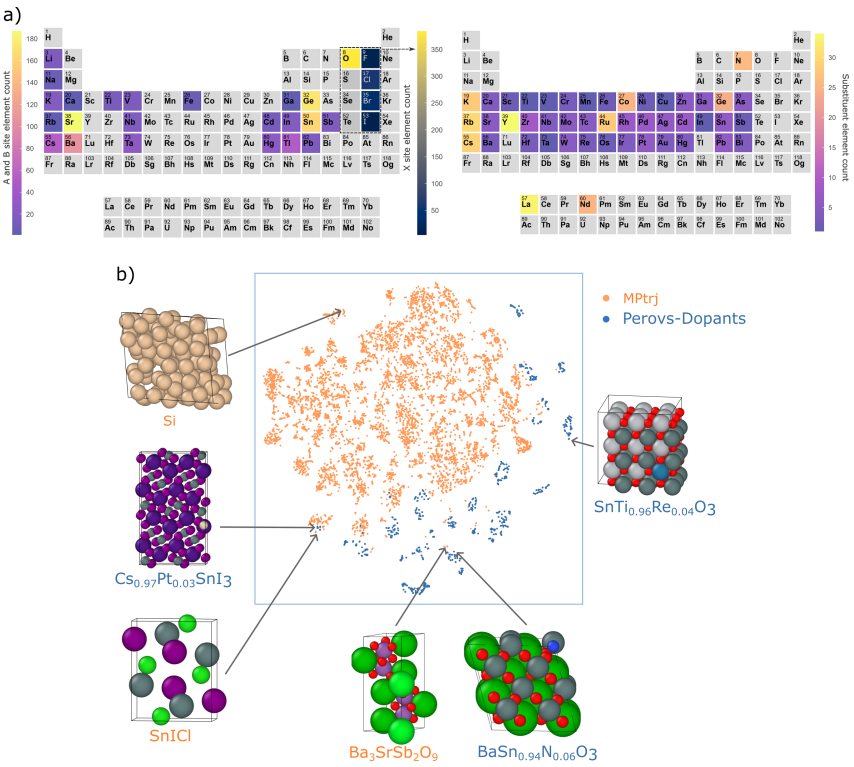

Figure 3: (a) Distribution of elements in the defect dataset. A and B site elements, and X site elements are shown in the left figure, and the substituent elements are shown in the right figure. (b) t-SNE plot comparing the chemical space covered by the MPtrj and Perovs-Dopants datasets.

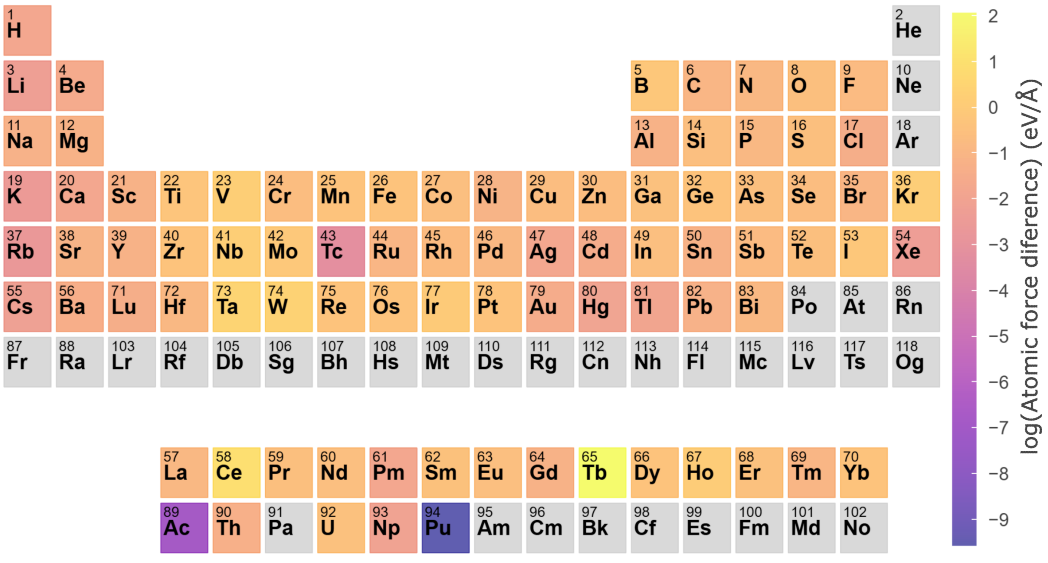

Figure A1: Average atomic force differences between CP2K and VASP for the MPtrj single-point calculations. Elements are color-coded to indicate the magnitude of force discrepancies.