Datasets for Benchmarking Machine Learning Models for Accelerated Search of Fast Ionic Conductors

Artem Dembitskiy*[a, b], **Innokentiy Humonen**[b], **Roman Eremin**[b], **Dmitry Aksyonov**[a], **Stanislav Fedotov**[a], **Semen Budennyy**[b]

[a] *Skolkovo Institute of Science and Technology, 121205 Moscow, Russian Federation* artem.dembitskiy@skoltech.ru

[b] *AIRI, 123112 Moscow, Russian Federation*

* Presenting author

1. Introduction

Fast ionic conductors are essential for advancing electrochemical devices, such as batteries, gas sensors, and ceramic membranes [1]. Traditional searches for these materials rely on costly high-throughput (HT) density functional theory (DFT) calculations across diverse chemical and structural spaces. To accelerate technological advancement in this area, HT schemes must be optimized by replacing DFT with accurate and fast surrogate models for rapid materials screening. Here, we benchmark existing machine learning (ML) models for HT search of Li-ion conductors.

2. Datasets

For the benchmarks, we introduce four datasets with calculated characteristics of Li-ion conductors. The descriptive statistics of each dataset is given in Table 1.

Table 1: The description of datasets for benchmarking ML models.

Dataset	Specs
nebDFT2k	target: migration barrier, geometry # of samples: 1,681
MPLiTrj	target: energy, forces, stress tensor # of samples: 929,066
BVEL13k	target: 1-3D percolation barrier # of samples: 12,807
nebBVSE122k	target: migration barrier # of samples: 122,421

The BVEL13k dataset is designed to benchmark graph neural network (GNN) models in predicting Li-ion 1, 2, and 3D percolation barriers corresponding to an energy threshold required for a mobile ion to percolate through the unit cell in 1, 2, or 3 dimensions, enabling macroscopic diffusion. This property is calculated using the bond valence site energy (BVSE) method utilizing tabulated empirical parameters [2]. The nebBVSE122k and nebDFT2k datasets are used to evaluate GNNs for predicting Li-ion migration barriers, defined as the relative height of the energy profile along the migration trajectory, using the BVSE and density functional theory (DFT) levels, respectively. Finally, the nebDFT2k and MPLiTrj datasets are employed to test the universal ML interatomic potentials (uMLIPs) for correct prediction of the geometry and energetics of minimum energy Li-ion migration pathways as predicted by DFT-based nudged elastic band (NEB) calculations.

3. Benchmark

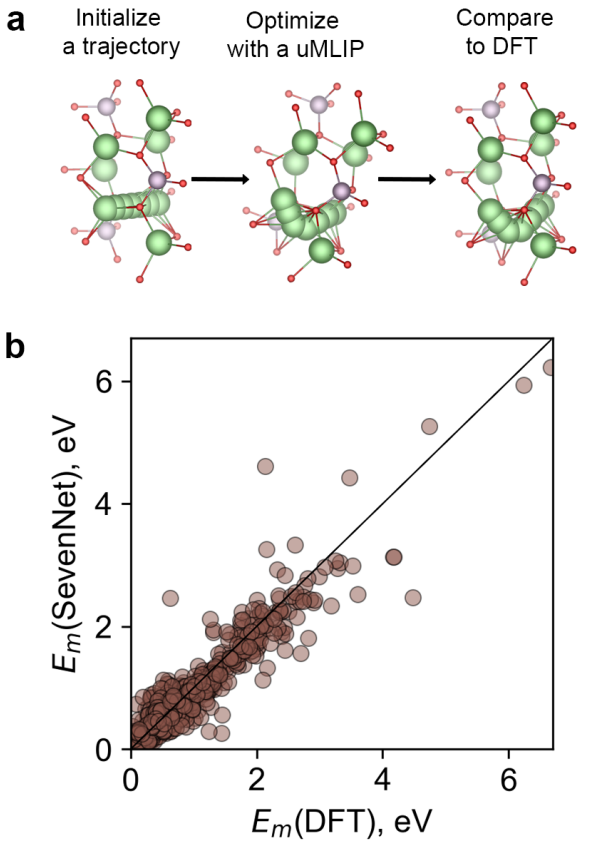

Fig. 1: a) A scheme for benchmarking uMLIPs on the nebDFT2k dataset, b) Parity plot for the predictions of the SevenNet model.

We show that the GNNs for the structure-to-property prediction of the BVSE-calculated 1-3D percolation and migration barriers achieve a reasonable accuracy, which is sufficient for identifying "fast" and "poor" ionic conductors for reducing candidate list in the HT scheme. Among the studied pre-trained uMLIPs, MACE-MP-0 [3], and SevenNet [4] possess the highest accuracy when solving the trajectory optimization task. An illustrative scheme for benchmarking uMLIPs and parity plots

for the SevenNet model predictions is given in Figure 1. Fine-tuning this uMLIP on a subsample of the MPLiTrj dataset, representing the crystal structure configurations obtained during DFT optimization of the trajectories, yields near-DFT accuracy (mean absolute error = 0.1 eV).

We discuss the practical application of the uMLIPs, and suggest methodology for integrating these steps into the HT screenings of fast ionic conductors. The datasets will be made available at `https://github.com/AIRI-Institute/LiTraj` after having published the manuscript related to the present study.

4. Related work

The classical ML and GNN models were previously utilized for predicting BVSE-calculated Li-ion (or other mobile specie) percolation barriers [5, 6]. Our models achieve comparable accuracy even in the face of a higher variance of the target property. As for the DFT part of the study, to the best of our knowledge, this is the first structured dataset with optimized Li-ion migration trajectories.

Acknowledgments

A.D., D.A., and S.F. acknowledge the financial support of Russian Science Foundation project No. 24-73-10204.

References

[1] Liwen Zhang, Bing He, Qian Zhao, Zheyi Zou, Shuting Chi, Penghui Mi, Anjiang Ye, Yajie Li, Da Wang, Maxim Avdeev, et al. A database of ionic transport characteristics for over 29 000 inorganic compounds. *Advanced Functional Materials*, 30(35):2003087, 2020.

[2] Stefan Adams and R Prasada Rao. High power lithium ion battery materials by computational design. *physica status solidi (a)*, 208(8):1746–1753, 2011.

[3] Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *arXiv preprint arXiv:2401.00096*, 2023.

[4] Yutack Park, Jaesun Kim, Seungwoo Hwang, and Seungwu Han. Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations. *Journal of chemical theory and computation*, 20(11):4857–4868, 2024.

[5] Masanobu Nakayama, Kenta Kanamori, Koki Nakano, Randy Jalem, Ichiro Takeuchi, and Hisatsugu Yamasaki. Data-driven materials exploration for li-ion conductive ceramics by exhaustive and informatics-aided computations. *The Chemical Record*, 19(4):771–778, 2019.

[6] Zhilong Wang, Yanqiang Han, Junfei Cai, An Chen, and Jinjin Li. An end-to-end artificial intelligence platform enables real-time assessment of superionic conductors. *SmartMat*, 4(6):e1183, 2023.