Dual Benchmarking of Universal Machine Learning Interatomic Potentials Against DFT and Experimental EXAFS Data

Pjotrs Žguns, **Inga Pudza**, **Alexei Kuzmin**

Institute of Solid State Physics, University of Latvia, Kengaraga street 8, LV-1063, Riga, Latvia,

pjotrs.zguns@cfi.lu.lv, *a.kuzmin@cfi.lu.lv*

1. Introduction

Universal machine learning interatomic potentials (uMLIPs) [1–5] offer *ab initio* accuracy in energy and force calculations at minimal computational cost, making them invaluable for materials modeling. Although pre-trained on extensive *ab initio* datasets, uMLIPs often exhibit systematic softening [6] and require fine-tuning using compound-specific data for optimal performance. Key questions are how many *ab initio* data points are needed for fine-tuning, and whether uMLIPs can achieve high accuracy consistent with experimental benchmarks.

2. Approach

We fine-tune uMLIPs using Density Functional Theory (DFT) data, progressively increasing the dataset size, and evaluate uMLIP performance against both DFT and experimental benchmarks. For experimental validation, we derive theoretical EXAFS spectra (extended X-ray absorption fine structure) from molecular dynamics (MD) simulations [7]. EXAFS, being sensitive to thermal fluctuations in bond lengths and angles, provides a robust benchmark for validating uMLIP accuracy. We first apply this approach to layered 2H$_c$-WS$_2$ and 2H$_c$-MoS$_2$ dichalcogenides (Fig. 1), which offer a balance between high symmetry and structural complexity due to their van der Waals gaps, making them ideal model systems.

3. Results

Our results reveal a clear softening of the original uMLIP [2], as evidenced by both DFT and EXAFS data, particularly, the Fourier Transform (FT) of the EXAFS spectra derived from uMLIP–MD (Fig. 2). EXAFS effectively detects uMLIP softening, as seen in the underestimation of the FT amplitudes. Fine-tuning with compound-specific *ab initio* data corrects this systematic softening, resulting in excellent agreement between MD-derived and experimental EXAFS spectra.

Using one to ten DFT frames for fine-tuning, as suggested in Ref. [6], improves the original uMLIP. However, achieving accurate EXAFS spectra and force predictions comparable to DFT requires approximately fifty structures in our case (Fig. 3). This difference is likely due to the increased complexity of the structure (vdW), which demands a larger DFT dataset. Thus, benchmarking against both DFT and experimental EXAFS data strengthens confidence in uMLIP reliability and helps determine the optimal dataset size for fine-tuning.

4. Outlook

In this ongoing work, we aim to extend our approach to complex materials, such as oxides with Jahn–Teller distortions, and assess uMLIP performance in modeling their thermal disorder and structural relaxations. Specifically, we will explore how fine-tuning with advanced exchange-correlation functionals improves agreement with the experimental EXAFS data, advancing uMLIP validation and bringing us closer to achieving experimental-level accuracy for complex materials.

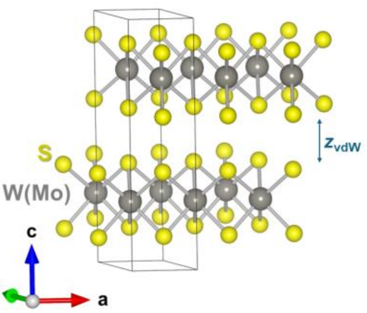

Fig. 1: Crystallographic structure of layered 2H$_c$-W(Mo)S$_2$ dichalcogenides [8]. The vdW gap, z_{vdW}, is about 3.02 Å.

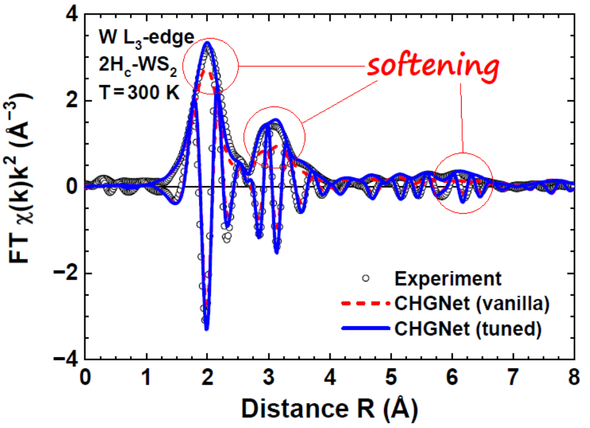

Fig. 2: Fourier transforms (FT) of the experimental and MD-derived EXAFS spectra of W L$_3$-edge ($\chi(k)k^2$), highlighting the evident softening of the vanilla CHGNet uMLIP [2].

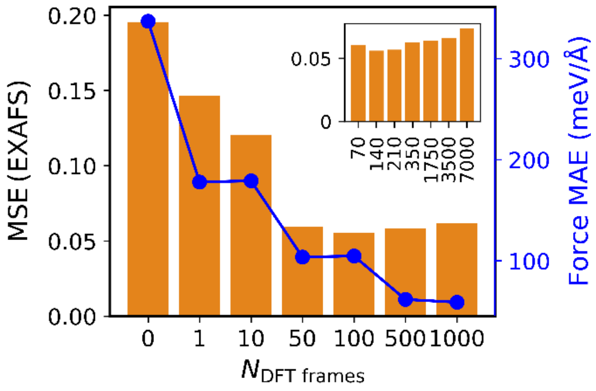

Fig. 3: Performance of the fine-tuned uMLIP as a function of DFT frames. Left axis: Mean squared error (MSE) relative to the experimental EXAFS spectrum. Right axis: Mean absolute error (MAE) of forces compared to DFT.

Dual Benchmarking of Universal Machine Learning Interatomic Potentials Against DFT and Experimental EXAFS Data

Pjotrs Žguns, **Inga Pudza**, **Alexei Kuzmin**

Institute of Solid State Physics, University of Latvia, Kengaraga street 8, LV-1063, Riga, Latvia,

pjotrs.zguns@cfi.lu.lv, a.kuzmin@cfi.lu.lv

Acknowledgments

P.Ž. and I.P. thank the support of the Latvian Council of Science under, respectively, projects 1.1.1.9/LZP/1/24/016 (agreement no. 1.1.1.9/1/24/I/001) and LZP-2023/1-0528.

References

[1] C. Chen, S.P. Ong, A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* 2, 718–728, 2022.

[2] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C.J. Bartel, G. Ceder, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling, *Nat. Mach. Intell.* 5, 1031–1041, 2023.

[3] A. Merchant, S. Batzner, S.S. Schoenholz, M. Aykol, G. Cheon, E.D. Cubuk, Scaling deep learning for materials discovery, *Nature* 624, 80–85, 2023.

[4] I. Batatia, P. Benner, Y. Chiang, A.M. Elena, D.P. Kovács, J. Riebesell, X.R Advincula, M. Asta, W.J. Baldwin, et al. A foundation model for atomistic materials chemistry, arXiv:2401.00096, 2023.

[5] H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen, C. Zeni, et al. Mattersim: A deep learning atomistic model across elements, temperatures and pressures, arXiv:2405.04967, 2024.

[6] B. Deng, Y. Choi, P. Zhong, J. Riebesell, S. Anand, Z. Li, K. Jun, K.A. Persson, G. Ceder, Systematic softening in universal machine learning interatomic potentials, *npj Comput. Mater.* 11, 9, 2025.

[7] A. Kuzmin and J. Chaboy, EXAFS and XANES analysis of oxides at the nanoscale, *IUCrJ* 1, 571-589, 2014.

[8] W. Schutte, J. De Boer, F. Jellinek, Crystal structures of tungsten disulfide and diselenide. *J. Solid State Chem.* 70, 207–209, 1987.