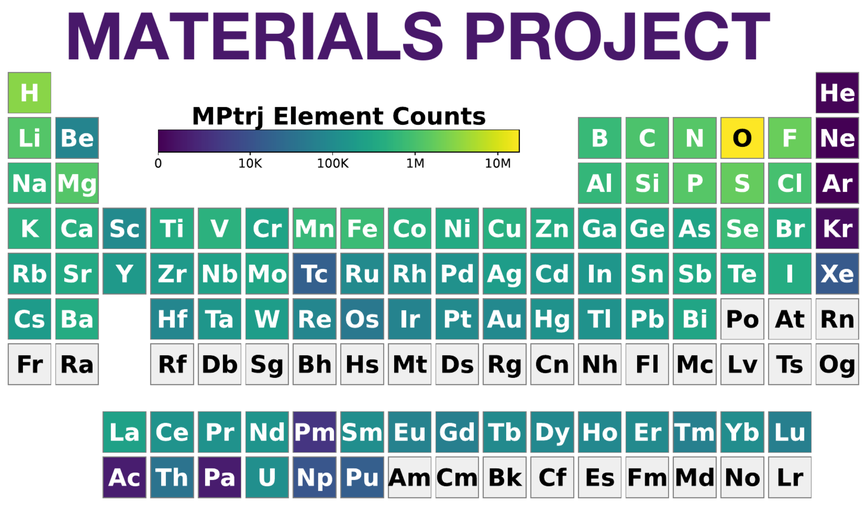
Benchmarking universal machine learning potentials on defects in bcc metals

Petr Grigorev[1], Thomas D. Swinburne[2] and James R. Kermode[3]

[1]CNRS, INSA Lyon, Universite Claude Bernard Lyon 1, MATEIS, UMR5510, 69621 Villeurbanne, France
[2]Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA
[3]Warwick Centre for Predictive Modelling, School of Engineering, University of Warwick, Coventry, United Kingdom

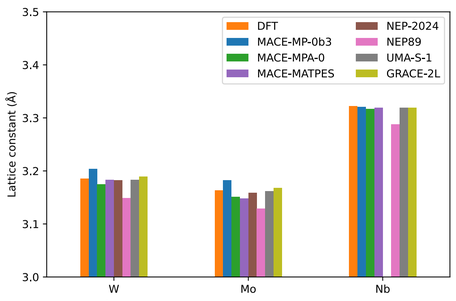
Introduction

Recent advancement of machine learning force fields allowed the development of universal models capable of describing large number of elements and chemical environments. Here we test a few universal models with different architectures and trained on different databases: MACE-MP0b3, MACE-MPA-0, MACE-MATPES: three versions of MACE potential [1], neuroevolution potentias NEP-2024 [2] and NEP89 [3], Universal Model for Atoms UMA-S- 1 [4], Graph Atomic Cluster Expansion model GRACE-2L: [5]. We look at properties of point and extended defects in three BCC metals: W, Mo, Nb and comare the results with available DFT and QM/MM data.

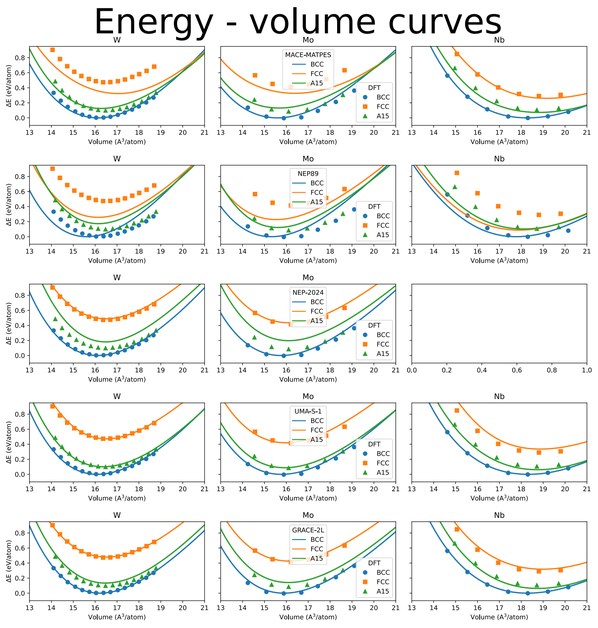


Elastic properties and point defects

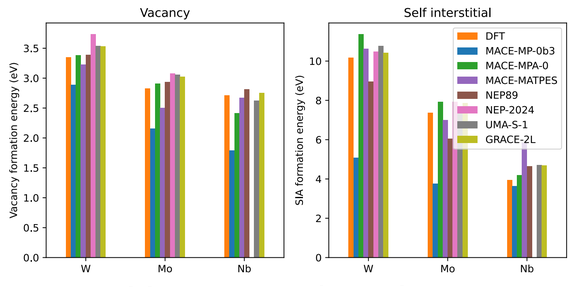
Lattice constants



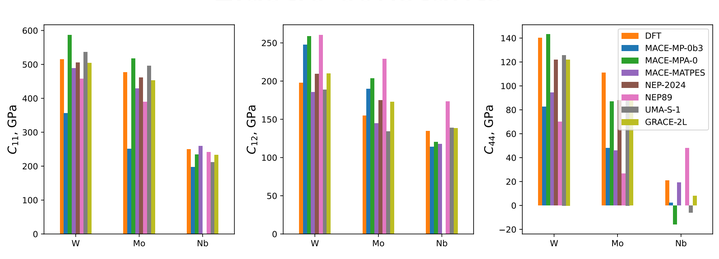
Energy - volume curves



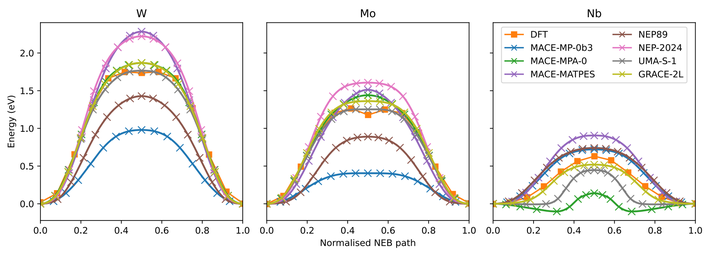
Point defect formation



Elastic constants



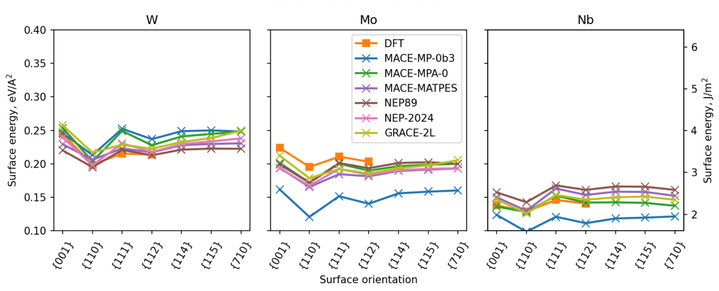
Vacancy migration



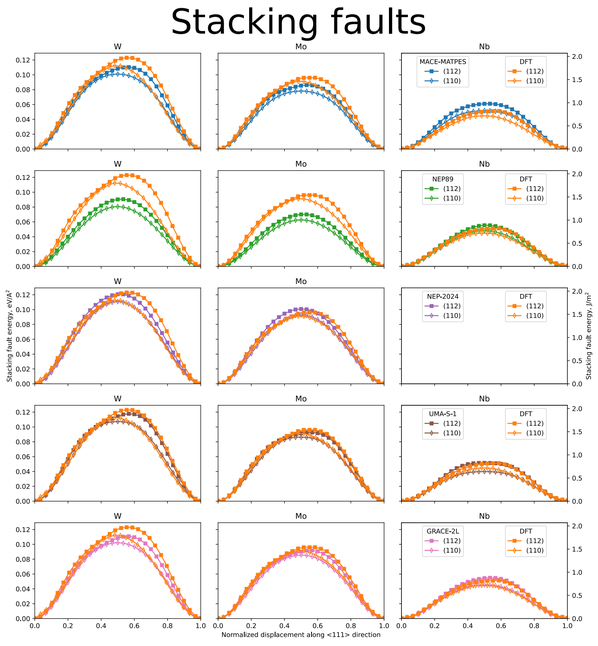
*Details of the calculation methods and sources of reference ab initio data can be found in [1], section A.13: Point and extended defects in BCC metals and supplementary materials of [6]..

Extended defects

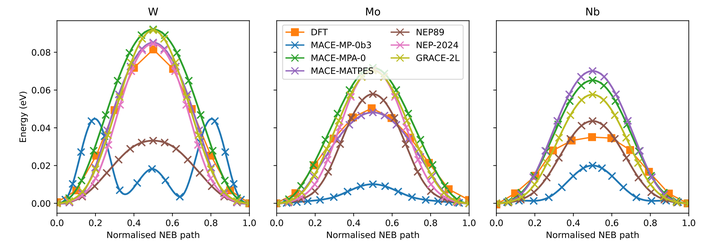
Free surfaces



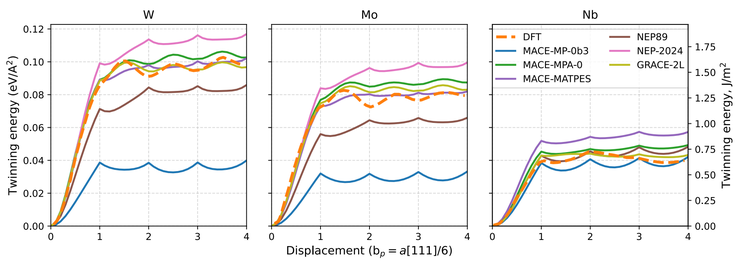
Stacking faults



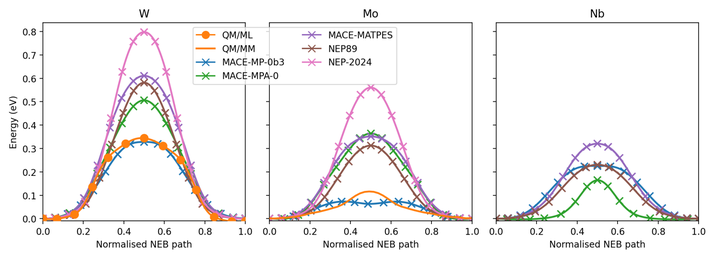
Screw 1/2<111>{110}dislocation



Twinning GSF {110}



Edge <100>{001} dislocation



*UMA-S-1 model has limited capacity to treat free boundary conditions and free surfaces needed for cells containing extended defects, thus some results are missing.

Discussion

Testing the universal models for W, Nb, Mo revealed:

- Systematic softening for MACE-MP0b3 model
- Models trained on *larger* databases show better performance
- Can we use the spread in the results with different models for uncertainty quantification?

Future work:

- Testing impurity properties to tackle desctiption of *chemistry*
- Fine tuning on *optimal* DFT data
- Using fine tuned models in a hybrid QM/ML scheme in a QM region

References:

[1] I. Batatia et al, A foundation model for atomistic materials chemistry, J. Chem. Phys. 163, 184110 (2025)
[2] Song, K. et al. General-purpose machine-learned potential for 16 elemental metals and their alloys. Nature Communications 15, 10208 (2024).
[3] Liang, T. et al. NEP89: Universal neuroevolution potential for inorganic and organic materials across 89 elements. (2025). arXiv:2504.21286
[4] Wood, B. M. et al. UMA: A family of universal models for atoms. (2025). arXiv:2506.23971
[5] Lysogorskiy, Y., Bochka[rev, A. & Drautz, R. Graph atomic cluster expansion for foundational machine learning interatomic potentials. (2025). arXiv:2508.17936
[6] Starikov, S., et. al Angular-dependent interatomic potential for large-scale simulation of bcc and hcp multi-component refractory alloys. (2026). Computational Materials Science, 262, 114369.

For updates visit pgrigorev.github.io and tomswinburne.github.io