Digital experiments for molecular passivation of hybrid perovskite surfaces

Laura-Bianca Paşca[1], Henry J. Snaith[2],
Volker L. Deringer[1]
[1]Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford, UK
[2]Department of Physics, University of Oxford, Oxford, UK

laura-bianca.pasca@queens.ox.ac.uk

Machine-learned interatomic potentials (MLIPs) can now be used to study chemically complex material systems, such as the passivation of undercoordinated atoms at hybrid perovskite surfaces. We investigate whether carefully designed, material-specific datasets can guide the continual fine-tuning (CFT) of MLIPs for challenging modelling tasks involving heterogeneous and interfacial systems.

CFT approaches are established in research on large language models [1] and have recently been proposed for MLIPs, with a focus on optimising training protocols [2]. However, the role which curated, material-specific datasets may play in the process of fine-tuning foundation models (FMs) for complex chemistry applications has not been systematically explored. Such datasets may complement large pre-training data by supplying highly-specific structural motifs, such as surface terminations or defects, which are often sparsely represented and remain difficult for general-purpose models to describe accurately [3]. We examine the extent to which such curated datasets provide additional benefits when applied to broadly transferable MLIPs whose accuracy may already approach a plateau for many applications.

Our approach proceeds in two stages: (i) we construct a diverse, curated dataset for hybrid perovskite systems (*hy-PV-26*), spanning crystalline, defective, surface, and disordered structures; and (ii) we apply this dataset within a CFT protocol to generate an MLIP specialised for the passivation of hybrid perovskite surfaces by an amino-silane molecular passivator. Molecular surface passivation of hybrid perovskites is widely studied experimentally as a strategy to improve the operational stability of perovskite solar cells [4], yet its chemical and structural complexity poses a significant challenge for conventional atomistic simulations. It is therefore chosen here as a representative test case of a heterogeneous, multicomponent system.

We compare four MLIP fitting strategies: a directly trained model, the zero-shot FM (without fine-tuning), an FM fine-tuned on the specialised passivation dataset, and a CFT approach in which the FM is first adapted to *hy-PV-26* and then to the passivation system. To probe the influence of the pre-training data, we compare models derived from different FMs, *viz.* MACE-MP-0b3 and MACE-MH-1 [5].

By combining numerical validation with chemically meaningful tests of surface–adsorbate interactions, we show that the CFT approach substantially improves model performance when starting from an FM whose pre-training data underrepresent surfaces and organic molecular chemistry. In contrast, for broadly transferable FMs that already approach a limit in general accuracy, CFT provides only marginal benefits over conventional fine-tuning. Finally, we use the best-performing MLIP to simulate amino-silane passivation at hybrid perovskite surfaces, yielding atomistic insights that complement experimental observations [4]: the MLIP can be used to relax the surface—adsorbate system, revealing the binding motifs that underpin the effective performance of specific amino-silane passivators. Moreover, the MLIP enables large-scale simulations offering a route to explore collective effects and surface dynamics.

Our work outlines a practical strategy for integrating curated, material-specific datasets into MLIP fine-tuning workflows to enable reliable simulations of complex chemical interfaces. At the same time, it introduces the *hy-PV-26* dataset as a resource for modelling hybrid perovskite materials.

References

[1] Zhang, Zihan, et al. ‘CITB: A Benchmark for Continual Instruction Tuning’, arXiv (2023), arxiv:2310.14510.

[2] Kim, Jaesun, et al. ‘Data-Efficient Multifidelity Training for High-Fidelity Machine Learning Interatomic Potentials’. *Journal of the American Chemical Society* (2025) 147(1): 1042–54.

[3] Kempen, Luuk H. E., et al. ‘How Accurate Are Foundational Machine Learning Interatomic Potentials for Heterogeneous Catalysis?’, arxiv (2025), arXiv:2512.16702.

[4] Lin, Yen-Hung, et al. ‘Bandgap-Universal Passivation Enables Stable Perovskite Solar Cells with Low Photovoltage Loss’. *Science* (2024) 384: 767–75.

[5] Batatia, Ilyes, et al. ‘Cross Learning between Electronic Structure Theories for Unifying Molecular, Surface, and Inorganic Crystal Foundation Force Fields’, arxiv (2025), arXiv:2510.25380

Figures

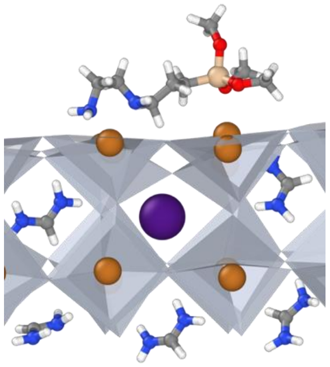

Figure 1. Snapshot from an MLIP-driven molecular dynamics simulation, showing the passivation of a hybrid halide perovskite surface by N-(2-aminoethyl)-3-aminopropyltrimethoxysilane (AEAPTMS).