



Motivation

- Grand Canonical Monte Carlo simulations (GCMC) enable the discovery of structure-property relationships and identification of top-performing candidates prior to synthesis.
- Traditional high-throughput screening relies on classical force fields (UFF, DREIDING), often *unable* to capture subtle electronic effects or strong interactions at open metal sites.
- While *ab initio* methods are highly accurate, they are computationally too expensive.

MLIPs as a potential middle ground

Benchmark

Framework: MLIP-MC^[1]

- Open-source Python framework built on the Atomic Simulation Environment (ASE).
- GCMC with three MLIP backends (FAIRChem, MACE-Torch, Orbital).
- Metropolis algorithm (molecular moves) + Peng-Robinson equation of state.

Models:

Name	Class	Datasets
ORB-v3-inf-consv-omat	GNS (Graph Network Simulator)	OMat
MACE-MP-0a	GNN (Graph Neural Network)	MPtrj
MACE-MP-0b3	GNN	MPtrj
MACE-MPA-0	GNN	MPtrj + sAlex
fairchem ODAC23	GT (Graph Transformer)	ODAC23 (single adsorbates)
fairchem ODAC25 (full)	GT	ODAC25 (multiple adsorbates)

Reference data: finetuned model (Mg-MOF-74) provided by Goeminne et al.^[2]

Systems:

- CO₂ x 3 ZIFs (each a different adsorption mechanism).
- ZIF-8 (physisorption) + Mg-MOF-74 (open metal sites) + ZIF-4 (dense framework).

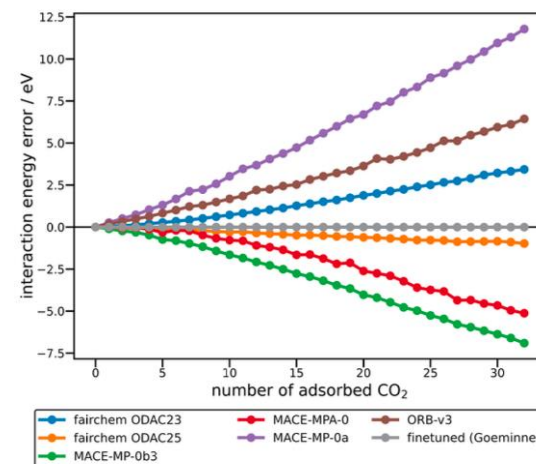
Simulation Details:

- Widom insertion simulations, 100k steps → isosteric heat of adsorption at infinite dilution.
- GCMC simulations (100k + 1,000k steps, 273K or 298K, 14 pressure points) → isosteric heat of adsorption.

Result: rather than exhibiting framework-specific errors, the models display a systematic bias

Training Data:

- ODAC family showed the best agreement with reference data, ODAC25 performs best.
- ODAC23 fails at higher pressures.



Error in MLIP-predicted interaction energy of CO₂ molecules with ZIF-8 relative to the reference DFT data.^[2]

Image reproduced from Ref. [1] under CC BY-NC-ND license.

Interaction Energies (IE):

- The models either systematically underestimate (ODAC, MACE-MP-0a and ORB-v3) or overestimate (MACE-MPA-0, MACE-MP-0b3) the IE.
- Error in IE grows approximately linearly with increasing CO₂ uptake.
- Both low (CO₂-framework interactions) and high (CO₂-CO₂ interactions) uptake regimes are described with comparable inaccuracy.

Computational efficiency:

- Fastest universal model (MACE-MP-0a) ≈ finetuned Mg-MOF-74 model.
- ODAC25 is 14 times slower.

Existing “universal” MLIPs are unable to reliably describe gas adsorption across diverse systems

Outlook and solutions

- Finetuned models are required to capture the correct thermodynamic behaviour of adsorbed gases.
 - Incorporating more adsorption structures in training datasets.
 - Cover systems beyond one adsorbed CO₂
- importance of high-quality and diverse training data for future MLIPs.

References

[1] arXiv:2602.13725, 2026. [2] *J. Chem. Theory Comput.* **2023**, *19*, 6313.