CLUSTER-BASED LEARNING TO DESCRIBE DISORDERED METAL-ORGANIC FRAMEWORKS AT THE MESOSCALE

P. Dobbelaere, S. Vandenhaute and V. Van Speybroeck

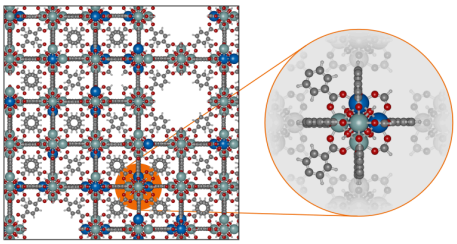
Center for Molecular Modeling, Ghent University, Technologiepark-Zwijnaarde 46, 9052 Zwijnaarde, Belgium

e-mail: pidobbel.dobbelaere@ugent.be

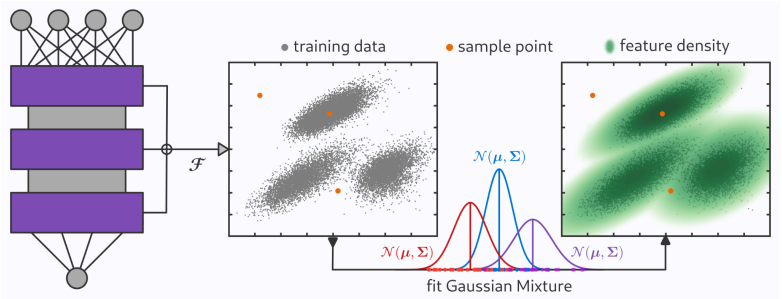
Keywords: *metal-organic frameworks, spatial disorder, machine learning potentials*

Metal-organic frameworks (MOFs) are a spectacularly versatile class of nanoporous materials, with potential applications in (heterogeneous) catalysis, gas sorption, nanosensing and energy storage.[1] By incorporating local spatial defects into their atomic structure, MOFs can be fine-tuned to incorporate desired chemical functionalities. These targeted modifications require a fundamental understanding of the microscopic structure of a material and its macroscopic emergent properties. Successful defect engineering warrants detailed research.

In MOFs, spatial disorder expresses itself across multiple length scales.[2] At the nanoscale, we find point defects such as missing linkers, metal ion substitutions and slightly larger node defects. Zooming out, synthesised crystallites can contain mesopores, regions of phase coexistence and eventually finite crystal boundaries. However, existing computational models often struggle to comprehensively describe disordered phenomena. *Ab initio* methods are restricted by the quantum scaling limit at extended length scales (> 10 nm), and force fields are hampered by a lack of functional flexibility. The recent addition of machine learning potentials (MLPs) to the modelling toolbox promises the combination of high accuracy and efficient evaluation, on the condition that we can collect representative training data. This remains prohibitively expensive, bringing us back to square one.



Here, we introduce a cluster-based learning methodology to develop state-of-the-art MLPs for spatially disordered MOFs. Following a divide-and-conquer approach, we deconstruct large supercells into smaller molecular fragments to circumvent *ab initio* limitations (see figure above). By assembling an appropriate set of finite clusters, it is possible to describe every (local) interaction of the original supercell. Our approach consists of two main components: (i) scan molecular systems and identify out-of-dataset chemical environments using a data-driven uncertainty model exploiting MLP feature space (see figure below), and (ii) encapsulate the corresponding atoms into clusters that replicate the supercell bulk environments. This scheme is embedded into an automated active learning workflow. Our recent preprint shows that cluster-based learning delivers models capable of accurately describing spatial defects in mesoscopic systems with over twenty thousand atoms.[3] Using this methodology, we can simulate (almost) arbitrarily disordered frameworks; a valuable step towards the study of realistic crystallites and an understanding of the role of disorder on MOF behaviour.



References

[1] H. Furukawa, K.E. Cordova et al. *Science*. **2013**, 341, 1230444

[2] A. K. Cheetham, T. D. Bennet et al. *Dalton Trans*. **2016**, 45, 4113-4126

[3] P. Dobbelaere, S. Vandenhaute et al. *arxiv*. **2025**, arXiv:2504.03881 (Preprint)