

Data-driven characterisation of 2D polymers

Author:

Aleksander Jakub Szewczyk

Co-authors:

Leonardo Medrano Sandonas, Gianaurelio Cuniberti

Covalent organic frameworks (COFs) are a class of advanced porous materials that can be precisely engineered for diverse applications, including catalysis, flexible electronics, gas separation, and sensors. A key advantage of COFs is their modular construction from discrete molecular building blocks, which in principle allows vast regions of chemical space to be explored combinatorially. However, the scale of this space makes exhaustive experimental exploration infeasible, necessitating computational approaches to fully optimise the material properties.

We are developing a high-throughput computational pipeline for data-driven design and inverse design of 2D COFs at scale. Candidate structures are generated combinatorially from a curated monomer library, targeting a search space of approximately 10,000 frameworks. The pipeline applies a multi-stage screening workflow: inexpensive structural stability filters act as early sieves, followed by evaluation of thermal and mechanical properties; electronic properties are reserved for structures that pass these earlier stages. Mapping this design space is intended to uncover structure–property relationships and identify COFs with optimised target properties — informing inverse design strategies that work backwards from desired functionality to candidate structures. We further aim to extend the screening to chiroptical properties, which have not yet been computationally explored in COFs and present a methodological challenge.

Machine learning interatomic potentials (MLIPs) based on the MACE architecture underpin the computationally demanding stages of this workflow. Building on the QCOF models introduced in our recent preprint — trained on non-equilibrium conformations of 2D COFs and benchmarked against DFT and general-purpose MACE potentials — we are extending the coverage of our model to a broader chemical space, including elements such as oxygen and sulphur, potentially through active learning approaches integrated within the screening pipeline itself.