

Graph models and fine-tuned machine learning potentials for microkinetic analyses in heterogeneous catalysis

Raffaele Cheula¹, John Kitchin², and Mie Andersen¹

¹ Center for Interstellar Catalysis, Department of Physics and Astronomy, Aarhus University, 8000 Aarhus, Denmark

² Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States

mie@phys.au.dk

Heterogeneous catalysis is nowadays expected to solve the challenges of our society related to the diversification of energy sources and the reduction of greenhouse gases. The processes of CO₂ hydrogenation have the potential to sustainably produce chemicals and fuels from green hydrogen and waste CO₂ [1]. To make such processes economically advantageous, R&D in catalysis relies on the discovery and optimization of catalytic materials based on experimental testing and computational analyses.

The *in-silico* modeling and design of catalyst materials must tackle the extreme complexity of chemical reactions at catalytic surfaces. This makes the direct application of density-functional theory (DFT) computationally prohibitive, especially when targeting a wide combinatory space of elements of the periodic table. This problem can be addressed with machine learning (ML) techniques, which can significantly reduce the number of DFT calculations required.

In this contribution, we combine DFT and ML models to investigate CO₂ hydrogenation reactions across a broad range of catalyst materials, including metals, single-atom alloys [2], and doped ZrO₂ surfaces [3]. As illustrated in Figure 1, we explore different levels of modeling, from accurate but computationally demanding DFT calculations to increasingly efficient ML-based strategies. These include descriptor-based models such as WWL-GPR (a graph-based Gaussian Process Regression framework [4]), as well as universal ML potentials (MLPs) from the Open Catalyst Project [5], which we apply to predict energies of reaction intermediates and transition states (TS) of reaction mechanisms.

Beyond predicting energies with ML models, we also apply active learning strategies to efficiently achieve DFT accuracy. For example, we combine single-ended TS-search methods with MLPs fine-tuning to estimate DFT activation barriers at a fraction of the usual computational cost. As illustrated in Figure 2, a pretrained MLP is first used to locate a TS at low computational cost. A DFT single-point calculation is then performed on the MLP-optimized structure, and the resulting forces are used to refine the MLP. This procedure is repeated until the DFT forces fall below a chosen threshold. This sequential refinement improves the accuracy of the MLP in the vicinity of

saddle points, enabling identification of TSs with DFT accuracy while drastically reducing the number of expensive DFT calculations.

Finally, we apply mean field microkinetic modeling to calculate the catalytic activity of the catalyst materials, accounting for the contribution of the different active sites of the catalyst surfaces. The application of our frameworks to CO₂ hydrogenation reactions allows us to rationalize how reaction mechanisms and catalytic performances (i.e., activity and selectivity) change with the catalyst composition, paving the way toward the design and nano-engineering of catalyst materials [2, 3].

References

- [1] M. I. Alam, R. Cheula, G. Moroni, L. Nardi, M. Maestri, *Cat. Sci. Tech.* 11, 20, 6601–6629 (2021).
- [2] R. Cheula and M. Andersen, *ACS Catalysis* 15, 13, 11377–11388 (2025).
- [3] R. Cheula, T. A. M. Q. Tran, and M. Andersen, *ACS Catalysis* 14, 13126–13135 (2024).
- [4] W. Xu, K. Reuter, M. Andersen, *Nature Computational Science* 2, 443–450 (2022).
- [5] L. Chanussot, et al., *ACS Catalysis*, 11, 10, 6059–6072 (2021).

Figures

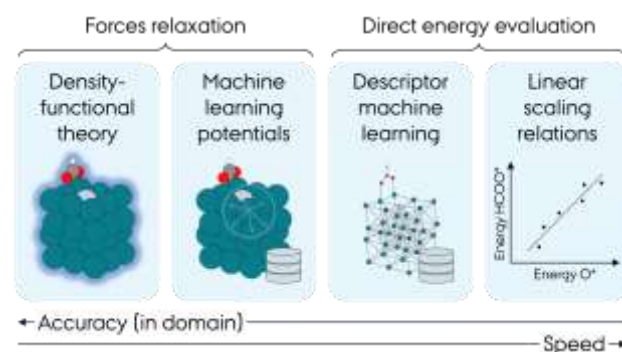


Figure 1. Schematic overview of the computational strategies employed in this work.

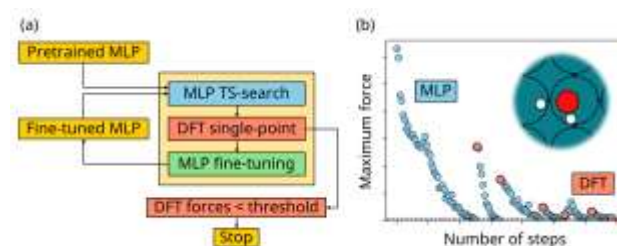


Figure 2. Active-learning refinement of machine-learning potentials for TS searches. (a) Workflow of the iterative procedure. (b) Evolution of the MLP and DFT calculated maximum forces during the active learning process, for the reaction of H₂O* dissociation on Rh(100).