AI-Assisted Reaction Network and Catalyst Exploration for CO_2 Conversion

Mikhail V. Polynski [1]

[1] *Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585*

Correspondence to: polynskimikhail@gmail.com (M.V.P.)

1. Introduction

Accurate prediction of catalytic activity is often limited by the tradeoff between model realism and the accuracy of quantum chemical calculations. Realistic nanocatalyst models often should capture distributions of catalytically relevant site ensembles rather than ideal low-index slabs, allow for condition- and composition-driven surface restructuring [1], account for co-adsorption effects that shift step energetics, and frequently include extended reaction networks rather than a small, hand-picked subset of steps. This is particularly critical for CO_2 conversion: for example, CO_2 hydrogenation proceeds via densely connected networks with competing hydroge-nation or protonation steps, C–C coupling, and C–O cleavage, often yielding multiple products under similar conditions [2]. Indeed, recent works highlight that reliable mechanistic insight and prediction of activity and selectivity demand analysis of large reaction networks [3], which often is hardly tractable.

However, recent advances in AI have created two complementary opportunities [4]. First, agentic AI can accelerate the scientific software engineering needed to build and maintain the underlying computational workflows, from transition state search pipelines to automated data handling and reaction network analysis. Second, machine learning can reduce the cost of activation barrier prediction, once an appropriate mechanistic representation has been established. Here we present this combined strategy for CO_2 conversion to C1-2 oxygenates on metal nanoparticles, in which AI is used as an accelerator of both workflow construction and ML model-based inference.

2. Results and Discussion

We introduced a rational ansatz for constructing catalytic reaction networks on nanoparticle models and applied it to CO_2 hydrogenation to C1-2 oxygenates on Cu and Pd nanoparticles, using an integrated active site model in which terrace and edge regions can act cooperatively rather than as isolated idealized sites [5]. Transition state structures were generated through a rapid semiempirical-to-DFT workflow, which made it possible to explore a mechanistically rich network without brute force DFT sampling of every possible step. The resulting Cu network explained several key features of Cu reactivity. CO_2 activation preferentially proceeds through the formate pathway, leading to the formation of methanol on pure Cu. The same network also clarifies why Cu intrinsically disfavors methane. Within the CH_n (n = 1-4) set of intermediates, the barriers for CH* to CH_2*, CH_2* to CH_3*, and CH_3* to CH_4* rise from 0.29 to 0.63 to 1.15 eV, whereas CH_2* can instead undergo kinetically facile C-C coupling with CH_2O*, CO*, and CO_2 with barriers of 0.18, 0.77, and 0.41 eV, respectively. This leads to two key mechanistic pathways: a CH_2-CH_2O-mediated route toward ethanol precursors and a CH_2-CO_2 route toward acetate-type intermediates, both of which help explain why Cu can favor C2 oxygenate-forming C-C chemistry once CH_2* is reached even though entry into the CH_x pool itself remains difficult. However, the introduction of dopants or co-catalysts facilitating the CH_2* can shift Cu-containing catalytic systems towards C2 oxygenate production. The underlying workflow has been released in our GitHub repository [6].

To move beyond Cu and Pd and avoid prohibitive DFT sampling for every barrier on every metal, we then generalized the study to fcc-Co, Ni, Rh, Pd, Pt, Au, and Cu nanoparticles using a transition-state-aware neural network model for barrier prediction [7]. The extended reaction network spans over 40 elementary steps per metal and includes the key C1 and C2 oxygenate pools, while the ML dataset was assembled from DFT-calculated forward and reverse activation barriers. This analysis showed that conventional or broadly pooled BEP relations are too crude for the reaction network under consideration. For example, C-O cleavage and C-C coupling BEP relationships showed RMSE values of about 0.39 and 0.40 eV, whereas selected single-TS-family or more tightly defined step classes can lead to accurate BEP relationships, with RMSE values as low as 0.06-0.09 eV. These trends motivated us to train a neural network model of original CNN-based architecture with an attention mechanism that explicitly encodes metal identity, transition state class, reaction free energy, and reactant/product structure

through Coulomb-matrix-derived representations. In this low-data regime, the neural network substantially outperformed a generally linear BEP-like baseline. The latter gave a 5-fold cross-validation RMSE/MAE of 0.42/0.35 eV, whereas the final neural-network architecture reached average 5-fold CV values of about 0.30/0.23 eV for forward barriers and 0.33/0.25 eV for reverse barriers, with statistically meaningful gains from the inclusion of the attention block. In practical terms, this reduced the number of expensive DFT transition-state searches needed to complete the catalytic networks while capturing the complexity of the underlying reaction network.

For activity and selectivity analysis, we reformulated PATHFINDER for catalytic reaction networks and coupled it to the energetic span model [8,9]. In our implementation, each metal-specific network was represented as a digraph with 42 nodes, where nodes corresponded to surface or gas-phase intermediates and edges carried activation barriers, reaction free energies, and explicit requirements for co-reactants such as H*, CO*, CH_2O*, or CO_2. Edges had Eyring-like kinetic weights with iteratively updated compound costs, so the algorithm evaluated kinetically plausible pathways that are consistent with reagent availability under reaction conditions.

We then used the lowest-cost candidate cycles as mechanistic inputs for energetic-span-based TOF estimation at 523.15 K with $p(H_2)$ = 30 bar and $p(CO_2)$ = 10 bar. This framework enabled rapid construction of Sabatier volcano plots for methane, ethanol, and acetic acid using CO and OH binding energies as descriptors and used them to identify the characteristics of the catalyst optimal for ethanol synthesis. The volcanoes for ethanol and acetic acid were nearly coincident, indicating a common Sabatier optimum for C2 oxygenate production. This similarity arised because the dominant pathways to both products shared closely related TOF-determining states, particularly steps involving CH_x formation. In contrast, the methane Sabatier volcano was shifted mainly toward weaker CO binding at the optimum, which distinguishes the catalyst requirements for methanation from those for C2 oxygenate synthesis. In summary, the combination of graph-based pathway selection and energetic span analysis provided a mechanistically grounded route from extended reaction networks to the target binding properties of Sabatier-optimal catalysts.

The key practical part of this work was the use of agentic AI systems, especially ChatGPT and Ardor Cerebrum [10], under close human supervision, for writing and refactoring chemistry-relevant code, automating parts of data extraction and structuring, and accelerating the workflow from PDFs and tabulated data to reaction-network analysis and model training. Rapid potential energy surface exploration, ML-driven prediction of activation barriers, and graph-based kinetic analysis were combined into a semi-automated workflow for CO_2 hydrogenation on metal nanoparticles. In this setting, the main benefit of AI-assisted research is the reduction of the technical and computational overhead required to develop the underlying software workflow, expand the reaction network, and perform graph-based analysis of catalytic activity and selectivity. This makes it more practical to account for multiple site types, edge and terrace cooperation, surface diffusion of intermediates, and competing elementary steps within extended reaction networks, rather than restricting the analysis to idealized surfaces and a small set of manually selected pathways.

2.1 Related work

Previous studies have shown that activation barriers on metal surfaces can be estimated with generalized BEP-type or descriptor-augmented ML models without explicit classification by transition state type. Early works in this direction focused on surface reactions of C-, O-, N-, and H-containing species [11-12]. In more recent works, the authors applied ML to C-C and C-H scission barriers in hydrogenolysis reactions of larger hydrocarbons on Ru, Pt, and Rh, showing that ML can outperform conventional BEP relationships and developed an ML model for activation energy prediction across diverse catalytic surfaces in the water gas shift reactions [13-14].

3. Summary

This study shows that extended catalytic reaction networks can be analyzed in a computationally tractable way without sacrificing mechanistic complexity of the underlying reaction network. This framework is now being extended to electrocatalytic CO2 reduction, and this presentation will highlight our recent advances in these areas.

Acknowledgments

We acknowledge the financial support provided by the National Research Foundation Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) program

(Development of Advanced Catalysts for Electrochemical Carbon Abatement, Project Code: 370184872). Computational work was performed using resources of the National Supercomputing Centre, Singapore.

References

[1] W. Liu, H. Zhao, X. Wu, J.-F. Wu, L. Chou, G. Dury, W. Hu, M.V. Polynski, A. Subramanian, S.M. Kozlov. W. Liu, ACS Catalysis (2026), DOI: 10.1021/acscatal.5c07737

[2] Y. Sheng, M.V. Polynski, M.K. Eswaran, B. Zhang, A.M.H. Lim, L. Zhang, J. Jiang, W. Liu, S.M. Kozlov, Appl. Cat. B, 343 (2024) 123550.

[3] S. Nitopi, E. Bertheussen, S.B. Scott, X. Liu, A.K. Engstfeld, S. Horch, B. Seger, I.E.L. Stephens, K. Chan, C. Hahn, J.K. Nørskov, T.F. Jaramillo, I. Chorkendorff, Chem. Rev., 119, 12 (2019) 7610.

[4] D.A. Boiko, M.V. Polynski, Pondering the Future of Chemical Research Amid the Wider Adoption of Artificial Intelligence Technologies, Chem. Eur. J. (2026), DOI: 10.1002/chem.202503630

[5] M.V. Polynski, S.M. Kozlov, EES Catal. (2026), DOI: 10.1039/D5EY00338E

[6] https://github.com/Computational-Nanocatalysis-Group/Reaction_network_exploration_tools

[7] M.V. Polynski, S.M. Kozlov, ChemRxiv (2025), DOI: 10.26434/chemrxiv-2025-5w31t

[8] P.L. Türtscher, M. Reiher, J. Chem. Inf. Model., 63, 1 (2023) 147.

[9] S. Kozuch, WIREs Comput Mol Sci., 2 (2012), 795.

[10] https://ardor.cloud/

[11] F. Goltl, M. Mavrikakis, ChemCatChem, 14 (2022) e202201108.

[12] D.J. Hutton, K.E. Cordes, C. Michel, F. Goltl, J. Chem. Inf. Model., 63 (2023) 6006.

[13] X. Zong, T. Xie, D.G. Vlachos, Appl. Catal. B Environ. Energy, 353 (2024) 124070.

[14] J.L.C. Fajin, A.K. Halder, M.N.D.S. Cordeiro, J. Phys. Chem. C, 129, 11 (2025) 5353-5360.