Machine Learning-Powered Exploration of Catalytic Reaction Networks for CO_2 Conversion into Value-Added Chemicals

Mikhail V. Polynski [a], **Sergey M. Kozlov** [a]

[a] *Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585 ; polynskimikhail@gmail.com (M.V.P.), sergey.kozlov@nus.edu.sg (S.M.K.)*

1. Introduction

The catalytic conversion of CO_2 into valuable chemicals is one of the major current directions in sustainable chemistry. However, the inherent chemical stability of CO_2 and the complexity of the associated catalytic reaction networks present significant challenges. These networks involve a vast number of intermediates and competing reaction pathways, complicating both mechanistic understanding and catalyst design.[1,2] Traditional computational approaches, while providing valuable insight, are often constrained by their inability to systematically explore the full reaction space. The need for a framework capable of handling this reaction pathway diversity has driven the integration of machine learning into reaction network modeling.[3]

ML models, trained on high-throughput DFT calculations, enable the automated exploration of extensive catalytic reaction networks, identifying key intermediates, transition states, and energetic trends with unprecedented efficiency.[4] These models extend beyond simplistic thermodynamic descriptors, facilitating microkinetic modeling that brings computational predictions closer to experimentally relevant conditions.[5] By refining Brønsted-Evans-Polanyi (BEP) relationships and accelerating the identification of kinetically viable reaction pathways, machine learning transforms catalyst screening into a predictive science, vastly improving the rational design of catalysts for selective CO_2 conversion.[6,7] In this presentation, we will address the construction of the microkinetic model and the underlying reaction network using a recently developed machine learning model that predicts activation (free) energies, trained on extensive DFT data.

2. Results and Discussion

We modeled nanostructured catalysts using M_{79} truncated octahedral nanoparticles with the fcc structure and extended (111) facets. Such choice of models is in line with the thermodynamic preference for (111) terminations in late transition metal nanoparticles,[8] observed experimentally. The obtained transition states (TS) were generally located on nanoparticle facets, while certain key steps—such as CO_2 activation and C–O cleavage in methanediolate intermediates—required TS at edge sites due to their distinct electronic and steric environments.

Spin-polarized DFT calculations for the generation of the training and test datasets were performed using the revPBE functional with a 415 eV plane-wave cutoff and the PAW method, including D3 dispersion corrections. We formulated BEP relationships for Cu and Pd nanoparticle systems, clustering elementary steps based on transition state structures rather than conventional reaction categories such as hydrogenation or C–C coupling.

These BEP relationships formed the basis of the trained ML models, where TS structure and location on nanoparticle terraces or edge sites were included as model features alongside metal identity and coordination type. This allowed us to encode the structural and energetic characteristics of transition states systematically, enabling the inclusion of data beyond Cu and Pd. While Cu and Pd datasets encompassed full catalytic reaction networks, the datasets for Ni, Co, Rh, Pt, and Au were limited to key transformations: C–C couplings, C–O cleavages, and selected hydrogenation and protonation reactions, chosen for their mechanistic significance in tuning selectivity and activity.

A baseline one-hot encoding scheme was implemented to categorize reaction class (hydrogenation, protonation, C–C coupling, or C–O cleavage), TS identity (nucleophilic addition to C(sp^2), sp^3-like, hydroxyl or carboxyl acidity, or metathesis-like), and metal identity. After hyperparameter optimization via grid search, kernel ridge regression (KRR) and CatBoost models were trained (Figure 1) with RMSE values for KRR and CatBoost on the test set equal to 0.30 and 0.32 eV, respectively, and with mean absolute error (MAE) of 0.25 eV. Thus, the obtained precision is significantly better than ~0.5 eV precision obtained in previous studies for activation energies with the same type of DFT functionals.[9]

An advanced model was developed to further improve prediction accuracy by incorporating reaction-specific structural descriptors into a neural network (NN). The NN was designed to leverage Coulomb matrices as numerical representations of reactant and product structures. The matrices were processed using a convolutional neural network architecture to extract spatially localized patterns relevant to activation barrier predictions.

A squeeze-and-excitation attention mechanism was implemented, which adaptively reweighted convolutional features based on their learned importance. This mechanism enhanced the model's focus on chemically relevant interactions while suppressing noise, improving predictive stability. The extracted convolutional features were combined with categorical one-hot encoded reaction descriptors, including metal and TS identities.

Cross-validation was employed to evaluate the robustness of the NN model, with model performance evaluated across five folds. The average MAE across the folds was 0.23 eV, indicating prediction stability. The final model was trained on the entire dataset and evaluated on a held-out test set, yielding RMSE = 0.24 eV and MAE = 0.18 eV, demonstrating improved generalizability compared to the conventional KRR and CatBoost models.

The trained models enabled inference predictions of activation barriers across extended catalytic reaction networks, covering 46 surface reactions and 12 adsorption-desorption equilibria on Ni, Co, Rh, Pt, and Au (with the data on Cu and Pd available from DFT computations). This framework facilitated the

Machine Learning-Powered Exploration of Catalytic Reaction Networks for CO_2 Conversion into Value-Added Chemicals

Mikhail V. Polynski[a], **Sergey M. Kozlov**[a]

[a] *Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585 ; polynskimikhail@gmail.com (M.V.P.), sergey.kozlov@nus.edu.sg (S.M.K.)*

construction of microkinetic models that informed our recommendations for multimetallic catalyst design.

2.1 Related work

Previous studies have explored BEP-based ML models for activation energy prediction relying on pre-reaction intermediate structures without explicitly distinguishing transition state types in the involved reactions. Notably, Göltl, Mavrikakis, and co-workers developed BEP-based ML models that categorized reactions based on reactant structures. [6,7]

2.2 Figures and tables

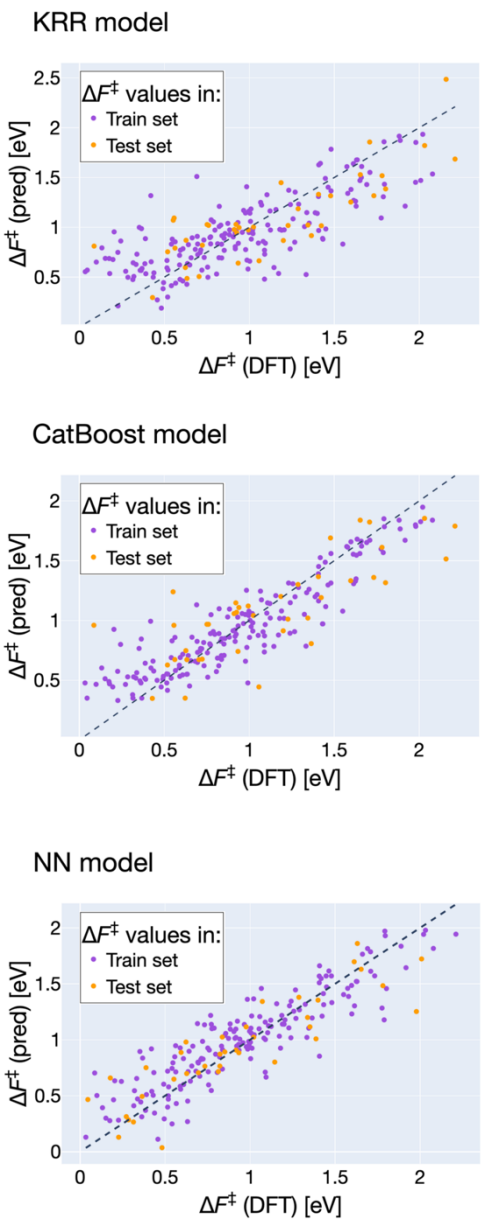

Fig. 1: Parity plots comparing DFT-calculated and ML-predicted activation free energies of various elementary steps in CO_2 hydrogenation to ethanol on transition metal catalysts using KRR, CatBoost, NN models.

Acknowledgments

This work is supported by Agency for Science, Technology and Research (A*STAR) through Low Carbon Energy Research Finding Initiative (LCERFI01–0033 | U2102d2006). Computational work was performed using resources of the National Supercomputing Centre, Singapore.

References

[1] Zhu, J. *et al.* Dynamic structural evolution of iron catalysts involving competitive oxidation and carburization during CO2 hydrogenation. *Sci Adv* 8, 3629 (2022).

[2] Sheng, Y. *et al.* A review of mechanistic insights into CO2 reduction to higher alcohols for rational catalyst design. *Appl Catal B* 343, 123550 (2024).

[3]Morandi, S. *et al.* A Foundational Model for Reaction Networks on Metal Surfaces. (2024) doi:10.26434/CHEMRXIV-2024-BFV3D.

[4] Chen, D., Shang, C. & Liu, Z. P. Machine-learning atomic simulation for heterogeneous catalysis. *npj Computational Materials 2023 9:1* 9, 1–9 (2023).

[5] Ulissi, Z. W., Medford, A. J., Bligaard, T. & Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat Commun* 8, 14621 (2017).

[6] Göltl, F., Mavrikakis, M., Göltl, F. & Mavrikakis, M. Generalized Brønsted-Evans-Polanyi Relationships for Reactions on Metal Surfaces from Machine Learning. *ChemCatChem* 14, e202201108 (2022).

[7] Hutton, D. J., Cordes, K. E., Michel, C. & Göltl, F. Machine Learning-Based Prediction of Activation Energies for Chemical Reactions on Metal Surfaces. *J Chem Inf Model* 63, 6006–6013 (2023).

[8] Vitos, L., Ruban, A. V., Skriver, H. L. & Kollár, J. The surface energy of metals. *Surf Sci* 411, 186–202 (1998).

[9] Goerigk, L. *et al.* A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Physical Chemistry Chemical Physics* 19, 32184–32215 (2017).